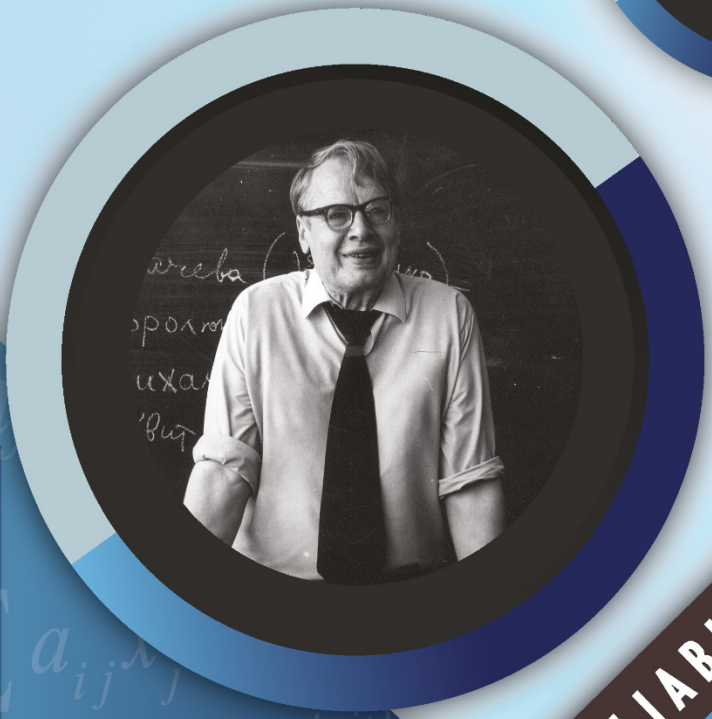
RTA

ISSN 1932-2321

JOURNAL IS REGISTERED IN THE LIBRARY OF THE U.S. CONGRESS

RELIABILITY: THEORY & APPLICATIONS

INTERNATIONAL GROUP ON RELIABILITY



RELIABILITY
RISK ANALYSIS
MAINTENANCE
SAFETY

GNEDENKO FORUM PUBLICATIONS

#3

(74) VOL.18

SEPTEMBER

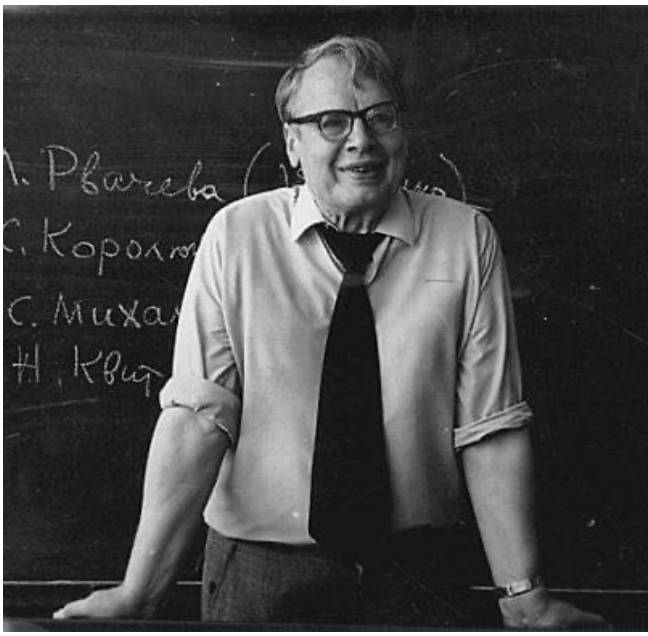
2023

SAN DIEGO

ISSN 1932-2321

http://www.gnedenko.net/Journal/index.htm





RELIABILITY:

THEORY & APPLICATIONS

Vol.18 No.3 (74),
September 2023

San Diego
2023

Editorial Board

e-Journal ***Reliability: Theory & Applications*** publishes papers, reviews, memoirs, and bibliographical materials on Reliability, Quality Control, Safety, Survivability and Maintenance.

Theoretical papers must contain new problems, finger practical applications and should not be overloaded with clumsy formal solutions.

Priority is given to descriptions of case studies.
General requirements for presented papers.

1. Papers must be presented in English in MS Word or LaTeX format.
2. The total volume of the paper (with illustrations) can be up to 15 pages.
3. A presented paper must be spell-checked.
4. For those whose language is not English, we kindly recommend using professional linguistic proofs before sending a paper to the journal.

The manuscripts complying with the scope of journal and accepted by the Editor are registered and sent for external review. The reviewed articles are emailed back to the authors for revision and improvement.

The decision to accept or reject a manuscript is made by the Editor considering the referees' opinion and considering scientific importance and novelty of the presented materials. Manuscripts are published in the author's edition. The Editorial Board are not responsible for possible typos in the original text. The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Publication in this e-Journal is equal to publication in other International scientific journals.

Papers directed by Members of the Editorial Boards are accepted without referring.
The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Send your papers to Alexander Bochkov, e-mail: a.bochkov@gmail.com

Table of Contents

distribution with four parameters is introduced. The mathematical and statistical properties of the proposed distribution, such as the quantile function, moments, moment generating function, survival function, hazard function, odds function, and reversed hazard function, were studied to understand its nature. The probability density function of the order statistics for this distribution was also obtained. The parameters of the model were estimated using the maximum likelihood method of estimation. The proposed model was applied to two real datasets relating to the relief times of twenty patients receiving an analgesic and the sum of skin folds in 202 athletes collected at the Australian Institute of Sports. The results showed that the new model outperformed its comparators and provides better fit than Topp-Leone exponentiated inverse exponential, Topp-Leone inverse exponential, exponentiated inverse exponential, inverse exponential and exponential distributions.

This research work focuses on Bayesian inference in this study to detect a change in the rate of a Maxwell distribution model with independent random variables. The paper specifically analyzes a single rate shift and demonstrates how the Bayesian framework can be used to efficiently solve this problem. To produce samples from Maxwell distribution and evaluate the datasets, simulation techniques were used, and the R programming language was used. Although the model looks to be simple, no analytical solutions are available for parameter inference, necessitating the use of approximations. The study emphasizes the Gibbs sampler's applicability for change-point analysis using a Markovian updating approach. The simulation research findings show that the predicted rate is near to the true value, confirming the consistency and stability of the Bayesian estimator.

Extropy is a complimentary dual of Shannon's entropy, which has many applications. The maximum likelihood and Bayesian approaches are used in this article to explore the weighted extropy and weighted residual extropy of the Pareto type II distribution. Using unified hyper-censoring data, we calculate the maximum likelihood estimation of weighted extropy and its residual measures. Based on symmetric and asymmetric loss functions, Bayesian estimators of weighted extropy and its residual measure are developed based on unified hyper censoring data. To do some complex calculations, Markov chain Monte Carlo methods are used. To test the performance of the estimators, a Monte Carlo simulation study and an illustration using real data sets were carried out. The outcomes of the study showed that as the sample size increases, maximum likelihood and Bayesian estimators of the weighted extropy and its residual measure perform well. Also, Bayesian estimators of the weighted extropy and its residual under the general entropy loss function are superior to the Bayesian estimators under the others in most cases. Theoretical and empirical findings are generally in good agreement.

Fuzzy graph and Fuzzy soft graph are indispensable computing modules for presenting membership and non - membership values in the world of uncertain situations and incidents. In this research article, we introduce the new module of Alternate Quadra Submerging Polar Fuzzy Soft Graph with four co –ordinates with membership and non – membership values. The aim of this new fuzzy soft graph is to find the single output from different uncertain parametric sets of subjects and events, between the range [-1, 1]. The submerge level of fixed four co ordinates is a tool to find the precise and reliable membership degree values from uncertain problems and outcomes. In this artifact, we also investigate the different types of Alternate Quadra Submerging Polar Fuzzy

Soft Graphs, corresponding parametric fuzzy values and submerge membership and non – membership values. We discussed Strong, Complete, Complement and m complement properties of Alternate Quadra Submerging Polar Fuzzy Soft Graphs. We use this fuzzy soft graph in the Analysis of water related diseases to find the result of most and least affected diseases with the symptoms among the hostel students in the same locality. We find the maximum and minimum membership and non - membership value of the water related diseases in an unique way by using this Alternate Quadra Submerging Polar Fuzzy Soft Graph score function values.

Bilal Ahmad Bhat, M. Sultan Shah, M.A.K Baig

Recently, the measurement of uncertainty has attracted the attention of researchers. In this article, we introduce a new weighted uncertainty measure known as weighted generalized entropy. We also study its dynamic (residual) version which is known as weighted generalized residual entropy. These are length-biased shift-dependent uncertainty measures. It is shown that the proposed dynamic uncertainty measure uniquely determines the survival function. The various significant properties and the relationship with other well-known reliability measures of the proposed dynamic uncertainty measure are also studied. Finally, a real life data set is used to illustrate the usefulness of our proposed uncertainty measures.

Binay Kumar

In the present investigation, we consider a bulk queue model with the assumption that the server may stop working due to random failure during any stage of the service. As soon as the server fails, it is immediately sent for repair. The server offers all incoming units the first mandatory service and any one of the optional services as per the unit's requirements. For computation purposes, we assume that the server offers m+1 services, of which the first one is essential and the remaining are optional. The server may take a vacation in accordance with the Bernoulli vacation schedule with probability p as soon as both service phases of a unit are completed. As the system empties, the server idles and needs some time to set up before initiating the next service. In order to analyse the model and derive various steady-state queue length distributions, we incorporated the supplementary variables corresponding to service time, vacation time, and repair time and applied the probability generating function technique to determine the various system state distributions. Using these probability distributions, we derive the explicit form of various performance indices. To discuss the validity of the present model, we obtained some well-known results from the queueing literature as a special case of the present model by setting appropriate parameters. Finally, to analyse the sensitivity of several performance indices, a numerical demonstration is provided.

K.M. Sakthivel, Vidhya G

In this research, we present an approach to model lifetime data by a weighted three-parameter probability distribution utilizing the exponential and gamma distributions. We have presented some of the essential characteristics such as the shapes of pdf, cdf, moments, incomplete moments, survival function, hazard function, mean residual life, stochastic ordering, and order statistics of the proposed distribution. Furthermore, we also presented the Bonferroni index and Lorenz curve of the proposed distribution. The maximum likelihood approach is used to estimate the parameters of the distribution. Finally, the proposed probability distribution is compared to goodness of fit with Lindley, Akash, exponential, two-parameter Lindley, cubic transmuted Rayleigh, and Exponential-Gamma distributions for the real-time data set.

varieties of signs accepted in the classifiers without control of expediency. The causes of errors arising in the evaluation of Kolmogorov statistics as the largest in absolute deviation are analyzed the deviation between statistical distribution functions of the population and sample using simulation modeling, fiducially intervals and the theory of testing statistical hypotheses. These erroneous calculations with a small number and multidimensionality of sampling implementations double increases of the Type II Error. Finally, the result showed the advantages of the new method in comparison with Kolmogorov' criterion via checking representativeness of sample.

B. Srinivasa Rao, M. Rami Reddy, K. Rosaiah

In this article an attribute control chart is designed for the Exponentiated Inverse kumaraswamy distribution under a time truncated life test by assuming the life-time of the item follow the selected Exponentiated Inverse Kumaraswamy distribution with known parameters. In order to limit the cost of checking the quality of an item in any industrial process with time truncation, this process is much useful. By considering the average number of defective items from a specified lot that are failed before the time limit, the attribute control limits are constructed. The control chart is determined using Binomial distribution based on the Upper and Lower control limits. The functioning of the designed control chart is examined with the average run length (ARL) values. The control chart constants and limits are calculated for specific ARL values with assumed parameters at different sample sizes for an in-control process. These control chart constants are obtained by considering different combinations of parameters of the assumed distribution. With these in-control limits the ARL values are observed by shifting the parameter values. A simulation analysis is developed by taking a specific number of observations in each sample and the average number of failures from each sample is considered as a statistic to establish the execution of the control chart for a specified ARL at a particular shift in parameter. With that statistic of average number of failures from the samples the control chart is prepared. It is observed a specific change in defective number when there is shift in parameter values. The results are illustrated with an example.

Olalekan Akanji Bello, Sani Ibrahim Doguwa, Abukakar Yahaya, Haruna Mohammed Jibril

In the area of distribution theory, statisticians have proposed and developed new models for generalizing the existing ones so as to make them more flexible and to aid their application in a variety of fields. In this article, we present a new distribution called the Type I Half-Logistic Exponentiated Weibull (TIHLEtW) Distribution with four positive parameters, which extends the Weibull distribution by two parameters. Some statistical properties of the TIHLEtW distribution, such as explicit expressions for the quantile function, probability weighted moments, moments, generating function, Reliability function, hazard function, and order statistics are discussed. A maximum likelihood estimation technique is employed to estimate the model parameters and the simulation study is presented. The superiority of the new distribution is illustrated with an application to two real data sets. The results showed that the new distribution fits better in the two real data sets amongst the range of distributions considered.

Maryam Gholami Arjenaki, Dr. Mahdi Karbasian, Amin Kazemi Manesh, Mohammadreza Jafari

Aim. Compressed air vessels are responsible for injecting compressed air to the mechanical flying device. It should

be noted that the pressure level inside these vessels is very important in conducting all operational stages successfully; therefore, it is of high significance to be assured of the quality of the vessels being used. This study was done in 2017 in order to calculate and estimate the reliability of compressed air vessels in mechanical flying device system with proposing a potential failures assessment and accelerated test approach taking into consideration the current methods. Methods. The paper uses methods of Fault Tree Analysis, Failure Mode & Effect Analysis and Accelerator tests. Initially, the interactions among the components were identified using the Design Structure Matrix in order to design a matrix to help estimate the reliability; consequently, improve equipment performance. Next, failures root recognition was done using Fault Tree Analysis diagram, then, failures reasons prioritization was done using Failure Mode & Effects Analysis tables. Accelerator tests were designed and applied on failure mechanisms such as leakage by pressure on vessels, corrosions on steal head, nipples, O-ring creeping, O-ring ozone cracking and liner chemical degradation. After that, the average of failure rates was calculated within the taking after stage for each test. Within the conclusion, the result of failure rates from the accelerator tests was compared with the result of failure rates from the process approach. Consequently, the most elevated amount of these two approaches was defined as the total failure rate; product reliability and lifetime were calculated utilizing this amount. Results. The following finding were obtained using the proposed methods. When Windy Liner was used, vessel lifetime was six years and half and vessel reliability in ten years was 0/22. Whereas, when Rotational Liner was used, vessel lifetime was eight years and three months and vessel reliability in ten years was 0/3. Conclusion.The approach proposed in the paper allows accelerator degradation test can also be used instead of accelerated test in order to calculate the reliability of failure mechanisms. In the event that high-quality and legitimate O-rings are used, vessel reliability and lifetime can be increased.

The paper introduces a new type of fractional integral transform called the N-transform of fractional order. This transform is utilized to derive various results for a more generalized function of fractional calculus known as the Aleph-function. The authors present several useful findings and explore the relationship between the N-transform and other existing fractional transforms. Additionally, the paper discusses the relationship between the N-transform of fractional order and other existing fractional transforms. It likely explores how this new transform relates to established transforms in fractional calculus. The authors have also examined special cases or specific examples to further illustrate the applications and properties of the N-transform of fractional order. These cases could involve particular functions or parameter values that offer insight into the behavior of the transform.

In this article, we introduce a new member to Poisson-X family namely, the Poisson-power Lindley distribution. The statistical as well as the distributional properties of the new distribution are studied. The flexibility of the distribution is illustrated by means of real data sets. We also introduce a reliability test plan for acceptance or rejection of a lot of products submitted for inspection when lifetimes follow the new distribution. The minimum sample size using binomial and Poisson approximations, operating characteristic values and minimum ratios of the true value and the required value of the parameter with a given producer's risk are also developed with respect to the newly introduced sampling plans. A real data example is also given to illustrate the sampling plan developed.

The production stage of the manufacturing process contains numerous subsystems, and the failure of one might have an impact on the entire system. Thus, a manufacturing plant needs to be reliable and well-maintained. This paper examines the profitability and reliability of a production plant for utensils while taking the effect of temperature into account. The plant processes raw materials through several subsystems in series including cutting, pressing, spinning, and polishing & packing. Winter production requires a significant amount of heat which could damage the machinery. As a result, production is low and preventive maintenance is carried out during the winter. For both the summer and winter seasons, many system measures have been assessed. The time distributions have been assumed to be exponential. The model has been analysed using the Markov and Regenerative processes. The production fluctuation between the summer and winter seasons have been illustrated using a numerical example with specific values for the parameters.

Queueing models in which the services are provided in groups (or blocks or batches) have found to be very useful in real-world applications and such queues been extensively analysed in the literature. In this paper we see one such group service queueing model with balking, reneging and immediate feedback. The arrival processes is a Markovian arrival , where, the arriving customer may balk the system while the server is idle and the pool is empty. Customers are provided service in groups of varying size from 1 to the fixed constant, say, N. The service time of a batch follows the phase type distribution corresponding to the each size of the group. A group's service time is taken as the highest of the service times of each customers who make up the group. The group of customers who are dissatisfied with the service then that group will get the service immediately. Here, the feedback of a group is defined as the average of the feedback of each customers who make up the group. During the admission period the customers may renege. We calculated the steady state probabilities by using the matrix geometric method, then, by using it we computed few performance measures. We have studied the busy period and the distribution of waiting time is derived. Results are illustrated with some graphical representations.

In factorial experiments, treatment combinations increase as the number of factors increases. While handling a large number of factors, many difficulties are encountered. Moreover, mechanical errors like mistaken identification of plots, wrong labeling of treatments, etc., may creep in. To overcome these difficulties, only a fraction of treatment combinations can be tested. This technique is known as fractional replication. The design with fractional replication is known as fractional factorial design (FFD). In FFD, the choice of the fraction of treatment depends on what type of information is sacrificed. Usually, the interactions with higher-order are omitted, and all main effects and two-factor interactions are estimated without loss of information. The procedure for the layout of FFD is closely related to the concept of confounding. The analysis of fractional factorials is similar to the analysis of full factors. FFD is used to reduce treatment combinations by a fraction. FFD plays a significant role when the experiment is too large. When compared to classical designs, FFD yields a cost-benefit relationship. Fuzzy theory is used to deal with the imprecise observations in this design. This paper proposes the statistical analysis of fuzzy fractional factorial design with numerical illustration.

function as a generator. We referred to the proposed family as the continuous Bernoulli-generated family of distributions. Basic statistical treatments of the proposed family such as the density and cumulative distribution functions, survival and hazard rate functions, quantile, moments, moment generating function, and Renyi entropy are derived. The method of maximum likelihood is employed to estimate the unknown parameters of the family and the asymptotic behaviour of the parameter estimates is investigated via Monte Carlo simulation study. The waiting time (in minutes) of 100 Bank customers and the tensile strength measured in GPa, of 69 carbon fibers data sets formed the basis for real-life data fittings. Results obtained from the fitting of the two data sets when compared with some existing non-nested models revealed that the fittings were in favor of the continuous-Bernoulli Weibull distribution over the rest competing distributions.

Ayed AL e'damat, Khalid Ul Islam Rather

Utilizing auxiliary information effectively in sample surveys can enhance the accuracy of estimations by capitalizing on the relationship between the main variable under study and the auxiliary variable. Estimators such as ratio, product, exponential, and regression estimators are frequently employed either during the estimation process, the design phase, or both. In everyday situations, it is common to incorporate information from one or two auxiliary variables to improve the precision of estimators. Auxiliary information has been in practice in sampling theory since the advent of modern sample surveys. Information on auxiliary variable having high correlation with the variable under study is quite useful in improving the sampling design. Cochran (1940) used the highly positively correlated study and auxiliary variable to propound the ratio estimator. Product estimator requires a high negative correlation between study and auxiliary variable. By reviewing the literature, it is concluded that applying the auxiliary information enhances the efficiencies of the estimators for estimating any parameter under consideration. So it is well established fact that the use of auxiliary variable technique improves the estimation process for target population. It is also noticed that ratio method of estimation is relatively simple and one of the commonly used methods of estimation. Due to limitations in terms of time and cost, sample surveys are often preferred over census surveys for collecting primary data. In these sample surveys, the ratio, product, and regression estimators are frequently employed to estimate the mean or other parameters of interest for the variable under study. To assess their efficiency, these estimators are compared based on their approximate mean squared errors. In this paper we proposed an exponential ratio type estimator for the estimation of finite population mean under systematic sampling. The mean square error of the proposed estimator is computed up to the first order of approximation and we find proposed estimator is efficient as compared to other existing estimators. Furthermore this theoretical result is supported by numerical examples as well.

Sofiene Fenina, Souheyl Jendoubi, Faker Bouchoucha,

This paper presented a study of the failure rate by introducing the effect of influencing variables. These variables have a random effect which depends on the external environment of the system. There are a multitude of variables and their modeling is difficult. The perturbation, to the failure rate, caused by external factors, has a direct impact on the time scale by the acceleration (or deceleration) of the degradation of the system. Therefore, the adopted methodology consists in introducing a perturbation on the Weibull parameters and studying its effect on the failure rate. Weibull parameters are considered random variables with a Gaussian distribution. The failure rate formulation in a random environment is offered through Weibull distribution. A case study of the hemodialysis machine is offered to illustrate the proposed approach and validate the results. The simulations presented show the failure rate statistics for different configurations of the Weibull distribution. The validation of the results was done using Monte Carlo simulations.

Principal Component Analysis and Cluster Analysis for feature extraction and finally a Generalized Linear model for prediction. Basic Tables and descriptive plots, Principal Component Analysis, Factor Analysis, and Generalized Linear Models were employed. Results: In the Asejire and Eleyele reservoirs, respectively, the PCA yields 5 significant main components explaining 76.56% and 60.97% of the variance, while the FA yields 5 significant major components explaining 94.90% and 79.97%. A generalized linear model (GLM) was used to study the variability in turbidity level, and the results indicate that two parameters "Iron and Silicon" in the Asejire reservoir are crucial for understanding turbidity variation and four "Colour, Alkaline, Silica, and Solids "contribute significantly to turbidity in the water level in the Eleyele Reservoir. With the exception of dissolved oxygen from either reservoir (Eleyele or Asejire) and iron from Eleyele Reservoir, many metrics in Asejire are within SON and WHO acceptable limits. This suggests that the water in the Eleyele reservoir is more contaminated than the Asejire reservoir.

Hosenur Rahman Prodhani, Rama Shanker

In this paper some important statistical properties of three-parameter Sujatha distribution including descriptive measures based on moments, reliability properties, mean deviations, stochastic ordering and Bonferroni and Lorenz curves have been discussed. The estimation of parameters using maximum likelihood estimation has been discussed. Finally, the goodness of fit of the distribution has been presented for two real lifetime datasets and compared with several one and two-parameter well- known lifetime distributions.

G. Ayyappan, G. Archana @ Gurulakshmi

A single server retrial queueing model with non-preemptive priority was examined in this research. The arrival of priority consumers follow a marked Markovian arrival pattern, and both high priority and low priority service times are according to phase type distribution. Matrix analytic method are used to examine the steady state analysis of this model. Various system performance measures, cost analysis and busy period analysis also examined in this model. In additionally, by using some system performance measures we provide the numerical illustration with numerically and graphically.

Dr. V.Kaviyarasu, A.Nagarajan

Statistical Quality Control is an important field in production and maintenance of quality product in manufacturing environments. Reliability sampling plans (RSP) were widely employed in the sectors of manufacturing to monitor the quality of products in order to safe guard the producer as well as the consumer also the experimental costs and time can be saved. This article is developed on the reliability sampling plan when the evaluating life of the product is set to be truncated at pre-determined time follows Exponential-Poisson (EP) distribution. The probability of acceptance criteria for the single sampling is designed to achieve the lowest sample size for such proposed two parameter probability distribution with the corresponding decision rule. This study is conducted to design plan parameters on the basis of desired quality levels such as Acceptable Reliability Quality Level (ARQL), Indifference Reliability Quality Level (IRQL) and Rejectable Reliability Quality Level (RRQL). This study computes the median life for the specified producer's risk, its OC curve is provided along with the

minimum ratio values. Furthermore, it determines the minimum size of the samples and the acceptance number. Table values have been obtained and provided for single sampling plan. Additionally, suitable examples are provided to conduct a study on a real time situations.

The authors of this study set out to build a software reliability growth model (SRGM). Software reliability is a crucial attribute that has to be quantified and evaluated. In most cases, software errors happen at unpredictable times. In this article, the failure intensity of the single parameter length-biased exponential class SRGM has been characterized taking into account the Poisson process of the incidence of software faults. The parameters of the proposed SRGM under investigation are the scale parameter ($\theta 1$) and the total number of failures ($\theta 0$). It is considered that the experimenter may have previous knowledge of the parameters from past or earlier experiences in the form of gamma priors. The posterior probability may be obtained by combining the prior probability with the likelihood of the data, and Bayes estimators can then be suggested.

In this paper a power weighted Sujatha distribution, which includes power Sujatha distribution, weighted Sujatha distribution and Sujatha distribution as particular cases, has been proposed. Its statistical properties including behavior of density function, moments, hazard rate function, and mean residual life function have been discussed. Estimation of parameters has been discussed using the method of maximum likelihood. A simulation study has been presented to know the performance of maximum likelihood estimates of parameters. Application of the proposed distribution have been explained with a real lifetime data relating to patients suffering from head and neck cancer and goodness of fit shows quite satisfactory fit.

This article introduces a new four-parameter probability distribution called the q-Exponential-Weibull distribution based on the q-Exponential-G family of distribution. The proposed new distribution has to decrease and increase failure rates which are more common in reliability scenarios and can be used instead of Weibull and the exponential distribution. It also includes some sub-models like q-Exponential-Exponential, q-Exponential-Rayleigh, Exponential-Weibull, Exponential-Exponential and Exponential-Rayleigh lifetime distributions. Various Mathematical and statistical Properties are investigated, which include Limiting behavior, Moments and Moment Generating functions, Quantile function and Order Statistics. The Maximum Likelihood estimator is used for estimating the model parameters. This new distribution is compared with other lifetime distributions using different kinds of real-life failure time data.

significance. This paper presents and illustrates a novel fuzzy triangular number-based approach to fuzzy CRD analysis. This paper also explores how flexible a CRD may be when handling uncertain elements. This study provides an example of a new method for fuzzy CRD analysis employing TFN.

The aim of this paper is to develop a probabilistic model for a cold standby system that consists of an imperfect switching device and a servicing facility. The model aims to address the issue of unexpected random failures of the switch by implementing preventive maintenance measures. The system has two identical units. It starts with one unit in active operation and another unit in cold standby mode. In standby mode the unit remains in perfect state. No failure is allowed in standby mode. As the operating unit fails, the standby unit needs to be switched into operation, to keep the system working. A servicing facility is present in the system to perform necessary servicing related tasks. The servicing facility referred to as the server, also takes care of all necessary remedial activities like preventive maintenance and repairs. The switch used as switching mechanism to place the standby unit into operation may found imperfect when needed. Similarly, the server too can fail while doing job. A preventive maintenance scheme is used for the switch whereas treatment is given to server. The method of semi-Markov process and regenerative point technique is used for model developing and solving, respectively. The expressions are derived to determine different system performance measures such as mean time to system failure, availability, busy period, expected number of preventive maintenances and the profit. The distributions of random time elapsed in repairs, replacements, preventive maintenances and treatments are general. This study highlights the usefulness of switch's preventive maintenance in long run. To study the asymptotic behavior of the system model, all the expressions for system performance measures are obtained in steady state. A simulation study is conducted using a presumed data set and assuming a Weibull probability distribution. The numerical results are shown in tabular form. The simulation results serve to highlight the significance of preventive maintenance for the switch. The findings of the paper can provide guidelines to the people engaged in designing, framing and implementing standby switching systems in real applications.

Bootstrap method was initially used to determine accuracy measures for sample estimates of independent and identical distributions (i.i.d.). In order to apply bootstrap method to time-dependent data, blocking technique is introduced to preserve serial correlation of the original time series data. In the past, resampling techniques for time-dependent data were implemented using Non-overlapping Block Bootstrap (NBB) method but its dichotomous block arrangement restricts the number of blocks. As a result, improvement becomes necessary. Although the Moving Block Bootstrap (MBB) method improves upon NBB with regard to many more blocks, it introduces an uneven representation of the time series elements which eventually influences its accuracy. In this paper, an innovative method called Moving Block Bootstrap Method with better element Representation (MBBR) is developed to ensure that the time series elements within the block are better represented with minimum number of elements. To compare MBB and MBBR, simulated studies were carried out on some set time series data following each classes of Autoregressive Moving Average (ARMA) model with different parameters, sample sizes and standard deviation using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Results show that by improving the representation of time series data in the blocking arrangement, the accuracy of the proposed method (MBBR) consistently outperforms the existing one (MBB) and thus, provides more efficient estimates of the dependent variable.

cut function of a particular form; it has some ambiguity or imprecision. The estimated numbers are very similar to the actual ones. This approach may generally be used for any real-time triangle fuzzy number calculation. In this proposed methodology, it is obvious that if the value of the observed fuzzy test statistics is similar to real numbers in the testing crisp hypotheses, then fuzzy RBD is very sensitive for making the determinations as to whether to accept or reject the fuzzy null hypotheses and also debates the application of the method for example.

This article is dedicated to the study of the new class of distributions and one of its particular members. Based on the ratio of CDF G(x) and 1+G(x) of the baseline distribution, we have developed the new trigonometric family of distributions by transforming the sine function, and we named it the new class sin-G (NCS-G) family of distributions. The general properties of the suggested family of distributions are provided. Using the inverted Weibull distribution as a baseline distribution, we have introduced a member of the suggested family having a reverse-j or increasing, or inverted bathtub-shaped hazard function. Some statistical properties of this NCS-IW distribution are explored. The associated parameters of the new distribution are estimated through the MLE method. To assess the estimation procedure, we conducted a Monte Carlo simulation and found that even for small samples, biases and mean square errors decreased as the size of the sample increased. Two real medical data sets are considered for the application of the NCS-IW distribution. Using some criteria for model selection and goodness of fit test statistics, we empirically proved that the suggested model performs better than six other existing models (most of which have more parameters).

The purpose of the work is to build, on the basis of multi-component and multi-phase models of queuing systems (QS), mathematical models of maintenance and repair of vehicles by repair enterprises to increase the efficiency of their use. Results. The article considers multi-component and multi-stage mathematical models of QSs with the distribution of the arrival flow simultaneously between the system components, which consist of a certain number of service channels and waiting places in the queue. The same service channels can have different performance depending on the type of customers which they serve. Customers go through several stages of service and waiting. Considered are service of customers without a lack of time to stay in the service channel and waiting and with a lack of such time. The service process in the QS of each component consists of several (ke) stages with the corresponding duration, the full-service period will be equal to the sum of such time intervals. Stage durations have certain probability distributions with appropriate parameters, then the total duration of the service process will have a generalized Erlang distribution with parameters of probability distributions of stages of order ke. The number of components and their parameters correspond to the similar characteristics of the production divisions of the repair enterprise. The study of the effectiveness of the repair enterprise operation as a multi-component and multi-stage QS consists in determining the probability of service and the probability of failure of QS components and the system as a whole, the number of service channels, the number of customers in components, the number of customers in component queues, the duration of maintenance of customers in components and the system, the duration of being customers in queues of components and QSs, duration of stay of requirements in QSs and duration of customer waiting in QS queues. The model is implemented using Any Logic University Researcher. The AnyLogic University Researcher development environment allowed to combine the principles of system dynamics with the paradigms of agent and discrete-event modeling. In addition, thanks to the built-in Java SE compiler, a library of ready-made solutions is available, including generators of random variables, which significantly expands the possibilities of developing and implementing experiments. In particular, experiments on optimization (relative to a defined criterion), sensitivity of the model, stability of the model, etc. are available.

Analysis of Retrial Queue with Two Way Communication, Working Breakdown and Collision

B. Somasundaram
•
Department of Mathematics,
Vel Tech Rangarajan Dr. Sagunthala R&D Institution of Science and Technology,
Tamilnadu, India
somu.b92@gmail.com

G. Archana @ Gurulakshmi
•
Department of Mathematics,
Puducherry Technological University, Puducherry, India
archanagurulakshmi@gmail.com

G. Ayyappan
•
Department of Mathematics,
Puducherry Technological University, Puducherry, India
ayyappanpec@hotmail.com

Abstract

This study examines a two-way communication retrial queue with collision and working breakdown. Retrial incoming calls may interfere with service if the server is busy with primary incoming calls. Arriving primary incoming calls are sent to the orbit while the server is busy or enter for service if the server is discovered to be idle. The server places calls while it is idle. Incoming calls are given high priority, and outgoing calls are given low priority. During regular service, the system may fails at any time. The server will continue to provide service after a breakdown even though at a slower rate rather than shutting down completely. We assume that outgoing calls and service time distributions are modelled in terms of various PH distributions, while incoming calls are assumed to follow a Markovian arrival process (MAP). The matrix-analytic approach will be used to investigate the resulting QBD process in steady state. Some of the performance metrics' calculations have been figured out. At the end, the results are presented in both numerical and graphical form.

Keywords: Retrial queues, Priority queue, Two way communication, Working breakdown, Collision.

1. Introduction

This study examines a queueing model of the form $MAP, PH/PH, PH/1$ beneficial in the call centre service sector. If the server is idle and a call comes on connectivity issues, a change in the network, etc., the customer will receive service right away. If not, the server will enter orbit and

the customer will have to attempt again afterward. The server may also call customers to inform them of the special offers during idle period.

Neuts [21] made a significant contribution to stochastic process theory.He created the adaptable Markovian point process, which made it feasible to create the Markovian arrival process (MAP) and batch Markovian arrival process ($BMAP$). The two arrival mechanisms were further extended by Lucantoni [19]. The most important MAP properties are those that allow for a matrix analytic solution to a stochastic model. [7] has also enhanced this useful tool to make it a more understandable in the encyclopaedia of operations research and management science. There are definitions for MAP's discrete and continuous situations. The parameters for MAP are D_0 and D_1 of dimension "m" in continuous time, where $D_0$is a non-singular stable matrix governing the transition corresponding to no arrival and D_1 regulates the transition corresponding to arrival. Take into consideration the generating matrix D, which is represented by the equation $D = D_0 + D_1$.

Retrial queues are given a lot of attention in recent years because they can be used to analyse the operation of many different systems, including telephone, call centres, computer networks, and communications systems. When customers who arrive but are unable to receive service enter the orbit and execute later, arbitrary time. The majority of queuing systems offer customers the option to leave the service area temporarily and join an orbit before returning to complete their request after a set amount of time. A customer's orbit is the time in between trials. The server solely serves incoming arrivals made by regular customers, according to the most of literature on retrials queues. Then there are real situations where servers can make outgoing phone calls while not conversing, simulating, for example, two-way communication.

Artalejo and Gomez-Correl [1] have analyzed the comparison of classical and retrial queues as well as advance retrial queues. Chakravarthy and Dudin [9] have examined a retrial queuing system in single server which are two types of customers arrives according to Markovian arrival process and service times follows exponential distribution. A retrial queues in finding a various kind of important problems has investigated by Falin and Templeton [12]. Ke and Chang [14] has examined the multi-server retrial queuing system with vacation and balking.

Artalejo and Phung-Duc [2] have examined single server $M/G/1$ retrial queues with two-way communication. During their analysis of the $M/G/1$ retrial queue, they found that the inward and outward service time distributions changed. An $M/G/1$-type mixed priority retrial queue with two-way communication, Bernoulli vacation, and discussions of collisions, working breakdown, negative arrival, repair, immediate feedback, and reneging was described by Ayyappan and Udayageetha [6]. The $MAP, PH_2^O/PH_1^I, PH_2^O/1$ retrial queue with vacation, feedback, two-way communication, and dissatisfied customers has been investigated by Ayyappan and Gowthami [4]. [20] . Krishnamoorthy et al. [16] has been investigated the multi server queueing system that has infinite capacity in which waiting customers generate into priority and non-preemptive service discipline.

The $M/M/1$ retrial queue with two-way communication and exponential service time distributions of ingoing and outgoing calls has been thoroughly examined by Artalejo and Phung-Duc [3], who also expanded their analysis to two-way communication for multi-server retrial queueing models. A multi-server queuing model with Markovian arrivals and multiple thresholds was researched by Chakravarthy [8]. A multi-server retrial queue with two types of customers arriving in accordance with MAP with type 1 customers having preemptive priority over type 2 customers was investigated by Kumar et al, [17]. A queueing model for automatic teller machines was created by Chakravarthy and Subramanian, [10]. These service systems are subject to failure because of catastrophic events. In this study, the effectiveness and accessibility of an ATM system during failures, repairs, and replacement were examined. Jeganathan [13] investigated a $M_1, M_2/M/1$ retrial inventory system with non-preemptive priority service.

Ayyappan and Thilagavathy [5] have investigated the $MAP/PH/1$ queueing model, consisting of setup, closedown, multiple vacations, standby servers, breakdown, repair, and reneging. They assume two different types of servers, especially main and standby server. For their model, standby server carries over the service at lower rates than main server. A queueing model with

server breakdowns, repairs, vacations and backup server has been examined by Chakravarthy et al [11]. In their model, used two types of server, one is main server and another one is backup server, customers arrive as MAP. Suganya et al [22] explored a perishable inventory system with a limited waiting capacity. A multi-server retrial queue with batch Markovian arrival process and breakdowns have examined by Kim et al [15]. They considered the flow of breakdowns according to MAP and the repair period has PH type distributions.

2. Model Description

A single server retrial queueing model with two different categories of arrivals is explored. Some of the arrival as incoming calls that comes after a MAP which contains the parameters of matrices D_0 and D_1 in the order m_1. The server that places the outgoing calls follows the PH type distribution (α, R) in order m_2 while it is idle. At the arriving incoming calls get the service immediately, if the server is available. Whenever the server is busy, those calls are enter into the orbit. Both incoming and outgoing call service times are distributed corresponding to the PH type with (β, S) and (γ, T) in the positions n_1 and n_2, respectively. The server may struck with breakdown during the busy period with rate η. When the service is broken down, customers using the server will pay a lower rate (θ) for the service. The server will instantly begin the repair operation with the ξ parameter. If server is available, an arriving retrial customer receives the service immediately with probability π. When the server is busy, the arriving retrial customer which collides the current customer who are getting service and the customers enter into the orbit with probability p.

3. The Steady-State Analysis

The queueing model's steady-state analysis is discussed in this section. We need to start with a few notations. Let $N(t)$ denote number of customers in the orbit, $Y_1(t)$ denoted working mode of the server (if $Y_1(t) = 0$ normal working mode and if $Y(t) = 1$ slower working mode), $Y_2(t)$ denoted status of the server (if $Y_2(t) = 0$ server is idle, $Y_2(t) = 1$- server busy with incoming call, $Y_2(t) = 2$- server busy with outgoing call), $S_1(t)$ and $S_2(t)$ represented for the phases of service when the server was busy with incoming calls and outgoing calls, respectively. $A_1(t)$ and $A_2(t)$ represented for the phases of arrival of incoming calls and outgoing calls, respectively. A continuous-time Markov chain ($CTMC$) is applied in the process $\{N(t), Y_1(t), Y_2(t), S_1(t), S_2(t), A_1(t), A_2(t)\}$, where the state space is provided by

$$\begin{aligned}\Omega = \{&(i,j,0,0,0,a_1,a_2):\ i\epsilon\mathbb{Z}^+,\ j=0 \text{ or } 1,\ 1\le a_1\le m_1,\ 1\le a_2\le m_2\}\\ \bigcup\{&(i,j,1,s_1,0,a_1,0):\ i\epsilon\mathbb{Z}^+,\ j=0 \text{ or } 1,\ 1\le s_1\le n_1,\ 1\le a_1\le m_1\}\\ \bigcup\{&(i,j,2,0,s_2,a_1,0):\ i\epsilon\mathbb{Z}^+,\ j=0 \text{ or } 1,\ 1\le s_2\le n_2,\ 1\le a_1\le m_1\}\end{aligned}$$

3.1. The Infinitesimal Generator Matrix

The infinitesimal generator matrix Q has the following structure of level dependent quasi birth-and-death ($LDQBD$).

$$Q=\begin{bmatrix} A_{00} & A_{01} & 0 & 0 & 0 & 0 & \dots\\ A_{10} & B_1 & B_0 & 0 & 0 & 0 & \dots\\ 0 & B_2 & B_1 & B_0 & 0 & 0 & \dots\\ 0 & 0 & B_2 & B_1 & B_0 & 0 & \dots\\ \vdots & \vdots & 0 & \ddots & \ddots & \ddots & \dots \end{bmatrix}$$

where the (block) matrices appearing in Q are as follows.

$$A_{00} = \begin{bmatrix} R \bigoplus D_0 & e_{m_2} \bigotimes \beta \bigotimes D_1 & R_0 \bigotimes \gamma \bigotimes I & 0 & 0 & 0 \\ \alpha \bigotimes S_0 \bigotimes I & (S \bigoplus D_0) - \eta I & 0 & 0 & \eta I & 0 \\ \alpha \bigotimes T_0 \bigotimes I & 0 & (T \bigoplus D_0) - \eta I & 0 & 0 & \eta I \\ \zeta I & 0 & 0 & (R \bigoplus D_0) - \zeta I & e_{m_2} \bigotimes \beta \bigotimes D_1 & R_0 \bigotimes \gamma \bigotimes I \\ 0 & \zeta I & 0 & \alpha \bigotimes (\theta S_0) \bigotimes I & \theta S \bigoplus D_0 - \zeta I & 0 \\ 0 & 0 & \zeta I & \alpha \bigotimes \theta T_0 \bigotimes I & 0 & \theta T \bigoplus D_0 - \zeta I \end{bmatrix}$$

$$A_{01} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I \bigotimes D_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & I \bigotimes D_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I \bigotimes D_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & I \bigotimes D_1 \end{bmatrix}$$

$$A_{10} = \begin{bmatrix} 0 & (1-p)\pi(e_{m_2} \bigotimes \beta \bigotimes I) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-p)\pi(e_{m_2} \bigotimes \beta \bigotimes I) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} (R \bigoplus D_0) - (\pi I) & e_{m_2} \bigotimes \beta \bigotimes D_1 & R_0 \bigotimes \gamma \bigotimes I & 0 & 0 & 0 \\ \alpha \bigotimes S_0 \bigotimes I & (S \bigoplus D_0) - \eta I & 0 & 0 & \eta I & 0 \\ \alpha \bigotimes T_0 \bigotimes I & 0 & (T \bigoplus D_0) - \eta I & 0 & 0 & \eta I \\ \zeta I & 0 & 0 & (R \bigoplus D_0) - (\zeta + \pi) I & e_{m_2} \bigotimes \beta \bigotimes D_1 & R_0 \bigotimes \gamma \bigotimes I \\ 0 & \zeta I & 0 & \alpha \bigotimes (\theta S_0) \bigotimes I & \theta S \bigoplus D_0 - \zeta I & 0 \\ 0 & 0 & \zeta I & \alpha \bigotimes \theta T_0 \bigotimes I & 0 & \theta T \bigoplus D_0 - \zeta I \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ p\pi(e_{n_1} \bigotimes \alpha \bigotimes I) & I \bigotimes D_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & I \bigotimes D_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p\pi(e_{n_1} \bigotimes \alpha \bigotimes I) & I \bigotimes D_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & I \bigotimes D_1 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & (1-p)\pi(e_{m_2} \bigotimes \beta \bigotimes I) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-p)\pi(e_{m_2} \bigotimes \beta \bigotimes I) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The boundary blocks B_1 and B_2 are of order $2mn$. A_0, A_1 and A_2 are square matrices of order mn.

4. System Analysis

4.1. Stability Condition

Using the definitions $B = B_0 + B_1 + B_2$ and δ as the steady-state probability vector of the irreducible matrix B, it can be confirmed that the vector δ satisfies

$$\delta B = 0, \quad \delta e = 1$$

The vector δ, partitioned as $delta = (\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5)$is evaluated with the assistance of the

following equations:

$$\begin{aligned}
&\delta_0[(R \oplus D_0) - \pi I] + \delta_1[\alpha \otimes S_0 \otimes I] + \delta_2[\alpha \otimes T_0 \otimes I] + \delta_3[\zeta I] = 0 \\
&\delta_0[e_{m_2} \otimes \beta \times (D_1 + (1-p)\pi I)] + \delta_1[(S \oplus D) - \eta I] + \delta_4[\zeta I] = 0 \\
&\delta_0[R_0 \otimes \gamma \otimes I] + \delta_2[(T \oplus D) - \eta I] + \delta_5[\zeta I] = 0 \\
&\delta_3[(R \oplus D_0) - (\zeta + \pi) I] + \delta_4[(\alpha \otimes \theta S_0 \otimes I) + p\pi(e_{n_1} \otimes \alpha \otimes I)] + \delta_5[\alpha \otimes \theta T_0 \otimes I] = 0 \\
&\delta_1[\eta I] + \delta_3[e_{m_2} \otimes \beta \otimes (D + (1-p)\pi I)] + \delta_4[(\theta S \oplus D) - \zeta I] = 0 \\
&\delta_2[\eta I] + \delta_3[R_0 \otimes \gamma \times I] + \delta_5[(\theta T \oplus D) - \zeta I] = 0
\end{aligned}$$

subject to

$$\delta_0 e_{m_1 m_2} + \delta_1 e_{m_1 n_1} + \delta_2 e_{m_1 n_2} + \delta_3 e_{m_1 m_2} + \delta_4 e_{m_1 n_1} + \delta_5 e_{m_1 n_2} = 1$$

The condition $\delta B_0 e < \delta B_2 e$, require to maintain the queueing model's stability. i.e.,

$$\begin{aligned}
\delta_1[e_{n_1} \otimes (p\pi\alpha e_{m_1} + D_1 e_{m_1})] + \delta_2[e_{n_2} \otimes e_{m_1} D_1] + \delta_4[e_{n_1} \otimes (p\pi\alpha e_{m_1} + D_1 e_{m_1})] + \delta_5[e_{n_2} \otimes e_{m_1} D_1] \\
< \delta_0[(1-p)\pi e_{m_2} \otimes \beta e_{m_1}] + \delta_3[(1-p)\pi e_{m_2} \otimes \beta e_{m_1}]
\end{aligned}$$

4.2. The Invariant Probability Vector

Let x indicate the infinitesimal generator Q's transition probability vector. The dimensions of this probability vector's subdivisions are $x_i(i > 0) = 2(m_1 m_2 + m_1 n_1 + m_1 n_2)$, which can be written as $x = x_0, x_1, x_2$.

If x is a transition probability vector for Q, then it will also satisfy the following two requirements:

$$xQ = 0 \qquad \text{and} \qquad xe = 1$$

Once stability has been reached the following equations can be solved to determine the steady-state probability vector x.

$$x_{i+1} = x_1 R^i, \quad i \geq 1$$

When R is the least non-negative solution to the equation,

$$R^2 B_2 + R B_1 + B_0 = 0$$

Then the final two vectors, x_0 and x_1, can be obtained by resolving the following equations:
$x_0 A_{00} + x_1 A_{10} = 0$,
$x_0 A_{01} + x_1[B_1 + R B_2] = 0$
with the normalizing condition

$$x_0 e_{2(m_1 m_2 + m_1 n_1 + m_1 n_2)} + x_1 [I - R]^{-1} e_{2(m_1 m_2 + m_1 n_1 + m_1 n_2)} = 1$$

The "Logarithmic Reduction Algorithm" described by Latouche et al. [18] may be used to generate the rate matrix R.

5. Busy Period Analysis

The length of time between customers entering a system that is void and the first time the system becomes empty again is referred to as the busy period. As a result, it is the first instance where the QBD process has considered a transition from level i to level $i-1, i \geq 2$. It is necessary to discuss both $i = 0$ and $i = 1$ individually for the boundary states. There are $(m_1 m_2 + n_1 m_1 + n_2 m_1)$ states for every level i, where i is 1. Thus, the j^{th} state of level i may be denoted as (i, j) when the states are arranged in lexicographic order. Notations:

1. $G_{j,j'}(k, x)$ Given that it started in the state (i, j) at time $t = 0$, the conditional probability that starts in the state (i, j) at time $t = 0$ entered the level $(i-1)$ by making precisely k transitions to the left.

2. $\tilde{G}_{j,j'}(z,s) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-sx} dG_{j,j'}(k,x); \qquad |z| \leq 1, \quad Re(s) \geq 0$
3. The matrix $\tilde{G}(z,s) = \tilde{G}_{j,j'}(z,s)$
4. The matrix that concerns the initial flow of time without taking into consideration boundary states is $G = \tilde{G}(1,0)$.
5. $G_{i,j'}^{(1,0)}(k,x)$ - The conditional probability that, given that the QBD process began in level 1 at time $t = 0$, it will make precisely k transitions to the left to attain level 0.
6. $G_{j,j'}^{(0,0)}(k,x)$- conditional probability that was described for the initial return to level 0.
7. E_{1j} - Expected initial passage time from level i to level $i-1$, assuming a time $t = 0$ and the process in the state (i,j).
8. E_1 -column vector with E_{1j} as its entries.
9. E_{2j} -Number of customers who should have been served during the first passage time from level i to level $i-1$, assuming that the first passage time starts in the state (i,j).
10. E_2 -column vector with E_{2j} as its entries.
11. $E_1^{1,0}$ - The vector which indicates expected initial passage times between level 1 and level 0.
12. $E_2^{1,0}$ - The vector that indicates the expected amount of service completions during the initial passage from level 1 to level 0.
13. $E_1^{(0,0)}$ - The expected initial return to level 0 time.
14. $E_2^{(0,0)}$ - The expected amount of services to be completed during the first return time to level 0.

It can be easily seen that the matrix $\tilde{G}(z,s)$ satisfies the following equations

$$\tilde{G}(z,s) = z(sI - B_1)^{-1}B_2 + (sI - B_1)^{-1}B_0\tilde{G}^2(z,s)$$

once the rate matrix R is evaluated, we can easily find the matrix G by making the result

$$G = -(B_1 + RB_2)^{-1}B_2.$$

Another method for evaluating the matrix G is the logarithmic reduction algorithm. The following equations are satisfied by $\tilde{G}^{(1,0)}(z,s)$ and $\tilde{G}^{(0,0)}(z,s)$, respectively, when it comes to the boundary states, which are 0 and 1.

$$\tilde{G}^{(1,0)}(z,s) = Z(sI - B_1)^{-1}A_{10} + (sI - B_1)^{-1}B_0\tilde{G}(z,s)\tilde{G}^{(1,0)}(z,s)$$

$$\tilde{G}^{(0,0)}(z,s) = (sI - A_{00})^{-1}A_{01}\tilde{G}^{(1,0)}(z,s).$$

Therefore following the three matrices namely, G, $\tilde{G}^{(1,0)}(1,0)$ and $\tilde{G}^{(0,0)}(1,0)$ are stochastic, we may easily calculate the following moments

$$\tilde{E}_1 = -\frac{\partial \tilde{G}(z,s)}{\partial s}\Big|_{s=0,z=1} = -(B_1 + B_0(G+I))^{-1}e$$
$$\tilde{E}_2 = \frac{\partial \tilde{G}(z,s)}{\partial z}\Big|_{s=0,z=1} = -(B_1 + B_0(G+I))^{-1}e$$
$$\tilde{E}_1^{(1,0)} = -\frac{\partial \tilde{G}^{(1,0)}(z,s)}{\partial s}\Big|_{s=0,z=1} = -(B_1 + B_0G)^{-1}(B_0E_1 + e)$$
$$\tilde{E}_2^{(1,0)} = \frac{\partial \tilde{G}^{(1,0)}(z,s)}{\partial z}\Big|_{s=0,z=1} = -[B_1 + B_0G]^{-1}[A_{10}e + B_0E_2]$$
$$\tilde{E}_1^{(0,0)} = -\frac{\partial \tilde{G}^{(0,0)}(z,s)}{\partial s}\Big|_{s=0,z=1} = -A_{00}^{-1}(e + A_{01}E_1^{(1,0)})$$
$$\tilde{E}_2^{(0,0)} = \frac{\partial \tilde{G}^{(0,0)}(z,s)}{\partial z}\Big|_{s=0,z=1} = -A_{00}^{-1}A_{01}E_2^{(1,0)}$$

6. Performance Measure

- Probability that making outgoing calls by the server

$$P_{MOA} = \sum_{i=0}^{\infty}\sum_{a_2=1}^{m_2}\sum_{a_1=1}^{m_1} x_{i00a_2a_1}$$

- Probability that busy with incoming calls by the server in regular service rate

$$P_{BIS} = \sum_{i=0}^{\infty}\sum_{s_1=1}^{n_1}\sum_{a_1=1}^{m_1} x_{i01s_1a_1}$$

- Probability that server is busy with outgoing calls in regular service rate

$$P_{BOS} = \sum_{i=0}^{\infty}\sum_{s_2=1}^{n_2}\sum_{a_1=1}^{m_1} x_{i02s_2a_1}$$

- Probability that busy with incoming calls by the server in slower service rate

$$P_{WIS} = \sum_{i=0}^{\infty}\sum_{s_1=1}^{n_1}\sum_{a_1=1}^{m_1} x_{i11s_1a_1}$$

- Probability that server is busy with outgoing calls in slower service rate

$$P_{WOS} = \sum_{i=0}^{\infty}\sum_{s_2=1}^{n_2}\sum_{a_1=1}^{m_1} x_{i12s_2a_1}$$

- Expected number of customers in the orbit

$$E_N = \sum_{i=0}^{\infty}\sum_{j=0}^{1}\sum_{a_2=1}^{m_2}\sum_{a_1=1}^{m_1} ix_{ij0a_2a_1} + \sum_{i=0}^{\infty}\sum_{j=0}^{1}\sum_{s_1=1}^{n_1}\sum_{a_1=1}^{m_1} ix_{ij1s_1a_1} + \sum_{i=0}^{\infty}\sum_{j=0}^{1}\sum_{s_2=1}^{n_2}\sum_{a_1=0}^{m_1} ix_{ij2s_2a_1}$$

7. Particular case

We consider an exponential distribution of both arriving calls such as incoming and outgoing and service times. Consider the following

$$D_0 = [-\lambda],\ D_1 = [\lambda],\ \ \alpha = [1],\ R = [\delta],$$
$$\beta = [1],\ S = [\mu_1],\ \ \gamma = [1],\ T = [\mu_2].$$

From this assumption, the infinitesimal generator matrix becomes

$$Q = \begin{bmatrix} A_{00} & A_{01} & 0 & 0 & 0 & 0 & \dots \\ A_{01} & B_1 & B_0 & 0 & 0 & 0 & \dots \\ 0 & B_2 & B_1 & B_0 & 0 & 0 & \dots \\ 0 & 0 & B_2 & B_1 & B_0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \dots \end{bmatrix}$$

The entries of the Q matrix are defined by

$$A_{00} = \begin{bmatrix} -\lambda-\delta & \lambda & \delta & 0 & 0 & 0 \\ \mu_1 & -\mu_1-\lambda-\eta & 0 & 0 & \eta & 0 \\ \mu_2 & 0 & -\mu_2-\lambda-\eta & 0 & 0 & \eta \\ \zeta & 0 & 0 & -\lambda-\delta-\zeta & \lambda & \delta \\ 0 & \zeta & 0 & \theta\mu_1 & -\theta\mu_1-\lambda-\zeta & 0 \\ 0 & 0 & \zeta & \theta\mu_2 & 0 & \theta\mu_2-\lambda-\zeta \end{bmatrix}$$

$$A_{01} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$A_{10} = \begin{bmatrix} 0 & \pi(1-p) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi(1-p) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \pi p & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi p & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$B_1 = \begin{bmatrix} -\lambda-\delta-\pi(1-p) & \lambda & \delta & 0 & 0 & 0 \\ \mu_1 & -\mu_1-\lambda-\eta-\pi p & 0 & 0 & \eta & 0 \\ \mu_2 & 0 & -\mu_2-\lambda-\eta & 0 & 0 & \eta \\ \zeta & 0 & 0 & -\lambda-\delta-\zeta-\pi(1-p) & \lambda & \delta \\ 0 & \zeta & 0 & \theta\mu_1 & -\theta\mu_1-\lambda-\zeta-\pi p & 0 \\ 0 & 0 & \zeta & \theta\mu_2 & 0 & \theta\mu_2-\lambda-\zeta \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 0 & \pi(1-p) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi(1-p) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In this sequel, the generator matrix B becomes

$$B = \begin{bmatrix} -\lambda-\delta-\pi(1-p) & \lambda-\pi(1-p) & \delta & 0 & 0 & 0 \\ \mu_1+\pi p & -\mu_1-\eta-\pi p & 0 & 0 & \eta & 0 \\ \mu_2 & 0 & -\mu_2-\eta & 0 & 0 & \eta \\ \zeta & 0 & 0 & -\lambda-\delta-\zeta-\pi(1-p) & \lambda-\pi(1-p) & \delta \\ 0 & \zeta & 0 & \theta\mu_1+\pi p & -\theta\mu_1-\zeta-\pi p & 0 \\ 0 & 0 & \zeta & \theta\mu_2 & 0 & \theta\mu_2-\zeta \end{bmatrix}$$

8. Numerical Result

In this section, we examine the outcome of our system with aid of numerical and graphical illustrations. For the arrival process, let us take four different MAP representations have same mean value, which is 1.

Arrival of incoming call in Erlang (ERL-A):

$$D_0 = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}$$

Arrival of incoming call in Exponential (EXP-A):

$$D_0 = [-1] \quad D_1 = [1]$$

Arrival of incoming call in Hyper exponential (HYP-A):

$$D_0 = \begin{bmatrix} -2.8 & 0 \\ 0 & -0.28 \end{bmatrix} \quad D_1 = \begin{bmatrix} 2.24 & 0.56 \\ 0.224 & 0.056 \end{bmatrix}$$

Arrival of incoming call in MAP-Negative Correlation (MAP-NC):

$$D_0 = \begin{bmatrix} -1.00222 & 1.00222 & 0 \\ 0 & -1.00222 & 0 \\ 0 & 0 & -225.75 \end{bmatrix} \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0.01002 & 0 & 0.99220 \\ 223.4925 & 0 & 2.2575 \end{bmatrix}$$

Arrival of outgoing call in Erlang (ERL-A):

$$\alpha = [1, 0] \quad R = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

Arrival of outgoing call in Exponential (EXP-A):

$$\alpha = [1] \quad R = [-1]$$

Arrival of outgoing call in Hyper exponential (HYP-A):

$$\alpha = [0.8, 0.2] \quad R = \begin{bmatrix} -2.8 & 0 \\ 0 & -0.28 \end{bmatrix}$$

For service times let us consider phase type distribution as follows:

Service of incoming call in Erlang (ERL-S):

$$\beta = [1, 0] \quad T = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

Service of incoming call in Exponential (EXP-S):

$$\beta = [1] \quad T = [-1]$$

Service of incoming call in Hyper exponential (HYP-S):

$$\beta = [0.8, 0.2] \quad T = \begin{bmatrix} -2.8 & 0 \\ 0 & -0.28 \end{bmatrix}$$

Service of outgoing call in Erlang (ERL-S):

$$\gamma = [1, 0] \quad S = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

Service of outgoing call in Exponential (EXP-S):

$$\gamma = [1] \quad S = [-1]$$

Service of outgoing call in Hyper exponential (HYP-S):

$$\gamma = [0.8, 0.2] \quad S = \begin{bmatrix} -2.8 & 0 \\ 0 & -0.28 \end{bmatrix}$$

8.1. Illustrative Example 1

From Tables 1,2, 3 and 4 explore the effectiveness of the breakdown rate (η) on the traffic intensity (ρ) and expected queue size (EN). We fix $\lambda = 0.1$, $\delta = 0.01$, $\mu_1 = 2$, $\mu_2 = 1$, $\zeta = 2$, $\theta = 0.6$, $\pi = 0.9$ and $p = 0.4$. We observe from tables 1, 2, 3 and 4 as increasing breakdown rate (η), the traffic intensity increases and also mean queue length (EN) increases for various arrival and service distributions.

Table 1: *Breakdown rate (η) vs ρ and EN - $EXP-A$*

	$EXP-S$		$ERL-S$		$HYP-S$	
η	ρ	EN	ρ	EN	ρ	EN
1.1	0.3641	1.16678	0.33822	1.16428	0.49312	1.18627
1.2	0.3666	1.16913	0.34063	1.16661	0.49634	1.18903
1.3	0.3691	1.17142	0.34297	1.16889	0.49943	1.19170
1.4	0.3715	1.17364	0.34523	1.17110	0.50238	1.19431
1.5	0.3738	1.17581	0.34742	1.17325	0.50522	1.19684
1.6	0.3761	1.17791	0.34954	1.17535	0.50794	1.19930
1.7	0.3782	1.17996	0.35160	1.17739	0.51055	1.20170
1.8	0.3803	1.18196	0.35359	1.17938	0.51306	1.20403
1.9	0.3824	1.18391	0.35553	1.18132	0.51548	1.20631

Table 2: *Breakdown rate (η) vs ρ and EN - $ERL-A$*

	$EXP-S$		$ERL-S$		$HYP-S$	
η	ρ	EN	ρ	EN	ρ	EN
0.1	0.40496	2.60789	0.36928	2.61353	0.55800	2.60590
0.2	0.40742	2.61448	0.37161	2.62046	0.56098	2.61231
0.3	0.40978	2.62088	0.37385	2.62722	0.56382	2.61855
0.4	0.41207	2.62712	0.37603	2.63381	0.56655	2.62462
0.5	0.41427	2.63321	0.37813	2.64023	0.56915	2.63055
0.6	0.41640	2.63913	0.38017	2.64650	0.57165	2.63632
0.7	0.41846	2.64491	0.38215	2.65261	0.57405	2.64196
0.8	0.42044	2.65055	0.38407	2.65858	0.57635	2.64745
0.9	0.42237	2.65605	0.38593	2.66440	0.57856	2.65282

Table 3: *Breakdown rate (η) vs ρ and EN - $HYP-A$*

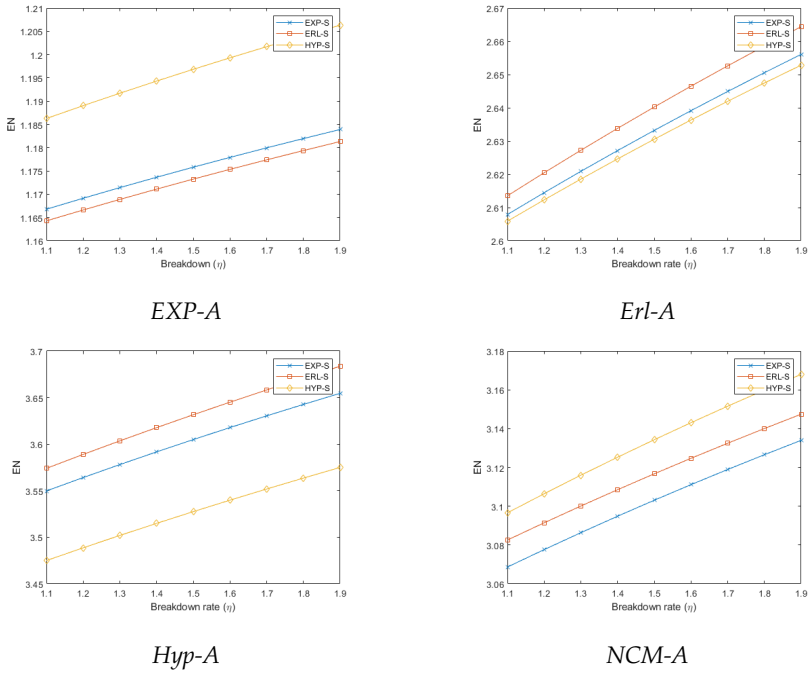
	$EXP-S$		$ERL-S$		$HYP-S$	
η	ρ	EN	ρ	EN	ρ	EN
0.1	0.26239	3.54972	0.25352	3.57397	0.31647	3.47495
0.2	0.26460	3.56408	0.25560	3.58893	0.31926	3.48861
0.3	0.26674	3.57805	0.25761	3.60351	0.32195	3.50192
0.4	0.26881	3.59165	0.25957	3.61772	0.32454	3.51488
0.5	0.27082	3.60490	0.26147	3.63157	0.32704	3.52751
0.6	0.27276	3.61780	0.26331	3.64507	0.32944	3.53982
0.7	0.27465	3.63038	0.26510	3.65824	0.33177	3.55182
0.8	0.27648	3.64264	0.26683	3.67109	0.33401	3.56354
0.9	0.27825	3.65460	0.26852	3.68363	0.33618	3.57497

Table 4: *Breakdown rate (η) vs ρ and EN - $NCM-A$*

	$EXP-S$		$ERL-S$		$HYP-S$	
η	ρ	EN	ρ	EN	ρ	EN
0.1	0.76028	3.06864	0.73819	3.08255	0.85856	3.09657
0.2	0.76097	3.07762	0.73880	3.09144	0.86000	3.10642
0.3	0.76165	3.08634	0.73940	3.10009	0.86139	3.11597
0.4	0.76230	3.09483	0.73998	3.10851	0.86271	3.12526
0.5	0.76293	3.10308	0.74054	3.11670	0.86398	3.13427
0.6	0.76354	3.11111	0.74109	3.12469	0.86519	3.14304
0.7	0.76413	3.11893	0.74162	3.13246	0.86635	3.15157
0.8	0.76471	3.12655	0.74214	3.14004	0.86746	3.15988
0.9	0.76526	3.13397	0.74264	3.14743	0.86853	3.16796

8.2. Illustrative Example 2

Two dimensional graphs are illustrated in the Figure 1. we fix $\lambda = 1$, $\delta = 0.1$, $\mu_1 = 2$, $\mu_2 = 1$, $\zeta = 3$, $\theta = 0.4$, $\pi = 0.6$ and $p = 0.5$. The illustration shows that the effect of the breakdown rate increase then the mean queue length is increase for various arrival and service distributions.



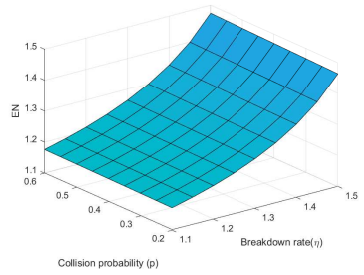
EXP-A *Erl-A*

Hyp-A *NCM-A*

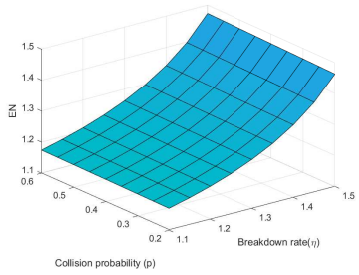
Figure 1: *Breakdown rate (η) vs Mean queue length (EN) with various arrival distributions*

8.3. Illustrative Example 3

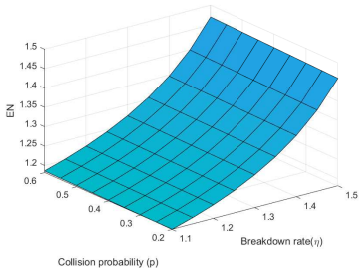
Three dimensional graphs are illustrated in the Figure 2. We examine the influence of the breakdown rate (η) and collision probability (p) on the mean queue length. We fix $\lambda = 0.1$, $\delta = 0.01$, $\mu_1 = 2$, $\mu_2 = 1$, $\zeta = 6$, $\theta = 0.6$, and $\pi = 0.3$. we could observe that the surface denotes upward trend as expected to propagate the values of breakdown rate (η) and the collision probability (p) against the expected queue size (EN) for various arrival and service times.



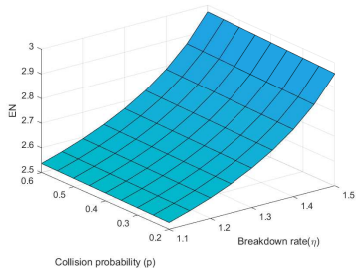
MM/MM/1



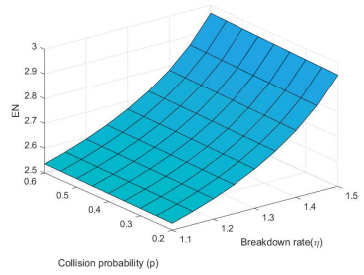
MM/EE/1



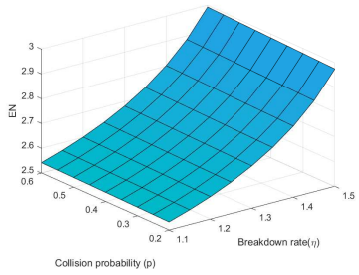
MM/HH/1



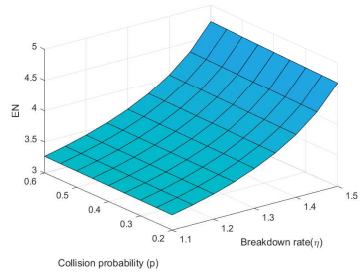
EE/MM/1



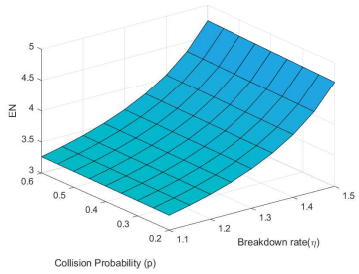
EE/EE/1



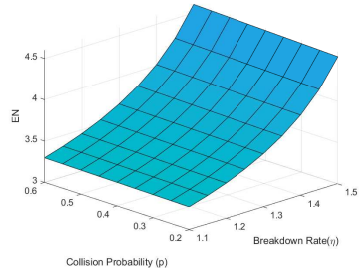
EE/HH/1



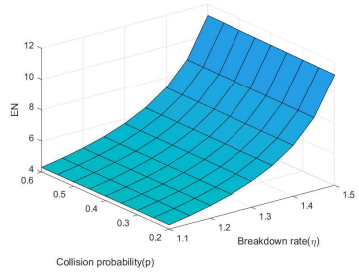
HH/MM/1



HH/EE/1



HH/HH/1



NC/MM/1

Figure 2: *Breakdown and collision vs mean queue length*

References

[1] Artalejo, J. R. and Gomez-corral, A. Retrial queueing system, Springer Berlin Heideiberg, Berlin, 2008.

[2] Artalejo, J. and Phung-Duc, T. (2013). Single server retrial queues with two way communication, *Appl Math Model*, 37(4): 1811–1822.

[3] Artalejo, J. and Phung-Duc, T. (2012). Markovian retrial queues with two way communication *Journal of Industrial & Management Optimization* 8(4):781–806.

[4] Ayyappan, G. and Gowthami, R. (2020). Analysis of $MAP, PH_2^{OA}/PH_1^l, PH_2^O/1$ retrial queue with vacation, feedback, two-way communication and impatient customers, *Soft computing*, 25(15): 9811–9838.

[5] Ayyappan, G. and Thilagavathy, K. (2020). Analysis of $MAP/PH/1$ queueing model with setup, closedown, multiple vacations, standby server, breakdown, repair and reneging. *Reliability: Theory & Applications* 15(2): 104–143.

[6] Ayyappan, G. and Udayageetha, J. (2017). Analysis of mixed priority retrial queueing system with two way communication, collisions, working breakdown, Bernoulli vacation, negative arrival, repair, immediate feedback and reneging, *Stochastic Model Application* 21(2):67–83.

[7] Chakravarthy, S. R. Markovian Arrival Processes, Wiley Encyclopedia of Operations Research and Management Science, Published Online: 15 JUN 2010.

[8] Chakravarthy, S.R. (2007). A multi-server queueing model with Markovian arrival and multiple threshoulds. *Asia-Pacific Journal of Operational Research* 24(2): 223–243.

[9] Chakravarthy, S. R. and Dudin, A. (2003). Analysis of a retrial queueing model with MAP arrivals and two types of customers. *Math computational Model* 37(3-4): 343–363.

[10] Chakravarthy, S.R. and Subramanian, S. (2018). A stochastic model for automated teller machines subject to catastrophic failures and repairs. *Queueing Models and Service Management*, 1(1):75–94.

[11] Chakravarthy, S. R., Shruti and Kulshrestha, R. (2020). A queueing model with server breakdowns, repairs, vacations, and backup server. *Operations Research Perspectives*, 7(1):1–41.

[12] Falin, G. I. and Templeton, J.G.C. Retrial queues. Chapman and Hall, London, New york, NY, 1997.

[13] Jeganathan, K., Anbazhagan, N., Kathiresan. J. (2013). A retrial inventory system with non-preemptive priority service. *International Journal of Information and Management Sciences*, 24(1): 57-77.

[14] Ke, J. C. and Chang ,F. M. (2017). $M/M/C$ balking retrial queue with vacation, *Quality Technology and Quantitative Management*, 16(1):54–66.

[15] Kim, C., Klimenok, V. I. and Orlovsky, D. S. (2008). The $BMAP/PH/N$ retrial queue with Markovian flow of breakdowns. *European Journal of Operational Research*, 189(3): 1057–1072.

[16] Krishnamoorthy , A., Babu, S. and Narayanan, V.C. (2008). $MAP/(PH/PH)/c$ queue with self-generation of priorities and non-Preemptive service, *Stochastic Analysis and Applications*, 26(6): 1250–1266.

[17] Kumar, M. S., Chakravarthy, S.R. and Arumuganathan, R. (2013). Preemptive resume priority retrial queue with two classes of MAP arrivals. *Applied Mathematical Sciences*, 7(52): 2569–2589.

[18] Latouche G, Ramaswami V. (1993). A logarithmic reduction algorithm for quasi-birth-death processes, *J. Appl. Probab.*, 30(3): 650–674.

[19] Lucantoni, D. (1991). New results on the single serverqueue with a batch Markovian arrival process. *Stochastic Models*, 7, 1-46.

[20] Madan, K. C. (2003). An $M/G/1$ type queue with time-homogeneous breakdowns and deterministic repair times, *Soochow Journal of Mathematics*, 29(1): 103–110.

[21] Neuts, M. F. (1979). A versatile Markovian point process. *J.Appl. Prob.*, 16, 764-779.

[22] Suganya, R., Lewis, N., Anbazhagan, N., Amutha, S., Kameswari, M., Acharya, S. and Joshi, G. P. (2021). Perishable inventory system with N-Policy, MAP arrivals, and impatient customers, *Mathematics* 9(13): 1514(1-15).

LSA: A LIGHTWEIGHT SYMMETRIC ENCRYPTION ALGORITHM FOR RESOURCE-CONSTRAINED IOT SYSTEMS

Amita Shah[1], Sanjay Shah[2], Hiren Patel[3], Namit Shah[4]

•

[1]Ph.D Scholar, Computer/IT Engineering, Gujarat Technological University, Gujarat, India
[2]Professor & Head, Computer Engineering Dept., Government Engineering College, Rajkot, Gujarat, India
[3]Principal, Vidush Somany Institute of Technology and Research, Sarva Vidyalaya Kelavani Mandal, Kadi, Gujarat, India.
[4]Student, Computer Engineering Dept., L D College of Engineering, Ahmedabad, Gujarat, India

Abstract

Today, Internet of Things (IoT) systems are being employed in a wide variety of domains, such as education, healthcare, industrial equipment automation, etc. With gigabytes of data being generated and processed by even the average IoT system, securing this generated data is crucial task. It requires a low-cost, high-performance encryption system for constrained IoT systems. The Advanced Encryption Standard (AES) is widely used for many cryptographic domains because of its strong security characteristics. AES is designed for general-purpose symmetric encryption algorithm but there is a need for a lighter algorithm that is specifically tuned for the needs of IoT devices with limited computation capabilities. Aim: This paper is proposing Lightweight Symmetric Algorithm (LSA) as a faster and lighter alternative to the standard AES-128 for IoT applications. The primary objective of its design was to minimize the time and memory usage required for encryption and decryption processes while retaining the strong security characteristics. Method: The research also demonstrates the comparative analysis of LSA and AES based on efficiency and resource usage. It also proves the difficulty of performing a successful brute force attack, confusion and diffusion properties, and avalanche criterion satisfiability are identical for AES and LSA algorithms. Findings: The comparison analysis of LSA and AES suggests a 14.68% lower memory usage for encryption and decryption as well as more than a 50% decrease, on average, in the required time for encryption or decryption of differently-sized files consisting of the same 128 bit data blocks. The comparisons and empirical observations show that AES and LSA are both almost identical in terms of their security characteristics such as the difficulty of performing a successful brute force attack, confusion and diffusion properties, and avalanche criterion satisfiability. Conclusion: The proposed LSA algorithm is compared with various available lightweight cipher technologies with respect to time, memory, and security properties suggests the suitability of LSA for resource constrained IoT devices with strong security requirements.

Keywords: IoT, Data security, Cryptography, Lightweight Algorithms, Security Encryption.

1. INTRODUCTION

The Internet of Things (IoT) is becoming an increasingly significant aspect of daily life as more and more devices with digital identities are connected to the Internet. The IoT paradigm is based on the connection between widely used and extremely diverse networked "things" like sensors, actuators, smartphones, etc., whose widespread use is due to recent advances in communication, sensor technologies, networking capabilities, mobile devices, cloud computing, etc. Data security is now a major necessity for many organisations. Security and privacy needs must be met because the IoT devise are integrated in users' daily lives [1-2].

However, due to the multiple standards and communication technologies involved, the IoT does not directly support conventional security measures. The term "lightweight cryptography" refers to a branch of cryptography that aims to create algorithms for use on hardware without the necessary resources like memory, power, and operational capacity to carry out the operation [3]. Only a few reliable hybrid cryptosystems are available to protect IoT smart devices. The objective is to create hybrid cryptosystems that can match the high performance demands of these constrained environments while possessing similar encryption capabilities. Even though many other new lightweight algorithms have been developed, there is always potential for development in terms of security and overhead reduction. [4-6].

The encryption of IoT data can be achieved by two ways, Symmetric and Asymmetric Cryptography. Symmetric encryption techniques are effective at protecting data, but communicating a secret key requires a separate mechanism. The key distribution issue is solved by asymmetric encryption techniques, although they are slower and consume far more resources than symmetric encryption. According to NIST, information security, like any other information technology management system, relies on three fundamental aspects: confidentiality, availability, and integrity. [8-10].

In this research, LSA aims to enhance the time and memory efficiency of AES without compromising the security features. The remaining sections of this paper are organized as follows: Section 2 provides an overview of current lightweight symmetric algorithms suitable for IoT environments. Section 3 outlines the design of the LSA algorithm being proposed. Section 4 showcases the performance and security analysis of the implemented algorithms, along with a comparison based on specific key parameters. At last, in Section 5, the paper is concluded by proving the security requirements of IoT system along with communication efficiency, resource utilization and strong encryption methodology.

2. RELATED WORKS

This section presents an overview of the existing lightweight symmetric encryption algorithms and offers a comparative analysis among them. Additionally, it investigates the suitability of the AES algorithm for enhancement, specifically to cater to the requirements of resource-constrained devices.

2.1 Lightweight Cryptography

Embedded systems, Internet of Things, and mobile computing devices are used across many industries [13], which is lacking the security mechanism for resource-constrained network. Lightweight cryptography is a trade-off between communication efficiency and data security. Due to its applicability to IoT systems with limited battery life, space, and memory size, lightweight cryptography has gained popularity [11-12]. Different approaches can be taken to implement lightweight cryptographic techniques, with some relying on software while others on hardware. Hardware-based lightweight cryptography aims to address performance limitations such as device size and power consumption. On the other hand, software-based lightweight cryptography focuses

on reducing CPU/memory usage, calculation complexity, and energy/power consumption [14]. Lightweight cryptography can be achieved through various methods, such as modifying or enhancing existing algorithms or creating new algorithms with lightweight characteristics. [15].

2.2. Overview of AES Algorithm

The AES algorithm [20] is built upon the SPN (substitution-permutation network) structure and incorporates key features such as high sensitivity to initial round and control parameters, random-like behaviors, and simplicity [18]. In the SPN structure, even slight modifications in the initial state and parameter configurations within the round function can result in significant and unpredictable changes in the final state [18][20]. AES operates on 128-bit (16-byte) blocks for both encryption and decryption of data. Figure 1 illustrates the basic block diagram of the AES algorithm. AES supports three key sizes: 128, 192, or 256 bits. For 128-bit keys, AES employs 10 rounds, for 192-bit keys it uses 12 rounds, and for 256-bit keys, it utilizes 14 rounds. Each round, except the final one, incorporates the SubBytes, ShiftRows, AddRoundKey, and MixColumns operations [17], [18], and [19]. It is worth noting that in this context, the term AES specifically refers to AES-128.

Key Expansion Routine:

The Key Expansion Routine of the standard AES-128 is used without any modification, which generates 11 keys of 128 bit from one single encryption/decryption key. It generates an array of 11 keys from the original seed key, which now becomes key 0. For our variations with a reduced number of rounds (7, instead of 10), only the first 8 keys are used. As each key consists of 4 words of 4 bytes each, we need 44 words in total for 10 rounds of encryption or decryption.

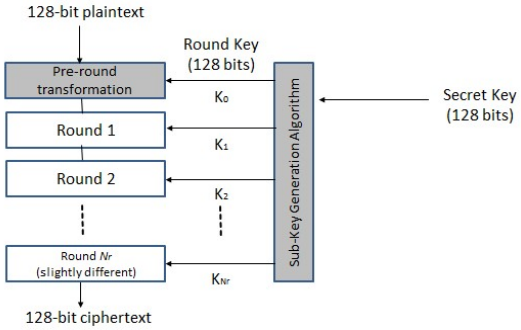


Figure 1: *Block diagram of AES-128 [16]*

AES-128 Key Expansion

```
for (i = 0 ; i < 4 ; i++)
        w[i] = key[i];
for (i = 4; i < 44 ; i++)
temp = w[i – 1];
if (i Mod 4 == 0)
temp = SBox ( RotWord ( temp ) )  Xor  Rcon[i / 4];
w[i] = w[i – 4]  Xor temp
```

RotWord(): Performs right shift on the word by 1 byte (0, 1, 2, 3 => 1, 2, 3, 0).
SBox(): Substitution from the Rijndael S-Box.
Xor: Bitwise Xor.
Rcon: The round constant array consists of successive powers of 2, one value for each round.
1 word = 4 bytes.
4 words -> 16 bytes -> the key for one round.

2.3. Study of other lightweight algorithms

Examples of some lightweight symmetric algorithms include AES [21], CAST-128 [22], PRESENT [23], TEA [24], HIGHT [25], BCC [26], MCBB [27], RC5 [28] etc. With the least amount of resource usage possible, lightweight cryptography strives to provide proper security levels. TEA's initial release was followed by a subsequent version that included additional features aimed at improving its security. Meanwhile, Block TEA was introduced as a complement to XTEA, and it operates on blocks of any size, unlike the original's 64-bit blocks [24] TEA exhibits some vulnerability, primarily its susceptibility to equivalent keys. In other words, each key is interchangeable with three other keys, reducing the effective key size to just 126 bits. Consequently, TEA is not suitable for use as a cryptographic method. Several researchers, including in [28], [29], and [30], focused on reducing the complexity of common algorithms, and based on their findings, these approaches can be applied in an IoT environment. Since the S-Box is crucial to AES and causes confusion during the encryption process, numerous researchers, including those in [31] and [32], have attempted to develop new S-Boxes to replace the old ones to increase the security of the AES algorithm. The IoT often uses a high number of resource-constrained nodes, necessitating the adoption of lightweight cryptographic primitives [33]. PRESENT employs bit-oriented permutations, which make it hardware-oriented and less suitable for software implementations. Bit permutations can be easily accomplished in hardware through straightforward wiring, whereas software implementations struggle to achieve similar performance. The FELICS (Fair Evaluation of Lightweight Cryptographic Systems) benchmarking framework is used to assess the performance of PRESENT when executed on microcontroller software environments. However, the results of software-only implementations may be significantly slower due to the inherent hardware orientation of the algorithm [34]. A hybrid approach was employed in [35] to merge the symmetric cipher AES, asymmetric cipher RSA, and the hashing function MD5 to provide confidentiality, data integrity, and authentication. However, the use of AES in processing occupies a considerable amount of ROM and RAM, while the MD5 algorithm is vulnerable to differential attacks and the RSA key requires a significant amount of memory for processing.

Table 1: *Comparison of Lightweight Algorithms for IoT Devices*

Lightweight Algorithms	Structure	No of Rounds	Key Size	Block Size
AES	SPN	10	128	128
PRESENT	SPN	32	80	64
TEA	Feistel	32	128	64
HEIGHT	GFS	32	128	64
RC5	ARX	20	16	32

To address the limitations of the cryptographic models outlined in Tab 1, the research explores a variety of suggested cryptosystems that incorporate various mathematical calculations. Subsequently, it also proposes a resilient and secure lightweight symmetric algorithm that offers efficient protection for IoT smart devices, as detailed in the upcoming sections.

3. THE PROPOSED ALGORITHM: LSA

In this paper, we propose a lightweight, secure, and fast symmetric encryption algorithm – Lightweight Symmetric Algorithm (LSA), to provide confidentiality in resource constrained IoT Devices. LSA can encrypt and decrypt data more quickly than AES.

In the context of the IoT environment, the importance of time and memory usage is on par with security considerations. This research focuses on reducing the time complexity of the algorithm while maintaining its security measures. The proposed algorithm aims to provide a lower-

complexity encryption method compared to AES, making it suitable for resource-constrained wireless devices. Additionally, it offers enhanced resilience against attacks compared to PRESENT and TEA. The proposed lightweight symmetric encryption algorithm adopts a substitution-permutation structure and builds upon the widely-used AES algorithm. By reducing the number of rounds and replacing the mixcolumn operation with junction jumping in the proposed LSA, performance improvements are achieved without compromising the security properties of the algorithm. Further details of the LSA are discussed in the following subsections.

Considering the constraints and requirements of IoT, there is a need to improve the AES algorithm in terms of time and energy consumption. With this objective in mind, we conducted tests and evaluations to identify the most time-consuming parts of the AES algorithm, which could be potential areas for optimization.

Each round of the AES algorithm involves four operation calls: Substitution, Shift Rows, Mix Columns, and Add Round Key. While AES can be implemented efficiently and cost-effectively in hardware [36], its software implementation tends to be more computationally intensive in terms of processing time.

Analysis of Modification in AES Algorithm:

To improve the performance of the algorithm, we developed three different versions of each operation of AES with the following variations. The research shows modified compute-intensive operations to make them lighter and examined nine more versions of modified AES.

Three different versions of SubBytes (Disabling ShiftRows and MixColumns).

- SubBytes_v0> Substitution bytes with 100% Substitution (Original)
- SubBytes_v1> Substitution bytes with 50% Substitution (checkerboard pattern)
- SubBytes_v2> Substitution bytes with 25% Substitution (only 1 in every four elements in the block)

depicts execution time analysis of different variants of SubBytes operation from which SubBytes_v2 is a relatively lightweight operation as per the performance.

Three different versions of ShiftRows (Disabling SubBytes and MixColumns)

- ShiftRows_v0> keep the 1st row unchanged and shift 2nd, 3rd and 4th row by 1,2 and 3 bytes subsequently (Original)
- ShiftRows_v1> keep the 1st and 3rd row unchanged and shift the 2nd and 4th row by 1 byte
- ShiftRows_v2> Let us keep the 1st, 2nd and 3rd row unchanged and shift the 4th row by 1 byte

The experiment shows the execution time analysis of different variants of ShiftRows operation from which SubBytes_v2 is a relatively lightweight operation.

Three different versions of MixColumns (Disabling ShiftRows and SubBytes).

- MixColumns_v0> Matrix Multiplication with the constant matrix (Original)

MixColumns_v1> Matrix Addition with the constant

- matrix
- MixColumns_v2> Matrix Subtraction with the constant matrix

This demonstrates that MixColumns_v1 is a lightweight component as per the experimental analysis of MixColumns variants. The constant matrix, here, refers to the AES Multiplication Matrix [19]. For V0, the Inverse Multiplication Matrix [19] is also required, while for V1 and V2, the same matrix is used during decryption as well.

Based on the various performed variations we have developed AES with combination of the fastest versions of all the 3 operations (SubBytes, ShiftRows & MixColumns) along with AddRoundKey) which incorporates V2_SubBytes (Substitution bytes with 25% Substitution), V1_ShiftRows (keeps the 1st and 3rd row unchanged and shift the 2nd and 4th row by 1 byte), V1_MixColumn (Matrix Addition with the pre-defined constant and simple XORing with the key in the AddRoundKey operation. It definitely reduces execution time, but at the same time, compromises certain level of

security.

Optimization Operations: Lightweight Security Algorithm for IoT

Reducing the complexity of operations in the proposed algorithm can potentially compromise its security level. To address this, the next improvement focuses on enhancing the security measures. The analysis of this test primarily serves the purpose of benchmarking and facilitating future experimentation. In the context of this research, the proposed algorithm is denoted as LSA-v1, which integrates the fastest versions of all the stages in AES002E

To investigate the effect of various operations on the encryption time, the experiment removes the operations one by one from the encryption process in AES.

First of all, to find heavy components, AES is modified and created the following variations:

- Only keeping ShiftRows and MixColumns and disabling SubBytes
- Only keeping SubBytes and MixColumns and disabling ShiftRows
- Only keeping SubBytes and ShiftRows and disabling MixColumns

The experiment depicts that execution time will be decreased if we remove MixColumns operation from AES.

The Mix Columns operation is generally the most resource-intensive operation in AES, and its removal leads to an overall improvement in the algorithm's execution time. The results demonstrate a significant reduction in encryption time for 1024-byte data, decreasing from 70 milliseconds to 15 milliseconds. Additionally, the Shift Rows operation is identified as the second most time-consuming operation after Mix Columns. Consequently, it becomes necessary to either remove or optimize the Shift Rows operation to make the algorithm more lightweight.

For this reason, the research extended experiment by further removing two operations.

- Only keeping SubBytes and disabling ShiftRows and MixColumns
- Only keeping ShiftRows and disabling SubBytes and MixColumns
- Only keeping MixColumns and disabling ShiftRows and SubBytes

In the optimization mentioned earlier, if all the operations are executed in isolation, it becomes apparent that MixColumns consumes the most time compared to other operations, reaffirming its heavyweight nature. Therefore, to enhance execution time performance, the most effective solution is to exclude MixColumns from the main core of the encryption operation. To generate lighter versions of the AES algorithm and for the replacement of the mixcolumn operation, we have introduced one more operation - Junction Jumping, which plays an important role to make the algorithm lightweight in terms of processing needs and to achieve a certain level of security.

Junction Jumping

This stage's main objective is propagating change from one byte to the next, thus introducing interdependence and linkage. Unlike the Mix Columns Operation, which is inherently exponential, this operation is linear, and primarily uses one of the most cost-effective CPU operations, bitwise XOR [29]. Fig. 2 shows the overall functioning of the newly introduced Junction Jumping operation on the 128 bits of input data. 'U' and 'L' refer to the upper and lower nibbles (4 bit groups) of all the bytes. For this stage, we consider the current state as an array of 16 bytes rather than a 4x4 matrix. Its main objective is propagating change from one byte to the next, thus introducing interdependence and linkage. Unlike the MixColumns Operation, which is inherently exponential, this operation is linear. It primarily uses one of the most cost-effective CPU operations, bitwise XOR. The proposed algorithm incorporates all the standard operations except MixColumns. It replaces MixColumns operation with Junction Jumping. The research achieved an improvement in the security characteristics, whereas the time complexity of the algorithm increases. Fig. 2 exhibits the process of AES with JJ, which for the research is referred as LSA-v2.

The Jumping Junction Algorithm

Prev ← Initial State of the 16 byte Word.
Next ← The Resultant State.
For 'i' in range(16):
Next[i] = Lower_Nibble(Prev[i]) + Xor(Upper_Nibble(Prev[i]), Lower_Nibble(Prev[i – 1]));
Since, only one half of every byte changes, there are no actual additional memory requirements as Prev is just a temporary state.
The process is also cyclic (1→2, ..., n – 1→n, n→1).

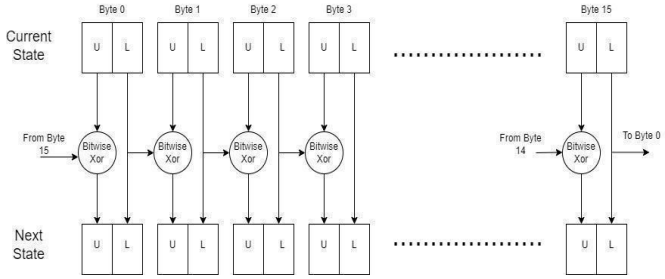


Figure 2: *Junction Jumping Operation*

The evaluations indicate that by replacing mixcolumn with JJ, time consumption has only decreased by 15%. To further explore the experimental possibilities, the proposed algorithm removes the shiftrow operation and reduces the number of rounds from 10 to 7, while maintaining the same block size of 128 bits. This modification improves the time complexity; however, it comes at the cost of compromised confusion and diffusion characteristics. The findings of the investigation into time consumption are depicted in Figure 7. It is important to note that while the round key generation process in the proposed algorithm resembles that of AES, it possesses inferior security properties. As a result, the enhanced version of LSA is denoted as LSA-v3.

In order to broaden the range of the experiments, and to maintain the trade-off between security and performance, a new additional operation - Parity Transformation is introduced which improves confusion and diffusion characteristics and is performed just after Round 0 during encryption, that is, once per the encryption / decryption process for a block. Parity Transformation plays an important role to make the algorithm lightweight in terms of processing needs and to achieve a certain level of security. It adds non-linearity to the system so that we can improve security measures, specifically the Average Avalanche Criteria that improved significantly from 32.5% to 43.33%.

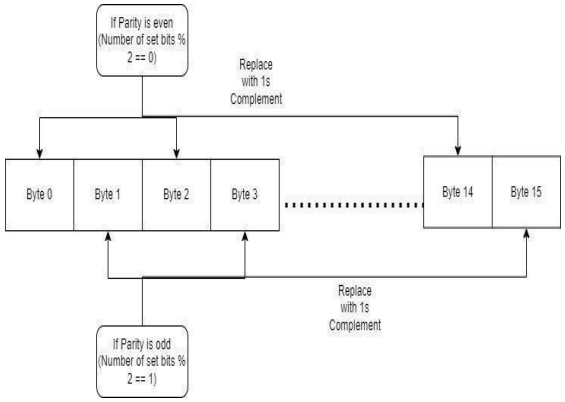


Figure 3: *Parity Transformation Operation*

Parity Transformation

This stage is performed only once and is incorporated in Round 0 to add non-linearity to the system so that we can improve security measures. It works on the principle that if an even number of bits are flipped, the resultant parity remains the same. This ensures that the transformation remains reversible. Fig. 3 shows the functioning of it.

The Steps: Parity Transformation Operation

_____

Run a loop to find the parity of the word.
Run another loop and 1s complement those bytes whose indices Mod 2 = The_Original_Parity.
Mod here refers to the modulo operation (the remainder).
The_Original_Parity will be 1 if the number of bits in the input block were odd; otherwise it would be 0.
Basically, if the parity is 1, the bytes at odd indices will be flipped and if it is even, then the bytes at even indices will be flipped.

_____

The research has advanced with the reduced number of rounds (seven), but because of the Parity Transformation operation there is a huge improvement in the confusion and diffusion characteristics of the algorithm. In this version it reached the required result, and so the paper is proposing it as LSA - Lightweight Symmetric Algorithm. The process of proposed LSA is displayed in Fig. 4 for visualization and clarity purposes. Execution time to process average 20MB data is 1.8s, 1.3s, 1.5s, 0.9s and 0.95s for AES, AES-fastest operations(LSA-v1), AES-JJ(LSA-v2) , AES-JJ-7 rounds(LSA-v3) and LSA, respectively. Fig. 5 and Fig. 6 illustrate the performance comparison of LSA with experimental versions. It has also compared with various lightweight symmetric algorithms like PRESENT and TEA, respectively; which shows that LSA performs better among all in terms of execution time. The proposed encryption process in LSA is designed and executed as shown in Fig. 4.

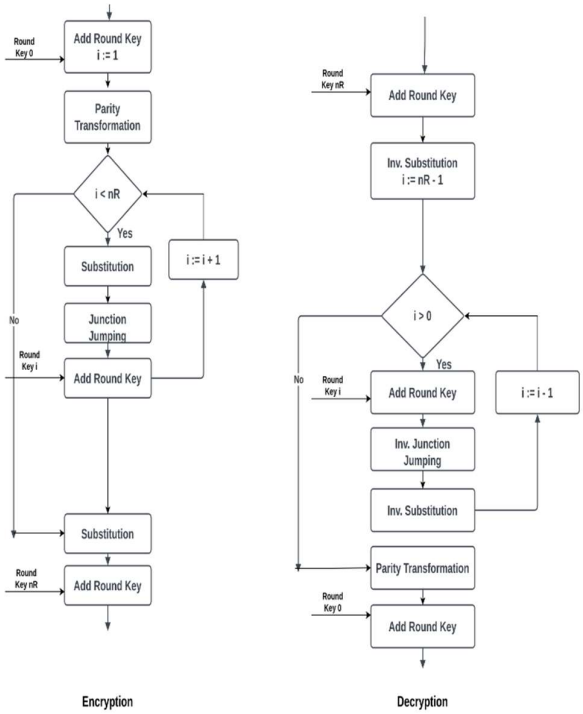


Figure 4: *Process of LSA*

The encryption process in LSA is outlined and illustrated in Figure 4. It closely resembles the encryption process in AES with some notable differences. In LSA, a parity transformation operation is conducted on the initial state during Round-0, and the resulting output becomes the input for the subsequent stages. In each encryption round of LSA, a round key is applied to encrypt a data block, similar to the Add Round Key operation in AES. However, instead of using MixColumns and ShiftRows as in AES, LSA employs the junction jumping operation as replacements.

LSA: Lightweight Security Algorithm

LSA-v1: The research has created a new AES variation with each of the 3 standard stages being replaced by a version of them from above with the 'best' performance characteristics. This version is used mostly for the purposes of benchmarking and further experimentation. While exhibiting impressive time complexity and average runtime characteristics, it had dismal security properties.
LSA-v2: In this version, a single modification has made to the Standard AES algorithm, replacing the high-cost Mix Columns Stage, which happens to be the most time-expensive stage, with the Junction Jumping Stage. This greatly improved the security characteristics but came at the cost of significantly longer runtimes and greater time complexity than LSA-v1.
LSA-v3: In this version, the number of rounds is reduced by 3 (7 instead of 10) because of expected early obfuscation (of a satisfactory level) and to do away with the mixcolumns and shiftrows stages in LSA-v2. Instead, it is replaced with junction jumping. This, as expected, came at the cost of the algorithm's security properties.

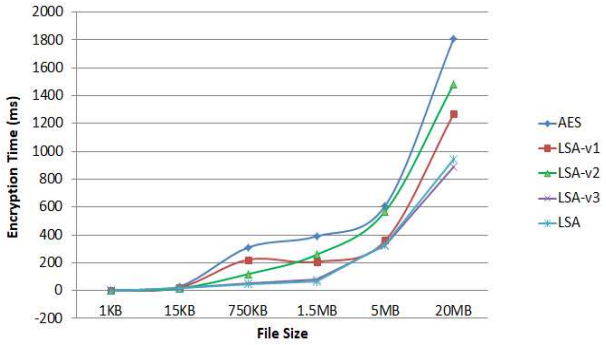


Figure 5: *Performance of different Versions of Lightweight Symmetric Algorithm (LSA)*

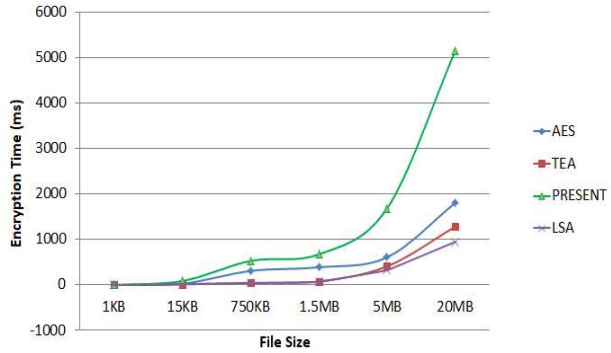


Figure 6: *Comparison of LSA with other lightweight symmetric algorithms*

LSA: This is the final algorithm that is proposed as an optimised solution. It is introduced in one new stage, Parity Transformation, into LSA-v3 for better confusion and diffusion characteristics (specifically the Average Avalanche Criteria that improved significantly from 32.5% to 43.33%).

4. IMPLEMENTATION RESULT ANALYSIS

Cryptographic algorithms are commonly implemented as hardware modules on sensor nodes. However, for off-the-shelf nodes lacking hardware security implementation, software implementation or hardware/software co-design approaches are considered suitable alternatives. It is often impractical to add new hardware circuitry to these nodes, making software implementation and evaluation of encryption algorithms more feasible. In software implementations, the primary design objectives are to minimize memory usage and optimize processor throughput and power efficiency. Consequently, the focus lies on reducing memory occupation while achieving improved performance and power savings. The forthcoming sections will delve into the analysis and results of performance and security metric comparisons between LSA, AES, PRESENT, and TEA algorithms. These discussions will shed light on the achieved performance and security levels of these algorithms.

4.1. Performance Metrics

Performance metrics hold significance in the comparison of various cipher algorithms. Consequently, it is essential to establish a consistent platform and mutually agreed-upon metrics. As part of our study, we have successfully implemented the proposed LSA algorithm and conducted a comparative analysis with existing algorithms, namely AES, PRESENT, and TEA. Given the limitations imposed by memory, power, and processing resources in wireless nodes, our evaluation primarily focuses on measuring time and memory consumption parameters. These metrics allow us to assess the overall performance of the implemented encryption algorithms in the context of wireless node constraints.

1) Encryption Time

The performance of an algorithm improves as its speed increases. Based on the findings presented in Figure 7, it is observed that PRESENT requires the longest time to execute the encryption operation, followed by AES. However, by reducing the number of rounds from 10 to 7 and replacing the ShiftRows and MixColumn operations with Junction Jumping, the LSA algorithm demonstrates a substantial improvement in time complexity. In this section, a more detailed analysis of the LSA algorithm's security properties will be conducted.

2) Memory Usage

IoT devices, particularly sensors, often have limited storage capacity. This storage space is primarily allocated to the operating system and the data collected by the sensors. Consequently, there is little room available for implementing security algorithms. Due to these constraints, it is not feasible to employ complex encryption algorithms on IoT nodes. In this study, we evaluate the RAM and ROM usage of each of the aforementioned algorithms. ROM usage refers to the space occupied by permanent code on the sensor nodes. On the other hand, RAM usage pertains to the space required for runtime code, including the storage of the stack and variables for intermediate calculation results. Since RAM directly impacts sensor performance during runtime, it holds greater significance than ROM [38].

The set up of a test bed for experiments and implemented these algorithms in raspberry pi to observe the usage of memory. Fig. 7 shows the memory use records for the encryption and decryption techniques. Compared to other algorithms, the PRESENT algorithm uses the most ROM. LSA uses less RAM and ROM than AES, PRESENT, and TEA do but less RAM and ROM than TEA do as well. The implementation of the RAM involves sophisticated technology and is more expensive than the ROM memory [38]. In LSA compared to AES, round-key generation is reduced, resulting in decreases in ROM and RAM utilisation of 13.15% and 14.68%, respectively. According to [39], low-cost implementations call for up to 4 KB ROM and 8 KB RAM, and lightweight implementations call for up to 4 KB ROM and 256 bytes RAM.

4.2. Security Metrics

The energy consumption and latency of the encryption operation are increased when the packet size is increased [37]. As the data packets transmitted by sensor nodes are typically small in size, the performance evaluation metrics focus on these small-sized packets. For the purpose of evaluating performance, we consider 10,000 randomly generated blocks, each sized at 128 bits. The encrypted outputs produced by each algorithm are then used for the analysis of security metrics.

1) Key Size (Length)

The size of the initial key plays a crucial role in determining the security level of encryption algorithms, particularly in their resistance against brute force attacks. The longer the key size, the more secure the encryption algorithm becomes. However, longer key sizes also result in increased key processing time and memory space requirements. Thus, selecting an appropriate key size depends on the desired security levels and the available resources. According to Table 1, PRESENT and TEA have a key size of 64 bits, while AES and LSA employ a key size of 128 bits. Among the algorithms discussed in this paper, LSA has a significantly lower likelihood of a successful brute force attack due to its larger key size compared to the other algorithms.

2) Information Entropy Analysis

Information entropy is a measure of the probability distribution of random events and can be utilized to assess diffusion characteristics. A higher level of system diffusion corresponds to a greater entropy value. In the analysis of entropy, random events can be represented as sequence values in bytes. In our case, the ideal entropy value is 4, which indicates that the values of the encrypted string are fully distributed [40]. To calculate the system entropy, we consider each nibble in the output as a unique symbol, resulting in a total of 24 possible symbols. The Shannon entropy value reflects the prevalence of diffusion, with a maximum possible value of 4 in our setup. Table 2 presents the results of the security are parameters for LSA, which comparable to those of AES. These security parameter results are obtained from the evaluation of 10,000 randomly generated data blocks.

Shannon Entropy equation:

$$H(X) = -\sum_{i=1}^{n}(p_i . log_2(p_i)) \quad (1)$$

Where P_i is the probability of every symbol.

Table 2: *Comparison of LSA and AES in term of Security metrics*

Algorithm	Hamming Distance	Shannon Entropy	Avalanche Effect
AES	50%	3.611	49%
LSA	50%	3.612	46.66%

In the results achieved similarly with the character frequency distribution domain metric, AES and LSA are the two algorithms with almost equivalent entropy.

3) Diffusion and Confusion Analysis

The design of ciphers incorporates two fundamental principles: diffusion and confusion [41]. These principles aim to complicate the statistical relationship between the key and the ciphertext, ensuring that each input bit affects multiple ciphertext bits [42]. Confusion and diffusion serve to prevent the deduction of secret data and secret keys, and their effectiveness disrupts statistical and other cryptanalytic methods. Confusion obscures the connection between the ciphertext and the key, while diffusion conceals the relationship between the plaintext and the ciphertext. Furthermore, the properties of diffusion and confusion in AES and LSA will be investigated in relation to text sensitivity. This investigation will consider metrics such as completeness, the avalanche effect, and the strict avalanche effect. These metrics provide insights into the extent to which AES and LSA exhibit diffusion and confusion properties.

4) Avalanche Effect

The avalanche effect metric measures the extent to which a change in a single input bit influences the output of an encryption algorithm. A secure algorithm is expected to exhibit an avalanche effect where a single bit change in the input causes approximately half of the output bits to change, reflecting the desired confusion and diffusion properties. In the multi-order avalanche effect analysis, LSA demonstrates slightly less growth compared to AES. This decline in metric growth in LSA can be attributed to the reduction in the number of rounds, which results in lower energy consumption. The nth order avalanche criterion quantifies the change in the output when n bits are altered in the input. For AES, the avalanche metric values remain at 49% for the 1st order, 2nd order, and 3rd order avalanche criteria. On the other hand, LSA exhibits metric values of 49% for the 1st order, 42% for the second order, and 49% for the 3rd order avalanche criteria. The average avalanche criterion value for LSA is extremely close to the optimal value of 50%, indicating a high level of diffusion and confusion in the algorithm.

5) Hamming Distance

The Hamming Distance metric is employed to determine the number of bits that change when data is transformed. When more than 50% of the bits are flipped, the complement percentage is considered. This is because any value above 50% (denoted as x%) is equivalent to (100 - x)%. Thus, 50% represents the maximum possible value. The Hamming Distance metric is utilized to assess confusion and observe the degree of obfuscation in the relationship between the input and output. Both AES and LSA exhibit results that align with the optimal expectations for secure algorithms in terms of the Hamming Distance metric.

4.3. Trade-Off Points

The fair trade-off between security and performance is crucial in identifying effective solutions based on specific applications. In the case of the proposed LSA algorithm, modifications have been made to AES to improve time complexity and memory utilization. While there are other encryption algorithms like PRESENT and TEA that are designed for energy-limited systems, they are susceptible to certain types of security attacks. In comparison, LSA aims to provide better security in specific areas compared to PRESENT and TEA. LSA demonstrates a higher level of security against statistical attacks, as indicated by the tested security metrics including entropy, balance analysis, avalanche effect, and Hamming distance, similar to AES. Moreover, LSA's use of a nonlinear structure in the substitution box and the Junction Jumping operation enhances its resistance against differential attacks. Although AES may have slightly stronger security characteristics, LSA's security properties are very close and only marginally weaker. As depicted in Figure 7, LSA exhibits significantly lower time and memory overhead compared to the AES algorithm, while maintaining nearly the same level of security. Furthermore, when compared to low-energy algorithms such as PRESENT and TEA, the security improvements offered by LSA outweigh the associated increase in time and memory consumption. Therefore, LSA can be considered as a suitable encryption algorithm for battery-operated sensors and other resource-constrained IoT nodes, offering robust security properties.

5. CONCLUSION

Encryption techniques play a crucial role in safeguarding data privacy in IoT devices. However, due to the limited resources of IoT nodes, it is essential to use algorithms that are energy and memory efficient. In this paper, we propose LSA, a lightweight symmetric encryption algorithm that is based on the well-known AES algorithm. Our initial observations indicate that LSA exhibits improved resistance against specific differential and statistical attacks compared to algorithms like PRESENT and TEA. This is mainly attributed to the inclusion of nonlinear elements and a larger key space in

LSA. The algorithm leverages junction jumping and parity transformation stages to reduce the overall operation time when compared to AES. In terms of performance, LSA demonstrates a significant decrease in encryption and decryption execution time, averaging over 50% improvement compared to AES, for a variety of file sizes. Considering the resource consumption and performance of sensor network nodes, LSA appears to be a more suitable choice than AES. In our security attack analysis, we evaluated the avalanche effect, which measures the sensitivity of algorithms to changes in plaintext. The results indicate that AES exhibits slightly higher sensitivity than LSA. Additionally, LSA shows a marginally higher vulnerability to differential attacks compared to AES. To further enhance the proposed algorithm, future studies could focus on conducting performance analyses of the security metrics at different rounds and stages within LSA.

References

[1] Hernández-Ramos, J. L., García-Teodoro, P., Díaz-Verdejo, J. E., Luna-Ramírez, F., García-Hernández, Á., & Sandoval Orozco, A. L. (2018). Protecting personal data in IoT platform scenarios through encryption-based selective disclosure. *Computer Communications*, 130:20-37.

[2] Naru, E. R., Saini, H., & Sharma, M. (2017). A recent review on lightweight cryptography in IoT. In 2017 I*nternational Conference on I-SMAC IoT in social, mobile, analytics and cloud) (I-SMAC)* (pp. 1-6)

[3] Xin, M. (2015). A mixed encryption algorithm used in internet of things security transmission system. In *2015 International Conference on Cyber-enabled Distributed Computing and Knowledge Discovery* (pp. 221-225). IEEE.

[4] Goyal, T. K., & Sahula, V. (2016). Lightweight security algorithm for low power IoT devices. In 2016 *International Conference on Advances in Computing, Communications and Informatics* (ICACCI) (pp. 2074-2079). IEEE.

[5] Ragab, A., He, Y., Khan, M. I., Tao, X., & Alghathbar, K. (2019). Robust hybrid lightweight cryptosystem for protecting IoT smart devices. In Y. Pan, J. Chen, T.-H. Kim, X. Li, & R. Niedermeier (Eds.), *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage* (pp. 155-170). Springer.

[6] Singh, S., Sharma, A., Singh, D., Tyagi, S., & Rodrigues, J. J. (2017). Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions. *Journal of Ambient Intelligence and Humanized Computing*, 8(1):1-18.

[7] Pérez, S., Fuentes, E., & Roa, L. M. (2018). A lightweight and flexible encryption scheme to protect sensitive data in smart building scenarios. *IEEE Access*, 6:11738-11750.

[8] Dhanda, S. S., Singh, B., & Jindal, P. (2020). Lightweight cryptography: a solution to secure IoT. Wireless Personal Communications, 112(3):1947-1980.

[9] Yousuf, T., Malik, H., Abdullah, A., Alzahrani, A. I., & Alghathbar, K. (2015). Internet of things (IoT) security: current status, challenges and countermeasures. *International Journal for Information Security Research (IJISR),* 5(4):608-616.

[10] Thabit, F., Alhomdy, S., Al-Ahdal, A. H., & Jagtap, S. (2021). A new lightweight cryptographic algorithm for enhancing data security in cloud computing. Global Transitions Proceedings, 2(1): 91-99.

[11] Rao, V., & Prema, K. V. (2021). A review on lightweight cryptography for Internet-of-Things based applications. *Journal of Ambient Intelligence and Humanized Computing*, 12:8835-8857.

[12] Prakasam, P., Sivaramakrishnan, S., Iqbal, A. T. M., et al. (2021). An enhanced energy efficient lightweight cryptography method for various IoT devices. *ICT Express*, 7(4):487-492.

[13] Eceiza, M., Flores, J. L., & Iturbe, M. (2021). Fuzzing the internet of things: a review on the techniques and challenges for efficient vulnerability discovery in embedded systems. *IEEE Internet of Things Journal*, 8(13):10390-10411.

[14] Prakasam, P., Sivaramakrishnan, S., Iqbal, A. T. M., et al. (2021). An enhanced energy

efficient lightweight cryptography method for various IoT devices. *ICT Express*, 7(4):487-492.

[15] Roy, S., Rawat, U., & Karjee, J. (2019). A lightweight cellular automata based encryption technique for IoT applications. *IEEE Access*, 7:39782-39793.

[16] Fadhil, M. S., Khalaf, Z. A., Al-Sultani, Z. M., & Dheyab, W. R. (2020). A New Lightweight AES Using a Combination of Chaotic Systems. In 2020 *1st. Information Technology To Enhance e-learning and Other Applications (IT-ELA)* (pp. 1-6). IEEE.

[17] Naif, J. R., Abdul-Majeed, G. H., & Farhan, A. K. (2019). Secure IOT system based on chaos-modified lightweight AES. In *2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 1-8)*. IEEE.

[18] Salman, R. S., Farhan, A. K., & Shakir, A. (n.d.). Lightweight modifications in the *Advanced Encryption Standard* (AES) for IoT applications: a comparative survey.

[19] Chatterjee, R., Chakraborty, R., & Mondal, J. K. (2019). Design of lightweight cryptographic model for end-to-end encryption in IoT domain. *IRO Journal on Sustainable Wireless Systems*, 1(4):215-224.

[20] Lee, A. (1999). NIST Special Publication 800-21, Guideline for Implementing Cryptography in the Federal Government. *National Institute of Standards and Technology*.

[21] Sadkhan, S. B., & Salman, A. O. (2018). Fuzzy logic for performance analysis of AES and lightweight AES. In 2018 *International Conference on Advanced Science and Engineering (ICOASE)* (pp. 1-6). IEEE.

[22] Muthavhine, K. D., & Sumbwanyambe, M. (2021). Modifying CAST algorithm in order to Increase Encryption Strength and to Reduce Memory Limitations. In *2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (pp. 1-6). IEEE.

[23] Chatterjee, R., & Chakraborty, R. (2020). A modified lightweight PRESENT cipher for IoT security. In 2020 *International Conference on Computer Science, Engineering and Applications (ICCSEA)* (pp. 1-6). IEEE.

[24] Shamala, L. M., Varghese, J., Chacko, V., et al. (2021). Lightweight cryptography algorithms for internet of things enabled networks: An overview. *Journal of Physics: Conference Series*, 1717(1):012072.

[25] Hui, Y., Xu, L., Zhang, Z., et al. (2021). BCC: Blockchain-based collaborative crowdsensing in autonomous vehicular networks. *IEEE Internet of Things Journal*, 9(6):4518-4532.

[26] Farajallah, M. (2022). Lightweight chaotic block cipher for IoT applications. *Journal of Theoretical and Applied Information Technology*, 100(15):2879-2889.

[27] Sharafi M, Eslami M, Safkhani M, et al. A low power cryptography solution based on chaos theory in wireless sensor nodes. *IEEE Access*. 2019; 7:8737-8753.

[28] Shahzadi, Romana, et al. "Chaos based enhanced RC5 algorithm for security and integrity of clinical images in remote health monitoring." *IEEE Access 7* (2019): 52858-52870.

[29] Yao X, Chen Z, Tian Y. A lightweight attribute-based encryption scheme for the Internet of Things. *Future Generation Computer Systems*. 2015; 49:104-112.

[30] Sevin A, Mohammed AAO. A survey on software implementation of lightweight block ciphers for IoT devices. *Journal of Ambient Intelligence and Humanized Computing*. 2021:1-15.

[31] Panahi P, Dehghantanha A, Conti M, et al. Performance evaluation of lightweight encryption algorithms for IoT-based applications. *Arabian Journal for Science and Engineering*. 2021;46(4):4015-4037.

[32] Alshammari BM, Alsulaiman FM, Alsulaiman MB, et al. Implementing a symmetric lightweight cryptosystem in highly constrained IoT devices by using a chaotic S-box. *Symmetry. 2021*;13(1):129.

[33] Guo Y, Li L, Liu B. Shadow: A lightweight block cipher for IoT nodes. *IEEE Internet of Things Journal*. 2021;8(16):13014-13023.

[34] Nath S, Som S, Negi MC. Cryptanalysis of a novel bitwise xor rotational algorithm and

security for IoT devices. *International Journal of Knowledge-based and Intelligent Engineering Systems.* 202

[35] Jang K, Lee J, Lee J, Kim K. Grover on GIFT. *Cryptology ePrint Archive*. 2020.

[36] Harini A, Krishnamurthy R, Venkatesan R, Murugan A. A novel security mechanism using hybrid cryptography algorithms. In: *Proceedings of the 2017 IEEE International Conference Electrical Instrumentation and Communication Engineering (ICEICE)*; 2017. p. 1-5.

[37] Zhao W, Ha Y, Alioto M. AES architectures for minimum-energy operation and silicon demonstration in 65nm with lowest energy per encryption. In 2015 *IEEE International Symposium on Circuits and Systems (ISCAS)* 2015 May 24 (pp. 2349-2352). IEEE.

[38] Tech Differences. (2017). Difference Between RAM and ROM Memory (With Comparison Chart)—*Tech Differences*. Accessed: Apr. 28, 2018. [Online]. Available: https://techdifferences.com/difference-between-ram-and-rom-memory.html

[39] C. Manifavas, G. Hatzivasilis, K. Fysarakis, and K. Rantos, ''Lightweight cryptography for embedded systems—A comparative analysis,'' in *Data Privacy Management and Autonomous Spontaneous Security. Berlin, Germany: Springer*, 2014, pp. 333–349.

[40]Anand K, Bianconi G. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E. 2009* Oct 13;80(4):045102.

[41] X.-Y. Wang and Q. Yu, ''A block encryption algorithm based on dynamic sequences of multiple chaotic systems,'' *Commun. Nonlinear Sci. Numer. Simul*., vol. 14, no. 2, pp. 574–581, 2009.

[42] X.-J. Tong, Z. Wang, Y. Liu, M. Zhang, and L. Xu, ''A novel compound chaotic block cipher for wireless sensor networks,'' *Commun. Nonlinear Sci. Numer. Simul*., vol. 22, nos. 1–3, pp. 120–133, 2015.

A STUDY ON THE PROPERTIES OF A NEW EXPONENTIATED EXTENDED INVERSE EXPONENTIAL DISTRIBUTION WITH APPLICATIONS

Sule Omeiza Bashiru
•
Department of Mathematical Sciences, Prince Abubakar Audu University, Anyigba, Kogi State, Nigeria.
Email: bash0140@gmail.com

Abstract

In this paper, a new continuous probability distribution called a new exponentiated extended inverse exponential distribution with four parameters is introduced. The mathematical and statistical properties of the proposed distribution, such as the quantile function, moments, moment generating function, survival function, hazard function, odds function, and reversed hazard function, were studied to understand its nature. The probability density function of the order statistics for this distribution was also obtained. The parameters of the model were estimated using the maximum likelihood method of estimation. The proposed model was applied to two real datasets relating to the relief times of twenty patients receiving an analgesic and the sum of skin folds in 202 athletes collected at the Australian Institute of Sports. The results showed that the new model outperformed its comparators and provides better fit than Topp-Leone exponentiated inverse exponential, Topp-Leone inverse exponential, exponentiated inverse exponential, inverse exponential and exponential distributions.

Keywords: Akaike information criterion, breast cancer, skin fold, inverse exponential, adequacy model

I. Introduction

The creation of novel, all-encompassing statistical models is an important field of study in distribution theory. Such distributions, which are extremely valuable in forecasting and simulating real-world phenomena, are abundant in the literature. The modeling of data in various practical domains, such as bio-medical analysis, reliability engineering, economics, forecasting, astronomy, demography, and insurance, has extensively used a number of classical distributions throughout the past few decades.

The majority of exponential distribution generalizations have constant, non-increasing, non-decreasing, and bathtub hazard rates. However, in real-world situations, it is possible for the data to display a unimodal (first increasing and then decreasing) inverted bathtub hazard rate. In the analysis of breast cancer data, we found that the mortality rises early, reaches a peak after some time, and then drops gradually; the related hazard rate is thus inverted bathtub-shaped or notably unimodal. For this type of data, the one parameter inverted exponential (IEx) distribution, which

has the inverted bathtub hazard rate, has been proposed as another extension of the exponential distribution in statistical literature.

In the literature, an inverted exponential distribution which was introduced in [1]. The distribution has an inverted bathtub hazard rate and can be used to simulate real-world events that have inverted bathtub failure rates. [2] have also addressed an example of its use with breast cancer data. According to [3], it has also been described as a model that is helpful in survival analysis.

To increase the modeling flexibility of current probability distributions utilizing various families of distributions, recent research in this field has focused on extending existing probability distributions. Some families of distributions proposed in literature include Kumaraswamy generalized family of distributions by [4], Topp Leone generalized family of distributions by [5], exponentiated extended generalized family of distributions by [6], Power Lindley generalized family of distributions by [7], Topp Leone exponentiated generalized family of distributions by [8], Topp Leone Kumaraswamy generalized family of distributions by [9], Odd Chen generalized family of distributions by [10], Modi generalized family of distributions by [11], A new generalized family of distributions by [12], Type I half-Logistic exponentiated generalized family of distributions by [13], Type II half-Logistic generalized family of distributions by [14], etc.

In line with this, some of the recent developments and extensions of the inverse exponential distribution using generalized families of continuous distribution can be found in [15], [16], [17], [18], [19].

In this context, we developed a generalization of the inverse exponential distribution based on [6], which is derived from the following general construction: if *G* denotes the baseline of a cumulative distribution function, then a generalized family of distributions can be defined with cumulative distribution function and probability density function give respectively as

$$F(x;\alpha,\lambda,\theta)=\left[1-\left[1-G(x;\omega)\right]^{\alpha\lambda}\right]^{\theta} \tag{1}$$

and

$$f(x;\alpha,\lambda,\theta)=\alpha\lambda\theta g(x;\omega)\left[1-G(x;\omega)\right]^{\alpha\lambda-1}\left[1-\left[1-G(x;\omega)\right]^{\alpha\lambda}\right]^{\theta-1} \tag{2}$$

where ω is the vector of parameters of the baseline distribution.

where $G(x;\omega)$ is the cumulative distribution function (cdf) of the baseline distribution with vector of parameter ω.

for $x\geq 0,\alpha,\lambda,\theta,\omega\geq 0$, where equations (1) and (2) are the cumulative distribution function and probability density function (pdf) of the family of distributions proposed by [6].

The cdf and pdf of the inverse exponential (IEx) distribution are given by

$$G(x;\beta)=e^{-\left(\frac{\beta}{x}\right)} \tag{3}$$

$$g(x;\beta)=\frac{\beta}{x^2}e^{-\left(\frac{\beta}{x}\right)} \tag{4}$$

The major goal of this study is to build a more flexible model by adding three more shape parameters to the inverse exponential distribution to increase its goodness-of-fit to real-world data sets. The main reasons for creating the NEtEIEx distribution in practice are to make the kurtosis more flexible compared to the baseline inverse exponential model, to produce skewness for symmetrical distributions, to build heavy-tailed distributions that are not longer-tailed for modeling real data, to have distributions with symmetric, left-skewed, right-skewed, and inverted bathtub shapes, and to consistently offer better fits than other generated models under certain conditions.

II. Methods

2.1. A New Exponentiated Extended Inverse Exponential (NEtEIEx) Distribution

This section developed a new continuous probability distribution function called new exponentiated extended inverse exponential (NEtEIEx) distribution and provide some plots of its pdf, cdf survival function and hazard rate function (hrf) in order to assess the shape of the new distribution. The cdf of the NEtEIEx distribution is obtained by inserting (3) into (1) given as:

$$F(x;\alpha,\beta,\lambda,\theta)=\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta} \tag{5}$$

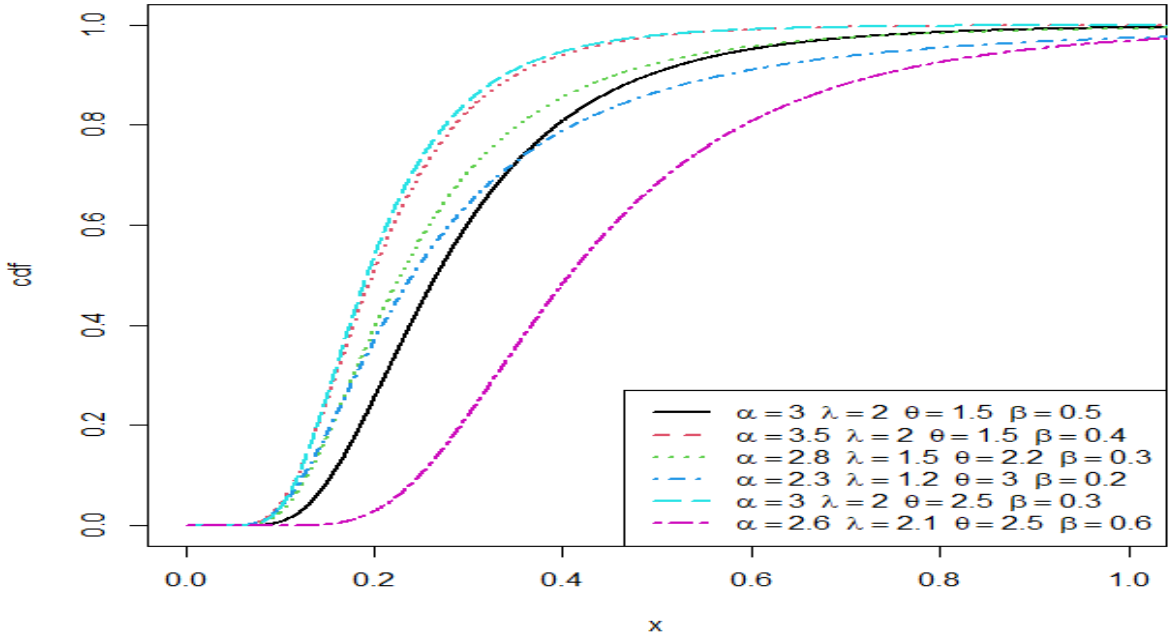


Figure 1: *Plots of CDF of the NEtEIEx distribution for different parameter values*

On differentiating equation (5) with respect to x, we obtained the pdf of a NEtEIEx distribution given in equation (6)

$$f(x;\alpha,\beta,\lambda,\theta)=\alpha\lambda\theta\frac{\beta}{x^2}e^{-\left(\frac{\beta}{x}\right)}\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda-1}\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta-1} \tag{6}$$

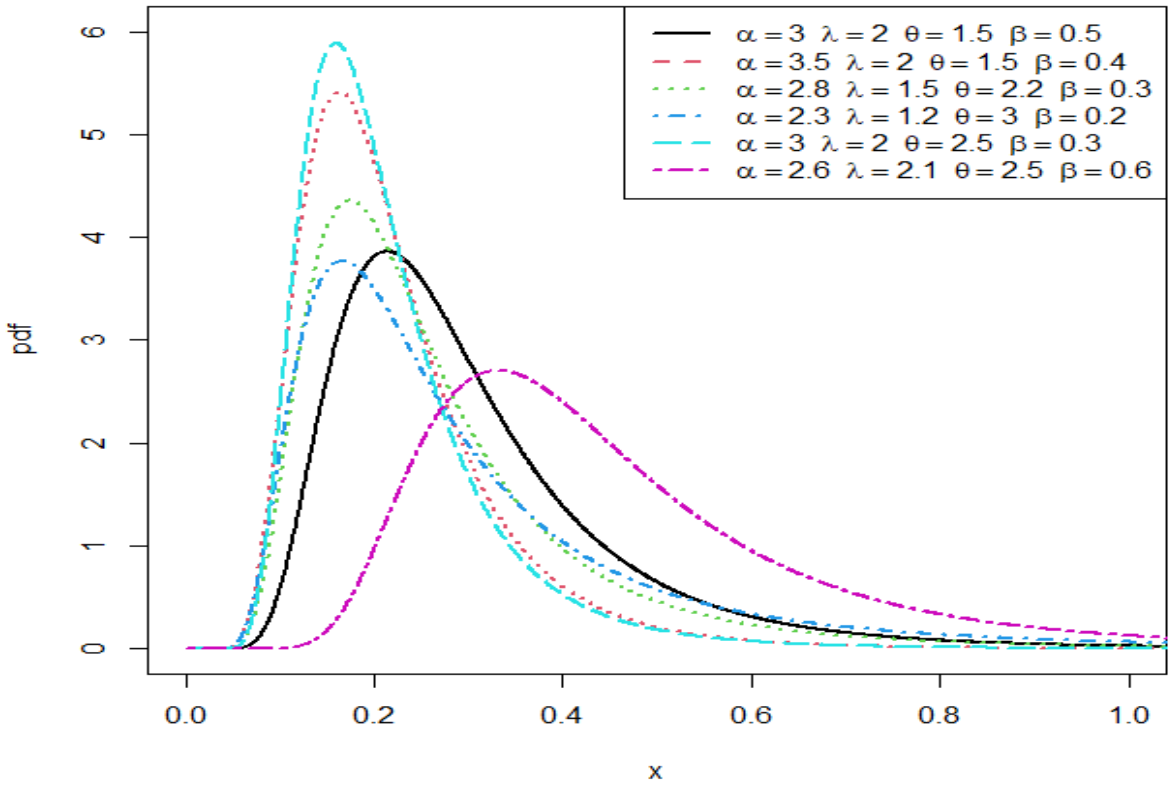


Figure 2: *Plots of PDF of the NEtEIEx distribution for different parameter values*

Where $x \geq 0$, $\beta > 0$ is the scale parameter and $\alpha, \theta, \lambda > 0$ are the shape parameters respectively.

2.1.1. Expansion of Density

In this section the pdf in equation (6) is expanded using binomial expansion. Expanding the last term in equation (6), we have

$$\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta-1} = \sum_{i=0}^{\infty}(-1)^{i}\binom{\theta-1}{i}\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda i}$$

$$\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda(i+1)-1} = \sum_{j=0}^{\infty}(-1)^{j}\binom{\alpha\lambda(i+1)-1}{j}\left[e^{-\left(\frac{\beta}{x}\right)}\right]^{j}$$

$$f(x;\alpha,\lambda,\theta) = \alpha\lambda\theta\frac{\beta}{x^{2}}\sum_{i,j=0}^{\infty}(-1)^{i+j}\binom{\theta-1}{i}\binom{\alpha\lambda(i+1)-1}{j}\left[e^{-\frac{\beta}{x}}\right]^{j+1} \quad (7)$$

Equation (7) is the expansion of equation (6) which will be used to derive some of the properties of the distribution.

2.1.2. Properties of the New Exponentiated Extended Inverse Exponential (NEtEIEx) Distribution

In this section, some of the mathematical and statistical properties of NEtEIEx distribution such as the quantile function, moments, moment generating function, reliability measure, odds function, reversed hazard function and order statistics are derived.

2.1.2.1. Moments

$$E(X^r)=\int_0^\infty x^r f(x)dx \tag{8}$$

$$E\left(X^r\right)=\alpha\lambda\theta\frac{\beta}{x^2}\sum_{i,j=0}^{\infty}\left(-1\right)^{i+j}\binom{\theta-1}{i}\binom{\alpha\lambda\left(i+1\right)-1}{j}\int_0^\infty x^r\left[e^{-\left(\frac{\beta}{x}\right)}\right]^{1+j} \tag{9}$$

Consider the integral part of equation (9), we have

$$\int_0^\infty x^r\left[e^{-\left(\frac{\beta}{x}\right)}\right]^{1+j}$$

Let $y=(1+j)\left(\frac{\beta}{x}\right)$; $x=(1+j)\left(\frac{\beta}{y}\right)$ and $\mathrm{dx}=\frac{dyx^2}{\beta(1+j)}$

$$E\left(X^r\right)=\alpha\lambda\theta\frac{\beta}{x^2}\sum_{i,j=0}^{\infty}\left(-1\right)^{i+j}\binom{\theta-1}{i}\binom{\alpha\lambda\left(i+j\right)-1}{j}\int_0^\infty\left[\left(1+j\right)\left(\frac{\beta}{y}\right)\right]^r e^{-y}\frac{dyx^2}{\beta\left(1+j\right)}$$

$$E\left(X^r\right)=\alpha\lambda\theta\beta^r\left(1+j\right)^{r-1}\sum_{i,j=0}^{\infty}\left(-1\right)^{i+j}\binom{\theta-1}{i}\binom{\alpha\lambda\left(i+j\right)-1}{j}\int_0^\infty y^{-r}e^{-y}dy$$

Where $\int_0^\infty y^{-r}e^{-y}dy=\Gamma\left(1-r\right)$

Therefore

$$E\left(X^r\right)=\alpha\lambda\theta\beta^r\left(1+j\right)^{r-1}\sum_{i,j=0}^{\infty}\left(-1\right)^{i+j}\binom{\theta-1}{i}\binom{\alpha\lambda\left(i+j\right)-1}{j}\Gamma\left(1-r\right) \tag{10}$$

Equation (10) is the moments of NEtEIEx distribution. To obtain the mean, we set *r = 1* in equation (10).

2.1.2.2. Moment Generating Function (mgf)

$$M_x(t)=\int_0^\infty e^{tx}f(x)dx \tag{11}$$

$$M_x(t)=\alpha\lambda\theta\beta^m\left(1+j\right)^{m-1}\frac{t^m}{m!}\sum_{i,j=0}^{\infty}\left(-1\right)^{i+j}\begin{pmatrix}\theta-1\\ i\end{pmatrix}\begin{pmatrix}\alpha\lambda\left(i+j\right)-1\\ j\end{pmatrix}\Gamma\left(1-m\right)$$

where the expansion of $e^{tx}=\sum_{m=0}^{\infty}\frac{t^m x^m}{m!}$

and following the process of moments above, we have

$$M_x(t)=\alpha\lambda\theta\frac{\beta}{x^2}\sum_{i,j=0}^{\infty}\left(-1\right)^{i+j}\begin{pmatrix}\theta-1\\ i\end{pmatrix}\begin{pmatrix}\alpha\lambda\left(i+1\right)-1\\ j\end{pmatrix}\int_0^{\infty}e^{tx}\left[e^{-\left(\frac{\beta}{x}\right)}\right]^{1+j} \tag{12}$$

2.1.2.3. Quantile Function

Quantile function has a significant position in probability theory and it is the inverse of the cdf. The quantile function is obtained using

$$Q(u)=F^{-1}(u) \tag{13}$$

Using the inverse of equation (5), we have the quantile function given as

$$x=\frac{\beta}{-\log\left[1-\left[1-u^{\frac{1}{\theta}}\right]\right]^{\frac{1}{\alpha\lambda}}} \tag{14}$$

The median is obtained by setting $u = 0.5$ in equation (14) given as

$$x_{median}=\frac{\beta}{-\log\left[1-\left[1-0.5^{\frac{1}{\theta}}\right]\right]^{\frac{1}{\alpha\lambda}}} \tag{15}$$

2.1.2.4. Hazard Function

Hazard function is given as

$$\tau(x;\alpha,\beta,\lambda,\theta)=\frac{f(x;\alpha,\beta,\lambda,\theta)}{R(x;\alpha,\beta,\lambda,\theta)} \tag{16}$$

The hazard function of the NEtEIEx distribution is given as

$$\tau(x;\alpha,\beta,\lambda,\theta)=\frac{\alpha\lambda\theta\frac{\beta}{x^2}e^{-\left(\frac{\beta}{x}\right)}\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda-1}\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta-1}}{1-\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta}} \tag{17}$$

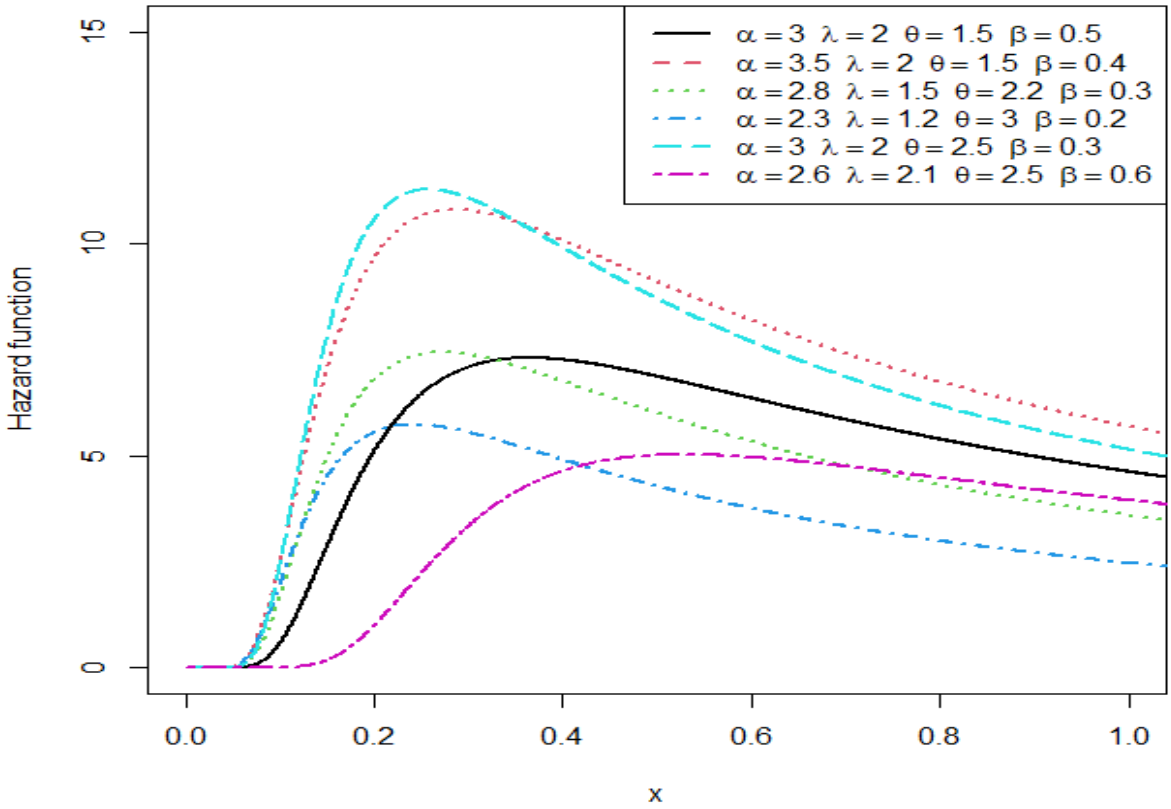


Figure 3: *Plots of hazard function of the NEtEIEx distribution for different parameter values*

2.1.2.5. Survival Function

The reliability function is also known as survival function, which is the probability that a system will survive beyond a specified time [20]. It can be defined as

$$R(x;\alpha,\beta,\lambda,\theta)=1-F(x;\alpha,\beta,\lambda,\theta) \tag{18}$$

The survival function of the NEtEIEx distribution is given as

$$R(x;\alpha,\lambda,\theta,\beta)=1-\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta} \tag{19}$$

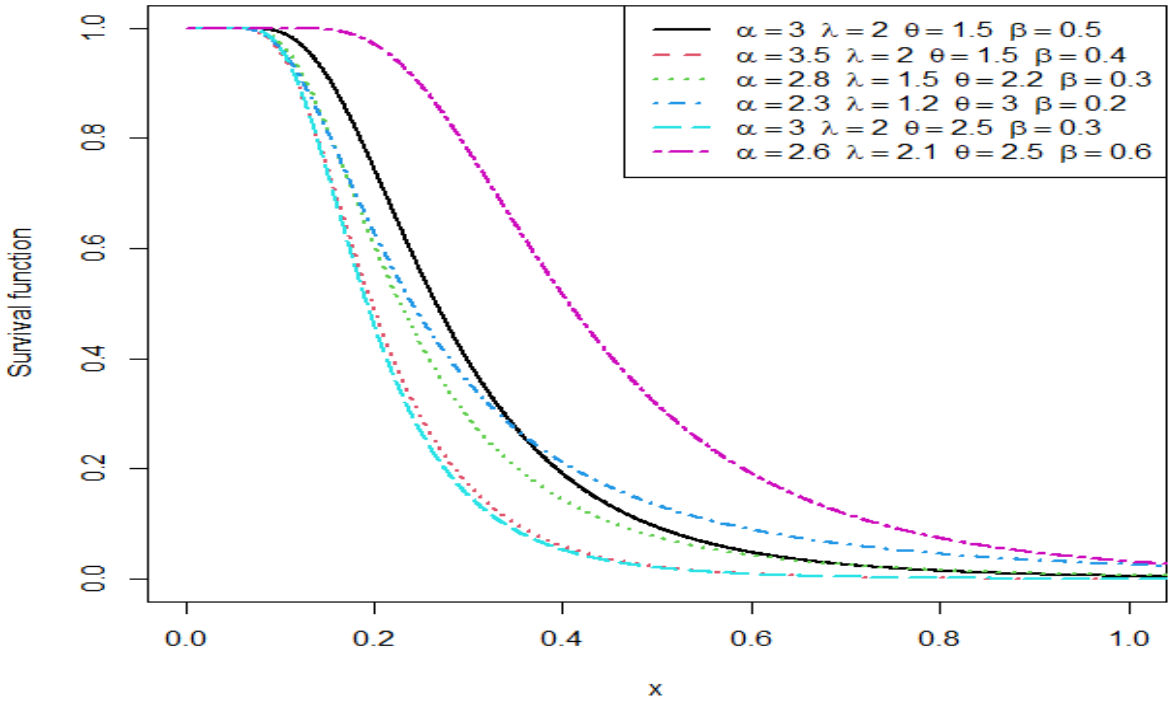


Figure 4: *Plots of survival function of the NEtEIEx distribution for different parameter values*

2.1.2.6. Reversed Hazard Function

Reversed hazard function of a random variable x is given as

$$\Re(x;\alpha,\beta,\lambda,\theta)=\frac{f(x;\alpha,\beta,\lambda,\theta)}{F(x;\alpha,\beta,\lambda,\theta)} \tag{20}$$

The reverse hazard rate function of the NEtEIEx distribution is given as

$$\Re(x;\alpha,\beta,\lambda,\theta)=\frac{\alpha\lambda\theta\frac{\beta}{x^2}e^{-\left(\frac{\beta}{x}\right)}\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda-1}\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta-1}}{\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta}} \tag{21}$$

2.1.2.7. Odds Function

The odds function of the NEtEIEx distribution is given as

$$\prod(x;\alpha,\beta,\lambda,\theta)=\frac{\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta}}{1-\left[1-\left[1-\left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta}} \tag{22}$$

2.2. Order Statistics

Let $X_1, X_2, ..., X_n$ be n independent random variable from the NEtEIEx distributions and let $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$ be their corresponding order statistic. Let $F_{r:n}(x)$ and $f_{r:n}(x)$, $r = 1,2,3,...n$ denote the cdf and pdf of the r^{th} order statistics $X_{r:n}$ respectively. The pdf of the r^{th} order statistics of $X_{r:n}$ is given as

$$f_{r:n}(x) = \frac{1}{B(r, n-r+1)} \sum_{i=0}^{\infty} \frac{(-1)^i \binom{n-r+1}{i}}{i! \binom{n-r+1-i}{}} [F(x)]^{r+i-1} f(x) \tag{23}$$

$$f_{r:n}(x) = \frac{\alpha\lambda\theta \frac{\beta}{x^2} e^{-\left(\frac{\beta}{x}\right)}}{B(r, n-r+1)} \sum_{i=0}^{\infty} \frac{(-1)^i \binom{n-r+1}{i}}{i! \binom{n-r+1-i}{}} \left[1 - \left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda-1} \left[1 - \left[1 - \left[e^{-\left(\frac{\beta}{x}\right)}\right]\right]^{\alpha\lambda}\right]^{\theta(r+i)-1} \tag{24}$$

Equation (24) is the r^{th} order statistics of the NEtEIEx distribution. To obtain the maximum and minimum order statistics, we set $r = 1$and $r = n$ in equation (24) respectively.

2.3. Estimation Method

The method of maximum likelihood estimation (MLE) is used in this section to estimate the parameters of the NEtEIEx distribution. For a random sample, $X_1, X_2, ..., X_n$ of size n from the NEtEIEx $(\alpha, \beta, \theta, \lambda)$, the log-likelihood function $L(\alpha, \beta, \theta, \lambda)$ of (6) is given as

$$\begin{aligned} \log L = n\log(\alpha) + n\log(\lambda) + n\log(\theta) + n\log(\beta) + \sum_{i=1}^{n} \frac{1}{x_i^2} + (\alpha\lambda - 1)\sum_{i=1}^{n} \log\left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right] \\ + (\theta - 1)\sum_{i=1}^{n} \log\left[1 - \left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha\lambda}\right] \end{aligned} \tag{25}$$

Differentiating the log-likelihood with respect to $\lambda, \alpha, \theta, \beta$ and equating the result to zero, we have

$$\frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} + (\lambda - 1)\sum_{i=1}^{n} \log\left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right] + \lambda(\theta - 1)\sum_{i=1}^{n} \frac{\left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha(\lambda-1)} \left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha} \log\left[e^{-\left(\frac{\beta}{x_i}\right)}\right]}{1 - \left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha\lambda}} = 0 \tag{26}$$

$$\frac{\partial L}{\partial \lambda} = \frac{n}{\lambda} + (\alpha - 1)\sum_{i=1}^{n} \log\left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right] - (\theta - 1)\sum_{i=1}^{n} \frac{\left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha\lambda} \log\left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha}}{1 - \left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha\lambda}} = 0 \tag{27}$$

$$\frac{\partial L}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log\left[1 - \left[1 - \left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha\lambda}\right] = 0 \tag{28}$$

$$\frac{\partial L}{\partial \beta}=\frac{n}{\beta}+(\alpha\lambda-1)\sum_{i=1}^{n}\frac{e^{-\left(\frac{\beta}{x_i}\right)}\frac{1}{x_i}}{1-\left[e^{-\left(\frac{\beta}{x_i}\right)}\right]}-\alpha\lambda(\theta-1)\sum_{i=1}^{n}\frac{\left[1-\left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha(\lambda-1)}\left[1-\left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha-1}e^{-\left(\frac{\beta}{x_i}\right)}\frac{1}{x_i}}{1-\left[1-\left[e^{-\left(\frac{\beta}{x_i}\right)}\right]\right]^{\alpha\lambda}}=0 \quad (29)$$

Now, equations (26), (27), (28) and (29) do not have a simple analytical form and are therefore not tractable. As a result, we have to resort to non-linear estimation of the parameters using iterative method.

III. Results

3.1. Applications

This section tests the new distribution's flexibility against a few other existing distributions using two actual data sets. AdequacyModel, a package in the R software, is used to produce the analyses' results in this study. Using the Akaike information criterion (AIC) and Bayesian information criterion (BIC), respectively, the performance of the distribution was compared to other existing distributions that were consistent with the baseline distribution in terms of providing good parametric fit to the data sets.

$$AIC=-2ll+2k \quad (30)$$

$$BIC=-2ll+k\log(n) \quad (31)$$

The model selection is carried out using the AIC and the BIC. Where ll denotes the log-likelihood function evaluated at the maxinum likelihood estimates, k is the number of parameters, and n is the sample size from the data. The model with minimum value of AIC and BIC is chosen as the best model to fit the data set. The comparators presented are Topp-Leone exponentiated inverse exponential (TLExIEx), Topp-Leone inverse exponential (TLIEx), exponentiated inverse exponential (ExIEx), inverse exponential (IEx) and exponential (Ex) distributions.

The first data set represents the relief times of twenty patients receiving an analgesic. This data set has been used by [21]. The data set is given as

1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3, 1.7, 2.3, 1.6, 2.0.

The second data set represents the sum of skin folds in 202 athletes collected at the Australian Institute of Sports, has been used by [22]. The data set is given as

28.0, 98, 89.0, 68.9, 69.9, 109.0, 52.3, 52.8, 46.7, 82.7, 42.3, 109.1, 96.8, 98.3, 103.6, 110.2, 98.1, 57.0, 43.1, 71.1, 29.7, 96.3, 102.8, 80.3, 122.1, 71.3, 200.8, 80.6, 65.3, 78.0, 65.9, 38.9, 56.5, 104.6, 74.9, 90.4, 54.6, 131.9, 68.3, 52.0, 40.8, 34.3, 44.8, 105.7, 126.4, 83.0, 106.9, 88.2, 33.8, 47.6, 42.7, 41.5, 34.6, 30.9, 100.7, 80.3, 91.0, 156.6, 95.4, 43.5, 61.9, 35.2, 50.9, 31.8, 44.0, 56.8, 75.2, 76.2,101.1, 47.5, 46.2, 38.2, 49.2, 49.6, 34.5, 37.5, 75.9, 87.2, 52.6, 126.4, 55.6, 73.9, 43.5, 61.8, 88.9, 31.0, 37.6,52.8, 97.9, 111.1, 114.0, 62.9, 36.8, 56.8, 46.5, 48.3, 32.6, 31.7, 47.8, 75.1, 110.7, 70.0, 52.5, 67, 41.6, 34.8, 61.8, 31.5, 36.6, 76.0, 65.1, 74.7, 77.0, 62.6, 41.1, 58.9, 60.2, 43.0, 32.6, 48, 61.2, 171.1, 113.5, 148.9, 49.9,59.4, 44.5, 48.1, 61.1, 31.0, 41.9, 75.6, 76.8, 99.8, 80.1, 57.9, 48.4, 41.8, 44.5, 43.8, 33.7, 30.9, 43.3, 117.8, 80.3, 156.6, 109.6, 50.0, 33.7, 54.0, 54.2, 30.3, 52.8, 49.5, 90.2, 109.5, 115.9, 98.5, 54.6, 50.9, 44.7, 41.8, 38.0, 43.2,70.0, 97.2, 123.6, 181.7, 136.3, 42.3, 40.5, 64.9, 34.1, 55.7, 113.5, 75.7, 99.9, 91.2, 71.6, 103.6, 46.1, 51.2, 43.8, 30.5, 37.5, 96.9, 57.7, 125.9, 49.0, 143.5, 102.8, 46.3, 54.4, 58.3, 34.0, 112.5, 49.3, 67.2, 56.5, 47.6, 60.4, 34.9.

Table 1: *The ML estimates and goodness of fit measurement for the first data set.*

Models	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	$-l$	AIC	BIC
NEtEIEx	3.735	1.832	13.235	1.831	15.575	39.150	43.133
TLExIEx	2.772	0.322	1.791	-	46.038	98.077	101.064
ExIEx	1.309	1.317	-	-	32.669	69.337	71.329
TLIEx	-	12.432	0.526	-	22.432	49.984	51.976
IEx	-	1.725	-	-	32.669	67.337	68.333
Ex	-	0.526	-	-	32.837	67.674	68.670

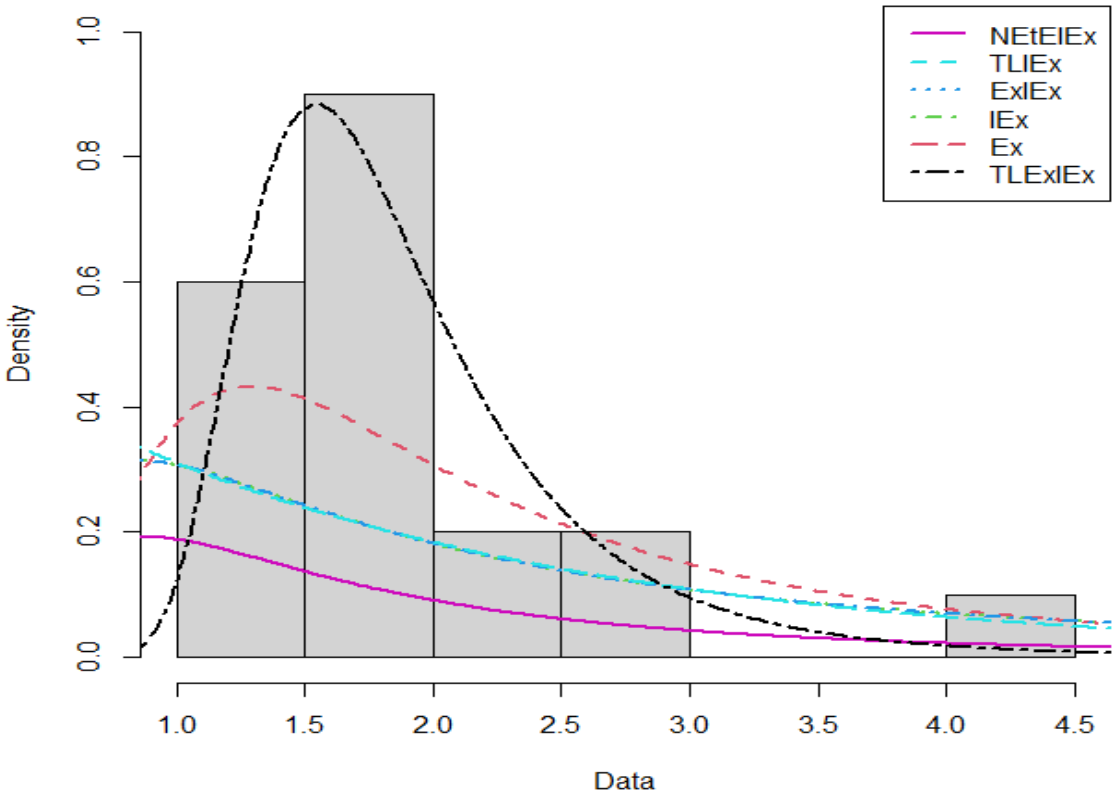


Figure 5: *Histogram and fitted pdfs for the NetEIEx, TLIEx, ExIEx, TLExIEx, IEx and Ex models for the first data set*

Table 2: *The ML estimates and goodness of fit measurement for the second data set*

Models	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	$-l$	AIC	BIC
NEtEIEx	0.096	13.103	79.636	30.736	955.251	1918.502	1931.735
TLExIEx	3.883	1.789	6.995	-	1521.690	3049.381	3059.305
ExIEx	9.867	5.771	-	-	1055.772	2115.544	2122.160
TLIEx	-	25.015	6.607	-	980.481	1964.962	1971.578
IEx	-	56.953	-	-	1055.772	2113.544	2116.852
Ex	-	0.014	-	-	1057.353	2116.707	2120.015

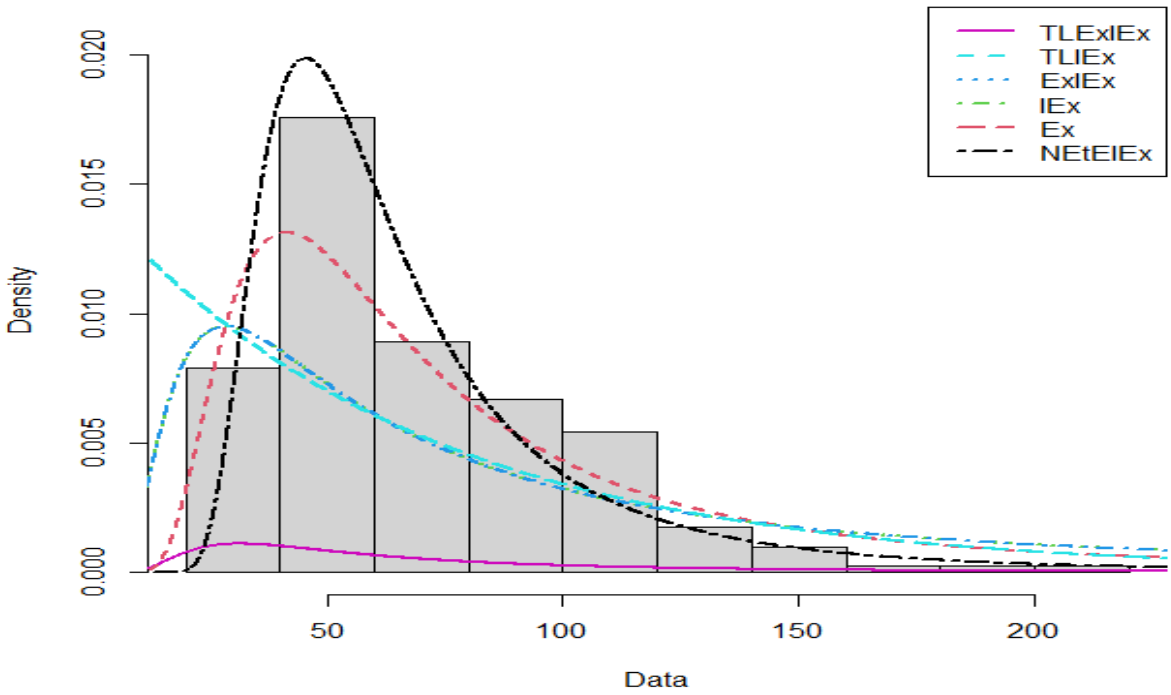


Figure 6: *Histogram and fitted pdfs for the NetEIEx, TLIEx, ExIEx, TLExIEx, IEx and Ex models for the second data set*

IV. Discussion

The estimated values for each parameter and the models' goodness of fits are shown in Tables 1 and 2. AIC and BIC are two metrics for goodness of fits. The model performs better when the AIC and BIC values are lower. Tables 1 and 2 show that the NEtEIEx distribution has the lowest AIC and BIC, respectively. This property makes the new model more adaptable and suitable for handling biomedical data sets.

The new model's forms, fit, and adaptability in connection to the data sets under consideration are shown in Figures 5 and 6. The black line, which represents the new model, more closely matched the data's pattern than the competitors. The histogram and fitted plots make it clear that the black line, which represents the NEtEIEx distribution, matches the two data sets under consideration better.

This study extends the inverse exponential distribution by creating a new continuous distribution known as the new exponentiated extended inverse exponential distribution. It was possible to obtain the survival function, hazard rate function, quantile function, inverted hazard function, odds function, and order statistics from the new distribution. Plotting the pdf and hazard rate function graphs revealed the contours of the suggested distribution. It was discovered that the hazard function is shaped like an upside-down bathtub. Adequacy Model is a package in R that was used to estimate the model parameters using the maximum likelihood method. The proposed distribution was applied to two real life data sets, and the outcomes are shown in Tables 1 and 2. The findings demonstrated that the new extended inverse exponential distribution with exponentiation is much more potent and superior at fitting the two data sets under consideration. The density graphs in figures 5 and 6 for the two data sets further show how adaptable the new model is.

References

[1] Keller, A.Z. and Kamath, A.R. (1982). Reliability analysis of CNC Machine Tools. *Reliability Engineering*, 3, 449-473.

[2] Singh, S.K., U. Singh and M. Kumar, (2013). Estimation of parameters of generalized inverted exponential distribution for progressive type-II censored sample with binomial removals. *Journal of Probability and Statistics*, Article ID 183652.

[3] Lin, C.T., B.S. Duran and T.O. Lewis, (1989). Inverted gamma as a life distribution. *Microelectron. Reliab.*, 29: 619-626.

[4] Cordeiro, G. M., and de Castro, M., (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81(7): 883–898.

[5] Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S. and Shahbaz, M. Q., (2016). Topp-Leone family of distributions: Some properties and application. *Pakistan Journal of Statistics and Operation Research*, XII, 3, 443-451.

[6] Elgarhy, M., Muhammad, A. H., Gamze, O. and Muhammad, A. N. (2017). A new exponentiated extended family of distributions with applications. *Gazi University Journal of Science*, 30(3): 101 - 115

[7] Hassan, S. and Nassr, S.G., (2019). Power Lindley-G Family of Distributions. *Annals of Data Science*, 6: 189-210.

[8] Ibrahim, S., Doguwa, S.I., Audu, I. and Jibril, H.M., (2020). On the Topp Leone exponentiated-G Family of Distributions: Properties and Applications. *Asian Journal of Probability and Statistics*, 7(1): 1-15.

[9] Ibrahim, S., Doguwa S. I., Audu, I. and Jibril, H. M., (2020). The Topp Leone Kumaraswamy-G Family of Distributions with Applications to Cancer Disease Data. *Journal of Biostatistics and Epidemiology*, 6(1):37-48.

[10] Anzagra, L., Sarpong, S. and Nasiru, S., (2020). Odd Chen-G family of distributions. *Annals of Data Science*, doi.org/10.1007/s40745-020-00248-2.

[11] Modi, K., Kumar, D. and Singh, Y., (2020). A New Family of Distribution with Application on Two Real Data sets on Survival Problem. *Science and Technology Asia*, 25(1): 1-10.

[12] Rasheed, N. (2020). A new generalized-G class of distributions and its applications with Dagun distribution. *Research Journal of Mathematical and Statistical Science*, 8(3), 1-13.

[13] Bello, O. A., Doguwa, S. I., Yahaya, A. and Jibril, H. M. (2020) A type I half Logistic exponentiated-G family of distributions: Properties and application. *Communication in Physical Sciences*, 7(3): 147–163.

[14] Bello, O. A., Doguwa, S. I., Yahaya, A. and Jibril, H. M. (2020) A type II half Logistic exponentiated-G family of distributions with applications to survival analysis. *FUDMA Journal of Sciences*, 7(3): 147 – 163.

[15] Eraikhuemem, I. B., Onwuka G, I., Yakura B. S., and Allahde H. (2021). An exponentiated odd Lindley inverse exponential distribution and its application to infant mortality and HIV transmission rates in Nigeria. *International STD Research and Reviews*, 10, 12-30

[16] Bashiru, O. S. (2021). A new extended generalized inverse exponential distribution: properties and applications. *Asian Journal of Probability and Statistics*, 11(2), 30-46

[17] Oguntunde, P.E., Babatunde, O.S. and Ogunmola, A.O. (2014).Theoritical Analysis of the Kumaraswamy Inverse Exponential Distribution. I*nternational Journal of Statistics and Applications*, 4(2) : 113-116.

[18] Ibrahim, S., Akanji, B. O. and Olanrewaju, L. H. (2020). On the extended generalized inverse exponential distribution with its applications. *Asian Journal of Probability and Statistics*, 7(3):14-27.

[19] Adekunle, I. K., Sule, I., Doguwa, S. I. and Yahaya, A. (2023). On the Properties and Applications of Topp-Leone Kumaraswamy Inverse Exponential Distribution. *Communication in Physical Sciences*, 8(4): 590- 603.

[20] Rasool, S.U and Ahmad S.P. (2022). Power Length biased weighted lomax distribution. *Reliability: Theory & Applications*, 4(71): 543-558.

[21] Sule, O.Y. and Halid, O.Y. (2023). On Gompertz Exponentiated Inverse Rayleigh Distribution. *Reliability: Theory & Applications*, 1(72): 412-424.

[22] Hosseini, B., Afshari, M. and Alizadeh M. (2018). The generalized odd gamma-G family of distributions: properties and applications. *Austrian Journal of Statistics*, 47(2):69-89.

ON CHANGE-POINT ANALYSIS OF MAXWELL DISTRIBUTION USING BAYESIAN TECHNIQUES

*[1]Taiwo M. Adegoke & [2]Oladapo M. Oladoja

•

[1,2]Department of Mathematics and Statistics, First Technical University, Ibadan, Nigeria.
[1]taiwo-adegoke@tech-u.edu.ng, [2]oladapo.oladoja@tech-u.edu.ng

Abstract

This research work focuses on Bayesian inference in this study to detect a change in the rate of a Maxwell distribution model with independent random variables. The paper specifically analyzes a single rate shift and demonstrates how the Bayesian framework can be used to efficiently solve this problem. To produce samples from Maxwell distribution and evaluate the datasets, simulation techniques were used, and the R programming language was used. Although the model looks to be simple, no analytical solutions are available for parameter inference, necessitating the use of approximations. The study emphasizes the Gibbs sampler's applicability for change-point analysis using a Markovian updating approach. The simulation research findings show that the predicted rate is near to the true value, confirming the consistency and stability of the Bayesian estimator.

Keywords: Gibbs sampling, Change-point, Bayes factor, Bayesian method, Conjugate prior distribution

1. Introduction

Change-point analysis (CPA) is a statistical technique for detecting and quantifying changes in data across time. CPA identifies data points when there is a significant shift in the underlying structure or behavior of the process producing the data. CPA finds applications in various domains. It is utilized in fault detection and reliability [1], insurance, econometric timeseries, and malware software detection [2]. Furthermore, it plays a role in signal detection, surveillance, security systems, meteorology, and climatology [3] . Changepoint analysis is also employed in graphical models [4], gynecology [5], communication network evolution [6, 7, 8], oceanography [9], sparse VAR models [10, 11], macrosociological processes and historical changes [12], medicine [13], and functional magnetic resonance recordings [14, 15] , among others.

CPA can be traced back to the work of [16, 17, 18], where cumulative sums (CUMSUM) approach was used to identify points of change in a sequence of normally distributed observations. Since then, several methods have been proposed for performing CPA, including Bayesian methods (see [19, 20, 21, 22, 23, 24, 25], likelihood-based methods (see [26, 27, 28, 29, 30, 31] and non-parametric methods (see [32, 33, 34, 35]). These techniques involve various statistical models and algorithms to estimate the change points accurately. The choice of method depends on the characteristics of the data and the specific objectives of the analysis.

In their study,[36] investigated the change point analysis (CPA) in the Maxwell distribution using Bayesian methods. They examined both informative and non-informative prior distributions and utilized two distinct loss functions, namely Linex LF and General Entropy LP, to detect the change point and estimate its magnitude. However, the present work introduces a novel approach by utilizing Bayes' factor techniques to detect a single change point in a series of observations following a Maxwell distribution. The method utilizes a conjugate prior distribution and employs a Monte Carlo Gibbs sampling approach to estimate the parameters involved.

2. Method

2.1. Bayesian Techniques

The definition of a change-point, as proposed by [16, 17, 18] and [26], involves a test to determine whether a sequence of independent observations, arranged in a successive order $x_1, x_2, \ldots, x_n$, are drawn from the same probability density function $F(x|\theta)$, which is characterized by the likelihood function.

$$L(x;\theta) = \prod_{i=1}^{n} f(x;\theta) \tag{1}$$

as against set of observations with a single change-point k represented as $x_1, x_2, \ldots x_k$ and $x_{k+1}, x_{k+2}, \ldots x_n$ before and after the change having two distinct probability density functions $F(x|\theta_1)$ and $F(x|\theta_2)$, where $\theta_1 \neq \theta_2$. The likelihood function for the alternate hypothesis can be expressed as

$$L(x;\theta_1,\theta_2) = \prod_{i=1}^{k} f(x;\theta_1) \prod_{i=\gamma+1}^{n} f(x;\theta_2) \tag{2}$$

In the Bayesian perspective, a joint prior distribution $p(\theta_1,\theta_2)$ is assumed for the parameters. Bayes theorem then provides the joint posterior distribution

$$p(\theta_1,\theta_2,\gamma|x,y) = \frac{p(x,y|\theta_1,\theta_2)p(\theta_1,\theta_2)}{\int\int p(x,y|\theta_1,\theta_2)p(\theta_1,\theta_2)\partial\theta_1\partial\theta_2} \tag{3}$$

The prior distribution $p(\theta_1,\theta_2)$ reflects the beliefs about the parameters before experimentation, whereas the posterior distribution $p(\theta_1,\theta_2,k|x)$ reflects the updated beliefs about the parameters after observing the sample data.

2.2. Bayes' Factor

Bayesian statisticians perceive hypothesis testing as a process of comparing models ([37], [38]) rather than focusing on whether a specific hypothesis is true, the emphasis is placed on determining which model, described under one hypothesis, is more favorable compared to another. The Bayesian approach to hypothesis testing was initially developed by [39, 40] as a fundamental component of scientific inference. A central aspect of Jeffreys' framework involved the concept of the *Bayes factor*, which represents the posterior odds of the null hypothesis when the prior probability on the null is one-half. Jeffreys employed this approach to compare predictions made by two competing scientific theories. In this methodology, statistical models are introduced to represent the likelihood of the observed data according to each theory, and Bayes' theorem is utilized to compute the posterior probability that one theory is correct.

In their study, [37] consider a dataset D, which is assumed to be generated under two hypotheses: H_1 and H_2. The probability densities $\zeta(D/H_1)$ and $\zeta(D/H_2)$ describe the data under each hypothesis, respectively. Prior probabilities, $\zeta(H_1)$ and $\zeta(H_2) = 1 - \zeta(H_1)$, are assigned to H_1 and H_2, respectively. By applying Bayes' theorem, the authors obtain the posterior probabilities $\zeta(H_1/D)$ and $\zeta(H_2/D)$ as follows:

$$\zeta(H_i/D) = \frac{\zeta(D/H_i)\zeta(H_i)}{\zeta(D/H_1)\zeta(H_1) + \zeta(D/H_2)\zeta(H_2)}, \quad (i = 1,2) \tag{4}$$

$$\zeta(H_1/D) = \frac{\zeta(D/H_1)\zeta(H_1)}{\zeta(D/H_1)\zeta(H_1) + \zeta(D/H_2)\zeta(H_2)}$$
$$\zeta(H_2/D) = \frac{\zeta(D/H_2)\zeta(H_2)}{\zeta(D/H_1)\zeta(H_1) + \zeta(D/H_2)\zeta(H_2)}$$

In certain applications, such as testing hypotheses regarding the presence of a change-point, it is often more informative to consider the odds in favor of H_2 compared to H_1 ([41, 42]).

$$\frac{\zeta(H_2/D)}{\zeta(H_1/D)} = \frac{\zeta(D/H_2)}{\zeta(D/H_1)} \times \frac{\zeta(H_2)}{\zeta(H_1)} \tag{5}$$

and the transformation is simply multiplication of the prior odds by

$$B_{12} = \frac{\zeta(D/H_2)}{\zeta(D/H_1)} = \frac{\int_{\theta_2} \zeta(D/\theta_2)\zeta(\theta_2)\mathrm{d}\theta_2}{\int_{\theta_1} \zeta(D/\theta_1)\zeta(\theta_1)\mathrm{d}\theta_1} \tag{6}$$

which is the baye's factor. Thus,in words,
posterior odd = bayes factor $\times$ prior odds
and the bayes factor is the ratio of the posterior odd of H_1 to its prior odds, regardless of the value of the prior odds.

By analogy with the likelihood ratio obtained from Equation (6) (i.e the quantity $\log B_{12}$) is often used to summarize the evidence for H_2 compare to H_1, with the rough interpretation shown in Table 1. This contrasts with the interpretation of a likelihood ratio, whose null χ^2 distribution for nested models would depend on the difference in their degree of freedom p ([37, 39, 40, 41, 42]). The log Bayes factor $2\log B_{12}$ is sometimes called the weight of evidence

Table 1: *Rough Interpretation of Bayes factor B_{12} given by Davison(2003) and Peter(2006)*

B_{12}	$2\log_e B_{12}$	Interpretation
Under 1	Negative	Supports model 1
1 - 3	0 - 2	Weak support for model 2
3 - 20	2 - 6	Support for model 2
20 - 150	6 - 10	Strong evidence for model 2
Over 150	Over 10	Very strong support for model 2

2.3. The Proposed Change Point Model

This section introduces a change-point model based on the Maxwell distribution. Consider a series of observations of size n (n>3) drawn from a Maxwell distribution with parameter θ whose null hypothesis H_1 can be stated as

$$H_1 : \theta_1 = \theta_2 = \theta \tag{7}$$

whose likelihood function can be expressed as

$$f(x \mid \theta) = \prod_{i=1}^{n} \frac{\pi}{2\theta^{2/3}} x^2 e^{-2\theta x^2} \theta > 0 \tag{8}$$

This means that the there is no shift in the parameter θ of the model.
Also, consider a series of observations $x_1, x_2, \ldots x_k, x_{k+1}, \ldots x_n$ with a single shift at point k drawn from different population with parameters θ_1 and θ_2. The alternate hypothesis can be stated as

$$H_2 : \theta_1 \neq \theta_2 \tag{9}$$

having the likelihood function

$$f(x \mid \theta) = \prod_{i=1}^{k} \frac{\pi}{2\theta_1^{2/3}} x^2 e^{-2\theta_1 x^2} \prod_{i=1}^{k} \frac{\pi}{2\theta_2^{2/3}} x^2 e^{-2\theta_2 x^2} \tag{10}$$

2.4. Bayesian Analysis for the Change-Point Model

For the no change-point model (8), we consider a conjugate prior distribution for the parameter θ having $Gamma(\theta|a,b)$ with probability density function and uniform prior for the parameter k with parameter value Uniform(1,n).

$$\pi(\theta) = \frac{b^a}{(n-1)\Gamma(a)}\theta^{a-1}e^{-b\theta} \qquad a>0, b>0 \tag{11}$$

The posterior distribution for the null hypothesis model (7) can be obtained by combining the likelihood function (8) with the prior distribution (11) given in Equation (3) as

$$\pi(\theta|x) \propto \frac{\left(b+\frac{\sum_{i=1}^{n} x-i^2}{2}\right)^{\frac{3n}{2}+a}\Gamma(a)}{\Gamma\left(\frac{3n}{2}+a\right)b^a} \tag{12}$$

Also, for the alternate hypothesis model (9), we consider a conjugate prior distribution for θ_1 and θ_2 having $Gamma(\theta_1|a_2,b_2)$ and $Gamma(\theta_2|a_2,b_2)$ having probability density function

$$\pi(\theta_1\theta_2) = \frac{b_1^{a_1}}{\Gamma(a_1)}\theta^{a_1-1}e^{-b_1\theta_1}\frac{b_2^{a_2}}{\Gamma(a_2)}\theta^{a_2-1}e^{-b_2\theta_2} \qquad a_1>0, b_1>0, a_2>0, b_2>0 \tag{13}$$

The posterior density function can be derived by combining the likelihood function (10) and prior distribution (13) and thus we have

$$\pi(\theta_1,\theta_2|x) \propto \frac{\left(b_1+\frac{\sum_{i=1}^{k} x_i^2}{2}\right)^{\frac{3k}{2}+a_1}\Gamma(a_1)\left(b_2+\frac{\sum_{i=k+1}^{n} x_i^2}{2}\right)^{\frac{3(n-k)}{2}+a_2}\Gamma(a_2)}{\Gamma\left(\frac{3k}{2}+a_1\right)b_1^{a_1}\Gamma\left(\frac{3(n-k)}{2}+a_2\right)b_2^{a_2}} \tag{14}$$

3. Results and Discussions

In this section, we carried out a simulation studies to demonstrate the proposed change-point model. To conduct the diagnostic successfully, we generated five sequences (chains), each consisting of 30,000 elements. A burn-in period of 10,000 observations was implemented, and thinning was applied, considering every 100th observation, using the Markov Chain Monte Carlo (MCMC) scheme.

3.1. Simulation Study

We simulated datasets having a single shift in the parameter θ drawn from a Maxwell distribution with predefined values expressed in model (15)

$$x_i \sim \begin{cases} dMax(1.5) & 1 \le i \le 41 \\ dMax(0.5) & 41 < i \le 80 \end{cases} \tag{15}$$

Table 2: *Summary Statistics for the Posterior Quantities*

Parameters	Mean	SD	Mc_Error	97.5% Credible Interval
k	41.1	1.137	0.0038	[0.000002–0.0000002]
θ_1	1.77	0.2321	0.0019	[1.348 – 2.255]
θ_2	0.564	0.0743	0.0005	[0.427 – 0.7194]

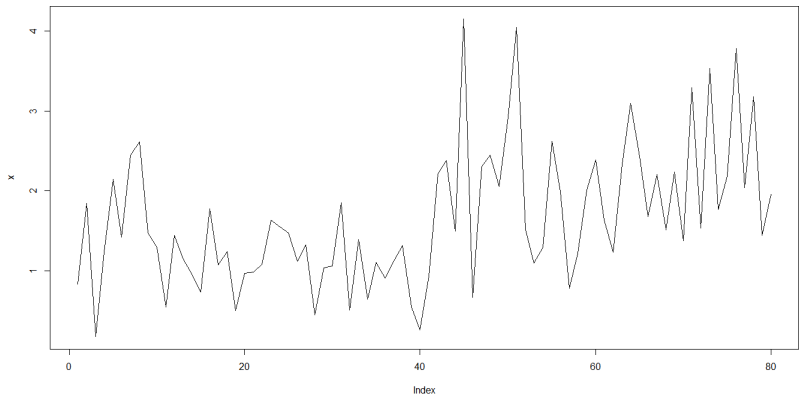


Figure 1: *Line plot for the simulation study*

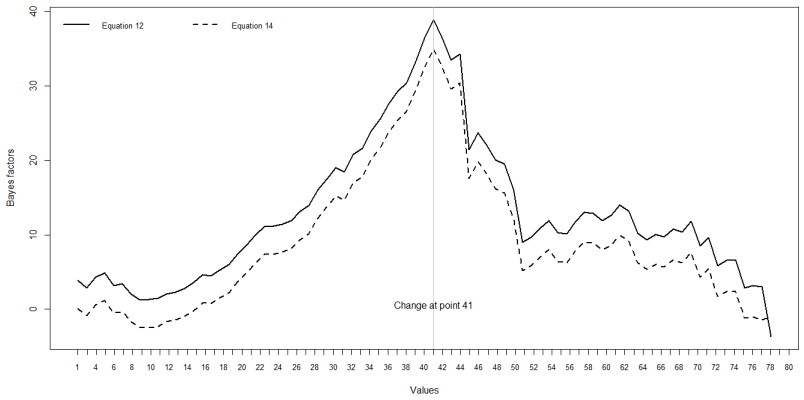


Figure 2: *Bayes Factor Plot for the simulated dataset*

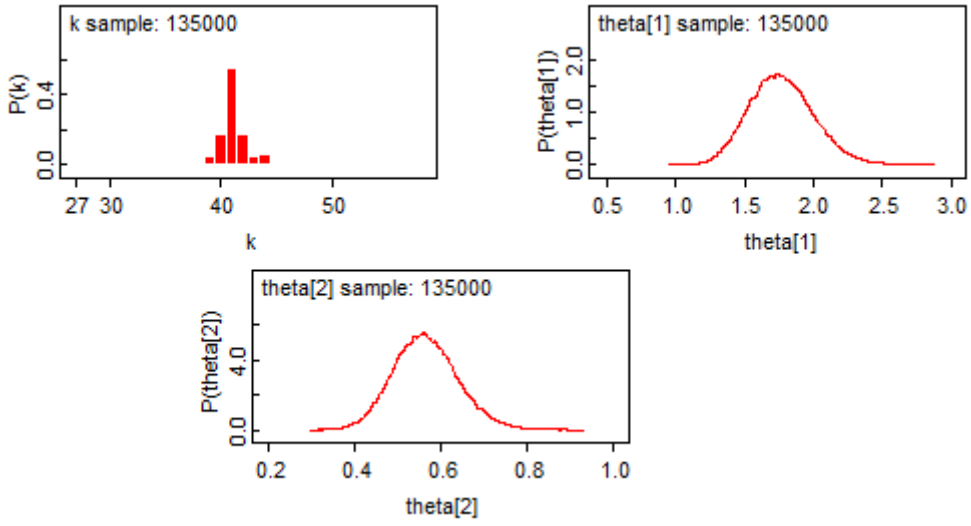


Figure 3: *Posterior densities for the parameter θ_1, θ_2 and k*

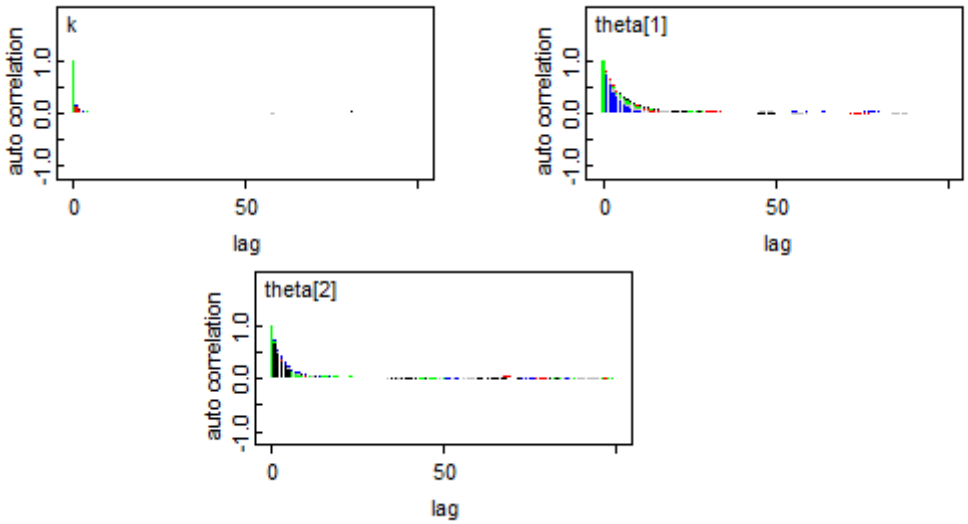


Figure 4: *Autocorrection plot for the Posterior densities for the parameter θ_1, θ_2 and k*

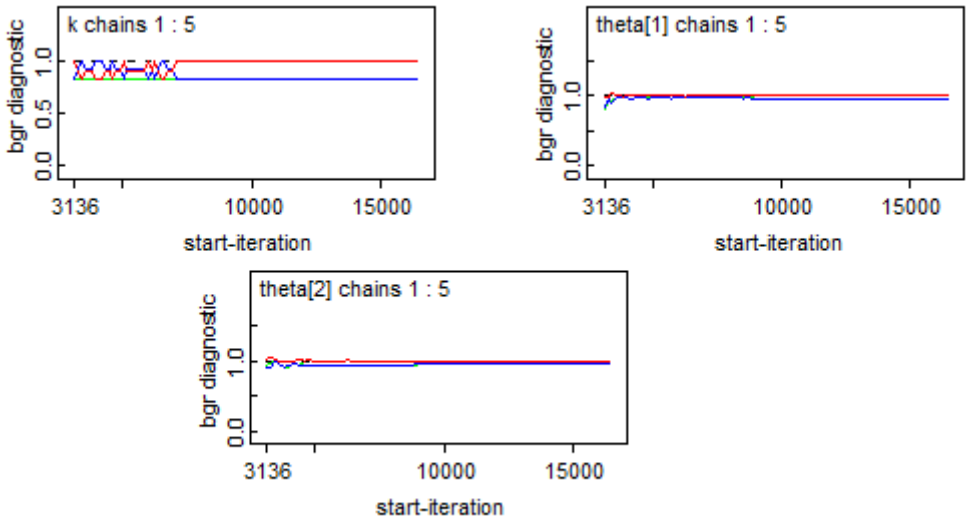


Figure 5: *Brooks-Gelman-Rubin plot for the Posterior densities for the parameter θ_1, θ_2 and k*

Figures 1 and 2 depict the line plot and the Bayes' factor plot for the simulated dataset. From Figure 2, we determined that the shift point was detected at the predefined point 41. Summary statistics from the Gibbs sampling MCMC are shown in Table 2, revealing that the change point k was identified at approximately 41. Posterior densities for all parameters are displayed in Figure 3, confirming that the density of k indicates a change occurring around point 41. The autocorrelation plot in Figure 4 demonstrates noticeable autocorrelation in lag 1 for all parameters. In Figure 5, the Brooks-Gelman plot for the posterior quantities suggests that the chain moves randomly from one iteration to the next, with the Brooks-Gelman-Rubin (BGR) plot for each parameter closely approaching 1. According to Gelman (2003), an acceptable limit is 1.1. Therefore, the BGR plots indicate excellent results. Considering the evidence from Figures 4 and 5, we conclude that the results obtained from the Gibbs sampler exhibit convergence to the posterior distribution and are accurate.

4. Discussion

In this paper, we introduce a single change-point model for datasets that follow a Maxwell distribution, using informative Bayes' Factor techniques. We employ the Bayesian method to detect the time at which a shift occurs in the dataset and apply this approach to both simulated datasets. One key advantage of the Bayesian approach over the frequentist approach is its ability to estimate uncertainty without relying on asymptotic sampling arguments that require large sample sizes. The main objective of this research is to develop a change-point model for detecting a single change-point in a series of observations that follow a Maxwell distribution. We accomplish this using a Bayesian method, which provides a more objective approach compared to subjective methods.

References

[1] Spokoiny, V. Multiscale Local Change Point Detection with Application to Value Art Risk . *Annual Of Statistics*. **37**, 1405-1436 (2009)

[2] Yan, G., Xiao, Z. & Eidenbenz, S. Catching instant messaging worms with change-point detection technique. *In Proceedings Of The USENIX Workshop On Large-scale Exploits And Emergent Threats*. (2008)

[3] Jaxk, R., Cheng, J., Wang, X., Lun, R. & Qiqi, L. A Review and Comparison of Change Point Detection for Climate Data. *Journal Of Applied Meteology And Climatology*. **6** pp. 900-915 (2007)

[4] Londschien, M., Kovacs, S. & Buhlmann, P. Change-point detection for graphical models in the presence of missing values. *Journal Of Computational And Graphical Statistics*. pp. 1-12 (2021)

[5] Erdman, C. & Emerson, J. A Fast Bayesian Change Point Analysis for the Segementation of Micro Arrary Data. *Bioinformatics*. **24**, 2143-2148 (2008)

[6] Kossinets, G. & Watts, D. Empirical analysis of an evolving social network. *Science*. **311** pp. 88-90 (2006)

[7] Eagle, N., Pentland, A. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings Of The National Academy Of Sciences*. pp. 15274-15278 (2009)

[8] Peel, L. & Clauset, A. Detecting change points in the large-scale structure of evolving networks. *In AAAI*. pp. 2914-2920 (2015)

[9] Killick, R., Eckley, I., Jonathan, P. & Ewan, K. Detection of Changes in the Characteristics of Oceanographic Time-Series Using Statistical Change Point Analysis. *Ocean Engineering*. **37**, 1120-1126 (2010)

[10] Wang, D., Yu, Y., Rinaldo, A. & Willett, R. Localizing changes in high-dimensional vector autoregressive processes. *ArXiv Preprint ArXiv:1909.06359* . (2019)

[11] Safikhani, A. & Shojaie, A. Joint Structural Break Detection and Parameter Estimation in High-Dimensional Nonstationary VAR Models. *Journal Of The American Statistical Association*. **117**, 251-264 (2022)

[12] Isaac, L. & Griffin, L. A historicism in time-series analyses of historical process: Critique, redirection, and illustrations from u.s. labor history. *American Sociological Review*. **54**, 873-890 (1989)

[13] Taylor, C. Change-Point Analysis: A Powerful New Tool For Detecting Changes.. (2010)

[14] Barnett, I. & Onnela, J. Change point detection in correlation networks. *Scientific Reports*. pp. 1-11 (2016)

[15] Zambon, D., Alippi, C. & Livi, L. Change-point methods on a sequence of graphs. *IEEE Transactions On Signal Processing*. **67** pp. 6327-6341 (2019)

[16] Page, E. Continous Inspection Schemes. *Biometrika*. **41** pp. 100-115 (1954,6)

[17] Page, E. A Test for a Change in a Parameter Occuring at an Unknown Point. *Biometrika*. **42** pp. 523-527 (1955)

[18] Page, E. On Problem in which a Change in Parameter Occurs at an Unknown Point. *Biometrika*. **44** pp. 248-252 (1957)

[19] Obisesan, K. Modelling multiple changepoints detection. (University of Ibadan,2015)
[20] Son, Y. & Kim, S. Bayesian single change point detection in a sequence of multivariate normal observations.. *Statistics*. **39**, 373-387 (2005)
[21] Perreault, L., Hache, M., Slivitsky, M. & Bobee, B. Detection of changes in precipitation and runoff over eastern Canada and US using a Bayesian approach. *Stochastic Environmental Research And Risk Assessment*. **13** pp. 201-216 (1999)
[22] Adegoke, T., Bakari, H. & Yahya, A. Bayesian approach for change-point detection of exponential models. *International Journal Of Basic And Applied Sciences*. **17**, 1-6 (2017)
[23] Adegoke, T. & Yahya, ;. A Non-informative Approach to Change-point Detection in a Sequence of Normally Distributed Data with Applications. *Annals Of Statistical Theory And Applications*. **1** pp. 61-70 (2019), www.pssng.org
[24] Adegoke, T. & Yahya, W. A Bayesian Multiple Change-point Analysis: an Application to Air Temperature and Rainfall Data. *Proceedings Of 2nd International Conference Of Professional Statitsical Society Of Nigeria*. **2** pp. 322-327 (2018)
[25] Yahya, W., Obisesan, K. & Adegoke, T. Bayesian Change-point Modelling of Hydrometeorological Data In Nigeria. *Proceedings Of 1st International Conference Of Nigeria Statistical Society*. **1** pp. 59-63 (2017)
[26] Hinkley, D. Inference about the change-point in a sequence of random variables. *Biometrika*. **57** pp. 1-17 (1970)
[27] Fotopoulos, S. & Jandhuala, V. On Hinkley's estimator: Inference about the change point. *Statistics And Probability Letters*., 1449-1458 (2007)
[28] Bisai, D., Chatterjee, S., Khan, A., Trend, N. & Midnapore Weather Observatory, I. Statistical Analysis of Trend and Change Point in Surface Air Temperature Timeseries for Midnapore Weather Observatory, West Bengal, India. *Hydrology Current Research*. **5**, 1-7 (2014)
[29] Liu, P. Maximum Likelihood Estimation of an Unknown Change-Point in the Parameters of a Multivariate Gaussian Series with Applications to Environmental Monitoring. . (Washington State University.,2010)
[30] Dibal, N., Mustapha, M., Adegoke, T. & Yahaya, A. Statistical change point analysis in air temperature and rainfall time series for cocoa research institute of Nigeria, Ibadan, Oyo State, Nigeria. *International Journal Of Applied Mathematics And Theoretical Physics*. **3**, 92-96 (2017)
[31] Ninomiya, Y. Construction of conservative test for change-point problem in two-dimensional random fields. *Journal Of Multivariate Analysis*. pp. 219-242 (2004)
[32] Račkauskas, A. & Suquet, C. Hölder norm test statistics for epidemic change. *Journal Of Statistical Planning And Inference*., 495-520 (2004)
[33] Miller, M. Adapting to Climate Change: Water Management for Urban Resilience. *Environ Urban*. **19** pp. 99-113 (2007)
[34] Grégoire, G. & Hamrouni, Z. Change point estimation by local linear smoothing. *Journal Of Multivariate Analysis*., 56-83 (2002)
[35] Horváth, L. & Kokoszka, P. Change-point detection with non-parametric regression. *Statistics*., 9-31 (2002)
[36] Pandya, M. & Pandya, H. Bayes Estimation of Change Point in Discrete Maxwell Distribution.. *International Journal Of Quality, Statistics, And Reliability*. pp. 1-9 (2011)
[37] Kass, R. & Raftery, A. Bayes Factor. *Journal Of The American Statistical Association*., 773-779 (1995)
[38] Berger, J. & Pericchi The intrinstic Bayes FActor for model selection and prediction. *Journal Of Amer. Stat. Ass.*. **91** pp. 109-202 (1996)
[39] Jeffreys, H. Theory of Probability. (Clarendon Press,1939)
[40] Jeffreys, H. Theory of Probability. (Clarendon Press,1961)
[41] Congdon, P. Bayesian Statistical Modelling. (John Wiley & Sons Ltd,2006)
[42] Davison, A. Statistical Models. (Cambridge University Press,2003)
[43] Gelman, A., and Hill, L.H. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

RT&A, No 3 (74)
Volume 18, September 2023

ANALYSIS OF UNCERTAINTY WEIGHTED MEASURES FOR PARETO II DISTRIBUTION

Baria A. Helmy[1], Amal S. Hassan[2], Ahmed K. El-Kholy[1]

•

[1]Faculty of Science, Al-Azhar University (girl's branch), Cairo, 11651, Egypt
BareaHelmy.2059@azhar.edu.eg
[2]Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, 12613, Egypt
amal52_soliman@cu.edu.eg

Abstract

Extropy is a complimentary dual of Shannon's entropy, which has many applications. The maximum likelihood and Bayesian approaches are used in this article to explore the weighted extropy and weighted residual extropy of the Pareto type II distribution. Using unified hyper-censoring data, we calculate the maximum likelihood estimation of weighted extropy and its residual measures. Based on symmetric and asymmetric loss functions, Bayesian estimators of weighted extropy and its residual measure are developed based on unified hyper censoring data. To do some complex calculations, Markov chain Monte Carlo methods are used. To test the performance of the estimators, a Monte Carlo simulation study and an illustration using real data sets were carried out. The outcomes of the study showed that as the sample size increases, maximum likelihood and Bayesian estimators of the weighted extropy and its residual measure perform well. Also, Bayesian estimators of the weighted extropy and its residual under the general entropy loss function are superior to the Bayesian estimators under the others in most cases. Theoretical and empirical findings are generally in good agreement.

Keywords: weighted extropy, weighted residual extropy, Bayesian estimate, Markov chain Monte Carlo method.

1. Introduction

The Shannon entropy, or differential entropy of a random variable Z with the probability density function (PDF) $f(z)$ is a fundamental notion in measuring discrimination and information. It is defined as

$$H(Z) = -\int_{-\infty}^{\infty} f(z) \log f(z) dz. \tag{1}$$

Shannon's entropy measure has become one of the most widely used uncertainty measurements, with several applications in various fields like reliability, survival analysis, and actuarial sciences, among others.

Entropy measurement has benefits in a variety of fields, including industrial engineering and the financial performance of the companies (see Marvizadeh [1] and Liew et al. [2]). Estimation studies for Shannon entropy with various censoring and distribution strategies can be found in Cho et al. [3], Liu and Gui [4], Hassan and Zaky [5], and Yu et al. [6]. The Bayesian estimator of dynamic residual entropy was investigated by Ahmadini et al. [7], Al-Babtain et al. [8], and Almarashi et al. [9]. Some researchers have looked at estimating entropy measures based on record data; see, for instance, Hassan and Zaki [10] and Al-Omari et al. [11]. Helmy et al. [12] studied the Shannon entropy for Lomax distribution in the context of unified hybrid censored

samples. Hassan et al. [13] studied estimation of information measures for power-function distribution in the presence of outliers.
Lad et al. [14] recently demonstrated the complementary dual of entropy called extropy, a different measure of uncertainty. The positive and negative pictures of a photographic film are related to entropy and extropy metrics. Extropy is utilized in voice recognition and to score the predicting distribution (see Becerra et al. [15]). Extropy, often known as differential extropy, is defined as

$$J(Z) = \frac{-1}{2}\int_0^\infty f^2(z)dz. \tag{2}$$

Qiu [16] recently investigated the characterization results, lower bounds, monotone and symmetric features of order statistics extropy, and record values. Extropy features of ranked set sampling were investigated by Raqab and Qiu [17]. In terms of extropy, Qiu et al. [18] studied information characteristics of mixed systems. Extropy, on the other hand, is ineffective for estimating the remaining lifetime of a unit that has lived for some units of time, $Z_t = [Z - t|Z \geq t]$ is the residual life function of Z at time t. As a result, Qiu and Jia [19] suggested the residual extropy to assess the residual uncertainty of Z_t and described the characterization and monotonic features of order statistics:

$$J_t(Z) = \frac{-1}{2\overline{F^2}(t)}\int_t^\infty f^2(z)dz. \tag{3}$$

where $\overline{F}(z) = 1 - F(z)$ is the survival function of Z. The fundamental disadvantage of the preceding information measures is that they only consider the probability density of the random variable rather than the values it takes. On the right side of (2), the integrand measure is shift-independent since it is only reliant on z through $f(z)$. This shift-independent quality, on the other hand, appears to be a disadvantage in many applications, such as mathematical neurobiology and reliability. In such cases, the random variable's value, as well as the probabilities, should be taken into consideration. Analogous to the weighted entropy see (Belis and Guiasu [20]), in order to efficiently model statistical data Balakrishnan et al. [21] introduced a new measure of information named weighted extropy (WEx). For a non-negative random variable Z with PDF is $f(z)$, the WEx is defined as

$$J^w(Z) = \frac{-1}{2}\int_0^\infty zf^2(z)dz. \tag{4}$$

Now we'll look at two distributions that have the same extropy but differently weighted extropies. Let X and Y be two random variables such that $X \sim U(a,b)$, $Y \sim U(2a, a+b)$, where $a, b > 0$. We have $f_X(x) = \frac{1}{b-a}$, for $x \in (a,b)$, and $f_Y(y) = \frac{1}{b-a}$, for $y \in (2a, a+b)$, and then

$$J(X) = \frac{-1}{2}\int_a^b \frac{1}{b-a}dx = \frac{-1}{2}, J(Y) = \frac{-1}{2}\int_{2a}^{b+a} \frac{1}{b-a}dx = \frac{-1}{2}.$$

and

$$J^w(X) = \frac{-1}{2}\int_a^b x\frac{1}{b-a}dx = \frac{-1}{2(b-a)}[\frac{b^2}{2} - \frac{a^2}{2}] = \frac{-(b+a)}{4}$$

$$J^w(Y) = \frac{-1}{2}\int_{2a}^{b+a} x\frac{1}{b-a}dx = -\frac{(b+a)^2 - (2a)^2}{4(b-a)} = \frac{-(b^2+2ba-3a^2)}{4(b-a)}.$$

Extropies are the same in the two cases, but weighted extropies are different, hence weighted extropies can also be used as a measure of uncertainty. The concept of residual and the past life of random variables were combined to create WEx, given the necessity of weighted measures as previously described. When an item is working at time t, it may be worthwhile to investigate its longevity beyond that time. The residual lifetime is the set of interest in such cases, thus the concept of weighted residual extropy (WREx), which is defined as follows:

$$J_t^w(Z) = \frac{-1}{2\overline{F^2}(t)}\int_t^\infty zf^2(z)dz. \tag{5}$$

In the literature, we couldn't find any research on WEx and WREx estimation problems using a unified hyper censoring scheme (UHCS). In this paper, UHCS is used for estimating weighted

extropy and its residual for Pareto II distribution (P-IID). The maximum likelihood (ML) and Bayesian estimators are used to investigate WEx and WREx measures. The Bayesian estimator is provided using symmetric and asymmetric loss functions and uses the Markov chain Monte Carlo (MCMC) method to estimate the posterior distribution. Both the application to real data and simulation concerns are covered.

The rest of the paper is organized in the following way. The WEx and WREx expressions of the P-IID are produced in Section 2. In Section 3, using ML and Bayesian estimation methods to assess WEx and WREx for P-IID using UHCS, the study in Bayesian happened under symmetric and asymmetric loss functions using the MCMC method. Section 4 discusses the simulation problem and its application to real data. The paper comes to a close with some concluding comments in Section 5.

2. Model Description

One of the vital lifetime models is the P-IID. It was introduced by Lomax [22], which is valuable in many fields, such as actuarial science and economics. It's been beneficial in issues of reliability and life testing (see Hassan and Al-Ghamdi [23]). Harris [24], and Atkinson et al. [25] used the P-IID to analyze income and wealth data. A random variable Z has a P-IID if its PDF is given by:

$$f(z) = \frac{\theta\xi^{\theta}}{(z+\xi)^{(\theta+1)}}, \quad \theta > 0, \xi > 0, z \geq 0. \tag{6}$$

The cumulative distribution function (CDF) of Z is specified by:

$$F(z) = 1 - \frac{\xi^{\theta}}{(z+\xi)^{\theta}}. \tag{7}$$

Several authors have published studies on P-IID in the literature. Balakrishnan and Ahsanullah [26] investigated certain recurrence relations between the moments of record values from the P-IID. Singh et al. [27] used Lindley's approximation to investigate the Bayesian estimate of P-IID under T-II HCS. Estimation of reliability for P-IID using ranked set sampling was provided by Hassan et al. [28].
The WEx of the P-IID is calculated by substituting (6) in (4) as following:

$$J^{w}(z) = \frac{-1}{2}\int_0^{\infty} z\Big[\frac{\theta\xi^{\theta}}{(\xi+z)^{\theta+1}}\Big]^2 dz$$

$$= \frac{-\theta^2\xi^{2\theta}}{2}\int_0^{\infty} z(\xi+z)^{-2\theta-2}dz.$$

Using integration by parts, then $J^{w}(z)$ is as follows

$$J^{w}(z) = \frac{-\theta^2\xi^{2\theta}}{2(2\theta+1)}\int_0^{\infty}(\xi+z)^{-2\theta-1}dz = \frac{-\theta}{4(2\theta+1)}. \tag{8}$$

Furthermore, WREx of the P-IID is calculated by substituting (6) in (5) as below:

$$J_t^{w}(z) = \frac{-(\xi+t)^{2\theta}}{2\xi^{2\theta}}\theta^2\xi^{2\theta}\int_t^{\infty} z(\xi+z)^{-2\theta-2}dz. \tag{9}$$

Using integration by parts, then $J_t^{w}(z)$ is as follows

$$J_t^{w}(z) = \frac{-\theta^2(\xi+t)^{2\theta}}{2}\Big[\frac{z(\xi+z)^{-2\theta-1}}{-2\theta-1}\Big|_t^{\infty} - \int_t^{\infty}\frac{(\xi+z)^{-2\theta-1}}{-2\theta-1}dz\Big]$$

$$= \frac{-\theta^2(\xi+t)^{2\theta}}{2}\Big[\frac{t(\xi+t)^{-2\theta-1}}{2\theta+1} + \frac{1}{2\theta+1}\frac{(\xi+t)^{-2\theta}}{2\theta}\Big].$$

After simplification, then $J_t^{w}(z)$ takes the following form:

$$J_t^{w}(z) = \frac{-\theta(2\theta t + t + \xi)}{4(2\theta+1)(\xi+t)}. \tag{10}$$

3. Estimation Methods

In this section, we study two methods of estimations, namely the ML and Bayesian to investigate WEx and WREx of the P-IID. In the Bayesian estimation method, we obtain the WEx and WREx estimators under different types of loss functions and use the MCMC method to calculate these estimators.

3.1. Maximum Likelihood Estimator

According to many life-testing experiments, censorship is crucial to lower the expense of the experiment and to shorten the amount of time spent testing. Childs et al. [29] merged type-I (T-I) and type-II (T-II) censoring and came up with a hybrid censoring system (HCS), which includes two categories (T-I HCS & T-II HCS) and these two types have been used in many types of research. The generalized T-I HCS and generalized T-II HCS were created by Chandrasekar et al. [30] by combining these two types. The UHCS, developed by Balakrishnan et al. [31], is a combination of generalized T-I and T-II hybrid censoring schemes.
Fix $r_1, r_2 \in \{1, 2, \ldots, n\}$ and the time points $T_1, T_2 \in (0, \infty)$ in this scheme.
$T^* = min\Big(max(x_{r_2:n}, T_1), T_2\Big)$, if the r_1^{th} failure happened prior to time T_1. $T^* = min(x_{r_2:n}, T_2)$ if the r_1^{th} failure happens between T_1 and T_2, and $T^* = x_{r_1:n}$, if the r_1^{th} failure happens after T_2. Consequently, we now have six cases covered by the UHCS. We can confirm that the test will be completed in time T_2 with at least r_1 failures using this scheme technique, if not, exactly r_1 failures.
We obtain the ML estimator of the P-IID parameters under UHCS. Let $Z_{1:n}, Z_{2:n}, \ldots, Z_{n:n}$ be n identical ordered failure times have P-IID, with fixed integer $r_1, r_2 \in 1, 2, ..., n$ where $r_1 < r_2 < n$ and time points $T_1, T_2 \in (0, \infty)$ where $T_1 < T_2$. Then, the likelihood function of θ and ξ, under UHCS for 6 situations is represented by

$$L(\underline{z}|\theta,\xi) = \begin{cases} \frac{n!}{(n-D_1)!}\prod_{i=1}^{D_1} f(z_{i:n})[1-F(T_1)]^{n-D_1}, & \text{for case 1} \\ \frac{n!}{(n-r_2)!}\prod_{i=1}^{r_2} f(z_{i:n})[1-F(r_2)]^{n-r_2} & \text{for case 2 and 4} \\ \frac{n!}{(n-D_2)!}\prod_{i=1}^{D_2} f(z_{i:n})[1-F(T_2)]^{n-D_2} & \text{for case 3 and 5} \\ \frac{n!}{(n-r_1)!}\prod_{i=1}^{r_1} f(z_{i:n})[1-F(r_1)]^{n-r_1} & \text{for case 6} \end{cases}, \tag{11}$$

where D_1 and D_2 are number of failures related to T_1 and T_2, respectively and $D_1 < D_2$. The likelihood function (11) can be written as:

$$L(\underline{z}|\theta,\xi) = \frac{n!}{(n-m)!}[\prod_{i=1}^{m} f(z_{i:n})][1-F(C)]^{n-m}, \tag{12}$$

where C denotes the experiment's end time and m denotes the number of observations made until the experiment's end time C and is given by:

$$(m, C) = \begin{cases} (D_1, T_1) & \text{at case 1} \\ (r_2, z_{r_2:n}) & \text{at case 2 and 4} \\ (D_2, T_2) & \text{at case 3 and 5} \\ (r_1, z_{r_1:n}) & \text{at case 6} \end{cases}, \tag{13}$$

From (6) and (7) by substituting in (12) we get likelihood function of P-IID under UHCS,

$$L(\underline{z}|\theta,\xi) = \frac{n!}{(n-m)!}[\theta^m \xi^{m\theta}\prod_{i=1}^{m}\frac{1}{(\xi+z_i)^{\theta+1}}][\frac{\xi^\theta}{(C+\xi)^\theta}]^{n-m}. \tag{14}$$

Take the logarithm of both sides, denoted by l,

$$\begin{aligned} l \propto\ & m\ln\theta + m\theta\ln\xi - (\theta+1)\sum_{i=1}^{m}\ln(\xi+z_i) \\ & + \theta(n-m)[\ln\xi - \ln(C+\xi)]. \end{aligned} \tag{15}$$

Take derivatives of (15) with respect to θ and ξ, we can get

$$\frac{\partial l}{\partial\theta} = \frac{m}{\theta} + m\ln\xi - \sum_{i=1}^{m}\ln(\xi+z_i) + (n-m)[\ln\xi - \ln(C+\xi)], \tag{16}$$

and

$$\frac{\partial l}{\partial\xi} = \frac{m\theta}{\xi} - (\theta+1)\sum_{i=1}^{m}\frac{1}{(\xi+z_i)} + \theta(n-m)[\frac{1}{\xi} - \frac{1}{(C+\xi)}]. \tag{17}$$

Set (16) & (17) to zero and solve them to get the ML estimator of θ and ξ. Then equation (16) is written as:

$$\hat{\theta} = A(\hat{\xi}), \tag{18}$$

where

$$A(\hat{\xi}) = \frac{-m}{m\ln\hat{\xi} - \sum_{i=1}^{m}\ln(\hat{\xi}+z_i) + (n-m)[\ln\hat{\xi} - \ln(C+\hat{\xi})]}.$$

By substituting from (18) into (17) after putting them equal zero to get

$$\frac{mA(\hat{\xi})}{\hat{\xi}} - (A(\hat{\xi})+1)\sum_{i=1}^{m}\frac{1}{(\xi+z_i)} + (n-m)A(\hat{\xi})[\frac{1}{\hat{\xi}} - \frac{1}{(C+\hat{\xi})}] = 0. \tag{19}$$

We may acquire the ML estimator of θ using an iterative process by computing the ML estimator from (19) and then substituting it into (18). As a result of the invariance property of ML estimation, the estimator of $J^w(z)$ and $J_t^w(z)$, becomes

$$\hat{J}^w(z) = \frac{-\hat{\theta}}{4(2\hat{\theta}+1)}, \tag{20}$$

and

$$\hat{J}_t^w(z) = \frac{-\hat{\theta}(2\hat{\theta}t + t + \hat{\xi})}{4(2\hat{\theta}+1)(\hat{\xi}+t)}. \tag{21}$$

3.2. Bayesian Estimator

Here, we consider both θ and ξ unknown, indicating that there is no natural conjugate bivariate prior distribution. As a result, we assume that the independent priors for θ and ξ are gamma (a,b) and gamma (c,d), respectively, with $\frac{a}{b}$ and $\frac{c}{d}$ as means. The θ and ξ priors are written as follows:

$$\pi_1(\theta) \propto \left(\theta^{a-1}e^{-b\theta}\right), \qquad \theta > 0,$$
$$\pi_2(\xi) \propto \left(\xi^{c-1}e^{-\xi d}\right), \qquad \xi > 0,$$

where a, b, c, and d are positive hyperparameters that carry prior knowledge. As a result, the joint prior distribution is as follows:

$$\pi(\theta,\xi) \propto \theta^{a-1}\xi^{c-1}e^{-(b\theta+d\xi)}. \tag{22}$$

The posterior distribution is given by

$$= E_1\theta^{m+a-1}\xi^{m\theta+c-1}e^{-\xi d}e^{-\theta[b+\sum_{i=1}^{m}\ln(\xi+z_i)-(n-m)\ln(\frac{\xi}{C+\xi})]}e^{-\sum_{i=1}^{m}\ln(\xi+z_i)}, \tag{23}$$

where $E_1^{-1} = \int_0^\infty \int_0^\infty L(\underline{z}|\theta,\xi)\pi(\theta,\xi)d\theta d\xi$, is the normalizing constant.

Then, the marginal posterior distributions of θ and ξ, are given below

$$\pi_1^*(\theta|\underline{z}) \propto \theta^{m+a-1}e^{-\theta b}\int_0^\infty \xi^{(m\theta+c-1)}e^{-\xi d-\theta[\sum_{i=1}^m \ln(\xi+z_i)-(n-m)\ln(\frac{\xi}{C+\xi})]}e^{-\sum_{i=1}^m \ln(\xi+z_i)}d\xi, \tag{24}$$

and

$$\pi_2^*(\xi|\underline{z}) \propto e^{-\xi d}\xi^{c-1}e^{-\sum_{i=1}^m \ln(\xi+z_i)}\int_0^\infty \theta^{m+a-1}\xi^{m\theta}e^{-\theta[b+\sum_{i=1}^m \ln(\xi+z_i)-(n-m)\ln(\frac{\xi}{C+\xi})]}d\theta. \tag{25}$$

As seen from (23) and (25) that $\pi_1^*(\theta|\xi)$ is calculated as following

$$\pi_1^*(\theta|\xi) \propto \theta^{m+a-1}e^{-\theta[b-m\ln\xi+\sum_{i=1}^m \ln(\xi+z_i)-(n-m)\ln(\frac{\xi}{C+\xi})]}. \tag{26}$$

As a result, the gamma distribution with shape parameter $(m+a-1)$ and scale parameter $[b-m\ln\xi+\sum_{i=1}^m \ln(\xi+z_i)-(n-m)\ln(\frac{\xi}{C+\xi})]$ is the posterior density function of θ given ξ. As a result, any gamma-producing technique can be used to generate θ samples with ease.
From (23) and (24) we can calculate $\pi_2^*(\xi|\theta)$, as following

$$\pi_2^*(\xi|\theta) \propto \xi^{m\theta+c-1}e^{-\xi d}e^{-\theta[\sum_{i=1}^m \ln(\xi+z_i)-(n-m)\ln(\frac{\xi}{C+\xi})]}e^{-\sum_{i=1}^m \ln(\xi+z_i)}. \tag{27}$$

Appropriate sampling techniques cannot be able to sample directly because this equation can never be solved to very well distributions. To generate an estimator for the following loss functions, use the MCMC approach.

3.2.1. Loss Functions

We will look at Bayesian estimators for both symmetric and asymmetric loss functions. The squared error loss (SEL) function is one of the most extensively utilized symmetric loss functions. The following is the SEL function:

$$L_1(\phi,\delta) = (\delta-\phi)^2,$$

where δ is an estimator of ϕ. The Bayesian estimator, based on the SEL function, is calculated as follows:

$$\hat{\phi}_{SEL} = E(\phi|data). \tag{28}$$

In terms of asymmetric loss functions, we chose the linear-exponential (LINEX) and the general entropy (GE) loss functions, which are the two most often used asymmetric loss functions. The following is a definition of the LINEX loss function:

$$L_2(\phi,\delta) = e^{-h(\delta-\phi)} - h(\delta-\phi) - 1,$$

where h is the sign presenting the asymmetry (see Parsian and Kirmani, [33]). Under the LINEX loss function, the Bayesian estimator is provided by

$$\hat{\phi}_{LINEX} = \frac{-1}{h}\ln[E(e^{-h\phi}|data)]. \tag{29}$$

The following is the definition of the GE loss function:

$$L_2(\phi,\delta) = (\frac{\delta}{\phi})^h - h(\frac{\delta}{\phi}) - 1.$$

In this case, the Bayesian estimator is:

$$\hat{\phi}_{GE} = [E(\phi^{-h}|data)]^{-\frac{1}{h}}. \tag{30}$$

Under the SEL, LINEX, and GE loss functions, the Bayesian estimators of WEx and WREx are as follows:

$$\hat{g}(\theta,\xi)_{SEL} = E_1 \int_0^\infty \int_0^\infty g(\theta,\xi)\theta^{m+a-1}\xi^{m\theta+c-1}e^{-\xi d-\sum_{i=1}^m \ln(\xi+z_i)}e^{-\theta Y_i(\theta,\xi,z_i)}d\theta d\xi, \quad (31)$$

$$\hat{g}(\theta,\xi)_{LINEX} = \frac{-1}{h}\ln[E_1 \int_0^\infty \int_0^\infty e^{-hg(\theta,\xi)}\theta^{m+a-1}\xi^{m\theta+c-1}e^{-\xi d-\sum_{i=1}^m \ln(\xi+z_i)}e^{-\theta Y_i(\theta,\xi,z_i)}d\theta d\xi], \quad (32)$$

$$\hat{g}(\theta,\xi)_{GE} = \big[E_1 \int_0^\infty \int_0^\infty (g(\theta,\xi))^{-h}\theta^{m+a-1}\xi^{m\theta+c-1}e^{-\xi d-\sum_{i=1}^m \ln(\xi+z_i)}e^{-\theta Y_i(\theta,\xi,z_i)}d\theta d\xi\big]^{\frac{-1}{h}}, \quad (33)$$

where $Y_i(\theta,\xi,z_i) = [b+\sum_{i=1}^m \ln(\xi+z_i) - (n-m)\ln(\frac{\xi}{C+\xi})]$, E_1 is the normalizing constant and to calculate the WEx put $g(\theta,\xi) = J^w(z)$ and to find the WREx put $g(\theta,\xi) = J_t^w(z)$.
It should be observed that all Bayesian entropy estimators are expressed as a ratio of two integrals, which cannot be simplified or directly computed. To compute the estimators, we use the MCMC approach.

3.2.2. MCMC Method

Consider using the MCMC approach to produce samples from posterior distributions and then using the (SEL, LINEx, GE) loss functions to compute Bayesian estimates (BEs) of WEx and WREx. There are many different MCMC schemes to select from, and it can be difficult to decide which one to use. Gibbs samplers as well as Metropolis-within-Gibbs samplers are key subclasses of MCMC algorithms. The advantage of the MCMC technique over the ML method is that by creating probability intervals based on the empirical posterior distribution, we can always obtain an acceptable interval estimate of the parameters. ML estimation frequently lacks this feature.

Algorithm

1) $\xi_0 = \hat{\xi}$.
2) $\theta^{(l)}$ is obtained from Gamma $\pi_1^*(\theta|\xi)$ as shown in (26).
3) To generate $\xi^{(l)}$ from $\pi_2^*(\xi|\theta)$ using Metropolis-Hastings (M-H) algorithm, see Metropolis et al. [32].
4) Calculate $\theta^{(l)}$ and $\xi^{(l)}$.
5) Repeat steps 2-4 N times.
6) Calculate the WEx and WREx via the loss functions using the Bayesian estimators of θ and ξ.

4. Simulation Investigation and Results

In this section, we study the performance of all previously proposed estimators for WEx and WREx, so we can use a simulation study to estimate WEx and WREx and use an Illustrative example.

4.1. Simulation Study

In this subsection, we look into the efficiency of the ML estimates (MLEs) and BEs of WEx and WREx, for the P-IID in terms of mean squared errors (MSEs) under different loss functions by using the Monte Carlo simulation.
• The UHCS from the P-IID are generated for sample sizes $n = 200$ and 100, using these samples, the MSEs of MLEs and BEs are computed.
• The BEs using the suggested loss functions when $(h = -1, 1)$ are calculated.
• The simulation runs $N = 1,000$ times and for the WREx, take $t = 0.5$.
• The prior parameters used in Bayesian inference are selected $(a, b, c, d) = (3.5, 5.5, 4, 2)$.
The MLEs and BEs of WEx and WREx were studied under the following cases where in (Table 1) the results of WREx and in (Table 2) the results of WEx:

1- Values of n, r_2, T_2 are taken as $(n = 100, r_2 = 50, T_2 = 1.2)$ and $r_1 = (80, 70, 60)$ at different values of T_1, where $T_1 = (0.2, 0.9)$.
2- Values of n, r_2, T_2 are taken as $(n = 200, r_2 = 140, T_2 = 1.2)$ and $r_1 = (190, 170, 150)$ at different values of T_1, where $T_1 = (0.2, 0.9)$.
3- Values of n, r_2, T_1 are taken as $(n = 100, r_2 = 50, T_1 = 0.9)$ and $r_1 = (80, 70, 60)$ at different values of T_2, where $T_2 = (1.2, 3)$.
4- Values of n, r_2, T_1 are taken as $(n = 200, r_2 = 140, T_1 = 0.9)$ and $r_1 = (190, 170, 150)$ at different values of T_2, where $T_2 = (1.2, 3)$.
5- Values of n, r_1, T_1 are taken as $(n = 100, r_1 = 80, T_1 = 1.2)$ and $r_2 = (70, 50, 30)$ at $T_2 = 5$.
6- Values of n, r_1, T_1 are taken as $(n = 200, r_1 = 170, T_1 = 1.2)$ and $r_2 = (160, 150, 120)$ at $T_2 = 5$.
The MSE for N samples is calculated using

$$\mathrm{MSE}(\hat{\phi}) = \sum_{i=1}^{N} \frac{(\hat{\phi}_i - \phi)^2}{N}, \tag{34}$$

where $\hat{\phi} = \hat{J}^w(z)$ and $\hat{J}_t^w(z)$. All results are given in Tables 1 and 2. Using tables, we may conclude the following:

1-The results in Tables 1 and 2 reveal that for all proposed estimates, the MSEs generally decrease with increasing value of n.
2- The MSE of weighted measures in ML and Bayesian estimates decreases as r_1 increases while the sample size n is fixed (see Figures 1 and 2).
3- When both the sample size n and the number of failures r_1 are fixed, the MSE of weighted measures in ML and Bayesian estimates decrease as the specified observation number r_2 increases (Figures 3 and 4).
4-The MSE of weighted measures in ML and Bayesian estimates generally decreases when the predetermined time T_1 and the extended time T_2 increase (see Figures 5 and 6).
5-The BEs of WEx and WREx viz LINEX loss function at $h = -1$ have a lot of information and the BEs using GE loss function at $h = 1$ have a lot of information since they have a small value of MSE.
6- The BEs of the $J^w(z)$ and $J_t^w(z)$ under the GE loss function are superior to the BEs under the other loss functions in most of the cases (Figures 1 and 2).
7-The BE of WEx and WREx is the best value under the GE loss function in most cases.
8-The amount of data obtained under the GE loss function is more in BEs than obtained under other loss functions.
9-By increasing the number of failures r_1 or r_2, the BEs of WEx and WREX are raised.

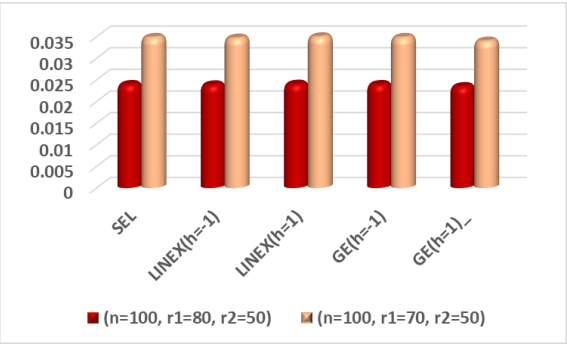


Figure 1: *The MSE of WREx estimate for various values of r_1.*

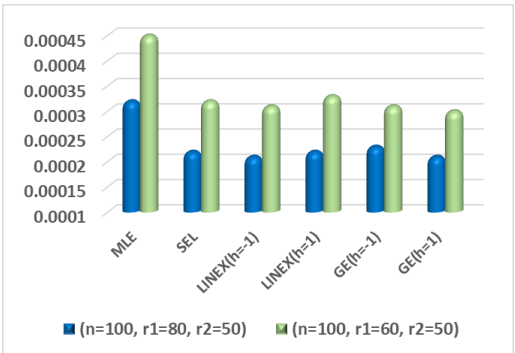


Figure 2: *MSE of WEx estimate for various values of r_1.*

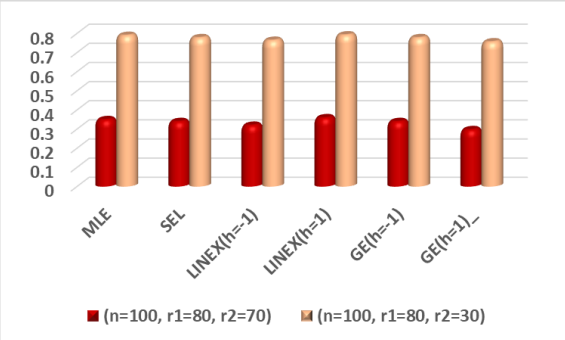


Figure 3: *The MSE of WREx estimate for different values of* r_2.

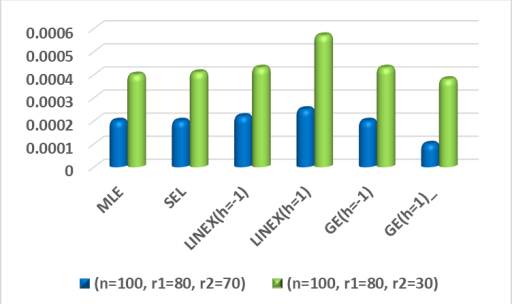


Figure 4: *The MSE of WEx estimate for various values of* r_2.

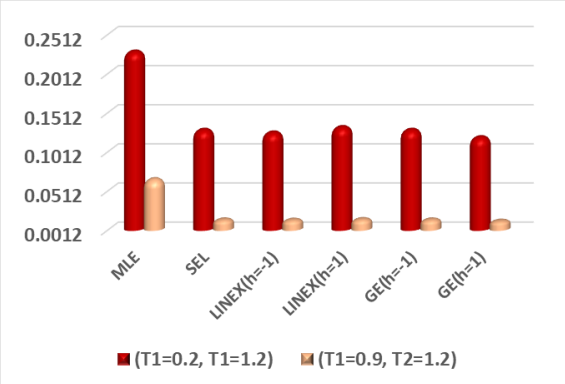


Figure 5: *The MSE of WREx estimate for various values of* T_1 *when* $n = 200, r_1 = 190, r_2 = 140$.

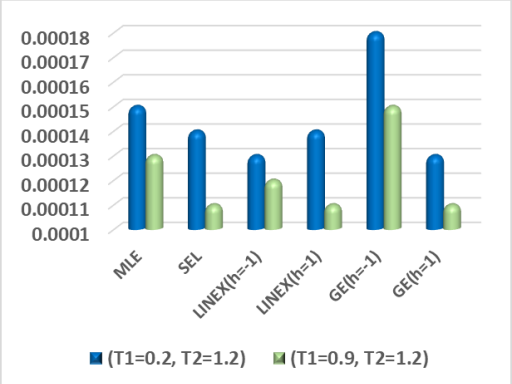


Figure 6: *The MSE of WEx estimate for various values of* T_1 *when* $n = 200, r_1 = 190, r_2 = 140$.

Table 1: *MLEs and BEs of WREx and associated MSE.*

n	r_1	r_2	MLE	SEL	LINEX		GE	
					$h=(-1)$	$h=(1)$	$h=(-1)$	$h=(1)$
					$(T_1=0.2$	$T_2=1.2)$		
100	80	50	-0.16189	-0.32887	-0.32861	-0.32914	-0.32887	-0.32731
			0.00015	0.02395	0.02387	0.02403	0.02395	0.02347
	70		-0.2097	-0.36075	-0.36034	-0.36117	-0.36075	-0.3585
			0.00127	0.03484	0.03468	0.03499	0.03484	0.034
	60		-0.16821	-0.37014	-0.36963	-0.37066	-0.37014	-0.36745
			0.00003	0.03843	0.038237	0.03863	0.03843	0.0373
200	190	140	-0.18586	-0.37053	-0.37032	-0.37073	-0.37053	-0.36944
			0.00014	0.03858	0.0385	0.03866	0.03858	0.03816
	170		-0.16035	-0.38768	-0.38744	-0.38792	-0.38768	-0.38647
			0.00019	0.04561	0.04551	0.04571	0.04561	0.0451
	150		-0.19091	-0.4324	-0.4321	-0.4327	-0.4324	-0.43103
			0.00028	0.16672	0.16656	0.16687	0.16672	0.16601
					$(T_1=0.9$	$T_2=1.2)$		
100	80	50	-0.16824	-0.30625	-0.30601	-0.30649	-0.30625	-0.30472
			0.00003	0.01746	0.0174	0.01753	0.01746	0.01706
	70		-0.17107	-0.31911	-0.3188	-0.31942	-0.31911	-0.3172
			0.00001	0.02103	0.02093	0.02112	0.02103	0.02047
	60		-0.30542	-0.35315	-0.35282	-0.3534	-0.35315	-0.351
			0.01724	0.03206	0.03194	0.03218	0.03206	0.0314
200	190	140	-0.7929	-0.73629	-0.73048	-0.74198	-0.73629	-0.71956
			0.0665	0.01465	0.01424	0.01507	0.01465	0.01294
	170		-0.64563	-0.5423	-0.5429	-0.54735	-0.54563	-0.53836
			0.0486	0.03817	0.03388	0.04233	0.03817	0.02721
	150		-0.58563	-0.48563	-0.4839	-0.48735	-0.48563	-0.47836
			0.1487	0.13817	0.13388	0.14243	0.13817	0.12601
					$(T_1=0.9$	$T_2=3)$		
100	80	50	-1.11252	-1.01252	-1.00171	-1.02322	-1.01252	-0.99011
			0.51982	0.41982	0.40594	0.4338	0.41982	0.39128
	70		-1.1324	-1.03264	-1.02677	-1.06183	-1.03264	-1.0296
			0.5188	0.4348	0.455	0.4903	0.4348	0.4112
	60		-1.1564	-1.06564	-1.05777	-1.07383	-1.06564	-1.05096
			0.5148	0.49148	0.4805	0.50303	0.49148	0.47112
200	190	140	-0.5662	-0.53222	-0.53067	-0.53375	-0.53222	-0.52614
			0.0981	0.0281	0.02759	0.02862	0.0281	0.0261
	170		-0.6997	-0.67367	-0.66779	-0.67942	-0.67367	-0.65475
			0.09774	0.09554	0.09193	0.09912	0.09554	0.0842
	150		-0.7897	-0.7767	-0.72779	-0.7752	-0.75367	-0.653
			0.11674	0.11354	0.11093	0.11812	0.11354	0.10842

Continue Table 1

n	r_1	r_2	MLE		($T_1 = 1.2$ LINEX	$T_2 = 5$)	GE	
				SEL	$h = (-1)$	$h = (1)$	$h = (-1)$	$h = (1)$
100	80	70	-0.999	-0.94699	-0.93012	-0.9638	-0.94699	-0.90949
			0.3492	0.3392	0.31983	0.35908	0.3392	0.29693
		50	-1.12431	-1.3531	-1.34752	-1.3638	-1.35631	-1.34385
			0.5694	0.5544	0.5436	0.5774	0.5744	0.5352
		30	-1.34631	-1.24631	-1.23852	-1.25438	-1.24631	-1.23385
			0.7894	0.77744	0.76376	0.79174	0.77744	0.75562
200	170	160	-0.80514	-0.70514	-0.70141	-0.70881	-0.70514	-0.694
			0.1348	0.11598	0.11345	0.11849	0.11598	0.10852
		150	-1.6424	-1.4524	-1.4499	-1.4635	-1.4524	-1.4381
			1.222	1.1322	1.1537	1.1752	1.13332	1.1145
		120	-1.75824	-1.67524	-1.66099	-1.68935	-1.67524	-1.65781
			1.882	1.71782	1.68067	1.75502	1.71782	1.67245

Table 2: *MLEs and BEs of WEx and associated MSE.*

n	r_1	r_2	MLE		($T_1 = 1.2$ LINEX	$T_2 = 5$)	GE	
				SEL	$h = (-1)$	$h = (1)$	$h = (-1)$	$h = (1)$
100	80	70	-0.1182	-0.1182	-0.11823	-0.11826	-0.11823	-0.11819
			0.0002	0.0002	0.00022	0.00025	0.0002	0.0001
		50	-0.1218	-0.1215	-0.1219	-0.1221	-0.1245	-0.1215
			0.0003	0.00021	0.00023	0.00037	0.00033	0.00028
		30	-0.1248	-0.12485	-0.12495	-0.12505	-0.12495	-0.12475
			0.0004	0.00041	0.00043	0.00057	0.00043	0.00038
200	170	160	-0.1152	-0.1152	-0.11521	-0.11527	-0.11521	-0.11519
			0.00012	0.00012	0.00012	0.00019	0.00012	0.00011
		150	-0.1242	-0.12422	-0.12422	-0.12422	-0.12422	-0.12422
			0.0004	0.0008	0.0005	0.0009	0.0008	0.0004
		120	-0.1352	-0.13522	-0.13522	-0.1422	-0.1432	-0.12522
			0.0014	0.0018	0.0014	0.0016	0.0014	0.0010

Continue Table 2

n	r_1	r_2	MLE					
					($T_1 = 0.2$ $T_2 = 1.2$)			
						LINEX	GE	
					$h = (-1)$	$h = (1)$	$h = (-1)$	$h = (1)$
100	80	50	-0.1276	-0.12482	-0.12482	-0.12492	-0.12482	-0.12470
			0.00032	0.00023	0.00023	0.00027	0.00024	0.00021
	70		-0.1015	-0.12494	-0.12494	-0.12497	-0.12494	-0.12491
			0.00006	0.00097	0.00095	0.00099	0.00091	0.0009
	60		-0.08959	-0.12476	-0.12476	-0.12478	-0.12476	-0.12465
			0.00002	0.00096	0.00096	0.00116	0.00096	0.00082
200	190	140	-0.09573	-0.1238	-0.1248	-0.1252	-0.1248	-0.1242
			0.00093	0.0009	0.00096	0.00098	0.00096	0.00091
	170		-0.09744	-0.12471	-0.12477	-0.12482	-0.12477	-0.1241
			0.00015	0.00086	0.00096	0.00106	0.00096	0.00043
	150		-0.12133	-0.12122	-0.12138	-0.12144	-0.12134	-0.12132
			0.00025	0.00025	0.00039	0.00029	0.00029	0.00027
					($T_1 = 0.9$ $T_2 = 1.2$)			
100	80	50	-0.09412	-0.12491	-0.12493	-0.12498	-0.12493	-0.12492
			0.00057	0.00094	0.00097	0.00099	0.00097	0.00095
	70		-0.136	-0.1249	-0.12495	-0.12498	-0.12495	-0.12491
			0.00037	0.00097	0.00097	0.00099	0.00097	0.00096
	60		-0.11962	-0.12498	-0.12498	-0.12498	-0.12498	-0.1249
			0.00067	0.00098	0.00098	0.00098	0.00098	0.00098
200	190	140	-0.1172	-0.11622	-0.11623	-0.11628	-0.11623	-0.11619
			0.00013	0.00011	0.00012	0.00015	0.00015	0.00011
	170		-0.119	-0.1097	-0.10972	-0.1097	-0.10982	-0.10968
			0.00003	0.00003	0.00003	0.00004	0.00003	0.00002
	150		-0.121	-0.1117	-0.1142	-0.1137	-0.11182	-0.11168
			0.00013	0.00022	0.00019	0.00027	0.00015	0.00011
					($T_1 = 0.9$ $T_2 = 3$)			
100	80	50	-0.13540	-0.12340	-0.12347	-0.12357	-0.12347	-0.12342
			0.00041	0.0003	0.00037	0.00047	0.00037	0.00033
	70		-0.1341	-0.12251	-0.12233	-0.12288	-0.12233	-0.12222
			0.00042	0.0003	0.00035	0.00037	0.00035	0.00032
	60		-0.1321	-0.12211	-0.12213	-0.12218	-0.12213	-0.12212
			0.00043	0.00031	0.00032	0.00035	0.00032	0.00031
200	190	140	-0.1107	-0.11079	-0.1108	-0.11087	-0.11081	-0.11078
			0.00004	0.00004	0.00004	0.00004	0.00004	0.00004
	170		-0.1142	-0.11421	-0.1145	-0.1146	-0.1146	-0.1142
			0.00008	0.00009	0.00008	0.00005	0.00009	0.00002
	150		-0.1152	-0.11521	-0.11525	-0.11526	-0.11526	-0.1152
			0.0001	0.00011	0.00012	0.00015	0.00012	0.00010

4.2. Data Analysis

Consider the data from Lee and Wang [34], which represent the remission times (in months) of a random sample of 128 bladder cancer patients. For these real data, the Kolmogorov-Smirnov (K-S) test is used, and the p-value indicates that the P-IID best matches the data where (p-value=0.336703 and K-S distance = 0.0820311). Figure 7 illustrates the estimated PDF and CDF of P-IID.

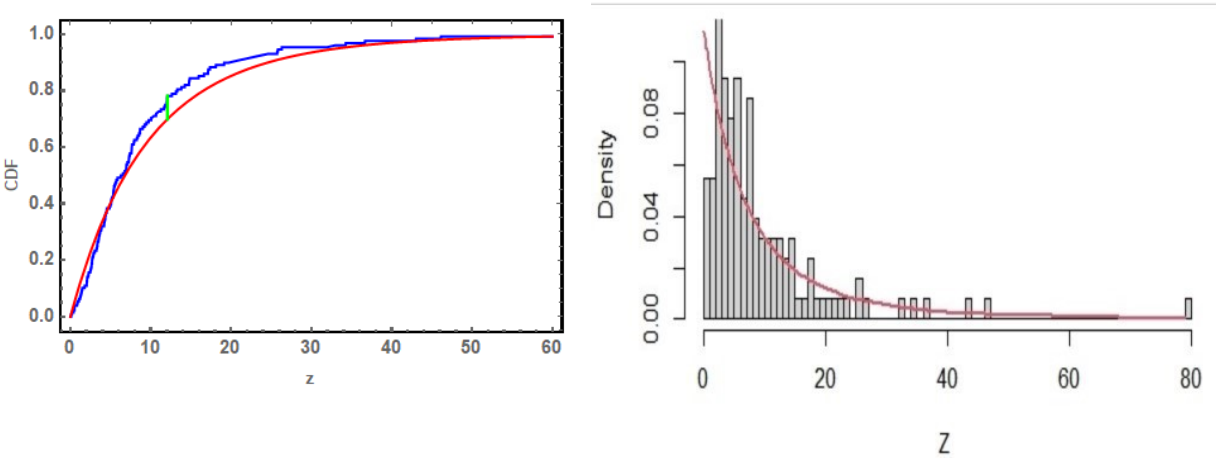


Figure 7: *Estimated PDF and CDF of P-IID*

In this application, we assume that the distribution of this data is the P-IID based on the UHCS. We take
Case 1: $(r_1 = 70, r_2 = 80, T_1 = 10, T_2 = 15)$,
Case 2: $(r_1 = 70, r_2 = 95, T_1 = 10, T_2 = 15)$,
Case 3: $(r_1 = 70, r_2 = 115, T_1 = 10, T_2 = 15)$,
Case 4: $(r_1 = 93, r_2 = 100, T_1 = 10, T_2 = 15)$,
Case 5: $(r_1 = 93, r_2 = 115, T_1 = 10, T_2 = 15)$,
Case 6: $(r_1 = 120, r_2 = 125, T_1 = 10, T_2 = 15)$.
We use a non-informative prior to calculating the BEs under (SEL, LINEX, GE) loss functions with $h = (-1, 1)$, because we don't know anything about the priors, therefore we choose $a = 0, b = 0, c = 0$, and $d = 0$. The results for the real data are listed in Tables 3 and 4.

Table 3: *Estimation of WEx when ($T_1 = 10$, $T_2 = 15$).*

Case	n	r_2	r_1	MLE	BSL	LINEX		GE	
						$h = (-1)$	$h = (1)$	$h = (-1)$	$h = (1)$
1		80	70	-0.125	-0.12183	-0.12184	-0.12183	-0.12175	-0.12183
2		95	70	-0.1252	-0.12183	-0.12184	-0.12183	-0.12175	-0.12183
3	128	115	70	-0.12	-0.12179	-0.1218	-0.12179	-0.12171	-0.12179
4		100	93	-0.125	-0.12179	-0.1218	-0.12179	-0.12171	-0.12179
5		115	93	-0.125	-0.12182	-0.12183	-0.12182	-0.12174	-0.12182
6		125	120	-0.1190	–0.11886	-0.11886	-0.11886	-0.11886	-0.11886

Table 4: *Estimation of WREx when* ($T_1 = 10$, $T_2 = 15$).

Case	n	r_2	r_1	MLE	BSL	LINEX		GE	
						$h=(-1)$	$h=(1)$	$h=(-1)$	$h=(1)$
1		80	70	-1.5305	-0.82458	-1.05544	-0.59372	-0.22025	-0.82458
2		95	70	-1.46202	-0.79034	-1.00075	-0.57993	-0.21951	-0.79034
3	128	115	70	-1.40151	-0.76005	-0.95306	-0.56705	-0.21867	-0.76005
4		100	100	-1.42467	-0.77163	-0.97124	-0.57203	-0.21895	-0.77163
5		115	115	-1.41513	-0.76689	-0.96375	-0.57003	-0.21893	-0.76689
6		125	120	-0.8634	-0.84222	-0.84318	-0.84123	-0.83977	-0.84222

The trace plot and histogram of the first 1000 MCMC results for the posterior distribution of WEx and WREx for case 1 are shown in Figures 8 and 9.

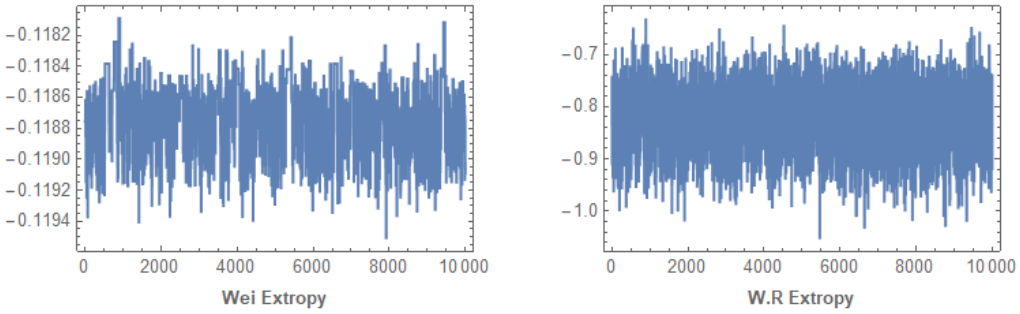


Figure 8: *The posterior sample trace plot for case* 1.

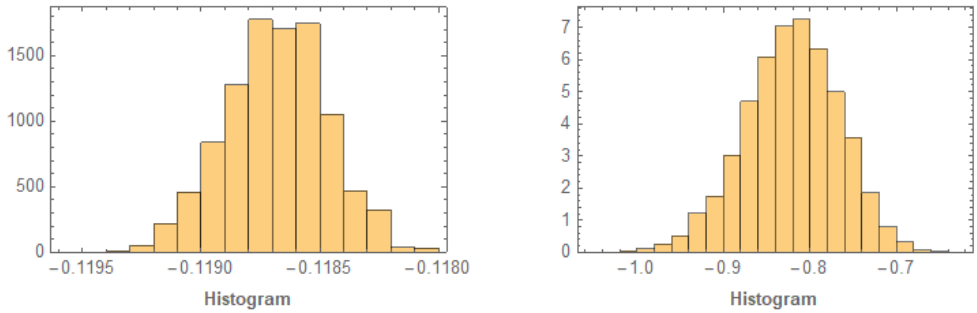


Figure 9: *The posterior sample histogram for case* 1.

We note from the study of this application that:
• The BE of WEx and WREx is less than the value of MLE in most cases.
• The BE of WEx and WREx via LINEX loss function at $h = -1$ have a small value. Furthermore, the BEs via the GE loss function at $h = -1$ have a large value. Finally, we reach the conclusion that the simulated research is supported by real data.

5. Conclusion

In this study, using UHCS, we investigated Bayesian and non-Bayesian estimators of WEx and WREx for the P-IID. For the weighted extropies measures under study, we found ML and Bayesian estimators for both symmetric and asymmetric loss functions. The MCMC techniques were used to calculate the Bayes estimates based on the M-H algorithm. In terms of accuracy measures, the performance of weighted extropy and its residual estimates for P-IID were explored. One

application to real data was considered, as well as a simulation issue. In general, the MSE values of ML and Bayesian estimators of weighted measures decrease as the number of failures rises in most cases, according to the results of the study. When compared to different estimates, the Bayesian estimate of WEx and WREx under the general entropy loss function performs well in the majority of situations. By increasing the number of failures r_1 or r_2, the BEs of WEx and WREx are raised.

Funding: Not applicable.

Authors' contributions: All authors read and approved the final manuscript.

References

[1] Marvizadeh, S.Z.(2013). Entropy Applications in Industrial Engineering. Ph.D. Thesis. The University of Nebraska.
[2] Liew, K.F., Lam, W.S., Lam, W.H.(2022). Financial network analysis on the performance of companies using an integrated entropy-DEMATEL-TOPSIS model. *Entropy*, 24: 1056.
[3] Cho, Y., Sun, H., Lee, K.(2014). An estimation of the entropy for a Rayleigh distribution based on doubly-generalized Type-II hybrid censored samples *Entropy*, 16(7):3655-3669.
[4] Liu, S., Gui, W.(2019). Estimating the entropy for Lomax distribution based on generalized progressively hybrid censoring. *Symmetry*, 11(10):1219.
[5] Hassan, A.S., Zaky, A.N.(2019). Estimation of entropy for inverse Weibull distribution under multiple censored data. *Journal of Taibah University for Science*, 13: 331-337.
[6] Yu, J., Gui, W., Shan, Y.(2019). Statistical inference on the Shannon entropy of inverse Weibull distribution under the progressive first-failure censoring. *Entropy*, 21(12): 1209.
[7] Ahmadini, A.A.H., Hassan, A.S., Zaki, A. N. Alshqaq, S.S.(2020). Bayesian inference of dynamic cumulative residual entropy from Parto II distribution with application to Covid 19. *AIMS Mathematics*, 6(3):2196-2216
[8] Al-Babtain, A.A., Hassan, A.S., Zaky, AN., Elbatal, I., Elgarhy, M.(2021). Dynamic cumulative residual Rényi entropy for Lomax distribution: Bayesian and non-Bayesian methods. *AIMS Mathematics*, 6(3): 3889-3914.
[9] Almarashi, A.M., Algarni, A., Hassan, A.S., Zaky, A.N., Elgarhy, M.(2021). Bayesian analysis of dynamic cumulative residual entropy for Lindley distribution. *Entropy*, 23: 1256.
[10] Hassan, A.S., Zaky, A.N.(2021). Entropy Bayesian Estimation for Lomax Distribution Based on Record. *Thailand Statistician*, 19(1):96-115.
[11] Al-Omari, A.I., Hassan, A.S., Nagy, H.F., Al-Anzi, A.R.A., Alzoubi, L.(2021). Entropy Bayesian analysis for the generalized inverse exponential distribution based on URRSS. *Computers, Materials, and Continua*, 69: 3795-3811.
[12] Helmy, B.A., Hassan, A.S., El-Kholy, A.K. (2021). Analysis of uncertainty measure using unified hybrid censored data with applications. *Journal of Taibah University for Science*, 15(1):1130-1143.
[13] Hassan, A.S., Elsherpieny, E.A., Mohamed, R.E.(2022). Estimation of information measures for power-function distribution in presence of outliers and their applications. *International Journal of Information and Communication*, 21(1): 1-25.
[14] Lad, F., Sanfilippo, G., Agro, G.(2015). Extropy: Complementary dual of entropy. *Statistical Science*, 30 (1):40-58.
[15] Becerra, A., de la Rosa, J.I., González, E., Pedroza, A.D., Escalante, N.I.(2018). Training deep neural networks with a non-uniform frame-level cost function for automatic speech recognition. Multimed. *Tools Application*, 77: 27231-27267
[16] Qiu, G.(2017). The extropy of order statistics and record values. *Statistics and Probability Letters,* 120:52-60.
[17] Raqab, M.Z. Qiu, G.(2019). On extropy properties of ranked set sampling. *Statistics*, 53 (1): 210-226.

[18] Qiu, G., Wang, L., Wang, X.(2019). On extropy properties of mixed systems. *Probability in the Engineering and Informational Sciences*, 33(3): 471-486.

[19] Qiu, G., Jia, K.(2018). The residual extropy of order statistics. *Statistics and Probability Letters*, 133:15-22.

[20] Belis, M. Guiasu, S.(1968). A quantitative-qualitative measure of information in cybernatic systems. *IEEE Transactions on Information Theory*, 14(4): 593-594.

[21] Balakrishnan, N., Buono, F., Longobardi, M.(2020). On weighted extropies. *Communications in Statistics - Theory and Methods*, 1-31.

[22] Lomax, K.S. (1954). Business Failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, 49(268): 847-852.

[23] Hassan, A.S., Al-Ghamdi, A.N.(2009). Optimum step stress accelerated life testing for Lomax distribution. *Journal of Applied Sciences Research*, 5: 2153-2164.

[24] Harris, C.M.(1968). The Pareto distribution as a queue service discipline. *Operations Research*, 16(2): 307-313.

[25] Atkinson, A.B., Harrison, A.J., James, A.(1978). Distribution of Personal Wealth in Britain. Cambridge: Cambridge University Press.

[26] Balakrishnan, N., Ahsanullah, M.(1994).Relations for single and product moments of record values from Lomax distribution. *Sankhya A, Series B*, 56(2): 140-146.

[27] Singh, S.K., Singh, U., Yadav, A.S.(2017). Bayesian estimation of Lomax distribution under type-II hybrid censored data using Lindley's approximation method. *International Journal of Data Science*, 2:352-368.

[28] Hassan, A.S., Almanjahie, I. M., Al-Omari, A. I., Alzoubi, L. Nagy, H.F., (2023). Stress-Strength modeling using median-ranked set sampling: estimation, simulation, and application. *Mathematics*, 11: 318.

[29] Childs, A., Chandrasekar, B., Balakrishnan, N., Kundu, D.(2003). Exact likelihood inference based on type-I and type-II hybrid censored samples from the exponential distribution. *Annals of the Institute of Statistical Mathematics*, 55: 319-330.

[30] Chandrasekar, B., Childs, A., Balakrishnan, N.(2004). Exact likelihood inference for the exponential distribution under generalized type-I and Type-II hybrid censoring. *Naval Research Logistics*, 51(7): 994-1004

[31] Balakrishnan, N., Rasouli, A., Sanjari, F.N.(2008). Exact likelihood inference based on a unified hybrid censoring sample from the exponential distribution. *Journal of Statistical Computation and Simulation*, 78(5): 475-488.

[32] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.(1953). Equations of state calculations by fast computing machines. *Chemical Physics*, 21: 1087-1091.

[33] Parsian, A. Kirmani, S.N.U.A.(2002). Estimation under LINEX loss function, In A. Ulla, Ed., Handbook of Applied Econometrics and Statistical Inference, 165 Dekker, New York, 53-76.

[34] Lee, E. T., Wang, J.(2003). Statistical Methods for Survival Data Analysis. 476. John Wiley, Sons.

RT&A, No 3 (74)
Volume 18, September 2023

ALTERNATE QUADRA SUB - MERGING POLAR FUZZY SOFT GRAPH AND ITS APPLICATION

Anthoni Amali A[1*], J.Jesintha Rosline [2]

•

Auxilium College(Autonomous), Vellore - 632006,
Affiliated to Thiruvalluvar University, Serkadu, Tamil Nadu, India,
anthoniamaliasir@gmail.com, jesi.simple@gmail.com

Abstract

Fuzzy graph and Fuzzy soft graph are indispensable computing modules for presenting membership and non - membership values in the world of uncertain situations and incidents. In this research article, we introduce the new module of Alternate Quadra Submerging Polar Fuzzy Soft Graph with four co – ordinates with membership and non – membership values. The aim of this new fuzzy soft graph is to find the single output from different uncertain parametric sets of subjects and events, between the range [-1, 1]. The submerge level of fixed four co ordinates is a tool to find the precise and reliable membership degree values from uncertain problems and outcomes. In this artifact, we also investigate the different types of Alternate Quadra Submerging Polar Fuzzy Soft Graphs, corresponding parametric fuzzy values and submerge membership and non – membership values. We discussed Strong, Complete, Complement and μ complement properties of Alternate Quadra Submerging Polar Fuzzy Soft Graphs. We use this fuzzy soft graph in the Analysis of water related diseases to find the result of most and least affected diseases with the symptoms among the hostel students in the same locality. We find the maximum and minimum membership and non - membership value of the water related diseases in an unique way by using this Alternate Quadra Submerging Polar Fuzzy Soft Graph score function values.

Keywords: AQSP fuzzy graph , AQSP Fuzzy Soft Graph, Strong and Complete AQSP fuzzy soft graph, Complement and μ – Complement proprties of AQSP fuzzy soft graphs.

1. Introduction

The future is parametric uncertain universal set, but this uncertainty is at the very heart of human creativity. Mathematicians and Scientists have a lot of experience with ignorance, doubt, and uncertainty. In 1965 Prof.Lotfi.A.Zadeh[20], invented Fuzzy set theory with membership values to solve uncertain subjects and events. The concept of fuzzy graph was first introduced by Rosenfeld[16]. Kaufmann's[10] initial definition of a fuzzy graph was based on Zadeh's fuzzy relations. Bhattacharya[6] gave some remarks on fuzzy graphs. In 1994, Moderson[14] and Peng introduced several notations on fuzzy graphs and the concept of complement of fuzzy graphs. In 1999, Molodtsov[12] introduced the concept of soft set theory to deal with uncertainties. It has been applied in the field of Applied Mathematics, Artificial Computation intelligent, Engineering, Smoothness of functions, Medical Science and Environment. Since the research on soft fuzzy sets has been very active and received much attention from researchers in worldwide.

In this current computing era, a few research studies contributed into fuzzification of soft set theory. Feng et al, combined soft sets with rough sets and fuzzy sets, obtaining three types of hybrid models, rough, soft sets, soft, rough sets, and soft – rough fuzzy sets. In 2001, Maji et al, initiated the concept of fuzzy soft sets which is a combination of soft sets. In 2002, M.S.Sunitha [19] and Vijayakumar gave a modified definition of Complement of fuzzy graph.

In 2006, Nagooorgani[8] and Chandrasekaran defined. μ – Complement of fuzzy graph, which is different from the definition of M.S.Sunitha's Complement of fuzzy graph.In 2015 , Samanta and Mohinta[13] investigated the notions of fuzzy soft graphs, Operation of union,intersection of two fuzzy soft graphs with properties related to this fuzzy soft graph module. Akram[1],[2] and Nawaz introduced the notions of fuzzy soft graph, strong, complete fuzzy soft graph and regular fuzzy soft graph with properties are investigated.

In this paper, we introduce certain types of Alternate Quadra Sub – merging Polar Fuzzy soft graphs, μ – Complement of AQSP fuzzy soft graphs and some properties of μ – Complement of AQSP fuzzy soft graphs. And we explore some results of strong and complete AQSP fuzzy soft graphs and isolated AQSP fuzzy soft graphs with theorems, examples, and applications. Using submerging level of fixation method in four quadrant membership and non - membership values, [- 0.5, 0] $\subset$ [-1, 0], [- 0.5, 0.5] $\subset$ [-1, 1], [0, 0.5] $\subset$ [0,1] and [0.5, 0.5] $\subset$ [1,1] will provide the solution from uncertain membership values. It is the module of medical and psychological studies to interpret a particular type of Uncertainty with parametric set.

2. Preliminaries

2.1. Fuzzy Graph[16].

Let V be a nonempty finite set and $\sigma : V \longrightarrow [0,1]$. And, let $\mu : V \times V \longrightarrow [0,1]$ such that $\mu(x,y) \leq \sigma(x) \wedge \sigma(y), \forall (x,y) \in V \times V$. Then the ordered pair $G = (\sigma, \mu)$ is called a fuzzy graph over the set V, where σ and μ are fuzzy vertex and edge of fuzzy graph $G = (\sigma, \mu)$.

2.2. Fuzzy Soft Set[13].

Let X be an initial universe set and E be the set of parameters. Let $A \subset E$. A pair (F,A) is called fuzzy soft set over X, where F is a mapping given by $F : A \longrightarrow I^X$ and I^X denotes the collection of all fuzzy subsets of X.

2.3. Complete Fuzzy Graph [14]

A Complete fuzzy graph is a pair of functions $G : (\sigma, \mu)$, where σ is a fuzzy subset of X and μ is a symmetric fuzzy relation on σ. Here $\sigma : X \to [0,1]$ and $\mu : X \times X \to [0,1]$ such that $\mu(x,y) = \wedge(\sigma(x), \sigma(y)) \ \forall \ x,y \ \in \sigma^*$.

2.4. Strong Fuzzy Graph [14]

A strong fuzzy graph is a pair of functions $G : (\sigma, \mu)$ where σ is a fuzzy subset of X and μ is a symmetric fuzzy relation on σ . Here $\sigma : X \to [0,1]$ and $\mu : X \times X \to [0,1]$ such that $\mu(x,y) = \wedge(\sigma(x), \sigma(y)) \ \forall \ x,y \ \in \mu^*$

2.5. Complement of Fuzzy Graph [14]

Let $G : (\sigma, \mu)$ be a fuzzy graph. The complement of G is defined as $\bar{G} = (\sigma, \bar{\mu})$, where $\bar{\mu}(x,y) = \sigma(x) \wedge \sigma(y) - \mu(x,y) \forall x,y \in V$. When G is a fuzzy graph, $\overline{G} = (\sigma, \overline{\mu})$ is complement of fuzzy graph.

2.6. μ- Complement of Fuzzy Graph [19]

Let $G : (\sigma, \mu)$ be a fuzzy graph. The μ- complement of G is defined as $G^\mu = (\sigma, \bar{\mu^\mu})$, where $\bar{\mu^\mu}(x,y) = \sigma(x) \wedge \sigma(y) - \mu(x,y), \ if \ \mu(x,y) > 0, \ \text{and} \ \mu^\mu(x,y), \ if \mu(x,y) = 0.$

3. Method

The essential definition of AQSP fuzzy soft graph method is deliberated with an examples.

3.1. Alternate Quadra Sub - merging Polar(AQSP) Fuzzy Graph

An Alternate Quadra - Submerging Polar (AQSP) Fuzzy Graph $G = (\sigma_{AQSP}, \mu_{AQSP})$ is a fuzzy graph with crisp graph $G^* = (\sigma^*_{AQSP}, \mu^*_{AQSP})$ is given as $V = (\sigma^P(x), \sigma^N(x), \rho^P(x), \rho^N(x))$ which is the membership value of vertices along with the uncertain membership value of edges is given as, $E = V \times V = (\mu^P(x,y), \mu^N(x,y), \gamma^P(x,y), \gamma^N(x,y))$. Here the vertex set V is defined with the given condition in a unique method which is an alternate contrast submerging polarized uncertain transformation.Here $\sigma^P = V \to [0,1]$, $\sigma^N = V \to [-1,0]$, $\rho^P = d \mid 0.5, \sigma^P(x) \mid$ and $\rho^N = -d \mid -0.5, \sigma^N(x) \mid$. Here (-0.5, 0.5) is the fixation of uncertain alternate contrast polarized submerging transformation into certain consistent preferable position. And the edge set E satisfies the following sufficient conditions.

$$(i) \quad \mu^P(x,y) \leq \min(\sigma^P(x), \sigma^P(y)), \quad (ii) \ \mu^N(x,y) \geq \max(\sigma^N(x), \sigma^N(y))$$

$$(iii) \ \gamma^P(x,y) \leq \min(\rho^P(x), \rho^P(y)) \quad (iv) \ \gamma^N(x,y) \geq \max(\rho^N(x), \rho^N(y)),$$

$\forall (x,y) \in E$. By definition, $\mu^P = V \times V \to [0,1] \times [1,0]$, $\mu^N = V \times V \to [-1,0] \times [0,-1]$ and the submerging mappings, $\gamma^P = V \times V \to [0,0.5] \times [0.5,0]$,
$\gamma^N = V \times V \to [-0.5,0] \times [0,-0.5]$, which denotes the impact of the alternate quadrant polarized fuzzy mapping. The maximum of submerging presumption to be at the level of confidence $[0,0.5] \subseteq [0,1]$ and the minimum of submerging presumption level of confidence is $[-0.5,0] \subseteq [-1,0]$ extension of the graph with its membership and non - membership values portrait the unique level of submerging destination in an AQSP fuzzy graph.

Also it must satisfy the condition, $-1 \leq \sigma^P(x) + \sigma^N(x) \leq 1$ *and* $|\rho^P(x) + \rho^N(x)| \leq 1$ with constrains $0 \leq \sigma^P(x) + \sigma^N(x) + |\rho^P(x) + \rho^N(x)| \leq 2$ such that the uncertain status of submerging presumption, transform into its precise consistent level with fixation mid - value 0.5, which implies that level of confidence 0.5 in an AQSP as the valuable membership of its position which is real and valid in the fuzzification. The example of AQSP fuzzy graph is given in Fig.1.

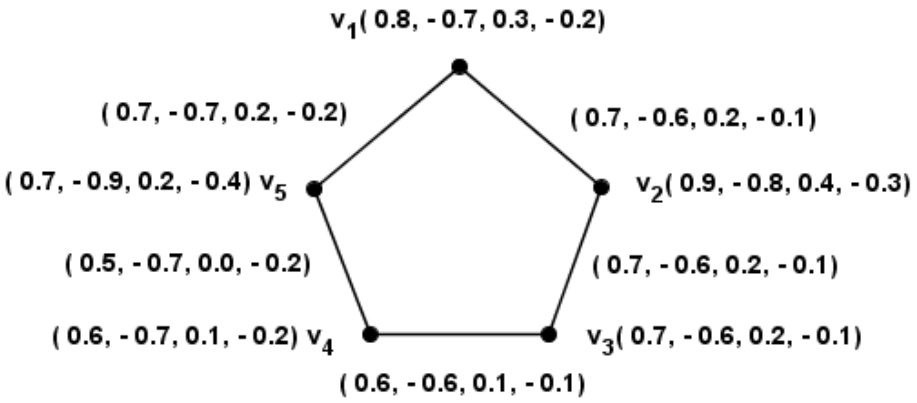


Figure 1: *AQSP Fuzzy Graph* $G = (\sigma_{AQSP}, \mu_{AQSP})$

3.2. AQSP Fuzzy Soft Graph

Let $V = ((\sigma_1^P(x), \sigma_1^N(x), \rho_1^P(x), \rho_1^N(x)), (\sigma_2^P(x), \sigma_2^N(x), \rho_2^P(x), \rho_2^N(x))...(\sigma_n^P(x), \sigma_n^N(x), \rho_n^P(x), \rho_n^N(x)))$ be a nonempty AQSP fuzzy set. E (Parameters set) and $A_{AQSP} \subset E$. Also let,

(i) $\sigma^P : A_{AQSP} \longrightarrow F_{AQSP}(V)$(Collection of all AQSP fuzzy subsets in V), $e \longmapsto \sigma_e^P$, and $\sigma_e^P : V \longrightarrow [0,1]$, $v_i \longmapsto \sigma_e^P$ then (A_{AQSP}, σ^P) : AQSP fuzzy soft vertex set.
(ii) $\sigma^N : A_{AQSP} \longrightarrow F_{AQSP}(V)$(Collection of all AQSP fuzzy subsets in V), $e \longmapsto \sigma_e^N$, and $\sigma_e^N : V \longrightarrow [-1,0]$, $v_i \longmapsto \sigma_e^N$ then (A_{AQSP}, σ^N) : AQSP fuzzy soft vertex set.

(iii) $\rho^P : A_{AQSP} \longrightarrow F_{AQSP}(V)$(Collection of all AQSP fuzzy submerge subsets in V), $e \longmapsto \rho_e^P$, and $\rho_e^P : V \longrightarrow [0,0.5]$, $v_i \longmapsto \rho_e^P$ then (A_{AQSP}, ρ^P) : AQSP fuzzy soft vertex set.
(iv) $\rho^N : A_{AQSP} \longrightarrow F_{AQSP}(V)$(Collection of all fuzzy submerge subsets in V), $e \longmapsto \rho_e^N$, and $\rho_e^N : V \longrightarrow [-0.5,0]$, $v_i \longmapsto \rho_e^N$ then (A_{AQSP}, ρ^N) : AQSP fuzzy soft vertex set.

(v) $\mu^P : A_{AQSP} \longrightarrow F_{AQSP}(V \times V)$(Collection of all AQSPfuzzy subsets in $V \times V$), $e \longmapsto \mu_e^P$, $\mu_e^P : V \times V \longrightarrow [0,1]$, $(v_i, v_j) \longmapsto \mu_e^P(v_i, v_j)$ then (A_{AQSP}, μ^P) :
AQSP fuzzy soft membership edge set.

(vi) $\mu^N : A_{AQSP} \longrightarrow F_{AQSP}(V \times V)$(Collection of all AQSPfuzzy subsets in $V \times V$), $e \longmapsto \mu_e^N$, and $\mu_e^N : V \times V \longrightarrow [-1,0]$, $(v_i, v_j) \longmapsto \mu_e^N(v_i, v_j)$ then (A_{AQSP}, μ^N) : AQSP fuzzy soft non - membership edge set.

(vii) $\gamma^P : A_{AQSP} \longrightarrow F_{AQSP}(V \times V)$(Collection of all AQSPfuzzy subsets in $V \times V$), $e \longmapsto \gamma_e^P$, and $\gamma_e^P : V \times V \longrightarrow [0,0.5]$, $(v_i, v_j) \longmapsto \gamma_e^P(v_i, v_j)$ then (A_{AQSP}, γ^P) : AQSP fuzzy soft submerge membership edge set.
(viii) $\gamma^N : A_{AQSP} \longrightarrow F_{AQSP}(V \times V)$(Collection of all AQSPfuzzy subsets in $V \times V$), $e \longmapsto \gamma_e^N$, and $\gamma_e^N : V \times V \longrightarrow [-0.5,0]$, $(v_i, v_j) \longmapsto \gamma_e^N(v_i, v_j)$ then (A_{AQSP}, γ^N) : AQSP fuzzy soft submerge membership edge set. Then the AQSP fuzzy soft graph is, $((A_{AQSP}),\ (\sigma^P, \sigma^N, \rho^P, \rho^N)),\ ((A_{AQSP}), (\mu^P, \mu^N, \gamma^P, \gamma^N))$ if the conditions are satisfied

(a) $\mu_e^P(x,y) \leq \sigma_e^P(x) \wedge \sigma_e^P(y)$, (b) $\mu_e^N(x,y) \geq \sigma_e^N(x) \vee \sigma_e^N(y)$,

(c) $\gamma_e^P(x,y) \leq \rho_e^P(x) \wedge \rho_e^P(y)$, (d) $\gamma_e^N(x,y) \geq \rho_e^N(x) \vee \rho_e^N(y)$, for all $e \in A_{AQSP}$ and for all values of $x, y = 1, 2, 3, ..., n$ and this AQSP fuzzy soft graph is denoted as $G_{AQSP}(A, V)$.

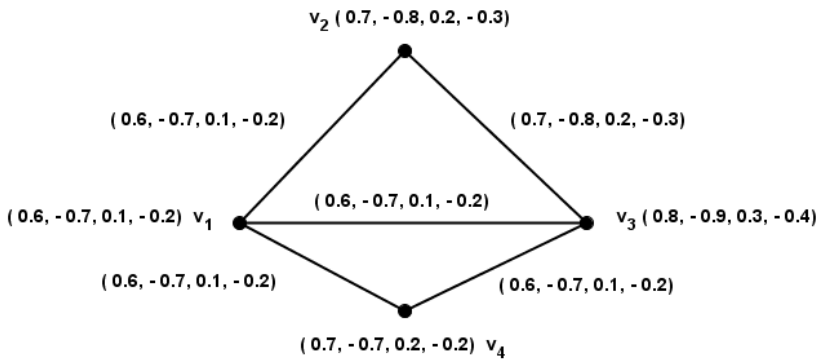


Figure 2: $G_{AQSP}(A, V)$ - *Corresponding to the parameter* e_1

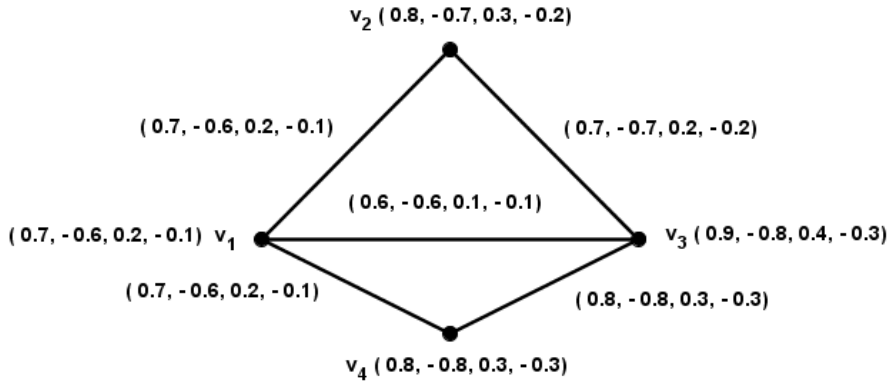


Figure 3: $G_{AQSP}(A, V)$ - *Corresponding to the parameter* e_2

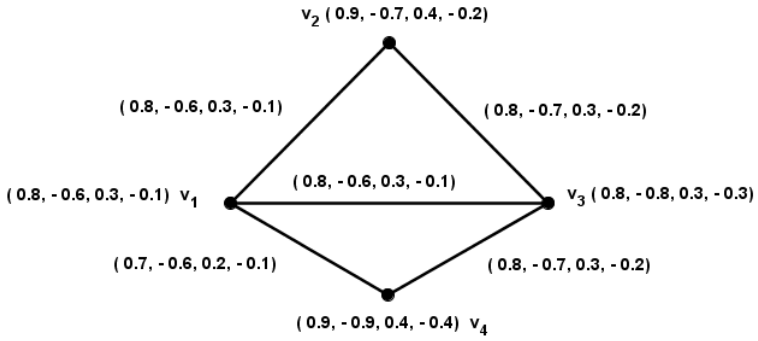


Figure 4: $G_{AQSP}(A, V)$ - *Corresponding to the parameter* e_3

3.3. Example of AQSP Fuzzy Soft Graph

Consider an AQSP fuzzy soft graph $G_{AQSP}(A, V)$, where $V = (v_1, v_2, v_3, v_4)$ and $E = (e_1, e_2, e_3)$. Here $G_{AQSP}(A, V)$ is described in Table.1. and $\mu_e(v_i, v_j) = 0$, $\forall (v_i, v_j) \in V \times V$ $\{(v_1, v_2), (v_2, v_3), (v_3, v_4), (v_1, v_4), (v_1, v_3)\}$ for all $e \in E$.

Table 1: *Tabular representation of AQSP Fuzzy Soft Graph parameter vertex set.*

(σ, ρ)	v_1	v_2	v_3	v_4
e_1	(0.6, - 0.7, 0.1,- 0.2)	(0.7, - 0.8, 0.2, -0.3)	(0.8, - 0.9, 0.3, - 0.4)	(0.6, - 0.7, 0.1,- 0.2)
e_2	(0.7, - 0.6, 0.2,- 0.1)	(0.8, - 0.7, 0.3, -0.2)	(0.9, - 0.8, 0.4, - 0.3)	(0.8, - 0.8, 0.3,- 0.3)
e_3	(0.8, - 0.6, 0.3,- 0.1)	(0.9, - 0.7, 0.4, -0.2)	(0.8, - 0.8, 0.3, - 0.3)	(0.9, - 0.9, 0.4,- 0.4)

Table 2: *Tabular representation of AQSP Fuzzy Soft Graph parameter edge set.*

(μ,γ)	v_1,v_2	v,v_3	v_3,v_4	v_4,v_1	v_1,v_3
e_1	(0.6, - 0.7, 0.1,- 0.2)	(0.7, - 0.8, 0.2, -0.3)	(0.6, -0.7, 0.1, - 0.2)	(0.6, - 0.7, 0.1,- 0.2)	(0.6, - 0.7, 0.1,- 0.2)
e_2	(0.7, - 0.6, 0.2,- 0.1)	(0.7, - 0.7, 0.2, -0.2)	(0.8, - 0.8, 0.3, - 0.3)	(0.7, - 0.6, 0.2,- 0.1)	(0.6, - 0.6, 0.1,- 0.1)
e_3	(0.8, - 0.6, 0.3,- 0.1)	(0.8, - 0.7, 0.3, -0.2)	(0.8, - 0.7, 0.3, - 0.2)	(0.7, - 0.6, 0.2,- 0.1)	(0.8, - 0.6, 0.3,- 0.1)

Table. 2. represents the AQSP fuzzy graph with parametric membership and non - membership with submerge values.

4. Results of Complete and μ - Complement of AQSP Fuzzy Soft Graph

4.1. Crisp graph of AQSP Fuzzy Soft Graph

Let $G_{AQSP}(A,V) = ((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP}),(\mu^P,\mu^N,\gamma^P,\gamma^N))$ be an AQSP fuzzy soft graph with underlying crisp graph is, $G^* = (\sigma^*,\mu^*)$, where
$\sigma^* = (v_i \in V : \sigma_e^P(v_i) > 0, \sigma_e^N(v_i) < 0, \rho_e^P(v_i) > 0, \rho_e^N(v_i) < 0)$
for some $e \in E$. $\mu^* = (v_i,v_j \in V \times V : \mu_e^P(v_i,v_j) > 0, \mu_e^N((v_i,v_j)) < 0,$
$\gamma_e^P((v_i,v_j)) > 0, \gamma_e^N((v_i,v_j)) < 0)$, $e \in E$.

4.2. Strong and Complete AQSP Fuzzy Soft Graph

Let $G_{AQSP}(A,V) = ((A_{AQSP},\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP},(\mu^P,\mu^N,\gamma^P,\gamma^N))$ is called as strong and complete AQSP fuzzy soft graph if,

(i) $\mu_e^P(x,y) = \sigma_e^P(x) \wedge \sigma_e^P(y)$, (ii) $\mu_e^N(x,y) = \sigma_e^N(x) \vee \sigma_e^N(y)$,
(iii) $\gamma_e^P(x,y) = \rho_e^P(x) \wedge \rho_e^P(y)$, (iv) $\gamma_e^N(x,y) = \rho_e^N(x) \vee \rho_e^N(y)$, for all $e \in A_{AQSP}$
and for all values of $x,y \in \mu^*$ is for strong AQSP fuzzy soft graph and for complete AQSP fuzzy soft graph is for all values of $x,y \in \sigma^*$.

4.3. Complement and μ- Complement of AQSP Fuzzy Soft Graph

Let $G_{AQSP}(A,V) = ((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP}),(\mu^P,\mu^N,\gamma^P,\gamma^N))$
be the AQSP fuzzy soft graph. The complement of AQSP fuzzy soft graph $G_{AQSP}(A,V)$ is defined as, $\overline{G}_{AQSP}(A,V) = ((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ (A_{AQSP}),(\mu^P,\mu^N,\gamma^P,\gamma^N))$, with the following sufficient conditions,

(i) $\overline{\mu}_e^P(x,y) = \sigma_e^P(x) \wedge \sigma_e^P(y) - \mu_e^P(x,y)$, (ii) $\overline{\mu}_e^N(x,y) = \sigma_e^N(x) \vee \sigma_e^N(y) \vee - \mu_e^N(x,y)$,
(iii) $\overline{\gamma}_e^P(x,y) = \rho_e^P(x) \wedge \rho_e^P(y) - \gamma_e^P(x,y)$, (iv) $\overline{\gamma}_e^N(x,y) = \rho_e^N(x) \vee \rho_e^N(y) - \gamma_e^N(x,y)$,
for all $e \in A_{AQSP}$ and for all values of $x,y \in V, e \in A_{AQSP}$.

4.4. Example of μ- Complement of AQSP Fuzzy Soft Graph

Consider an AQSP fuzzy soft graph $G_{AQSP}(A,V)$, where $V = (v_1,v_2,v_3,v_4)$ and

$E = (e_1,e_2,e_3)$. Here $G_{AQSP}(A,V)$ is described in Table.5. and $\mu_e(v_i,v_j) = 0$,
$\forall (v_i,v_j) \in V \times V\ \{(v_1,v_2),(v_2,v_3),(v_3,v_4),(v_2,v_4),(v_1,v_3)\}$ for all $e \in E$.

Table 3: *Tabular representation of AQSP Fuzzy Soft Graph parameter vertex set.*

(σ, ρ)	v_1	v_2	v_3	v_4
e_1	(0.7, - 0.8, 0.2,- 0.3)	(0.8, - 0.8, 0.3, -0.3)	(0.9, - 0.9, 0.4, - 0.4)	(0.9, - 0.6, 0.4,- 0.1)
e_2	(0.6, - 0.6, 0.1,- 0.1)	(0.7, - 0.7, 0.2, -0.2)	(0.8, - 0.8, 0.3, - 0.3)	(0.7, - 0.9, 0.2,- 0.4)
e_3	(0.8, - 0.8, 0.3,- 0.3)	(0.6, - 0.6, 0.1, -0.1)	(0.7, - 0.7, 0.2, - 0.2)	(0.9, - 0.9, 0.4,- 0.4)

Table 4: *Tabular representation of AQSP Fuzzy Soft Graph parameter edge set.*

(μ, γ)	v_1, v_2	v_2, v_3	v_3, v_4	v_4, v_2	v_1, v_3
e_1	(0.6, - 0.7, 0.1,- 0.2)	(0.8, - 0.7, 0.3, -0.2)	(0.8, -0.5, 0.3, 0.0)	(0.7, - 0.6, 0.2,- 0.1)	(0.7, - 0.7, 0.2,- 0.2)
e_2	(0.5, - 0.6, 0.0,- 0.1)	(0.6, - 0.6, 0.1, -0.1)	(0.6, - 0.7, 0.1, - 0.2)	(0.5, - 0.6, 0.0,- 0.1)	(0.5, - 0.5, 0.0, 0.0)
e_3	(0.6, - 0.6, 0.1,- 0.1)	(0.5, - 0.6, 0.0, -0.1)	(0.6, - 0.6, 0.1, - 0.1)	(0.5, - 0.5, 0.0, 0.0)	(0.6, - 0.6, 0.1,- 0.1)

Table. 5. represents the AQSP fuzzy graph with parametric membership and non - membership with submerge values.

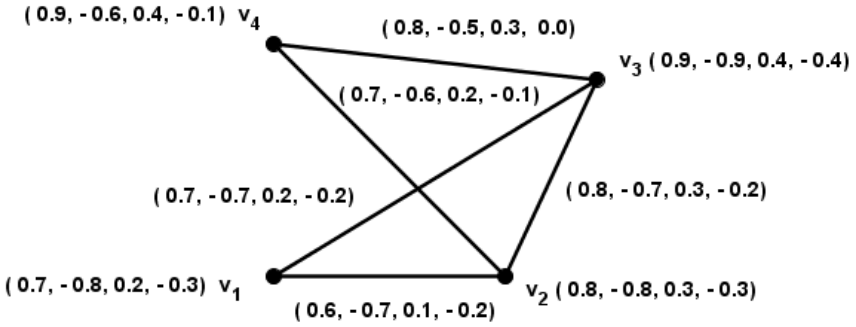


Figure 5: $G_{AQSP}(A, V)$ - *Corresponding to the parameter* e_1

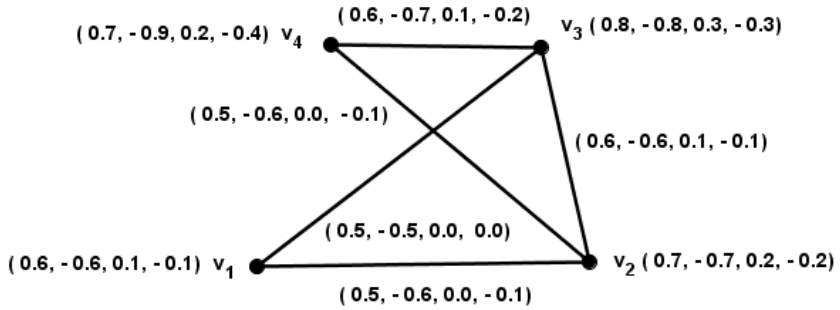


Figure 6: $G_{AQSP}(A, V)$ - *Corresponding to the parameter* e_2

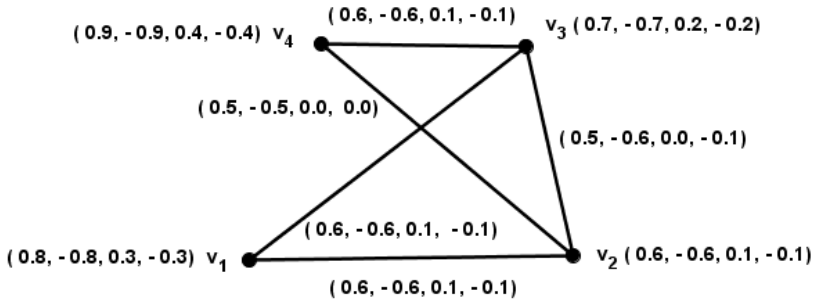


Figure 7: $G_{AQSP}(A,V)$ - *Corresponding to the parameter* e_3

4.5. Complement of AQSP Fuzzy Soft Graph

Consider an AQSP fuzzy soft graph $G_{AQSP}(A,V)$, where $V=(v_1,v_2,v_3,v_4)$ and $E=(e_1,e_2,e_3)$. Here $G_{AQSP}(A,V)$ is described in Figure 8,9,10 and we prove the Complement of AQSP fuzzy soft graph. $\mu_e(v_i,v_j)=0$,
$\forall(v_i,v_j)\in V\times V\ \{(v_1,v_2),(v_2,v_3),(v_3,v_4),(v_2,v_4),(v_1,v_3)\}$ for all $e\in E$.

Theorem 1. Sum of the Size of $(\overline{G}_{AQSP})$ and the size of $(G_{AQSP}(A,V))$ is equal to the result to twice the sum of its minimum and maximum membership and non - membership submerging AQSP fuzzy soft graph values.Then we prove the following result such as,

(i) $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y))$,
(ii)$S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^N(x)\vee\sigma_e^N(y))$,
(iii) $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^P(x)\wedge\rho_e^P(y))$,
(iv)$S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^N(x)\vee\rho_e^N(y))$,
$\forall x,y\in V,e\in A_{AQSP}$.

Proof. The order of the complement of AQSP fuzzy soft graph of $S(\overline{G}_{AQSP})$ is equal to the order of the AQSP fuzzy soft graph $G_{AQSP}(A,V)$ is obvious.

$$\mu_e^P(x,y)\leq\sigma_e^P(x)\wedge\sigma_e^P(y)\ \forall x,y\in V,e\in A_{AQSP} \tag{1}$$

$$\overline{\mu}_e^P(x,y)=\sigma_e^P(x)\wedge\sigma_e^P(y)-\mu_e^P(x,y)\ \forall x,y\in V,e\in A_{AQSP}$$

$$\mu_e^P(x,y)\leq\sigma_e^P(x)\wedge\sigma_e^P(y)\ \forall x,y\in V,e\in A_{AQSP} \tag{2}$$

From (1) and (2), we get $\mu_e^P(x,y)\leq\sigma_e^P(x)\wedge\sigma_e^P(y)\ \forall x,y\in V,e\in A_{AQSP}$.

(i) $\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\mu_e^P(x,y)+\overline{\mu}_e^P(x,y))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y))$
$\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\mu_e^P(x,y)+\sum_{e\in A_{AQSP}}\sum_{x\neq y}\overline{\mu}_e^P(x,y))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y))$.

Hence, $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y))$,

$$\mu_e^N(x,y)\geq\sigma_e^N(x)\vee\sigma_e^N(y)\ \forall x,y\in V,e\in A_{AQSP} \tag{3}$$

$$\overline{\mu}_e^N(x,y)=\sigma_e^N(x)\vee\sigma_e^N(y)-\mu_e^N(x,y)\ \forall x,y\in V,e\in A_{AQSP}$$

$$\mu_e^P(x,y)\geq\sigma_e^N(x)\wedge\sigma_e^N(y)\ \forall x,y\in V,e\in A_{AQSP} \tag{4}$$

From (3) and (4), we get $\mu_e^N(x,y)\geq\sigma_e^N(x)\vee\sigma_e^P(y)\ \forall x,y\in V,e\in A_{AQSP}$.

(ii) $\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\mu_e^N(x,y)+\overline{\mu}_e^N(x,y))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^(x)\wedge\sigma_e^N(y))$
$\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\mu_e^N(x,y)+\sum_{e\in A_{AQSP}}\sum_{x\neq y}\overline{\mu}_e^N(x,y))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^N(x)\wedge\sigma_e^N(y))$.
Hence, $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^N(x)\vee\sigma_e^N(y))$,

Similarly we get the result for submerging membership and non - membership values.
(iii) $\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\gamma_e^P(x,y)+\overline{\gamma}_e^P(x,y))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^P(x)\wedge\rho_e^P(y))$
$\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\gamma_e^P(x,y)+\sum_{e\in A_{AQSP}}$
$\sum_{x\neq y}\overline{\gamma}_e^P(x,y))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^P(x)\wedge\rho_e^P(y))$.
Hence, $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^P(x)\wedge\rho_e^P(y))$,

(iv) $\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\gamma_e^N(x,y)+\overline{\gamma}_e^N(x,y))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^N(x)\vee\rho_e^N(y))$
$\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\gamma_e^N(x,y)+\sum_{e\in A_{AQSP}}\sum_{x\neq y}\overline{\gamma}_e^N(x,y))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^N(x)\vee\rho_e^N(y))$.
Hence, $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^N(x)\vee\rho_e^N(y))$,

4.6. Example of Complement of AQSP fuzzy graph

Consider the AQSP fuzzy soft graph $(G_{AQSP}(A,V))$ in Figure. 5,6,7. and its complement of AQSP fuzzy soft graph $S(\overline{G}_{AQSP})$ Figure. 8,9,10. we get,the order of the complement of AQSP fuzzy soft graph, $S(\overline{G}_{AQSP})$ is equal to the order of AQSP fuzzy soft graph $S(G_{AQSP}(A,V))$.i.e. $O(\overline{G}_{AQSP})=O(G_{AQSP}(A,V))=(9.1,-9.1,3.1,-3.1)$

And, $S(\overline{G}_{AQSP})=(0.4,-0.6,0.4,-0.4), S(G_{AQSP}(A,V))=(6.1,-8.6,1.6,-1.8)$
$(\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y)=2(10.2)), \sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^N(x)\vee\sigma_e^N(y)=2(-9.2)),$
$\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^P(x)\wedge\rho_e^P(y)=2(2.7)), \sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^N(x)\vee\rho_e^N(y)=2(2.7))$.

Then we have 2(10.2, - 9.2, 2.7, - 2.7) = (20.4, - 18.4, 5.4, - 5.4). Therefore,

(i) $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y))$,
(ii)$S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^N(x)\vee\sigma_e^N(y))$,
(iii) $S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^P(x)\wedge\rho_e^P(y))$,
(iv)$S(\overline{G}_{AQSP})+S(G_{AQSP}(A,V))\geq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\rho_e^N(x)\vee\rho_e^N(y))\ \forall x,y\in V, e\in A_{AQSP}$.

4.7. Complement of AQSP fuzzy soft graph

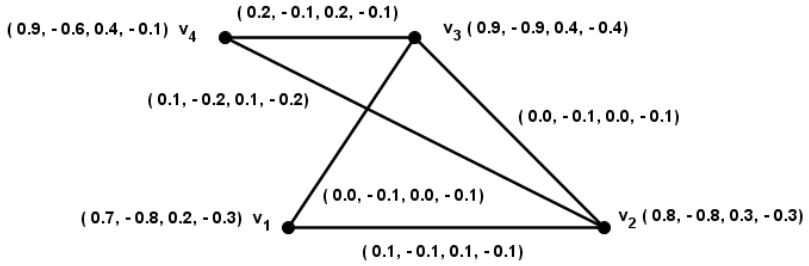


Figure 8: *Complement of $G_{AQSP}(A,V)$ - Corresponding to the parameter e_1*

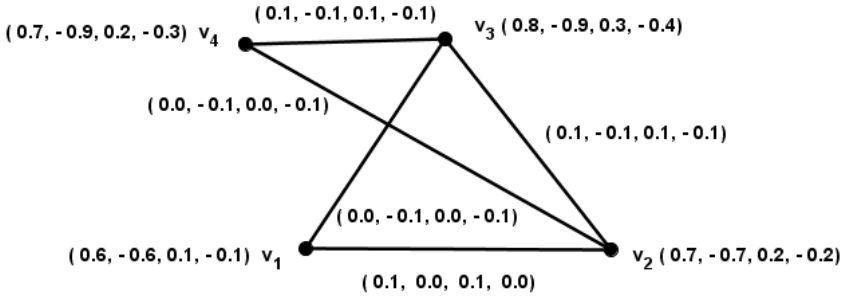


Figure 9: *Complement of $G_{AQSP}(A, V)$ - Corresponding to the parameter e_2*

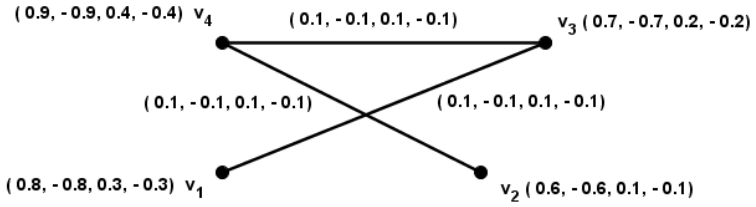


Figure 10: *Complement of $G_{AQSP}(A, V)$ - Corresponding to the parameter e_3*

4.8. Remark

The order of the complement of AQSP fuzzy soft graph $\overline{G}_{AQSP}(A, V)$ is equal to the order of the AQSP fuzzy soft graph $G_{AQSP}(A, V)$.

5. Properties of μ - Complement of AQSP Fuzzy Soft Graph

(i) The order of the complement AQSP fuzzy soft graph, $O(\overline{G}_{AQSP})$ is equal to the order of AQSP fuzzy soft graph $O(G_{AQSP}(A, V))$. And, $O(\overline{G^{\mu}}_{AQSP}) = O(G_{AQSP}(A, V))$ is presented in the Example.4.7. of AQSP fuzzy soft graph.

(ii) Vertex set of $(\overline{G^{\mu}}_{AQSP}) = (G_{AQSP}(A, V)$

(iii) The number of elements to the edge set of $\overline{G^{\mu}}$ is less than the number of elements in the node set of $(G_{AQSP}(A, V)$.

(iv) $\mu_e^{\mu^P}(x, y) > 0$ if $(x, y) \in \mu^*$, *otherwise* $\mu_e^{\mu^P}(x, y) = 0$

(v) $\mu_e^{\mu^N}(x, y) < 0$ if $(x, y) \in \mu^*$, *otherwise* $\mu_e^{\mu^N}(x, y) = 0$

(vi) $\mu_e^{\gamma^P}(x, y) > 0$ if $(x, y) \in \gamma^*$, *otherwise* $\gamma_e^{\gamma^P}(x, y) = 0$

(vii) $\mu_e^{\gamma^N}(x, y) < 0$ if $(x, y) \in \gamma^*$, *otherwise* $\gamma_e^{\gamma^N}(x, y) = 0$

For the size of $G_{AQSP}(A, V)$, μ complement AQSP fuzzy soft graph membership value is,
(viii) $S(G_{AQSP}(A, V) = \sum_{e \in A_{AQSP}} \sum_{x \neq y} (\mu_e \mu^P(x, y))$
$= \sum_{e \in A_{AQSP}} (\sum_{x,y \in \mu^*} (\sigma_e^P(x) \wedge \sigma_e^P(y) - \mu_e^P(x, y) \ \forall x, y \in V, e \in A_{AQSP}))$
$= \sum_{e \in A_{AQSP}} (\sum_{x,y \in \mu^*} (\sigma_e^P(x) \wedge \sigma_e^P(y) - \sum_{e \in A_{AQSP}} (\sum_{x,y \in \mu^*} \ x, y \in \mu_e^P(x, y)$

$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*}(\sigma_e^P(x)\wedge\sigma_e^P(y) - S(G_{AQSP}(A,V))$
i.e. $S(\overline{G}_{AQSP}) + S(G_{AQSP}(A,V)) \leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^P(x)\wedge\sigma_e^P(y))$.

For the $G_{AQSP}(A,V)$, μ complement AQSP fuzzy soft graph non - membership value is,

(ix) $S(G_{AQSP}(A,V) = \sum_{e\in A_{AQSP}}\sum_{x\neq y}(\mu_e\mu^N(x,y))$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*}(\sigma_e^N(x)\vee\sigma_e^N(y) - \mu_e^N(x,y)\ \forall x,y\in V, e\in A_{AQSP}))$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*}(\sigma_e^N(x)\vee\sigma_e^N(y) - \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*} x,y\in\mu_e^N(x,y)$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*}(\sigma_e^N(x)\vee\sigma_e^N(y) - S(G_{AQSP}(A,V))$
i.e. $S(\overline{G}_{AQSP}) + S(G_{AQSP}(A,V)) \leq 2\sum_{e\in A_{AQSP}}\sum_{x\neq y}(\sigma_e^N(x)\vee\sigma_e^N(y))$

For the size of $G_{AQSP}(A,V)$, μ complement AQSP fuzzy soft graph membership value is,

(x) $S(G_{AQSP}(A,V) = \sum_{e\in A_{AQSP}}\sum_{x\neq y}(\gamma_e\gamma^P(x,y))$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\gamma^*}(\rho_e^P(x)\wedge\rho_e^P(y) - \gamma_e^P(x,y)\ \forall x,y\in V, e\in A_{AQSP}))$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\gamma^*}(\rho_e^P(x)\wedge\rho_e^P(y) - \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*}\ x,y\in\gamma_e^P(x,y)$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\gamma^*}(\rho_e^P(x)\wedge\rho_e^P(y) - S(G_{AQSP}(A,V))$.

For $G_{AQSP}(A,V)$, μ complement AQSP fuzzy soft graph submerging value is

(xi) $S(G_{AQSP}(A,V) = \sum_{e\in A_{AQSP}}\sum_{x\neq y}(\gamma_e\gamma^N(x,y))$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\gamma^*}(\rho_e^N(x)\vee\rho_e^N(y) - \gamma_e^N(x,y)\ \forall x,y\in V, e\in A_{AQSP}))$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\gamma^*}(\rho_e^N(x)\vee\rho_e^N(y) - \sum_{e\in A_{AQSP}}(\sum_{x,y\in\mu^*}\ x,y\in\gamma_e^N(x,y)$
$= \sum_{e\in A_{AQSP}}(\sum_{x,y\in\gamma^*}(\rho_e^N(x)\vee\rho_e^N(y) - S(G_{AQSP}(A,V))$

Theorem 2. The complement of a strong AQSP fuzzy soft graph $\overline{G}_{AQSP}(A,V)$ is also strong AQSP fuzzy soft graph $G_{AQSP}(A,V)$.

Proof. Let $G_{AQSP}(A,V)$ be an strong AQSP fuzzy soft graph by definition 4.3.of the complement of a strong AQSP fuzzy soft graph for the membership values,

$$\begin{aligned}
\overline{\mu}^P(x,y) &= \sigma_e^P(x)\wedge\sigma_e^P(y) - \mu_e^P(x,y)\ \forall x,y\in V\times V, e\in A_{AQSP},\\
&= \sigma_e^P(x)\wedge\sigma_e^P(y) - (\sigma_e^P(x)\wedge\sigma_e^P(y)), \mu_e^P(x,y) > 0,\\
&\sigma_e^P(x)\wedge\sigma_e^P(y), \qquad \mu_e^P(x,y) = 0,\\
&= 0, \qquad \mu_e^P(x,y) > 0,\\
&\sigma_e^P(x)\wedge\sigma_e^P(y), \qquad \mu_e^P(x,y) = 0,\\
&= 0, \qquad \overline{\mu}_e^P(x,y) = 0,\\
&\sigma_e^P(x)\wedge\sigma_e^P(y), \qquad \overline{\mu}_e^P(x,y) > 0.\\
\overline{\mu}^P(x,y) &= \sigma_e^P(x)\wedge\sigma_e^P(y), \overline{\mu}_e^P(x,y) = 0\ \forall x,y\in V\times V,
\end{aligned}$$

where (x,y) is the edge $\forall,\ (x,y)\in\overline{\mu}$.

The complement of a strong AQSP fuzzy soft graph for the non - membership values,

$$\begin{aligned}
\overline{\mu}^N(x,y) &= \sigma_e^N(x)\vee\sigma_e^N(y) - \mu_e^N(x,y)\ \forall x,y\in V\times V, e\in A_{AQSP},\\
&= \sigma_e^N(x)\vee\sigma_e^N(y) - (\sigma_e^N(x)\vee\sigma_e^N(y)), \mu_e^N(x,y) < 0,\\
&\sigma_e^N(x)\vee\sigma_e^N(y), \qquad \mu_e^N(x,y) = 0,\\
&= 0, \qquad \mu_e^N(x,y) < 0,\\
&\sigma_e^N(x)\vee\sigma_e^N(y), \qquad \mu_e^N(x,y) = 0,\\
&= 0, \qquad \overline{\mu}_e^N(x,y) = 0,\\
&\sigma_e^N(x)\vee\sigma_e^N(y), \qquad \overline{\mu}_e^N(x,y) < 0.\\
\overline{\mu}^N(x,y) &= \sigma_e^N(x)\vee\sigma_e^N(y), \overline{\mu}_e^N(x,y) = 0\ \forall x,y\in V\times V,
\end{aligned}$$

where (x,y) is the edge $\forall,\ (x,y)\in\overline{\mu}$.

Similarly the complement of the submerging AQSP fuzzy soft graph membership values are,

$$\begin{aligned}\overline{\gamma}^P(x,y) &= \rho_e^P(x) \wedge \rho_e^P(y) - \gamma_e^P(x,y)\ \forall x,y \in V \times V, e \in A_{AQSP},\\ &= \rho_e^P(x) \wedge \rho_e^P(y) - (\rho_e^P(x) \wedge \rho_e^P(y)), \gamma_e^P(x,y) > 0,\\ \overline{\gamma}^P(x,y) &= \rho_e^P(x) \wedge \rho_e^P(y), \overline{\gamma}_e^P(x,y) = 0\ \forall x,y \in V \times V,\end{aligned}$$

where (x,y) is the edge $\forall,\ (x,y) \in \overline{\gamma}$.

For the complement of the submerging AQSP fuzzy soft graph non - membership values are,

$$\begin{aligned}\overline{\gamma}^N(x,y) &= \rho_e^N(x) \wedge \rho_e^N(y) - \gamma_e^N(x,y)\ \forall x,y \in V \times V, e \in A_{AQSP}\ ,\\ &= \rho_e^N(x) \wedge \rho_e^N(y) - (\rho_e^N(x) \wedge \rho_e^N(y)), \gamma_e^N(x,y) > 0,\\ \overline{\gamma}^N(x,y) &= \rho_e^N(x) \wedge \rho_e^N(y), \overline{\gamma}_e^N(x,y) = 0\ \forall x,y \in V \times V,\end{aligned}$$

where (x,y) is the edge $\forall,\ (x,y) \in \overline{\gamma}$. Hence,the theorem is completed.

Theorem 3. The complement of a complete AQSP fuzzy soft graph $\overline{G}_{AQSP}(A,V)$ is also complete AQSP fuzzy soft graph $G_{AQSP}(A,V)$.

Proof. Let $G_{AQSP}(A,V) = ((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP}),(\mu^P,\mu^N,\gamma^P,\gamma^N))$ be the complete AQSP fuzzy soft graph by definition 4.3.of the complement of a complete AQSP fuzzy soft graph for the membership values,

$$\begin{aligned}\overline{\mu}^P(x,y) &= \sigma_e^P(x) \wedge \sigma_e^P(y) - \mu_e^P(x,y)\ \forall x,y \in V, e \in A_{AQSP}\ ,\\ &= \sigma_e^P(x) \wedge \sigma_e^P(y) - (\sigma_e^P(x) \wedge \sigma_e^P(y)), \mu_e^P(x,y) > 0,\\ &\sigma_e^P(x) \wedge \sigma_e^P(y), \qquad \mu_e^P(x,y) = 0,\\ &= 0, \qquad \mu_e^P(x,y) > 0,\\ &\sigma_e^P(x) \wedge \sigma_e^P(y), \qquad \mu_e^P(x,y) = 0,\\ &= 0, \qquad \overline{\mu}_e^P(x,y) = 0,\\ &\sigma_e^P(x) \wedge \sigma_e^P(y), \qquad \overline{\mu}_e^P(x,y) > 0.\\ \overline{\mu}^P(x,y) &= \sigma_e^P(x) \wedge \sigma_e^P(y), \overline{\mu}_e^P(x,y) = 0\ \forall x,y \in V, \forall\ (x,y) \in \overline{\sigma^*}.\end{aligned}$$

The complement of a Complete AQSP fuzzy soft graph for the non - membership values,

$$\begin{aligned}\overline{\mu}^N(x,y) &= \sigma_e^N(x) \vee \sigma_e^N(y) - \mu_e^N(x,y)\ \forall x,y \in V, e \in A_{AQSP}\ ,\\ &= \sigma_e^N(x) \vee \sigma_e^N(y) - (\sigma_e^N(x) \vee \sigma_e^N(y)), \mu_e^N(x,y) < 0,\\ &\sigma_e^N(x) \vee \sigma_e^N(y), \qquad \mu_e^N(x,y) = 0,\\ &= 0, \qquad \mu_e^N(x,y) < 0,\\ &\sigma_e^N(x) \vee \sigma_e^N(y), \qquad \mu_e^N(x,y) = 0,\\ &= 0, \qquad \overline{\mu}_e^N(x,y) = 0,\\ &\sigma_e^N(x) \vee \sigma_e^N(y), \qquad \overline{\mu}_e^N(x,y) < 0,\\ \overline{\mu}^N(x,y) &= \sigma_e^N(x) \vee \sigma_e^N(y), \overline{\mu}_e^N(x,y) = 0\ \forall x,y \in V, \forall\ (x,y) \in \overline{\sigma^*}.\end{aligned}$$

For the complement of the submerging Complete AQSP fuzzy soft graph membership values ,

$$\begin{aligned}\overline{\gamma}^P(x,y) &= \rho_e^P(x) \wedge \rho_e^P(y) - \gamma_e^P(x,y)\ \forall x,y \in V, e \in A_{AQSP}\ ,\\ &= \rho_e^P(x) \wedge \rho_e^P(y) - (\rho_e^P(x) \wedge \rho_e^P(y)), \gamma_e^P(x,y) > 0\\ \overline{\gamma}^P(x,y) &= \rho_e^P(x) \wedge \rho_e^P(y), \overline{\gamma}_e^P(x,y) = 0\ \forall x,y \in V, \forall\ (x,y) \in \overline{\sigma^*}.\end{aligned}$$

Similarly we get result for the complement of the submerging Complete AQSP fuzzy soft graph non - membership values ,$\overline{\gamma}^N(x,y) = \rho_e^N(x) \wedge \rho_e^N(y), \overline{\gamma}_e^N(x,y) = 0\ \forall x,y \in V$, where (x,y) denoted the vertices for all $(x,y) \in \overline{\rho^*}$. Hence the proof.

Theorem 4. Let the AQSP fuzzy soft graph be $G_{AQSP}(A,V)$. Then $G_{AQSP}(A,V)$ is an isolated AQSP fuzzy soft graph if and only if $\overline{G}_{AQSP}(A,V)$ is a complete AQSP fuzzy soft graph.

Proof. $G_{AQSP}(A,V) = ((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP}),(\mu^P,\mu^N,\gamma^P,\gamma^N))$ be the AQSP fuzzy soft graph. Then $G_{AQSP}(A,V)$.Let $\overline{G}_{AQSP}(A,V)$ be its complement of AQSP fuzzy soft graph. Then the given isolated AQSP fuzzy soft graph is $G_{AQSP}(A,V)$. Then,
$\mu_e^P(x,y) = 0, \forall (x,y) \in V \times V, e \in A_{AQSP}$.
Since,$\overline{\mu}^P(x,y) = \sigma_e^P(x) \wedge \sigma_e^P(y) - \mu_e^P(x,y)\ \forall x,y \in V \times V, e \in A_{AQSP}$,
$\overline{\mu}^P(x,y) = \sigma_e^P(x) \wedge \sigma_e^P(y), \forall\ x,y \in V \times V, e \in A_{AQSP}$,
Hence , $\overline{G}_{AQSP}(A,V)$ is complete AQSP fuzzy soft graph. Conversely, Given $\overline{G}_{AQSP}(A,V)$ to be a complete AQSP fuzzy soft graph. $\overline{\mu}^P(x,y) = \sigma_e^P(x) \wedge \sigma_e^P(y), \forall x,y \in V \times V, e \in A_{AQSP}$,
Since,$\mu^P(x,y) = \sigma_e^P(x) \wedge \sigma_e^P(y) - \overline{\mu}_e^P(x,y)\ \forall x,y \in V \times V, e \in A_{AQSP}$,
$\overline{\mu}^P(x,y) - \overline{\mu}^P(x,y), \forall\ x,y \in V \times V, e \in A_{AQSP}$,
$= 0$, $\forall\ x,y \in V \times V, e \in A_{AQSP}$,
$\mu^P(x,y) = 0, \forall\ x,y \in V \times V, e \in A_{AQSP}$.
Similarly, we get the result for non - membership values and submerge values such as,
$\overline{\mu}^N(x,y) = \sigma_e^N(x) \vee \sigma_e^N(y) - \mu_e^N(x,y)\ \forall x,y \in V \times V, e \in A_{AQSP}$,
$\overline{\mu}^N(x,y) = \sigma_e^N(x) \vee \sigma_e^N(y), \forall\ x,y \in V \times V, e \in A_{AQSP}$.
Hence $\overline{G}_{AQSP}(A,V)$ is complete AQSP fuzzy soft graph.

6. Application of AQSP Fuzzy Soft Graph

6.1. Analysis of AQSP fuzzy soft graph in Water - related diseases.

The analysis of the Water related diseases is done for different hostel students in the same locality. This kind of diseases occur by drinking polluted water . Children make up the majority of harmed diseases by contaminated water. This leads to a number of common ailments such as Diarrehea, Dysentery, Cholera, and Typhoid fever. We use the AQSP fuzzy soft graph module to find the most common diseases that the students are affected. And the corresponding parameteric symptoms of diesases is presented in AQSP fuzzy soft edges. The following descriptions will pave the way to find the cause of this sicknesess to precise the correct medicine .

6.2. Description of the Analysis

1. Let us consider the AQSP fuzzy soft sets such as, $((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP}),\ (\mu^P,\mu^N,\gamma^P,\gamma^N))$. Which is the parametric set taken as the different symptoms of Water diseases .

2. Specify the vertex and edge sets of AQSP fuzzy soft graphs $G_{AQSP}(A,V)$, which corresponds to the symptoms of Water related diseases of the students in the hostel .

3. Measure the most common symptoms of this sickness by taking AQSP fuzzy soft graph membership and non - membership values with submerging level.

4. Calculate the score values of the $((A_{AQSP}),\ (\sigma^P,\sigma^N,\rho^P,\rho^N)),\ ((A_{AQSP}),(\mu^P,\mu^N,\gamma^P,\gamma^N))$ by using the score function, $\frac{1}{2}\ \ (\frac{1}{S^P_{AQSP}}\ \sum\theta^P_x\ -\ \frac{1}{S^N_{AQSP}}\ \sum\theta^N_x\ )$.

5. The maximum score membership value in AQSP fuzzy soft graph $G_{AQSP}(A,V)$, is the most common symptoms of Water related diseases.

6. Consider AQSP fuzzy soft vertex set, v_1 = Typhoid fever, v_2 = Diarrehea, v_3 = Dysentery, and v_4 = Cholera.

6.3. Discussion of the New AQSP fuzzy soft graph

We consider AQSP fuzzy soft graph corresponding to the parameter e_1 as,

v_1, v_2 = Fever, v_2, v_3 = Diarrehea, v_3, v_4 = Muscles aches, v_1, v_4 = Sweating, v_1, v_3 = Fatigue.

The AQSP fuzzy soft graph Corresponding to the parameter e_2, is

v_1, v_2 = Vomiting, v_2, v_3 = Muscles Cramps, v_3, v_4 = Nausea,
v_1, v_4 = Diarrehea, v_1, v_3 = Head ache.

The AQSP fuzzy soft graph Corresponding to the parameter e_3 is, v_1, v_2 = Cramps and Bloating, v_2, v_3 = Weightloss,

v_3, v_4 = Nausea, v_1, v_4 = Diarrehea, v_1, v_3 = Abdomend pain. The following Table.5. presents the membership and non - membership submerging values .

Table 5: *Tabular representation of AQSP Fuzzy Soft Graph parameter vertex set.*

(σ, ρ)	v_1	v_2	v_3	v_4
e_3	(0.8, - 0.8, 0.3,- 0.3)	(1.0, - 1.0, 0.5, -0.5)	(0.8, - 0.8, 0.3, - 0.3)	(0.9, - 0.9, 0.4,- 0.4)
e_1	(0.7, - 0.8, 0.2,- 0.3)	(0.8, - 0.8, 0.3, -0.3)	(0.9, - 0.9, 0.4, - 0.4)	(0.9, - 0.6, 0.4,- 0.1)
e_2	(0.6, - 0.6, 0.1,- 0.1)	(0.7, - 0.7, 0.2, -0.2)	(0.8, - 0.8, 0.3, - 0.3)	(0.7, - 0.9, 0.2,- 0.4)

Table 6: *Tabular representation of AQSP Fuzzy Soft Graph parameter edge set.*

(μ, γ)	v_1, v_2	v_2, v_3	v_3, v_4	v_4, v_1	v_1, v_3
e_1	(0.6, - 0.7, 0.1,- 0.2)	(0.8, - 0.7, 0.3, -0.2)	(0.8, -0.5, 0.3, 0.0)	(0.7, - 0.6, 0.2,- 0.1)	(0.7, - 0.7, 0.2,- 0.2)
e_2	(0.5, - 0.6, 0.0,- 0.1)	(0.6, - 0.6, 0.1, -0.1)	(0.6, - 0.7, 0.1, - 0.2)	(0.5, - 0.6, 0.0,- 0.1)	(0.5, - 0.5, 0.0, 0.0)
e_3	(0.6, - 0.6, 0.1,- 0.1)	(0.5, - 0.6, 0.0, -0.1)	(0.6, - 0.6, 0.1, - 0.1)	(0.5, - 0.5, 0.0, 0.0)	(0.6, - 0.6, 0.1,- 0.1)

Table 7: ***D****ifferent Hostel students affected by Water related diseases Score values.*

Hostel - 1 $(\sigma, \rho)v_1$) Score	**Hostel - 2** $(\sigma, \rho)v_2$) Score	**Hostel -3** $(\sigma, \rho)v_3$ Score	**Hostel - 4** $(\sigma, \rho)v_4$ Score
0.576	1.000	0.543	0.990
0.502	0.476	0.499	0.456
0.476	0.733	0.654	0.630

(i) The most affected common diseases from different water - related disease is

v_2=Diarrehea, which is the main symptoms of the students in different hostels in the same locality. The score value of the disease Diarrehea is, v_2 = 1.000 .

(ii) Corresponding to the parameter e_1 score value of (v_2, v_3) = 0.648.

Table 8: ***D****ifferent Hostel students affected by Water related diseases Score values.*

Hostel - 1	**Hostel - 2**	**Hostel -3**	**Hostel - 4**
($\mu, \gamma)v_1, v_2$ Score	$(\mu, \gamma)v_2, v_3$ Score	$(\mu, \gamma)v_3, v_4$ Score	$(\mu, \gamma)v_1, v_4$ Score
0.489	0.648	0.509	0.634
0.478	0.500	0.487	0.466
0.485	0.646	0.500	0.633

Many students are affected by these sicknesess such as, v_1, v_2 = Fever, v_2, v_3 = Diarrehea, v_3, v_4 = Muscles aches, v_1, v_4 = Sweating, v_1, v_3 = Fatigue.

(iii)The least affected common diseases from different water - related disease is

v_4= Cholera, which is the main symptoms of the students in different hostels in the same locality. The score value of the disease Diarrehea is, v_4 = 0.456 .

(iv) Corresponding to the parameter e_1 score value of (v_1, v_4) = 0.466.

7. Conclusion

In this artifact , AQSP fuzzy soft graph definitions, complement of AQSP fuzzy soft graphs are introduced with theorems and examples. Some results about the strong AQSP fuzzy soft graph, complete AQSP fuzzy soft graph with μ - complement AQSP fuzzy soft graph and isolated AQSP fuzzy soft graph with complements is constructed.The analysis of water - related deseases result is the invention of AQSP fuzzy soft graph module.

Declarations
Acknowledgements
The authors do thankful to the editor for giving an opportunity to submit our research article in this esteemed journal. And grateful to the Institution for providing SEED Money and MATLAB software for the research purpose.
Conflict of interest
The authors declared that they have no conflict of interest regarding the publication of the research article.

References

[1] Akram, Muhammad. *Bipolar fuzzy graphs* Information sciences 181, no. 24 (2011): 5548-5564.
[2] Akram, Muhammad, and Saira Nawaz. *Fuzzy soft graphs with applications* Journal of Intelligent and Fuzzy Systems 30, no. 6 (2016): 3619-3632.
[3] Akram, Muhammad, and Sundas Shahzadi. *Novel intuitionistic fuzzy soft multiple-attribute decision-making methods* Neural Computing and Applications 29 (2018): 435-447.
[4] Akram, Muhammad, and Saira Nawaz. *On fuzzy soft graphs* Italian journal of pure and applied mathematics 34 (2015): 497-514.
[5] Atanassov, Krassimir T., and Krassimir T. Atanassov. *Intuitionistic fuzzy relations (IFRs)* On intuitionistic fuzzy sets theory (2012): 147-193.
[6] Bhattacharya, Prabir. *Some remarks on fuzzy graphs* Pattern recognition letters 6, no. 5 (1987): 297-302.
[7] Chellamani, P., D. Ajay, Said Broumi, and T. Antony Alphonse Ligori. *An approach to decision-making via picture fuzzy soft graphs* Granular Computing (2021): 1-22.
[8] Gani, A. Nagoor, R. Jahir Hussain, and S. Yahya Mohamed. *Irregular Intuitionistic fuzzy graph* IOSR Journal of Mathematics (IOSR-JM) 9 (2014): 47-51.

[9] Gani, A. Nagoor, and K. Radha. *Regular property of fuzzy graphs* Bulletin of Pure and Applied Science 27, no. 2 (2008): 415-423.
[10] Kaufmann, Arnold. *Introduction to the theory of fuzzy subsets* Academic press, 1975.
[11] Maji, Pabitra Kumar. *More on intuitionistic fuzzy soft sets*. In Rough Sets, Fuzzy Sets, Data Mining and Granular Computing: 12th International Conference, RSFDGrC 2009, Delhi, India, December 15-18, 2009. Proceedings 12, pp. 231-240. Springer Berlin Heidelberg, 2009.
[12] Molodtsov, Dmitriy. *Soft set theory—first results* Computers and mathematics with applications 37, no. 4-5 (1999): 19-31.
[13] Mohinta, Sumit, and T. K. Samanta. *An introduction to fuzzy soft graph* Mathematica Moravica 19, no. 2 (2015): 35-48.
[14] Moderson, J. N., and P. S. Nair. *Fuzzy Graphs and Fuzzy Hypergraphs* Heidelberg, Germany: Physica-Verlag Heidelberg, 2000. DOI: 10.1007.
[15] Nawaz, Hafiza Saba, and Muhammad Akram. *Oligopolistic competition among the wireless internet service providers of Malaysia using fuzzy soft graphs* Journal of Applied Mathematics and Computing (2021): 1-36.
[16] Rosenfeld, Azriel. *Fuzzy graphs* In Fuzzy sets and their applications to cognitive and decision processes, pp. 77-95. Academic press, 1975
[17] Rosline, J. Jesintha, and T. Pathinathan. *Triple layered fuzzy graph* International Journal of Fuzzy Mathematical Archieve 8, no. 1 (2015): 36-42.
[18] Shahzadi, Gulfam, Muhammad Akram, and Bijan Davvaz. *Pythagorean fuzzy soft graphs with applications* Journal of Intelligent and Fuzzy Systems 38, no. 4 (2020): 4977-4991.
[19] Sunitha, M. S., and A. Vijayakumar. *Studies on fuzzy graphs* PhD diss., Department of Mathematics, 2001.
[20] Zadeh, L. *Fuzzy sets* Inform Control 8 (1965): 338-353.
[21] Zhang, Wen-Ran, and Lulu Zhang. *YinYang bipolar logic and bipolar fuzzy logic* Information Sciences 165, no. 3-4 (2004): 265-287.

WEIGHTED GENERALIZED ENTROPY: PROPRTIES AND APPLICATION

Bilal Ahmad Bhat[1*], M. Sultan Shah[2] and M.A.K Baig[3]

•

Department of Statistics, University of Kashmir, Srinagar, J&K-190006, India
bhatbilal3819@gmail.com
msultanshahh01@gmail.com
baigmak@gmail.com
*Corresponding Author

Abstract

Recently, the measurement of uncertainty has attracted the attention of researchers. In this article, we introduce a new weighted uncertainty measure known as weighted generalized entropy. We also study its dynamic (residual) version which is known as weighted generalized residual entropy. These are length-biased shift-dependent uncertainty measures. It is shown that the proposed dynamic uncertainty measure uniquely determines the survival function. The various significant properties and the relationship with other well-known reliability measures of the proposed dynamic uncertainty measure are also studied. Finally, a real life data set is used to illustrate the usefulness of our proposed uncertainty measures.

Keywords: Weighted entropy, weighted residual entropy, hazard rate function and characterization results.

1. Introduction

The notion of entropy that was introduced by Shannon [1] is a very important and well known concept in the area of information theory. For an absolutely continuous non-negative r.v U having p.d.f $g(u)$, the Shannon's entropy (SE) is defined as

$$H_U(g) = -\int_0^\infty g(u)\log g(u)du = -E[\log(U)]. \quad (1)$$

Throughout this article, the notations r.v and p.d.f stands for an absolutely continuous non-negative random variable and the probability density function respectively.

If a lifetime component has survived up to an age t, then the SE is not useful for measuring the uncertainty about its remaining life. To overcome this problem, Ebrahimi [2] has introduced the concept of residual entropy and is defined as

$$H_U(g;t) = -\int_t^\infty \frac{g(u)}{\bar{G}(t)}\log\frac{g(u)}{\bar{G}(t)}du, \quad (2)$$

where, $\bar{G}(t) = 1 - G(t)$ is the survival function (s.f) of the r.v U.

It is clear that the SE is well-known by means of its applications in the area of information theory, but it is a shift-independent uncertainty measure (UM) because it remains unchanged, if for instance U is uniformly distributed in (c,d) or $(c+h,d+h)$ for any $h \in \mathcal{R}$. However, in some applied contexts, such as reliability or mathematical neurobiology, the shift-dependent UM's are

desirable. To fulfill this requirement, Belis and Guiasu [3] have introduced the concept of weighted entropy (a shift-independent UM) and is defined as

$$H_{(U,w)}(g) = -\int_0^\infty w(u)g(u)\log g(u)du$$
$$= -ug(u)\log g(u)du\,, \tag{3}$$

where, the coefficient u (i.e the length of the system or component under consideration) represents the weight function of the elementary events.

Similarly, Di Crescenzo and Longobardi [4] have introduced the weighted version of residual entropy (2) and is given by

$$H_{(U,w)}(g;t) = -\int_t^\infty u\frac{g(u)}{\bar{G}(t)}\log\frac{g(u)}{\bar{G}(t)}du\,. \tag{4}$$

In the recent literature, it is seen that the study of weighted UM's have attracted the attention of researchers for introducing the new flexible weighted UM's. For more details see Misagh et al. [5], Misagh and Yari [6], Nourbakhsh and Yari [7], Mirali and Baratpour [8], Kayal [9], Nair et al. [10], Rajesh et al. [11], Khammar and Jahanshahi [12], Bhat and Baig [13] and Bhat et al. [14] etc. Motivated with this research literature, here in this article, our objective is to introduce a new weighted UM and its dynamic (residual) version on the basis of the following new generalization of SE

$$H_U^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left(\int_0^\infty g^{2\frac{\eta}{\mu}-1}(u)du\right),\ \frac{\mu}{2} < \eta < \mu,\ \mu \geq 1, \tag{5}$$

where,

$\lim_{\substack{\eta\to 1\\ \mu=1}} H_U^{(\eta,\mu)}(g) = -\int_0^\infty g(u)\log g(u)du$, which is the SE given in (1).

Analogous to (2) and on the basis of (5), the generalized residual entropy can be defined as

$$H_U^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\log\left(\int_t^\infty\left(\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du\right),\ \frac{\mu}{2} < \eta < \mu, \mu \geq 1. \tag{6}$$

The rest of the article is organized as follows: In section 2, we discuss the weighted generalized entropy (WGE) of order η and type μ in the form of its definition and some properties. The section 3 presents the weighted generalized residual entropy (WGRE) and also some of its significant characterization results. In section 4, we study the various important properties of WGRE and also its relationship with other well-known reliability measures. In section 5, an application of the WGE and WGRE by using a real life data set is presented. Finally, we illustrate some concluding remarks in section 6.

2. Weighted Generalized Entropy (WGE)

In this section, we introduce the weighted version of (5) which is known as weighted generalized entropy (WGE) of order η and type μ.

Definition 2.1 For a r.v U having p.d.f $g(u)$, the WGE of order η and type μ denoted by $H_{(U,w)}^{(\eta,\mu)}(g)$ is defined as

$$H_{(U,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left(\int_0^\infty\left(ug(u)\right)^{2\frac{\eta}{\mu}-1}du\right),\frac{\mu}{2} < \eta < \mu \geq 1, \tag{7}$$

where, the coefficient u in the integrand denotes the weight function as in (3).

In the following example, we illustrate the importance of WGE.

Exumple 2.1. Let U and V be two r.v's distributed as

$$g_U(u) = \begin{cases} 2u, & 0 < u < 1 \\ 0, & otherwise \end{cases}, \qquad g_V(v) = \begin{cases} 2(1-v), & 0 < v < 1 \\ 0 & otherwise \end{cases}.$$

Here, we can see that

$$H_U^{(\eta,\mu)}(g) = H_V^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left(\frac{\mu 2^{2\left(\frac{\eta}{\mu}-1\right)}}{\eta}\right),$$

But, the WGE's of U and V are different with each other as follows

$$H_{(U,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left(\frac{\mu 2^{2\frac{\eta}{\mu}-1}}{4\eta-\mu}\right)$$

and

$$H_{(V,w)}^{(\eta,\mu)} = \frac{1}{2\eta(\mu-\eta)}\log\left(2^{2\frac{\eta}{\mu}-1}\mathrm{B}\left(2\frac{\eta}{\mu}, 2\frac{\eta}{\mu}\right)\right),$$

where,

$$\mathrm{B}(c,d) = \int_0^1 y^{c-1}(1-y)^{d-1},\ c,d>0 = \frac{\Gamma(c)\Gamma(d)}{\Gamma(c+d)}.$$

Thus, even though $H_U^{(\eta,\mu)}(g) = H_V^{(\eta,\mu)}(g)$, but $H_{(U,w)}^{(\eta,\mu)}(g) \neq H_{(V,w)}^{(\eta,\mu)}(g), \forall\ \frac{\mu}{2} < \eta < \mu, \mu \geq 1$.

Example 2.2. Let $g(u)$ be the p.d.f of a r.v U distributed as:

(a) Exponentially with $g(u) = \beta e^{-\beta u}, u>0, \beta>0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left[\frac{\Gamma\left(2\frac{\eta}{\mu}\right)}{\beta\left(2\frac{\eta}{\mu}-1\right)^{2\frac{\eta}{\mu}}}\right].$$

(b) Gamma with $g(u) = \frac{1}{\Gamma(\beta)}e^{-u}u^{\beta-1}, 0<u<\infty, \beta>0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left[\frac{\Gamma\left(\beta\left(2\frac{\eta}{\mu}-1\right)+1\right)}{(\Gamma(\beta))^{2\frac{\eta}{\mu}-1}\left(2\frac{\eta}{\mu}-1\right)^{\beta\left(2\frac{\eta}{\mu}-1\right)+1}}\right].$$

(c) Lomax with $g(u) = \frac{m}{(1+u)^{1+m}}, u>0, m>0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left[\frac{m^{2\frac{\eta}{\mu}-1}\Gamma\left(2\frac{\eta}{\mu}\right)\Gamma\left(m\left(2\frac{\eta}{\mu}-1\right)-1\right)}{\Gamma\left(\left(2\frac{\eta}{\mu}-1\right)(m+1)\right)}\right], m\left(2\frac{\eta}{\mu}-1\right)>1\ .$$

(d) Rayleigh with $g(u) = \beta u e^{-\frac{\beta}{2}u^2}, u \geq o, \beta>0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log\left[\frac{2^{2\frac{\eta}{\mu}-\frac{3}{2}}\Gamma\left(2\frac{\eta}{\mu}-\frac{1}{2}\right)}{\sqrt{\beta}\left(2\frac{\eta}{\mu}-1\right)^{2\frac{\eta}{\mu}+\frac{3}{2}}}\right].$$

Lemma 2.1. If $Z = mU$, with $m>0$, then

$$H_{(Z,w)}^{(\eta,\mu)}(g) = \frac{1}{2\eta(\mu-\eta)}\log m + H_{(U,w)}^{(\eta,\mu)}(g).$$

Theorem 2.1. For a r.v U having SE $H_U(g)$, we obtain

$$H_{(U,w)}^{(\eta,\mu)}(g) \geq \frac{1}{\eta\mu}\left[H_U(g) - \left(\frac{\mu-2\eta}{2(\mu-\eta)}\right)E(\log U)\right].$$

Proof. By applying the log-sum inequality, we obtain

$$\int_0^\infty g(u)\log\frac{g(u)}{(ug(u))^{2\frac{\eta}{\mu}-1}}du \geq \int_0^\infty g(u)du\log\frac{\int_0^\infty g(u)du}{\int_0^\infty (ug(u))^{2\frac{\eta}{\mu}-1}du}$$

$$= -\log\int_0^\infty (ug(u))^{2\frac{\eta}{\mu}-1}du.$$

Due to (7), the desired result is satisfied.

3. Weighted Generalized Residual Entropy (WGRE)

In this section, we introduce the dynamic (residual) version of (7) which is known as weighted generalized residual entropy (WGRE) of order η and type μ. Some important characterization results of this UM are also discussed.

Definition 3.1 Let U be a r.v with p.d.f $g(u)$ and s.f $\bar{G}(t)$, then the WGRE of order η and type μ is defined as

$$H_{(U,w)}^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\log\left[\int_t^\infty \left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du\right],\ \frac{\mu}{2}<\eta<\mu, \mu\geq 1\ . \quad (8)$$

Here, we evaluate the WGRE of some lifetime distributions.

Example 3.1. Let a r.v U be distributed as:

(a) Exponentially with p.d.f $g(u) = \beta e^{-\beta u}, u\beta > 0, \beta > 0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\left[R\beta t + \log\left(\frac{\Gamma(R+1,R\beta t)}{\beta R^{R+1}}\right)\right],$$

(b) Gamma with p.d.f $g(u) = \frac{1}{\Gamma(\beta)}e^{-u}u^{\beta-1},\ 0 < u < \infty, \beta > 0,$

$$H_{(U,w)}^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\log\left[\frac{\Gamma(R\beta+1,Rt)}{R^{R\beta+1}(\Gamma(\beta,t))^{R}}\right],$$

(c) Weibull with p.d.f $g(u) = \frac{1}{m}e^{-\left(\frac{u-n}{m}\right)},\ u > n, m > 0, n > 0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\left[R\,\frac{t}{m} + \log\left(\frac{m\Gamma\left(R+1,R\frac{t}{m}\right)}{R^{R+1}}\right)\right],$$

(d) Rayleigh with p.d.f $g(u) = \beta u e^{-\frac{\beta}{2}u^2},\ u \geq 0, \beta > 0$, then

$$H_{(U,w)}^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\left[\frac{R\beta t^2}{2} + \log\left\{\frac{2^{R-\frac{1}{2}}\Gamma\left(R+\frac{1}{2},\frac{R\beta t^2}{2}\right)}{\sqrt{\beta}R^{R+\frac{1}{2}}}\right\}\right],$$

where, $\Gamma(n,mz) = m^n\int_z^\infty e^{-mx}x^{n-1}dx,\ m,n > 0$ is an upper incomplete gamma function and $R = 2\frac{\eta}{\mu} - 1$ respectively.

Theorem 3.1 If $H_U^{(\eta,\mu)}(g)$ and $H_{(U,w)}^{(\eta,\mu)}(g;t)$ denotes the GRE and WGRE of a r.v U, then for all $t > 0$, we have

$$H_{(U,w)}^{(\eta,\mu)}(g;t) = \frac{1}{2\eta(\mu-\eta)}\log\left[t^{2\frac{\eta}{\mu}-1}exp\left(2\eta(\mu-\eta)H_U^{(\eta,\mu)}(g;t)\right)\right.$$
$$\left. + \left(2\frac{\eta}{\mu}-1\right)\int_{x=t}^\infty x^{2\left(\frac{\eta}{\mu}-1\right)}\left(\frac{\bar{G}(x)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}exp\left(2\eta(\mu-\eta)H_U^{(\eta,\mu)}(g;x)dx\right)\right].$$

Proof. From (8), we have

$$\int_t^\infty\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du = \int_t^\infty\left[\int_0^u\left(2\frac{\eta}{\mu}-1\right)y^{2\left(\frac{\eta}{\mu}-1\right)}dy\right]\left(\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du$$
$$= \left(2\frac{\eta}{\mu}-1\right)\int_t^\infty\left[\int_0^t y^{2\left(\frac{\eta}{\mu}-1\right)}dy + \int_t^u y^{2\left(\frac{\eta}{\mu}-1\right)}dy\right]\left(\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du$$
$$= t^{2\frac{\eta}{\mu}-1}\int_t^\infty\left(\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du + \left(2\frac{\eta}{\mu}-1\right)\int_{y=t}^\infty\left[y^{2\left(\frac{\eta}{\mu}-1\right)}\left(\int_{u=y}^\infty\left(\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du\right)\right]dy. \quad (9)$$

From (6), we have

$$\int_t^\infty\left(\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du = exp\left[2\eta(\mu-\eta)H_U^{(\eta,\mu)}(g;t)\right]. \quad (10)$$

and

$$\int_t^\infty g^{2\frac{\eta}{\mu}-1}du = \bar{G}^{2\frac{\eta}{\mu}-1}(t)exp\left[2\eta(\mu-\eta)H_U^{(\eta,\mu)}(g;t)\right]. \quad (11)$$

Using (9), (10) and (11) in (8), we obtain the required result.

Here, we show that $\bar{G}(t)$ is uniquely determined by $H_{(U,w)}^{(\eta,\mu)}(g;t)$.

Theorem 3.2. Let U be a r.v having p.d.f $g(u)$, s.f $\bar{G}(t)$ and WGRE $H_{(U,w)}^{(\eta,\mu)}(g;t) < \infty, \frac{\mu}{2} < \eta < \mu, \mu \geq 1$ respectively. If $H_{(U,w)}^{(\eta,\mu)}(g;t)$ is increasing in t, then $H_{(U,w)}^{(\eta,\mu)}(g;t)$ uniquely determines $\bar{G}(t)$.

Proof. Rewriting (8) as

$$exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right] = \int_t^\infty\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du. \quad (12)$$

Differentiating (12) w.r.t t, we have

$$\frac{\partial}{\partial t}exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t) = \left(2\frac{\eta}{\mu}-1\right)\lambda_G(t)\int_t^\infty\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du - \left(t\lambda_G(t)\right)^{2\frac{\eta}{\mu}-1}\right], \quad (13)$$

where, $\lambda_G(t) = \frac{g(t)}{\bar{G}(t)}$ represents the hazard rate of U. Using (12), we can rewrite (13) as

$$\left(t\lambda_G(t)\right)^{2\frac{\eta}{\mu}-1} - \left(2\frac{\eta}{\mu}-1\right)exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\lambda_G(t)\right]$$

$$2\eta(\mu-\eta)exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right]\frac{\partial}{\partial t}H_{(U,w)}^{(\eta,\mu)}(g;t)=0. \quad (14)$$

Hence for fixed $t>0$, $\lambda_G(t)$ is a solution of $\psi(u_t)=0$, where

$$\psi(u_t)=t^{2\frac{\eta}{\mu}-1}-\left(2\frac{\eta}{\mu}-1\right)exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right]u_t$$
$$+2\eta(\mu-\eta)exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right]\frac{\partial}{\partial t}H_{(U,w)}^{(\eta,\mu)}(g;t).$$

Differentiating both sides w.r.t u_t, we have

$$\frac{\partial}{\partial u_t}\psi(u_t)=\left(2\frac{\eta}{\mu}-1\right)t^{2\frac{\eta}{\mu}-1}u_t^{2\left(\frac{\eta}{\mu}-1\right)}-\left(2\frac{\eta}{\mu}-1\right)exp\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right].$$

Also,

$$\frac{\partial^2}{\partial u_t^2}\psi(u_t)=\left(2\frac{\eta}{\mu}-2\right)\left(2\frac{\eta}{\mu}-1\right)t^{2\frac{\eta}{\mu}-1}u_t^{2\frac{\eta}{\mu}-3}.$$

Now, $\frac{\partial}{\partial u_t}\psi(u_t)=0$ gives

$$u_t=\left[\frac{exp(2\eta(\mu-\eta))H_{(U,w)}^{(\eta,\mu)}(g;t)}{t^{2\frac{\eta}{\mu}-1}}\right]^{2\left(1-\frac{\eta}{\mu}\right)}=u_0 \text{ (say)}.$$

For $\frac{\mu}{2}<\eta<\mu,\ \mu\geq 1$, $\frac{\partial^2}{\partial u_t^2}\psi(u_0)<0$. Thus, $\psi(u_t)$ attains maximum at u_0. Also, $\psi(0)>0$ and $\psi(\infty)=-\infty$. Further it can be easily observed that $\psi(u_t)$ first increases for $0<u_t<u_o$ and then decreases for $u_t>u_o$. So, the unique solution to $\psi(u_t)=0$ is given by $u_t=\lambda_G(t)$. Thus, $H_{(U,w)}^{(\eta,\mu)}(g;t)$ uniquely determines $\lambda_G(t)$ which in turns determines $\bar{G}(t)$.

4. Properties and Inequalities of WGRE

This section presents some interesting properties and inequalities of weighted generalized residual entropy .

Definition 4.1. Let U and V be two r.v's having WGRE's $H_{(U,w)}^{(\eta,\mu)}(g;t)$ and $H_{(V,w)}^{(\eta,\mu)}(g;t)$, then U is said to be smaller than V in WGRE of order η and type μ (denoted by $U \overset{WGRE}{\leq} V$), if$H_{(U,w)}^{(\eta,\mu)}(g;t)\leq H_{(V,w)}^{(\eta,\mu)}(g;t),\ \forall\, t>0$.

Definition 4.2. A r.v U or a s.f $\bar{G}$ will be said to have increasing (decreasing) WGE for residual life of order η and type μ IWGERL (DWGERL), if $H_{(U,w)}^{(\eta,\mu)}(g;t)$ is increasing (decreasing) in t, $t>0$.

Lemma 4.1. If $Y=aU$, with $a>0$ is a constant, then

$$H_{(Y,w)}^{(\eta,\mu)}(g;t)=\frac{1}{2\eta(\mu-\eta)}\log a+H_{(U,w)}^{(\eta,\mu)}\left(g,\frac{t}{a}\right).$$

Proof.

$$H_{(Y,w)}^{(\eta,\mu)}(g;t)=\frac{1}{2\eta(\mu-\eta)}\log\int_t^\infty\left(y\frac{g(y)}{Pr(Y>t)}\right)^{2\frac{\eta}{\mu}-1}dy.$$

Setting $Y=aU$, a strictly increasing function of U, we have

$$H_{(Y,w)}^{(\eta,\mu)}(g;t)=\frac{1}{2\eta(\mu-\eta)}\log\left[a\int_{\frac{t}{a}}^\infty\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du\right].$$

By using (8), the desired result is obtained.

Theorem 4.1. For two r.v's U and V, let us define $Y_1=a_{1U}$ and $Y_2=a_2V$ with $a_1,a_2>0$. Let $U \overset{WGRE}{\leq} V$ and $a_1\leq a_2$. Then $Y_1 \overset{WGRE}{\leq} Y_2$ $Y_1 \overset{WGRE}{\leq} Y_2$ if $H_{(U,w)}^{(\eta,\mu)}(g;t)$ or $H_{(V,w)}^{(\eta,\mu)}(g;t)$ is decreasing in $t>0$.

Poof. Suppose $H_{(U,w)}^{(\eta,\mu)}(g;t)$ is decreasing in t.

Now, $U \overset{WGRE}{\leq} V$ implies

$$H_{(U,w)}^{(\eta,\mu)}\left(g;\frac{t}{a_2}\right)\leq H_{(V,w)}^{(\eta,\mu)}\left(g;\frac{t}{a_2}\right). \quad (15)$$

Further, since $\frac{t}{a_1}\geq\frac{t}{a_2}$, we have

$$H_{(U,w)}^{(\eta,\mu)}\left(g;\frac{t}{a_1}\right) \leq H_{(U,w)}^{(\eta,\mu)}\left(g;\frac{t}{a_2}\right). \tag{16}$$

Combining (15) and (16), we obtain

$$H_{(U,w)}^{(\eta,\mu)}\left(g;\frac{t}{a_1}\right) \leq H_{(V,w)}^{(\eta,\mu)}\left(g;\frac{t}{a_2}\right). \tag{17}$$

Using Lemma 4.1 in (17), we have $Y_1 \overset{WGRE}{\leq} Y_2$.

Theorem 4.2. For a r.v U having support $(0,k], k>0$, p.d.f $g(u)$ and s.f $\bar{G}(t), t>0$, then for $\frac{\mu}{2}<\eta<\mu,\ \mu\geq 1$, the following upper bound of $H_{(U,w)}^{(\eta,\mu)}(g;t)$ holds

$$H_{(U,w)}^{(\eta,\mu)}(g;t) \leq \frac{1}{2\eta(\mu-\eta)}\left[\frac{\int_t^k\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}\log\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du}{\int_t^k\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du}+\log(k-t)\right].$$

Proof. From log-sum inequality and (8), we have

$$\int_t^k\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}\log\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du \geq \int_t^k\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du\log\frac{\int_t^k(ug(u))^{2\frac{\eta}{\mu}-1}du}{\int_t^k(\bar{G}(t))^{2\frac{\eta}{\mu}-1}du}$$
$$=\int_t^k\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du\left[2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)-\log(k-t)\right].$$

After simplification, we get the desired result.

Theorem 4.3. Let $\bar{G}$ be a IWGRE (DWGRE) and $\mu>\eta$, then

$$\lambda_G(t) \leq (\geq)\left[\frac{\left(2\frac{\eta}{\mu}-1\right)exp\left\{2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}}{t^{2\frac{\eta}{\mu}-1}}\right]^{\frac{\mu}{2(\eta-\mu)}}.$$

Proof. From (14), we have

$$2\eta(\mu-\eta)\frac{\partial}{\partial t}H_{(U,w)}^{(\eta,\mu)}(g;t)=\left(2\frac{\eta}{\mu}-1\right)\lambda_G(t)-exp\left\{2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}\left(t\lambda_G(t)\right)^{2\frac{\eta}{\mu}-1}.$$

Since $\bar{G}$ is IWGERL (DWGERL), therefore, we have

$$\lambda_G(t)\left[\left(2\frac{\eta}{\mu}-1\right)-t^{2\frac{\eta}{\mu}-1}\lambda_G^{2\left(\frac{\eta}{\mu}-1\right)}(t)exp\left\{2\eta(\eta-\mu)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}\right] \geq (\leq)0.$$

which leads to

$$\lambda_G(t) \leq (\geq)\left[\frac{\left(2\frac{\eta}{\mu}-1\right)exp\left\{2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}}{t^{2\frac{\eta}{\mu}-1}}\right]^{\frac{\mu}{2(\eta-\mu)}}.$$

Theorem 4.4. If U is IWGERL (DWGERL), then

$$H_{(U,w)}^{(\eta,\mu)}(g;t) \leq (\geq)\frac{1}{2\eta(\mu-\eta)}\log\left[\frac{t^{2\frac{\eta}{\mu}-1}}{2\frac{\eta}{\mu}-1}\left(\frac{1+\frac{\partial}{\partial t}m_G(t)}{m_G(t)}\right)^{2\left(\frac{\eta}{\mu}-1\right)}\right],$$

where $m_G(t)$ is the mean residual life function of U.

Proof. From (14), we have

$$\frac{\partial}{\partial t}H_{(U,w)}^{(\eta,\mu)}(g;t)=\frac{1}{2\eta(\mu-\eta)}\left[\left(2\frac{\eta}{\mu}-1\right)\lambda_G(t)-\left(t\lambda_G(t)\right)^{2\frac{\eta}{\mu}-1}exp\left\{2\eta(\eta-\mu)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}\right].$$

Using $\lambda_G(t)=\frac{1+\frac{\partial}{\partial t}m_G(t)}{m_G(t)}$, we have

$$\frac{\partial}{\partial t}H_{(U,w)}^{(\eta,\mu)}(g;t)=\frac{1}{2\eta(\mu-\eta)}\left[\left(2\frac{\eta}{\mu}-1\right)\left(\frac{1+\frac{\partial}{\partial t}m_G(t)}{m_G(t)}\right)\right.$$
$$\left.-t^{2\frac{\eta}{\mu}-1}\left(\frac{1+\frac{\partial}{\partial t}m_G(t)}{m_G(t)}\right)^{2\frac{\eta}{\mu}-1}exp\left\{2\eta(\eta-\mu)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}\right].$$

Since, $\bar{G}$ is IWGERL (DWGERL), therefore, after simplification, we have

$$H_{(U,w)}^{(\eta,\mu)}(g;t) \geq (\leq)\log\left[\frac{t^{2\frac{\eta}{\mu}-1}}{2\frac{\eta}{\mu}-1}\left(\frac{1+\frac{\partial}{\partial t}m_G(t)}{m_G(t)}\right)^{2\left(\frac{\eta}{\mu}-1\right)}\right].$$

Theorem 4.5. Let U be the lifetime of a system with p.d.f $g(u)$ and s.f $\bar{G}(t), t > 0$, then $H_{(U,w)}^{(\eta,\mu)}(g;t)$ attains a lower bound as follows

$$H_{(U,w)}^{(\eta,\mu)}(g;t) \geq \frac{1}{2\eta(\mu-\eta)}\left[\left(2\frac{\eta}{\mu}-1\right)\int_t^\infty \frac{g(u)}{\bar{G}(t)}\log u du + 2\left(1-\frac{\eta}{\mu}\right)H_U(g;t)\right]. \quad (18)$$

Proof. From log-sum inequality, we have

$$\int_t^\infty g(u)\log\frac{g(u)}{\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}}du \geq \int_t^\infty g(u)du\log\frac{\int_t^\infty g(u)du}{\int_t^\infty\left(u\frac{g(u)}{\bar{G}(t)}\right)^{2\frac{\eta}{\mu}-1}du}$$

$$= \bar{G}(t)\left[\log\bar{G}(t) - \log\left\{2\eta(\mu-\eta)H_{(U,w)}^{(\eta,\mu)}(g;t)\right\}\right]. \quad (19)$$

where (19) is obtained from (8).

The L.H.S of (19) leads to

$$2\left(1-\frac{\eta}{\mu}\right)\int_t^\infty g(u)\log g(u)du - \left(2\frac{\eta}{\mu}-1\right)\int_t^\infty g(u)\log u du + \left(2\frac{\eta}{\mu}-1\right)\bar{G}(t)\log\bar{G}(t). \quad (20)$$

Using (20) in (19), we obtain (18).

5. Application

To illustrate the effectiveness and importance of our proposed UM's, we consider a real life data set. The data set represents the remission times (in months) of a random sample of 128 bladder cancer patients given in Lee and Wang [15] and is given as follows:

0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

Afaq et al. [16] have shown that the length biased Lomax distribution (LD) provides a better fit for this data. Now, in order to compute the entropy of this data set, it is necessary to apply the weighted entropy technique rather than the simple entropy. For the weighted entropy, we need to consider the basic model (i.e LD) of the length biased LD. The MLEs of the parameters of LD from this data set are obtained as: $\theta = 8.431393$(shape parameter) and $\lambda = 70.289624$(scale parameter) respectively. Now, for $\eta = 1.5, \mu = 2.5$ and $t = 10$, we have $H_{(U,w)}^{(\eta,\mu)}(g) = 1.638$ and $H_{(U,w)}^{(\eta,\mu)}(g;t) = 1.694$. Similarly, $\eta = 2.5, \mu = 3$ and $t = 20$, we obtain $H_{(U,w)}^{(\eta,\mu)}(g) = 1.164$ and $H_{(U,w)}^{(\eta,\mu)}(g;t) = 1.481$ respectively.

6. Conclusion

In this article, we have introduced the concepts of weighted generalized entropy and also its dynamic (residual) version which is known as weighted generalized residual entropy. It has been shown that the proposed residual entropy uniquely determines the survival function. The various important properties and the relationship with other well-known reliability measures of the proposed residual entropy are also obtained. Finally, a real data set has been used to investigate the usefulness of the proposed entropy functions.

References

[1] Shannon, C. E. (1948). A mathematical theory of communications, *Bell System Technical Journal*, 27, 379-423.

[2] Ebrahimi, N. (1996). How to measure uncertainty in the residual lifetime distribution. *Sankhya Series A*, 58, 48-56.

[3] Belis, M. and Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems. *IEEE Transactions on Information Theory*, IT., 4, 593–594.

[4] Di, Crescenzo, A. and Longobardi, M. (2006). On weighted residual and past entropies. *Scientiae Mathematicae Japonicae*, 64, 255–266.

[5] Misagh, F., Panahi, Y., Yari, G. H. and Shahi, R. (2011). Weighted cumulative entropy and its estimation. In Quality and Reliability (ICQR), IEEE International conference, 477-480.

[6] Misagh, F. and Yari, G. H. (2011). On weighted interval entropy. Statistics and Probability Letters, 81, 188–194.

[7] Nourbakhsh, M. and Yari, G. (2016). Weighted Renyi's entropy for lifetime distributions. *Communications in Statistics-Theory and Methods*, doi: 10.1080 /03610926.2016.1148729.

[8] Mirali, M. and Baratpour, S. (2017). Dynamic version of weighted cumulative residual entropy. *Communications in Statistics-Theory and Methods*, *46*(22), 11047-11059.

[9] Kayal, S. (2017). On weighted generalized cumulative residual entropy. *Springer Science+Business Media New York*, 1-17.

[10] Nair, R. S., Sathar, E. I. A. and Rajesh, G. (2017). A study on dynamic weighted failure entropy of order α. *American Journal of Mathematical and Management Sciences*, 36(2), 137-149.

[11] Rajesh, G., Abdul-Sathar, E., and Rohini, S. N. (2017). On dynamic weighted survival entropy of order α. *Communications in Statistics-Theory and Methods, 46*(5), 2139–2150.

[12] Khammar, A., and Jahanshahi, S. (2018). On weighted cumulative residual Tsallis entropy and its dynamic version. *Physica A: Statistical Mechanics and Its Applications,* 491, 678–692.

[13] Bhat, B.A. and Baig, M. A. K. (2019). A New Two Parametric Weighted Generalized Entropy for Lifetime Distributions. *Journal of Modern Applied Statistical Methods*, *18*(2).

[14] Bhat, B. A., Mudasir, S., and Baig. M. A. K. (2019). Some Characterization Results on Length-Biased Generalized Interval Entropy for Lifetime Distributions. *Pakistan Journal of Statistics*, *35*(2), 155-170.

[15] Lee, E. T., and Wang, J. (2003). *Statistical methods for survival data analysis*. *John Wiley & Sons*, Vol. 476.

[16] Ahmad, A., Ahmad, S. P. and Ahmed, A. (2016). Length-biased Weighted Lomax Distribution: Statistical Properties and Application. *Pakistan Journal of Statistics and Operation Research*, 245-255.

BULK ARRIVAL QUEUEING MODEL WITH SETUP AND m OPTIONAL SERVICE UNDER BERNOULLI VACATION SCHEDULE AND SERVER FAILURE

Binay Kumar
•
Magadh Mahila College, Patna University, Patna
bkmathasr@gmail.com

Abstract

In the present investigation, we consider a bulk queue model with the assumption that the server may stop working due to random failure during any stage of the service. As soon as the server fails, it is immediately sent for repair. The server offers all incoming units the first mandatory service and any one of the optional services as per the unit's requirements. For computation purposes, we assume that the server offers m+1 services, of which the first one is essential and the remaining are optional. The server may take a vacation in accordance with the Bernoulli vacation schedule with probability p as soon as both service phases of a unit are completed. As the system empties, the server idles and needs some time to set up before initiating the next service. In order to analyse the model and derive various steady-state queue length distributions, we incorporated the supplementary variables corresponding to service time, vacation time, and repair time and applied the probability generating function technique to determine the various system state distributions. Using these probability distributions, we derive the explicit form of various performance indices. To discuss the validity of the present model, we obtained some well-known results from the queueing literature as a special case of the present model by setting appropriate parameters. Finally, to analyse the sensitivity of several performance indices, a numerical demonstration is provided.

Keywords: queue, bulk, essential service, optional service, supplementary variable, queue length

I. Introduction

Most queueing literature makes the assumption that the server in the service station is always available and that the service station never fails. These presumptions, meanwhile, are notably irrational. In real-world systems, it frequently happens that service stations break down and need to be fixed. We frequently experience situations where the entire system pauses owing to a random failure of a unit in computer communication networks, flexible manufacturing systems, production systems, and other areas.

Due to the potential impact on system performance, these types of systems with a repairable service facility are highly worth investigating from both an operational and queueing theory perspective. For detailed related work on queueing models with unreliable servers, we may refer to the work done by Avi-Itzhak and Naor [2], Li et al. [13], Wang and Yang [22], etc. Chaudhury and Tadj [10] discussed the linear cost procedure to obtain the optimal stationary policy of an unreliable queueing model with a Bernaulli vacation schedule. Rajadurai et al. [17] investigated an unreliable queueing model with a modified vacation schedule and applied the supplementary variable technique to obtain the study state queue size distribution. Yang and Wu [23] discussed the M/M/1 queueing model with the assumption that there is a state-dependent breakdown rate under N policy. They assumed that as the system became empty, the server would take a working vacation. Further, Chakravarthy et al. [5] generalised the model of the working-repair-vacation queue by

assuming the concept of backup servers, which work at a relatively slow rate during the absence of the main server. Recently, Meena et al. [16] applied the supplementary variable technique to analyse the unreliable non-Markovian machine system, which comprised both operating and standby machines under N policy.

In some queueing situations, servers are unavailable for services for occasional intervals of time; such queueing models are termed vacation models. During vacation, the server may perform other types of service or may perform scheduled maintenance. Due to their variety of applications in computer systems, communication networks, and production and inventory systems, queuing systems with vacations have been extensively investigated. A comprehensive and detailed review of the vacation models can be found by Doshi [11], Choudhury [6] and Tian and Zhang [20] and Takagi [19]. Yang et al. [24] investigated a retrial queueing model with a constant retrial rate under the assumption that as orbit becomes empty, the server takes its first essential vacation. Further, the server may take additional option vacations after availing of the first essential vacation. Ayyapan and Karpagam [3] discussed an unreliable non Markovian queue model with a standby server under Bernoulli's schedule vacation policy. It is assumed that when the main server stops working due to random failure, a standby server starts serving the arriving unit. Ahuga et al. [1] applied the Runge-Kutta method to investigate a Markovian queueing system with multiple stages of service and vacation, where it is assumed that the server may breakdown during the busy period and vacation period. Recently, Rani et al. [18] applied recursive approach to find the steady-state queue size distribution of a finite population Markovian queueing model with vacation and discouragement factors. They apply the particle swarm optimisation technique to determine the optimal total cost.

It happens frequently in various queueing circumstances that when units use the first essential service, they subsequently need further services, or more than one service. For a better understanding, we will use the example of a car's service centre. Here, units arrive for routine maintenance, and if a serious problem is found with any element of the vehicle while it is being serviced, they go for repair or replacement of that component. For some comprehensive work in phase service, we may refer to Madan [14], Wang [21], Choudhury and Paul [8], Ke [12], etc. Choudhury and Deka [9] discussed a queueing model based on the assumption that units arrive one by one and the server is unreliable. But in real life situations where units arrive in groups of random size, units may demand more than one type of optional service apart from the essential one. Further, there may be a need for startup time to start the service again. Such situations motivated us to extend the model of Choudhury and Deka [9] by assuming that

- Units arrive in batches of random size.
- Second-phase services may choose among the available optional services.
- Server need start up time to start the service again.
- Server may go on vacation under Bernoulli's vacation schedule.

The remaining paper is organised as follows: In Section II, we describe the brief model description by making some basic assumptions. In Section III, the governing equations of the present model are described. In Section IV, we derive the steady-state queue size distribution function. In Section V, the performance measures of the present model are carried out. In Section VI, some well-known results are established as special cases of the present model. Finally, in Section VII, numerical illustration and sensitivity analysis of performance measures are done.

II. Medel description

In the present model, we consider a non-Markovian queueing model with the assumption that units arrive in batches of random size, according to poisson arrival fashion. There is a single server that provides the first essential services as well as one of the optional services to each arriving unit. As soon as the system becomes empty, the server gets turned off and needs startup time to start again when at least one or more units arrive. The brief description of notations used for the present model is as follows:

λ : Batch arrival rate of the unit.

$S(x)$: Distribution function of set up time.

$B_0(x)$: Distribution function of essential service time.

$B_i(x)$: Distribution function of $i^{th}(i=1,2,...,m)$ optional service time.

$V(x)$: Distribution function of vacation time.

$G_0(x)$: Distribution function of repair when its fails during essential service of a unit .

$G_i(x)$: Distribution function of repair when its fails during $i^{th}(i=1,2,...,m)$ optional service.

$g_0^{(k)}$:The k^{th} moment of repair time when its fails during essential service of a unit

$g_i^{(k)}$:The k^{th} moment of repair time when its fails during $i^{th}(i=1,2,...,m)$ optional service of a unit

r_i : Probability to opt $i^{th}(i=1,2,...,m)$ optional service after essential service.

p : Probability to opt optional vacation after service completion of a unit.

$N_q(t)$ **:** Denote the queue size in system at time t .

$S^0(t)$: Elapsed set up time at time t ..

$B_0^0(t)$: Elapsed service time of essential service at time t .

$B_i^0(t)$: Elapsed service time of $i^{th}(i=1,2,...,m)$ optional service at time t .

$V^0(t)$: Elapsed vacation time at time t ..

$G_0^0(t)$: Elapsed repair time at t time when its fails during essential service of a unit .

$G_i^0(t)$: Elapsed repair time at t time when its fails during $i^{th}(i=1,2,...,m)$ optional service.

Let $\gamma(t)$ denote the state of server at time t, where

$$\gamma(t)=\begin{cases}0 & \text{if the server is idle at time } t,\\ 1 & \text{if the server is start up at time } t,\\ 2 & \text{if the server is busy with essential service at time } t,\\ 2+i & \text{if the server is busy with ith (i=1,2,...,m) optional service at time } t,\\ 3+m & \text{if the server is on vacation at time } t,\\ 4+m & \text{if the server is under repair when it breakdown during essential service at time } t,\\ 4+m+j & \text{if the server is under repair when it breakdown jth (j=1,2,...,m) service.}\end{cases}$$

The variables $S^0(t)$, $B_0^0(t)$, $B_i^0(t)$ $(i=1,2,...,m)$, $V^0(t)$, $G_0^0(t)$ and $G_i^0(t)$ $(i=1,2,...,m)$ are added as supplementary variable in order to obtain a bivariate markav process $\{N_q(t), X(t)\}$ where $X(t)$ assumes values,

$0,\ S^0(t),\ B_0^0(t),\ B_i^0(t)$ *if* $\gamma(t)=0,\ 1,\ 2,\ 2+i\ \ (i=1,2..m)$ respectively and values

$V^0(t),\ \ G_0^0(t),\ G_i^0(t)$ *if* $\gamma(t)=3+m,\ \ 4+m,\ \ 4+m+j\ \ \ (j=1,2..m)$ respectively.

To construct the model, we define the following probabilities

$$L_n(t)=\Pr\{N_q(t)=n, X(t)=0\};\ n\geq 0, \tag{2.1}$$

$$S_n(x,t)=\Pr\{N_q(t)=n,\ X(t)=S^0(t);\ x\leq S^0(t)\leq x+dx\};\ x>0,\ n\geq 1, \tag{2.2}$$

$$P_n^{(0)}(x,t)=\Pr\{N_q(t)=n,\ X(t)=B_0^0(t);\ x\leq B_0^0(t)\leq x+dx\};\ x>0,\ \ n\geq 1, \tag{2.3}$$

$$P_n^{(i)}(x,t)=\Pr\{N_q(t)=n,\ X(t)=B_i^0(t);\ x\leq B_i^0(t)\leq x+dx\};\ x>0,\ \ n\geq 1, 1\leq i\leq m, \tag{2.4}$$

$$V_n(y,t)=\Pr\{N_q(t)=n,\ X(t)=V^0(t);\ y\leq V^0(t)\leq y+dy\};\ y>0,\ \ n\geq 1, \tag{2.5}$$

$$R_n^{(0)}(x,y,t)=\Pr\{N_q(t)=n,\ X(t)=R_0^0(t);\ y\leq R_0^0(t)\leq y+dy/\ B_0^0(t)=x\};\quad x>0,\ \ n\geq 1, \tag{2.6}$$

$$R_n^{(i)}(x,y,t)=\Pr\{N_q(t)=n,\ X(t)=R_i^0(t);\ y\leq R_i^0(t)\leq y+dy/\ B_i^0(t)=x\};\ x>0,\quad n\geq 1, 1\leq i\leq m. \tag{2.7}$$

Further it is assume that

$$V(0)=0, V(\infty)=1, S(0)=0, S(\infty)=1, B_i(0)=0, B_i(\infty)=1,\ \ G_i(0)=0, G_i(\infty)=1.$$

Further it is assume that $G\ (y)$, $V(y)$ functions are continuous at $y=0$,while $B_i(x)$, $S(x)$ are continuous at $x=0.$

The hazard rate functions for present system is given by

$$\eta(x)dx=\frac{dS(x)}{1-S(x)},\ \mu_i(x)dx=\frac{dB_i(x)}{1-B_i(x)},\ \nu(y)dy=\frac{dV(y)}{1-V(y)},\ g_i(y)dy=\frac{dG_i(y)}{1-G_i(y)}\ \ \text{for}\ 0\leq i\leq m.$$

Further we define the following probability generating functions for $i=0,1,2,...,m$ as follows.

$$R^{(i)}(x,y,z)=\sum_{n=1}^{\infty}z^n R_n^{(i)}(x,y)\ ;\qquad R^{(i)}(x,0,z)=\sum_{n=1}^{\infty}z^n R_n^{(i)}(x,0)\quad ;\qquad S(x,z)=\sum_{n=1}^{\infty}z^n S_n(x)$$

$$S(0,z)=\sum_{n=1}^{\infty}z^n S_n(0) \quad ; \quad P^{(i)}(x,z)=\sum_{n=1}^{\infty}z^n P_n^{(i)}(x) \quad ; \quad P^{(i)}(0,z)=\sum_{n=1}^{\infty}z^n P_n^{(i)}(0)$$

$$V(y,z)=\sum_{n=1}^{\infty}z^n V_n(y) \quad ; \quad V(0,z)=\sum_{n=1}^{\infty}z^n V_n(0) \quad ; \quad L(z)=\sum_{n=0}^{\infty}z^n L_n$$

III. Governing Equations

The governing equations of the system are

$$\lambda L_0 = q\left[r_0\int_0^{\infty}\mu_0(x)P_1^{(0)}(x)dx+\int_0^{\infty}\mu_1(x)P_1^{(1)}(x)dx+....+\int_0^{\infty}\mu_m(x)P_1^{(m)}(x)dx\right] + \int_0^{\infty}\nu(y)V_1(y)dy, \tag{3.1}$$

$$\lambda L_1 + q\left[r_0\int_0^{\infty}\mu_0(x)P_1^{(0)}(x)dx+\int_0^{\infty}\mu_1(x)P_1^{(1)}(x)dx+....+\int_0^{\infty}\mu_m(x)P_1^{(m)}(x)dx\right] + \int_0^{\infty}\nu(y)V_1(y)dy = \lambda c_1 L_0, \tag{3.2}$$

$$\lambda L_n = \lambda\sum_{k=1}^{n}c_k L_{n-k}\ ,\ n\geq 2, \tag{3.3}$$

$$\frac{d}{dx}S_n(x)+[\lambda+\eta(x)]S_n(x)=\lambda\sum_{j=1}^{n}c_j S_{n-j}(x);\ \ x>0, n\geq 1, \tag{3.4}$$

$$\frac{d}{dx}P_n^{(i)}(x)+[\lambda+\alpha_i+\mu_i(x)]P_n^{(i)}(x)=\lambda\sum_{j=1}^{n}c_j P_{n-j}^{(i)}(x)+\int_0^{\infty}g_i(y)R_n^{(i)}(x,y)dy; \quad x>0, y>0,\ 0\leq i\leq m, \tag{3.5}$$

$$\frac{d}{dy}V_n(y)+[\lambda+\nu(y)]V_n(x)=\lambda\sum_{j=1}^{n}c_j V_{n-j}(y);\ \ n\geq 1, y>0, \tag{3.6}$$

$$\frac{d}{dy}R_n^{(i)}(x,y)+[\lambda+g_i(y)]R_n^{(i)}(x,y)=\lambda\sum_{j=1}^{n}c_j R_{n-j}^{(i)}(x,y); \quad n\geq 1, x>0, y>0,\ 0\leq i\leq m, \tag{3.7}$$

We will solve the equations (3.1)-(3.7) under the following boundary condition at $x=0$ and $y=0$ given by:

$$S_1(0)=\lambda L_0, \tag{3.8}$$

$$S_n(0)=0;\quad n\geq 2, \tag{3.9}$$

$$P_n^{(0)}(0)=q\left[r_0\int_0^{\infty}\mu_0(x)P_{n+1}^{(0)}(x)dx+\int_0^{\infty}\mu_1(x)P_{n+1}^{(1)}(x)dx+...+\int_0^{\infty}\mu_m(x)P_{n+1}^{(m)}(x)dx\right] + \int_0^{\infty}\eta(x)S_n(x)dx+\int_0^{\infty}\nu(y)V_{n+1}(y)dy\ ; \qquad n\geq 1, \tag{3.10}$$

$$P_n^{(i)}(0)=r_i\int_0^{\infty}\mu_0(x)P_n^{(0)}(x)dx; \quad n\geq 1, \quad 1\leq i\leq m. \tag{3.11}$$

at $y=0$:

$$V_n(0)=p\left[r_0\int_0^{\infty}\mu_0(x)P_n^{(0)}(x)dx+\sum_{i=1}^{m}\int_0^{\infty}\mu_i(x)P_n^{(i)}(x)dx\right]; \quad n\geq 1 \tag{3.12}$$

and at $y=0$ for $i=0,1,2,...,m.$ and fixed value of x.

$$R_n^{(i)}(x;0)=\alpha_i P_n^{(i)}(x);\quad n\geq 1, \quad i=0,1,2,...,m. \tag{3.13}$$

The normalizing condition for present system is given by

$$\sum_{n=0}^{\infty} L_n + \sum_{n=1}^{\infty}\sum_{i=0}^{m}\left[\int_0^{\infty} P_n^{(i)}(x)dx + \int_0^{\infty}\int_0^{\infty} R_n^{(i)}(x,y)dxdy\right] + \sum_{n=1}^{\infty}\int_0^{\infty} S_n(x)dx + \sum_{n=1}^{\infty}\int_0^{\infty} V_n(y)dy = 1 \tag{3.14}$$

IV. Mathematical Analysis

Apply summation formula after multiplying equation (3.2) and (3.3) by appropriate power of z, we get

$$\lambda L(z) + z\left[q\left\{r_0\int_0^{\infty}\mu_0(x)P_1^{(0)}(x)dx + \sum_{i=1}^{m}\int_0^{\infty}\mu_i(x)P_1^{(i)}(x)dx\right\} + \int_0^{\infty}\nu(y)V_1(y)dy\right] \tag{4.1}$$
$$= \left[q\left\{r_0\int_0^{\infty}\mu_0(x)P_1^{(0)}(x)dx + \sum_{i=1}^{m}\int_0^{\infty}\mu_i(x)P_1^{(i)}(x)dx\right\} + \int_0^{\infty}\nu(y)V_1(y)dy\right] + \lambda X(z)L(z)$$

Substitutes the value of (3.1) into equation (4.1) we get

$$L(z) = \frac{L_0(1-z)}{1-X(z)} \tag{4.2}$$

Solving equation (3.4), (3.6) and (3.7) in usual manner we get

$$S(x,z) = S(0,z)[1-S(x)]\exp\{-a_1(z)x\}; \qquad x>0, \tag{4.3}$$

$$V(y,z) = V(0,z)[1-V(y)]\exp\{-a_1(z)y\}; \qquad y>0, \tag{4.4}$$

$$R^{(i)}(x,y,z) = R^{(i)}(x,0,z)[1-G_i(y)]\exp\{-a_1(z)y\}; \qquad y>0,\ 0\le i\le m. \tag{4.5}$$

On multiplying equations (3.8),(3.9) and (3.13) by appropriate power of z, then after little simplification, we get

$$R^{(i)}(x,0,z) = \alpha_i P^{(i)}(x,z)\ ;\ i=0,1,2,...,m, \tag{4.6}$$

$$S(0,z) = z\lambda L_0. \tag{4.7}$$

On simplifying equations (4.3) and (4.7) we have

$$S(x,z) = z\lambda L_0[1-S(x)]\exp\{-a_1(z)x\}; \qquad x>0, \tag{4.8}$$

On simplifying equations (3.5) and (4.5), we have

$$\frac{d}{dx}P^{(i)}(x,z) + (a_1(z)+\alpha_i+\mu_i(x))P^{(i)}(x,z) = R^{(i)}(x,0,z)\overline{G}_i(a_1(z)); \qquad 0\le i\le m. \tag{4.9}$$

Solving equations (4.6) and (4.9), we get

$$P^{(i)}(x,z) = P^{(i)}(0,z)[1-B_i(x)]\exp\{-\phi_i(z)x\}; \quad x>0,\ 0\le i\le m, \tag{4.10}$$

where $\phi_i(z) = a_1(z) + \alpha_i(1-\overline{G}_i(a_1(z)))$ and $a_1(z) = \lambda(1-X(z))$.

From equation (4.10), (4.6) and (4.5) we have

$$R^{(i)}(x,y,z) = \alpha_i P^{(i)}(0,z)[1-B_i(x)]\exp\{-\phi_i(z)x\}[1-G_i(y)]\exp\{-a_1(z)y\}; \tag{4.11}$$
$$(x,y)>0,\ 0\le i\le m.$$

Further, multiplying equations (3.11), (3.12) by suitable power of z and after simplification we have

$$P^{(i)}(0,z) = r_i P^{(0)}(0,z)\overline{B}_0(\phi_0(z)); \qquad 1\le i\le m. \tag{4.12}$$

$$V(0,z) = pP^{(0)}(0,z)\overline{B}_0(\phi_0(z))\left[r_0 + \sum_{i=1}^{m} r_i\overline{B}_i(\phi_i(z))\right]. \tag{4.13}$$

Similarly, from equation (3.10) we have

$$P^{(0)}(0,z) = \frac{q}{z}\left[r_0\int_0^{\infty}\mu_0(x)P^{(0)}(x,z)dx + \sum_{i=1}^{m}\int_0^{\infty}\mu_i(x)P^{(i)}(x,z)dx\right] \tag{4.14}$$
$$+\int_0^{\infty}\eta(x)S(x,z)dx + \frac{1}{z}\int_0^{\infty}\nu(y)V(y,z)dy - \lambda L_0$$

Substituting the value of equation (4.4), (4.8), (4.10) in (4.14) and then using the value of equations (4.7), (4.12) - (4.13) we get

$$P^{(0)}(0,z) = \frac{\lambda L_0 z[1-z\overline{S}(a_1(z))]}{\left[\overline{B}_0(\phi_0(z))\{r_0 + \sum_{i=1}^{m} r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\} - z\right]} \tag{4.15}$$

The limiting value of equation (4.15) when $z\to 1$, is given by

$$P^{(0)}(0,1) = \frac{\lambda L_0[1+\lambda E(X)E(S)]}{(1-\rho)} \tag{4.16}$$

where $\rho = \lambda E(X)\{E(B_0)(1+\alpha_0 g_0^{(1)}) + \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + pE(V)\}$.

Evaluating $z \to 1$ in equation (4.2),(4.4), (4.8)- (4.13) and using the equation (4.16) we have

$$L(1) = \frac{L_0}{E(X)}, \tag{4.17}$$

$$S(x,1) = \lambda L_0[1-S(x)], \tag{4.18}$$

$$P^{(0)}(x,1) = \frac{\lambda L_0[1+\lambda E(X)E(S)][1-B_0(x)]}{(1-\rho)}\ ;\ x>0, \tag{4.19}$$

$$P^{(i)}(x,1) = \frac{r_i \lambda L_0[1+\lambda E(X)E(S)][1-B_i(x)]}{(1-\rho)};\ x>0,\ \ 1\le i\le m, \tag{4.20}$$

$$V(y,1) = \frac{p\lambda L_0[1+\lambda E(X)E(S)][1-V(y)]}{(1-\rho)};\ y>0, \tag{4.21}$$

$$R^{(0)}(x,y,1) = \frac{\alpha_0 \lambda L_0[1+\lambda E(X)E(S)][1-B_0(x)][1-G_0(y)]}{(1-\rho)};\quad (x,y)>0. \tag{4.22}$$

$$R^{(i)}(x,y,1) = \frac{\alpha_i r_i \lambda L_0[1+\lambda E(X)E(S)][1-B_i(x)][1-G_i(y)]}{(1-\rho)};\quad (x,y)>0. \tag{4.23}$$

From equations (4.17)-(4.23) and normalizing condition (3.14), we have

$$L_0 = \frac{E(X)(1-\rho)}{1+\lambda E(X)E(S)}. \tag{4.24}$$

Theorem 1: The joint probability distribution functions of system state and queue size, under stability condition, are given by

$$S(x,z) = \frac{z\lambda E(X)(1-\rho)[1-S(x)]\exp\{-a_1(z)x\}}{1+\lambda E(X)E(S)}, \tag{4.25}$$

$$P^{(0)}(x,z) = \frac{\lambda z E(X)(1-\rho)[1-z\overline{S}(a_1(z))][1-B_0(x)]\exp\{-\phi_0(z)x\}}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]}, \tag{4.26}$$

$$P^{(i)}(x,z) = \frac{r_i \lambda z E(X)(1-\rho)\overline{B}_0(\phi_0(z))[1-z\overline{S}(a_1(z))][1-B_i(x)]\exp\{-\phi_i(z)x\}}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]}, \tag{4.27}$$

$$1\le i\le m,$$

$$V(y,z) = \frac{p\lambda z E(X)(1-\rho)[1-z\overline{S}(a_1(z))]\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}[1-V(y)]\exp\{-a_1(z)y\}}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]}, \tag{4.28}$$

$$R^{(0)}(x,y,z) = \frac{\alpha_0 \lambda z E(X)(1-\rho)[1-z\overline{S}(a_1(z))][1-B_0(x)]\exp\{-\phi_0(z)x\}[1-G_0(y)]\exp\{-a_1(z)y\}}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]}, \tag{4.29}$$

$$R^{(i)}(x,y,z) = \frac{\alpha_i r_i \lambda z E(X)(1-\rho)[1-z\overline{S}(a_1(z))]\overline{B}_0(\phi_0(z))[1-B_i(x)]\exp\{-\phi_i(z)x\}[1-G_i(y)]\exp\{-a_1(z)y\}}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]}, \tag{4.30}$$

$$1\le i\le m,$$

Theorem 2: The marginal probability distribution function of system state queue size are given by

$$S(z) = \frac{z\lambda E(X)(1-\rho)[1-\overline{S}(a_1(z))]}{(1+\lambda E(X)E(S))a_1(z)}, \tag{4.31}$$

$$P^{(0)}(z) = \frac{\lambda z E(X)(1-\rho)[1-z\overline{S}(a_1(z))][1-\overline{B}_0(\phi_0(z))]}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]\phi_0(z)}, \tag{4.32}$$

$$P^{(i)}(z) = \frac{r_i \lambda z E(X)(1-\rho)\overline{B}_0(\phi_0(z))[1-z\overline{S}(a_1(z))][1-\overline{B}_i(\phi_i(z))]}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m} r_i \overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]\phi_i(z)}, \tag{4.33}$$

$$1\le i\le m,$$

$$V(z)=\frac{p\lambda zE(X)(1-\rho)[1-z\overline{S}(a_1(z))]\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}[1-\overline{V}(a_1(z))]}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]a_1(z)}, \tag{4.34}$$

$$R^{(0)}(z)=\frac{\alpha_0\lambda zE(X)(1-\rho)[1-z\overline{S}(a_1(z))][1-\overline{B}_0(\phi_0(z))][1-\overline{G}_0(a_1(z))]}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]a_1(z)\phi_0(z)}, \tag{4.35}$$

$$R^{(i)}(z)=\frac{\alpha_i r_i\lambda zE(X)(1-\rho)[1-z\overline{S}(a_1(z))]\overline{B}_0(\phi_0(z))[1-\overline{B}_i(\phi_i(z))][1-\overline{G}_i(a_1(z))]}{(1+\lambda E(X)E(S))[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]\phi_i(z)a_1(z)}, \tag{4.36}$$

$$1\le i\le m,$$

Proof: See appendix A.

Theorem 3: The stationary queue size distribution at random epoch is given by

$$P(z)=\frac{\lambda E(X)(1-z)(1-\rho)\{1-z\overline{S}(a_1(z))\}[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}]}{a_1(z)[1+\lambda E(X)E(S)][\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]} \tag{4.37}$$

Proof: Adding the equations (4.31)-(4.36) we get required result.

The equation (4.37) can be written as

$$P(z)=\xi(z)\times\omega_{M^X/G/1}^{J\,opt\,with\,Vaca.}(z) \tag{4.38}$$

Where $\xi(z)=\frac{\lambda E(X)\{1-z\overline{S}(a_1(z))\}}{a_1(z)[1+\lambda E(X)E(S)]}$

and $\omega_{M^X/G/1}^{J\,opt\,with\,Vaca.}(z)=\frac{(1-z)(1-\rho)[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}]}{[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]}$

Equation (4.38) shows that the queue size distribution divides into two independent random variables: the first $\omega_{M^X/G/1}^{J\,opt\,with\,Vaca.}(z)$, the stationary queue size distribution of the unreliable bulk queue with optional service including vacation and repair, and the second $\xi(z)$ is the number of arrivals during idle time including setup time.

Theorem 4: The stationary queue size distribution of system at departure epoch is given by

$$\pi(z)=\frac{(1-\rho)\{1-z\overline{S}(a_1(z))\}[\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}]}{[1+\lambda E(X)E(S)][\overline{B}_0(\phi_0(z))\{r_0+\sum_{i=1}^{m}r_i\overline{B}_i(\phi_i(z))\}\{q+p\overline{V}(a_1(z))\}-z]} \tag{4.39}$$

Proof: See appendix B.

Equation (4.39) can be written as

$$\pi(z)=\frac{1-X(z)}{E(X)(1-z)}\times P(z) \tag{4.40}$$

$$\pi(z)=\xi(z)\times\frac{1-X(z)}{E(X)(1-z)}\times\omega_{M^X/G/1}^{J\,opt\,with\,Vaca.}(z) \tag{4.41}$$

Thus, the queue size distribution at the departure epoch decomposes into three independent random variables: $\omega_{M^X/G/1}^{J\,opt\,with\,Vaca.}(z)$,the stationary queue size distribution of the unreliable bulk queue with optional service including vacation and repair; $\xi(z)$ the number of arrivals during idle time including setup time; and the third independent random variable $\frac{1-X(z)}{E(X)(1-z)}$, the number of customers placed before a tagged customer.

V. Performance measures

(a) System state probabilities

By considering limit $z \to 1$ in the marginal probability generating function of the server state queue distribution, it is possible to determine the system state probability of the server state.

- The probability that server is under startup is $P_S = \frac{\lambda E(X)E(S)(1-\rho)}{\{1+\lambda E(X)E(S)\}}$,
- The probability that server is busy with essential service $P_{B_0} = \lambda E(X)E(B_0)$,
- The probability that server is busy in providing the $i^{th}(1 \le i \le m)$ optional service $P_{B_i} = r_i \lambda E(X)E(B_i)$,
- The probability that server is under optional vacation $P_V = p\lambda E(X)E(V)$,
- The probability that server is under repair when its fail during essential service $P_{R_0} = \alpha_0 \lambda E(X)E(B_0)g_0^{(1)}$
- The probability that server is under repair when its fail during $i^{th}(1 \le i \le m)$ optional service $P_{R_i} = r_i \alpha_i \lambda E(X)E(B_i)g_i^{(1)}$,
- Probability that server is idle is given by $P_L = \frac{(1-\rho)}{1+\lambda E(X)E(S)}$.

(b) Average queue length

(i) The mean system size (L_q) at arbitrary epoch can be determined using

$$L_q = \frac{dP(z)}{dz}\Big|_{z=1}$$

$$L_q = \rho + \frac{2\lambda E(X)E(S) + (\lambda E(X))^2 E(S^2) + \lambda E(X^{(2)})E(S)}{2(1+\lambda E(X)E(S))}$$

$$+ \frac{(\lambda E(X))^2 \left\{ E(B_0^2)(1+\alpha_0 g_0^{(1)})^2 + \sum_{i=1}^{m} r_i E(B_i^2)(1+\alpha_i g_i^{(1)})^2 + \alpha_0 g_0^{(2)} E(B_0) + \sum_{i=1}^{m} r_i \alpha_i g_i^{(2)} E(B_i) + pE(V^2) \right.}{\left. + 2E(B_0)(1+\alpha_0 g_0^{(1)}) \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + 2pE(V) \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + 2pE(B_0)(1+\alpha_0 g_0^{(1)})E(V) \right\}}{2(1-\rho)} \quad (5.1)$$

$$+ \frac{\lambda E(X^{(2)}) \left\{ E(B_0)(1+\alpha_0 g_0^{(1)}) + \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + pE(V) \right\}}{2(1-\rho)}$$

(ii) The mean system size (L_D) at departure epoch can be determined using

$$L_D = \frac{d\pi(z)}{dz}\Big|_{z=1}$$

$$L_D = \rho + \frac{2\lambda E(X)E(S) + (\lambda E(X))^2 E(S^2) + \lambda E(X^{(2)})E(S)}{2(1+\lambda E(X)E(S))}$$
$$+ \frac{(\lambda E(X))^2 \left\{ E(B_0^2)(1+\alpha_0 g_0^{(1)})^2 + \sum_{i=1}^{m} r_i E(B_i^2)(1+\alpha_i g_i^{(1)})^2 + \alpha_0 g_0^{(2)} E(B_0) + \sum_{i=1}^{m} r_i \alpha_i g_i^{(2)} E(B_i) + pE(V^2) \right.}{2(1-\rho)}$$
$$\left. + 2E(B_0)(1+\alpha_0 g_0^{(1)}) \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + 2pE(V) \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + 2pE(B_0)(1+\alpha_0 g_0^{(1)})E(V) \right\} \quad (5.2)$$
$$+ \frac{\lambda E(X^{(2)}) \left\{ E(B_0)(1+\alpha_0 g_0^{(1)}) + \sum_{i=1}^{m} r_i E(B_i)(1+\alpha_i g_i^{(1)}) + pE(V) \right\}}{2(1-\rho)} + \frac{E(X^{(2)})}{2E(X)}$$

From (5.1) and (5.2) we can easily observe that $L_D = L_q + \frac{E(X^{(2)})}{2E(X)}$

(c) Average waiting time

The average waiting time can be obtained as

$$E(W_q) = \frac{L_q}{\lambda E(X)} \quad (5.3)$$

VI. Special cases

In this section, we evaluate some special case by setting appropriate parameter to validate our result with existing models.

Case (i): By setting $P(S=0)=1,\ P(X=1)=1,\ r_1=1, m=1$; equation (4.39) gives

$$\pi(z) = \frac{(1-\rho)(1-z)[\overline{B}_0(\phi_0(z))\overline{B}_1(\phi_1(z))\{q + p\overline{V}(\lambda(1-z))\}]}{[\overline{B}_0(\phi_0(z))\overline{B}_1(\phi_1(z))\{q + p\overline{V}(a_1(z))\} - z]}$$

where $\phi_i(z) = \lambda(1-z) + \alpha_i(1-\overline{G}_i(\lambda(1-z))),\ i = 0,1.$

The present model reduces to the model studied by Chaudhury and Deka [9].

Case (ii): By setting $P(S=0)=1,\ P(X=1)=1,\ r_1=1, m=1, \alpha_1=\alpha_2=\ldots=\alpha_m=0$; equation (4.39) gives

$$\pi(z) = \frac{(1-\rho)(1-z)[\overline{B}_0(\lambda(1-X(z)))\overline{B}_1(\lambda(1-X(z)))\{q + p\overline{V}(\lambda(1-X(z)))\}]}{[\overline{B}_0(\lambda(1-X(z)))\overline{B}_1(\lambda(1-X(z)))\{q + p\overline{V}(\lambda(1-X(z)))\} - z]}$$

The present model reduces to the model studied by Chaudhury and madan [7].

Case (iii): By setting $\alpha_1=\alpha_2=\ldots=\alpha_m=0,\ p=0$; equation (4.39) gives

$$\pi(z) = \frac{(1-\rho)\{1-z\overline{S}(a_1(z))\}[\overline{B}_0(\lambda(1-X(z)))\{r_0 + \sum_{i=1}^{m} r_i \overline{B}_i(\lambda(1-X(z)))\}]}{[1+\lambda E(X)E(S)][\overline{B}_0(\lambda(1-X(z)))\{r_0 + \sum_{i=1}^{m} r_i \overline{B}_i(\lambda(1-X(z)))\} - z]}$$

The present model reduces to model investigated by Ke [12].

Case(iv): By setting $\alpha_1=\alpha_2=\ldots=\alpha_m=0,\ p=0, r_1=r_2=\ldots=r_m=0,$; equation (4.39) gives

$$\pi(z) = \frac{(1-\rho)\{1-z\overline{S}(a_1(z))\}[\overline{B}_0(\lambda(1-X(z)))]}{[1+\lambda E(X)E(S)][\overline{B}_0(\lambda(1-X(z))) - z]}$$

The present model reduces to the model studied by Choudhury [6]

Case(v): By setting $\alpha_0=\alpha_1=\ldots=\alpha_m=0,\ p=0, r_2=\ldots=r_m=0, P(X=1)=1$; equation (4.39) gives

$$\pi(z)=\frac{(1-\rho)(1-z)\overline{B}_0(\lambda(1-z))\{r_0+\overline{B}_1(\lambda(1-z))\}}{[\overline{B}_0(\lambda(1-z))\{r_0+\overline{B}_1(\lambda(1-z))\}-z]}$$

The present model reduces to the model studied by Medhi [15].

VII. Numerical illustration

In present section, we will provide the numerical illustration and sensitivity analysis of the various performance measures on different parameters of the model. For this, it assume that the first two moments of the batch size distribution are given by $E(X)=\frac{b}{a},\ E(X^2)=\frac{b(1+b)}{a^2};\ b=1-a.$ It is assumed that the server's start-up time will follow an deterministic distribution with first and second moments $E(S)=\frac{1}{s},E(S^2)=\frac{1}{s^2}.$ The distribution of compulsory and elective service periods is assumed to be exponential, and its first and second moments are therefore derived as $E(B_i)=\frac{1}{\mu_i},\ E(B_i^2)=\frac{2}{\mu_i^2}\ ;\ i=01,2.$ where μ_i denote the service rate. Further, the distribution of vacation time is assumed to be Erlangian-2 and has parameter $\gamma_i(i=1,2).$ The first and second moments of vacation time distribution are $E(V)=\frac{1}{v},\ E(V^2)=\frac{3}{2v^2}$ .The repair time distribution is further assumed to follow an exponential distribution with a parameter g_i and having the first two moments $g_i^{(1)}=\frac{1}{g_i},\ g_i^{(2)}=\frac{2}{g_i^2};\ i=0,1,2;$ Coding in MATLAB is used to create computer programmes. We now present the numerical results in tables (1) -(5).

Table 1: $E(X)=2,\mu_1=\mu_2=2\mu_0,\alpha_0=0.01,\alpha_1=\alpha_2=2\alpha_0,r_0=r_1=r_1=1/3,v=15,$
$s=10,g_0=10,g_1=15,g_2=15.$

Table2: $E(X)=2,\mu_1=\mu_2=2\mu_0,\alpha_1=\alpha_2=2\alpha_0,r_0=r_1=r_1=1/3,v=15,$
$s=10,g_0=10,g_1=15,g_2=15,\lambda=0.7,\mu_0=2$

Table 3: $E(X)=2,\mu_1=\mu_2=2\mu_0,\alpha_1=\alpha_2=2\alpha_0,r_0=r_1=r_1=1/3,v=15,p=0.5,$
$s=10,g_0=10,g_1=15,g_2=15,\alpha_0=0.01.$

Table 4: $E(X)=2,\mu_1=\mu_2=2\mu_0,\alpha_1=\alpha_2=2\alpha_0,r_0=r_1=r_1=1/3,v=15,p=0.5,$
$s=10,g_0=10,g_1=15,g_2=15,\lambda=0.7,\mu_0=2.$

Table 5: $E(X)=2,\mu_1=\mu_2=2\mu_0,\alpha_1=\alpha_2=2\alpha_0,r_0=r_1=r_1=1/3,v=15,p=0.5,$
$s=10,g_0=10,g_1=15,g_2=15,\lambda=0.7,\alpha_0=0.01$

Table 1: *Effect of arrival rate and service rate on* $L_q\ (W_q)$ *for variation in* p

	$\mu=2$				$\mu=2.1$			
	$p=0.3$		$p=0.7$		$p=0.3$		$p=0.7$	
λ	L_q	W_q	L_q	W_q	L_q	W_q	L_q	W_q
0.61	14.174	11.618	18.492	15.157	10.632	8.715	13.275	10.881
0.63	17.922	14.224	24.894	19.758	12.851	10.199	16.634	13.201
0.65	23.607	18.160	36.393	27.994	15.857	12.197	21.611	16.624
0.67	33.253	24.816	63.102	47.091	20.159	15.044	29.755	22.205
0.69	53.215	38.562	194.278	140.781	26.829	19.441	45.498	32.969

Table 2: *Effect of* p *on* L_q (W_q) *for variation in failure rate and* m

		$m=2$		$m=1$		$m=0$	
	p	L_q	W_q	L_q	W_q	L_q	W_q
$\alpha_0 = 0.01$	0.1	27.349	19.535	17.130	12.236	7.094	5.067
	0.3	33.760	24.114	19.614	14.010	7.623	5.445
	0.5	43.930	31.379	22.869	16.335	8.218	5.870
	0.7	62.558	44.684	27.328	19.520	8.894	6.353
	0.9	107.754	76.967	33.814	24.153	9.668	6.906
$\alpha_0 = 0.05$	0.1	28.516	20.368	17.578	12.556	7.172	5.123
	0.3	35.519	25.371	20.188	14.420	7.710	5.507
	0.5	46.895	33.496	23.636	16.883	8.317	5.940
	0.7	68.622	49.016	28.406	20.290	9.006	6.433
	0.9	126.698	90.498	35.442	25.316	9.797	6.998

Table 3: *Effect of arrival rate on system state probabilities*

λ	P_L	P_S	P_{B_0}	P_{B_1}	P_{B_2}	P_V	P_{R_0}	P_{R_1}	P_{R_2}
0.61	0.12934	0.01578	0.61000	0.10167	0.10167	0.04067	0.00061	0.00014	0.00014
0.63	0.10399	0.01310	0.63000	0.10500	0.10500	0.04200	0.00063	0.00014	0.00014
0.65	0.07882	0.01025	0.65000	0.10833	0.10833	0.04333	0.00065	0.00014	0.00014
0.67	0.05382	0.00721	0.67000	0.11167	0.11167	0.04467	0.00067	0.00015	0.00015
0.69	0.02900	0.00400	0.69000	0.11500	0.11500	0.04600	0.00069	0.00015	0.00015

Table 4: *Effect of service rate on system state probabilities*

μ_0	P_L	P_S	P_{B_0}	P_{B_1}	P_{B_2}	P_V	P_{R_0}	P_{R_1}	P_{R_2}
2	0.01666	0.00233	0.70000	0.11667	0.11667	0.04667	0.00070	0.00016	0.00016
2.1	0.05569	0.00780	0.66667	0.11111	0.11111	0.04667	0.00067	0.00015	0.00015
2.2	0.09117	0.01276	0.63636	0.10606	0.10606	0.04667	0.00064	0.00014	0.00014
2.3	0.12356	0.01730	0.60870	0.10145	0.10145	0.04667	0.00061	0.00014	0.00014
2.4	0.15326	0.02146	0.58333	0.09722	0.09722	0.04667	0.00058	0.00013	0.00013

Table 5: *Effect of failure rate on system state probabilities*

α_0	P_L	P_S	P_{B_0}	P_{B_1}	P_{B_2}	P_V	P_{R_0}	P_{R_1}	P_{R_2}
0.01	0.01666	0.00233	0.70000	0.11667	0.11667	0.04667	0.00070	0.00016	0.00016
0.02	0.01577	0.00221	0.70000	0.11667	0.11667	0.04667	0.00140	0.00031	0.00031
0.03	0.01488	0.00208	0.70000	0.11667	0.11667	0.04667	0.00210	0.00047	0.00047
0.04	0.01400	0.00196	0.70000	0.11667	0.11667	0.04667	0.00280	0.00062	0.00062
0.05	0.01311	0.00184	0.70000	0.11667	0.11667	0.04667	0.00350	0.00078	0.00078

The impact of arrival rate and service rate on the average queue length (waiting time) L_q (W_q) is shown in Table 1. The table clearly shows that the L_q (W_q) increases with rising arrivals, however, there is a diminishing trend brought on by a rise in service rate. Additionally, there is an increasing tendency in L_q $(W_q$) with an increase in p for the fixed value of the arrival rate. Table 2 displays the impact of p on the average queue length (waiting time). The table clearly shows that there is an increasing tendency in L_q (W_q) as a consequence of the growth in p. Additionally seen is a decline in L_q (W_q) as a result of a reduction in the availability of optional services. The variation in system state probability caused by variations in arrival (service) rates is shown in Table 3(4). It is evident from the data that with an increase in arrival (service) rate P_{B_0} , P_{B_1} , P_{B_2} and P_V have growing (declining) trends, whereas P_L and P_S have decreasing (increasing)

trends. Table 5 demonstrates that as the failure rate rises, P_L and P_S tend to decline while P_{B_0}, P_{B_1}, P_{B_2} and P_V remain constants. Along with the rise in failure rates, increasing trends can be seen in P_{R_0}, P_{R_1}, and P_{R_2}.

VIII. Conclusion

In the present article, we investigated a queueing model with an unreliable server under the provision of Bernoulli vacation, setup time, and two-phase service, where the first service is essential and the second is optional, and we had to choose among the available options. In the current study, we use the supplementary variable approach to build the model and assess several performance indices expressions. Our model may be useful in more flexible queueing circumstances that occur in many manufacturing and production systems, where some services may be optional based on the customer's desire and where the manufacture of the items must be done in phases, such as assembling, testing, packing, etc. The model studied can be further generalised by incorporating feedback services as well as some more features such as N-Policy, retrial, and extended vacation policies.

References

[1] Ahuja, A, Jain., A. and Jain, M. (2022). Transient analysis and ANFIS computing of unreliable single server queueing model with multiple stage service and functioning vacation. *Mathematics and Computers Silulation,* 192: 464-490.

[2] Avi-Itzhak, B. and Naor, P. (1963). Some queueing problems with the service station subject to breakdowns. *Operations Research*, 11(3): 303–320.

[3] Ayyapan, G. and Karpagam, S. (2019). Analysis of a bulk queue with unreliable server, immediate feedback, N-policy, Bernoulli schedule multiple vacation and stand-by server. *Ain Shams Engineering Journal,* 10(4): 873-880

[4] Banik, A.D.(2013). Stationary distributions and optimal control of queues with batch Markovian arrival process under multiple adaptive vacations. *Computers & Industrial Engineering,* 65(3): 455-465.

[5] Chakravarthy, S.R., Shruti and Kulshrestha, R.(2020). A queueing model with server breakdowns, repairs, vacations, and backup server. *Operations Research Perspectives,* 7: 100131

[6] Choudhury, G. (2000). An M[x]/G/1 queueing system with a setup period and a vacation period. *Queueing systems,* 36:23-38.

[7] Choudhury, G. and Madan, K.C. (2004). A two phases batch arrival queueing system with a vacation time under Bernoulli schedule. *Applied Mathematics and Computation.*149:337-349.

[8] Choudhury,G. and Paul, M.(2005). A two phase queueing system with Bernoulli feedback. *International Journal of Information and Management Sciences.* 16 (1): 35–52.

[9] Choudhury, G. and Deka, M. (2012). A single server queueing system with two phases of service subject to server breakdown and Bernoulli vacation. *Applied Mathematical Modelling*, 36(12): 6050–6060.

[10] Choudhury,G. and Tadj. L. (2011). The optimal control of an M^X/G/1 unreliable server queue with two phases of service and Bernoulli vacation schedule. *Mathematical and Computer Modelling,* 54(1–2): 673-688.

[11] Doshi, B.T. (1986). Queueing systems with vacations: a survey. *Queueing Systems*, 1:29–66.

[12] Ke, J.C.(2008). An MX/G/1 system with startup server and J additional options for service, *Applied Mathematical Modelling,* 32: 443-458.

[13] Li, W., Shi, D. and Chao, X. (1997). Reliability analysis of M/G/1 queueing system with server breakdowns and vacations. *Journal of Applied Probability*, 34(2):546-555.

[14] Madan ,K.C.(2000). An M/G/1 queue with second optional service. *Queueing Systems*, 34: 37–46.

[15] Medhi,J. (2002). A single server Poisson input queue with a second optional channel. *Queueing Systems*, 42(3): 239–242.

[16] Meena, K.,Jain, M., Assad, A., Sethi, R. and Garg, D. (2022). Performance and cost comparative analysis for M/G/1 repairable machining system with N-policy vacation. *Mathematics and Computers in Simulation*, 200: 315-328

[17] Rajadurai, P., Saravanarajan, M.C., and Chandrasekaran (2014). Analysis of an $M^{[X]}/(G_1, G_2)/1$ retrial queueing system with balking, optional re-service under modified vacation policy and service interruption. *Ain Shams Engineering Journal*, 5(3): 935-950.

[18] Rani, S., Jain, M. and Meena, R.K.(2023). Queueing modeling and optimization of a fault-tolerant system with reboot, recovery, and vacationing server operating under admission control policy. *Mathematics and Computers in Simulation*, 209: 408-425.

[19] Takagi, H. Queueing Analysis' – A Foundation of Performance Evaluations, Amsterdam, 1991.

[20] Tian N, Zhang ZG. Vacation queueing models-theory and applications. Springer-Verlag, 2006.

[21] Wang, J. (2004). An M/G/1 queue with second optional service and server breakdowns. *Computer & Mathematics with Application*, 47: 1713–1723.

[22] Wang, K.H. and Yang, D.Y. (2009). Controlling arrivals for a queueing system with an unreliable server: Newton-Quasi method. *Applied Mathematics and Computation*, 213(1):92–101.

[23] Yang, D.Y. and Wu, C.H. (2015). Cost-minimization analysis of a working vacation queue with *N*-policy and server breakdowns. *Computers & Industrial Engineering* , 82:151-158.

[24] Yang, D.Y, Chang, F.M. and Ke, J.C.(2016). On an unreliable retrial queue with general repeated attempts and *J* optional vacations. *Applied Mathematical Modelling*, 40(4): 3275-3288.

Appendix A

Proof of theorem 1:

Integrating equations (4.25)-(4.27) with respect to x and ussing the result

$$\int_0^\infty e^{-sx}(1-M(x))dx = \frac{1-\overline{M}(s)}{s} \tag{A.1}$$

We get equations (4.31)-(4.33).

Similarly integrating equations (4.28) with respect to y and using (A.1) we get equation (4.34). On repeating the same process for equations (4.29) and (4.30) with variable x, y , and using equation (A.1), we get equations (4.35)-(4.36).

Appendix. B

Proof of theorem 2:

To obtain the queue size distribution at the departure epoch, on the line of Choudhury and Deka [9], we have

$$\pi_j = k_0\left\{r_0\int_0^\infty \mu_0(x)P_{j+1}^{(0)}(x)dx + \int_0^\infty \mu_1(x)P_{j+1}^{(1)}(x)dx + ... + \int_0^\infty \mu_m(x)P_{j+1}^{(m)}(x)dx + \int_0^\infty \nu(y)V_{j+1}(y)dy\right\} \tag{B.1}$$

where k_0 is the normalizing constant and $\{\pi_j; j = 0,1,2,...\}$ as the probability that there are j customers in the queue at a departure epoch.

Multiplying equation (B.1) by z^j and using $\pi(z) = \sum_{j=0}^{\infty}\pi_j z^j$ and after simplification,

We get

$$\pi(z) = \frac{k_0\lambda L_0\{1-z\overline{S}(a_1(z))\}[\overline{B}_0(\phi_0(z))\{r_0 + \sum_{i=1}^{m} r_i\overline{B}_i(\phi_i(z))\}\{q + p\overline{V}(a_1(z))\}]}{[\overline{B}_0(\phi_0(z))\{r_0 + \sum_{i=1}^{m} r_i\overline{B}_i(\phi_i(z))\}\{q + p\overline{V}(a_1(z))\} - z]} \tag{B.2}$$

Using the condition $\pi(1) = 1$, we get

$$k_0 = \frac{(1-\rho)}{\lambda L_0\{1+\lambda E(X)E(S)\}} \tag{B.3}$$

Using the value of equation (B.3) into (B.2), we get required result.

RT&A, No 3 (74)
Volume 18, September 2023

A NEW FINITE MIXTURE OF PROBABILITY MODELS WITH APPLICATION

K.M. Sakthivel and Vidhya G

•

Department of Statistics, Bharathiar University, Coimbatore 641046, Tamil Nadu, India
sakthithebest@buc.edu.in,vidhyastatistic96@gmail.com

Abstract

In this research, we present an approach to model lifetime data by a weighted three-parameter probability distribution utilizing the exponential and gamma distributions. We have presented some of the essential characteristics such as the shapes of pdf, cdf, moments, incomplete moments, survival function, hazard function, mean residual life, stochastic ordering, and order statistics of the proposed distribution. Furthermore, we also presented the Bonferroni index and Lorenz curve of the proposed distribution. The maximum likelihood approach is used to estimate the parameters of the distribution. Finally, the proposed probability distribution is compared to goodness of fit with Lindley, Akash, exponential, two-parameter Lindley, cubic transmuted Rayleigh, and Exponential-Gamma distributions for the real-time data set.

Keywords: Lifetime distribution, Hazard function, Mean residual life function, Order statistic, Maximum likelihood estimation.

1. Introduction

A scientific approach to the statistical modeling of a wide variety of random events has been made possible by finite mixture of probability models. Due to its adaptability in representing complicated data, finite mixture models have drawn significant interest recently, both from a theoretical and practical perspective. Karl Pearson [15] conducted one of the earliest significant analyses utilizing mixture models. He modeled a proportional combination of two normal probability density functions with varying means and variances. A variety of probability distributions were subsequently utilized by many authors to fit a combination of probability distributions. Similarly, Lindley [17] also modeled the 'Lindley distribution' which is a combination of an exponential distribution with a scale parameter of θ and a gamma distribution having a shape parameter of 2 and a scale parameter of θ with their corresponding mixing proportions, $\frac{\theta}{\theta+1}$ and $\frac{1}{\theta+1}$ respectively.

A probability density function (pdf) and cumulative distribution function (cdf) for the Lindley distribution were included below.

$$f(x) = \frac{\theta^2(1+x)e^{-\theta x}}{\theta+1}; x > 0, \theta > 0 \tag{1}$$

$$F(x) = 1 - \left[1 + \frac{\theta x}{\theta+1}\right] e^{-\theta x}; x > 0, \theta > 0 \tag{2}$$

Shanker [22] used the finite mixture model to propose the Akash distribution, which is described by its pdf and cdf.

$$f(x) = \frac{\theta^3(1+x^2)e^{-\theta x}}{\theta^2+2}; x > 0, \theta > 0 \tag{3}$$

$$F(x) = 1 - \left[1 + \frac{\theta x(\theta x+2)}{\theta^2+2}\right] e^{-\theta x}; x > 0, \theta > 0 \tag{4}$$

Furthermore, the finite mixing model is

$$f(x) = w_1 g_1(x) + w_2 g_2(x) \tag{5}$$

Where Shanker [22] uses the mixing proportion for Akash distribution with weights as $w_1 = \frac{\theta^2}{\theta^2+2}$ and $w_2 = \frac{2}{\theta^2+2}$. Here, $g_1(x)$ and $g_2(x)$ denotes pdf of exponential (θ) and gamma (3, θ) distribution respectively.

We make changes to the Akash distribution to make it more inclusive and adaptable. Shanker [22] used the term θ to describe the parameters of an exponential and a gamma distribution. In this study, we presented a new probability distribution, which we called the Exp-Gamma distribution. The proposed distribution is more flexible and it performs like the Generalized version of the Akash distribution. We did this by employing the scale parameter λ for the exponential distribution and shape parameter 3, and the scale parameter β for the gamma distribution with the mixture proportion of $\frac{\theta^2}{\theta^2+2}$ and $\frac{2}{\theta^2+2}$ respectively.

This paper is also arranged in the following manner. In section 2, we present the Exp-Gamma distribution. Section 3 contains the usual moments and their related measures for the Exp-Gamma distribution. Section 4 deals with reliability analysis. Log-odds rate is calculated in section 5. Section 6 discusses Entropy. Section 7 deals with stochastic ordering. The order statistics for the Exp-Gamma distribution are given in section 8. The Lorenz and Bonferroni curves are presented in Section 9. The section 10 Zenga index is derived. In section 11, it is discussed how to estimate the Exp-Gamma distribution's parameters using the maximum likelihood method. Finally, section 12's proposed distribution as an application makes use of real-time data.

2. Exponential-Gamma Distribution(Exp-Gamma)

The probability distribution of the Exp-Gamma distribution can be described by its probability density function and cumulative distribution function.

$$f(x;\theta,\lambda,\beta) = \frac{1}{\theta^2+2}\left[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}\right] \tag{6}$$

$$F(x) = \frac{\theta^2(1-e^{-\lambda x}) + 2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)}{\theta^2+2} \tag{7}$$

for, $x \geq 0, \theta \geq 0, \lambda \geq 0, \beta \geq 0$.

The following images Figure 1 and Figure 2 show a few potential pdf and cdf shapes for an Exp-Gamma distribution for various parameter values. The Akash and Gamma distributions are the special cases of the Exp-Gamma distribution when $\lambda = \beta = \theta$ and $\theta = 0$ respectively. According to Figure 1, the Exp-Gamma distribution presents a variety of pdf patterns, including right-skewed and reversed-J shaped, pdf parameters that have fixed values.

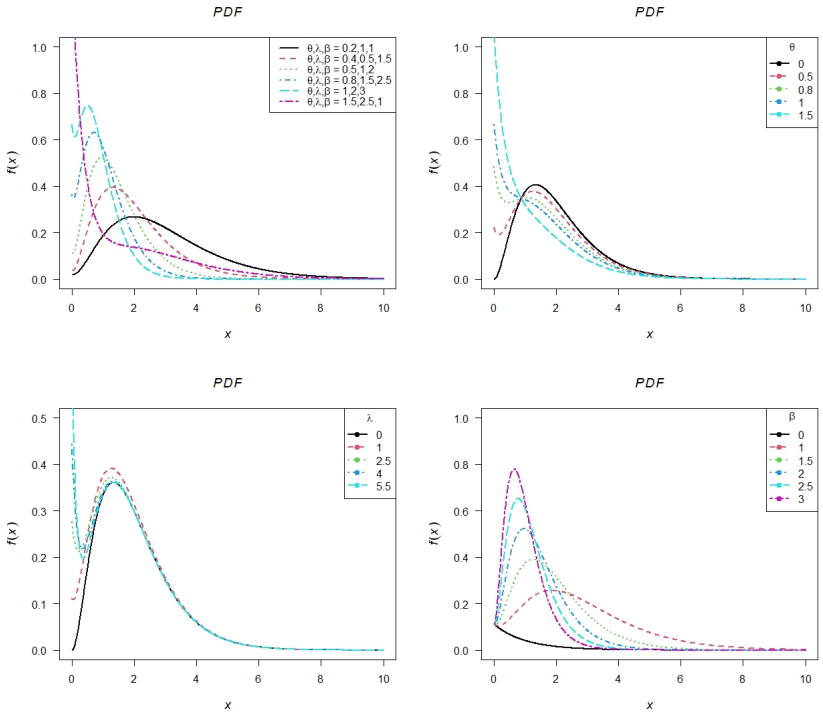


Figure 1: *The shape of the pdf of the Exp-Gamma distribution with varying parameter values.*

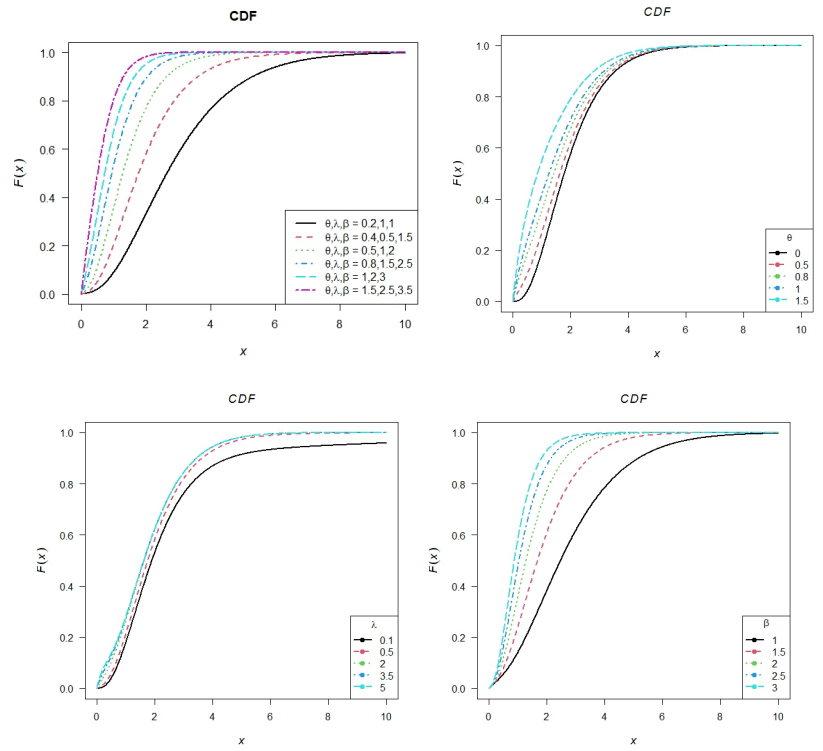


Figure 2: *The form of the Exp-Gamma distribution's cdf changes when the parameter values change.*

3. Moments and related measures

The r^{th} moment (raw moments) has been obtained as

$$\begin{aligned} E(X^r) &= \int_0^\infty x^r f(x)dx \\ &= \int_0^\infty x^r \frac{1}{\theta^2+2}[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}]dx \\ &= \frac{1}{\theta^2+2}\left[\frac{\theta^2\Gamma(r+1)}{\lambda^r} + \frac{\Gamma(r+3)}{\beta^r}\right] \end{aligned} \tag{8}$$

when $r = 1, 2, 3, 4$ then the results follow.

The Exp-Gamma distribution's first four moments are:

$$Mean(\mu) = E(X) = \frac{\beta\theta^2 + 6\lambda}{\beta\lambda(\theta^2+2)}$$

$$E(X^2) = \frac{2\beta^2\theta^2 + 24\lambda^2}{\beta^2\lambda^2(\theta^2+2)}$$

$$E(X^3) = \frac{6\beta^3\theta^2 + 120\lambda^3}{\beta^3\lambda^3(\theta^2+2)}$$

$$E(X^4) = \frac{24\beta^4\theta^2 + 720\lambda^4}{\beta^4\lambda^4(\theta^2+2)}$$

As a result, the Exp-Gamma distribution's central moments are calculated as

$$\mu_2 = Variance = \frac{12\lambda^2 + \theta^2(\beta^2\theta^2 + 24\lambda^2 + 4\beta^2 - 12\lambda\beta)}{\beta^2\lambda^2(\theta^2+2)^2}$$

$$\mu_3 = \frac{2[13\beta^3\theta^2 + 6\beta^3\theta^4 + 60\lambda^3\theta^4 + 24\lambda^3\theta^2 - 36\lambda^2\beta\theta^4 - 36\lambda\beta^2\theta^2 - 54\lambda^2\beta\theta^2 - 84\lambda^3]}{\beta^3\lambda^3(\theta^2+2)^3}$$

$$\begin{aligned} \mu_4 = \Big[&24\beta^4\theta^2(0.375\theta^6 + 3\theta^4 + 8\theta^2 + 8) - 48\lambda^4(93 + 114\theta^2 + 51\theta^4 - 1.5\theta^6) + \\ &72\lambda^2\beta^2\theta^2(12 + 2\theta^4 + \theta^2) - 72\beta^3\theta^2(8\lambda + 4\lambda\theta^2 + \theta^4) \\ &-192\lambda^3\beta\theta^2(5.5 + \theta^2 + 2.5\theta^4)\Big]\frac{1}{\beta^4\lambda^4(\theta^2+2)^4} \end{aligned}$$

With the use of the aforementioned moments, closed-form formulas for the Exp-Gamma distribution's skewness, kurtosis, variation, and index of dispersion are produced. The variance-to-mean ratio is known as the index of dispersion (DI). The model is appropriate for datasets with low dispersion if the DI value is less than 1. The model works well with overly distributed datasets if the DI value is greater than 1.

$$\begin{aligned} skewness(x) &= \frac{E(X^3) - 3E(X^2)\mu + 2\mu^3}{\sigma^3} \\ &= \frac{2[13\beta^3\theta^2 + 6\beta^3\theta^4 + 60\lambda^3\theta^4 + 24\lambda^3\theta^2 - 36\lambda^2\beta\theta^4 - 36\lambda\beta^2\theta^2 - 54\lambda^2\beta\theta^2 - 84\lambda^3]}{(12\lambda^2 + \theta^2(\beta^2\theta^2 + 24\lambda^2 + 4\beta^2 - 12\lambda\beta))^{\frac{3}{2}}} \\ Kurtosis &= \frac{E(X^4) - 4E(X^3)\mu + 6E(X^2)\mu^2 - 3\mu^4}{\sigma^4} \end{aligned}$$

$$Kurtosis = \Bigg[24\beta^4\theta^2(0.375\theta^6 + 3\theta^4 + 8\theta^2 + 8) - 48\lambda^4(93 + 114\theta^2 + 51\theta^4 - 1.5\theta^6) + 72\lambda^2\beta^2\theta^2(12 + 2\theta^4 + \theta^2) - 72\beta^3\theta^2(8\lambda + 4\lambda\theta^2 + \theta^4) - 192\lambda^3\beta\theta^2(5.5 + \theta^2 + 2.5\theta^4) \Bigg] \frac{1}{(12\lambda^2 + \theta^2(\beta^2\theta^2 + 24\lambda^2 + 4\beta^2 - 12\lambda\beta))^2}$$

$$COV = \frac{\sigma}{\mu} = \frac{(12\lambda^2 + \theta^2(\beta^2\theta^2 + 24\lambda^2 + 4\beta^2 - 12\lambda\beta))^{\frac{1}{2}}}{\beta\theta^2 + 6\lambda}$$

$$DOI(\gamma) = \frac{\sigma^2}{\mu} = \frac{12\lambda^2 + \theta^2(\beta^2\theta^2 + 24\lambda^2 + 4\beta^2 - 12\lambda\beta)}{(\beta\lambda(\theta^2 + 2))(\beta\theta^2 + 6\lambda)}$$

As seen in the table 1 to 5, the mean, variance, skewness, kurtosis, and index of dispersion are all expressed in quantitative terms.

From the tables, we can infer that the proposed distributions have the following features:

* The mean of the proposed function is a declining function of θ, λ, and, β.
* The Exp-Gamma distribution is positively skewed for all the parameter values.
* Every positively skewed set of data can fits the suggested distribution.
* When the parameter values of the Exp-Gamma distribution are less than 1, then the Exp-Gamma distribution belongs to the light-tailed distribution, and when it exceeds the value of 1, then it belongs to the heavy-tailed distribution.
* The Exp-Gamma distribution is appropriate for both over- and under-dispersed datasets, as evidenced by the increasing and diminishing DI behavior.

Table 1: *Mean values of the model*

θ	β	λ					
		0.5	1	1.5	2	2.5	3
1	0.5	4.6667	4.3333	4.2222	4.1667	4.1333	4.1111
	1	2.6667	2.3333	2.2222	2.1667	2.1333	2.1111
	1.5	2.0000	1.6667	1.5556	1.5000	1.4667	1.4444
	2	1.6667	1.3333	1.2222	1.1667	1.1333	1.1111
	2.5	1.4667	1.1333	1.0222	0.9667	0.9333	0.9111
	3	1.3333	1.0000	0.8889	0.8333	0.8000	0.7778
2	0.5	3.3333	2.6667	2.4444	2.3333	2.2667	2.2222
	1	2.3333	1.6667	1.4444	1.3333	1.2667	1.2222
	1.5	2.0000	1.3333	1.1111	1.0000	0.9333	0.8889
	2	1.8333	1.1667	0.9444	0.8333	0.7667	0.7222
	2.5	1.7333	1.0667	0.8444	0.7333	0.6667	0.6222
	3	1.6667	1.0000	0.7778	0.6667	0.6000	0.5556
3	0.5	2.7273	1.9091	1.6364	1.5000	1.4182	1.3636
	1	2.1818	1.3636	1.0909	0.9545	0.8727	0.8182
	1.5	2.0000	1.1818	0.9091	0.7727	0.6909	0.6364
	2	1.9091	1.0909	0.8182	0.6818	0.6000	0.5455
	2.5	1.8545	1.0364	0.7636	0.6273	0.5455	0.4909
	3	1.8182	1.0000	0.7272	0.5909	0.5091	0.4545

Table 2: *The variance of the model*

θ	β	λ					
		0.5	1	1.5	2	2.5	3
1	0.5	12.8889	13.8889	14.4691	14.8056	15.0222	15.1728
	1	3.5556	3.2222	3.3580	3.4722	3.5556	3.6173
	1.5	2.2222	1.4444	1.4321	1.4722	1.5111	1.5432
	2	1.8889	0.8889	0.8025	0.8056	0.8222	0.8395
	2.5	1.7956	0.6622	0.5314	0.5122	0.5156	0.5240
	3	1.7778	0.5556	0.3951	0.3611	0.3556	0.3580
2	0.5	10.2222	10.2222	10.6173	10.8889	11.0756	11.2099
	1	3.8889	2.5556	2.5062	2.5556	2.6089	2.6543
	1.5	3.1111	1.3333	1.1358	1.1111	1.1200	1.1358
	2	2.9722	0.9722	0.7006	0.6389	0.6256	0.6265
	2.5	2.9689	0.8356	0.5195	0.4356	0.4089	0.4010
	3	3.0000	0.7778	0.4321	0.3333	0.2978	0.2840
3	0.5	7.8347	6.7190	6.7769	6.8864	6.9779	7.0496
	1	3.9669	1.9587	1.7190	1.6798	1.6820	1.6942
	1.5	3.5152	1.2094	0.8705	0.7817	0.7542	0.7466
	2	3.4463	0.9917	0.6033	0.4897	0.4473	0.4298
	2.5	3.4552	0.9114	0.4932	0.3647	0.3134	0.2899
	3	3.4821	0.8788	0.4408	0.3023	0.2451	0.2176

Table 3: *Skewness of the model*

θ	β	λ					
		0.5	1	1.5	2	2.5	3
1	0.5	0.0252	0.0224	0.0205	0.0194	0.0187	0.0182
	1	0.1912	0.2015	0.1910	0.1794	0.1705	0.1639
	1.5	0.5940	0.6636	0.6800	0.6587	0.6306	0.6054
	2	1.0277	1.5293	1.5954	1.6119	1.5773	1.5279
	2.5	1.2964	2.9363	3.0324	3.1345	3.1482	3.0980
	3	1.4238	4.7520	5.1614	5.3091	5.4316	5.4400
2	0.5	0.0494	0.0592	0.0574	0.0554	0.0539	0.0528
	1	0.2053	0.3950	0.4645	0.4738	0.4680	0.4593
	1.5	0.3739	0.9375	1.3331	1.5255	1.5893	1.5990
	2	0.4567	1.6424	2.4939	3.1600	3.5401	3.7158
	2.5	0.4828	2.3873	3.9009	5.1726	6.1719	6.8048
	3	0.4856	2.9913	5.5430	7.5000	9.2689	10.6651
3	0.5	0.0877	0.1541	0.1641	0.1639	0.1620	0.1601
	1	0.2215	0.7013	1.0708	1.2328	1.2935	1.3129
	1.5	0.3088	1.2455	2.3669	3.2923	3.8550	4.1606
	2	0.3406	1.7724	3.5959	5.6103	7.3395	8.5668
	2.5	0.3483	2.1934	4.8130	7.8061	10.9577	13.7362
	3	0.3475	2.4707	5.9817	9.9640	14.4031	18.9349

Table 4: *Kurtosis of the model*

θ	β	λ					
		0.5	1	1.5	2	2.5	3
1	0.5	0.0291	0.0239	0.0212	0.0198	0.0189	0.0183
	1	0.4635	0.4655	0.4200	0.3824	0.3567	0.3388
	1.5	2.2113	2.3624	2.3568	2.2050	2.0551	1.9358
	2	4.5271	7.4158	7.5377	7.4486	7.0975	6.7211
	2.5	5.9951	18.3288	18.0734	18.4501	18.1850	17.5152
	3	6.6248	35.3808	37.5423	37.7987	38.2654	37.7084
2	0.5	0.0612	0.0709	0.0657	0.0617	0.0590	0.0571
	1	0.4662	0.9786	1.1455	1.1349	1.0926	1.0512
	1.5	1.0564	3.3125	4.9543	5.6862	5.8242	5.7456
	2	1.3369	7.4593	11.9326	15.6581	17.6348	18.3282
	2.5	1.4021	12.5799	22.7055	31.2223	38.2278	42.3995
	3	1.3871	16.9017	37.7625	53.0000	67.5222	79.2691
3	0.5	0.1318	0.2697	0.2848	0.2798	0.2730	0.2670
	1	0.4966	2.1087	3.6540	4.3159	4.5235	4.5574
	1.5	0.7823	4.7057	10.6753	16.4434	20.0144	21.8491
	2	0.8732	7.9452	18.9774	33.7393	48.0182	58.4636
	2.5	0.8850	10.7268	29.1432	52.9549	82.3714	110.9024
	3	0.8726	12.5162	40.2223	75.2908	119.4368	170.8054

Table 5: *Index of dispersion of the model*

θ	β	λ					
		0.5	1	1.5	2	2.5	3
1	0.5	2.7619	3.2051	3.4269	3.5533	3.6344	3.6907
	1	1.3333	1.3809	1.5111	1.6026	1.6667	1.7135
	1.5	1.1111	0.8667	0.9206	0.9815	1.0303	1.0684
	2	1.1333	0.6667	0.6566	0.6905	0.7255	0.7556
	2.5	1.2224	0.5843	0.5198	0.5299	0.5524	0.5751
	3	1.3333	0.5556	0.4444	0.4333	0.4444	0.4603
2	0.5	3.0667	3.8333	4.3434	4.6667	4.8863	5.0444
	1	1.6667	1.5333	1.7350	1.9167	2.0596	2.1717
	1.5	1.5556	1.0000	1.0222	1.1111	1.2000	1.2778
	2	1.6212	0.8333	0.7418	0.7667	0.8159	0.8675
	2.5	1.7128	0.7833	0.6152	0.5939	0.6133	0.6444
	3	1.8000	0.7778	0.5556	0.5000	0.4963	0.5111
3	0.5	2.8727	3.5195	4.1414	4.5909	4.9203	5.1697
	1	1.8182	1.4364	1.5758	1.7597	1.9273	2.0707
	1.5	1.7576	1.0233	0.9576	1.0116	1.0915	1.1732
	2	1.8052	0.9091	0.7374	0.7182	0.7455	0.7879
	2.5	1.8631	0.8794	0.6459	0.5814	0.5745	0.5906
	3	1.9151	0.8788	0.6061	0.5117	0.4814	0.4788

The r^{th} Incomplete moment for Exp-Gamma distribution has been obtained as

$$\begin{aligned}\phi_r(x) &= \int_0^t x^r f(x)dx \\ &= \int_0^r x^r \frac{1}{\theta^2+2}[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}]dx \\ &= \frac{1}{\theta^2+2}\left[\frac{\theta^2\gamma(r+1,\lambda t)}{\lambda^r} + \frac{\gamma(r+3,\beta t)}{\beta^r}\right]\end{aligned} \tag{9}$$

when $r = 1$ the first incomplete moment of the Exp-Gamma distribution is

$$\phi_1(x) = \frac{1}{\theta^2+2}\left[\frac{\beta\theta^2\gamma(2,\lambda t) + \lambda\gamma(4,\beta t)}{\lambda\beta}\right]$$

The related Exp-Gamma distribution moment-generating function is

$$\begin{aligned}M_X(t) = E(e^{tX}) &= \int_0^\infty e^{tX} f(x)dx \\ &= \sum_{i=0}^{\infty}\frac{t^i}{i!}\left(\frac{1}{\theta^2+2}\left[\frac{\theta^2\Gamma(i+1)}{\lambda^i} + \frac{\Gamma(i+3)}{\beta^i}\right]\right)\end{aligned} \tag{10}$$

The corresponding characteristic function of the Exp-Gamma distribution is

$$\begin{aligned}\phi_X(t) = E(e^{itX}) &= \int_0^\infty e^{itX} f(x)dx \\ &= \sum_{i=0}^{\infty}\frac{it^k}{k!}\left(\frac{1}{\theta^2+2}\left[\frac{\theta^2\Gamma(k+1)}{\lambda^k} + \frac{\Gamma(k+3)}{\beta^k}\right]\right)\end{aligned} \tag{11}$$

The Exp-Gamma distribution's associated cumulant-generating function is

$$\begin{aligned}K_X(t) &= \log_e M_X(t) \\ &= \prod_{i=0}^{\infty}\log_e\left(\frac{t^i}{i!}\left(\frac{1}{\theta^2+2}\left[\frac{\theta^2\Gamma(i+1)}{\lambda^i} + \frac{\Gamma(i+3)}{\beta^i}\right]\right)\right)\end{aligned} \tag{12}$$

Probability-weighted moments are derived using a different method for statistical distributions whose inverse form is difficult to define. The corresponding probability-weighted moment for the Exp-Gamma distribution can be found using the formula below.

$$\begin{aligned}\pi_{r,s} &= E(X^r F(x)^s) \\ &= \int_0^\infty x^r f(x)[F(x)]^s dx \\ &= \frac{1}{(\theta^2+2)^{s+1}}\int_0^\infty x^r[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}][\theta^2 - \theta^2 e^{-\lambda x} + 2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]^s dx\end{aligned} \tag{13}$$

The corresponding n^{th} conditional moment of the Exp-Gamma distribution is defined as

$$E[X^n / X > x] = \frac{1}{S(x)}\int_x^\infty x^n f(x)dx$$

$$
\begin{aligned}
E[X^n / X > x] &= \frac{\theta^2 + 2\int_x^\infty x^n \left(\frac{1}{\theta^2+2}[\theta^2 \lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}] \right) dx}{(\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]} \\
&= \frac{-\Gamma(n+1, \lambda x)\theta^2 \beta^n - \Gamma(n+3, \beta x)\lambda^n}{\lambda^n \beta^n \left((\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)] \right)}
\end{aligned} \tag{14}
$$

4. Reliability Analysis

4.1. Survival Function

The odds that an item won't fail before x is specified is the survival function $S(x)$.

$$
\begin{aligned}
S(x) &= P(X > x) = 1 - F(x) \\
&= 1 - \frac{\theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]}{\theta^2 + 2} \\
&= \frac{(\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]}{\theta^2 + 2}
\end{aligned} \tag{15}
$$

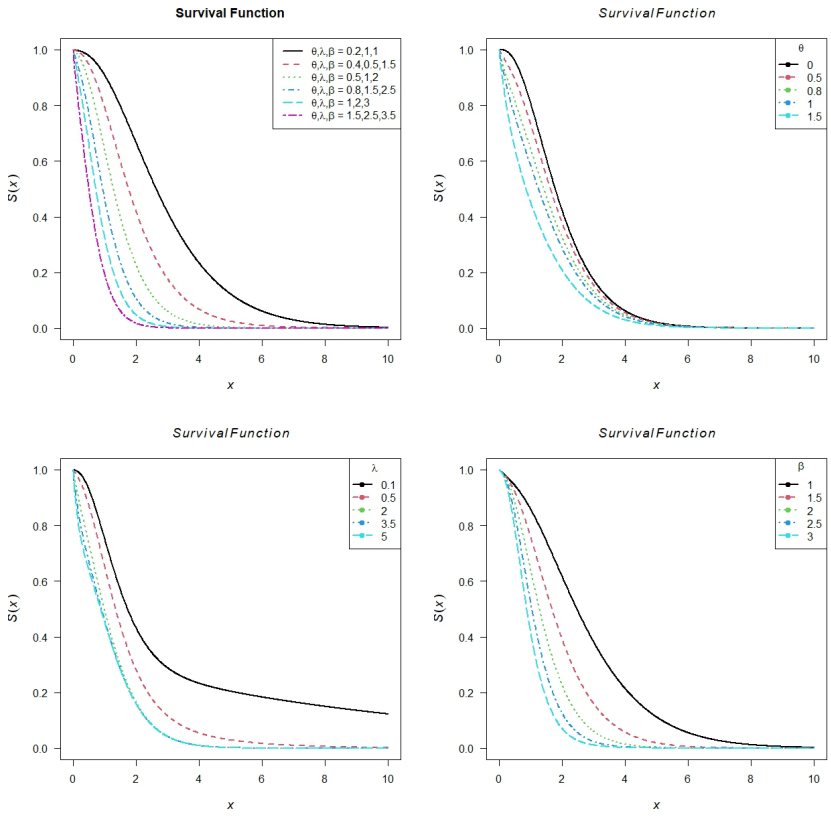


Figure 3: *The different shapes of the sf of an Exp-Gamma distribution for different parameter values.*

4.2. Hazard Rate Function

Assume that X is a continuous random variable with pdf $f(x)$ and cdf $F(x)$. The hazard function of X is

$$\begin{aligned} h(x) &= \frac{f(x)}{1-F(x)} \\ &= \frac{[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}]}{(\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]} \end{aligned} \tag{16}$$

4.3. Mean Residual Life Function

Assume that X is a continuous random variable with pdf $f(x)$ and cdf $F(x)$. According to X, the mean residual life function is

$$\begin{aligned} m(x) = E[X - x/X > x] &= \frac{1}{1-F(x)} \int_x^\infty [1-F(t)]dt \\ &= \frac{\beta\theta^2 e^{-\lambda x} + \lambda(x^2\beta^2 + 4x\beta + 6)e^{-\beta x}}{(\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]} \end{aligned} \tag{17}$$

The Exp-Gamma distribution's hazard function can take three different shapes: decreasing HF, unimodal HF, increasing HF, and decreasing-increasing HF. A declining function is also a property of the mean residual life function.

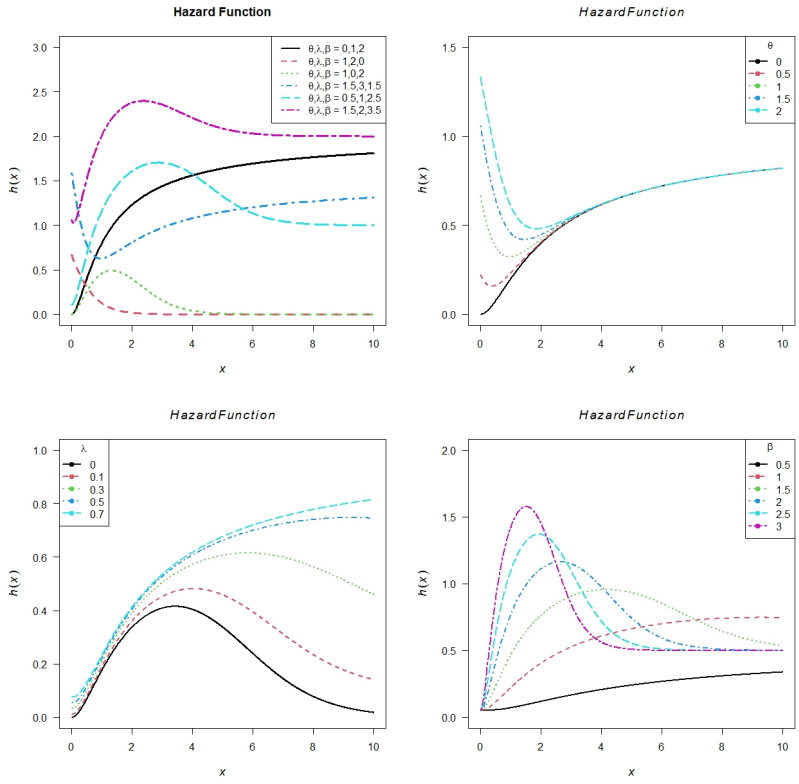


Figure 4: *Hazard function of the Exp-Gamma distribution for different parameter values. The shape of the hazard function changes as the parameter values are varied.*

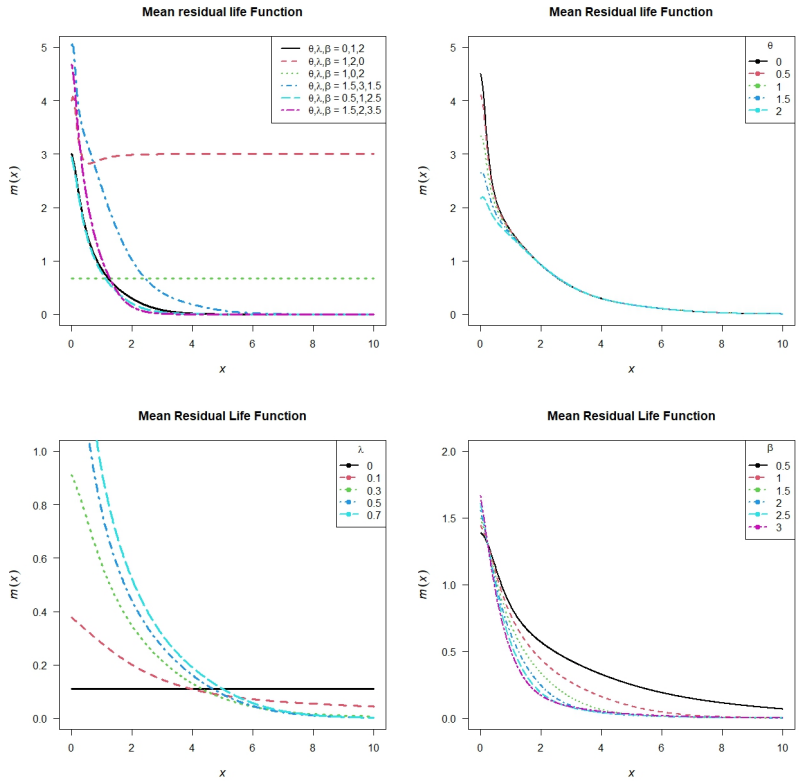


Figure 5: *The various forms of an Exp-Gamma distribution's mean residual life function for various parameter values.*

4.4. Mean Inactivity Time

The mean inactive time is the amount of time that has passed after an item's failure based on the premise that it failed in $(0, t)$.

$$\begin{aligned}\psi_x(t) &= E(X - t/X < t) \\ &= t - \frac{\phi_1(t)}{F(t)} \\ &= t - \frac{\beta\theta^2\gamma(2,\lambda t) + \lambda\gamma(4,\beta t)}{\lambda\beta\left(\theta^2(1-e^{-\lambda t}) + [2 - e^{-\beta t}(t^2\beta^2 + 2t\beta + 2)]\right)}\end{aligned} \tag{18}$$

4.5. Cumulative Hazard

The cumulative hazard function is

$$\begin{aligned}H(x) &= -\log(1 - F(x)) \\ &= \log(\theta^2 + 2) - \log\left((\theta^2 + 2) - \theta^2(1 - e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]\right)\end{aligned} \tag{19}$$

4.6. Reversed Hazard Rate

The Reversed Hazard Rate is

$$\begin{aligned}\tau(x) &= \frac{f(x)}{F(x)} \\ &= \frac{[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}]}{\theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]}\end{aligned} \tag{20}$$

5. Log-odds Rate

The log-odds rate was used by Wang et al. (2003) to propose a model for time to failure as well as some definition of failure time distributions. By simulating the failure process in terms of the log odds rate, the model may be used to analyze the distribution of time until failure.

The odds function is given by

$$\begin{aligned}\pi_o(x) &= \frac{F(x)}{S(x)} \\ &= \frac{\theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]}{(\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]}\end{aligned} \tag{21}$$

The log-odds function is given by

$$\begin{aligned}LO(x) = \log\frac{F(x)}{1-F(x)} = (\log(\theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]) - \log((\theta^2+2) - \\ \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]))\end{aligned} \tag{22}$$

The log-odds rate is defined as

$$\begin{aligned}LOR(x)(x) &= \frac{h(x)}{\overline{F}(x)} \\ &= \frac{[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}](\theta^2+2)}{(\theta^2+2) - \theta^2(1-e^{-\lambda x}) + [2 - e^{-\beta x}(x^2\beta^2 + 2x\beta + 2)]}\end{aligned} \tag{23}$$

6. Entropy

Entropy is a metric for describing the degree of uncertainty in a random variable (X) for the probability density function obtained from the lifetime distribution.

6.1. Renyi Entropy

Renyi entropy of a random variable $X \sim Exp - Gamma(\theta, \lambda, \beta)$ with pdf is defined as

$$\begin{aligned}I_R(\eta) &= \frac{1}{1-\eta}\log\int_0^\infty f^\eta(x)dx; \eta > 0, \eta \neq 1 \\ &= \frac{1}{1-\eta}\log\int_0^\infty \left(\frac{1}{\theta^2+2}[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}]\right)^\eta dx \\ &= \frac{1}{1-\eta}\log\left(\frac{1}{(\theta^2+2)^\eta}\int_0^\infty [\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}]^\eta\right)dx\end{aligned} \tag{24}$$

6.2. Shannon Entropy

The Shannon Entropy of $X \sim Exp - Gamma(\theta, \lambda, \beta)$ is given by

$$\begin{aligned} E[-\log f(X)] &= -\int_X f(x)\log f(x)dx \\ &= E\left[-log\left(\frac{1}{\theta^2+2}[\theta^2\lambda e^{-\lambda x}+\beta^3x^2e^{-\beta x}]\right)\right] \\ &= \log(\theta^2+2) - E\left[\log[\theta^2\lambda e^{-\lambda x}+\beta^3x^2e^{-\beta x}]\right] \\ &= -\frac{1}{\theta^2+2}\int_X\left[\theta^2\lambda e^{-\lambda x}+\beta^3x^2e^{-\beta x}\right]\log\left(\frac{1}{\theta^2+2}[\theta^2\lambda e^{-\lambda x}+\beta^3x^2e^{-\beta x}]\right)dx \end{aligned} \tag{25}$$

6.3. Generalized Entropy

The Generalized Entropy of $X \sim Exp - Gamma(\theta, \lambda, \beta)$ is given by

$$\begin{aligned} GE(w,\delta) &= \frac{1}{\delta(\delta-1)\mu^\delta}\left[\int_0^\infty x^\delta f(x)dx\right] - 1 \\ &= \frac{1}{\delta(\delta-1)\left(\frac{\beta\theta^2+6\lambda}{\beta\lambda(\theta^2+2)}\right)^\delta}\left[\int_0^\infty x^\delta\left(\frac{1}{\theta^2+2}[\theta^2\lambda e^{-\lambda x}+\beta^3x^2e^{-\beta x}]\right)dx\right] - 1 \\ &= \frac{\left(\beta^\delta\theta^2\Gamma(\delta+1)+\lambda^\delta\Gamma(\delta+3)\right)\left(\beta\lambda(\theta^2+2)\right)^\delta}{(\theta^2+2)\lambda^\delta\beta^\delta(\delta(\delta-1)(\beta\theta^2+6\lambda)^\delta)} - 1 \end{aligned} \tag{26}$$

7. Stochastic ordering

Stochastic ordering can be used to assess the relative performance of positive continuous random variables. The size of random variable X is less than that of random variable Y.

- Stochastic order ($X \leq_{st} Y$) if $F_X(x) \geq F_Y(y)$ for all x.
- Hazard rate order ($X \leq_{hr} Y$) if $h_X(x) \geq h_Y(y)$ for all x.
- Mean residual life order ($X \leq_{mrl} Y$) if $m_X(x) \geq m_Y(y)$ for all x.
- Likelihood ratio order ($X \leq_{lr} Y$) if $\frac{f_X(x)}{f_Y(y)}$ decreases in x.

The stochastic ordering of distributions was created by Shaked and Shanthi Kumar (1994) using the results.

The Exp-Gamma distribution is sorted according to the strongest 'likelihood ratio'. Let $X \sim Exp - Gamma(\theta_1, \lambda_1, \beta_1)$ and $Y \sim Exp - Gamma(\theta_2, \lambda_2, \beta_2)$. If, $\beta_1 \geq \beta_2$, then $X \leq_{lr} Y$ and hence $X \leq_{hr} Y, X \leq_{mlr} Y$ and $X \leq_{st} Y$.we have

$$\frac{f_X(x)}{f_Y(x)} = \frac{(\theta_2^2+2)[\theta_1^2\lambda_1 e^{-\lambda_1 x}+\beta_1^3x^2e^{-\beta_1 x}]}{(\theta_1^2+2)[\theta_2^2\lambda_2 e^{-\lambda_2 x}+\beta_2^3x^2e^{-\beta_2 x}]}$$

$$\begin{aligned} \log\frac{f_X(x)}{f_Y(x)} &= \log\left[\frac{\left(\theta_2{}^2+2\right)\left[\theta_1{}^2\lambda_1 e^{-\lambda_1 x}+\beta_1{}^3x^2e^{-\beta_1 x}\right]}{\left(\theta_1{}^2+2\right)\left[\theta_2{}^2\lambda_2 e^{-\lambda_2 x}+\beta_2{}^3x^2e^{-\beta_2 x}\right]}\right] \\ &= \log\left(\theta_2{}^2+2\right)+\log\left[\theta_1{}^2\lambda_1 e^{-\lambda_1 x}+\beta_1{}^3x^2e^{-\beta_1 x}\right]-\log\left(\theta_1^2+2\right)-\log\left[\theta_2{}^2\lambda_2 e^{-\lambda_2 x}+\beta_2{}^3x^2e^{-\beta_2 x}\right] \end{aligned}$$

$$\frac{d}{dx}\log\frac{f_X(x)}{f_Y(x)} = \frac{{\theta_2}^2\lambda_2{}^2e^{-\lambda_2 x} - {\beta_2}^3\left(2xe^{-\beta_2 x} - x^2\beta e^{-\beta_2 x}\right)}{\left[{\theta_2}^2\lambda_2 e^{-\lambda_2 x} + {\beta_2}^3x^2e^{-\beta_2 x}\right]} - \frac{{\theta_1}^2\lambda_1{}^2e^{-\lambda_1 x} + \beta_1^3\left(2xe^{-\beta_1 x} - x^2\beta e^{-\beta_1 x}\right)}{\left[{\theta_1}^2\lambda_1 e^{-\lambda_1 x} + {\beta_1}^3x^2e^{-\beta_1 x}\right]} \tag{27}$$

Now if $\theta_1 = \theta_2 = \theta, \lambda_1 = \lambda_2 = \lambda, \beta_1 \geq \beta_2$, then it implies $\frac{d}{dx}\log\frac{f_X(x)}{f_Y(x)} \leq 0$. This means that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y, X \leq_{mlr} Y$ and $X \leq_{st} Y$.

8. Order Statistics

If $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denotes the order statistic of a random sample $X_1, X_2, \ldots, X_n$ from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$ then the pdf $X_{(r)}$ is given by

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}f_X(x)\left[F_X(x)\right]^{(r-1)}\left[1-F_X(x)\right]^{(n-r)}$$

For, $r = 1, 2, \ldots n$. The pdf of the r^{th} order statistic for the Exp-Gamma distribution is calculated, and the pdf of the largest order statistic $X_{(n)}$ and smallest order statistic $X_{(1)}$ are given below.

n^{th} order statistics

$$\begin{aligned} f_{X_{(n)}}(x) &= nf_X(x)\left[F_X(x)\right]^{(n-1)} \\ &= \frac{n}{\theta^2+2}\left[\theta^2\lambda e^{-\lambda x} + \beta^3x^2e^{-\beta x}\right]\left[\frac{\theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]}{\theta^2+2}\right]^{(n-1)} \end{aligned} \tag{28}$$

1^{st} order statistics

$$\begin{aligned} f_{X_{(1)}}(x) &= nf_X(x)\left[1-F_X(x)\right]^{(n-1)} \\ &= \frac{n}{\theta^2+2}\left[\theta^2\lambda e^{-\lambda x} + \beta^3x^2e^{-\beta x}\right]\left[\frac{\left(\theta^2+2\right) - \theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]}{\theta^2+2}\right]^{(n-1)} \end{aligned} \tag{29}$$

The pdf of a median of order statistic is given as

$$\begin{aligned} f_{m+1:n}(x) &= \frac{(2m+1)}{m!m!}f_X(x)\left[F_X(x)\right]^m\left[1-F_X(x)\right]^m \\ &= \frac{(2m+1)}{m!m!}\left(\frac{1}{\theta^2+2}\left[\theta^2\lambda e^{-\lambda x} + \beta^3x^2e^{-\beta x}\right]\right)\left[\frac{\theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]}{\theta^2+2}\right]^m \\ &\quad \left[\frac{\left(\theta^2+2\right) - \theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]}{\theta^2+2}\right]^m \end{aligned} \tag{30}$$

9. Lorenz and Bonferroni Curves

The Bonferroni and Lorenz curves (Bonferroni, 1930) are used in a variety of sectors, including economics, demography, insurance, and medicine. An Exp-Gamma distribution's Bonferroni and Lorenz curves are calculated as follows:

$$B_o(x) = \frac{1}{\mu F(x)} \int_0^t x f(x) dx = \frac{L_0(x)}{F(x)}$$
$$= \frac{\beta\theta^2\gamma(2,\lambda t) + \lambda\gamma(4,\beta t)}{\lambda\beta\mu\left(\theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]\right)}$$
$$L_o(x) = \frac{1}{\mu} \int_0^t x f(x) dx = \frac{\phi_1(x)}{E(X)}$$
$$= \frac{\left[\beta\theta^2\gamma(2,\lambda t) + \lambda\gamma(4,\beta t)\right]}{\lambda\beta\mu\left(\theta^2+2\right)}$$

10. Zenga index

The Gini index is commonly used to account for the extent of income inequality in a population. The Zenga index (Zenga, 2007) is a relatively new metric and a novel alternative to the Gini index and other current inequality measurements and curves, and the Zenga index is denoted by z.

$$z = 1 - \frac{\mu^-_{(x)}}{\mu^+_{(x)}}$$

where,

$$\mu^-_{(x)} = \frac{1}{F(x)} \int_0^x x f(x) dx = \left[\frac{\beta\theta^2\gamma(2,\lambda x) + \lambda\gamma(4,\beta x)}{\lambda\beta\left(\theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]\right)}\right]$$
$$\mu^+_{(x)} = \frac{1}{1-F(x)} \int_0^\infty x f(x) dx = \frac{\beta\theta^2 + 6\lambda}{\beta\lambda\left(\left(\theta^2+2\right) - \theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]\right)}\Bigg]$$
$$z = 1 - \left[\frac{\beta\theta^2\gamma(2,\lambda x) + \lambda\gamma(4,\beta x)\left(\beta\lambda\left(\left(\theta^2+2\right) - \theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]\right)\right)}{\left(\beta\theta^2 + 6\lambda\right)\lambda\beta\left(\theta^2\left(1-e^{-\lambda x}\right) + \left[2 - e^{-\beta x}\left(x^2\beta^2 + 2x\beta + 2\right)\right]\right)}\right]$$

11. Estimation of Parameters

In this section, the MLE approach is used to estimate the parameters θ, λ, and β. Consider a sample drawn at random from the Exp-Gamma distribution. Then the log-likelihood function is provided by

$$g(x) = \frac{1}{\theta^2+2}\left[\theta^2\lambda e^{-\lambda x} + \beta^3 x^2 e^{-\beta x}\right]$$
$$L\left(x_i,\theta,\lambda,\beta\right) = \prod_{i=1}^n g\left(x_i,\theta,\lambda,\beta\right)$$
$$L\left(x_i,\theta,\lambda,\beta\right) = \prod_{i=1}^n \left(\frac{1}{\theta^2+2}\left[\lambda\theta^2 e^{-\lambda x_i} + \beta^3 x_i^2 e^{-\beta x_i}\right]\right)$$
$$= \left(\frac{n}{\theta^2+2}\prod_{i=1}^n \left[\lambda\theta^2 e^{-\lambda x_i} + \beta^3 x_i^2 e^{-\beta x_i}\right]\right)$$

The respective sample log-likelihood function is

$$\log L(x_i,\theta,\lambda,\beta) = \log n - \log(\theta^2+2) + \sum_{i=1}^n \log[\lambda\theta^2 e^{-\lambda x_i} + \beta^3 x_i^2 e^{-\beta x_i}]$$

Now that we have differentiating w.r.t. θ, λ, and β, we can write

$$\frac{\partial \log L}{\partial \theta} = \frac{-2\theta}{(\theta^2+2)} \sum_{i=1}^{n} \frac{2\theta\lambda e^{-\lambda x_i}}{\left[\theta^2\lambda e^{-\lambda x_i} + \beta^3 x_i^2 e^{-\beta x_i}\right]} = 0$$

$$\frac{\partial \log L}{\partial \lambda} = \sum_{i=1}^{n} \frac{\theta^2\left(e^{-\lambda x_i} - \lambda x_i e^{-\lambda x_i}\right)}{\left[\theta^2\lambda e^{-\lambda x_i} + \beta^3 x_i^2 e^{-\beta x_i}\right]} = 0$$

and

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{x_i^2\left(3\beta^2 e^{-\beta x_i} - \beta^3 x_i e^{-\beta x_i}\right)}{\left[\theta^2\lambda e^{-\lambda x_i} + \beta^3 x_i^2 e^{-\beta x_i}\right]} = 0$$

The MLEs are obtained by solving this system of nonlinear equations. The sample likelihood function can be quantitatively improved by using nonlinear optimization techniques, which are frequently more practical. R programming can be used to solve these equations numerically.

12. Application

Biomedical science lifespan data sets have been fitted with Exp-Gamma distribution. This section compares the goodness of fit of the Exp-Gamma model to the one-parameter Akash [22], Lindley [17], Exponential, two-parameter Lindley [26], Cubic transmuted Rayleigh, and Exponential-Gamma [18] distributions on a real-life data set. A density comparison diagram is also included in this section.

The data, according to Gross and Clark (1975, P.105), represents the lifetime data on the minutes of pain alleviation experienced by 20 people who received an analgesic. The details are as follows:

1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3.0, 1.7, 2.3, 1.6, 2.0

For a real lifetime dataset, the $-2lnL, AIC, AICC, BIC, K-S, CVM, andAD$ statistics have been calculated and shown in Table 7 to compare the goodness of fit of the Exp-Gamma, Akash, Lindley, Exponential, Cubic transmuted Rayleigh, Two parameter Lindley, Exponential-Gamma distributions.

Table 6: *Estimated parameter values of the distributions for the dataset*

Model	**Parameter Estimate**	**Log-Lik**
Exp-Gamma	$\hat{\theta} = 5.3520e^{-05}$, $\hat{\lambda} = 0.2914$	
	$\hat{\beta} = 1.5789$	-22.8873
Akash	$\hat{\theta} = 1.1569$	-29.7613
Lindley	$\hat{\theta} = 0.8161$	-30.2496
Exponential	$\hat{\lambda} = 0.5263$	-32.8371
Cubic transmuted Rayleigh	$\hat{\sigma} = 2.63597$	
	$\hat{\lambda} = 2.5971$	-24.9371
Two parameter Lindley	$\hat{\theta} = 1.48$	-25.8862
	$\hat{\alpha} = -0.2914$	
Exponential-Gamma	$\hat{\lambda} = 0.7361$	-62.2516
	$\hat{\alpha} = 1.7971$	

The variance-covariance matrix of the MLEs is computed as

$$I(\hat{\theta})^{-1} = \begin{pmatrix} 1.1362e^{-01} & -9.5914e^{-06} & 8.3587e^{-07} \\ -9.5914e^{-06} & 8.0969e^{-10} & -8.1641e^{-11} \\ 8.3587e^{-07} & -8.1641e^{-11} & 4.1550e^{-02} \end{pmatrix}$$

The variances of the MLEs of the parameters of Exp-Gamma θ, λ and β are $\text{var}(\hat{\theta}) = 0.1136$, $\text{var}(\hat{\lambda}) = 8.0969e^{-10}$ and $\text{var}(\hat{\beta}) = 0.0415$. And 95% confidence intervals of θ, λ and β are $[-6.60597, 6.60704]$, $[0.29136, 0.29147]$ and $[1.1794, 1.9785]$ respectively.

Table 7: *Criteria for comparison*

Model	-2lnL	AIC	AICC	BIC	AD	K-S statistic	CVM
Exp-Gamma	45.7745	51.7745	53.2747	54.7617	1.9324 (0.097)	0.2587 (0.1007)	0.3508 (0.1375)
Akash	59.5226	61.5226	61.7471	62.5206	3.3554 (0.0185)	0.3705 (0.0082)	0.6555 (0.0154)
Lindley	60.4991	62.4991	62.7213	63.4948	3.7504 (0.0118)	0.3911 (0.0044)	0.7550 (0.0086)
Exponential	65.6742	67.6742	67.8964	68.6699	4.6035 (0.0046)	0.4395 (0.0009)	0.9630 (0.0026)
Cubic transmuted Rayleigh	49.8742	53.8742	54.5801	55.8657	2.216 (0.0707)	0.26534 (0.1196)	0.3873 (0.0772)
Two parameter Lindley	51.7724	55.4375	54.7785	55.8564	3.7822 (0.0085)	0.4102 (0.0075)	0.5275 (0.0058)
Exponential -Gamma	124.503	128.5032	130.4946	129.2091	41.855 (0.0000)	1.000 (0.0000)	5.4779 (0.0000)

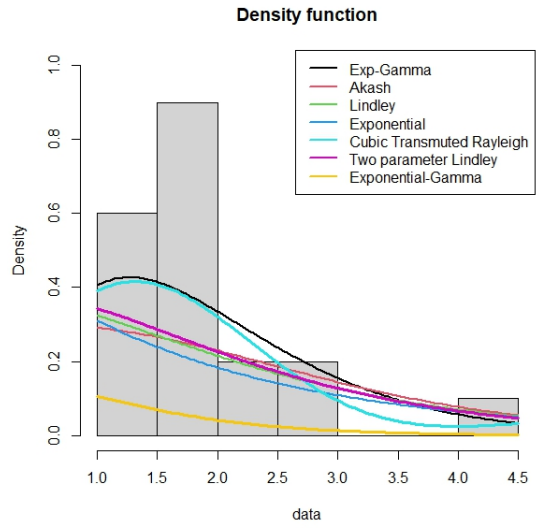


Figure 6: *Comparison of model fit for the distributions.*

The Exp-Gamma distribution fits the dataset better than the Akash, Lindley, exponential, two-parameter Lindley, Cubic transmuted Rayleigh, and Exponential-Gamma distributions as observed from Table 7.

13. Conclusion

A weighted three-parameter probability distribution is developed in this study for modelling skewed lifetime data. We derive expansions of important statistical measures like mean, variance, moments, and moment generating function, etc., as well as maximum likelihood estimation is used to estimate the Exp-Gamma distribution's parameters and hazard and reliability functions are used to examine the distribution's properties. The proposed distribution was fitted using real-time data.

References

[1] Abdulaziz S. Alghamdi, Muhammad Ahsan-ul-Haq, Ayesha Babar, Hassan M. Aljohani, and Ahmed Z. Afify (2022). The discrete power-Ailamujia distribution: properties, inference, and applications. *AIMS Mathematics*, 7(5): 8344-8360.

[2] Ali, M., Khalil, A., Ijaz, M., and Saeed, N. (2021). Alpha-power Exponentiated inverse Rayleigh distribution and its applications to real and simulated data. *PLoS ONE*, 16(1).

[3] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 7(11): 63-79.

[4] Anita Abdollahi Nanvapisheh, MirMostafaee, S., M., T., K., and Emrah Altun (2019). A new two-parameter distribution: properties and applications. *Journal of Mathematical Modeling*, 7(1): 35-48.

[5] Bohning, D. (1999). Computer Assisted of Mixtures and Applications: A meta-analysis, Disease Mapping, and others. *Chapman & Hall, London.*

[6] Chesneau, C., Kumar, V., Khetan, M.,and Arshad, M. (2022). On a modified weighted Exponential distribution with applications. *Math. Comput. Appl.*, 27(17).

[7] Debasis Kundu and Ramashwar D. Gupta (2006). Estimation of P[Y<X] for weibull distribution. *IEEE Transactions on Reliability*, 55(2).

[8] Everitt, B.S., and Hand, D.J. (1981). Finite mixture distributions. *Chapman & Hall, New York*.

[9] Geoffrey McLachlan and David Peel (2000). Finite mixture models *Wiley Series in Probability and Statistics, A Wiley-Interscience Publication.*.

[10] Geoffrey J. McLachlan and Kaye E. Basford (1988). Mixture models: inference and applications to clustering. *Marcel Dekker, INC. New York.*

[11] Geoffrey J. McLachlan, Sharon X. Lee, and Suren I. Rathnayake (2019). Finite Mixture Models. *Annual Review of Statistics and its Applications*, 6: 11.1-11.24.

[12] Ghitany, M. E., Atieh, B., and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics Computing and Simulation*, 78: 493-506.

[13] Giorgi, G. M., and Nadrajah, S. (2010). Bonferroni and Gini indices for various parametric families of distributions. *Metron*, 68: 23-46.

[14] Ijaz, M., Mashwani, W. K., and Belhaouari, S. B. (2020). A novel family of lifetime distribution with applications to real and simulated data. *PLoS ONE*, 15(10).

[15] Karl Pearson (1894). Contribution to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A.*, 185: 71-110

[16] Kerrie L. Mengersen, Christian Robert, and Mike Titterington (2011). Mixtures: Estimation and Applications. *John Wiley & Sons, U.K.*

[17] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B*, 20: 102-107.

[18] Ogunwale, O. D., Adewusi, O. A., and Ayeni, T. M. (2019). Exponential-Gamma distribution. *International Journal of Emerging Technology and Advanced Engineering*, 9(10): 245-249.

[19] Ospina, R.,and Ferrari, S. L. P. (2010). Inflated beta distributions. *Stat Papers*, 51(111).

[20] Paul D. McNicholas (2017). Mixture Model-Based Classification. *CRC Press, Taylor and Francis Group.*

[21] Peter Schlattmann (2009). Medical Applications of Finite Mixture Models. *Springer.*

[22] Shanker, R. (2015a). Akash distribution and its applications. *International Journal of Probability and Statistics*, 4(3): 65-75.
[23] Shanker, R. (2015b). Shanker distribution and its applications. *International Journal of Statistics and Applications*, 5(6): 338-348.
[24] Shanker, R. (2016a). Aradhana distribution and its applications. *International Journal of Statistics and Applications*, 6(1): 23-34.
[25] Shanker, R. (2016b). Sujatha distribution and its applications. *Statistics in Transition. New Series*, 17(3): 391-410.
[26] Shanker, R. and Mishra, A. (2013). A two-parameter Lindley distribution. *Statistics in Transition new series*, 14(1): 45-56.
[27] Shanker, R. and Shukla, K. K. (2017). Ishita distribution and Its Applications. *Biometrics and Biostatistics International Journal*, 5(2): 39-46.
[28] Sylvia Fruhwirth-Schnatter (2006). Finite Mixture and Markov Switching Models. *Springer.*
[29] Titterington, D. M., Adrian F. M. Smith, and Makov, U. E. (1985). Statistical Analysis of Finite Mixture Distributions. *John Wiley & Sons, New York.*
[30] Wahid, A. M. Shehata, Haitham M. Yousof, and Mohamed Aboraya (2021). A novel generator of continuous probability distributions for the asymmetric left-skewed bimodal real-life data with properties and copulas. *Pakistan Journal of Statistics and Operation Research*, 17(4): 943-961.
[31] Zhongjie Shen, Amani Alrumayh, Zubair Ahmad, Reman Abu-Shanab, Maha Al-Mutairi and Ramy Aldallal (2022). A new generalized Rayleigh distribution with analysis to big data of an online community. *Alexandria Engineering Journal*, 61: 11523-11535.

RT&A, No 3 (74)
Volume 18, September 2023

A NEW GENERALIZATION OF TWO PARAMETRIC DIVERGENCE MEASURE AND ITS APPLICATIONS

Fayaz Ahmed

Department of Statistics, University of Kashmir, Srinagar, India
fayazahmed4095@gmail.com

Mirza Abdul Khalique Baig

Department of Statistics, University of Kashmir, Srinagar, India
baigmak@gmail.com

Abstract

In this communication, we proposed two parametric generalized divergence measures. The well-known divergence measures available in the literature are a particular case of our new proposed divergence measure. We also looked into its monotonic behaviour and characterization results. We applied the proposed measure to some life-time distributions and observed that the deviation has been reduced. We have shown the mortality rate of two different countries based on COVID-19 data sets.

Keywords: characterization result, Kullback Leibler divergence measure, Havrda and Charvats divergence, monotonic behavior, probability distribution, Renyi's divergence.

1. Introduction

Information measures play an important role in the field of information theory and other applied sciences.[12] pioneered the concept of information measure (uncertainty).He proposed a way to achieve the uncertainty associated with the probability distribution and established that it is an important part of information theory, which today has many applications in various disciplines. Suppose X is a continuous non-negative random variable, then the [12] entropy is defined as

$$H_S(X) = -\int_0^\infty f(x) log f(x) dx \tag{1}$$

where f is defined as the probability density function of X.
Furthermore, it can be written as

$$H_s(X) = E(-log f(x))$$

The $H_S(X)$ is equal to the expected value of (-logf(x)).
The significance of adequate distance measures between probability distributions stems from their function and has extensive application in entropy.The most prominent divergence used in information theory is relative entropy, also known as [6] divergence (KL divergence).It is widely used in contigency tables, ANOVA tables, statistical inference, etc.

If f(x) and g(x) are the two probability distributions for a continuous random variable X and Y, respectively, then the [6] divergence is given by

$$D_{KL}(f||g) = \int_0^\infty f(x) log \frac{f(x)}{g(x)} dx \tag{2}$$

Furthermore, it can be written as

$$D_{KL}(f||g) = E_{KL}\left[log \frac{f(x)}{g(x)}\right]$$

Remarks
1. If g(x)=1, then it just becomes [12].
2. If g(x)= f(x), then [6] divergence is reduced to zero.
In this direction the generalization of [6] divergence of order α was proposed by [11] and is defined as

$$D_R(f||g) = \frac{1}{\alpha - 1} log \int_0^\infty f(x)^\alpha g(x)^{1-\alpha} dx, \quad \alpha \neq 1, \quad \alpha > 0 \tag{3}$$

Remarks
1. If $\alpha \to 1$, then it simply becomes [6] divergence.
Several researchers have developed various generalizations of [6] divergence in different ways, and in this direction, [5] proposed a new generaliztion of the [6] divergence measure of order ' α' defined as follows

$$D_{HC}(f||g) = \frac{1}{\alpha - 1}\left[\int_0^\infty f(x)^\alpha g(x)^{1-\alpha} dx - 1\right], \quad \alpha \neq 1, \quad \alpha > 0 \tag{4}$$

Remarks
1. If $\alpha \to 1$, then it just becomes [6]
Our aim is to develop the new two-parametric divergence to reduce the deviations. Applied these proposed measure to life-time distributions having different density functions. Also, we obtained some characterization results.Our proposed new two-parametric measure is defined as

$$D_{\alpha,\beta}(f||g) = \frac{1}{\beta(\alpha - \beta)}\left[\int_0^\infty f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} dx - 1\right], \alpha \neq \beta, \quad \beta < \alpha + 1, \alpha, \beta > 0 \tag{5}$$

Remarks
1. When we take f(x) = g(x) then divergence became zero
2. if β = 1 (5) reduced to [5] of order α
3. if β = 1, $\alpha \to 1$ it converge to simply [6] divergence
4. if g(x) = 1, (5) reduces to simply [12]

1.1. Comparasion between known measure and new proposed measure

Example 1.2. Assume X and Y are two non-negative random variables with probability density functions as $f(x) = 2x; \quad 0 < x < 1$ and $\quad g(x) = 2(1-x); \quad 0 < x < 1$. The comparison was derived from (1.2.). The table (1) shows the comparison between the known measure and the proposed measure. From, table (1), we conclude that the divergence is reduced in the proposed divergence measure as compared to the known divergence measure. It means that when we introduce a parameter into a known measure, the distance reduces. We say that it is an alternate measure of the known divergence measure. Figure 1 demonstrates this.

Table 1: *Comparasion between known measure and new proposed measure.*

α	β	$D_{HC}(f\|\|g)$	$D_{\alpha,\beta}(f\|\|g)$
0.1	1	0.094	0.076
0.2	1	0.18	0.11
0.3	1	0.26	0.12
0.4	1	0.34	0.12
0.5	1	0.42	0.10
0.6	1	0.51	0.08
0.7	1	0.61	0.05
0.8	1	0.72	0.02
0.9	1	0.85	0.008

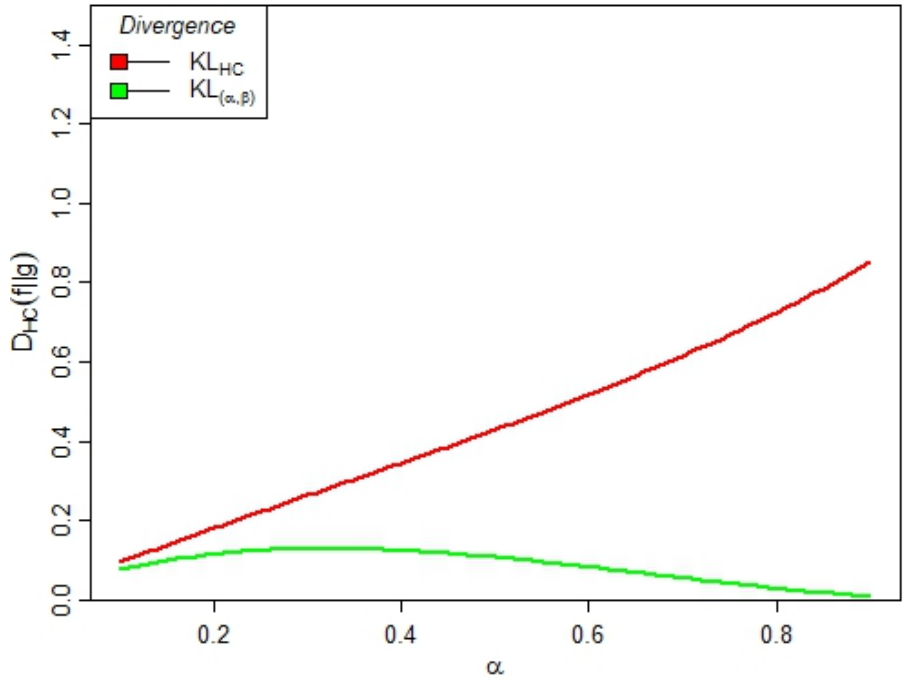


Figure 1: *Divergence between known measure and new proposed measure*

Theorem 1. Assume X and Y are two non-negative random variables with probability density functions f(x) and g(x), then

$$D_{\alpha,\beta}(f||g) \geq 0 \tag{6}$$

with equality if and only if f(x) = g(x).

Proof. By using Gibbs'inequality

$$\int_0^\infty f(x) log \frac{f(x)}{g(x)} dx \geq 0 \tag{7}$$

$$\int_0^\infty \left(f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right) log \frac{\left(f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)}{g(x)} dx \geq 0 \tag{8}$$

$$-\int_0^\infty \left(f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right) log \frac{g(x)}{\left(f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)} dx \geq 0 \tag{9}$$

then (-log) is a convex function and $\int_0^\infty g(x) = 1$ we get

$$\int_0^\infty \left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right) log\frac{g(x)}{\left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx \geq -log\int_0^\infty g(x) \tag{10}$$

and

$$\int_0^\infty \left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right) log\frac{g(x)}{\left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx \geq -log(1) \tag{11}$$

After simplification, we obtained the result. ■

Definition 1.1 Let X be an integer of finite measure. If f(x) and g(x) are density functions and x is integrable, then log-sum inequality defined as

$$\int f(x)log\frac{f(x)}{g(x)}dx \geq [f(x)dx]\, log\left[\frac{f(x)dx}{g(x)dx}\right]$$

Theorem 2. Assume X and Y are two non-negative random variables with the probability density functions f(x) and g(x) respectively and $\beta < \alpha + 1$, $\alpha,\beta > 0$ then

$$D_{\alpha,\beta}(f||g) \geq \frac{1}{log\beta}D_{KL}(f||g) \tag{12}$$

Proof. By using log-sum inequality we have

$$\int_0^\infty f(x)log\frac{f(x)}{\left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx \geq \int_0^\infty f(x)log\frac{f(x)}{\left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx \tag{13}$$

$$\int_0^\infty f(x)log\frac{f(x)}{\left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx \geq -log\int_0^\infty \left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}dx-1\right) \tag{14}$$

$$\int_0^\infty f(x)log\frac{f(x)}{\left(f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx = \int_0^\infty f(x)logf(x)dx - (\alpha-\beta+1)\int_0^\infty f(x)logf(x)dx - (\beta-\alpha)\int_0^\infty f(x)logg(x)dx + \int_0^\infty f(x)log(1) \tag{15}$$

After simplification, then

$$-log\beta(\alpha-\beta)D_{\alpha,\beta}(f||g) \geq [(\alpha-\beta)H(X) + (\beta-\alpha)I(X,Y)] \tag{16}$$

where, I(X,Y) is a [7]. Hence,we get the desired result. ■

We proposed the weighted generalized divergence measure (WGDM) in Section 2, We studied the montonic properties in Section 3, and in Section 4, we identified divergence for the different life-time distributions.We evaluate this article's conclusion in the final part.

2. Weighted generalized divergence measure(WGDM)

In this section,we propose the weighted generalized divergence measure.In real-life situations,[12] and [6] divergence give equal importance to the random variable, but in practical situations, this may cause problems.To overcome this problem, first [2] introduced a measure known as weighted entropy.The weighted entropy is defined as

$$H_S^w(X) = -\int_0^\infty xf(x)logf(x)dx \tag{17}$$

Remarks 1. If x= 1 then, it becomes simply [12].
The weight function is represented by a factor x that gives more weight to the larger value of the random variable.This measure is known as shift-dependent. Many researchers have proposed various weighted measures [13], [8] and [9] In the recent past, based on the concept of weighted entropy, [14] gave weight to the [6] divergence, defined as

$$D_{KL}^w(f||g) = \int_0^\infty xf(x)log\frac{f(x)}{g(x)}dx \tag{18}$$

Remarks
1. If x=1 then, it becomes simply [6].
Furthermore, it can be written as

$$D_{KL}^w(f||g) = E_{KL}\left[Xlog\frac{f(x)}{g(x)}\right]$$

Definition Similar to (2) and based on (5), the weighted proposed measure is defined as

$$D_{KL}^w(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\int_0^\infty x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}dx - 1\right], \alpha \neq \beta, \quad \beta < \alpha+1, \alpha, \beta > 0 \tag{19}$$

Remarks
1. If x^α = 1 then, it became reduced to (5).
To show the importance of random variables in the new two-parametric generalized divergence measure, we consider the following example:

Example 2.1. Suppose X and Y are two non-negative continuous random variables with the density function as follows.
1. $f_1(x) = 1, \quad 0 < x < 1$ and $\quad g_1(x) = nx^{n-1} \quad 0 < x < 1$
2. $f_2(x) = 1, \quad 0 < x < 1$ and $\quad g_2(x) = n(1-x)^{n-1} \quad 0 < x < 1$
Then,the weighted generalized divergence measure characterized the distribution function uniquely.
Using (5) after simplification, we get

$$D_{1(\alpha,\beta)}(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\frac{n^{\beta-\alpha}}{(\beta-\alpha)(n-1)+1} - 1\right] = D_{2(\alpha,\beta)}(f||g) \tag{20}$$

Again using (8) after simplification, we get

$$D_{1(\alpha,\beta)}^w(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\frac{n^{\beta-\alpha}}{\alpha+(\beta-\alpha)(n-1)+1} - 1\right] \tag{21}$$

$$D^{w}_{2(\alpha,\beta)}(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\frac{n^{\beta-\alpha}\Gamma(\alpha+1)\Gamma(t-s)}{\Gamma(\alpha+t-s+1)} - 1\right] \tag{22}$$

Where, t = n(β-α)+1, s = (β-α), and B(u,v) = $\int_0^1 x^{u-1}(1-x)^{v-1}dx =$

$\int_0^1 \frac{x^{u-1}}{(1+x)^{u+v}} = \frac{\Gamma u \Gamma v}{\Gamma(u+v)}$

which is known as the "complete beta function."
We can see from the preceding example that, without weight, our proposed measure has the same value, but given weight, the value is different, so we conclude that the weighted measure uniquely determines the distribution.
From the table (2.1), we conclude that the proposed generalized divergence measure is equal but the weighted generalized divergence measure is different. It can be seen that when different values of alpha, beta, and n are used,
the $D_1(\alpha,\beta)(f||g) = D_2(\alpha,\beta)(f||g)$, but when the proposed divergence measure is weighted, the $Dw_1(\alpha,\beta)(f||g) < D^{w}_{2(\alpha,\beta)}(f||g)$.

Theorem 3. If X and Y are two non-negative continuous random variables with probability density functions f(x) and g(x), then the inequality is as follows

$$D^{w}_{\alpha,\beta}(f||g) \geq \frac{1}{log\beta} D_{KL}(f||g) + \alpha \int_0^\infty f(x) log x, \alpha \neq \beta, \beta < \alpha+1, \alpha, \beta > 0 \tag{23}$$

Proof. By using log-sum inequality, we have

$$\int_0^\infty f(x) log \frac{f(x)}{\left(x^{\alpha} f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)} dx \geq \int_0^\infty f(x) log \frac{f(x)}{\left(x^{\alpha} f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)} dx \tag{24}$$

$$\int_0^\infty f(x) log \frac{f(x)}{\left(x^{\alpha} f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)} dx = -log \int_0^\infty \left(x^{\alpha} f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right) dx \tag{25}$$

$$\int_0^\infty f(x) log \frac{f(x)}{\left(x^{\alpha} f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)} dx = -log\left[\beta(\alpha-\beta) D^{w}_{\alpha,\beta}(f||g)\right] \tag{26}$$

Now from L.H.S of (24), we have

$$\int_0^\infty f(x) log \frac{f(x)}{\left(x^{\alpha} f(x)^{\alpha-\beta+1} g(x)^{\beta-\alpha} - 1\right)} dx = H(x)\left[(\alpha-\beta+1) - 1\right] + (\beta-\alpha) I(X,Y) - \alpha \int_0^\infty f(x) log x \tag{27}$$

$$\int_0^\infty f(x)log\frac{f(x)}{\left(x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx=\frac{1}{log\beta(\alpha-\beta)} \\ [H(x)(\alpha-\beta)+(\beta-\alpha)I(X,Y)]-\alpha\int_0^\infty f(x)logx \tag{28}$$

and

$$\int_0^\infty f(x)log\frac{f(x)}{\left(x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}-1\right)}dx=-\frac{1}{log\beta}\left[-H(x)-I(X,Y)\right] \\ +\alpha\int_0^\infty f(x)logx \tag{29}$$

Using (26) and (30), we obtained the result. ■

Theorem 4. Let X and Y be two random variables with weighted generalized divergence(WGD) $D^w_{\alpha,\beta}(f||g)$ and $\alpha\neq\beta$, α, $\beta>0$, then

$$D^w_{\alpha,\beta}(f||g)\leq\frac{1}{\beta(\alpha-\beta)}\int_0^\infty x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}dx-\left[\frac{1}{\beta(\alpha-\beta)}+1\right] \tag{30}$$

Proof. Since we know that for any pasitive number(for any $x>0$) then by using this inequality logx $\leq$ x-1 we get

$$D^w_{\alpha,\beta}(f||g)=log\frac{1}{\beta(\alpha-\beta)}\left[\int_0^\infty x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}dx-1\right] \tag{31}$$

$$D^w_{\alpha,\beta}(f||g)=\frac{1}{\beta(\alpha-\beta)}\left[\int_0^\infty x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}dx-1\right]-1 \tag{32}$$

$$D^w_{\alpha,\beta}(f||g)=\frac{1}{\beta(\alpha-\beta)}\int_0^\infty x^\alpha f(x)^{\alpha-\beta+1}g(x)^{\beta-\alpha}dx-\frac{1}{\beta(\alpha-\beta)}-1 \tag{33}$$

After simplification, we obtained the result. ■

3. Montonic properties

Definition 3.1 A function's monotonicity gives insight into how it will behave. If the graph of a function increases only as the equation's values increase, the function is said to be monotonically increasing. Similar to this, a function is said to be monotonically declining if its values exclusively decrease. In this section, we demonstrate the monotonic properties of the proposed divergenec measure. Consider the following numerical examples:
Example 2.2. Assume X and Y are two non-negative random variables with probability density functions as

1. $f_1(x)=2x;\quad 0<x<1$ and $\quad g_1(x)=2(1-x);\quad 0<x<1$
2. $f_2(x)=\frac{x}{2};\quad 0<x<2$ and $\quad g_1(x)=\frac{(2-x)}{2};\quad 0<x<2$

Using (5) after simplification, we get

$$D_{1(\alpha,\beta)}(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\Gamma(\alpha-\beta+2)\Gamma(\beta-\alpha+1)-1\right] = D_{2(\alpha,\beta)}(f||g) \tag{34}$$

Again using (19) after simplification, we get

$$D^{w}_{1(\alpha,\beta)}(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\left(\frac{2\Gamma(2\alpha-\beta+2)}{\Gamma(\alpha+3)}\right)-1\right]$$

$$D^{w}_{2(\alpha,\beta)}(f||g) = \frac{1}{\beta(\alpha-\beta)}\left[\left(\frac{2^{(2\alpha+1)}\Gamma(2\alpha-\beta+2)}{\Gamma(\alpha+3)}\right)-1\right]$$

Where,

B(u,v) = $\int_0^1 x^{u-1}(1-x)^{v-1}dx = \int_0^1 \frac{x^{u-1}}{(1+x)^{u+v}} = \frac{\Gamma u \Gamma v}{\Gamma(u+v)}$ which is known as complete beta function.

Here, from the below graphs (a), (b), and (c), we obtain that for different values of α, β and n then the measure $D_{1(\alpha,\beta)}(f||g), D_{2(\alpha,\beta)}(f||g), D^{w}_{1(\alpha,\beta)}(f||g)$, and $D^{w}_{2(\alpha,\beta)}(f||g)$ respectively, indicate increased behavior.

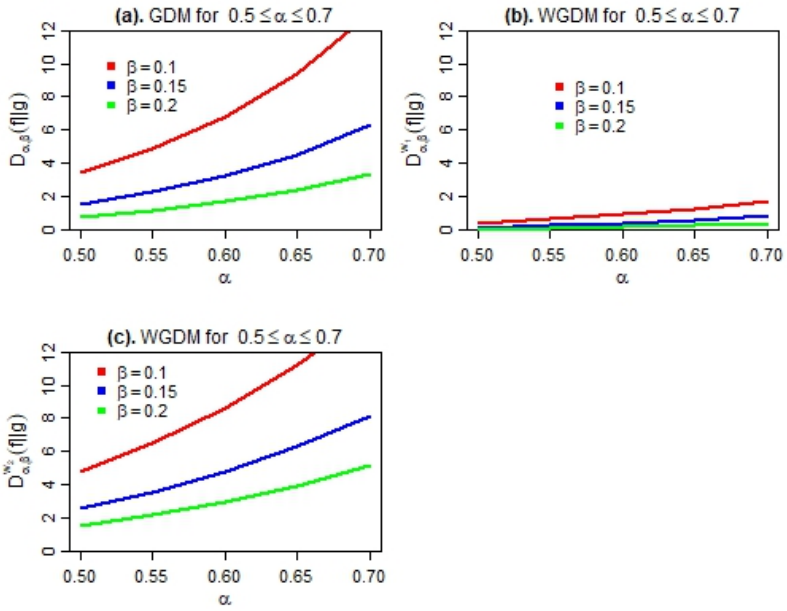


Figure 2: *Monotonic behavior of proposed weighted and non-weighted measure*

4. Divergence measures for some well-known life-time distribution

In this section, we obtained the divergence measures for some life-time distributions using the new proposed divergence measure.

Where,

U = $\frac{1}{\beta(\alpha-\beta)}$, p= $\beta-\alpha$, w= $\theta p+1$, and B(u,v) = $\int_0^1 x^{u-1}(1-x)^{v-1}dx = \int_0^1 \frac{x^{u-1}}{(1+x)^{u+v}} = \frac{\Gamma u \Gamma v}{\Gamma(u+v)}$ beta function.

Table 2: *Proposed divergence measure for some life-time distribution*

Distribution	f(x)	g(x)	x	Proposed Divergence measure
Uniform	$\frac{1}{m}$	$\frac{\theta(m-x)^{(\theta-1)}}{m^{\theta}}$	$0 \leq x \leq m$	$U\left[\frac{\theta^p}{p(\theta-1)+1}-1\right]$
Exponential	ne^{-nx}	$n\theta e^{-n\theta x}$	$x \geq n$, $\theta > 0$	$U\left[\frac{\theta^p}{(\alpha-\beta+w)}-1\right]$
Finite range	$r(1-x)^{r-1}$	$r\theta(1-x)^{(r\theta-1)}$	$0 < x < 1; r, \theta > 0$	$U\left[\frac{\theta^p}{(\alpha-\beta+w)}-1\right]$
Beta	sx^{s-1}	$s\theta x^{s\theta-1}$	$0 < x < 1$, $\theta, s > 0$	$U\left[\frac{\theta^p}{(\alpha-\beta+w)}-1\right]$
Power	$\frac{bx^{b-1}}{c^b}$	$\frac{b\theta x^{b\theta-1}}{c^{b\theta}}$	$0 < x < c; b, \theta > 0$	$U\left[\frac{\theta^p}{\alpha-\beta+w}-1\right]$

Table 3: *Proposed weighted divergence measure for some life-time distribution*

Distribution	f(x)	g(x)	x	Proposed Divergence measure
Uniform	$\frac{1}{m}$	$\frac{\theta(m-x)^{(\theta-1)}}{m^{\theta}}$	$0 \leq x \leq m$	$U\left[\frac{\theta^p m^{\alpha}\Gamma(\alpha+1)\Gamma(w-p)}{\Gamma(\alpha+w-p+1)}-1\right]$
Exponential	ne^{-nx}	$n\theta e^{-n\theta x}$	$x \geq n$,	$U\left[\frac{\theta^p\Gamma(\alpha+1)}{(\alpha-\beta+w+1)^{\alpha+1}}-1\right]$
Finite	$r(1-x)^{r-1}$	$r\theta(1-x)^{(r\theta-1)}$	$0 < x < 1$	$U\left[\frac{\theta^p r\Gamma(\alpha+1)\Gamma(rw+\alpha-\beta)}{\Gamma(2\alpha+rw-\beta+1)}-1\right]$
Beta	sx^{s-1}	$s\theta x^{s\theta-1}$	$0 < x < 1$	$U\left[\frac{\theta^p}{(\alpha-\beta+w)+\alpha}-1\right]$
Power	$\frac{bx^{b-1}}{c^b}$	$\frac{b\theta x^{b\theta-1}}{c^{b\theta}}$	$0 < x < c$	$U\left[\frac{\theta^p bc^{\alpha}}{\alpha-b\theta(\alpha-\beta)+b(\alpha-\beta+1)}-1\right]$

5. Application

Concerning the applicability of the newly proposed divergence measure, we analyzed two sets of actual data published by Almongy et al.[1] based on COVID-19. The first data set was taken over 108 days from Mexico country. Data was collected from March 4 to July 20, 2020.This data set represents the mortality rate. We consider only 30 observations from 108 observations using a random number table.

Dataset-1

1.041,2.988,5.242,7.903,6.327,7.840,7.267,6.370,2.926,5.985,7.854,3.233,7.151,4.292,2.326,3.298,5.459, 3.440,3.215,4.661,3.499,3.395,2.070,2.506,3.029,3.359,3.778,3.219,4.120,8.551.

The second data set was taken over 30 days from the Netherlands country. Data was collected from March 31 to April 30, 2020.This data set also shows mortality rates.

Dataset-2

1.273,6.027,10.656,12.274,1.974,4.960,5.555,7.584,3.883,4.462,4.235,5.307,7.968,13.211,3.611,3.647,6.940, 7.498,5.928,7.099,2.254,5.431,10.289,10.832,4.097,5.048,1.416,2.857,3.461,14.918. With the parameters θ_1 and θ_2, both sets of data can be fitted as an exponential distribution.Here we used MLE method for unknown parameter estimation.The estimated value of parameter $\hat{\theta}_1$ = 0.220 and $\hat{\theta}_2$=0.1624 with different standard error. The estimated value of weighted proposed divergence measure are $\hat{D}^{w}_{1\alpha,\beta}(f||g)$ = 1.543

$\hat{D}^{w}_{2\alpha,\beta}(f||g)$ = 0.024. Our analysis demonstrates that Mexico has a higher mortality rate than the Netherlands.

6. Conclusions

In this communication, we proposed a new two parametric weighted generalized divergence measure of orderα and type β. The characterization result is justified by the numerical example that it uniquely determines the distribution functions, and we also studied the mononic behaviour

of the proposed divergence measure.Finally, we derived some expressions for some life-time distributions and also showed the mortality rate of two different countries.

References

[1] Almongy, H. M., Almetwally, E. M., Aljohani, H. M., Alghamdi, A. S., and Hafez, E. H. (2021). A new extended Rayleigh distribution with applications of COVID-19 data *Results in Physics*, 23:104012.

[2] Belis, M. and Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems (Corresp.). *IEEE Transactions on Information Theory*, 14(4):593–594.

[3] Di Crescenzo, A. and Longobardi, M. (2007). On weighted residual and past entropies. *arXiv preprint math/0703489*.

[4] Ebrahimi, N. and Kirmani, S. N. A. (1996). A characterisation of the proportional hazards model through a measure of discrimination between two residual life distributions. *Biometrika*, 83(1):233–235.

[5] Havrda, J. and Charvat, F. (1967). Quantification method of classification processes. Concept of structural *a*-entropy *Kybernetika*, 3(1):30–35.

[6] Kullback, S. and Leibler, R. A. (1951). information and sufficiency. *Annals of mathematics statistics*, 22(1):79–86.

[7] Kerridge, D. F. (1961). Inaccuracy and inference. *Journal of the Royal Statistical Society. Series B (Methodological)*,184–194.

[8] Mirali, M.A.L.I.H.E H. and Fakoor, V. (2017). On weighted cumulative residual entropy. *Communications in Statistics-Theory and Methods*, 46(6):2857–2869.

[9] Moharana, R. and Kayal, S. (2019). On Weighted Kullback–Leibler Divergence for Doubly Truncated Random Variables. *REVSTAT-Statistical Journal*, 17(3):297–320.

[10] Gupta, R. D. and Nanda, A. (2002). α and β entropies and relative entropies of distribution. *Journal of statistics theory of application*, 1(30):177–190.

[11] Renyi, A. (1961). On measures of entropy and information. Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. *In proceeding of the fourth Berkeley symposium on mathematical statistics and probability*, 1:547–561.

[12] Shannon, C. E. (1948). A mathematical theory of communications. *Bell System Technical Journal*, 27:379–423.

[13] Suhov, Y. and Salimeh, S. Y. (2015). Entropy-power inequality for weighted entropy. *arXiv preprint arXiv:1502.02188*.

[14] Yasaei Sekeh, S. and Mohtashami Borzadaran, G. R. (2013). On Kullback-Leibler Dynamic Information. *Available at SSRN 2344078*.

ESTIMATION OF A PARAMETER OF FARLIE-GUMBEL-MORGENSTERN BIVARIATE BILAL DISTRIBUTION BY RANKED SET SAMPLING

M. R. Irshad[1], R. Maya[2], A.I. Al-Omari[3], Ahmad A. Hanandeh[4], S. P. Arun[5]

•

[1]Cochin University of Science and Technology, Kochi-22, Kerala.
irshadm24@gmail.com
[2]University College, Trivandrum-695 034.
publicationsofmaya@gmail.com
[3]Department of Mathematics, Faculty of Science, Al al-Bayt University, Mafraq, Jordan.
alomari_amer@yahoo.com
[4]Yarmouk University, Irbid 21163, Jordan.
ahmad.hanandeh@yu.edu.jo
[5]University Library, Research Centre, University of Kerala, Trivandrum-695 034.
arunsptvm@gmail.com

Abstract

A bivariate version of the Bilal distribution has been proposed in the literature, called the Farlie-Gumbel-Morgenstern bivariate Bilal (FGMBB) distribution. In this article, we have dealt with the problem of estimation of the scale parameter associated with the study variable Z of primary interest, based on the ranked set sample defined by ordering the marginal observations on an auxiliary variable W, when (W, Z) follows a FGMBB distribution. When the dependence parameter ϕ is known, we have proposed the following estimators, viz., an unbiased estimator based on the Stoke's ranked set sample and the best linear unbiased estimator based on the Stoke's ranked set sample for the scale parameter of the variable of primary interest. The efficiency comparison of the proposed estimators with respect to the maximum likelihood estimator have been carried out.

Keywords: Farlie-Gumbel-Morgenstern bivariate Bilal distribution, Concomitants of order statistics, Ranked set sampling, Best linear unbiased estimator

1. Introduction

The Bilal distribution was introduced by [1], as a member of the families of distributions for the median of a random sample arising from an arbitrary lifetime distribution. Also, he shows that, this distribution belongs to the class of new better than average renewal failure rates and its probability density function (pdf) is always unimodal and has less of skewness and kurtosis than the pdf of the exponential distribution by about 25% and 28% respectively. The cumulative distribution function (cdf) of the Bilal distribution with the scale parameter σ is given by

$$F(x;\sigma) = 1 - e^{-\frac{2x}{\sigma}}\left(3 - 2e^{-\frac{x}{\sigma}}\right);\ \sigma > 0,\ x > 0. \tag{1}$$

The corresponding pdf is given by

$$f(x;\sigma) = \frac{6}{\sigma}e^{-\frac{2x}{\sigma}}\left(1 - e^{-\frac{x}{\sigma}}\right);\ \sigma > 0,\ x > 0. \tag{2}$$

Furthermore, the author obtained the closed form expressions for the quantile function, the hazard rate function and simple expression for moments in terms of the exponential function. Even though the Bilal distribution has only one parameter, this distribution possess high fitting ability compared to other competing models for two different real datasets, namely, the dataset consisting of thirty successive values of precipitation (in inches) given by [14] and the data for waiting times before service of 100 bank customers reported by [13]. Based on type-2 censored sample, [2] provide certain estimators of the parameter of the Bilal distribution. According to [3], the one parameter Bilal model can be generalized into the two parameter Bilal model, whose applications are elaborately discussed. Now the Proficiency of univariate Bilal distribution compared to other competing models well established in the literature in the theoretical as well as applied perspective. But even a single work is not been seen so far in the available literature on bivariate Bilal model except the work of [17]. A bivariate extension of one parameter Bilal distribution using Morgenstern approach was proposed by [17], so-called the Farlie-Gumbel-Morgenstern Bivariate Bilal (FGMBB) distribution and elucidated its inferential aspects using concomitants of order statistics (COS).
A bivariate random variable (W, Z) is said to follow a FGMBB distribution, if its pdf is given by

$$f(w,z) = \begin{cases} \frac{36}{\sigma_1\sigma_2} e^{-\frac{2w}{\sigma_1}} \left(1 - e^{-\frac{w}{\sigma_1}}\right) e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \\ \times \left[1 + \phi \left(2e^{-\frac{2w}{\sigma_1}} \left\{3 - 2e^{-\frac{w}{\sigma_1}}\right\} - 1\right) \left(2e^{-\frac{2z}{\sigma_2}} \left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\} - 1\right)\right], \\ \qquad w > 0, z > 0; \sigma_1 > 0, \sigma_2 > 0; -1 \leq \phi \leq 1. \\ 0, \quad \text{otherwise.} \end{cases} \tag{3}$$

Clearly the marginal distributions of W and Z variables are univariate Bilal distributions with pdf's are respectively given by

$$f_W(w) = \begin{cases} \frac{6}{\sigma_1} e^{-\frac{2w}{\sigma_1}} \left(1 - e^{-\frac{w}{\sigma_1}}\right); \ \ if \ \ \sigma_1 > 0, \ w > 0, \\ 0, \qquad otherwise. \end{cases}$$

and

$$f_Z(z) = \begin{cases} \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right); \ \ if \ \ \sigma_2 > 0, \ z > 0, \\ 0, \qquad otherwise. \end{cases} \tag{4}$$

Clearly,

$$E(W) = \frac{5}{6}\sigma_1, Var(W) = \frac{13}{36}\sigma_1^2,$$

$$E(Z) = \frac{5}{6}\sigma_2, \tag{5}$$

$$Var(Z) = \frac{13}{36}\sigma_2^2. \tag{6}$$

The ranked set sampling (RSS) scheme was first developed by [19] as a process of increasing the precision of the sample mean as an estimator of the population mean. McIntyre's idea of ranking is possible whenever it can be done easily by a judgement method. For a detailed discussion on the theory and applications of RSS [11]. Basically the procedure involves choosing n sets of units, each of size n, and ordering the units of each of the set by judgement method or by applying some inexpensive method, without making actual measurement on the units. Then the unit ranked as one from the 1^{st} set is actually measured, the unit ranked as two from the 2^{nd} set is measured. The process continuous in this way until the unit ranked as n from the n^{th} set is measured. Then the observations obtained under the afore mentioned criterion is known as ranked set sample (*rss*) and the procedure is known as RSS. For recent developments in RSS, one can refer [6], [4] and [5].

In some practical problems, the variable of primary concern say Z, is more intricate to measure, but an auxiliary variable W related with Z is easily measurable and can be ordered exactly. In this case, [22] developed another scheme of RSS, which is as follows: Choose n independent bivariate sets, each of size n. In the first set of size n, the Z variate associated with smallest ordered W is measured, in the second set of size n, the Z variate associated with the second smallest, W is measured. This process is continued until the Z associated with the largest W from the n^{th} set is measured. The measurements on the Z variate of the resulting new set of n units chosen by the above method gives a *rss* as suggested by [22]. If $W_{(r:n)r}$ is the observation measured on the auxiliary variable W from the unit chosen from the r^{th} set, then we write $Z_{[r:n]r}$ to denote the corresponding measurement made on the study variable Z on this unit so that $Z_{[r:n]r}, r = 1, 2, \cdots, n$ form the *rss*. $Z_{[r:n]r}$ was referred by [12] as the concomitant of the r^{th} order statistic arising from the r^{th} sample.

The *rss* mean as an estimator for the mean of the study variate Z, when an auxiliary variable W is used for ranking the sample units has suggested by [22], under the assumption that (W, Z) follows a bivariate normal distribution. Based on *rss* obtained on the study variate Z, [10] have improved the estimator of [22] by deriving the best linear unbiased estimator (BLUE) of the mean of the study variate Z. COS and its applications in RSS from Farlie-Gumbel-Morgenstern bivariate Lomax distribution is elaborately elucidated by [20]. The estimation of a parameter of Morgenstern type bivarite Lindley distribution by RSS has been discussed in [15]. Parameter estimation of Cambanis-type bivariate uniform distribution with RSS is studied by [16]. For review of various variants of RSS and their application in parameter estimation [11].

The remaining part of this paper is assembled as follows. In section 2, we have proposed an unbiased estimator σ_2^* of σ_2 using Stoke's *rss*. As mentioned earlier if (W, Z) has a FGMBB distribution as defined in (3), then the marginal distributions of both W and Z have Bilal distributions and the pdf of Z is given in (4). We have evaluated the Cramer-Rao Lower Bound (CRLB) for the variance of an unbiased estimator of σ_2 involved in (4) based on a random sample of size n and is given by $\frac{13}{25}\frac{\sigma_2^2}{n}$. In this section, we have also shown that the variance of proposed unbiased estimator σ_2^* is strictly less than $\frac{13}{25}\frac{\sigma_2^2}{n}$, the CRLB for the variance of an unbiased estimator of σ_2 involved in (4), for all $\phi \in B$, where $B = [-1, 1] - \{0\}$. In this section, we have further discussed an efficiency comparison between σ_2^* and the maximum likelihood estimator (MLE) $\hat{\sigma_2}$ of σ_2 based on a random sample of size n arising from (3). In section 3, we have derived the BLUE $\tilde{\sigma_2}$ of σ_2 involved in FGMBB distribution based on Stoke's *rss* and made an efficiency comparison of $\tilde{\sigma_2}$ relative to $\hat{\sigma_2}$.

2. An unbiased estimator of σ_2 using Stoke's RSS.

Suppose the bivariate random vector (W, Z) follows a FGMBB distribution with pdf given in (3). Select a *rss* as per Stoke's RSS scheme. Let $W_{(r:n)r}$ be the observation obtained on the auxiliary variate W in the r^{th} unit of the *rss* and let $Z_{[r:n]r}$ be the measurement made on the variate related with $W_{(r:n)r}, r = 1, 2, \cdots, n$. Clearly $Z_{[r:n]r}$ is the r^{th} COS of a random sample of size n arising from the FGMBB distribution. Using the results of [21], we obtain the pdf of $Z_{[r:n]r}, r = 1, 2, \cdots, n$, and is given by

$$f_{[r:n]}(z) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \left[1 + \phi \frac{(n-2r+1)}{(n+1)} \left(2e^{-\frac{2z}{\sigma_2}} \left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\} - 1\right)\right]. \tag{7}$$

The mean and variance of $Z_{[r:n]r}$ for $r = 1, 2, \cdots, n$, is obtained as

$$E[Z_{[r:n]r}] = \sigma_2 \left[\frac{5}{6} - \frac{19}{60}\phi \frac{(n-2r+1)}{(n+1)}\right] \tag{8}$$

and

$$Var[Z_{[r:n]r}] = \sigma_2^2 \left[\frac{13}{36} - \frac{253}{1800}\phi \frac{(n-2r+1)}{(n+1)} - \frac{361}{3600}\phi^2 \frac{(n-2r+1)^2}{(n+1)^2}\right]. \tag{9}$$

Since $Z_{[r:n]r}$ and $Z_{[s:n]s}$ for $r \neq s$ are arising from two independent samples, we obtain

$$Cov[Z_{[r:n]r}, Z_{[s:n]s}] = 0, r \neq s.$$

Next, we derive an unbiased estimator of σ_2 and its variance using the *rss* observations $Z_{[r:n]r}$ for $r = 1, 2, \cdots, n$, on the variable Z of primary interest and are given by the following theorem.

Theorem 1. Let (W, Z) follows a FGMBB distribution with pdf given by (3). Let $Z_{[r:n]r}$, $r = 1, 2, \cdots, n$ be the *rss* observations on a study variate Z generated out of ranking made on an auxiliary variate W. Then

$$\sigma_2^* = \frac{6}{5n} \sum_{r=1}^{n} Z_{[r:n]r}$$

is an unbiased estimator of σ_2 and its variance is given by

$$Var[\sigma_2^*] = \frac{\sigma_2^2}{n} \left[\frac{13}{25} - \frac{361}{2500} \frac{\phi^2}{n} \sum_{r=1}^{n} \left(\frac{n-2r+1}{n+1} \right)^2 \right]. \tag{10}$$

Proof By using the definition, we have

$$\begin{aligned} E[\sigma_2^*] &= \frac{6}{5n} \sum_{r=1}^{n} E[Z_{[r:n]r}] \\ &= \frac{6}{5n} \sum_{r=1}^{n} \left[\frac{5}{6} - \frac{19}{60} \phi \frac{(n-2r+1)}{(n+1)} \right] \sigma_2. \end{aligned} \tag{11}$$

Using the result,

$$\sum_{r=1}^{n} (n-2r+1) = 0. \tag{12}$$

Applying (12) in (11) we get,

$$E[\sigma_2^*] = \sigma_2.$$

Therefore, σ_2^* is an unbiased estimator of σ_2. The variance of σ_2^* is given by,

$$Var[\sigma_2^*] = \frac{36}{25n^2} \sum_{r=1}^{n} Var[Z_{[r:n]r}]. \tag{13}$$

Applying (9) and (12) in (13), we get

$$Var[\sigma_2^*] = \frac{\sigma_2^2}{n} \left[\frac{13}{25} - \frac{361}{2500} \frac{\phi^2}{n} \sum_{r=1}^{n} \left(\frac{n-2r+1}{n+1} \right)^2 \right].$$

Hence the proof.

As mentioned above, if (W, Z) has the FGMBB distribution as defined in (3), then the marginal distribution of both W and Z are Bilal distributions and the pdf of Z is given in (4). The CRLB for the variance of any unbiased estimator of σ_2 based on a random sample of size n drawn from (4) is obtained as $\frac{13}{25} \frac{\sigma_2^2}{n}$. Now we compare the the variance of σ_2^* with the CRLB for the variance of an unbiased estimator of σ_2 involved in (4). If we write $E_1(\sigma_2^*)$ to denote the ratio of $\frac{13}{25} \frac{\sigma_2^2}{n}$ with $Var(\sigma_2^*)$, then we have,

$$E_1(\sigma_2^*) = \frac{1}{\left[1 - \frac{361}{1300} \frac{\phi^2}{n} \sum_{r=1}^{n} \left(\frac{n-2r+1}{n+1} \right)^2 \right]}. \tag{14}$$

It is easily verified that

$$E_1(\sigma_2^*) \geq 1.$$

Thus we arrive at a conclusion that the estimator σ_2^* based on Stoke's *rss* is more efficient as it assert the statement that *rss* always provide more information than simple random sample even if ranking is imperfect [11]. It is very clear that $Var(\sigma_2^*)$ is a decreasing function of ϕ^2 and hence the gain in efficiency of the estimator σ_2^* increases as $|\phi|$ increases.
Again on simplifying (14) we get,

$$E_1(\sigma_2^*) = \frac{1}{1 - \frac{361\phi^2}{1300}\left[\frac{2}{3}\left(\frac{2+1/n}{1+1/n}\right) - 1\right]}.$$

Then,

$$\begin{aligned}\lim_{n\to\infty} E_1(\sigma_2^*) &= \lim_{n\to\infty} \frac{1}{1 - \frac{361\phi^2}{1300}\left[\frac{2}{3}\left(\frac{2+1/n}{1+1/n}\right) - 1\right]} \\ &= \frac{1}{1 - \frac{361\phi^2}{3900}}.\end{aligned}$$

From the above expression it is clear that the maximum value for $E_1(\sigma_2^*)$ is attained when $|\phi| = 1$ and in this case $E_1(\sigma_2^*)$ tends to 3900/3539.
Next we discuss the efficiency comparison of σ_2^* with the asymptotic variance of MLE of σ_2 involved in the FGMBB distribution. If (W, Z) follows a FGMBB distribution with pdf given in (3), then

$$\begin{aligned}\frac{\partial \log f(x,y)}{\partial \sigma_1} = \frac{1}{\sigma_1}\Bigg\{&-1 + \frac{2w}{\sigma_1} - \frac{we^{-\frac{w}{\sigma_1}}}{\sigma_1(1-e^{-\frac{w}{\sigma_1}})} \\ &+ \frac{4\phi w e^{-\frac{2w}{\sigma_1}}\left[-3 + 18e^{-\frac{2z}{\sigma_2}} - 12e^{-\frac{3z}{\sigma_2}} + 3e^{-\frac{w}{\sigma_1}} - 18e^{-\frac{w}{\sigma_1}}e^{-\frac{2z}{\sigma_2}} + 12e^{-\frac{w}{\sigma_1}}e^{-\frac{3z}{\sigma_2}}\right]}{\sigma_1\left\{1 + \phi\left[1 - 2e^{-\frac{2w}{\sigma_1}}(3-2)e^{-\frac{w}{\sigma_1}}\right]\left[1 - 2e^{-\frac{2z}{\sigma_2}}(3-2)e^{-\frac{z}{\sigma_2}}\right]\right\}}\Bigg\}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \log f(x,y)}{\partial \sigma_2} = \frac{1}{\sigma_2}\Bigg\{&-1 + \frac{2z}{\sigma_2} - \frac{ze^{-\frac{z}{\sigma_2}}}{\sigma_2(1-e^{-\frac{z}{\sigma_2}})} \\ &+ \frac{4\phi z e^{-\frac{2z}{\sigma_2}}\left[-3 + 18e^{-\frac{2w}{\sigma_1}} - 12e^{-\frac{3w}{\sigma_1}} + 3e^{-\frac{z}{\sigma_2}} - 18e^{-\frac{z}{\sigma_2}}e^{-\frac{2w}{\sigma_1}} + 12e^{-\frac{z}{\sigma_2}}e^{-\frac{3w}{\sigma_1}}\right]}{\sigma_2\left\{1 + \phi\left[1 - 2e^{-\frac{2w}{\sigma_1}}(3-2)e^{-\frac{w}{\sigma_1}}\right]\left[1 - 2e^{-\frac{2z}{\sigma_2}}(3-2)e^{-\frac{z}{\sigma_2}}\right]\right\}}\Bigg\}.\end{aligned}$$

Then we have,

$$\begin{aligned}I_{\sigma_1}(\phi) &= E\left(\frac{\partial \log f(x,y)}{\partial \sigma_1}\right)^2 \\ &= \frac{36}{\sigma_1^2}\int_0^\infty\int_0^\infty e^{-2u}(1-e^{-u})\Big\{-1 + 4u + ue^{-u}(u-2+2e^{-u})(1-e^{-u})^{-2} \\ &- \frac{12\alpha u^2 e^{-2u}\left[-2 + 12e^{-2v} - 8e^{-3v} + 3e^{-u} - 18e^{-u}e^{-2v} + 12e^{-u}e^{-3v}\right]}{\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}} \\ &+ \frac{24\alpha u e^{-2u}\left[-1 + 6e^{-2v} - 4e^{-3v} + e^{-u} - 6e^{-u}e^{-2v} + 4e^{-u}e^{-3v}\right]}{\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}^2} \\ &\times\{1+\alpha[1-2e^{-2v}(3-2e^{-v})][1-6e^{-2u}-6ue^{-2u}+4e^{-3u}+6ue^{-3u}]\} \\ &\}\, e^{-2v}(1-e^{-v})\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}dudv,\end{aligned}$$

$$
\begin{aligned}
I_{\sigma_2}(\phi) &= E\left(\frac{\partial \log f(x,y)}{\partial \sigma_2}\right)^2 \\
&= \frac{36}{\sigma_2^2}\int_0^\infty\int_0^\infty e^{-2u}(1-e^{-u})\Big\{-1+4v+ve^{-v}(v-2+2e^{-v})(1-e^{-v})^{-2} \\
&-\frac{12\alpha v^2 e^{-2v}\left[-2+12e^{-2u}-8e^{-3u}+3e^{-v}-18e^{-v}e^{-2u}+12e^{-v}e^{-3u}\right]}{\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}} \\
&+\frac{24\alpha v e^{-2v}\left[-1+6e^{-2u}-4e^{-3u}+e^{-v}-6e^{-v}e^{-2u}+4e^{-v}e^{-3u}\right]}{\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}^2} \\
&\times\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-6e^{-2v}-6ve^{-2v}+4e^{-3v}+6ve^{-3v}]\} \\
&\}\, e^{-2v}(1-e^{-v})\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}dudv
\end{aligned}
$$

and

$$
\begin{aligned}
I_{\sigma_1\sigma_2}(\phi) &= E\left(\frac{\partial^2 \log f(x,y)}{\partial \sigma_1 \partial \sigma_2}\right) \\
&= \frac{36}{\sigma_1\sigma_2}\int_0^\infty\int_0^\infty e^{-2u}(1-e^{-u})\Big\{144\alpha uve^{-2u}e^{-2v}\left[1-e^{-u}-e^{-v}+e^{-u}e^{-v}\right] \\
&-\frac{144\alpha^2 uve^{-2u}e^{-2v}\left[-1+6e^{-2v}-4e^{-3v}+e^{-u}-6e^{-u}e^{-2v}+4e^{-u}e^{-3v}\right]}{\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}^2} \\
&\times\{[1-2e^{-2u}(3-2e^{-u})][e^{-v}-1]\} \\
&\}\, e^{-2v}(1-e^{-v})\{1+\alpha[1-2e^{-2u}(3-2e^{-u})][1-2e^{-2v}(3-2e^{-v})]\}dudv.
\end{aligned}
$$

Thus the Fisher information matrix associated with the random variable (W, Z) is given by,

$$
I(\phi) = \begin{bmatrix} I_{\sigma_1}(\phi) & -I_{\sigma_1\sigma_2}(\phi) \\ -I_{\sigma_1\sigma_2}(\phi) & I_{\sigma_2}(\phi) \end{bmatrix}. \tag{15}
$$

We have computed the values of $\sigma_1^{-2}I_{\sigma_1}(\phi)$ and $\sigma_1^{-1}\sigma_2^{-1}I_{\sigma_1\sigma_2}(\phi)$ numerically for $\phi = \pm 0.25, \pm 0.50, \pm 0.75, \pm 1$ (clearly $\sigma_1^{-2}I_{\sigma_1}(\phi)$=$\sigma_2^{-2}I_{\sigma_2}(\phi)$) and are given below:

ϕ	$\sigma_1^{-2}I_{\sigma_1}(\phi)$	$\sigma_1^{-1}\sigma_2^{-1}I_{\sigma_1\sigma_2}(\phi)$	ϕ	$\sigma_1^{-2}I_{\sigma_1}(\phi)$	$\sigma_1^{-1}\sigma_2^{-1}I_{\sigma_1\sigma_2}(\phi)$
0.25	1.9381	0.1373	-0.25	1.9381	-0.1373
0.50	1.9795	0.2772	-0.50	1.9795	-0.2773
0.75	2.0530	0.4230	-0.75	2.0530	-0.4236
1.00	2.1705	0.5815	-1.00	2.1705	-0.5841

Thus from (15), the asymptotic variance of the MLE $\hat{\sigma_2}$ of σ_2 involved in the FGMBB distribution under a bivariate sample of size n is obtained as

$$
Var(\hat{\sigma_2}) = \frac{1}{n}I_{\sigma_2}^{-1}(\phi), \tag{16}
$$

where $I_{\sigma_2}^{-1}(\phi)$ is the (2,2)th element of the inverse of $I(\phi)$ given by (15).

We have compute the efficiency $E(\sigma_2^*|\hat{\sigma_2}) = \frac{Var(\hat{\sigma_2})}{Var(\sigma_2^*)}$ of σ_2^* relative to $\hat{\sigma_2}$ for $n = 2(2)20$; $\phi = \pm 0.25, \pm 0.50, \pm 0.75, \pm 1$ and are given in table 1. From the table, one can infer that the estimator σ_2^* is more efficient than $\hat{\sigma_2}$ and efficiency increases with n and $|\phi|$ for $n \geq 4$.

Remark 2.1. For given value of $\phi \in (0, 1]$, once the variance of σ_2^* is evaluated, then this variance is equal to the variance of σ_2^* for $-\phi$ because the variance given in (10) depends only on ϕ by a term containing ϕ^2 only.

3. Best linear unbiased estimator of σ_2 using Stoke's RSS

In this section we derive the BLUE of σ_2 provided the dependence parameter ϕ is known. Suppose $Z_{[r:n]r}$ for $r = 1, 2, \cdots, n$, are the rss observation generated from (3) as per Stoke's RSS scheme. Let

$$\zeta_{r,n} = \frac{5}{6} - \frac{19}{60}\phi\frac{(n-2r+1)}{(n+1)}, \tag{17}$$

$$\psi_{r,r,n} = \frac{13}{36} - \frac{253}{1800}\phi\frac{(n-2r+1)}{(n+1)} - \frac{361}{3600}\phi^2\frac{(n-2r+1)^2}{(n+1)^2}. \tag{18}$$

Using (17) and (18), we get

$$E[Z_{[r:n]r}] = \sigma_2\zeta_{r,n},\ 1 \le r \le n \tag{19}$$

and

$$Var[Z_{[r:n]r}] = \sigma_2^2\psi_{r,r,n},\ 1 \le r \le n. \tag{20}$$

Also we have

$$Cov[Z_{[r:n]r}, Z_{[s:n]s}] = 0, r, s = 1, 2, \cdots, n \text{ and } r \ne s. \tag{21}$$

Let $\mathbf{Z}_{[n]} = (Z_{[1:n]1}, Z_{[2:n]2}, \cdots, Z_{[n:n]n})'$ denote the column vector of n rss observations. Then from (19), (20) and (21), we can write,

$$E[\mathbf{Z}_{[n]}] = \sigma_2\zeta \tag{22}$$

and the dispersion matrix of $\mathbf{Z}_{[n]}$,

$$D[\mathbf{Z}_{[n]}] = \sigma_2^2\mathbf{G}, \tag{23}$$

where $\zeta = (\zeta_{1,n}, \zeta_{2,n}, \cdots, \zeta_{n,n})'$ and $\mathbf{G} = diag(\psi_{1,1,n}, \psi_{2,2,n}, \cdots, \psi_{n,n,n})$, where $\zeta_{r,n}$ and $\psi_{r,r,n}$ for $r = 1, 2, \cdots, n$ are respectively given by equations (17) and (18). If ϕ contained in ζ and $\mathbf{G}$ are known then (22) and (23) together defines a generalized Gauss-Markov setup and then the BLUE of σ_2 is given by

$$\tilde{\sigma_2} = (\zeta'\mathbf{G}^{-1}\zeta)^{-1}\zeta'\mathbf{G}^{-1}\mathbf{Z}_{[n]}$$

and the variance of σ_2 is given by

$$Var(\tilde{\sigma_2}) = \frac{\sigma_2^2}{\zeta'\mathbf{G}^{-1}\zeta}.$$

On simplifying, we get

$$\tilde{\sigma_2} = \frac{\sum_{r=1}^{n}\frac{\zeta_{r,n}}{\psi_{r,r,n}}}{\sum_{r=1}^{n}\frac{\zeta_{r,n}^2}{\psi_{r,r,n}}}Z_{[r:n]r} \tag{24}$$

and

$$Var(\tilde{\sigma_2}) = \frac{\sigma_2^2}{\sum_{r=1}^{n}\frac{\zeta_{r,n}^2}{\psi_{r,r,n}}}. \tag{25}$$

From (24), we have $\tilde{\sigma_2}$ is a linear functions of the *rss* observations $Z_{[r:n]r}, r = 1, 2, \cdots, n$ and hence $\tilde{\sigma_2}$ can be written as $\tilde{\sigma_2} = \sum_{r=1}^{n} a_r Z_{[r:n]r}$, where

$$a_r = \frac{\frac{\zeta_{r,n}}{\psi_{r,r,n}}}{\sum_{r=1}^{n}\frac{\zeta_{r,n}^2}{\psi_{r,r,n}}}, r = 1, 2, \cdots, n.$$

We have evaluated the numerical values of means and variances using the expressions (17) and (18) respectively for $\phi = 0.25(0.25)1$ and for $n = 2(2)20$. Using these values we have evaluated

the variance of BLUE $\tilde{\sigma_2}$ for $\phi = 0.25(0.25)1$ and for $n = 2(2)20$. Also we have computed the ratio $E(\tilde{\sigma}_2|\hat{\sigma}_2) = \frac{Var(\hat{\sigma}_2)}{Var(\tilde{\sigma}_2)}$ to measure the efficiency of our estimator $\tilde{\sigma}_2$ relative to $\hat{\sigma}_2$ for $n = 2(2)20$ and $\phi = 0.25(0.25)1$ and are presented in table 1. From the table it is clear that, BLUE of σ_2 performs well compared to the MLE of σ_2, namely $\hat{\sigma}_2$.

Remark 3.1. As in the case of variance of an unbiased estimator given in (10), for a given value of $\phi \in (0, 1]$, once the variance of $\tilde{\sigma_2}$ is evaluated, then there is no need to again evaluate the variance of $\tilde{\sigma_2}$ when $\phi = -\phi$. To establish this argument we prove the following theorem.

Theorem 2. Let (W, Z) follows a FGMBB distribution with pdf given by (3). For a given $\phi \in (0, 1]$, $Var[\tilde{\sigma}_2(\phi_0)]$ is the variance of the BLUE $\tilde{\sigma}_2$ of σ_2 involved in the FGMBB distribution, then

$$Var[\tilde{\sigma}_2(-\phi_0)] = Var[\tilde{\sigma}_2(\phi_0)]. \tag{26}$$

Proof The terms $\zeta_{r,n}$ and $\psi_{r,r,n}$ defined by (17) and (18) are functions of ϕ, r and n and hence $\zeta_{r,n}$ and $\psi_{r,r,n}$ can be denoted as $\zeta_{r,n}(\phi)$ and $\psi_{r,r,n}(\phi)$ respectively. From (17) and (18), it is clear that

$$\zeta_{r,n}(\phi) = \zeta_{n-r+1,n}(-\phi),\ 1 \le r \le n \tag{27}$$

and

$$\psi_{r,r,n}(\phi) = \psi_{n-r+1,n-r+1,n}(-\phi),\ 1 \le r \le n. \tag{28}$$

As a consequence of (27) and (28), we get

$$\begin{aligned} Var[\tilde{\sigma}_2(\phi)] &= \frac{\sigma_2^2}{\sum_{r=1}^{n} \frac{\zeta_{r,n}^2(\phi)}{\psi_{r,r,n}(\phi)}} = \frac{\sigma_2^2}{\sum_{r=1}^{n} \frac{\zeta_{n-r+1,n}^2(-\phi)}{\psi_{n-r+1,n-r+1,n}(-\phi)}} \\ &= Var[\tilde{\sigma}_2(-\phi)]. \end{aligned}$$

Hence the proof.

Remark 3.2. *For FGMBB distribution defined in (3), we have evaluated the correlation coefficient between the two variables and is given by $\rho = \frac{361}{1300}\phi$. But in certain real life situations our assumption that ϕ is known may viewed as unrealistic. Hence if we have a situation with ϕ unknown, we compute the sample correlation τ from $(W_{r:n}, Z_{[r:n]})$ for $r = 1, 2, \cdots, n$ and introduce a moment type estimator $\hat{\phi}$ for ϕ as,*

$$\hat{\phi} = \begin{cases} -1, & if\ \ \tau < \frac{-361}{1300} \\ \frac{1300}{361}\tau, & if\ \ \frac{-361}{1300} \le \tau \le \frac{361}{1300} \\ 1, & if\ \ \tau > \frac{361}{1300}. \end{cases}$$

Table 1: Efficiencies of the estimators σ_2^* and $\tilde{\sigma_2}$ relative to $\hat{\sigma}_2$.

n	ϕ	$e(\sigma_2^*\|\hat{\sigma}_2)$	$e(\tilde{\sigma}_2\|\hat{\sigma}_2)$	ϕ	$e(\sigma_2^*\|\hat{\sigma}_2)$	$e(\tilde{\sigma}_2\|\hat{\sigma}_2)$
	0.25	0.9992	0.9992	-0.25	0.9992	0.9992
2	0.50	0.9984	0.9984	-0.50	0.9984	0.9984
	0.75	0.9957	0.9957	-0.75	0.9957	0.9957
	1.00	0.9849	0.9849	-1.00	0.9850	0.9853
	0.25	1.0008	1.0008	-0.25	1.0008	1.0008
4	0.50	1.0047	1.0047	-0.50	1.0029	1.0029
	0.75	1.0103	1.0111	-0.75	1.0047	1.0047
	1.00	1.0106	1.0139	-1.00	1.0110	1.0147
	0.25	1.0012	1.0012	-0.25	1.0012	1.0012
6	0.50	1.0082	1.0082	-0.50	1.0082	1.0082
	0.75	1.0168	1.0180	-0.75	1.0168	1.0180
	1.00	1.0223	1.0273	-1.00	1.0230	1.0286
	0.25	1.0015	1.0015	-0.25	1.0015	1.0015

n	ϕ	$e(\sigma_2^*\|\hat{\sigma}_2)$	$e(\tilde{\sigma}_2\|\hat{\sigma}_2)$	ϕ	$e(\sigma_2^*\|\hat{\sigma}_2)$	$e(\tilde{\sigma}_2\|\hat{\sigma}_2)$
8	0.50	1.0094	1.0094	-0.50	1.0094	1.0094
	0.75	1.0192	1.0209	-0.75	1.0192	1.0209
	1.00	1.0282	1.0351	-1.00	1.0300	1.0367
	0.25	1.0019	1.0019	-0.25	1.0019	1.0019
10	0.50	1.0098	1.0098	-0.50	1.0098	1.0098
	0.75	1.0221	1.0241	-0.75	1.0221	1.0241
	1.00	1.0312	1.0398	-1.00	1.0330	1.0419
	0.25	1.0023	1.0023	-0.25	1.0023	1.0023
12	0.50	1.0094	1.0094	-0.50	1.0094	1.0094
	0.75	1.0242	1.0266	-0.75	1.0242	1.0266
	1.00	1.0376	1.0455	-1.00	1.0376	1.0455
	0.25	1.0000	1.0000	-0.25	1.0000	1.0000
14	0.50	1.0110	1.0110	-0.50	1.0110	1.0110
	0.75	1.0225	1.0254	-0.75	1.0225	1.0254
	1.00	1.0380	1.0472	-1.00	1.0380	1.0472
	0.25	1.0031	1.0031	-0.25	1.0031	1.0031
16	0.50	1.0126	1.0126	-0.50	1.0126	1.0126
	0.75	1.0258	1.0291	-0.75	1.0258	1.0291
	1.00	1.0403	1.0473	-1.00	1.0403	1.0473
	0.25	1.0035	1.0035	-0.25	1.0035	1.0035
18	0.50	1.0106	1.0106	-0.50	1.0106	1.0106
	0.75	1.0291	1.0291	-0.75	1.0291	1.0291
	1.00	1.0415	1.0534	-1.00	1.0415	1.0534
	0.25	1.0000	1.0000	-0.25	1.0000	1.0000
20	0.50	1.0118	1.0157	-0.50	1.0118	1.0157
	0.75	1.0242	1.0283	-0.75	1.0242	1.0283
	1.00	1.0420	1.0508	-1.00	1.0420	1.0508

4. Estimation of σ_2 based on censored ranked set sample

In this section, we obtain some estimators of σ_2 using censored RSS scheme. Suppose k units are censored in the Stoke's RSS scheme, then we may represent the *rss* observations on the study variate Z as $\delta_1 Z_{[1:n]1}, \delta_2 Z_{[2:n]2}, \cdots, \delta_n Z_{[n:n]n}$ where,

$$\delta_i = \begin{cases} 0, & if \text{ the } i^{th} \text{ unit is censored,} \\ 1, & \text{otherwise.} \end{cases}$$

and hence $\sum_{i=1}^{n} \delta_i = n - k$. In this case the usual unbiased estimator of σ_2 is equal to $\frac{6\sum_{i=1}^{n} \delta_i Z_{[i:n]i}}{5(n-k)}$. It may be noted that one need not get $\delta_i = 0$ for $i = 1, 2, \cdots, k$ and $\delta_i = 1$ for $i = k+1, k+2, \cdots, n$. Hence if we write m_i, $i = 1, 2, \cdots, n-k$ as the integers such that $1 \leq m_1 < m_2 < \cdots m_{n-k}$ and for which δ_{m_i}=1, then,

$$E\left[\frac{6\sum_{i=1}^{n} \delta_i Z_{[i:n]i}}{5(n-k)}\right] = \sigma_2\left[1 - \frac{19\phi}{50(n+1)(n-k)}\sum_{i=1}^{n-k}(n - 2m_i + 1)\right].$$

Thus it is clear that the in the censored case the usual unbiased estimator is not an unbiased estimator of σ_2. However we can construct an unbiased estimator of σ_2 based on $\frac{6\sum_{i=1}^{n} \delta_i Z_{[i:n]i}}{5(n-k)}$ is given in the following theorem.

Theorem 3. Suppose that the random variable (W, Z) has a FGMBB distribution as defined in (3). Let $Z_{[m_i]m_i}$, $i = 1, 2, \cdots, n-k$ be the *rss* observations on the study variate Z resulting out of censoring applied on the auxiliary variable W. Then an unbiased estimator of σ_2 based on

$\frac{6}{5(n-k)}\sum_{i=1}^{n-k} Z_{[m_i]m_i}$ is given by

$$\sigma_2^*(k) = \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]}\sum_{i=1}^{n-k} Z_{[m_i]m_i}$$

and its variance is given by

$$Var[\sigma_2^*(k)] = \frac{3600(n+1)^2\sigma_2^2}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]^2}\sum_{i=1}^{n-k}\psi_{m_i}$$

where ψ_{m_i} is as defined in (18).

Proof We have

$$\begin{aligned}
E[\sigma_2^*(k)] &= \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]}\sum_{i=1}^{n-k} E[Z_{[m_i]m_i}] \\
&= \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]} \\
&\quad \times \sum_{i=1}^{n-k}\left[\frac{5}{6} - \frac{19}{60}\phi\frac{(n-2m_i+1)}{(n+1)}\right]\sigma_2 \\
&= \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]} \\
&\quad \times \left[\frac{5(n-k)}{6} - \frac{19\phi}{60(n+1)}\sum_{i=1}^{n-k}(n-2m_i+1)\right]\sigma_2 \\
&= \sigma_2.
\end{aligned}$$

Thus $\sigma_2^*(k)$ is an unbiased estimator of σ_2. The variance of $\sigma_2^*(k)$ is given by

$$\begin{aligned}
Var[\sigma_2^*(k)] &= \frac{3600(n+1)^2}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]^2}\sum_{i=1}^{n-k} Var(Z_{[m_i]m_i}) \\
&= \frac{3600(n+1)^2\sigma_2^2}{\left[50(n+1)(n-k) - 19\phi\sum_{r=1}^{n-k}(n-2m_i+1)\right]^2}\sum_{i=1}^{n-k}\psi_{m_i}
\end{aligned}$$

where ψ_{m_i} is as defined in (18). Hence the theorem.
As a competitor of the estimator $\sigma_2^*(k)$, next we propose the BLUE of σ_2 based on the censored *rss*, resulting out of ranking of observations on W.
If $\mathbf{Z}_{[n]}(k) = (Z_{[m_1]m_1}, Z_{[m_2]m_2}, \cdots, Z_{[m_{n-k}]m_{n-k}})'$, then the mean vector and the variance-covariance matrix of $\mathbf{Z}_{[n]}(k)$ are given by

$$E[\mathbf{Z}_{[n]}(k)] = \sigma_2\zeta(k), \tag{29}$$

$$D[\mathbf{Z}_{[n]}(k)] = \sigma_2 G(k), \tag{30}$$

where $\zeta(k) = (\zeta_{m_1}, \zeta_{m_1}, \cdots, \zeta_{m_{n-k}})'$, $G(k) = diag(\psi_{m_1}, \psi_{m_2}, \cdots, \psi_{m_{n-k}})$.
if the parameter ϕ involved in $\zeta(k)$ and $G(k)$ are known then (29) and (30) together defines a generalized Gauss-Markov setup and hence the BLUE $\tilde{\sigma}_2(k)$ of σ_2 is obtained as,

$$\tilde{\sigma}_2(k) = [(\zeta(k))'(G(k))^{-1}\zeta(k)]^{-1}(\zeta(k))'(G(k))^{-1}\mathbf{Z}_{[n]}(k) \tag{31}$$

and the variance of σ_2 is given by

$$Var(\tilde{\sigma}_2(k)) = [(\zeta(k))'(G(k))^{-1}\zeta(k)]^{-1}\sigma_2^2. \tag{32}$$

On substituting the values of $\zeta(k)$ and $G(k)$ in (31) and (32) and simplifying we get,

$$\tilde{\sigma}_2(k) = \frac{\sum_{i=1}^{n-k}(\zeta_{m_i}/\psi_{m_i})}{\sum_{i=1}^{n-k}(\zeta_{m_i}^2/\psi_{m_i})} Z_{[m_i]m_i} \tag{33}$$

and

$$Var(\tilde{\sigma}_2(k)) = \frac{1}{\sum_{i=1}^{n-k}(\zeta_{m_i}^2/\psi_{m_i})}\sigma_2^2. \tag{34}$$

Remark 4.1. Since both the BLUE $\tilde{\sigma}_2(k)$ and the unbiased estimator $\sigma_2^*(k)$ based on the censored ranked set sample utilize the distributional property of the parent distribution they lose the usual robustness property. Hence in this case the BLUE $\tilde{\sigma}_2(k)$ shall be considered as a more preferable estimator than $\sigma_2^*(k)$.

References

[1] Abd-Elrahman, A. M. (2013). Utilizing ordered statistics in lifetime distri- butions production: A new lifetime distribution and applications. *Journal of Probability and Statistical Science*, 11:153–164.

[2] Abd-Elrahman, A. M. and Niazi, S. F. (2016). Approximate Bayes estimators applied to the Bilal model. *Journal of the Egyptian Mathematical Society*, 1–6.

[3] Abd-Elrahman, A. M. (2017). A new two-parameter lifetime distribution with decreasing, increasing or upside-down bathtub-shaped failure rate. *Communications in Statistics-Theory and Methods*, 46:8865–8880.

[4] Al-Omari, A. I. (2021). Maximum likelihood estimation in location-scale families using varied L ranked set sampling. *RAIRO Operations Research*, 55:2759–2771.

[5] Al-Omari, A. I. and Abdallah, M. S. (2021). Estimation of the distribution function using moving extreme and minimax ranked set sampling. *Communications in Statistics-Simulation and Computation*.

[6] Al-Omari, A. I. and Almanjahie, I. M. (2021). New improved ranked set sampling design with an application to real data. *Computers, Materials and Continua*.

[7] Al-Saleh, M. F. (2004). Steady-state ranked set sampling and parametric estimation. *Journal of Statistical planning and Inference*, 123:83–95.

[8] Al-Saleh, M. F. and Al-Omari, A. I. (2002). Multistage ranked set sampling. *Journal of Statistical planning and Inference*, 273–286.

[9] Al-Saleh, M. F. and Al-Kadiri, M. (2000). Double-ranked set sampling. *Statistics and Probability Letters*, 205–212.

[10] Barnett, V. and Moore, K. (1997). Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering. *Communications in Statistics-Theory and Methods*, 24:697–710.

[11] Chen, Z., Bai, Z. and Sinha, B. K. (2004). Lecture Notes in Statistics, Ranked Set Sampling: Theory and Applications, Theory and Algorithms, Springer, New York.

[12] David, H. A. and Nagaraja, H. N. (2003). Order statistics: Third edition., John Wiley and Sons, New York.

[13] Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its applications. *Mathematics and Computers in Simulation*, 78:493–506.

[14] Hinkley, D. (1977). On quick choice of power transformation. *Applied Statistics*, 67–69.

[15] Irshad, M. R., Maya, R. and Arun S. P. (2019). Estimation of a parameter of Morgenstern type bivariate Lindley distribution by ranked set sampling. *iSTATISTIK: Journal of the Turkish Statistical Association*, 12:25–34.

[16] Koshti, R. D. and Kamalja, K. K. (2020). Parameter estimation of Cambanis-type bivariate uniform distribution with ranked set sampling. *Journal of Applied Statistics*.

[17] Maya, R., Irshad, M. R. and Arun, S. P. (2021). Farlie-Gumbel-Morgenstern bivariate Bilal distribution and its inferential aspects using concomitants of order statistics. *Journal of Probability and Statistical Science*, 19:1–20.

[18] McGhilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47:461–466.

[19] McIntyre, G. A. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3:385–390.

[20] Philip, A. and Thomas, P. Y. (2017). On concomitants of order statistics and its application in defining ranked set sampling from Farlie-Gumbel Morgenstern bivarite Lomax distribution. *Journal of the Iranian Statistical Society*, 16:67–95.

[21] Scaria, J. and Nair, N. U. (1999). On concomitants of order statistics from Morgenstern family. *Biometrical Journal*, 41:483–489.

[22] Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics-Theory and Methods*, 6:1207–1211.

COST-REVENUE ANALYSIS AND ANFIS COMPUTING OF HETEROGENEOUS QUEUING MODEL WITH A SECOND OPTIONAL SERVICE WITH FEEDBACK UNDER HYBRID VACATION

Divya K[1] and Indhira K*

Department of Mathematics, School of Advanced Sciences,
Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.
divya.k2020@vitstudent.ac.in, kindhira@vit.ac.in.

Abstract

This research article examines an $M/M/2$ heterogeneous queueing model that provides two services: a mandatory first essential service (FES) and an optional second optional service (SOS). The model incorporates breakdown, feedback, and a hybrid vacation policy. Matrix expressions are structured to evaluate the stationary probability distribution of the number of customers in the system and system performance measures using the matrix-geometric approach (MGA). Additionally, formulas are being developed to estimate the model's performance indicators. The cost function is being evaluated to determine the best values of the system's decision variables, and an adaptive neural fuzzy inference system (ANFIS) based on soft computing technology is being utilized to validate the obtained results.

Keywords: Markovian queue, Breakdown, Hybrid vacation, Matrix geometric approach, ANFIS.

1. Introduction

Queuing theory is the study of how people behave in service systems like telephone systems and waiting lines. It is a branch of operations research that focuses on analysing the arrival and departure of packets or customers from a service system. Queuing theory is used to analyse a wide range of scenarios, such as traffic control, banking, manufacturing, computer and telecommunications networks, production and transportation systems, and even healthcare systems. In general, queuing theory involves finding an appropriate mathematical model to describe the system and then analysing it to determine how the system performs and how it can be improved. Morse [17] was the first to consider the concept of heterogeneous servers in multi-server queueing models, which is more realistic than assuming all servers provide service at an equal rate. In queuing systems with human servers, this assumption is impossible to implement as different servers can provide services at varying rates. It's obvious that heterogeneous services are essential to the functioning of almost every industrial system. Li and Stanford [9] explored heterogeneous multi-server accumulating priority queues. Krishnamoorthy et al. [14] presented a two heterogeneous servers queueing model. Chang et al. [5] investigate an unreliable-server retrial queue with customer feedback and impatience. They analyse the performance of the system through simulations and derive analytical expressions for its measures of performance and cost. Recently, Wu and Yang [25] examined a two-phase heterogeneous service model. They initially performed a single objective optimization using Canonical PSO then developed a bi-objective cost optimization model for the system and

waiting time in tandem.

In queueing theory, vacation refers to a period of time in which a server is unavailable to serve customers. There are several types of vacations, including planned and random, as well as hybrid vacations which combine the two. Planned vacations occur at predetermined intervals, while random vacations occur randomly throughout the queueing system's operation. Hybrid vacations combine both planned and random vacations in order to optimize the system's performance. Servi and Finn [19] work in queuing theory focused on working vacations (WV) for $M/M/1$ queues. In this approach, the server can continue servicing requests at a slower rate when customers are not present. Bouchentouf et al. [3] researched a multi-station unstable machine model with customer impatience and a working vacation schedule. Ziad et al. [29] examined a $M/M/c$ queueing system with waiting servers, balking, reneging, and a K-variant working vacation interrupted by a Bernoulli schedule. Bouchentouf et al. [4] examine the performance and economics of a single server queueing model with feedback, impatient customers, and a vacation policy, finding an optimal control policy for vacation times, and providing their results, which can be used to make decisions regarding the system's parameters. Anshul Kumar et al. [16] investigated the hybrid holiday policy and a two-stage service procedure using matrix geometric techniques. This hybrid holiday is a combo of working vacation and complete vacation (CV), in which the server may begin in WV and continue to give service at a decreased rate when the server is idle. If there are clients present in WV, the server will linger in WV and offer service. If not, it will go to CV. The server will revert to its usual operations after the CV has been completed, and it will begin providing service to any customers who are ready at that time. Dual servers with varying service rates are a logical outgrowth of this concept.

The term "essential service" refers to services that servers must offer in order to satisfy customer needs. Optional services, on the other hand, are those that are provided based on customer demand. Essential services are usually core services that must be provided in order for the server to fulfil its role, while optional services are additional services that customers may choose to opt-in to receive. Laxmi and Jyothsna [22] examined a finite buffer impatient customer queue with WV, where the server offers two phases of service: essential and optional. Yang et al. [26] used SOS to an M/M/R queueing model and offered economic analysis. Several queueing models for optional services have been studied by researchers such as Anitha et al. [2], Chandrika and Kalaiselvi [6], Li and Wang [10], Yang and Chen [27], and others.

Realistically, it's not possible to have a completely reliable server because it could break at any time. Repair setups such as thresholds, backups, or restarts must be applied to restore service. Significant research on breakdown provides a model to investigate how such changes may reduce customer wait times in real-time service systems with a total failure server, such as those found in banks, manufacturing plants, contact centres, and so on.

Ye and Liu [28] applied the MGA to discover the steady-state solutions for a Markovian arrival single server queuing system with an operational breakdown. Vijayalakshmi et al. [23] used a matrix technique to examine the restricted capacity of a Markovian queueing model with working interruptions and two-phase service. The $M/M/1$ model with working breakdowns and recovery policies based on k-threshold recovery time and setup recovery was studied by Ezeagu et al. [8]. A queueing model for a service system with a secondary server was presented by Chakravarthy et al.

The Matrix Geometric Approch was first developed in the 1960s and has since become a fundamental tool in queueing theory. It was widely used in Markov chain models and queuing systems to solve difficult real-time problems. In particular, this method helps to determine the stationary behaviour of a queueing model by calculating the expected number of customers and expected times in the system. M.F. Neuts [18] was a key figure in the development of the Matrix Geometric Approach (MGA). Ke et al. [13] conducted research on an $M/M/R$ queueing system with SOS. Shekhar et al. [21] employed metaheuristics to find effective emergency

vacation queueing techniques. Anshul and Madhu Jain [16] conducted an investigation into an unreliable server, they studied the effects of an MGA-based Markovian queueing model for a two-stage service system that utilises a hybrid vacation policy.

ANFIS, or Adaptive Neuro-Fuzzy Inference System, is an intelligent system that combines neural networks and fuzzy logic to model complex non-linear systems. It can be used for many applications, such as classification, prediction, control, and optimization. ANFIS can also be used for transient analysis, where it can generate more accurate insights about system dynamics. Jang [11] proposed the ANFIS as a tool to model complex systems. He demonstrated the application of ANFIS in modelling the nonlinear dynamics of a continuous-time system and provided references on the uses of ANFIS in various research fields. The article was an important milestone in furthering the application of ANFIS in the sciences. The content of Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence was analysed by Jang et al. [12]. ANFIS has been used in many research papers, including Ahuja et al. study of the transient of an unreliable single-server queueing model with multi-stage service and a working vacation. Sethi et al. [20] proposed a mathematical model to analyse the system and demonstrated that the parameters of the system can be optimised using an ANFIS approach. They also discussed the performance of different policies in terms of cost and found that an N-policy was the most efficient for cost optimization. An analysis of an $M^X/G/1$ retrial queue with impatient customers, an unreliable server, a modified vacation policy, delayed repair, and a Bernoulli feedback system was presented by Upadhyaya and Kushwaha [24]. They also used an ANFIS computing approach to compare their numerical results to those obtained from explicit analytical formulas.

The structure of this paper is outlined in the following way: Section 2 outlines the model description and associated mathematical assumptions, as well as provides an explanation for the transition rate matrix. In Section 3, the stability condition is established and the matrix-geometric approach is explained. Performance measures such as the expected number of customers in the system in terms of system state probabilities and the total cost accumulated through different activities and cost elements are derived in Section 4. Section 5 provides numerical results, while the conclusions in Section 6 illustrate noteworthy features and potential areas for future research.

2. Model description

Under the hybrid vacation policy, we propose a heterogeneous Markovian queueing model with a second optional service with feedback and breakdown. This model is detailed as follows:

The arrival pattern: In this model, customers arrive according to a Poisson process with an arrival rate of λ.

The service pattern: The first-in-first-out policy directs the service discipline.We investigate a queueing model with two heterogeneous servers: the first server is constantly available and totally reliable, charges a service rate of μ_0. while the second server has two distinct phases - an first essential phase (FES) and a second optional phase (SOS) - and is only occasionally accessible and unstable. server 2 charges μ_1 for FES and μ_2 for SOS, with the service rates for both following an exponential distribution.

Breakdown and repair rule : The queueing system has two servers, with the first server consisting of a single, reliable phase, and the second server consisting of two phases: FES and SOS. If both phases of the second server experience breakdowns with rates η_1, η_2 and, it is immediately sent for repair with rates θ_1 and θ_2, respectively.

Feedback rule: After receiving a service, an unsatisfied customer can decide to re-enter the system for another service, with a probability $\bar{\kappa}$ referred to as "feedback", or they can choose to permanently exit the system with a probability κ (=1-$\bar{\kappa}$). Feedback service is consider as a new arival λ.

The hybrid vacation strategy: During vacation periods, the server starts WV, which has an exponential distribution with a slower service rate of μ_v. When the server gets empty during WV, it will switch to CV mode. While customers enter the system at CV, the server will return to its usual busy state and start serving customers. CV duration has an exponential distribution with a mean θ_v. The server enters the working period in the CV state with the probability ξ of providing service to the customer. which is always 1.

Let us consider $F(t) = \{(\mathfrak{W}(t), \mathfrak{J}(t)); t \geq 0\}$be the bivariate Markov process (BMP) with a

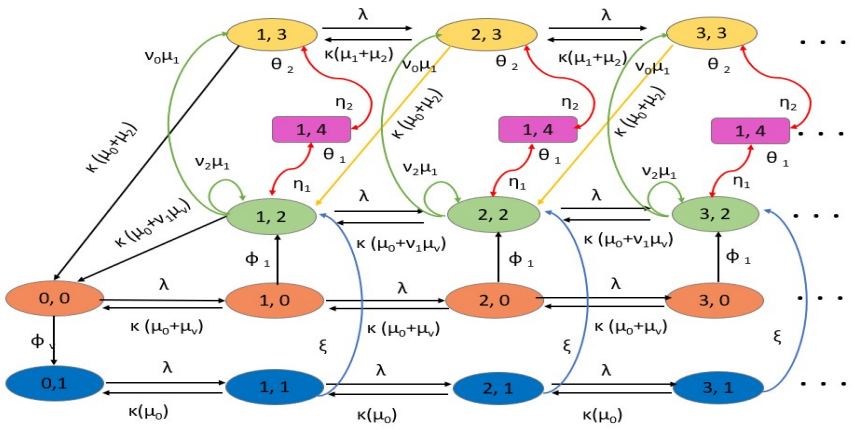


Figure 1: *State transition diagram of the model*

state space at time t. where $\mathfrak{W}(t)$ represents the number of customers in the system, and

$$\mathfrak{J}(t) = \begin{cases} 0, & \text{if server 2 is in } \textit{Working vacation} \\ 1, & \text{if server 2 is in } \textit{Complete vacation} \\ 2, & \text{if server 2 is in } \textit{FES} \\ 3, & \text{if server 2 is in } \textit{SOS} \\ 4, & \text{if server 2 is in } \textit{Breakdown} \end{cases}$$

All stochastic processes in the system are independent of one another. The structure of this model's transition diagram is depicted in the below figure. 1.

2.1. Governing equations

By using birth- death process, Governing equations can be formulated as follows:

$$\begin{aligned}
(\lambda + \phi_v)\iota_{0,0} &= (\mu_0 + \mu_v)\iota_{1,0} + \kappa\mu_0\iota_{0,1} + \kappa(\mu_0 + \nu_1\mu_2)\iota_{1,2} + \kappa(\mu_0 + \mu_2)\iota_{1,3} \\
[\kappa(\mu_0 + \mu_v) + \lambda + \phi_1]\iota_{1,0} &= \lambda\iota_{0,0} + \kappa(\mu_0 + \mu_v)\iota_{2,0} \\
[\kappa(\mu_0 + \mu_v) + \lambda + \phi_1]\iota_{m,0} &= \lambda\iota_{m-1,0} + \kappa(\mu_0 + \mu_v)\iota_{m+1,0} \\
(\kappa\mu_0 + \lambda)\iota_{0,1} &= \phi_v\iota_{0,0} + \kappa\mu_0\iota_{1,1} \\
(\kappa\mu_0 + \lambda)\iota_{1,1} &= \phi_v\iota_{0,1} + \kappa\mu_0\iota_{2,1} \\
(\kappa\mu_0 + \lambda)\iota_{m,1} &= \phi_v\iota_{m-1,1} + \kappa\mu_0\iota_{m+1,1} \\
[\lambda + \kappa(\mu_0 + \nu_1\mu_1) + \nu_0\mu_1 + \eta_1]\iota_{1,2} &= \xi\iota_{1,1} + \phi_1\iota_{0,1} + \nu_2\mu_1\iota_{1,2} + \theta_1\iota_{1,4} + \kappa(\mu_0 + \mu_1)\iota_{2,2} + \kappa(\mu_0 + \mu_2)\iota_{2,3} \\
[\lambda + \kappa(\mu_0 + \nu_1\mu_1) + \nu_0\mu_1 + \eta_1]\iota_{m,2} &= \xi\iota_{m,1} + \phi_1\iota_{m-1,1} + \nu_2\mu_1\iota_{m,2} + \theta_1\iota_{m,4} + \kappa(\mu_0 + \mu_1)\iota_{m+1,2} \\
&\quad + \kappa(\mu_0 + \mu_2)\iota_{m+1,3} \\
[\kappa(\mu_0 + \mu_2) + \lambda + \eta_2]\iota_{1,3} &= \nu_0\mu_1\iota_{1,2} + \theta_2\iota_{1,4} \\
[\kappa(\mu_0 + \mu_2) + \lambda + \eta_2]\iota_{m,3} &= \nu_0\mu_1\iota_{m,2} + \theta_2\iota_{m,4}
\end{aligned}$$

$$[\theta_1+\theta_2]\iota_{1,4} = \eta_1\iota_{1,2} + \eta_2\iota_{1,3}$$
$$[\theta_1+\theta_2]\iota_{m,4} = \eta_1\iota_{m,2} + \eta_2\iota_{m,3}$$

To make obtaining solutions to our model easier and faster, we are employing the MGA. This method is used to obtain steady-state probabilities when the state-space increases rapidly. This technique helps us to achieve effective and numerically stable solutions that would otherwise be difficult and time-consuming to obtain.

2.2. Matrix Geometric Solution

The system state is symbolized by $\mathfrak{W}(t)$ and $\mathfrak{J}(t)$. Let $\{(\mathfrak{W}(t), \mathfrak{J}(t)); t \geq 0\}$, with the state space organised in lexicographical manner as follows.

$$\mathrm{Y} = (0,0)\bigcup(0,1)\bigcup(m,n); n \geq 1, m = 0,1,2,3,4$$

The set of equations in section 2.1 are utilised to create the model's steady-state probability using the matrix-geometric approach. The block tridiagonal pattern is represented by the associated infinitesimal generator matrix $\mathbb{G}$ of this Markov chain, which is expressed as follows:

$$\mathbb{G} = \begin{bmatrix} S_0 & T_0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ V_0 & U_1 & U_0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & U_2 & U_1 & U_0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & U_2 & U_1 & U_0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & U_2 & U_1 & U_0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \cdots \end{bmatrix}$$

where

$$S_0 = \begin{bmatrix} -(\lambda+\phi_v) & \phi_v \\ \beta & -(\lambda+\beta) \end{bmatrix}; \quad T_0 = \begin{bmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \end{bmatrix}; \quad V_0 = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \\ \gamma & 0 \\ \delta & 0 \\ 0 & 0 \end{bmatrix}$$

$$U_0 = \begin{bmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}; \quad U_2 = \begin{bmatrix} \alpha & 0 & 0 & 0 & 0 \\ 0 & \beta & 0 & 0 & 0 \\ 0 & 0 & \gamma & 0 & 0 \\ 0 & 0 & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and $$U_1 = \begin{bmatrix} -[\lambda+\phi_1+\alpha] & 0 & \phi_1 & 0 & 0 \\ 0 & -[\lambda+\xi+\beta] & \xi & 0 & 0 \\ 0 & 0 & -[\lambda+\eta_1+\nu_1\mu_1+\gamma-\nu_2\mu_1] & \nu_0\mu_1 & \nu_1 \\ 0 & 0 & 0 & -[\lambda+\eta_2+\delta] & \eta_2 \\ 0 & 0 & \theta_1 & \theta_2 & -[\theta_1+\theta_2] \end{bmatrix}$$

Here, $\alpha = \kappa[\mu_0+\mu_v]$, $\beta = \kappa\mu_0$, $\gamma = \kappa[\mu_0+\nu_1\mu_1]$, $\delta = \kappa[\mu_0+\mu_2]$.

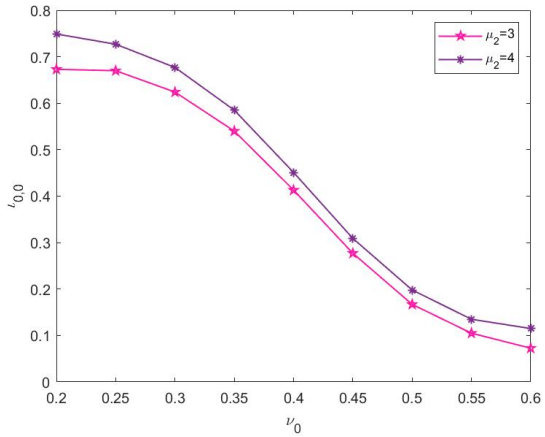


Figure 2: *Influence of ν_0 on the idle state $\iota_{0,0}$ of the servers*

3. Stability Condition

Theorem 1. The system is stable if and only if $\rho = \frac{\lambda}{\beta}\left(\frac{\mathbb{A}\theta_1+\mathbb{B}\theta_2}{\mathbb{A}\theta_1+\mathbb{B}\theta_2+\mathbb{C}}\right) < 1$

Proof. Let us define the matrix $\mathrm{U} = U_0 + U_1 + U_2$ given by

$$\mathrm{U} = \begin{bmatrix} -\phi_1 & 0 & \phi_1 & 0 & 0 \\ 0 & -\xi & \xi & 0 & 0 \\ 0 & 0 & -[\eta_1+(\nu_0-\nu_2)\mu_1] & \nu_0\mu_1 & \eta_1 \\ 0 & 0 & -\delta & \delta & 0 \\ 0 & 0 & \theta_1 & \theta_2 & -[\theta_1+\theta_2] \end{bmatrix} \tag{1}$$

There exists a stationary probability $\mathcal{I} = (\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4)$ of U such that

$$\mathcal{I}\mathrm{U} = 0; \quad \mathcal{I}\mathfrak{e} = 1 \tag{2}$$

where $\mathfrak{e} = [1,1,1,1,1]^T$. Using theorem 3.1.1 of Netus [18], the necessary and sufficient condition for the stability of the system is as follows:

$$\mathcal{I}U_0\mathfrak{e} < \mathcal{I}U_2\mathfrak{e} \tag{3}$$

Solving 1 and 2, we get

$$\lambda[\mathcal{I}_0 + \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4] < \alpha\mathcal{I}_0 + \beta\mathcal{I}_1 + \gamma\mathcal{I}_2 + \delta\mathcal{I}_3 \tag{4}$$

$$\frac{\lambda}{\sigma_1}[(\nu_2+\nu_0\mu_1+\delta))\theta_1 + (\delta+\eta_1+\mu_1(\nu_0+\nu_1-\nu_2))\theta_2] < \frac{\beta}{\sigma_1}[(\kappa\mu_1\nu_1+\delta+\nu_0\mu_1)\theta_1 + \kappa\mu_2 + (\delta+\nu_1+\mu_1(\nu_0+\nu_1-\nu_2)\theta_2)] \tag{5}$$

$$\frac{\lambda}{\sigma_1}[\mathbb{A}\theta_1 + \mathbb{B}\theta_2] < \frac{\beta}{\sigma_1}[\mathbb{A}\theta_1 + \mathbb{B}\theta_2 + \mathbb{C}] \tag{6}$$

$$\frac{\lambda}{\beta}\left[\frac{\mathbb{A}\theta_1 + \mathbb{B}\theta_2}{\mathbb{A}\theta_1 + \mathbb{B}\theta_2 + \mathbb{C}}\right] < 1 \tag{7}$$

Here

$$\begin{aligned} \mathbb{A} &= \nu_2 + \nu_0\mu_1 + \delta \\ \mathbb{B} &= \delta + \eta_1 + \mu_1(\nu_0 + \nu_1 - \nu_2) \\ \mathbb{C} &= \kappa(\mu_1\nu_1 + \mu_2) \end{aligned}$$

Table 1: *Impact of arrival rate λ on perfomance measures*

λ	$\mathrm{E}[L_s]$	$\mathrm{E}[L_q]$	P_{Idle}	$P_{WV}^{S_2}$	$P_{CV}^{S_2}$	$P_{FES}^{S_2}$	$P_{SOS}^{S_2}$	$P_{Bd}^{S_2}$
0.5	0.1862	0.0271	0.8393	0.9450	0.0365	0.0121	0.0032	0.0016
0.6	0.2378	0.0447	0.8069	0.9409	0.0358	0.0160	0.0046	0.0023
0.7	0.2904	0.0647	0.7744	0.9359	0.0349	0.0198	0.0062	0.0029
0.8	0.3480	0.0891	0.7416	0.9297	0.0347	0.0240	0.0081	0.0035
1	0.4780	0.1495	0.6757	0.9149	0.0321	0.0339	0.0134	0.0052
1.1	0.5539	0.1800	0.6423	0.9053	0.0310	0.0396	0.0172	0.0063
1.2	0.6411	0.2488	0.6085	0.8938	0.0299	0.0460	0.0218	0.0077
1.3	0.7384	0.3073	0.5757	0.8936	0.0185	0.0517	0.0266	0.0089
1.4	0.8442	0.3851	0.5398	0.8648	0.0274	0.0614	0.0344	0.0109
1.5	0.9796	0.4725	0.5044	0.8453	0.0261	0.0706	0.0431	0.0130

Table 2: *Impact of service rates $(\mu_0, \mu_1, \mu_2, \mu_v)$ on some perfomance measures*

$(\mu_0, \mu_1, \mu_2, \mu_v)$	$\mathrm{E}[L_s]$	$\mathrm{E}[L_q]$	P_{Idle}	$P_{WV}^{S_2}$	$P_{CV}^{S_2}$	$P_{FES}^{S_2}$	$P_{SOS}^{S_2}$	$P_{Bd}^{S_2}$
(**3**,2,1.5,1)	1.2737	0.6971	0.4440	0.8191	0.0236	0.0783	0.0611	0.0163
(**4**,2,1.5,1)	0.8209	0.3774	0.5652	0.9063	0.0228	0.0424	0.0208	0.0071
(**5,2.5**,1.5,1)	0.5589	0.2008	0.6414	0.9409	0.0204	0.0247	0.0098	0.0037
(**6,3,**1.5,1)	0.4409	0.1351	0.6941	0.9580	0.0181	0.0161	0.0054	0.0023
(**7,3.5,2**,1)	0.2693	0.0713	0.7334	0.9007	0.0161	0.0103	0.0028	0.0015
(**7.5,4,2.5**,1)	0.3323	0.0822	0.7495	0.9716	0.0151	0.0093	0.0024	0.0012
(8,4,3,**1.5**)	0.2884	0.0641	0.7752	0.9746	0.0146	0.0077	0.0017	0.0009

where

$$\mathcal{I}_0 = 0, \quad \mathcal{I}_1 = 0, \quad \mathcal{I}_2 = \frac{[\nu_2 + \delta]\theta_1 + \delta\theta_2}{\sigma_1}, \quad \mathcal{I}_3 = \frac{[\eta_1 + \mu_1(\nu_0 + \nu_1 - \nu_2)\theta_2 + \mu_1\nu_0\theta_1]}{\sigma_1},$$
$$\mathcal{I}_4 = \frac{[\nu_2(\eta_1 - \mu_1\nu_2) + \mu_1\nu_2(\nu_0 + \nu_1)] + [\kappa\nu_1 + \kappa\mu_1(\nu_1 - \nu_2)](\mu_0 + \mu_2)}{\sigma_1} \tag{8}$$

Here,

$$\sigma_1 = \eta_1(\nu_1 + \theta_2) + \nu_2(\theta_1 - \mu_1\nu_2) + \eta_1\delta + \mu_1\nu_2(\nu_0 + \nu_1) + (\theta_1 + \theta_2)(\delta + \nu_0\mu_1) + \mu_1(\theta_2 + \delta)(\nu_1 - \nu_2)$$

Hence, System stability is ensured that ρ is equal to or less than 1. ■

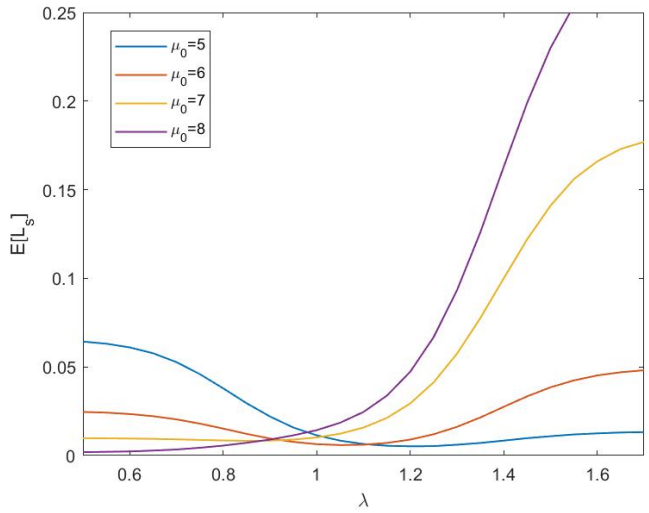


Figure 3: *Influence of* $\mathrm{E}[L_s]$ *for different* μ_0 *w.r.t* λ

3.1. Stationary probability distribution

We define $\mathfrak{P}_{mn} = \{(\mathfrak{W}(t) = m, \mathfrak{J}(t) = n)\}$ where m indicates the total number of customers in the queue and n reflects the server state. Under the stability condition $\rho < 1$. The prob. vector is described as follows: $\mathcal{I} = [\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3, ...]$, where $\mathcal{I}_0 = [i_{0,0}, i_{0,1}]$, $m = 0$; $\mathcal{I}_i = [i_{m,0}, i_{m,1}, i_{m,2}, i_{m,3}, i_{m,4}]$, $m = 1, 2, 3, 4, ...$ Since the steady-state criterion is achieved, then the

Table 3: *Impact of breakdown and repair rates* $(\eta_1, \theta_1, \eta_2, \theta_2)$ *on some perfomance measures*

$(\eta_1, \theta_1, \eta_2, \theta_2)$	$E[L_s]$	$E[L_q]$	P_{Idle}	$P_{WV}^{S_2}$	$P_{CV}^{S_2}$	$P_{FES}^{S_2}$	$P_{SOS}^{S_2}$	$P_{Bd}^{S_2}$
(**0.5**,2,0.5,3)	0.5071	0.8011	0.0227	0.1204	0.0397	0.0159	0.4323	0.0109
(**0.6**,2,**0.55**,3)	0.6796	0.9203	0.0322	0.0297	0.0128	0.0044	0.3679	0.0113
(**0.7**,2,**0.6**,3)	0.6808	0.9219	0.0323	0.0289	0.0124	0.0042	0.3685	0.0113
(**0.8,3,0.7,4**)	0.6744	0.9130	0.0321	0.0391	0.0120	0.0035	0.3665	0.0113
(0.8,**4,0.8,5**)	0.6825	0.9241	0.0323	0.0294	0.0110	0.0027	0.3718	0.0113
(0.8,**5,0.9**,5)	0.6824	0.9240	0.0323	0.0303	0.0103	0.0026	0.3718	0.0113

sub prob.vectors $\mathcal{I}_i$ satisfy the following equations:

$$\mathcal{I}_0 S_0 + \mathcal{I}_1 V_0 + 0 + \ldots = 0 \tag{9}$$
$$\mathcal{I}_0 T_0 + \mathcal{I}_1 U_1 + \mathcal{I}_2 U_2 + 0 + \ldots = 0 \tag{10}$$
$$\mathcal{I}_1 U_0 + \mathcal{I}_2 U_1 + \mathcal{I}_3 U_2 + 0 + \ldots = 0 \tag{11}$$
$$\mathcal{I}_2 U_0 + \mathcal{I}_3 U_1 + \mathcal{I}_4 U_2 + 0 + \ldots = 0 \tag{12}$$
$$\vdots$$
$$\mathcal{I}_i U_0 + \mathcal{I}_{i+1} U_1 + \mathcal{I}_{i+2} U_2 + \ldots = 0 \text{ where } i \geq 2 \tag{13}$$
$$\mathcal{I}_j = \mathcal{I}_1 \mathcal{R}^{j-1}, \text{ where } j \geq 2. \tag{14}$$

Let the matrix $\mathcal{R}$ represents the rate matrix. By substituting equation 12 into the equations 7 to 11, we get

$$\mathcal{I}_0 S_0 + \mathcal{I}_1 V_0 = 0 \tag{15}$$
$$\mathcal{I}_0 T_0 + \mathcal{I}_1 [U_1 + \mathcal{R} U_2] = 0 \tag{16}$$
$$\mathcal{I}_1 [U_0 + \mathcal{R} U_1 + \mathcal{R}^2 U_2] = 0 \tag{17}$$
$$\mathcal{I}_1 \mathcal{R} [U_0 + \mathcal{R} U_1 + R^2 U_2] = 0 \tag{18}$$
$$\mathcal{I}_1 \mathcal{R}^{i-1} [U_0 + \mathcal{R} U_1 + \mathcal{R}^2 U_2] = 0 \quad i \geq 2. \tag{19}$$

The normalizing equation can be expressed as

$$\mathcal{I}_0 \mathfrak{e} + \mathcal{I}_1 [I - \mathcal{R}]^{-1} \mathfrak{e} = 1 \tag{20}$$

Here $\mathfrak{e}$ is a column vector in which all elements are 1's in the corresponding column. By using methodologies from Neuts [18] and Latouche and Ramaswami [15]. we have estimated? the rate matrix $\mathcal{R}$. Thus, $\mathcal{R}$ is a minimal non-negative solution of the matrix quadratic equations.

$$U_0 + \mathcal{R} U_1 + \mathcal{R}^2 U_2 = 0 \tag{21}$$
$$\mathcal{R} = -U_0 U_1^{-1} - \mathcal{R}^2 U_2 U_1^{-1} \tag{22}$$

Where $\mathcal{R} \geq 0$ and it's an irreducible non-negative matrix of spectral radius smaller than one[18]. Matrix $\mathcal{R}$ can be calculated using an iterative approach as shown below.

$$\mathcal{R}_0 = 0 \tag{23}$$
$$\mathcal{R}_{n+1} = -U_0 U_1^{-1} - \mathcal{R}_k^2 U_2 U_1^{-1} \qquad k \geq 1 \tag{24}$$

All values of $\mathcal{R}$ will expand monotonically, and non-negative matrix $\mathcal{R}$ is converging to $-U_1^{-1}$ and $[U_0 + \mathcal{R}^2 U_2]$. The steady state is attained via the MGM.

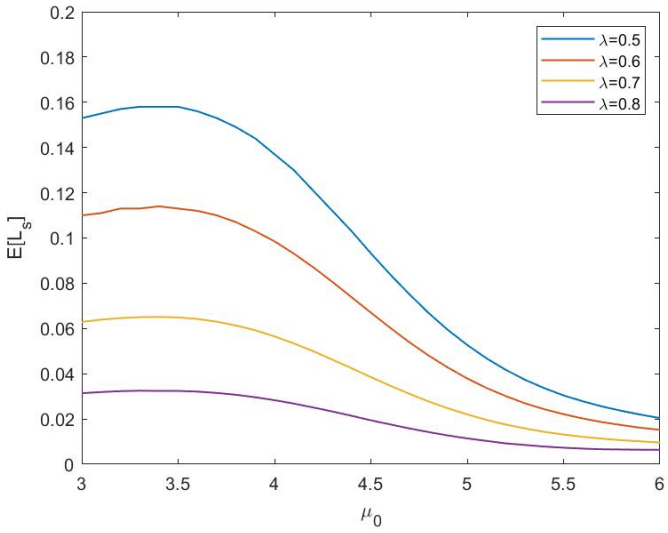


Figure 4: *Influence of* $\mathrm{E}[L_s]$ *for different* λ *w.r.t* μ_0

4. Measuring the Performance Characteristics

The following are some performance measurements for the queueing model under consideration in terms of steady-state probability.

4.1. System State Probabilities

- The probability that the servers are idle: $P_{Idle} = \iota_{0,0}$
- The probability of server 2 is in working vacation: $P_{WV}^{S_2} = \sum_{m=1}^{\infty} \iota_{m,0}$
- The probability of server 2 is in complete vacation: $P_{CV}^{S_2} = \sum_{m=1}^{\infty} \iota_{m,1}$
- The probability of server 2 is in FES: $P_{FES}^{S_2} = \sum_{m=1}^{\infty} \iota_{m,2}$
- The probability of server 2 is in SOS: $P_{SOS}^{S_2} = \sum_{m=1}^{\infty} \iota_{m,3}$
- The probability of server 2 is in breakdown: $P_{Bd}^{S_2} = \sum_{m=1}^{\infty} \iota_{m,4}$

4.2. Expected Numbers of Customers in the System and Queue

- The expected no.of customers in the system and queue:

$$\mathrm{E}[L_s] = \sum_{m=1}^{\infty} m\iota_{m,0} + \sum_{m=1}^{\infty} m\iota_{m,1} + \sum_{m=1}^{\infty} m\iota_{m,2} + \sum_{m=1}^{\infty} m\iota_{m,3} + \sum_{m=1}^{\infty} m\iota_{m,4}$$

$$\mathrm{E}[L_q] = \sum_{m=1}^{\infty} (m-1)\iota_{m,0} + \sum_{m=1}^{\infty} (m-1)\iota_{m,1} + \sum_{m=1}^{\infty} (m-1)\iota_{m,2} + \sum_{m=1}^{\infty} (m-1)\iota_{m,3} + \sum_{m=1}^{\infty} (m-1)\iota_{m,4}$$

- The expected no.of customers served is calculated by:

$$\mathrm{E}[SC] = \sum_{m=1}^{\infty} [\mu_0\iota_{m,0} + (\mu_0 + \mu_v)\mu_0\iota_{m,1} + (\mu_0 + \mu_1)\mu_0\iota_{m,2} + (\mu_0 + \mu_2)\mu_0\iota_{m,3}]$$

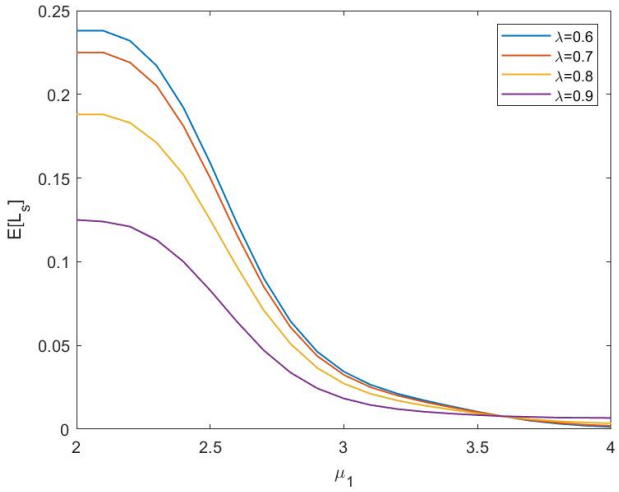


Figure 5: *Influence of* $\mathrm{E}[L_s]$ *for different* μ_0 *w.r.t* λ

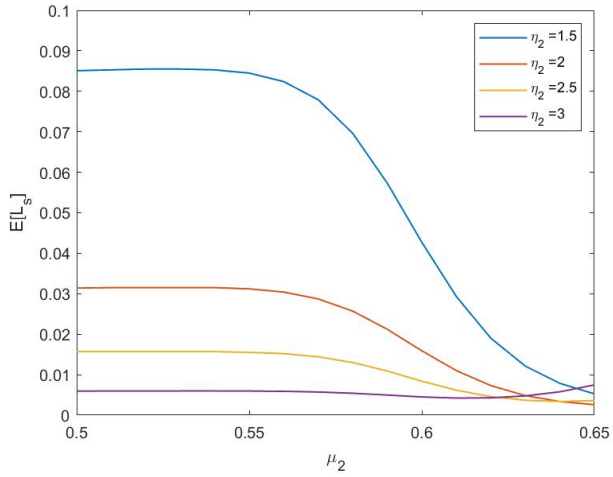


Figure 6: *Influence of* $\mathrm{E}[L_s]$ *for different* μ_0 *w.r.t* λ

4.3. Estimating the Cost and Revenue

Cost and revenue analysis play an important role in queuing systems as they provide an economic interpretation that can be applied to various technical and industrial situations. We define the total expected cost function per unit time and incorporate service rates as selection factors in order to find the optimal service rates that minimize the total cost f unction. The following factors are incorporated into our prediction:

C_{ls} = Each customer's holding cost per unit time in the system.
C_v = Cost per customer served in the vacation mode of the server 2.
C_F = Cost per customer served in the FES mode of the server 2.
C_S = Cost per customer served in the SOS mode of the server 2.
C_η = Cost per customer incurred when a broken down server is under repair.
C_{μ_0} = Cost per customer served in the busy mode of the server 1.
C_{μ_1} = Cost per customer served in the FES mode of the server 2.
C_{μ_2} = Cost per customer served in the SOS mode of the server 2.

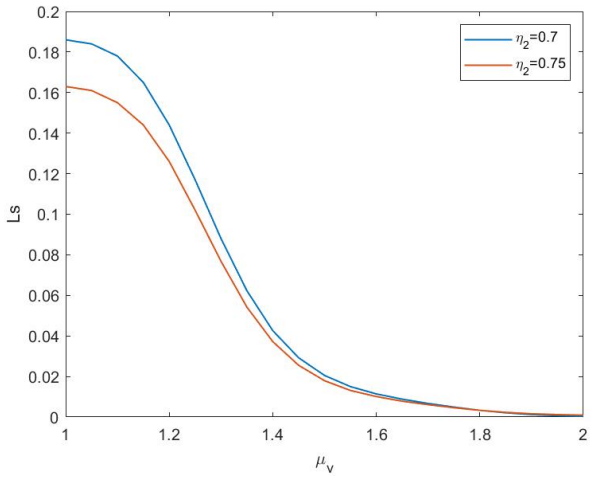
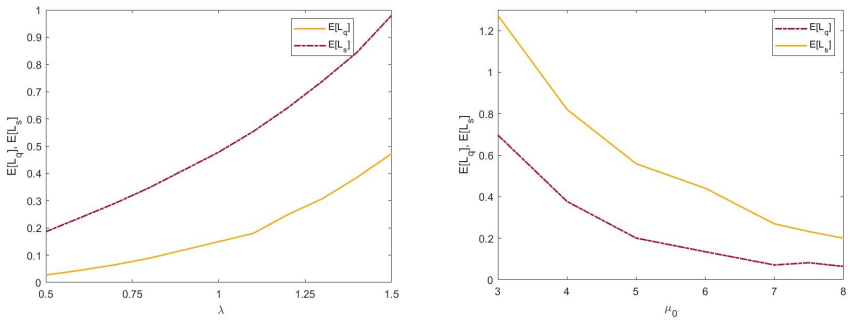
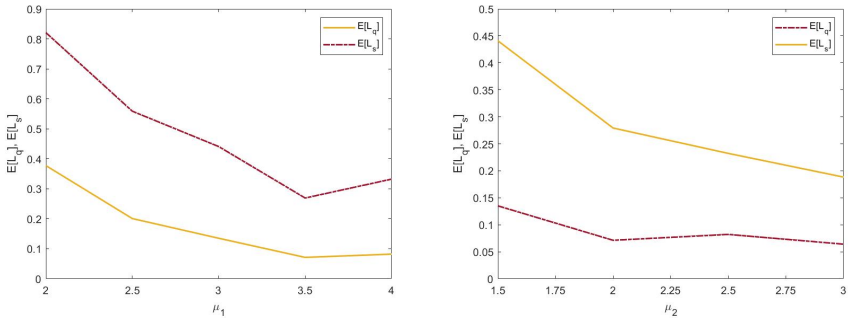


Figure 7: *Influence of* $\mathrm{E}[L_s]$ *for different* μ_0 *w.r.t* λ



(a) λ vs $E[L_q]$, $E[L_s]$ (b) μ_0 vs $E[L_q]$, $E[L_s]$



(c) μ_1 vs $E[L_q]$, $E[L_s]$ (d) μ_2 vs $E[L_q]$, $E[L_s]$

Figure 8: *Influence of a few parameters on 2D representation*

Total projected cost (TPC) is defined as:

$$TPC = C_{ls}L_s + C_v(P_{WV}^{S_2} + P_{CV}^{S_2}) + C_F P_{FES}^{S_2} + C_S P_{SOS}^{S_2} + C_\eta P_{Bd}^{S_2} + C_{\mu_0}\mu_0 + C_{\mu_1}\mu_1 + C_{\mu_2}\mu_2$$

If Rev represents customer service revenue, then the system's total anticipated revenue (TAR) is given by:

$$TAR = Rev * E[SC]$$

The total profit is given as:

$$T_{profit} = TAR - TPC$$

The complexity and non-linearity of the cost and revenue functions make it difficult to analyse the behaviour of the cost-revenue model and identify the most suitable values. All computations have been rounded off to two decimal places.

Table 4: *Cost set values for various cost aspects*

Cost set	C_{ls}	C_{wv}	C_{cv}	C_F	C_S	C_{Bd}	C_{μ_0}	C_{μ_1}	C_{μ_2}
I	45	20	10	30	25	15	20	15	10
II	40	20	10	30	25	15	20	15	10
III	45	25	10	30	25	15	20	15	10

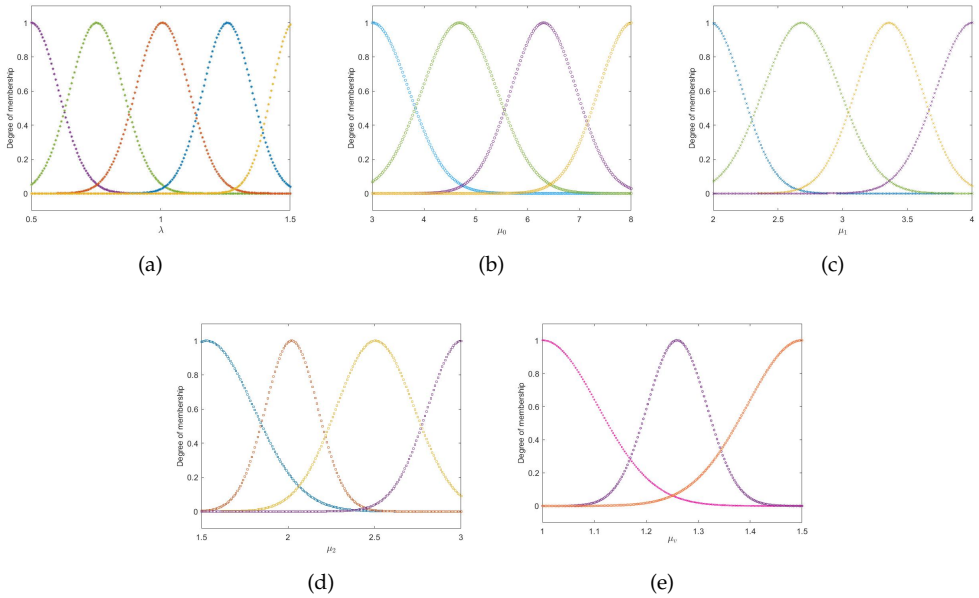


Figure 9: *ANFIS MF for (a) λ, (b) μ_0, (c) μ_1, (d) μ_2 (e) μ_v input variables*

4.4. Adaptive neuro-fuzzy inference system (ANFIS)

Jyh-Shing Roger Jang [11] was the person who first proposed ANFIS in 1992. The system combines both fuzzy logic and neural networks to capture much of the uncertainty and inexactness of real-life systems. In its simplest form, ANFIS consists of a number of inference rules that are used to make decisions or predictions. It can be used for classification, optimization, control, and other tasks where accurate predictions are needed. There are several types of ANFIS, including Type-1, Type-2, and hybrid systems. ANFIS is based on the principles of fuzzy logic, which allows it to consider multiple inputs and outputs simultaneously when making decisions. It also uses neural networks to adjust the strength of the rules. The Tagaki-Sugeno (TS) rule is a type of fuzzy inference system most commonly used for regression and control tasks. It is based on the Takagi-Sugeno-Kang formulation, which is an extension of the Mamdani type of fuzzy logic systems. TS rules combine both fuzzy sets and linear models to allow for more accurate predictions. The architecture of ANFIS can be briefly described with the use of fuzzy parameters. This is done using fuzzy "If-Then" rules, which allow us to perform ANFIS

input-output functions and input-output data pairs. Three components are needed for a fuzzy inference system:
(i) selection of fuzzy rules;
(ii) development of a data structure defining the membership functions (MF) used in the fuzzy rules;
(iii) a reasoning mechanism that performs the inference procedure based on the given fuzzy rules.
The fuzzy rules can be defined as
Rule 1 :"If x is X1 and y is Y1 then $f_1 = p_1x + q_1y + r_1$",
Rule 2 :"If x is X2 nd y is Y2 then $f_2 = p_2x + q_2y + r_2$".
ANFIS networks can be implemented by using the fuzzy toolbox of MATLAB software, where a Gaussian function is used to select fuzzy input parameters, like λ, μ_0, μ_1, μ_2 and μ_v. Moreover, linguistic variables are also defined for input parameters as seen in Table 5.

Table 5: *Values of the membership function for linguistics based on input parameters*

Input parameters	No. of membeship function	Linguistic Values
λ	5	very small, small, medium, large, very large
μ_0, μ_1, μ_2	4	small, medium, large, very large
μ_v	3	small, medium, large

5. Numerical Discussions

5.1. Sensitivity Analysis

In this section, MATLAB is utilized to illustrate how system behavior measurements are influenced by various parameters. The service time, vacation time, breakdown time, and repair time are assumed to be exponentially distributed, and the stability criterion is met by giving the parameters random values. Subsequently, numeric results for the primary performance indicators are obtained.

Variations in λ, (μ_0, μ_1, μ_2, μ_v) , $(\eta_1, \theta_1, \eta_2, \theta_2)$, are provided to reveal average system size ($\mathrm{E}[L_s]$), average queue size ($\mathrm{E}[L_q]$), and some performance measures in our queueing model.

Table 1 clearly depicts that as arrival rate (λ) rises, $\mathrm{E}[L_s]$, $\mathrm{E}[L_q]$ also escalates for the value of $\mu_0 = 3, \mu_1 = 2, \mu_2 = 1.5, \mu_v = 1, \eta_1 = 0.5, \eta_2 = 0.7, \theta_1 = 2, \theta_2 = 3, \kappa = 0.8, \xi = 1, \phi_v = 0.2, \phi_1 = 0.2, \nu_0 = 0.2, \nu_1 = 0.5, \nu_2 = 0.3$.

Table 2 clearly depicts that as $(\mu_0, \mu_1, \mu_2, \mu_v)$ escalates, $\mathrm{E}[L_s], \mathrm{E}[L_q]$ also diminish for the value of $\lambda = 1.7, \eta_1 = 0.5, \eta_2 = 0.7, \theta_1 = 2, \theta_2 = 3, \kappa = 0.8, \xi = 1, \phi_v = 0.2, \phi_1 = 0.2, \nu_0 = 0.2, \nu_1 = 0.5, \nu_2 = 0.3$.

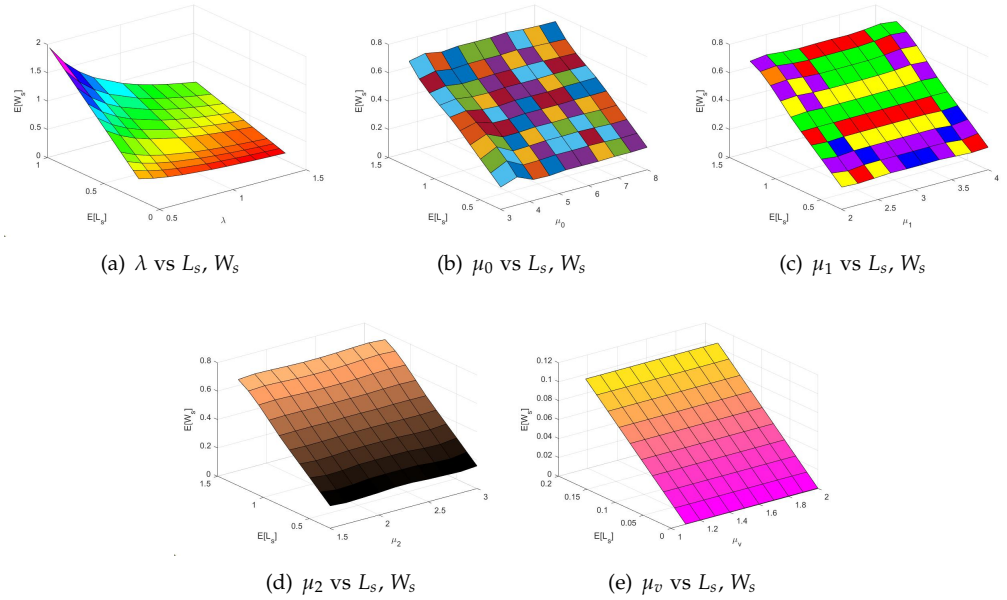
Table 3 clearly depicts that as $(\eta_1, \theta_1, \eta_2, \theta_2)$ varies, $\mathrm{E}[L_s], \mathrm{E}[L_q]$ are changed for the value of $\lambda = 1.7, \mu_0 = 3, \mu_1 = 2, \mu_2 = 1.5, \mu_v = 1, \kappa = 0.8, \xi = 1, \phi_v = 0.2, \phi_1 = 0.2, \nu_0 = 0.2, \nu_1 = 0.5, \nu_2 = 0.3$.

Figure 2 illustrates the impact of ν_0 for various values of μ_2 on the servers' idle state ($\iota_{0,0}$). The image illustrates that, when the value of μ_2 is held constant, the idle probability, shown by $\iota_{0,0}$, drops as one increases the value of ν_0. This is because when ν_0 rises, the number of customers choosing for SOS increases, and as a consequence, the server's idle probability drops.

Figure 3 shows the increasing nature of the number of customer in the system with varying values of μ_0. Figures. 4, 5, 6 and 7 demonstrate that an increase in the service rate of the second server in busy and vacation modes leads to a decrease in the number of customers ($\mathrm{E}[L_s]$) within the system. Moreover, these figures illustrate the effects on $\mathrm{E}[L_s]$ when varying values of

λ, η_1 and η_2 are applied.

Figure 8 (a) indicates that as the arrival rate (λ) increases, the expected queue length ($E[L_q]$) and expected system length ($E[L_s]$) also rises. On the other hand, Figures 8 (b–d) show that when the service rate of server 1 (μ_0) or servers 2 (μ_1, μ_2) increases, the expected queue length ($E[L_q]$) and the expected system length ($E[L_s]$) diminish.



(a) λ vs L_s, W_s (b) μ_0 vs L_s, W_s (c) μ_1 vs L_s, W_s

(d) μ_2 vs L_s, W_s (e) μ_v vs L_s, W_s

Figure 10: *Influence of a few parameters on 3D representation*

Figure 10(a – e) shows a three-dimensional graph depicting system performance measures. In Figure 10 (a), the surface illustrates an increase in the arrival rate (λ), expected system length ($E[L_s]$), and expected waiting time in the system ($E[W_s]$) increase. Figures 10 (b) and (c–d) demonstrate that as the service rates of server 1 (μ_0) and server 2 (μ_1, μ_2) increase, the expected system length ($E[L_s]$) and the expected waiting time in the system ($E[W_s]$) decrease. Figure 10(e) further shows that when the vacation rate μ_v increases, expected system length ($E[L_s]$) and expected waiting time in the system ($E[W_s]$) also decrease.

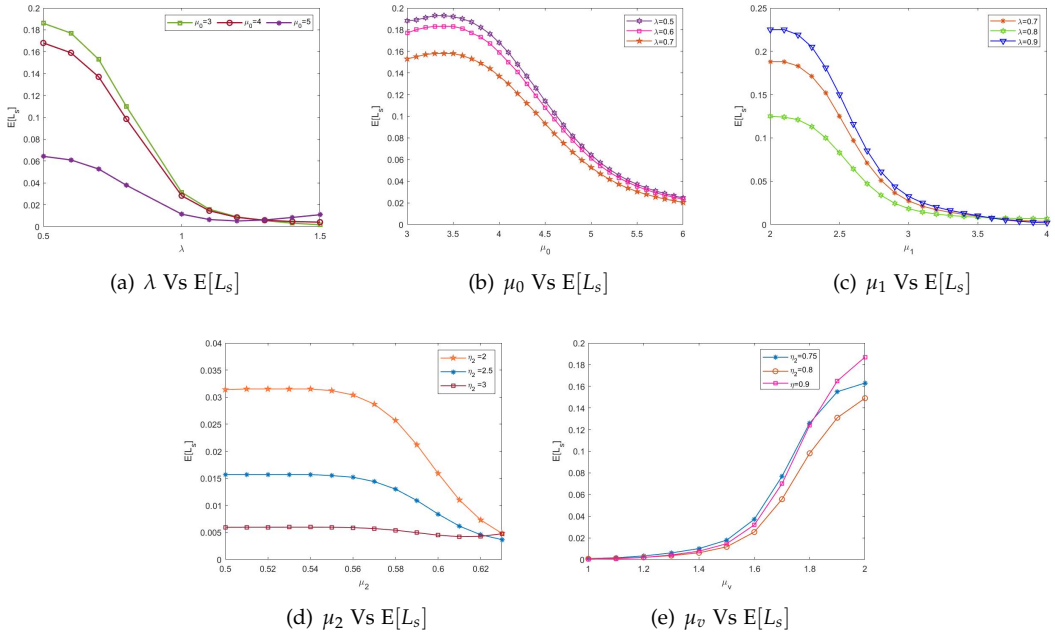
5.2. Anfis Computing and results

Tthe ANFIS results are constructed and verified by the Matlab software by executing the 'neuroFuzzyDesigner' command. The accuracy of the ANFIS outputs for $E[L_s]$ can be examined using the absolute percentage errors. Δ_a is provided by

$$\Delta_a = \frac{\left| E[L_s] - E[\overset{*}{L_s}] \right|}{E[L_s]} \times 100\%$$

where Δ_a is absolute percentage error, $E[L_s]$ exact value of the expected no.of customers in the system by analytical method, $E[\overset{*}{L_s}]$ estimated expected no.of customers in the system by ANFIS technique for varying values of (i) λ (ii) μ_0 (iii) μ_1 (iv) μ_2 (v) μ_v are recorded, and the absolute percentage errors (Δ_a) and accuracy of the estimated value in percentage of $E[L_s]$ are also summarized in Tables 6 - 7.

Lesser The Γ_e value indicates that our ANFIS method is closer to the analytical method's



(a) λ Vs E$[L_s]$ (b) μ_0 Vs E$[L_s]$ (c) μ_1 Vs E$[L_s]$

(d) μ_2 Vs E$[L_s]$ (e) μ_v Vs E$[L_s]$

Figure 11: E$[L_s]$ *Vs (a)* λ*, (b)* μ_0*, (c)* μ_1*, (d)* μ_2 *(e)* μ_v

results. Figures 7 (a-d) depict the graphs of membership functions for input values (i). *lambda* (ii) μ_0 (iii) μ_1 (iv) μ_v. The tick marks in Fig. 11 (a–d) represent the ANFIS results for the ESL forecast, while the continuous lines represent the analytical results. In these images, there are check marks almost completely covering the curved lines. This suggests that both outcomes are favourable.

Table 6: *Values of the* Δ_a, E$[L_s]$, E$^*[L_s]$ *by varying input parameters* λ, μ_0

λ	E$[L_s]$	E$^*[L_s]$	Δ_a	μ_0	E$[L_s]$	E$^*[L_s]$	Δ_a
0.5	0.1862	0.186	0.1074	3	1.2737	1.27	0.2905
0.6	0.2378	0.238	0.0841	4	0.8209	0.821	0.0122
0.7	0.2904	0.290	0.1377	5	0.5589	0.559	0.0179
0.8	0.348	0.348	0.0000	6	0.4409	0.441	0.0227
1.0	0.478	0.478	0.0000	7	0.2693	0.269	0.1114
1.1	0.5539	0.554	0.0181	7.5	0.3323	0.332	0.0903
1.2	0.6411	0.641	0.0156	8	0.2884	0.288	0.1387
1.3	0.7384	0.738	0.0542				
1.4	0.8442	0.844	0.0237				
1.5	0.9796	0.98	0.0408				
Average of Δ_a			0.0482	Average of Δ_a			0.0977
Accuracy in predicted value (%)			99.951	Accuracy in predicted value(%)			99.902

5.3. Cost Optimization

The estimated cost per unit of time TPC and total anticipated revenue TAR are found to be **$132.3475** and **$243.40** when certain values for the parameters are used such as λ=1, μ_0=6, μ_1=3, μ_2=3.5, μ_v=2, η_1=0.8, θ_1=5 , η_2=0.9, θ_2=5, κ=0.8, ξ=1, ϕ_v= 0.1, ϕ_1=0.2, ν_0=0.2, ν_1=0.5, ν_2=0.3,

Table 7: *Values of the* Δ_a, $E[L_s]$, $E[\overset{*}{L_s}]$ *by varying input parameters* μ_1, μ_2 μ_v

μ_1	$E[L_s]$	$E[\overset{*}{L_s}]$	Δ_a	μ_2	$E[L_s]$	$E[\overset{*}{L_s}]$	Δ_a	μ_v	$E[L_s]$	$E[\overset{*}{L_s}]$	Δ_a
2	0.8209	0.821	0.0122	1.5	0.4409	0.441	0.0227	1	0.8209	0.821	0.0122
2.5	0.5589	0.559	0.0179	2	0.2693	0.269	0.1114	1.1	0.5589	0.559	0.0179
3	0.4409	0.441	0.0227	2.5	0.3323	0.332	0.0903	1.2	0.4409	0.441	0.0227
3.5	0.2693	0.269	0.1114	3	0.2884	0.29	0.5548	1.3	0.2693	0.269	0.1114
4	0.3323	0.332	0.0903					1.4	0.3323	0.332	0.0903
								1.5	0.2884	0.29	0.5548
Average of Δ_a			0.0509	Average of Δ_a			0.1948	Average of Δ_a			0.1349
Accuracy in predicted value (%)			99.9491	Accuracy in predicted value (%)			99.8052	Accuracy in predicted value (%)			99.8651

Rev=250 are taken into consideration. To gauge the impact of changing the cost parameters, the cost function was examined for three different cost set values, as displayed in Table 4. In the feasible interval, the cost function is convex with regard λ, μ_1, μ_2 and η_1, η_2 as seen in Figs. 13(a-c) - 15(a-c).

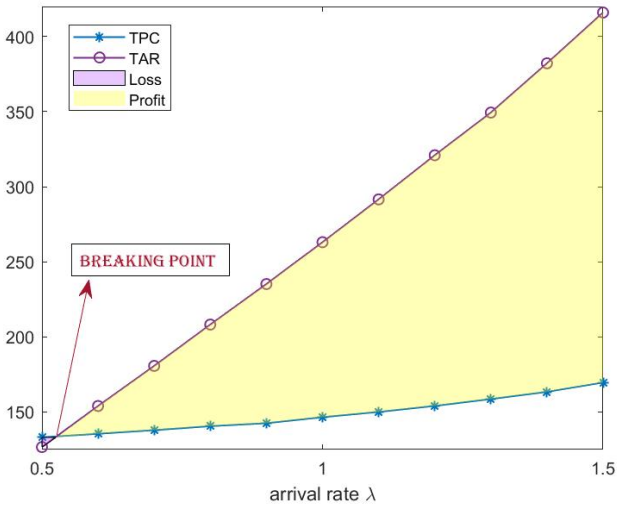
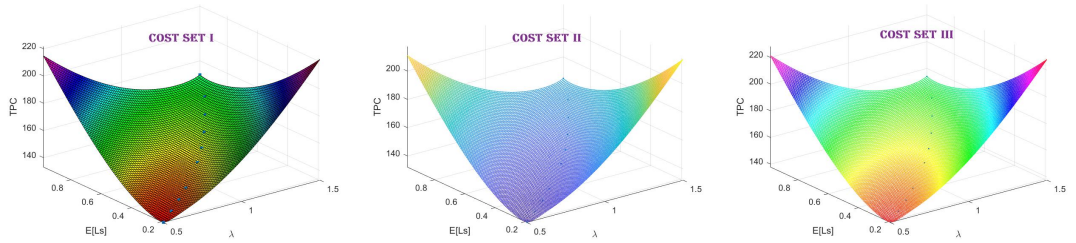
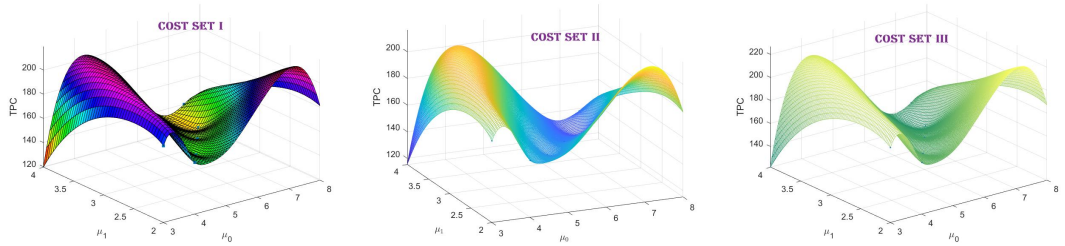


Figure 12: *Impact of* λ *on TPC, TAR*



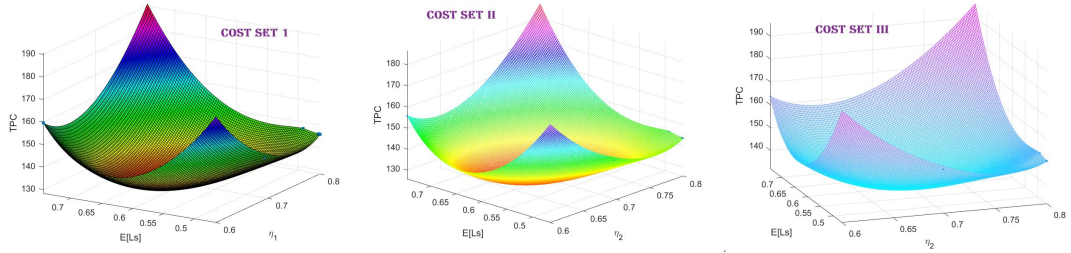
(a) TPC Vs λ, $E[L_s]$ (b) TPC Vs λ, $E[L_s]$ (c) TPC Vs λ, $E[L_s]$

Figure 13: *Total projected cost varying values of* λ, $E[L_s]$



(a) TPC Vs μ_0, μ_1 (b) TPC Vs μ_0, μ_1 (c) TPC Vs μ_0, μ_1

Figure 14: *Total projected cost varying values of* μ_0, μ_1



(a) TPC Vs η_1, E$[L_s]$ (b) TPC Vs η_1, E$[L_s]$ (c) TPC Vs η_1, E$[L_s]$

Figure 15: *Total projected cost varying values of* η_1, E$[L_s]$

The effect of λ on TPC and TAR in the model is depicted in Figure 12. As λ increases, the total elapsed cost and total expected response also increase. The point at which there is neither gain nor loss is seen around λ = 0.53, based on the set of cost values given for the model. If λ is lower than 0.53, there will be a loss, and when λ is greater than 0.53, the system will see a profit. Thus, with knowledge of the customer arrival rate, appropriate actions can be taken to reduce the TPC and maximize the TAR. Our model's TPC and TAR are shown with the relevant values in Table 8.

6. Conclusion

This research explores a heterogeneous and unreliable server queueing system with the addition of a second optional service, feedback, and breakdown in a hybrid vacation scheme. Through numerical examples, The efficiency of the matrix-geometric technique in determining steady-state probabilities and other performance metrics has been discovered. Utilizing the matrix-geometric approach and Adaptive Network-Based Fuzzy Interference System (ANFIS) for calculating performance indices and the cost function, it has been shown that the model can be used in a real-time system. By comparing the results from ANFIS and the numerical results, we can demonstrate the usefulness of a neural-fuzzy tool to assess the performance of queueing systems based on commercial and technological standards. Lastly, this strategy can be further applied to more realistic scenarios, such as those with the concepts of a phase-type arrival and impatient customer, standby servers and Markovian models like $MAP/PH/1$ and $GI^X/Geo/1$ queues.

Acknowledgement: Not applicable
Funding: Not applicable
Data Availability: Not applicable

Supplementary Materials: Not applicable
Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Ahuja, A. Jain, A. Jain, M. (2022) Transient analysis and ANFIS computing of unreliable single server queueing model with multiple stage service and functioning vacation. *Mathematics and Computers in Simulation*. 192:464-490. https://doi.org/10.1016/j.matcom.2021.09.011

[2] Anitha, K. Maragathasundari, S. Bala, M. (2017) Queuing system of bulk arrival model with optional services in third stage and two different vacation policies. *International Journal of Mathematics*. 5:711-721.

[3] Bouchentouf, A. A. Boualem, M. Yahiaoui, L. Ahmad, H. (2022) A multi-station unreliable machine model with working vacation policy and customers' impatience. *Quality Technology & Quantitative Management*. 1-31. DOI:10.1080/16843703.2022.2054088

[4] Bouchentouf, A. A. Cherfaoui, M. Boualem, M. (2019) Performance and economic analysis of a single server feedback queueing model with vacation and impatient customers. *Opsearch* . 56:300-323. http://dx.doi.org/10.1007/s12597-019-00357-4

[5] Chang, F. M. Liu, T. H. Ke. J. C. (2019) On an unreliable-server retrial queue with customer feedback and impatience. *Applied Mathematics and Modelling* . 55:171-182. http://dx.doi.org/10.1016/j.apm.2017.10.025

[6] Chandrika, K.U. Kalaiselvi. C. (2013) Batch arrival feedback queue with additional multi optional service and multiple vacation. *International Journal of Scientific Research Publications*. 3(3):1-8. https://www.ijsrp.org/research-paper-0313.php?rp=P15833

[7] Chakravarthy, S.R. Shruti, Kulshrestha. (2020) Queueing model with server breakdowns, repairs, vacations, and backup server. *Operations Research Perspectives*. 7:100131 https://doi.org/10.1016/j.orp.2019.100131

[8] Ezeagu, N.J. Orwa, G. O. Winckler. M.J. (2018) Transient analysis of a finite capacity $M/M/1$ queuing system with working breakdowns and recovery policies. *Global Journal of Pure and Applied Mathematics* . 14(8):1049-1065. DOI:10.12691/ajams-7-1-1

[9] Li, N. (2016) Stanford. Multi-server accumulating priority queues with heterogeneous servers. *European Journal of Operational Research* . 252(3):866-878. DOI:10.1016/j.ejor.2016.02.010

[10] Li, J. Wang, J. (2006) An $M/G/1$ retrial queue with second multi-optional service, feedback and unreliable server. *Journal of Applied Mathematics*. 21:252-262.

[11] Jang, J.S. (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernet ics*. 23(3):665-685. http://dx.doi.org/10.1109/21.256541

[12] Jang, J.S.R. Sun, C.T. Mizutani. E. (1997) Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence. *IEEE Transactions on Automat*. 42(10):1482-1484. http://dx.doi.org/10.1109/TAC.1997.633847

[13] Ke, J.C. Wu, C.H. Pearn, W.L. (2013) Analysis of an infinite multi-server queue with an optional service. *Computers & Industrial Engineering*. 65(2):216-225. http://dx.doi.org/10.1016/j.cie.2013.02.017

[14] Krishnamoorthy, A. Sreenivasan, C. (2012) An $M/M/2$ queueing system with heterogeneous servers including one with working vacation. *International Journal of Stochastic Analysis*. https://doi.org/10.28919/jmcs/6165

[15] Latouche, G. Ramaswami, V. (1999) Introduction to matrix analytic methods in stochastic modeling. *SIAM Review*. . https://doi.org/10.1155/S1048953399000362

[16] Kumar Anshul, Madhu Jain. (2022) Cost Optimization of an Unreliable server queue with two stage service process under hybrid vacation policy. *Mathematics and Computers in Simulation*.259-281. https://doi.org/10.1016/j.matcom.2022.08.007

[17] Morse, P. M. (2004) Queues, inventories and maintenance: the analysis of operational systems with variable demand and supply. *Courier Corporation*.

[18] Neuts, M.F. (1981) Matrix-Geometric solutions in stochastic models. *Johns Hopkins University Press* .

[19] Servi, L.D. Finn, S.G. (2002) $M/M/1$ queues with working vacations ($M/M/1/WV$), *Performance Evaluation*. 50(1):41-52. http://dx.doi.org/10.1016/S0166-5316(02)00057-3

[20] Sethi, R. Jain, M. Meena, R.K. Garg, D. (2020) Cost optimization and ANFIS computing of an unreliable $M/M/1$ queueing system with customers impatience under N-policy. *International Journal of Applied Mathematics and Statistics* . 6:1-14. https://link.springer.com/article/10.1007/s40819-020-0802-0

[21] Shekhar, C. Varshney, S. Kumar. A. (2021) Matrix-geometric solution of multi-server queueing systems with Bernoulli scheduled modified vacation and retention of reneged customers: A meta-heuristic approach. *Quality Technology & Quantitative Management*. 18(1):39-66. http://dx.doi.org/10.1080/16843703.2020.1755088

[22] Vijaya Laxmi, P. Jyothsna, K. (2022) Cost and revenue analysis of an impatient customer queue with second optional service and working vacations. *Communications in Statistics-Simulation and Computation*. 51(8:4799-4814. 10.1080/03610918.2020.175237

[23] Vijayalakshmi, V. Kalidass, K. Pavitha, K. (2018) An $M/M/1/N$ queue with working breakdowns and a two-phase service. International *Journal of Pure and Applied Mathematics* 119(15):2285-2297.

[24] Upadhyaya, S. Kushwaha, C. (2020) Performance prediction and ANFIS computing for unreliable retrial queue with delayed repair under modified vacation policy. *International Journal of Mathematics in Operational Research*. 17(4):437-466. https://doi.org/10.1504/IJMOR.2020.110843

[25] Wu, C.H. Yang, D.Y. (2021) Bi-objective optimization of a queueing model with two-phase heterogeneous service . *Computers & Operations Research* 130:105230. http://dx.doi.org/10.1016/j.cor.2021.105230

[26] Yang,D.Y. Wang, K.H. Kuo, Y.T. (2011) Economic application in a fnite capacity multi-channel queue with second optional channel. *Applied Mathematics and Computation* 217(18):7412-7419 . http://dx.doi.org/10.1016/j.amc.2011.02.031

[27] Yang, D.Y. Chen, Y.H. (2018) Computation and optimization of a working breakdown queue with second optional service. *Journal of Industrial and Production Engineering* 35(3):181-188. http://dx.doi.org/10.1080/21681015.2018.1439113

[28] Ye, Q. Liu, L. (2017) Analysis of $MAP/M/1$ queue with working breakdowns. *Communications in Statistics - Theory and Methods* 47(13) 3073-308. http://dx.doi.org/10.1080/03610926.2017.1346808

[29] Ziad, I. Laxmi, P.V. Bhavani, E.G. Bouchentouf, A. A. Majid . (2023) Matrix Geometric Solution of a Multi-Server Queue With Waiting Servers and Customers Impatience Under Variant Working Vacation and Vacation Interruption. *Yugoslav journal of operations research* . http://dx.doi.org/10.2298/YJOR220315001Z

Table 8: *Arrival and service rates and corresponding TPC, TAR*

λ	**1**		**1.1**		**1.2**		**1.3**		**1.4**		**1.5**	
	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR
I	146.55	263.05	150.05	291.57	154.06	320.95	158.63	349.36	163.42	382.17	169.63	415.98
II	144.16	263.05	147.28	291.57	150.85	320.95	154.94	349.36	159.20	382.17	164.74	415.98
III	151.13	263.05	154.58	291.57	158.53	320.95	163.10	349.36	167.74	382.17	173.86	415.98
$(\mu_0, \mu_1, \mu_2, \mu_v)$	(**4**, 2, 1.5, 1)		(**5**, **2.5**, 1.5, 1)		(**6**, **3**, 1.5, 1)		(**7**, **3.5**, **2**, 1)		(**7.5**, **4**, **2.5**, 1)		(8,4,3,**1.5**)	
	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR
I	162.19	458.12	150.21	463.68	144.83	470.60	135.69	355.26	139.89	478.33	137.90	456.97
II	158.08	458.12	147.41	463.68	142.62	470.60	134.34	355.26	138.23	478.33	136.46	456.97
III	166.72	458.12	154.91	463.68	149.62	470.60	140.19	355.26	144.75	478.33	142.77	456.97
$(\eta_1, \theta_1, \eta_2, \theta_2)$	(**0.6**,2,**0.55**,3)		(**0.7**,2,**0.6**,3)		(**0.8**,**3**,**0.7**,**4**)		(0.8,**4**,**0.8**,**5**)		(0.8,**5**,**0.9**,5)		(0.8,5,0.9,**5**)	
	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR	TPC	TAR
I	146.17	258.32	146.02	257.25	147.25	267.52	145.90	256.67	145.94	257.01	133.28	243.40
II	143.82	258.32	143.69	257.25	144.79	267.52	143.58	256.67	143.61	257.01	132.34	243.40
III	150.77	258.32	150.63	257.25	151.81	267.52	150.53	256.67	150.56	257.01	138.14	243.40

EVALUATION OF RELIABILITY' INDICES AND CHARACTERISTICS OF THE POWER SYSTEM' EQUIPMENT AND DEVICES BY NON TRADITIONAL METHOD

Farzaliyev Y.Z., Farhadzadeh E.M.

•

Azerbaijan Scientific - Research and Design - Prospecting Institute of Energetic, Baku city, H. Zardabi Avenue 94
yusif.farzaliyev@azerenerji.gov.az

Abstract

In the paper considered the research expediency classification of statistical data according to the given varieties of signs. The researching carried out based on modeling of small and multidimensional samples to statistical distribution functions. A discrepancy found in the estimation of the mathematical expectation of the average values of sample implementation, to overcome this inconsistency, a new method for modeling samples of random variables is proposed. It established that the classification in the literature data carried out according to the varieties of signs accepted in the classifiers without control of expediency. The causes of errors arising in the evaluation of Kolmogorov statistics as the largest in absolute deviation are analyzed the deviation between statistical distribution functions of the population and sample using simulation modeling, fiducially intervals and the theory of testing statistical hypotheses. These erroneous calculations with a small number and multidimensionality of sampling implementations double increases of the Type II Error. Finally, the result showed the advantages of the new method in comparison with Kolmogorov' criterion via checking representativeness of sample.

Keywords: representativeness, reliability indices, varieties of signs, statistical distribution function, simulation modeling, fiducially intervals, testing statistical hypothesis, type I and II errors, sample, population, multidimensionality

1. Introduction

In the paper presents the results of a study most difficult case of estimating the statistical distribution function for a given varieties of signs. An analogue of the problem solved is the Kolmogorov's criterion with the significant difference that the statistical distribution function here compared with the analytical distribution of a random variable. The practical application of this criterion is often erroneous, since it is not the Kolmogorov' statistics that is compared with the critical value, but the magnitude of the largest difference between the given distribution function and the statistical distribution function of a random sample, which is similar to it. It shown that reducing the risk of an erroneous decision in a situation where the deviation between the distributions functions is doubtful achieved by taking into account the magnitude of the Type II

Error. The study of this important issue allowed us to identify the main cause of this error and indicate the way to eliminate it. By the above criteria, the representativeness of the sample is recognized only under the condition that the significance of the statistics "the greatest value of a random variable" in comparison with other statistics is significantly higher.

The main assumption is the possibility of presenting the statistical data of operation via representative sample from the population of these data, i.e. these data appear homogeneous. The reliability indices calculated here are naturally averaged characters. In reality, the data belongs to the class of multidimensional data. However, due to the lack of methods for analyzing multidimensional data, they are mistakenly taken as an analogue of the general population, and calculations of reliability' indices and characteristics are carried out using methods that focus on analyzing samples from the general population, i.e. data with one constant distribution law. In turn, this assumption leads to erroneous decisions with all the ensuing consequences. Therefore, reliability ensuring provides for the possibility of comparing estimates of reliability indices of specific electrical equipment, i.e. the transition from the average reliability indices to indices of individual reliability.

In analyzing the reliability of electric power system equipment, the classification of operation statistics data carried out on one, and sometimes on two signs. Classification of statistical data on more than two signs not practiced. The reason for this is the diversity of the varieties of signs and the decrease in the accuracy estimates of reliability indices. The decrease in accuracy goes because of the assumption that the statistical data corresponds to a *random sample from general population.*

Statistical data characterizing the reliability of electric power system equipment (information on non-operating states) depend on a large number of passport and operational data (installation site, voltage class, design, service life and other signs.). That is why they *cannot be considered* either as an *analogue of the general population*, or as *a finite sample of homogeneous data*. Firstly, multidimensional data set not only by a set of random variables characterizing the reliability of the studies objects, but also by set of varieties of signs characterizing each random variable.

When classifying the multidimensional statistical data for a given varieties of signs, sample data is extracted non-randomly from a finite population of multidimensional data. A non-random sample consists of random variables and the distribution' features in the variation' interval of random variables of a finite population of multidimensional data depend on the varieties of signs. The type of distribution law for a finite population of multidimensional statistical data not known and systematically changes randomly as statistical data accumulated. The change' interval of a random variable in a sample from finite population of multidimensional statistical data for a given varieties of signs is no longer than the change' interval of a random variable in the finite population.

These features allow us to conclude that the use of *classical methods for analyzing samples from the general population for analyzing samples from a finite population of multidimensional and small volume data leads to an increase in the risk of an erroneous decision.*

2. Methods

The method and algorithm for calculating statistical distribution function (s.d.f.), which characterizes the largest deviation of $F_\Sigma(\xi)$ and $F_S^*(\xi)$, provided that $F_S^*(\xi)$ is unrepresentative, consists of the following sequence of calculations:

1. The next (from the required N implementations) sample of n random numbers is simulated;
2. Forming s.d.f. of $F_S^*(\xi)$;

3. The largest divergence between $F_\Sigma(\xi)$ and $F_S^*(\xi)$ is determined. We denote this value as $\Delta_{n,emp}$, where the index "emp" corresponds to the empirical character of the sample.
   Having determined the statistical characteristics of this sample { $F_S^*(\xi)$ and $\Delta_{n,emp}$}, we proceed to the formation of $F^*(\Delta_n^*)$ according to the realizations of the greatest divergence between the distribution functions $F_\Sigma(\xi)$ and the set (N). $F_S^*(\psi)$, modeled on $F_S^*(\xi)$. For what:
4. According to $F_S^*(\xi)$ is forming distribution: where i=1,(n+1); ξ - is random variable with a uniform distribution in the interval [0,1];

$$F_S^*(\psi) = \begin{cases} 0 & if\ \psi \le \psi_1 \\ \dfrac{i-1}{n+1} + \dfrac{(\psi - \psi_i)}{(\psi_{i+1} - \psi_i)(n+1)} & if\ \psi_1 < \psi < \psi_{n+1} \\ 1 & if\ \psi \ge \psi_{n+1} \end{cases} \quad (1)$$

5. On standard RAND() program is simulation random number ξ with a uniform distribution in the interval [0,1];
6. On the distribution (1) is calculation a random number ψ corresponding to probability ξ. Calculations are carried out according to the formula:

$$\psi = \psi_i + (\psi_{i+1} - \psi_i)[\xi \cdot (n+1) - (i-1)] \quad (2)$$

with i=1,(n+1)

7. Items 5 and 6 are repeated n times;
8. On the sampling of $\{\psi\}_n$ builds s.d.f. $F_S^*(\psi)$;
9. The largest divergence between $F_\Sigma(\xi)$ and $F_S^*(\psi)$ is determined. Denote it by Δ_n^*;
10. Items (5÷9) will repeat N times;
11. The average value of the random variable Δ_n^* is determined. Denote it by $M^*(\Delta_n^*)$;
12. According to N values of Δ_n^*, s.d.f. of $F^*(\Delta_n^*)$ is formed [1].

If we assume that distribution $F^*(\Delta_n^*)$ corresponds to the normal distribution law, and the average value of $M^*(\Delta_n^*)$ is equal to $\Delta_{n,emp}$ and corresponds to $F^*(\Delta_n^*) = \beta = 0{,}5$, then for all implementations of $\Delta_{n,emp}$, the probability of which is $0.1<\alpha<0.5$, preference should be given to assumption H_2. However, the assumption about distribution function of normal law of the $F^*(\Delta_n^*)$ does not correspond to reality. As an example, figure 1 shows a histogram of the distribution of implementations Δ_n^* for s.d.f. $F_S^*(\psi)$ given in table 1.

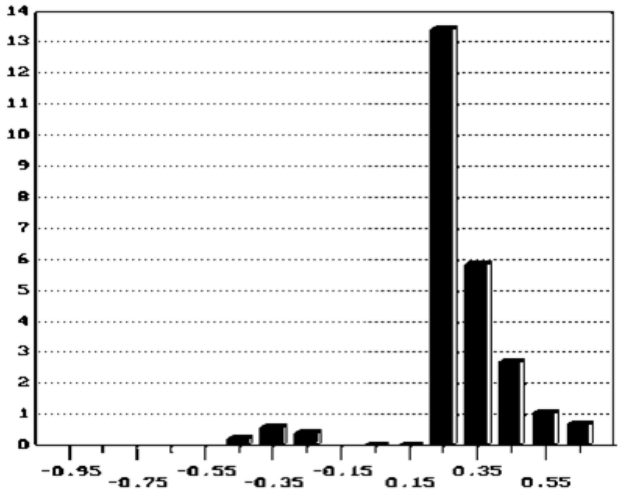


Figure 1. The histogram of implementations Δ_n^*

Experimental studies have established that:

– evaluating of the expediency of data classification by comparing the boundary values of confidence intervals of reliability indices is associated with an increase in the risk of erroneous decisions;

– the error in using the absolute value of the largest discrepancy between $F_{\Sigma}^{*}(\xi)$ and $F_{S}^{*}(\xi)$ instead of the calculated value of Kolmogorov's statistics is in the difference of their distribution functions, and, consequently, in the critical values [2].

– the regression equation of boundary values of fiducially intervals obtained by the standard program of the power transformation with determination coefficient R^2 (R^2>0,999) looks like:

$$\left.\begin{array}{c} \underline{\Delta(H_1)} = -An_S^{-0,5} \\ \overline{\Delta(H_1)} = -\left[\underline{\Delta(H_1)} - n_S^{-1}\right] \text{where A=0,652}\,\alpha^{-0,175} \\ \text{at } \alpha = 0,05 \\ \underline{\Delta(H_1)} = -1,12\,n_S^{-0,5} \\ \overline{\Delta(H_1)} = (1-1,12n_S^{0,5})n_S^{-1} \end{array}\right\}$$

– the reducing risk of an erroneous decision is achieved by taking into account the significance of the difference in the distribution of random variables of the divergence between the population's s.d.f. and data sampling.

In this regard, there is a need for modeling s.d.f. statistical parameters of random variables [3].

3. Results and Discussion

Classification of statistical data according to the given varieties of signs, firstly, presupposes the possibility to evaluating its expediency.

One of the ways to characterize the expediency of data classification is to assess the nature of the divergence of the s.d.f. a finite population of multidimensional data and a sample of this multidimensional data for a given varieties of signs.

Thus, the developed method and algorithm determines the most significant varieties of signs and, therefore, the working sample at each stage of the classification of multidimensional data. Reducing the risk of erroneous classification of multidimensional statistical data carried out by evaluating the expediency of such a classification. The basis of the comparison between $F_{\Sigma}^{*}(X)$ and $F_{S}^{*}(X)$ is the statistical modeling of n_s pseudorandom numbers ξ equal to the number of random variables of the sample, with a uniform distribution in the interval [0,1]. A precondition for this is the random nature of the difference between $F_{\Sigma}^{*}(\xi)$ and $F_{S}^{*}(\xi)$.

The representative nature of sample $\{\xi\}_{n_s}$, in solving the problem of evaluating the expediency of classifying multidimensional data controlled by the Kolmogorov's criterion. According to this criterion, sample $\{\xi\}_{n_s}$ is not representative if:

$$D_n > D_{n,(1-\alpha)} \tag{3}$$

where:

$$D_n = \max(D_n^+, D_n^-) \tag{4}$$

$$D_n^+ = \max\{D_i^+\}; \qquad D_n^- = \max\{D_i^-\}; \qquad 1 \le i \le n \tag{5}$$

$$D_i^+ = \left(\frac{i}{n} - \xi_i\right) \tag{6}$$

$$D_i^- = \left(\xi_i - \frac{i-1}{n}\right) \tag{7}$$

$D_{n,(1-\alpha)}$ - is the critical value of the statistics D_n, provided that $F_{\Sigma}^{*}(\xi)$ and $F_{s}^{*}(\xi)$ differ randomly.

In Kolmogorov's criteria it is noted that the evaluation of D_n by formula,

$$D_n^+ = \max\{|D_i^+|\}; \quad 1 \le i \le n \tag{8}$$

leads to incorrect decisions about the ratio of $F_{\Sigma}^{*}(\xi)$ and $F_{s}^{*}(\xi)$. The reason for this discrepancy not specified. For an indefinite in advance n, a decrease in the calculation time, is achieved by using the exact Stephens approximation, which tabulated critical values of $D_{n,(1-\alpha)}$, depending on n and α reduces to a dependence only on α. The sampling $\{\xi\}_n$ is unrepresentative if:

$$A \cdot D_{n_V} > S_{1-\alpha} \tag{9}$$

where

$$A = \left(\sqrt{n_V} + 0.12 + \frac{0.11}{\sqrt{n_V}}\right) \tag{10}$$

For example, for $n_s = 4$, the value of A = 2.175 and for α = 0.1 the critical value is $S_{1-\alpha} = 1.224$, and for α = 0.05, the value is $S_{1-\alpha} = 1.358$.

The application of the inverse problem-solving method, where it is known in advance that sample $\{\xi\}_n$ is unrepresentative, has shown that the criteria (3) and (8) for the most commonly used in practice values of α=0.05 and α=0.1 do not establishes the non-random nature of the divergence between $F_{\Sigma}^{*}(\xi)$ and $F_{S}^{*}(\xi)$ at less n_s only for those cases where it is not in doubt.

For example: To confirm this statement, consider the following example. Let random numbers ψ have a uniform distribution $F_{\Sigma}(\psi)$ in the interval [0.5; 1]. A random sample $\{\psi\}_n$ with n=4 is specified: {0.86346; 0.50672; 0.91424 and 0.67210}. Let us check the assumption of the representativeness of this sample for the uniform distribution law of the random variable ξ in the interval [0,1]. The calculations' results gave in table 1.

Table 1. An example of representativeness of the sample

i	$F_{\Sigma}(\psi_i)$	i/n	D_i^+	D_i^-	Note
1	0.507	0.25	-0.257	+0.506	$D_i^+ = 0.086$; $D_i^- = 0.506$
2	0.672	0.5	-0.172	+0.422	D_n=0.506; $D_n<D_{4;\,0.9}$=0.565
3	0.863	0.75	-0.113	+0.363	AD_n=1.101;
4	0.914	1.00	+0.086	+0.164	$AD_n<S_{0.9}$=1.224

As follows from table 1, sample $\{\psi\}_4$ does not contradict the assumption of representativeness with respect to $F_{\Sigma}(\xi)$ with $\alpha = 0.1$

These features and some assumptions about the reasons for their occurrence required to move from the analysis of the absolute values of the largest difference the discrepancy D_n, to the analysis of the distribution of the largest absolute value of the implementation of vertical discrepancy $F_{\Sigma}(\xi)$ and $F_s^*(\xi)$, which we denote as Δ_n. The use of type formulas:

$$\Delta_n = \max\left(\xi_i - \frac{i}{n}\right) \quad 1 \le i \le n \tag{11}$$

when calculated on a computer leads to erroneous results. For example, according to the data of Table 1, the maximum value among the four implementations of the value of D_i^+ will be $D_i^+ = 0.086$, and the largest in absolute value vertical divergence between $F_{\Sigma}(\xi)$ and $F_S^*(\xi)$ will be equal to $D_1^+ = -0.256$.

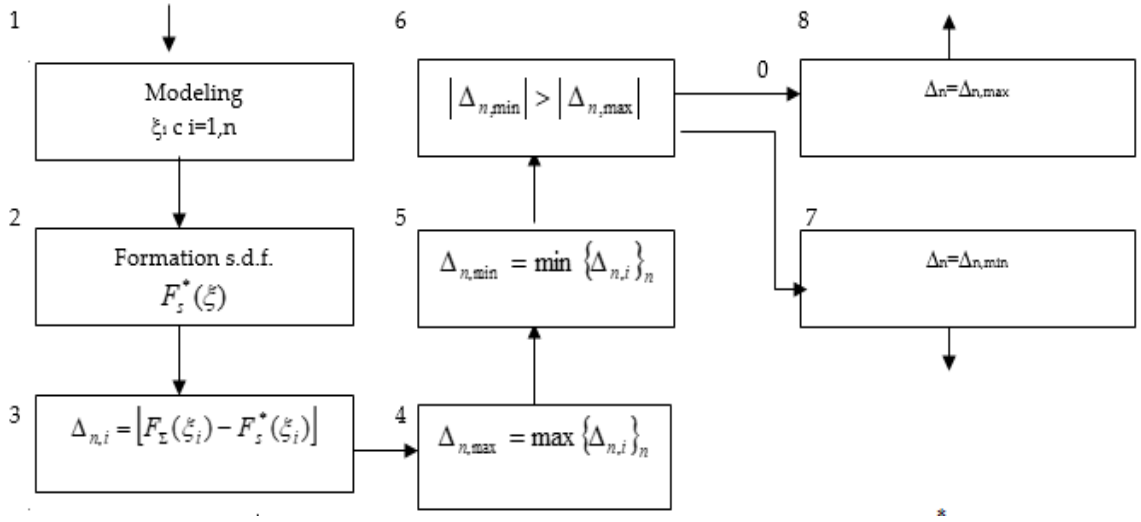


Figure 2. Block diagram of the algorithm for calculating the largest divergence distributions $F_{\Sigma}(\xi)$ and $F_S^*(\xi)$

The systematization' results of these implementations presented in Table 2 allow us to conclude:

Table 2. Some evaluation results of s.d.f. $F^*(\Delta_n)$

$F^*(\Delta_n)$ / n	0,025	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,975
2	-0.842	-0.775	-0.684	-0.551	-0.473	-0.149	-0.363	-0.304	-0.239	-0.060	0.184	0.285	0.343
3	-0.7094	-0.635	-0.566	-0.471	-0.400	-0.335	-0.296	-0.252	-0.200	-0.145	0.231	0.299	0.372
4	-0.623	-0.567	-0.494	-0.414	-0.355	-0.302	-0.253	-0.217	-0.173	0.155	0.240	0.319	0.377
5	-0.567	-0.511	-0.449	-0.370	-0.318	-0.274	-0.232	-0.190	-0.147	0.164	0.246	0.309	0.360
6	-0.523	-0.469	-0.411	-0.338	-0.292	-0.252	-0.215	-0.173	-0.127	0.171	0.244	0.303	0.358
7	-0.481	-0.438	-0.384	-0.318	-0.274	-0.235	-0.201	-0.162	-0.113	0.165	0.235	0.290	0.342
11	-0.389	-0.353	-0.309	-0.255	-0.219	-0.189	-0.110	-0.129	-0.097	0.160	0.216	0.260	0.302
16	-0.33	-0.295	-0.258	-0.215	-0.184	-0.158	-0.134	-0.103	0.107	0.150	0.194	0.232	0.264
22	-0.280	-0.253	-0.221	-0.183	-0.157	-0.135	-0.113	-0.083	0.105	0.137	0.176	0.210	0.235
29	-0.246	-0.219	-0.193	-0.160	-0.138	-0.119	-0.099	-0.068	0.098	0.126	0.158	0.186	0.212
40	-0.208	-0.187	-0.164	-0.136	-0.119	-0.102	-0.084	-0.050	0.089	0.112	0.140	0.164	0.185
60	-0.173	-0.156	-0.137	-0.114	-0.097	-0.083	-0.069	0.054	0.077	0.096	0.118	0.138	0.155
90	-0.142	-0.127	-0.111	-0.092	-0.079	-0.068	-0.055	0.051	0.067	0.081	0.100	0.116	0.130
120	-0.122	-0.110	-0.096	-0.080	-0.068	-0.059	-0.047	0.047	0.060	0.072	0.089	0.102	0.114
150	-0.110	-0.099	-0.086	-0.071	-0.062	-0.053	-0.042	0.041	0.053	0.065	0.079	0.092	0.104

1. The quantiles of the $F^*(\Delta_n)=\alpha$ distribution with $n\geq 2$ are equal in magnitude and opposite in sign (the difference in sign is due to the difference in formulas (6) and (11) to the quantiles of the distribution $F(D_n)=2\alpha$;
2. The distribution of $F^*(\Delta_n)$ is asymmetrical. For illustrative purposes in figure 3 are s.d.f. $F^*(\Delta_n)$ for n_s. It is precisely the assumption about the symmetry of the distribution $F(\Delta_n)$ that can explain the discrepancy between the probability of almost equal quantiles of the distributions $F^*(\Delta_n)$ and $F(D_n)$;
3. The smaller ξ_n is, the greater the negative value of Δ_n in sign, since $\Delta_n=(\xi_n-1)$. According to experimental data, the smallest Δ_n value for n = 2 turned out to be Δ_n= -0,992, and the largest Δ_n=+0,489 with supremum's equal to 1 and 0.5 respectively;
4. In the distribution $F^*(\Delta_n)$, we will distinguish between lower $\underline{\Delta_n}$ and upper $\overline{\Delta_n}$ boundary values with a significance level α, i.e.

$$\left.\begin{aligned} F^*\left(\underline{\Delta_n}\right) &= \alpha/2 \\ F^*\left(\overline{\Delta_n}\right) &= (1-\alpha/2) \end{aligned}\right\} \qquad (12)$$

5. It was established that at $0{,}25\geq F^*(\Delta_n)\geq 0{,}75$, i.e. at $\alpha\leq 0{,}5$

$$\overline{\Delta_n} = -\left(\frac{1}{n} + \underline{\Delta_n}\right) \qquad (13)$$

For example, for n = 4 and α=0.1 in accordance with the distribution of $F^*(\Delta_n)$ (see table 2) the value of $\underline{\Delta_4} = -0.567$, and $\overline{\Delta_4} = +0.319$. At the same time, according to the formula (13):

$$-(0{,}25-0{,}567)=0{,}317=\overline{\Delta_4}$$

If n = 29 and α = 0.2, then $\underline{\Delta_n} = -0.193$, and $\overline{\Delta_n} = 0.158$. The value of $\overline{\Delta_n}$ by the formula (13) is

$$-(0.034-0.193) = 0.159.$$

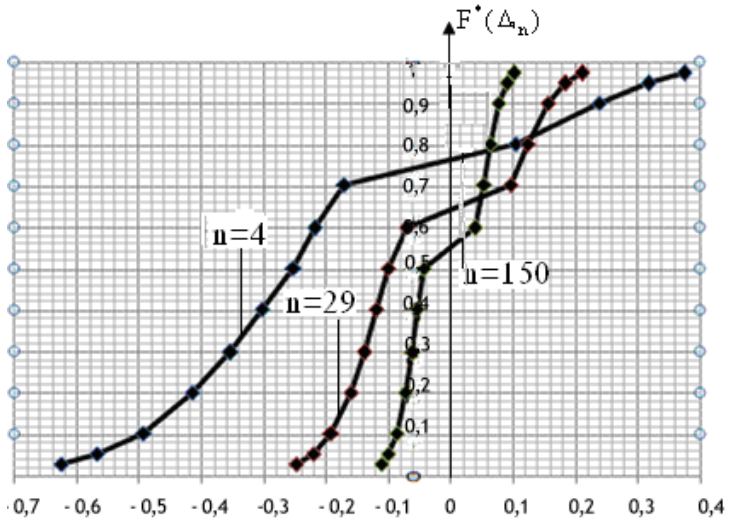


Figure 3. S.d.f. of $F^*(\Delta_n)$ for numbers n_s

In figure 4 shows the histograms of the distribution of negative and positive values of Δ_n for n = 4 and n = 29

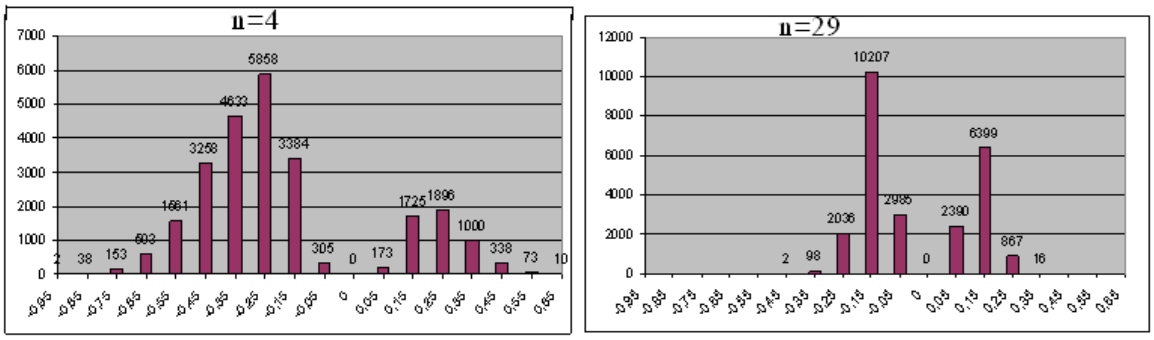


Figure 4. Histograms of the distribution of the greatest divergence of distributions $F_\Sigma(\xi)$ and $F_S^*(\xi)$

As follows from figure 4, negative values of Δ_n significantly exceed positive values of Δ_n in relative number and range of change. Based on paragraph 3, it is clear that this is not accidental and does not indicate that the sample is not representative. With increasing n, the ratio of negative and positive values of Δ_n decreases and tends to unity. For n = 2, negative values of Δ_n are 87.5%, and for n = 29, 61%, and for n = 150, 55%. Thus, even for n = 150, the quantiles of the distribution $F^*(\Delta_n)$ for $\alpha = 0.05$ and $\alpha = 0.95$ are not equal to [–0.099; +0.092]. The histograms also explain the patterns of distribution $F^*(\Delta_n)$ shown in Figure 3.

Figure 5 shows the curves of changes in the boundary values statistics Δ_n for a number of values s.d.f. $F^*(\Delta_n)$. The criterion for controlling the representativeness of sample $\{\xi\}_n$ with significance level α in this case is:

$$\underline{\Delta_n} < \Delta_n < \overline{\Delta_n} \qquad (14)$$

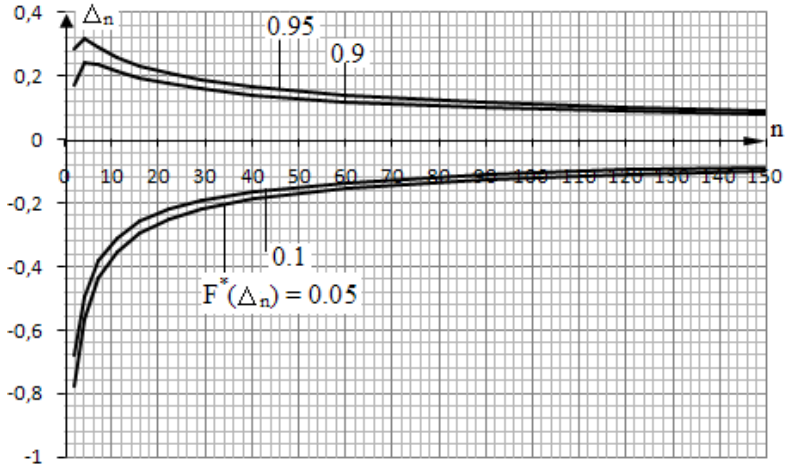


Figure 5. Regularity change of the boundary values to the greatest divergence of the distributions $F_\Sigma(\xi)$ and $F_S^*(\xi)$

Denote positive values of Δ_n by Δ_n^+, and negative values - Δ_n^-. Taking into account paragraph 1 and equation (13), a sample $\{\xi\}_n$ with a significance level of $\alpha \le 0{,}5$ can be taken as representative if:

$$\left.\begin{aligned} \Delta_n^+ &< \left[D_{n,(1-2\alpha)} - \frac{1}{n} \right] \\ \left|\Delta_n^-\right| &< D_{n,(1-2\alpha)} \end{aligned}\right\} \qquad (15)$$

whereas:

$$\left(\Delta_n^+ + \frac{1}{n} \right) = \left|\Delta_n^-\right|$$

criterion (14) for significance level α can be represented as:

$$\left(\overline{\Delta_n^+} + \frac{1}{n} \right) = \left|\underline{\Delta_n^-}\right| = D_{n,(1-2\alpha)} \qquad (16)$$

Here it is necessary to pay attention to the inconsistency of the equations of significance of Δ_n and $d_{n,(1-2\alpha)}$.

If we look at Table 1, it is easy to see that the interval criterion (12), which allows you to take into account the sign of the greatest divergence Δ_n, is also unable to establish the non-representative nature of the sample $\{\psi\}_n$.

It is known that reducing the risk of an erroneous decision when classifying data can be achieved by considering not only Type I Error, but also of the Type II Error [4].

The simplest solution to this problem would be to compare Δ_n between $F_\Sigma(\xi)$ and $F_s^*(\xi)$ with the boundary values of the interval $\left[\underline{\Delta_n}; \overline{\Delta_n}\right]$ corresponding to the significance level α=0,5. This is the limiting case of values α when $\Delta_n=0$. The Type II Error is β=(1-α), i.e. also equal to 0.5. If α is taken less than 0.5, then the Type II Error β increases. In real conditions:

- configurations $F_\Sigma(\xi)$ and $F_s^*(\xi)$ are different, i.e. $\Delta_n \neq 0$;
- for the same values of Δ_n, (α+β) is less than or equal to one;
- as Δ_n increases, (α+β) decreases, reaches its minimum ($\Delta_{n,opt}$) and then increases;
- if $\Delta_n < \Delta_{n,opt}$, then α>β, if $\Delta_n > \Delta_{n,opt}$, then α<β;
- the difference between α and β increases as the difference between Δ_n and $\Delta_{n,opt}$ increases.

Comparison of the implementations of Δ_n with the boundary values of $\underline{\Delta_n}$ and $\overline{\Delta_n}$, calculated respectively for $F^*(\underline{\Delta_n}) = 0{,}25$ and $F^*(\overline{\Delta_n}) = 0.75$, makes it possible not to calculate s.d.f., which determines the Type II Error β, which can be attributed to the advantages of this method. Its disadvantages are the need to double the number of simulated implementations of the distribution $F_s^*(\xi)$, the unjustified reduction of the disperse Δ_n, the heuristic approach.

Conclusions

1. Statistical data on the reliability of Electric Power System' equipment and devices are a finite population of multidimensional data. Therefore, the use of classical methods of analyzing samples from the general population for analyzing samples from multidimensional data leads to an increased risk of erroneous decisions;
2. Research of the accuracy of existing methods for modeling random variables according to s.d.f. showed that the discrepancy between the accuracy of the methods is manifested only when the number of implementations of the sample $n_s \le 20$. A new method of modeling random numbers by s.d.f. is recommended;
3. Experimental researches have established that with the significance levels of assumptions used in practice, if the critical value of Kolmogorov's statistics at significance level α is equal to the estimate of the magnitude of the largest divergence, then the significance level of this estimate will be equal to $0{,}5\alpha$;
4. Reducing the risk of an erroneous decision is achieved by taking into account the significance of the difference in the distribution of random variables of the divergence between the s.d.f. of population and data sampling;
5. Assessing the appropriateness of data classification requires the involvement of simulation modeling of realizations of random variables, involving the mathematical apparatus of the theory of testing statistical hypotheses and fiducial probabilities.

References

[1] Buslenko, N.P. Modelirovanie slozhnyh sistem / N.P. Buslenko. - Moskva: Nauka, - 1978, - 400 s. (In Russian).

[2] Gnedenko, B.V. and Belyaev, Yu.K. and Solovev A.D. (1965) Matematicheskie metodyi v teorii nadezhnosti. M. «Nauka», 524 s. (In Russian).

[3] Kendall, M.Dzh. and St'yuart A. (1966) Teoriya raspredeleniy. Per. s angl. Pod red. A.N.Kolmogorova. 587 s

[4] Ryabinin I.A. (1971) Osnovyi teorii i rascheta nadezhnosti sudovyih elektroenergeticheskih sistem. L., Sudostroenie. 454 s. (In Russian).

AN ATTRIBUTE CONTROL CHART FOR TIME TRUNCATED LIFE TESTS USING EXPONENTIATED INVERSE KUMARASWAMY DISTRIBUTION

B. Srinivasa Rao[1], M. Rami Reddy[2] and K. Rosaiah[3]

•

[1]Department of Mathematics & Humanities, R.V.R& J.C College of Engineering, Chowdavaram, Guntur- 522 019, A.P, India
[2]Freshman Engineering Department, Lakireddy Bali Reddy College of Engineering, Mylavaram-521 230, A.P, India
[3]Department of Statistics, Acharya Nagarjuna University, Guntur, A.P, India[3]

[1]boyapatisrinu1@gmail.com, [2]ramireddy.munnangi@gmail.com, [3]rosaiah1959@gmail.com[3]

Abstract

In this article an attribute control chart is designed for the Exponentiated Inverse kumaraswamy distribution under a time truncated life test by assuming the life-time of the item follow the selected Exponentiated Inverse Kumaraswamy distribution with known parameters. In order to limit the cost of checking the quality of an item in any industrial process with time truncation, this process is much useful. By considering the average number of defective items from a specified lot that are failed before the time limit, the attribute control limits are constructed. The control chart is determined using Binomial distribution based on the Upper and Lower control limits. The functioning of the designed control chart is examined with the average run length (ARL) values. The control chart constants and limits are calculated for specific ARL values with assumed parameters at different sample sizes for an in-control process. These control chart constants are obtained by considering different combinations of parameters of the assumed distribution. With these in-control limits the ARL values are observed by shifting the parameter values. A simulation analysis is developed by taking a specific number of observations in each sample and the average number of failures from each sample is considered as a statistic to establish the execution of the control chart for a specified ARL at a particular shift in parameter. With that statistic of average number of failures from the samples the control chart is prepared. It is observed a specific change in defective number when there is shift in parameter values. The results are illustrated with an example.

Keywords: Exponentiated Inverse Kumaraswamy distribution; attribute control chart; time truncated life test; average run length; simulation.

1. Introduction

To examine the quality of an article in industrial production, control charts are much helpful. Simultaneously it is very important to maintain the standards or even to improve the quality of article to meet customer satisfaction levels. It requires a regular monitoring to assess the quality of the articles. To face the competition in the market it is very important to complete this screening process with less cost and within a shortest possible time. It is a common practice that the process is considered to be in control when the examined statistic values lies within the control limits known as Upper and Lower control limits (UCL and LCL) that are also believed as the extremities for specification of the product's quality. If the sample points exceeding the limits then the production procedure is treated as intemperate and these items are considered to be defective or

imperfect. In terms of reducing the defective items or refining the quality of the items in a less span of time and with minimal cost, the control chart methods are much beneficial.

In general we have two kinds of control charts for variables and attributes. Variable control charts can be applied to any quality characteristic that is measurable. Whereas attribute control charts are useful in classifying defective and non-defective items in the production process.

There are several studies designed by various authors that the construction and implementation of different attribute and variable control charts. A few of them are Epprecht et al.[1] studied about the Adaptive control charts for attributes. Wu et al. [2] prepared an optimal np chart with curtailment. Ho and Quinino [3] discussed the monitoring process of variability through an attribute control chart. Wu and Wang [4] proposed np-control chart using double inspection. Further, some more attribute and variable controlcharts also established in Chiu and Kuo [5], A.D. Rodrigues et al. [6], Joekes and Barbosa [7], Arif et al. [8], and Shafqat et al. [9].

Generally most of the constructing processes of control charts are based on the supposition that the quality of items follow normality. While some circumstances where the its characteristic is unknown or doesn't follow the normal distribution. Various authors developed the procedure for the construction of control charts for non normal distributions, for example: Bai and choi [10], Chang and Bai, [11], Al-Oraini and Rahim [12], Aslam et al. [13] and Lin and Chou [14].

As it is essential for any industry to sustain in the competitive market by manufacturing more reliable products with less cost, with less manpower in a short period time it requires less time for inspection of defective products. To achieve this, it is necessary to have a time truncated life test based control chart. Hence, preparing a control chart for monitoring a non-normality characteristic product under the time truncated test is preferred to inspect the lifetime of the product.

Various authors proposed articles to expand the methodology of constructing control charts for various distributions under time truncated life test. A few references of such models are Aslam and Jun [15] developed a time truncated life test(TLT) based attribute control chart for Weibull distribution. Aslam et al. [16] designed a TLT based control chart for Pareto distribution of second kind. Similarly, Rao [17] proposed for exponentiated half-logistic distribution. Rosaiah et al. [18] considered for exponentiated Frechet distribution. Shruthi. G and O.S.Deepa [19] monitored ARL for Exponentiated distributions under TLT. Rao et al. [20] introduced TLT based chart for Dagum distribution. Adeoti and Ogundipe [21] developed for generalized exponential distribution under TLT. Rosaiah et al. [22] designed for type-II generalized log logistic distribution. Jafarian-Namin et al. [23] studied an efficient design of attribute control chart under TLT for weibull distribution. G. S .Rao and Al-Omari [24] designed for Length-Biased Weighted Lomax Distribution. Baklizi and Ghannam [25] proposed a TLT based attribute control chart for the inverse Weibull distribution.

Our interest is to develop an article of TLT based attribute control chart for monitoring the quality process when the lifetime of an article follow Exponentiated Inverse kumaraswamy distribution (EIKD). To monitor the functioning of a control chart ARL is general procedure which gives the average number of values that must be considered before an observation signals as out-of control. The ARL is determined as$ARL = \frac{1}{P}$; here P is the probability of any observation indicates out of control. The article is summarized in the following way: a concise introduction of the Exponentiated Inverse kumaraswamy distribution (EIKD) is provided in section-2, design and execution of control chart with calculations of ARL values when the parameter is shifted for EIKD is discussed with an application in Section-3. The process is evaluated with an analysis of simulation data in section-4. Few closing remarks are specified in Section-5.

2. Exponentiated Inverse Kumaraswamy distribution

The probability density function (pdf) of the Exponentiated inverse Kumaraswamy distribution (EIKD) is

$$f(x) = \alpha\beta\lambda(1+x)^{-(\alpha+1)}(1-(1+x)^{-\alpha})^{\beta\lambda-1} \; ; 0 < x < \infty, \;\; \alpha, \beta, \lambda > 0 \qquad (1)$$

Its cumulative distribution function (cdf) is given by,

$$F(x) = [1-(1+x)^{-\alpha}]^{\beta\lambda};\ 0<x<\infty,\ \ \alpha,\beta,\lambda>0 \tag{2}$$

Where α, β and λ are shape parameters

The mean of EIKD is $E(X) = \lambda\beta\ B\left(1-\frac{1}{\alpha},\beta\lambda\right),\quad \alpha>1$

The Reliability function is

$$R(x) = 1-[1-(1+x)^{-\alpha}]^{\beta\lambda} \tag{3}$$

and the Hazard function is

$$H(x) = \frac{\alpha\lambda\beta(1+x)^{-(\alpha+1)}[1-(1+x)^{-\alpha}]^{\beta\lambda-1}}{1-[1-(1+x)^{-\alpha}]^{\lambda\beta}},\ 0<x<\infty\ \ and\ \ \ \lambda,\alpha,\beta>0 \tag{4}$$

Graphs of the pdf, cdf, Reliability and hazard functions of EIKD for selected parameter values are plotted respectively.

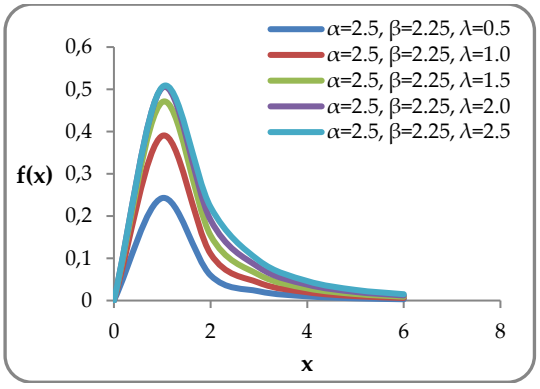


Figure 1: *The pdf plots of EIKD*

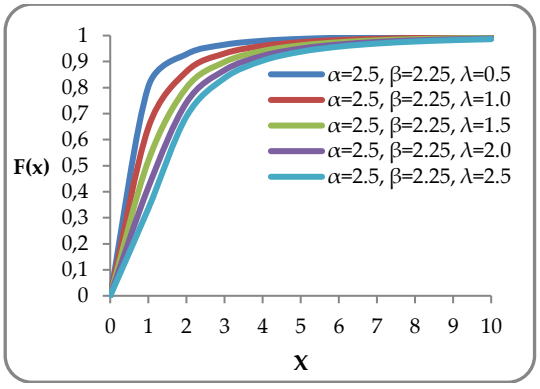


Figure 2: *The cdf plots of EIKD*

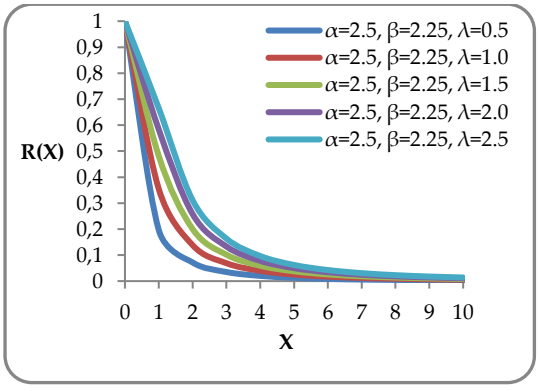


Figure 3: *Reliability function plots of EIKD*

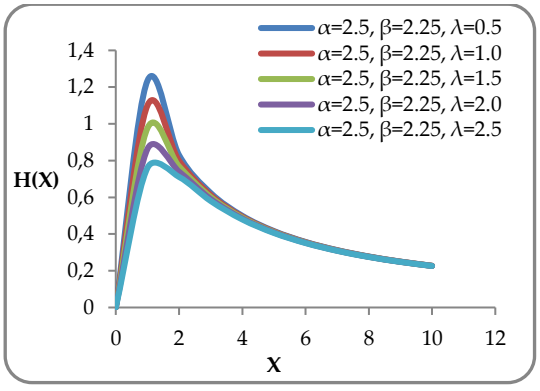


Figure 4: *Hazard function plots of EIKD*

3. Designing of the control chart

To construct the 'np' chart on the basis of defective articles in the production process the following methodology was implemented

Step1: A sample of 'n' articles is considered randomly from every subgroup lot and apply time truncated life test for these articles. Consider the number of articles (D) that are out of order (failure) within the termination time$t_0$stated as $t_0 = a\mu_0$, here $a$is constant related with target average life μ_0 when the process is supposed to be in-control.

Step2: Declare the process is under control if D rests in the limits of LCL and UCL, else, if D>UCL or D<LCL it is declared as not in control. Since D is a distinct count out of a sample of 'n' items, it can be considered as a "Binomial variate" as$n\ and\ p$ are parameters the control limits for in-control process are given as;

$$.UCL = np_0 + k\sqrt{np_0(1-p_0)} \tag{5a}$$

$$.LCL = MAX\left[0, np_0 - k\sqrt{np_0(1-p_0)}\right] \tag{5b}$$

Here$'P_0'$is the probability of an article is bungled earlier than$t_0$and it is determined from equation-(2) as"$p_0 = F(t_0)$"and k is the constant of the control chart. However, we state the process is in-control when$\mu = \mu_0$(or the parameters $\alpha = \alpha_0, \beta = \beta_0\ and\ \lambda = \lambda_0$). Then$p_0$ is obtained from equation (2) as

$$p_0 = F\left[t_0;\ \alpha_0, \beta_0, \lambda_0\right] = \left[1 - (1+t_0)^{-\alpha_0}\right]^{\beta_0\lambda_0}$$

$$= \left[1 - \left\{1 + a\ \beta_0\lambda_0 . B\left(1 - \frac{1}{\alpha_0},\ \beta_0\lambda_0\right)\right\}^{-\alpha_0}\right]^{\beta_0\lambda_0} \tag{6}$$

In real time applications, the probability$p_0$is typically unknown; then the limits for such situations are

$$UCL = \bar{D} + k\sqrt{\bar{D}\left(1 - \bar{D}/n\right)} \tag{7a}$$

$$LCL = Max\left[0,\ \bar{D} - k\sqrt{\bar{D}\left(1 - \bar{D}/n\right)}\right] \tag{7b}$$

Here$\bar{D}$is mean count of failure articles in the samples.

The probability$P_{in}^0$of confirming in control process of the planned chart is specified as

$$P_{in}^0 = \sum_{d=LCL+1}^{UCL} \binom{n}{d} (P_0)^d (1-P_0)^{n-d} \tag{8}$$

$$P_{in}^0 = P[LCL \le D \le UCL/P_0] =$$

$$\sum_{d=LCL+1}^{UCL} \binom{n}{d} \left\{\left[1 - \left[1 + a\beta_0\lambda_0 . B\left(1 - \frac{1}{\alpha_0}, \beta_0\lambda_0\right)\right]^{-\alpha_0}\right]^{\beta_0\lambda_0}\right\}^d \left\{\left[\left[1 + a\beta_0\lambda_0 . B\left(1 - \frac{1}{\alpha_0}, \beta_0\lambda_0\right)\right]^{-\alpha_0}\right]^{\beta_0\lambda_0}\right\}^{n-d} \tag{9}$$

The control charts efficacy can be examined by its "Average Run Length (ARL)", and is defined when the process is under control as

$$ARL_0 = \frac{1}{1 - P_{in}^0} \tag{10}$$

3.1. ARLs with a shift in Parameter

To examine the performance of the control chart with a shift in one of the parameter (λ) as $\lambda_1 = c\lambda_0$, here c is specified as shift constant

The probability of an article is out of order prior to the experimental time$t_0$is consider asp_1, and is obtained as

$$p_1 = F\left[t_0;\ \alpha_0, \beta_0, \lambda_1\right] = \left[1 - (1+t_0)^{-\alpha_0}\right]^{\beta_0\lambda_1}$$

$$= \left[1 - \left\{1 + a\ \beta_0\lambda_0 . B\left(1 - \frac{1}{\alpha_0},\ \beta_0\lambda_0\right)\right\}^{-\alpha_0}\right]^{\beta_0 c\lambda_0} \quad (11)$$

The in-control probability of the process with the parameter shift as

$$P_{in}^1 = \sum_{d=LCL+1}^{UCL} \binom{n}{d} (P_1)^d (1-P_1)^{n-d} \quad (12)$$

$P_{in}^1 = P[LCL \le D \le UCL/P_1]=$

$$\sum_{d=LCL+1}^{UCL} \binom{n}{d} \left\{\left[1 - \left[1 + a\beta_0\lambda_0 . B\left(1 - \frac{1}{\alpha_0}, \beta_0\lambda_0\right)\right]^{-\alpha_0}\right]^{\beta_0\lambda_1}\right\}^d \left\{\left[\left[1 + a\beta_0\lambda_0 . B\left(1 - \frac{1}{\alpha_0}, \beta_0\lambda_0\right)\right]^{-\alpha_0}\right]^{\beta_0\lambda_1}\right\}^{n-d} \quad (13)$$

The ARL for the process shift is given as

$$ARL_1 = \frac{1}{1 - P_{in}^1} \quad (14)$$

The approach for the calculations of intended chart is mentioned below

(1) Choose the ARL (sayr_0), parameter values ($\alpha_0,\ \beta_0\ and\ \lambda_0$), and the constant a.

(2) Control chart parameters to be determined for a specified sample size n, providedARL_0 which is specified in Eq-(10) very near to$r_0$that is$ARL_0 \ge r_0$.

(3) The parameters obtained in previous step are utilized to calculate ARL_1 as per the shift constant c using Eq-(14).

The control limits of the chart are determined for various parameter values and$r_0$values that are shown in the Tables 1 through 8.We have noticed a rapid reducing tend in ARL_1values with the decrement in shift value 'c'.

Table 1: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=20;* $\boldsymbol{\lambda_0 = 1.5, \alpha_0 = 2.5\ and\ \beta_0 = 2.5}$

L C L	2	1	1	2
U C L	14	13	14	15
a	0.3224	0.2891	0.3246	0.3512
k	2.8561	2.9154	2.9628	2.9965
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.088	250.258	300.372	370.897
0.9	118.620	120.348	119.549	178.751
0.8	49.315	46.023	46.680	71.215
0.7	20.178	18.085	19.162	28.765
0.6	8.803	7.716	8.452	12.287
0.5	4.212	3.674	4.090	5.664
0.4	2.273	2.014	2.231	2.888
0.3	1.434	1.319	1.420	1.686
0.2	1.095	1.058	1.091	1.178
0.1	1.005	1.002	1.004	1.014

Table 2: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=20;* $\boldsymbol{\lambda_0 = 1.5, \alpha_0 = 2\ and\ \beta_0 = 2.5}$

LCL	5	1	0	1
UCL	17	13	12	14
a	0.4858	0.3015	0.2724	0.3328
k	2.8126	2.8682	2.9864	2.9962
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.071	250.140	300.364	370.482
0.9	162.945	120.234	111.599	149.679
0.8	84.533	45.983	40.166	56.697
0.7	39.198	18.072	15.478	22.461
0.6	18.309	7.711	6.580	9.563
0.5	8.865	3.672	3.172	4.471
0.4	4.520	2.014	1.792	2.362
0.3	2.479	1.319	1.229	1.462
0.2	1.514	1.058	1.034	1.101
0.1	1.094	1.002	1.001	1.005

Table 3: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=20;* $\boldsymbol{\lambda_0 = 2,\ \alpha_0 = 2\ and\ \beta_0 = 3}$

LCL	5	1	1	0
UCL	17	13	14	13
a	0.5421	0.351	0.3909	0.3498
k	2.6423	2.8645	2.9641	3.0125
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.555	250.200	300.514	370.295
0.9	161.470	120.292	119.607	131.616
0.8	83.529	46.003	46.699	47.893
0.7	38.777	18.079	19.169	18.611
0.6	18.146	7.714	8.455	7.885
0.5	8.803	3.673	4.091	3.730
0.4	4.497	2.014	2.231	2.033
0.3	2.471	1.319	1.420	1.325
0.2	1.512	1.058	1.091	1.059
0.1	1.094	1.002	1.004	1.002

Table 4: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=20;* $\boldsymbol{\lambda_0 = 2.5,\ \alpha_0 = 2.5\ and\ \beta_0 = 3}$

LCL	2	0	1	0
UCL	14	12	14	12
a	0.4198	0.3598	0.4221	0.3488
k	2.7562	2.8163	2.9564	2.9688
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.081	250.221	300.745	370.192
0.9	118.599	92.574	119.700	141.897
0.8	49.306	34.284	46.731	49.380
0.7	20.175	13.640	19.179	18.264
0.6	8.802	5.984	8.458	7.455
0.5	4.211	2.972	4.092	3.456
0.4	2.273	1.726	2.231	1.885
0.3	1.434	1.210	1.420	1.256
0.2	1.095	1.031	1.091	1.039
0.1	1.005	1.001	1.004	1.001

Table 5: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=30;* $\boldsymbol{\lambda_0 = 1.5,\ \alpha_0 = 2.5\ and\ \beta_0 = 2.25}$

LCL	9	3	6	8
UCL	24	18	22	24
a	0.4381	0.2897	0.2685	0.4246
k	2.7864	2.8636	2.9238	2.9817
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.112	250.372	301.927	370.298
0.9	166.463	90.656	122.171	270.239
0.8	69.870	28.891	43.166	101.717
0.7	26.828	10.262	16.265	36.301
0.6	10.954	4.257	6.789	13.868
0.5	4.923	2.138	3.227	5.846
0.4	2.509	1.344	1.806	2.805
0.3	1.507	1.067	1.232	1.597
0.2	1.111	1.004	1.034	1.132
0.1	1.006	1.000	1.001	1.007

Table 6: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=30;* $\boldsymbol{\lambda_0 = 1.5,\ \alpha_0 = 2\ and\ \beta_0 = 2.5}$

LCL	9	3	2	4
UCL	24	18	17	20
a	0.4657	0.3021	0.27939	0.3409
k	2.7951	2.8783	2.9438	2.9857
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.157	250.489	300.090	370.029
0.9	166.398	90.720	96.226	136.724
0.8	69.836	28.908	29.187	43.416
0.7	26.817	10.267	10.039	15.051
0.6	10.951	4.258	4.087	5.950
0.5	4.922	2.138	2.045	2.774
0.4	2.509	1.344	1.301	1.584
0.3	1.507	1.067	1.053	1.141
0.2	1.111	1.004	1.003	1.014
0.1	1.006	1.000	1.000	1.000

Table 7: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=30;* $\boldsymbol{\lambda_0 = 2,\ \alpha_0 = 2\ and\ \beta_0 = 3}$

LCL	9	2	1	5
UCL	24	17	16	21
a	0.5207	0.3327	0.31	0.415
k	2.7642	2.8645	2.9368	2.9864
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.133	250.941	300.492	370.132
0.9	166.433	78.483	84.867	165.081
0.8	69.854	24.700	25.461	53.743
0.7	26.823	8.846	8.813	18.549
0.6	10.952	3.744	3.651	7.198
0.5	4.922	1.941	1.881	3.245
0.4	2.509	1.271	1.241	1.766
0.3	1.507	1.047	1.038	1.202
0.2	1.111	1.002	1.001	1.025
0.1	1.006	1.000	1.000	1.000

Table 8: $\boldsymbol{ARL_1}$ *Values for the designed chart with n=30;* $\boldsymbol{\lambda_0 = 2.5,\ \alpha_0 = 2.5\ and\ \beta_0 = 3}$

LCL	9	3	4	8
UCL	24	18	20	24
a	0.5421	0.3848	0.43039	0.5279
k	2.7123	2.8215	2.9452	2.9938
	ARLo=200	ARLo=250	ARLo=300	ARLo=370
C	ARL_1	ARL_1	ARL_1	ARL_1
1	200.041	250.287	300.702	370.128
0.9	166.565	90.609	103.844	270.558
0.8	69.926	28.878	34.402	101.855
0.7	26.846	10.259	12.545	36.343
0.6	10.960	4.256	5.211	13.880
0.5	4.925	2.137	2.545	5.849
0.4	2.510	1.344	1.513	2.806
0.3	1.507	1.067	1.122	1.598
0.2	1.111	1.004	1.012	1.132
0.1	1.006	1.000	1.000	1.007

3.2 Application of intended chart

To establish the applicability of the intended chart for the improvement of the quality of any manufactured article, we assume that the lifespan of the article follows the EIKD with parameters $\alpha_0 = 2.5, \beta_0 = 2.25\ and\ \lambda_0 = 1.5.$ Consider the aimed mean life of the article is set as 1000 hrs with size of the sample n=20 of each group. If the target in control ARL value is fixed as$r_0 = 300$for the control chart, as shown in Table 1, we obtain a= 0.3246, k=2.9628. From Equation-(6) we get $p_0 = 0.4221$. Using Equations-(5a & 5b) the LCL and UCL are determined as LCL =1 and UCL = 14. Then, the functioning of the prepared control chart is as follows:

Step 1: Take a sample of 20 articles from each subgroup and test their lifespan for period of 324.6 hours. Figure out the failure count of articles (D) through the test.

Step 2: We assert the process is in-control if 1 ≤ D ≤ 14, if not it is not in control.

4. Simulation study

A simulation study is presented to monitor the applicability of the designed control chart. It is executed using EIKD with specified parameters and the construction process is as follows:

The data is originated using EIKD with parameters $\alpha_0 = 2,\ \beta_0 = 2.5\ and\ \lambda_0 = 1.5.$A subgroup of 15 samples of size n=20 each are taken by considering the ARL as$r_0 = 300$.The process is affirmed as in control with these parameter values when$\mu_0 = 3.5483$.Another subgroup of 15 samples each of size n=20 are taken from EIKD with shift in parameter$\lambda_1 = c\lambda_0$with shift value c =0.7. The control chart coefficient k=2.9864 is considered from Table 2, with ARL value as $r_0 = 300$ and n=20 when the process is in control.

The termination time of the life-test will be$t_0 = a\mu_0 = 0.2724 \times 3.5483 = 0.9665$. The number of articles that are failed before the termination time$t_0$is considered as D which is calculated and presented in Table 9 for each sample. The number of average failures is$\overline{D} = 6$, the control limits are determined from equations (7a & 7b) are UCL=12 and LCL=0. The points of failure count (D) of each sample are exhibited in Figure 5. It is clearly observed that the planned chart indicated the shift at 18th sample (3rd sample after change in shift) while the corresponding

ARL value is 15. Thus the proposed chart effectively identifies the shift in this process.

5. Conclusion

In this article, we have designed a new ' np' control chart by assuming lifetime of the product follows EIKD to monitor the quality of manufactured articles under time truncation. The designed chart is then assessed by ARL's acquired from simulation study for different sample sizes; parameter values and objective in-control ARLs have been considered. The functioning of the intended chart is described with an explanatory example. For advance research, anyone can review applying the suggested control chart for any significant lifetime distribution. We have the feasibility to consider more quickened testing design to develop appropriate control charts for such circumstances.

Table 9: *The simulation Analysis*

	sample																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	D
1	2.10	3.23	1.14	0.53	2.34	2.41	0.84	0.06	2.06	3.02	2.27	0.45	0.52	3.45	3.21	2.51	2.11	0.15	1.56	1.33	6
2	2.79	0.30	3.43	0.94	0.58	0.17	3.25	1.83	0.85	0.38	1.44	2.11	2.09	0.34	2.60	2.52	0.73	1.91	0.31	2.12	9
3	1.18	2.15	3.23	1.61	0.08	3.37	1.51	2.20	2.84	2.84	0.52	2.69	0.94	2.10	3.52	3.30	1.82	2.98	1.28	0.52	4
4	3.38	1.38	3.03	0.10	0.88	0.35	0.91	0.54	2.80	2.49	0.64	2.73	1.56	2.54	2.64	1.22	1.43	3.46	0.77	3.24	7
5	3.25	0.64	2.58	0.69	0.36	0.81	3.55	2.96	1.15	1.07	1.07	2.01	3.00	1.40	2.68	0.88	0.69	2.57	1.74	3.25	6
6	1.74	0.11	2.75	0.05	1.55	2.19	0.28	1.78	2.05	2.11	2.30	2.83	0.84	0.03	1.31	1.60	0.34	2.75	1.28	2.28	6
7	3.35	2.75	1.77	0.32	2.32	3.27	2.15	1.76	1.65	1.07	0.17	2.60	1.33	3.36	0.53	2.39	2.90	3.14	3.54	0.44	4
8	2.32	1.80	0.74	3.31	1.48	3.35	0.38	2.65	0.97	1.42	1.86	3.03	1.73	0.03	3.55	0.21	1.98	0.40	0.99	2.38	5
9	1.23	1.05	3.49	3.45	0.49	1.87	1.97	2.78	2.67	1.76	1.96	1.12	2.32	1.43	3.04	1.46	2.34	0.62	3.15	1.10	2
10	2.57	0.08	0.51	0.32	2.34	3.34	0.13	0.09	3.16	2.60	0.57	1.12	1.21	2.31	2.57	2.71	1.34	0.85	1.66	0.52	8
11	2.65	2.46	2.20	2.70	2.64	0.13	1.10	1.54	1.93	0.19	3.26	2.41	2.40	1.84	0.20	1.58	2.89	0.18	0.84	2.53	5
12	1.17	3.47	1.14	2.79	1.30	1.64	0.34	2.53	3.30	2.31	3.18	3.30	1.93	2.75	2.64	1.33	1.00	2.49	0.61	2.96	2
13	1.18	0.58	1.24	0.61	2.04	2.93	1.52	2.79	2.60	1.00	1.72	2.39	3.10	1.96	1.32	2.87	0.62	2.25	1.76	1.63	3
14	2.88	2.56	1.20	2.04	1.54	1.54	1.81	0.43	0.25	3.02	0.41	1.61	2.06	1.21	2.84	3.36	1.71	0.69	1.83	2.09	4
15	2.23	2.80	3.48	0.95	3.34	2.70	2.64	3.13	2.66	0.99	0.06	0.28	2.66	2.00	2.27	2.09	0.56	3.37	0.04	1.95	5
16	1.63	1.61	0.46	2.62	2.37	0.45	0.63	1.96	2.03	1.30	0.92	2.73	1.71	0.26	1.58	2.08	1.30	0.32	2.84	1.82	6
17	2.12	0.04	0.22	0.05	0.60	2.12	0.94	1.23	2.58	1.21	1.37	0.95	2.19	0.30	1.20	1.85	0.53	0.39	0.93	1.46	10
18	0.93	0.43	0.94	0.95	0.86	0.37	0.65	0.98	1.60	2.93	0.17	2.51	0.94	0.70	0.87	1.19	2.17	1.66	0.58	0.25	13
19	2.51	1.74	0.31	2.66	1.58	0.57	1.58	0.80	2.33	2.90	1.34	2.04	1.41	1.77	2.82	0.13	2.81	1.82	0.34	2.58	5
20	1.67	1.34	2.03	0.09	1.06	2.78	1.98	2.22	0.81	2.24	1.48	0.19	0.13	0.85	1.35	1.90	0.16	1.04	1.31	2.93	6
21	1.16	1.00	0.57	2.45	2.17	2.91	0.15	2.08	1.70	2.94	0.50	1.62	2.04	1.47	0.86	2.75	2.36	1.61	0.69	2.68	5
22	2.46	0.60	2.16	1.92	1.30	1.85	2.61	0.53	0.40	2.45	1.78	2.85	1.45	2.54	0.43	0.63	2.63	1.14	1.72	2.24	5
23	0.99	0.25	2.70	2.79	2.41	2.95	1.26	1.82	2.30	2.06	0.88	1.18	2.45	2.53	0.45	2.36	2.09	0.18	0.34	2.95	5
24	2.12	1.94	1.25	2.93	0.33	2.31	2.35	0.86	2.54	2.21	0.86	0.22	0.54	1.56	0.97	0.40	1.36	2.54	1.51	1.92	6
25	2.10	0.03	1.26	1.60	0.69	0.16	1.06	0.51	0.95	0.94	0.95	1.92	0.59	0.70	2.65	2.39	0.56	0.10	0.54	2.53	12
26	1.68	1.18	2.49	2.17	0.67	0.61	0.24	0.48	1.28	2.68	0.98	1.07	2.97	0.73	2.14	0.75	2.37	1.81	1.46	1.50	6
27	2.81	1.35	1.02	1.20	1.76	0.85	0.88	2.29	2.27	2.44	1.44	0.73	0.28	2.75	1.40	0.65	1.42	2.37	1.10	1.94	5
28	0.05	2.43	2.41	0.47	2.15	2.54	0.23	1.57	1.39	0.00	1.70	1.37	1.10	2.03	2.48	0.43	0.75	0.80	0.06	0.17	9
29	2.99	2.57	1.53	2.41	2.16	0.99	0.44	1.77	2.12	0.95	1.03	1.96	2.72	0.22	2.01	1.88	1.59	1.22	2.19	0.33	4
30	2.35	0.50	2.97	0.13	2.69	2.30	0.64	2.77	1.70	2.56	1.03	0.08	2.93	0.83	0.17	2.91	0.39	2.65	2.28	1.03	7

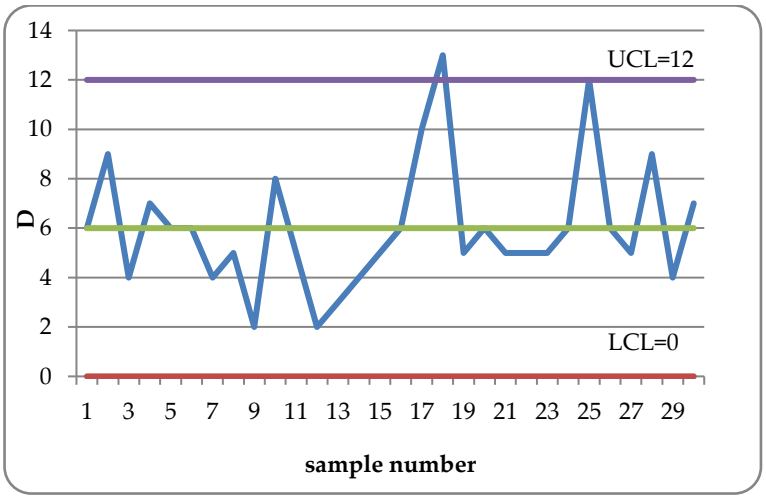


Figure 5: ***Control chart for simulation data.***

References

[1] E. K. Epprecht, A. F. B. Costa and F. C. T. Mendes (2003). Adaptive control charts for attributes., *IIE Transactions*,35(6): 567–582.

[2] Wu Z, Luo H and Zhang X. (2006). Optimal np control chart with curtailment.*Eur J Oper Res*, 174: 1723–1741.

[3] L. Lee Ho and R. C. Quinino (2013). An attribute control chart for monitoring the variability of a process.*International Journal of Production Economics*, 145(1): 263–267.

[4] Wu and Wang Q. (2007). An np control chart using double inspections.*J Appl Stat*, 34:843–855.

[5] J.E. Chiu and T. I. Kuo. (2007). Attribute control chart for multivariate Poisson distribution. *Communications in Statistics- Theory and Methods*, 37(1): 146–158.

[6] A.D. Rodrigues, E.K. Epprecht and M.S. De Magalhaes. (2011). Double-sampling control charts for attributes.*Journal of Applied Statistics*, 38: 87-112.

[7] Joekes S, Barbosa E.P. (2013). An improved attribute control chart for monitoring non-conforming proportion in high quality processes.*Control Eng Pract*, 21:407–412.

[8] O. H. Arif, M. Aslam, and C.H. Jun. (2017). EWMA np control chart for the Weibull distribution.*Journal of Testing and Evaluation*, 45(3): 1022–1028.

[9] Shafqat, A., Hussain, J., Al-Nasser, A. D., & Aslam, M. (2017). Attribute control chart for some popular distributions.*Communications in Statistics - Theory and Methods*, 47(8): 1978–1988.

[10] Bai D.S and Choi I.S, $\overline{X}$ and R control charts for skewed populations. (1995), *J Qual Tech*, 27: 120–131.

[11] Y.S. Chang and D.S. Bai. (2001). Control charts for positively skewed populations with weighted standard deviations.*Qual. Reliab, Eng. Int*, 17: 397-406.

[12] H.A. Al-Oraini and M.A. Rahim. (2003). Economic statistical design of control charts for systems with gamma (λ, 2) in-control times.*Journal of Applied Statistics*, 30: 397-409.

[13] Aslam M, Azam M, Khan N and Jun CH. (2015). A control chart for an exponential distribution using multiple dependent state sampling. *Qual. Quant*, 49: 455–462.

[14] Y.C. Lin and C.Y. Chou. (2007). Non-normality and the variable parameters X bar control charts.*J Oper Res Soc*, 176: 361- 373.

[15] M. Aslam and C.H. Jun. (2015). Attribute control charts for the Weibull distribution under truncated life tests.*Quality Engineering*, 27(3): 283–288.

[16] M. Aslam, N. Khan, and C. H. Jun, (2016). A control chart for time truncated life tests using Pareto distribution of the second kind.*Journal of Statistical Computation and Simulation*, 86(11): 2113–2122.

[17] G. S. Rao. (2018).A control chart for time truncated life tests using exponentiated half logistic distribution.*Applied Mathematics & Information Sciences*, 12(1): 125–131.

[18] K. Rosaiah, G. S. Rao and M. S. Babu. (2018).An attribute control chart under truncated life test for the exponentiated Frechet distribution. *Quality-Access to Success*, 19(163).

[19] Shruthi. G and O.S.Deepa. (2018). Average Run Length for Exponentiated distribution under truncated life test. *International Journal of Mechanical Engineering and Technology*, 9(6):1180-1188.

[20] G. S. Rao, A. K. Fulment and P. K. Josephat. (2019). Attribute control chart for the Dagum distribution under truncated life tests. *Life Cycle Reliability and Safety Engineering*, 8: 329–335.

[21] O. A. Adeoti and P. Ogundipe. (2021). A control chart for the generalized exponential distribution under time truncated life test. *Life Cycle Reliability and Safety Engineering*, 10(1): 53–59.

[22] K. Rosaiah, G. S. Rao, and S.V.S. Prasad. (2021). A control chart for time truncated life test using type II generalized log-logistic distribution.*Biometrics and Biostatistics International Journal*, 10(4): 138–143.

[23] S. Jafarian-Namin, M. Aslam, M. S. Fallah Nezhad and F. Eskandari-Kataki. (2021). Efficient designs of modeling attribute control charts for a Weibull distribution under truncated life tests. *Opsearch*, 58(4): 942–961.

[24] G. S .Rao, and Amer Ibrahim Al-Omari. (2022). Attribute Control Charts Based on TLT for Length-Biased Weighted Lomax Distribution. *Hindawi , Journal of Mathematics*, ID 3091850.

[25] Baklizi, A., Ghannam, S.A. (2022). An attribute control chart for the inverse Weibull distribution under truncated life tests. *HELIYON*, 8(2022): 1-5.

A TYPE I HALF -LOGISTIC EXPONENTIATED WEIBULL DISTRIBUTION: PROPERTIES AND APPLICATIONS

Olalekan Akanji Bello [1*]
Sani Ibrahim Doguwa[1]
Abukakar Yahaya[1]
Haruna Mohammed Jibril[2]

•

Department of Statistics, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria[1].
Department of Mathematics, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria[2].
olalekan4sure@gmail.com[1*], sidoguwa@gmail.com[1], ensiliyu2@yahoo.co.uk[1]
alharun2004@yahoo.com[2]

Abstract

In the area of distribution theory, statisticians have proposed and developed new models for generalizing the existing ones so as to make them more flexible and to aid their application in a variety of fields. In this article, we present a new distribution called the Type I Half-Logistic Exponentiated Weibull (TIHLEtW) Distribution with four positive parameters, which extends the Weibull distribution by two parameters. Some statistical properties of the TIHLEtW distribution, such as explicit expressions for the quantile function, probability weighted moments, moments, generating function, Reliability function, hazard function, and order statistics are discussed. A maximum likelihood estimation technique is employed to estimate the model parameters and the simulation study is presented. The superiority of the new distribution is illustrated with an application to two real data sets. The results showed that the new distribution fits better in the two real data sets amongst the range of distributions considered.

Keywords: Type I Half-Logistic Exponentiated-G, Weibull distribution, Quantile function, Reliability function, Maximum likelihood, Order Statistics.

1. Introduction

All parametric statistical techniques, such as inference, modeling, survival analysis, and reliability, are based on statistical distributions. Fitting the data to a statistical model is a critical step when analyzing lifetime data. For this reason, a number of lifespan distributions have been established in the literature. The majority of lifespan models have a limited set of behaviors. Such distributions are unable to provide a better fit for all real scenarios. As a result, a variety of distribution classes have been created by expanding common continuous distributions. The generated family of continuous distributions is a new enhancement for developing and expanding classic distributions. The newly generated distributions have been extensively researched in a variety of fields, and they provide greater application flexibility.

One of the most well-known lifetime distributions is the Weibull distribution. It adequately represents many distinct forms of failures, both in components and in general. To deal with bathtub-shaped failure rates, various generalizations and extensions of the Weibull distribution have been proposed in the statistical literature. Mudholkar *et al*. [25] pioneered *Exponentiated Weibull* distribution, the modified Weibull extension by Xie *et al*. [32] , flexible Weibull extension (FWEx) by Bebbington *et al.* [10], beta modified Weibull by Silva *et al.* [30], Kumaraswamy Weibull by Cordeiro *et al.* [13], transmuted Weibull by Aryal and Tsokos [8], truncated Weibull distribution by Zhang and Xie [34], Kumaraswamy inverse Weibull by Shahbaz *et al.* [29], exponentiated generalized Weibull by Cordeiro *et al.* [15], McDonald modified Weibull by Merovci and Elbatal [24], beta inverse Weibull by Hanook *et al.* [19], transmuted additive Weibull by Elbatal and Aryal [17], McDonald Weibull by Cordeiro *et al.* [12], Kumaraswamy modified Weibull by Cordeiro *et al.* [14], transmuted complementary Weibull geometric by Afify *et al.* [1], Kumaraswamy transmuted exponentiated additive Weibull by Nofal *et al.* [27], generalized transmuted Weibull by Nofal *et al.* [26], Topp-Leone generated Weibull by Aryal *et al.*,[9], Kumaraswamy complementary Weibull geometric by Afify *et al.* [2], Marshall-Olkin additive Weibull by Afify *et al.* [3], Zubair–Weibull by Ahmad [4], alpha power transformed Weibull by Ahmad *et al.* [5], Topp Leone exponentiated weibull by Ibrahim [20] distributions.

Bello *et al.* [11] proposed a new distribution family called the Type I Half-Logistic Exponentiated-G (TIHLEt-G) with two extra shape parameters. For any arbitrary cumulative distribution function as a baseline (cdf) $H(x,\vartheta)$, the TIHLEt-G family with two positive shape parameters λ and α has cumulative distribution function (cdf) and probability density function (pdf) given by

$$F_{TIHLEt-G}(x;\lambda,\alpha,\vartheta)=\frac{1-[1-H^{\alpha}(x;\vartheta)]^{\lambda}}{1+[1-H^{\alpha}(x;\vartheta)]^{\lambda}},\ x>0,\ \lambda,\alpha>0 \tag{1}$$

and

$$f_{TIHLEt-G}(x;\lambda,\alpha,\vartheta)=\frac{2\lambda\alpha h(x;\vartheta)H_{(x;\vartheta)}^{\alpha-1}[1-H_{(x;\vartheta)}^{\alpha}]^{\lambda-1}}{\left[1+[1-H_{(x;\vartheta)}^{\alpha}]^{\lambda}\right]^{2}},\ x>0,\ \lambda,\alpha>0 \tag{2}$$

The cdf and pdf of the Weibull distribution are given as

$$H(x;\theta,\beta)=1-e^{-\theta x^{\beta}},\ x>0,\ \theta,\beta>0 \tag{3}$$

$$h(x;\theta,\beta)=\theta\beta x^{\beta-1}e^{-\theta x^{\beta}},\ x>0,\ \theta,\beta>0 \tag{4}$$

The goal of this paper is to develop a more flexible model by extending the two parameter Weibull distribution. The Type II half logistic Weibull (TIHLEtW) distribution is the name given to the new model. We develop the TIHLEtW distribution from Bello *et al*. [11] and provide some essential statistical properties. The layout of this paper is organized as follows: Section 2 defines the TIHLEtW distribution. We obtained very useful and important representations for the TIHLEtW distribution in Section 3. Section 4, some statistical properties such as probability weighted moments, moments, moments generating function, quartile function, reliability function, hazard function and order statistics are derived. The parameters of the new model were estimated using the maximum likelihood estimation (MLEs) approach in Section 5. The simulation study was conducted to show that the estimates are efficient and consistent using MLE in Section 6. The application of the new model to two real data sets was shown in Section 7 to demonstrate the use of the new model. Finally Section 8 concludes the paper.

2. Type I Half-Logistic Exponentiated Weibull (TIHLETW) Distribution

In this section, we define a new model called TIHLEtW model, the random variable *X is* said to have a TIHLEtW model, if its cdf is obtained by inserting equation (3) in equation (1) as follows

$$F_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)=\frac{1-\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda}},\ x>0,\lambda,\alpha,\theta,\beta>0 \tag{5}$$

and its corresponding pdf is

$$f_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)=\frac{2\lambda\alpha\theta\beta x^{\beta-1}e^{-\theta x^{\beta}}\left[1-e^{-\theta x^{\beta}}\right]^{\alpha-1}\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda-1}}{\left[1+\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda}\right]^{2}} \tag{6}$$

where β is a scale parameter and λ,α,θ are shape parameters.

3. Important Representation

In this section, we derived a useful representation for the TIHLEtW pdf and cdf. Due to the fact that the generalized binomial series is

$$(1+Z)^{-\beta}=\sum_{i=0}^{\infty}(-1)^{i}\binom{\beta+i-1}{i}z^{i} \tag{7}$$

For $|z|<1$ and β is a positive real non integer. The density function of the TIHLEtW distribution is then obtained by using the binomial theorem (7) to (6).

$$f_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)=2\lambda\alpha\theta\beta x^{\beta-1}e^{-\theta x^{\beta}}\sum_{i=0}^{\infty}(-1)^{i}\binom{1+i}{i}\left[1-e^{-\theta x^{\beta}}\right]^{\alpha-1}\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda(i+1)-1}$$

Now, using the generalized binomial theorem, we can write

$$\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda(i+1)-1}=\sum_{j=0}^{\infty}(-1)^{j}\binom{\lambda(i+1)-1}{j}\left[1-e^{-\theta x^{\beta}}\right]^{\alpha j}$$

Also

$$\left[1-e^{-\theta x^{\beta}}\right]^{\alpha(j+1)-1}=\sum_{k=0}^{\infty}(-1)^{k}\binom{\alpha(j+1)-1}{k}\left(e^{-\theta x^{\beta}}\right)^{k}$$

Then, the pdf can be written as:

$$f_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)=\sum_{i,j,k=0}^{\infty}\eta_{p}x^{\beta-1}\left(e^{-\theta x^{\beta}}\right)^{k+1} \tag{8}$$

where $\eta_{p}=2\lambda\alpha\theta\beta(-1)^{i+j+k}\binom{i+1}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}$

In addition, an expansion for the $\left[F(x,\lambda,\alpha,\theta,\beta)\right]^{h}$ is produced, with h being an integer, and the binomial expansion is worked out once more.

$$\left[F_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)\right]^h = \underbrace{\left[1-\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^\lambda\right]^h}_{A}\underbrace{\left[1+\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^\lambda\right]^{-h}}_{B}$$

$$A=\left[1-\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^\lambda\right]^h=\sum_{m=0}^{h}(-1)^m\binom{h}{m}\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^{\lambda m}$$

$$B=\left[1+\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^\lambda\right]^{-h}=\sum_{p=0}^{h}(-1)^p\binom{h+p-1}{m}\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^{\lambda p}$$

Combining A and B, we obtain

$$\left[F_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)\right]^h=\sum_{p,m=0}^{h}(-1)^{p+m}\binom{h}{m}\binom{h+p-1}{p}\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^{\lambda(p+m)}$$

Consider

$$\left[1-\left[1-e^{-\theta x^\beta}\right]^\alpha\right]^{\lambda(p+m)}=\sum_{z=0}^{\infty}(-1)^z\binom{\lambda(p+m)}{z}\left[1-e^{-\theta x^\beta}\right]^{\alpha z}$$

$$\left[1-e^{-\theta x^\beta}\right]^{\alpha z}=\sum_{q=0}^{\infty}(-1)^q\binom{\alpha z}{q}(e^{-\theta x^\beta})^q$$

The cdf can be written as:

$$\left[F_{TIHLEtW}(x;\lambda,\alpha,\theta,\beta)\right]^h=\sum_{p,m=0}^{h}\varphi_t(e^{-\theta x^\beta})^q \tag{9}$$

where $\varphi_t=\sum_{z,q=0}^{\infty}(-1)^{p+m+z+q}\binom{h+p-1}{p}\binom{h}{m}\binom{\lambda(m+p)}{z}\binom{\alpha z}{q}$

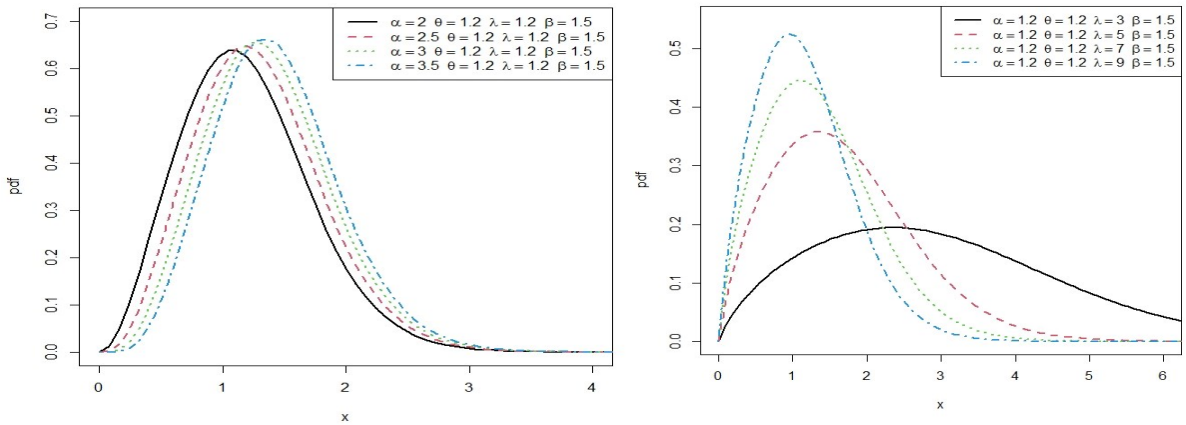


Figure 1: Plots of Pdf of TIHLEtW distribution for different values of parameters.

4. Statistical Properties

In this section, we derived some statistical properties of the new of distribution.

4.1. Probability weighted moments

Greenwood et al. [18] introduced a class of moments known as probability weighted moments (PWMs). This class is used to derive inverse form estimators for the parameters and quantiles of a

distribution. The PWMs, represented by $\tau_{r,s}$, can be derived for a random variable X using the following relationship.

$$\tau_{r,s} = E\left[X^r F(X)^s\right] = \int_{-\infty}^{\infty} x^r f(x)(F(x))^s dx \tag{10}$$

The PWMs of TIHLEtW distribution is derive by substituting (8) and (9) into (10), and replacing h with s, as follows

$$\tau_{r,s} = \sum_{i,j,k=0}^{\infty} \sum_{p,m=0}^{s} \eta_p \varphi_t \int_0^{\infty} x^r (e^{-\theta x^{\beta}})^{k+1+q} dx \tag{11}$$

Consider the integral

$$\int_0^{\infty} x^r e^{-(k+1+q)\theta x^{\beta}} dx$$

Let $y = (k+1+q)\theta x^{\beta} \Rightarrow x = \left[\frac{y}{(k+1+q)\theta}\right]^{\frac{1}{\beta}}; dx = \frac{dy}{(k+1+q)\theta\beta x^{\beta-1}}$

Then

$$\int_0^{\infty} \left[\frac{y}{(k+1+q)\theta}\right]^{\frac{r}{\beta}} e^{-y} \frac{dy}{(k+1+q)\theta\beta x^{\beta-1}} = \frac{1}{\left[(k+1+q)\theta\right]^{\frac{r}{\beta}+1}} \int_0^{\infty} y^{\frac{r}{\beta}} e^{-y} dy$$

$$\int_0^{\infty} y^{\frac{r}{\beta}} e^{-y} dy = \Gamma\left(\frac{r}{\beta}+1\right)$$

The PWMs of TIHLEtW can be written as follows

$$\tau_{r,s} = \sum_{i,j,k=0}^{\infty} \sum_{p,m=0}^{s} \frac{\eta_p \varphi_t \Gamma\left(\frac{r}{\beta}+1\right)}{(k+1+q)^{\frac{r}{\beta}+1} \theta^{\frac{r}{\beta}}} \tag{12}$$

Now,

$$\varphi_t = \sum_{z,q=0}^{\infty} (-1)^{p+m+z+q} \binom{s+p-1}{p} \binom{s}{m} \binom{\lambda(m+p)}{z} \binom{\alpha z}{q}$$

and

$$\eta_p = 2\lambda\alpha(-1)^{i+j+k} \binom{i+1}{i} \binom{\lambda(i+1)-1}{j} \binom{\alpha(j+1)-1}{k}$$

4.2. Moments

Since the moments are necessary and important in any statistical analysis, especially in applications. Therefore, we derive the r[th] moment for the new distribution.

$$\mu_r^{'} = E(x^r) = \int_0^{\infty} x^r f(x)dx \tag{13}$$

By using the important representation of the pdf in equation (8), we have

$$E(X^r) = \sum_{i,j,k=0}^{\infty} \eta_p \int_0^{\infty} x^r (e^{-\theta x^{\beta}})^{k+1} dx \tag{14}$$

Consider the integral

$$\int_0^{\infty} x^r e^{-(k+1)\theta x^{\beta}} dx$$

Let $w = (k+1)\theta x^{\beta} \Rightarrow x = \left[\frac{w}{(k+1)\theta}\right]^{\frac{1}{\beta}}; dx = \frac{dw}{(k+1)\theta\beta x^{\beta-1}}$

Then

$$\int_0^{\infty} \left[\frac{w}{(k+1)\theta}\right]^{\frac{r}{\beta}} e^{-w} \frac{dw}{(k+1)\theta\beta x^{\beta-1}} = \frac{1}{\left[(k+1)\theta\right]^{\frac{r}{\beta}+1}} \int_0^{\infty} w^{\frac{r}{\beta}} e^{-w} dw$$

$$\int_0^{\infty} w^{\frac{r}{\beta}} e^{-w} dw = \Gamma\left(\frac{r}{\beta}+1\right)$$

The r[th] moment for TIHLEtW distribution can be written as follows

$$E(X^r) = \sum_{i,j,k=0}^{\infty} \frac{\eta_p \Gamma\left(\frac{r}{\beta}+1\right)}{(k+1)^{\frac{r}{\beta}+1} \theta^{\frac{r}{\beta}}} \tag{15}$$

Now

$$\eta_p = 2\lambda\alpha(-1)^{i+j+k} \binom{i+1}{i} \binom{\lambda(i+1)-1}{j} \binom{\alpha(j+1)-1}{k}$$

The mean and variance of TIHLEtW distribution are as follows

$$E(X) = \sum_{i,j,k=0}^{\infty} \frac{\eta_p \Gamma\left(\frac{1}{\beta}+1\right)}{(k+1)^{\frac{1}{\beta}+1} \theta^{\frac{1}{\beta}}} \tag{16}$$

and

$$\operatorname{var}(X) = \sum_{i,j,k=0}^{\infty} \frac{\eta_p \Gamma\left(\frac{2}{\beta}+1\right)}{(k+1)^{\frac{2}{\beta}+1} \theta^{\frac{2}{\beta}}} - \left[\sum_{i,j,k=0}^{\infty} \frac{\eta_p \Gamma\left(\frac{r}{\beta}+1\right)}{(k+1)^{\frac{r}{\beta}+1} \theta^{\frac{r}{\beta}}}\right]^2 \tag{17}$$

4.3. Moment generating function (mgf)

The Moment Generating Function of x is given as:

$$M_x(t) = E\left(e^{tx}\right) = \int_0^{\infty} e^{tx} f(x) dx \tag{18}$$

where the expansion of $e^{tx} = \sum_{m=0}^{\infty} \frac{t^m x^m}{m!}$

The moment generating function of TIHLEtW distribution is given by

$$M_x(t) = \sum_{i,j,k=0}^{\infty} \sum_{m=0}^{\infty} \frac{t^m \eta_p \Gamma(\frac{m}{\beta} + 1)}{(k+1)^{\frac{m}{\beta}+1} \theta^{\frac{m}{\beta}} m!} \tag{19}$$

4.4. Reliability function

The reliability function which is also known as survivor function that gives the probability that a patient will survive longer than specified period of time. It is defined as

$$R(x;\lambda,\alpha,\theta,\beta) = \frac{2\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda}} \tag{20}$$

4.5. Hazard function

The hazard function is the probability of an event of interest occurring within a relatively short time frame and is defined as

$$T(x;\lambda,\alpha,\theta,\beta) = \frac{\lambda\alpha\theta\beta x^{\beta-1} e^{-\theta x^{\beta}}\left[1-e^{-\theta x^{\beta}}\right]^{\alpha-1}}{\left[1+\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]^{\lambda}\right]\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\alpha}\right]} \tag{21}$$

4.6. Quantile Function

The quantile function is a vital tool to create random variables from any continuous probability distribution. As a result, it has a significant position in probability theory. For x, the quantile function is F(x) = u, where u is distributed as U(0,1).The TIHLEtW distribution is easily simulated by inverting equation (5) which yields the Quantile function Q(u) defined as:

$$x = Q(u) = \left\{ \frac{-\log\left[1-\left[1-\left[\frac{1-U}{U+1}\right]^{\frac{1}{\lambda}}\right]^{\frac{1}{\alpha}}\right]}{\theta} \right\}^{\frac{1}{\beta}} \tag{22}$$

The first quartile, the median and the third quartile of TIHLEtW distribution are obtained by putting u = 0.25, 0.5 and 0.75, respectively in equation (22).

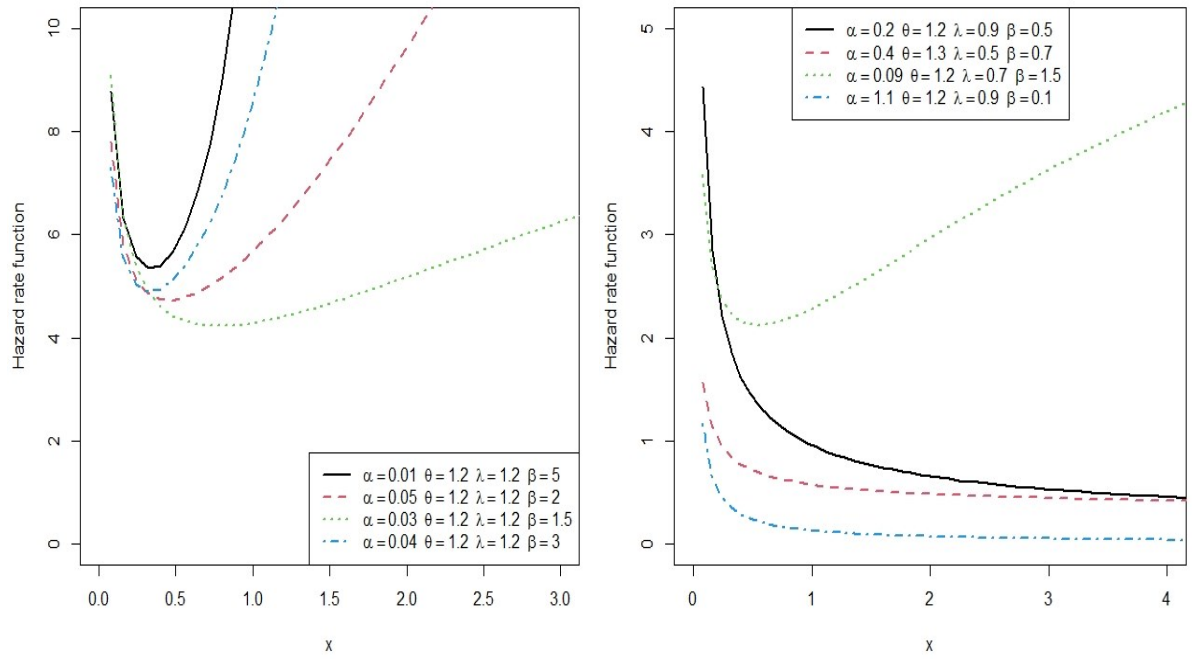


Figure 2: Plots of hazard of the TIHLEtW distribution for different valves of parameters.

4.7. Order Statistics

Many areas of statistics including reliability and life testing have made substantial use of order statistics. Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables with their corresponding continuous distribution function F (x). Let $X_1, X_2,.., X_n$ be n independently distributed and continuous random variables from the TIHLEtW distribution. Let $F_{r:n}(x)$ and $f_{r:n}(x)$, r = 1, 2, 3, ..., n denote the cdf and pdf of the r^{th} order statistics $X_{r:n}$ respectively. David [16] gave the probability density function of $X_{r:n}$ as:

$$f_{r:n}(x) = \frac{f(x)}{B(r, n-r+1)} \sum_{v=0}^{n-r} (-1)^v \binom{n-r}{v} F(x)^{v+r-1} \tag{23}$$

By substituting equation (8) and equation (9) into equation (23), also replacing h with v+r-1

in equation (9). We have

$$f_{r:n}(x;\lambda,\alpha,\theta,\beta)=\frac{1}{B(r,n-r+1)}2\lambda\alpha\theta\beta x^{\beta-1}\sum_{v=0}^{n-r}\sum_{i,j,k=0}^{\infty}\sum_{p,m=0}^{v+r-1}(-1)^{i+j+k}(-1)^{p+m+z+q}(-1)^{v}$$
$$\binom{n-r}{v}\binom{i+1}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k} \tag{24}$$
$$\binom{r+v+p-2}{p}\binom{r+v-1}{m}\binom{\lambda(m+p)}{z}\binom{\alpha z}{q}\varphi_t(e^{-\theta x^{\beta}})^{k+1+q}$$

The equation above is called the r[th] order statistics for the TIHLEtW distribution.

Let r = n, then the probability density function of the maximum order statistics of TIHLEtW distribution is

$$f_{n:n}(x;\lambda,\alpha,\theta,\beta)=2n\lambda\alpha\theta\beta x^{\beta-1}\sum_{i,j,k=0}^{\infty}\sum_{p,m=0}^{v+n-1}(-1)^{i+j+k}(-1)^{p+m+z+q}(-1)^{v}\binom{i+1}{i}$$
$$\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\binom{n+v+p-2}{p} \tag{25}$$
$$\binom{n+v-1}{m}\binom{\lambda(m+p)}{z}\binom{\alpha z}{q}(e^{-\theta x^{\beta}})^{k+1+q}$$

Also, let r = 1, then the probability density function of the minimum order statistics of TIHLEtW distribution is

$$f_{1:n}(x;\lambda,\alpha,\theta,\beta)=2n\lambda\alpha\theta\beta x^{\beta-1}\sum_{v=0}^{n-1}\sum_{i,j,k=0}^{\infty}\sum_{p,m=0}^{v}(-1)^{i+j+k}(-1)^{p+m+z+q}(-1)^{v}\binom{n-1}{v}\binom{i+1}{i}$$
$$\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\binom{v+p-1}{p} \tag{26}$$
$$\binom{v}{m}\binom{\lambda(m+p)}{z}\binom{\alpha z}{q}(e^{-\theta x^{\beta}})^{k+1+q}$$

5. Parameter Estimation

In this paper, we explore the maximum likelihood technique to estimate the unknown parameters of the TIHLEtW distribution for complete data. Maximum likelihood estimates (MLEs) have appealing qualities that may be used to generate confidence ranges and provide simple approximations that function well in finite samples. In distribution theory, the resulting approximation for MLEs is easily handled, either analytically or numerically. Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample of size n from the TIHLEtW distribution. Then, the likelihood function based on observed sample for the vector of parameter $(\lambda, \alpha, \theta, \beta)^T$ is given by

$$\log(\phi) = n\log(2) + n\log(\lambda) + n\log(\alpha) + n\log(\theta) + n\log(\beta)$$
$$+(\beta-1)\sum_{i=1}^{n}\log(x_i) - \theta\sum_{i=1}^{n}(x_i^{\beta})$$
$$+(\alpha-1)\sum_{i=1}^{n}\log[1-e^{-\theta x_i^{\beta}}] + (\lambda-1)\sum_{i=1}^{n}\log\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right] \quad (27)$$
$$-2\sum_{i=1}^{n}\log\left[1+\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda}\right]$$

The components of score vector $\Delta L(\phi) = \left(\frac{\partial L(\phi)}{\partial\lambda}, \frac{\partial L(\phi)}{\partial\alpha}, \frac{\partial L(\phi)}{\partial\theta}, \frac{\partial L(\phi)}{\partial\beta}\right)^{T}$ are given as

$$\frac{\partial L(\phi)}{\partial\lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n}\log\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]$$
$$-2\sum_{i=1}^{n}\frac{\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda}\log\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]}{1+\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda}} \quad (28)$$

$$\frac{\partial L(\phi)}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log\left[1-e^{-\theta x_i^{\beta}}\right]$$
$$+(\lambda-1)\sum_{i=1}^{n}\frac{\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\log\left[1-e^{-\theta x_i^{\beta}}\right]}{1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}} \quad (29)$$
$$+2\sum_{i=1}^{n}\frac{\lambda\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda-1}\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\log\left[1-e^{-\theta x_i^{\beta}}\right]}{1+\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda}}$$

$$\frac{\partial L(\phi)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n}x_i^{\beta} + (\alpha-1)\sum_{i=1}^{n}\frac{e^{-\theta x_i^{\beta}}x_i^{\beta}}{1-e^{-\theta x_i^{\beta}}}$$
$$-(\lambda-1)\sum_{i=1}^{n}\frac{\alpha\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha-1}e^{-\theta x_i^{\beta}}x_i^{\beta}}{1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}} \quad (30)$$
$$-2\sum_{i=1}^{n}\frac{\lambda\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda-1}\alpha\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha-1}e^{-\theta x_i^{\beta}}x_i^{\beta}}{1+\left[1-\left[1-e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda}}$$

$$\frac{\partial L(\phi)}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} \log(x_i) - \theta \sum_{i=1}^{n} x_i^{\beta} \log(x_i) + (\alpha - 1)$$
$$-(\lambda - 1) \sum_{i=1}^{n} \frac{\alpha \left[1 - e^{-\theta x_i^{\beta}}\right]^{\alpha - 1} e^{-\theta x_i^{\beta}} \theta x_i^{\beta} \log(x_i)}{1 - \left[1 - e^{-\theta x_i^{\beta}}\right]} \tag{31}$$
$$-2 \sum_{i=1}^{n} \frac{\lambda \left[1 - \left[1 - e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda - 1} \alpha \left[1 - e^{-\theta x_i^{\beta}}\right]^{\alpha - 1} e^{-\theta x_i^{\beta}} \theta x_i^{\beta} \log(x_i)}{1 + \left[1 - \left[1 - e^{-\theta x_i^{\beta}}\right]^{\alpha}\right]^{\lambda}}$$

The MLEs are obtained by setting $\frac{\partial L(\phi)}{\partial \lambda}, \frac{\partial L(\phi)}{\partial \alpha}, \frac{\partial L(\phi)}{\partial \theta}$ and $\frac{\partial L(\phi)}{\partial \beta}$ to zero and solving these equations simultaneously. These equations cannot be solved analytically, so we have to appeal to numerical method.

6. Simulation Study

In this section, a numerical analysis will be conducted to evaluate the performance of MLE for TIHLEtW Distribution.

Table 1: MLEs, biases and RMSE for some values of parameters

		(3,2,2.5,2)			(3,2,2.5,3)		
N	Parameters	Estimated Values	Bais	RMSE	Estimated Values	Bais	RMSE
20	λ	3.0715	0.0715	0.7353	3.0263	0.0263	0.7096
	α	2.1100	0.1100	0.8549	2.2121	0.2121	0.9444
	θ	2.8745	0.3745	0.9361	2.9235	0.4235	1.0310
	β	2.8745	0.8745	1.2251	3.9235	0.9235	0.9431
50	λ	3.0412	0.0412	0.5405	3.0242	0.0242	0.5472
	α	2.1008	0.1008	0.6248	2.1454	0.1454	0.7306
	θ	2.6588	0.1588	0.5958	2.7069	0.2069	0.6429
	β	2.6588	0.6588	0.8740	3.7069	0.7069	0.6757
100	λ	3.0094	0.0094	0.3889	3.0149	0.0149	0.3930
	α	2.1013	0.1013	0.4425	2.0947	0.0947	0.5194
	θ	2.5938	0.0938	0.3702	2.6052	0.1052	0.4013
	β	2.5398	0.5398	0.6934	3.6052	0.6052	0.5530
250	λ	3.0552	0.0552	0.2985	3.0133	0.0133	0.2430
	α	2.0518	0.0518	0.2713	2.0286	0.0286	0.2740
	θ	2.5044	0.0044	0.2159	2.5114	0.0114	0.1853
	β	2.4044	0.4044	0.5487	3.5114	0.5114	0.5225
500	λ	3.0261	0.0261	0.1840	3.0029	0.0029	0.1756
	α	2.0325	0.0325	0.1856	2.0116	0.0116	0.1831
	θ	2.5011	0.0011	0.1405	2.5033	0.0033	0.1406
		2.3011	0.3011	0.4205	3.2033	0.2033	0.4162

	β						
1000	λ	3.2149	0.0149	0.1380	3.0014	0.0014	0.1136
	α	2.0201	0.0201	0.1243	2.0006	0.0006	0.1067
	θ	2.5009	0.0009	0.1096	2.5010	0.0010	0.0891
	β	2.2029	0.2029	0.3147	3.1010	0.1010	0.2069

The table above shows the values of biases and RMSEs approach zero and the estimates tend to the initial (true) values as the sample increases, which indicates that the estimates are efficient and consistent.

7. Applications to Real Data

In this section, we fit the TIHLEtW distribution to two real data sets and give a comparative study with the fits to the Type II Exponentiated Half Logistic Weibull (TIIEHLW) distribution by Al-Mofleh *et al.* [7], Half-Logistic Generalized Weibull (HLGW) Distribution by Masood and Amna [23], Exponentiated Weibull (EW) by Pal *et al.* [28], Weibull Distribution by Xie and Lai [33] and Topp-Leone Generated Weibull (TLGW) Distribution by Aryal *et al.* [9] as comparator distributions for illustrative purposes.

The TIIEHLW distribution developed by Al-Mofleh *et al.* [7] has pdf defined as:

$$f(x;\alpha,\lambda,\beta,\theta)=2\alpha\lambda\beta\theta x^{\beta-1}e^{-\theta x^{\beta}}\left[1-e^{-\theta x^{\beta}}\right]^{\lambda-1}\frac{\left[1-\left[1-e^{-\theta x^{\beta}}\right]^{\lambda}\right]^{\alpha-1}}{\left[1+\left[1-e^{-\theta x^{\beta}}\right]^{\lambda}\right]^{\alpha+1}} \tag{32}$$

The HLGW distribution developed by Masood and Amna [23] has pdf defined as:

$$f(x;\lambda,\alpha,\theta)=\frac{2\lambda\alpha\theta x^{\alpha-1}\left[1+\theta x^{\alpha}\right]^{\lambda-1}exp\left[1-\left[1+\theta x^{\alpha}\right]^{\lambda}\right]}{\left[1+exp\left[1-\left[1+\theta x^{\alpha}\right]^{\lambda}\right]\right]^{2}} \tag{33}$$

The EW distribution proposed by Pal *et al.,* [28] has pdf given as:

$$f(x;\alpha,\lambda,\beta)=\alpha\lambda^{\beta}\beta x^{\beta-1}\left[1-exp(-\lambda x)^{\beta}\right]^{\alpha-1}exp(-\lambda x)^{\beta} \tag{34}$$

The Weibull Distribution proposed by Xie and Lai [33] has pdf given as:

$$f(x;\theta,\beta)=\theta\beta x^{\beta-1}e^{-\theta x^{\beta}} \tag{35}$$

The TLGW distribution developed by Aryal *et al.,* [9] has pdf defined as:

$$f(x;\alpha,\theta,\beta,\lambda)=2\alpha\theta\beta\lambda^{\beta}x^{\beta-1}e^{-(\lambda x)^{\beta}}\left[1-e^{-(\lambda x)^{\beta}}\right]^{\theta\alpha-1}$$
$$\left[1-\left[1-e^{-(\lambda x)^{\beta}}\right]^{\theta}\right]\left[2-\left[1-e^{-(\lambda x)^{\beta}}\right]^{\theta}\right]^{\alpha-1} \tag{36}$$

The two datasets used as examples in the application demonstrate the new proposed distribution flexibility, applicability, and "best fit" in modeling the datasets empirically when compared to the above comparator distributions. All of the calculations are performed using the R programming language.

Data set 1

The first data set shown below represents the remissions times (in months) of a random sample of one hundred and twenty-eight (128) bladder cancer patients, previously used by Lee and Wang [22]:

0.08, 2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26, 3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54, 3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77, 32.15, 2.64, 3.88, 5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34, 7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23, 5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26, 2.83, 4.33, 5.49, 7.66, 11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02,4.34, 5.71, 7.93, 11.79, 18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25,8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

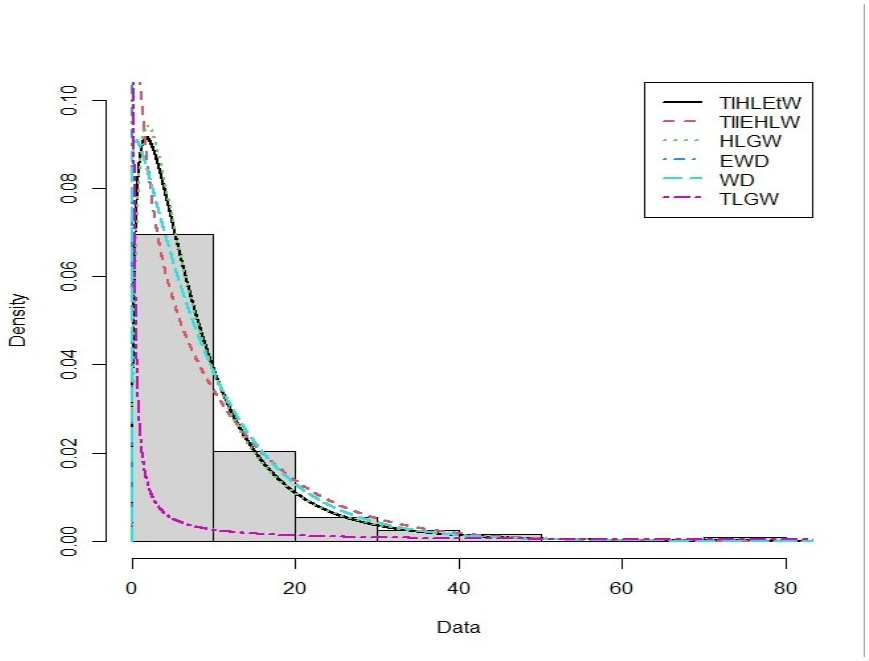


Figure 3: Fitted pdfs for the TIHLEtW, TIIEHLW, HLGW, EWD, WD, and TLGW distributions to the data set 1

Table 2: MLEs, Log-likelihoods and Goodness of Fits Statistics for the Data Set 1

Distributions	α	λ	θ	β	LL	AIC
TIHLEtW	3.4827	1.1097	0.7581	0.5426	-410.6609	829.3218
TIIEHLW	0.2368	0.8929	0.2634	1.1245	-418.4258	844.8516
HLGW	1.0581	0.6613	0.2868		-412.4861	830.9721
EWD	1.1545	0.1188		0.9861	-413.1202	832.2403
WD			0.0939	1.0478	-414.0869	832.1738
TLGW	6.6269	0.0219	4.1785	0.2522	-442.2653	880.5306

Table 2 presents the results of the Maximum Likelihood Estimation of the parameters of the new proposed distribution and the five comparator distributions. Based on the goodness of fit measure, the new proposed distribution reported the minimum AIC value, though followed closely by the HLGW. The visual inspection of the fit presented in Figure 3, also confirms the superiority of the proposed distribution amongst its comparators. Thus the new proposed distribution 'best fit' bladder cancer patients data set amongst the range of distributions considered.

Data set 2

The second data set shown below represents the life times data relating to times (in months from 1st January, 2013 to 31st July, 2018) of 105 patients who were diagnosed with hypertension and received at least one treatment related to hypertension in the hospital where death is the event of interest, previously used by Umeh and Ibenegbu [31]:

45, 37, 14, 64, 67, 58, 67, 55, 64, 62, 9, 65, 65, 43, 13, 8, 31, 30, 66, 9, 10, 31, 31, 31, 46, 37, 46, 44, 45, 30, 26, 28, 45, 40, 47, 53, 47, 41, 39, 33, 38, 26, 22, 31, 46, 47, 66, 61, 54, 28, 9, 63, 56, 9, 49, 52, 58, 49, 53, 63, 16, 67, 61, 67, 28, 17, 31, 46, 52, 50, 30, 33, 13, 63, 54, 63, 56, 32, 33, 37, 7, 56, 1, 67, 38, 33, 22, 25, 30, 34, 53, 53, 41, 45, 59, 59, 60, 62, 14, 57, 56, 57, 40, 44, 63.

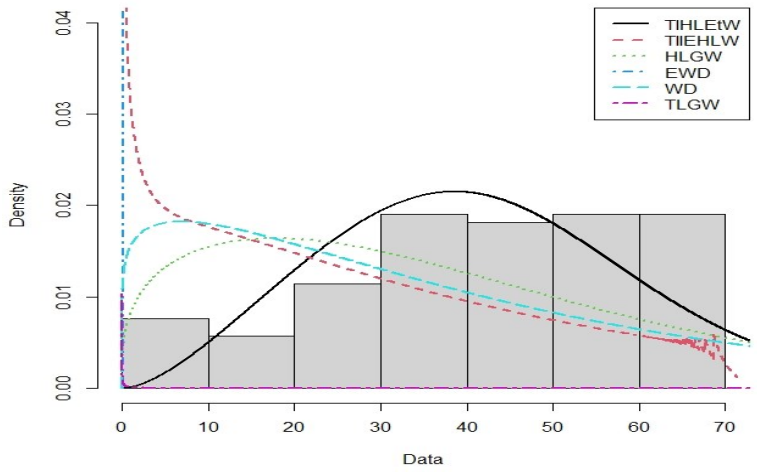


Figure 4: Fitted pdfs for the TIHLEtW, TIIEHLW, HLGW, EWD, WD, and TLGW distributions to the data set 2

Table 3: MLEs, Log-likelihoods and Goodness of Fits Statistics for the Data Set 2

Distributions	α	λ	θ	β	LL	AIC
TIHLEtW	0.3906	0.0877	0.0016	2.3771	-446.1673	900.3346
TIIEHLW	0.0487	0.5271	0.3378	1.1028	-495.9288	999.8576
HLGW	1.3181	0.7136	0.0175		-475.625	957.2501
EWD	5.0005	4.4779		0.1826	-464.4788	934.9575
WD			0.0138	1.1406	-487.8239	979.6479
TLGW	10.8957	0.0120	0.0624	8.2172	-471.7036	951.4072

Table 3 shows the results of the Maximum Likelihood Estimation of the parameters of the TIHLEtW distribution and the five comparator distributions. Based on the goodness of fit statistic AIC, the new distribution reported the minimum AIC value suggesting that the distribution is the 'best fit' to the hypertension patients. The visual inspection of the fit presented in Figure 4, also reaffirms the superiority of the new distribution amongst its comparators.

8. CONCLUSION

In this article, we proposed and studied a new distribution called the Type I Half-Logistic Exponentiated Weibull Distribution using the family of distribution proposed by Bello *et al.* (2021). Explicit quantile function, probability weighted moments, moments, generating function, reliability function, hazard function, and order statistics were examined as statistical components of the new proposed distribution. The parameters are estimated using the maximum likelihood technique. We present some simulation results to evaluate the new distribution's performance. In comparison to well-known models, two real data sets are evaluated to highlight the importance and flexibility of the new distribution. The findings reveal that the new distribution appears to be superior to the existing models considered, implying that it can be used to model data in a variety of applications.

References

[1] Afify, A.Z., Nofal, Z.M., and Butt, N.S. (2014). Transmuted complementary Weibull geometric distribution. *Pak J Stat Oper Res*; 10: 435-454.

[2] Afify, A.Z., Cordeiro, G.M., Butt, N.S., Ortega, E.M.M., and Suzuki, A.K. (2017). A new lifetime model with variable shapes for the hazard rate. *Braz J Probab Stat*; 31: 516-541.

[3] Afify, A.Z., Cordeiro, G.M., Yousof. H.M., Abdus, S., and Ortega, E.M.M. (2018). The Marshall-Olkin additive Weibull distribution with variable shapes for the hazard rate. *Hacettepe J Math Stat*; 47: 365-381.

[4] Ahmad, Z. (2018). The Zubair-G family of distributions: properties and applications. *Annals of Data Science*, 5, 1–14.

[5] Ahmad, Z., Ilyas, M., and Hamedani, G. G. (2019). The extended alpha power transformed family of distributions: properties and applications. *Journal of Data Science*, 17(4), 726–741.

[6] Ahmad, Z., Elgarhy, M., and Abbas, N. (2019). A new extended alpha power transformed family of distributions: properties and applications. *Journal of Statistical Modelling: theory and Applications*, 1(1), 13–28.

[7] Al-Mofleh Hazem, Elgarhy Mohamed, Afify Ahmed Z, and Zannon Mohammad (2020). Type II Exponentiated Half Logistic Generated Family of Distribution with Applications. *Electronic Journal of Applied Statistical Analysis*, 13(2), 536-561.

[8] Aryal, G.R., and Tsokos, C.P. (2011). Transmuted Weibull distribution: a generalization of the Weibull probability distribution. *Eur J Pure Appl Math*, 4: 89-102.

[9] Aryal, G.R., Ortega, E.M.M., Hamedani, G.G., and Yousof, H.M. (2017). The Topp-Leone generated Weibull distribution: regression model, characterizations and applications. *Int J Stat Probab*; 6(1) 126-141.

[10] Bebbington, M. Lai, C.D. and Zitikis, R. (2007). A flexible Weibull extension. *Reliability Engineering & System Safety*, 92(6), 719–726.

[11] Bello, O. A., Doguwa, S. I., Yahaya, A., and Jibril, H. M. (2021). A Type I Half Logistic Exponentiated-G Family of Distributions: Properties and Application. *Communication in Physical Sciences*, 7(3), 147-163.

[12] Codeiro, G.M., Hashimoto, E.M., and Ortega, E.M.M. (2014). The McDonald Weibull model. *J Theor Appl Stat*; 48: 256-278.

[13] Cordeiro, G.M, Ortega, E.M.M., and Nadarajah, S. (2010) The Kumaraswamy Weibull distribution with application to failure data. *J Franklin Inst*, 347: 1399-1429.

[14] Cordeiro, G.M., Ortega, E.M.M., and Silva, G.O. (2014). The Kumaraswamy modified Weibull distribution: theory and applications. *J Stat Comput Simul*; 84: 1387-1411.

[15] Cordeiro, G.M., Ortega, E.M.M., and Da Cunha D.C.C. (2013). The exponentiated generalized class of distributions. *J Data Sc*., 11: 1-27.

[16] David, H. A. (1970). Order statistics, Second edition. Wiley, New York.

[17] Elbatal, I., and Aryal, G. (2013). On the transmuted additive Weibull distribution. *Austrian J Stat;* 42: 117-132.

[18] Greenwood, J.A. Landwehr, J.M., and Matalas, N.C. (1979). Probability weighted moments: Definitions and relations of parameters of several distributions expressible in inverse form. *Water Resources Research*, 15, 1049-1054.

[19] Hanook, S., Shahbaaz, M.Q., Mohsin, M., and KIBRIA, G. (2013). A Note on Beta Inverse Weibull Distribution. *Commun Stat Theory Methods*; 42: 320-335.

[20] Ibrahim, S. (2021). The properties of Topp Leone exponentiated weibull distribution with application to survival data. *Research Journal of Mathematical and Statistical Sciences,* 9(1), 9-15.

[21] Lai, C. D., Xie, M., and Murthy, D. N. P. (2003) . A modified Weibull distribution. *IEEE Transactions on Reliability*, 52(1), 33–37.

[22] Lee, E. T. and Wang, J. W. (2003). Statistical methods for survival data analysis (3rd Edition), John Wiley and Sons, New York, USA, 535 Pages, ISBN 0-471-36997-7.

[23] Masood Anwar and Amna Bibi. (2018). The Half-Logistic Generalized Weibull Distribution. *Journal of Probability and Statistics*, 8767826, 12.

[24] Merovci, F., and Elbatal, I. (2013). The McDonald modified Weibull distribution: properties and applications. *arXiv preprint arXiv*:13092961. (In Press).

[25] Mudholkar, G.S., Srivastava, D.K., and Kollia, G.D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, **91**: 1575-1583.

[26] Nofal, Z.M., Afify, A.Z., Yousof, H.M., and Cordeiro, G.M. (2017). The generalized transmuted-G family of distributions. *Commun Stat Theory Methods*; 46: 4119-4136.

[27] Nofal, Z.M., Afify, A.Z., Yousof, H.M., Granzotto, D.C.T., and Louzada, F. (2016). Kumaraswamy transmuted exponentiated additive Weibull distribution. *Int J Stat Probab*; 5: 78-99.

[28] Pal, M., Ali, M.M., and Woo, J. (2006) Exponentiated Weibull distribution. *STATISTICA, anno* LXVI, 2.

[29] Shahbaz, M.G, Shahbaz, S., and Butt, N.M.(2012). The Kumaraswamy inverse Weibull distribution. *Pak J Stat Oper,* 8: 479-489.

[30] Silva, G. O., Ortega, E. M. M., and Cordeiro, G. M. (2010). The beta modified Weibull distribution. *Lifetime Data Analysis*, 16(3), 409–430.

[31] Umeh, E. and Ibenegbu, A. (2019). A Two-Parameter Pranav Distribution with Properties and Its Application; Journal of Biostatistics and Epidemiology, 5(1) : 74-90

[32] Xie, M., Tang, Y., and Goh, T. N. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 76 (3), 279–285.

[33] Xie, M., and Lai, C. D. (1996). On the increase of the expected lifetime by parallel redundancy. *Asia Pacific J. Oper. Res*. 13, 171179.

[34] Zhang, T. and Xie, M. (2011). On the upper truncated Weibull distribution and its reliability implications. *Reliability Engineering & System Safety*, 96(1), 194–20.

RT&A, No . 3 (74)
Volume 18, September 2023

ESTIMATION OF RELIABILITY AND LIFETIME OF COMPOSITE OVERWRAPPED PRESSURE VESSELS ADOPING POTENTIAL FAILURES ASSESSMENT AND ACCELERATED TESTS APPROACH

Maryam Gholami Arjenaki[1], Dr. Mahdi Karbasian[2]*, Amin Kazemi Manesh[3], Mohammadreza Jafari[4]

[1] University of Technology-Shahin Shahr Campus, Shahin Shahr, Isfahan, Iran.
[2]*Malek Ashtar University of Technology, Tehran, Iran.
[3]Islamic Azad University- Ahar branch, Ahar, Iran.
[4]Islamic Azad University-Zahedan Branch, Zahedan, Iran.
[1]maryamgholami6677@gmail.com, [2]*mkarbasian@yahoo.com, [3]kazemimaneshamin@gmail.com, [4]Jafari6340@gmail.com

Abstract

Aim. *Compressed air vessels are responsible for injecting compressed air to the mechanical flying device. It should be noted that the pressure level inside these vessels is very important in conducting all operational stages successfully; therefore, it is of high significance to be assured of the quality of the vessels being used. This study was done in 2017 in order to calculate and estimate the reliability of compressed air vessels in mechanical flying device system with proposing a potential failures assessment and accelerated test approach taking into consideration the current methods.* ***Methods.*** *The paper uses methods of Fault Tree Analysis, Failure Mode & Effect Analysis and Accelerator tests. Initially, the interactions among the components were identified using the Design Structure Matrix in order to design a matrix to help estimate the reliability; consequently, improve equipment performance. Next, failures root recognition was done using Fault Tree Analysis diagram, then, failures reasons prioritization was done using Failure Mode & Effects Analysis tables. Accelerator tests were designed and applied on failure mechanisms such as leakage by pressure on vessels, corrosions on steal head, nipples, O-ring creeping, O-ring ozone cracking and liner chemical degradation. After that, the average of failure rates was calculated within the taking after stage for each test. Within the conclusion, the result of failure rates from the accelerator tests was compared with the result of failure rates from the process approach. Consequently, the most elevated amount of these two approaches was defined as the total failure rate; product reliability and lifetime were calculated utilizing this amount.* ***Results.*** *The following finding were obtained using the proposed methods. When Windy Liner was used, vessel lifetime was six years and half and vessel reliability in ten years was 0/22. Whereas, when Rotational Liner was used, vessel lifetime was eight years and three months and vessel reliability in ten years was 0/3.* ***Conclusion****.The approach proposed in the paper allows accelerator degradation test can also be used instead of accelerated test in order to calculate the reliability of failure mechanisms. In the event that high-quality and legitimate O-rings are used, vessel reliability and lifetime can be increased.*

Keywords: Reliability, Accelerator Life Tests, Composite Overwrapped Pressure Vessel, Failure Modes and Effects Analysis, Fault Tree Analysis.

1- Introduction

Product reliability is an essential aspect of customer expectation. The expectation to increase reliability at the lowest cost has become an important issue for companies in offering competitive products [1]. Product reliability is about promoting the product to the needed quality over time, which is absolutely important in creating fame and maintaining competitive advantages for companies [2]. Traditional reliability evaluation methods were always based on failure data which were obtained from product lifetime, i.e., it was not possible to calculate the reliability as long as the product was functional. This lifetime failure data was obtained almost impossibly in a very short period of time for products with high reliability and extended lifetime. To solve this problem, creating quantitative data is absolutely significant and it can be achieved by evaluating the potential failures approach. Furthermore, using accelerator tests to estimate reliability in real conditions is considered to be practical [3].

Products with high reliability and lifetime are used extensively in spacecraft, aviation, and other fields, so, how to evaluate the reliability and predict their lifetime is one of the most important topics [4]. Pressure vessels inject compressed air into the mechanical flying device. The pressure level inside these vessels is very important in order to successfully proceed all operational sections. It is necessary to have these vessels always ready to operate with operating pressures (approximately 280 bars). It is highly essential for the user to know how many years the vessels will be functional with this operating pressure [4].

The used composite overwrapped pressure vessels (COPV), are made of metal or polymer core and is wrapped by low-density composite layers. The interior layers are responsible for providing the original structure, strength, rigidity and appropriate surface in contact with the gas inside the vessel. Composite coating is wrapped around the core with the aim of ensuring the mechanical strength of the inner layers against high levels of pressure as well as resisting against scratch, impact and other possible damages [5]. In terms of ratio of strength to density, due to severe decline in weight, COPV have noticeable advantages compared to the metal pressure vessels. However, some difficulties such as high production costs diminished this superiority [6].

Since by default, COPV were designed for flying machine equipment; high reliability and safety are considered in their design. But the safety considered in the design phase cannot be trusted because incidents and damages in composites have caused explosions due to the release of high energy of the compressed gas in the vessels. It is true that the possibility of such incidents occurring is rare, but if we consider the severity of the disaster; especially, in aerial and human-related projects, the high amount of reliability seems to be necessary [7].

This study was conducted in one of the flying machine device manufacturing industries in 2017 and the purpose of this study was to calculate the reliability of these compressed air pressure vessels in these systems.

2- Methods

Initially, the interactions among the components were identified using the design structure matrix [8]. Furthermore, failure roots recognition was done using fault tree analysis (FTA) diagram [9]. Occurrence and severity numbers were defined by experts using failure modes and effects analysis (FMEA) tables; eventually, risk priority number (RPN) was calculated [10]. As a result, priority of failure causes was clarified. Critical failure factors were then specified after designing risk analysis matrix.

Furthermore, accelerator tests were designed and performed. Next, the failure rate was calculated for each test. At the end, the failure rates average from the accelerator tests were compared with failure rates of the process approach. After that, the highest number derived from this comparison was the final failure rate of the product; consequently, product reliability and lifetime were calculated (Fig. 1).

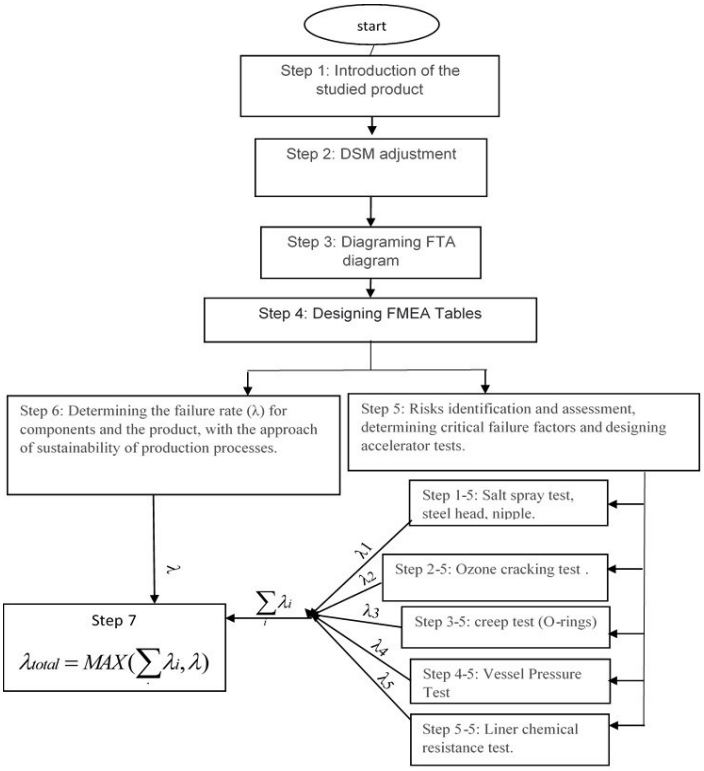


Figure 1: Research steps chart

2-1- Design Structure Matrix (DSM)

At the heading row and the body row of DSM, the product components are mentioned in the chart and the relations among them were clarified [supporting table 1].

2-2- Fault tree analysis (FTA)

In this stage, FTA was demonstrated for each part of the vessel's components. The example of fault tree for O-ring parts and steel head can be seen in figures 2 and 3. Also, a fault tree was drawn for the nipple, composite, rotary and wind liner parts accordingly.

2-3- Failure Modes and Effective Analysis (FMEA)

FMEA tables, in the next stage, were designed for FTA diagrams which were then designed for product components [supporting table 2]. FMEA table was completed according to the data from FTA diagrams; the top events in FTA diagrams were placed under potential failures column, and other forms of failures in FTA were placed under potential causes column according to their stage.

Kim et al. (2013) used FMEA tables in an article entitled 'A new reliability allocation weight for reducing occurrence of severe failure effects'. To complete FMEA tables, they used the standards which had been designed by the US Army. By the dedication of occurrence and severity number to potential failures in FMEA table, formula 1 can be used to calculate the failure rate of each potential failure [11]. According to brainstorming sessions and consultations with industry experts, risk priority number, severity and occurrence numbers were defined between 1-10. RPN was calculated by multiplying severity by occurrence. In FMEA tables, a special made-up code was given to every potential failure cause. This would not only make future analysis easier, but also would help to avoid explanations.

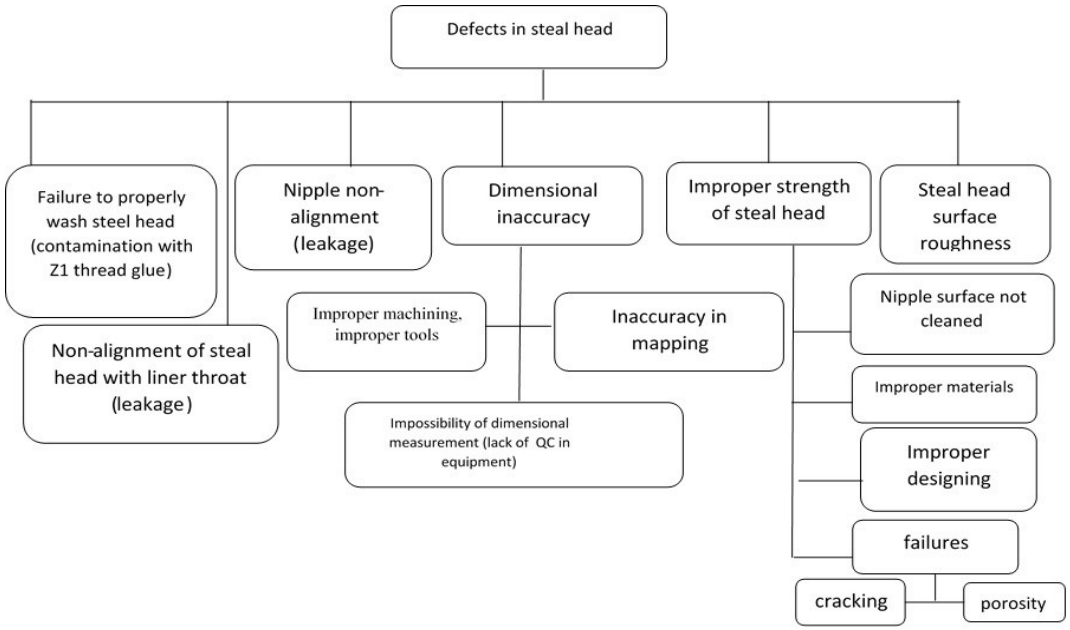


Figure 2: Steal head Fault Tree

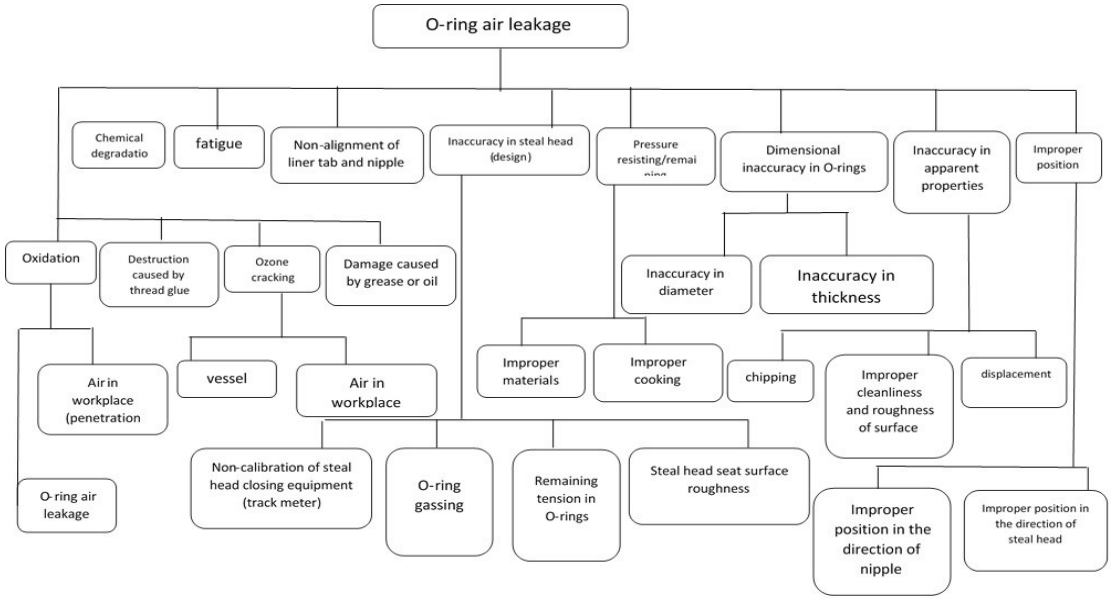


Figure 3: O-ring Fault Tree

$$\lambda=\text{EXP}\ (-9.993+0.7702*O) \quad (1)$$

In formula 1, O is Occurrence, and EXP is Exponential Function. λ as a failure rate was calculated based on any failure cause with occurrence number. Time unit for FMEA tables is 10 years [12].

2-4- The development of accelerator tests

After completing the risk analysis table, failure codes in the critical area of risk analysis table and potential failures with the highest score received in RPN were identified, therefore, wherever possible, accelerator tests were designed and performed on these mechanisms. Accelerator tests can be

performed on mechanisms such as: leakage caused by pressure on vessels, corrosion on the series of steal and nipple components, creeps in O-rings, Ozone cracking in O-rings, liner chemical degradation. In the next step, accelerator tests for the aforementioned mechanisms were designed; it is significant to realize that the samples were in 3 stress levels and in each level, three samples were evaluated and the overall of 9 samples for each test were examined. Selection of test parameters were based on product material, geometric tolerance and environmental conditions.

2-5- Salt Spray Test for Steel Head and Nipple

The Salt Spray Test is intended in corrosion-related failure mechanisms of the steel head and nipple. Formula 2 can be used to calculate the lifetime of steel head and nipple using the data derived from the corrosion.

$$A = \frac{TFop}{TFstress} = \exp\left[a.(RHstress - RHop)\right] * \exp\left[\frac{Ea}{Kb}\left(\frac{1}{Top} - \frac{1}{Tstress}\right)\right] \quad (2)$$

In formula 2, TFop is time to fail in normal condition, TFstress is time to fail under stress, a is constant and is equal to 0.1 (RH-1). RHstress is humidity under stress; RHop is humidity in normal conditions. Ea is activation energy in terms of electron volt. Kb is Boltzmann constant which is equal to 8.61*10-5 [12] [supporting table 3].

2-6- Ozone cracking test

In order to find failure in O-rings, Ozone cracking test was designed. In this test, to determine the accelerator factor, when temperature is not the main cause of acceleration, formula 3 which is reverse exponentiation relation is used.

$$AF = \left(\frac{\xi_{stress}}{\xi_{op}}\right)^{n} \quad (3)$$

In formula 3, ξ_{op} and ξ_{stress} are operational stress and stress under accelerated conditions, respectively. And n is power derived from the exponentiation law model. According to formula 4 and 5, lifetime or Time to Failure (TF) is calculated in operational conditions [12].

$$AF = \frac{TF_{operation}}{TF_{stress}} \quad (4)$$

$$TF_{operation} = AF \times TF_{stress} \quad (5)$$

2-7- Creep test for O-ring

Creep test was designed for failure-related creep mechanism in O-rings. O-ring lifetime is calculated based on formula 6 under creep condition.

$$TF = t_0^{\gamma\sigma} e^{(u \quad /RT)} \quad (6)$$

TF is the destruction time caused by creeping. t_0 is constant and is equal to 10-12 per second. u is active energy in fracture process. ɣ is a constant material structure. R is molar gas constant and is equal to 8/314 j/mol0k. T is absolute heat degree and σ is constant stress [12] [supporting Table 4 and 5].

2-8- Pressure test

Pressure test is designed for failure leakage due to pressure on vessel. Pressure and heat are the two failure factors used in pressure test. When the failure reason is pressure, formula 3 is used to calculate

accelerator factor; whereas, when heat is the failure reason, formula 7 is used to calculate accelerator factor. In general, acceleration coefficient can be defined for either heat stress or non-heat stress, then formula 8 can be used for the accumulation of stress in equipment. After calculating AF, formula 5 can be used to calculate lifetime in operational conditions [12][supporting Table 6 and 7].

$$AF_1 = EXP\left(\frac{Ea}{Kb}\left(\frac{1}{T2}-\frac{1}{T1}\right)\right) \qquad (8)$$

$$AF = AF_1 \times AF_2 \times \ldots \times AF_n$$

2-9- Chemical resistance test for liner

Chemical resistance test is designed for failure mechanisms in chemical degradation of liner. Formulas 3 and 4 are used to calculate liner lifetime in chemical resistance test. ξop demonstrates PH in normal environment and $TF_{operation} = AF \times TF_{stress}$ shows PH in stress environment[12] [supporting Table 8].

2-10- Calculation of reliability and failure rate

Failure rate is calculated after acceleration tests are done. According to the assumption by which the product lifetime function follows exponential distribution, after calculating the lifetime in operational conditions, formula 9 can be used to calculate failure rate. Mean Time to Failure (MTTF) is the calculated number in operational lifetime. By using this formula, failure rate in accelerator tests can be calculated [12].

$$MTTF = \frac{1}{\lambda} \rightarrow \lambda = \frac{1}{MTTF} \qquad (9)$$

In this study, the Lussar's Law was used to calculate failure rate in products and failure rate in product manufacturing process.

2-11- Calculation of reliability, failure rate and vessel lifetime.

In order to calculate the failure rate in mechanical components in FMEA table, all failure rates in the table must be summed up so that the total product failure rate is derived. For example, to calculate the total failure rates in steal head, all failure rates in FMEA table related to steal head were summed up. To calculate failure rate resulting from the accelerator tests of all vessel components, when using the wind liner, failure rates from 5 accelerator tests were summed up. On the other hand, when the rotary liner was used, all accelerator tests except pressure test were calculated since pressure test was used for wind liner test.

Finally, to calculate the total failure rate of vessel components, failure rates resulted from the FMEA tables were summed up. Then, the maximum of these two numbers, after summing up the failure rates in accelerator tests and comparing them with the total failure rate of FMEA, were considered as the total failure rate of vessels.

Then, total Reliability (total) (R(t)) was calculated using formula 10 and lifetime was calculated using formula 9. It is needless to mention that since the duration of this study was 10 years, 1 is replaced by t in formula 10 [12].

$$R(t) = EXP(-(\sum \lambda) * t) \qquad (10)$$

3- Result

3-1- Calculation of failure codes, RPN and failure rate

FMEA tables were completed for different parts of the vessel. Failure codes NR-3, TK-2, NO-1, NO-2, NO-11, N-3, N-4, N-14 and N-16 had the highest RPN among vessel failure codes, respectively. Accelerator tests, then, were designed and performed. In table 1, samples of- failure code, RPN and failure rate were calculated for O-ring. More tables can be seen in supporting tables. According to risk matrix analysis in figure 4, N-16 and N-14 were of the only failure codes placed in critical area.

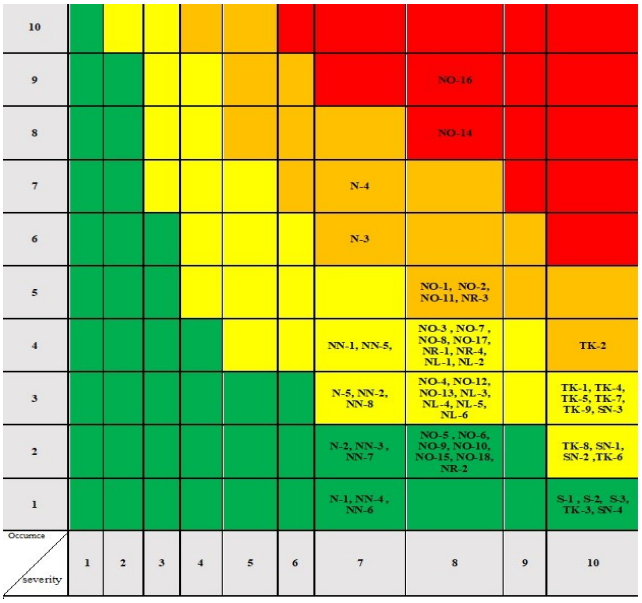


Figure 4: Risk analysis matrix for failure code sets of vessel components.

3-2- Calculation of accelerator test for potential failures

After performing the accelerator tests, according to the aforementioned formulas, operational component lifetime for each section was calculated in the methods section.

3-3- Salt Spray Test on Steal Head and Nipple

Operational lifetime was calculated by inserting the parameters listed in salt spray test parameters table into formula 2 [supporting Table 3]. Then by inserting the calculated lifetime in formula 9, the failure rate of salt spray test was calculated.

3-4- Ozone cracking test

By inserting the parameters listed in the ozone cracking test parameters table into formula 3, the number for accelerator factor (AF) was calculated [supporting Table 4]. Then by inserting AF into formula 4, the operational lifetime was calculated. The failure rate of ozone cracking test was derived by inserting the calculated lifetime into formula 9.

3-5- O-ring creep test

By inserting the parameters listed in creep test parameters table into formula 6, the operational lifetime was calculated. Then by inserting the derived amount of lifetime into formula 9, the failure rate of creep test was calculated [supporting Table 5] [Table 2].

3-6- Pressure test

By inserting the parameters listed in pressure test parameters (pressure failure factor) table into formula 3, the number for AF2 was calculated [supporting Table 6]. Then by inserting parameters into pressure test parameters (temperature failure factor) table, AF1 was calculated [supporting Table 7]. Next, by inserting the amount for accelerator test into formula 8, the amount of AF was calculated. Following that, by inserting AF into formula 4, the amount of lifetime was calculated; furthermore, by inserting the calculated lifetime into formula 9, the failure rate in pressure test was calculated.

3-7- Liner chemical resistance

By inserting the parameters related to liner chemical resistance test into formulas 3 and 4, the operational lifetime was calculated [supporting Table 8]. So, by inserting the calculated amount of lifetime into formula 9, the failure rate for chemical resistance was derived.

3-8- Calculation of vessel reliability, failure rate and lifelong

Table 3 shows failure rate derived from FMEA tables and accelerator test for vessel components. The reliability result, failure rate and total amount of lifetime for all vessels are shown in table 4.

Table 1: *Design of FMEA table for O-Ring*

Potential Failure	Severity	Failure Code	Potential Failure Reasons		Occurrence	RPN	Failure Rate
			Stage 1	Stage 2			
O-ring air leakage	8	NO-1	Improper positioning in the direction of the steal head		5	40	0.002150618
		NO-2	Improper positioning in the direction of the nipple		5	40	0.002150618
		NO-3	Displacement (defects in apparent coordinates)		4	32	0.000995565
		NO-4	Defects in surface smoothness and cleanliness (defects in apparent coordinates)		3	24	0.000460867
		NO-5	Chipped (defects in apparent coordinates)		2	16	0.000213345
		NO-6	Dimensional defects in O-rings	Thickness inaccuracy Diameter inaccuracy	2	16	0.000213345
		NO-7	Improper cocking (pressure resisting/remaining)		4	32	0.000995565
		NO-8	Improper materials (pressure resisting/remaining))		4	32	0.000995565
		NO-9	Roughness of steal head seat surface (defects in steal head (design))		2	16	0.000213345
		NO-10	Remaining tension in O-ring (defects in steal head (design))		2	16	0.000213345
		NO-11	O-ring biting (defects in steal head (design))		5	40	0.002150618
		NO-12	Non-calibration of steal head dosing equipment (track meter) (defects in steal head (design))		3	24	0.000460867
		NO-13	Non-alignment of the liner tab and nipple		3	24	0.000460867
		NO-14	Fatigue		8	64	0.021679243
		NO-15	Damage caused by grease or oil (chemical degradation)		2	16	0.000213345
		NO-16	Ozone cracking (chemical degradation)	Air in workplace Vessel	9	72	0.046831464
		NO-17	Damage caused by thread adhesive (chemical degradation)		4	32	0.000995565
		NO-18	Oxidization (chemical degradation)	Air in workplace due to the penetration through threads O-ring air leakage	2	16	0.000213345

Table 2: *Lifetime results of salt spray, ozone cracking, creep, pressure, liner chemical resistance test*

Salt spray test lifetime results					
TFop	Rate Failure				
72.6504467	0.01376454				
ozone cracking test results					
Failure Rate	TF Op (year)	TF Op (h)	AF		
0.0057239	174.707	1530432.236	5101.4408		
Creep test results					
TF	Failure Rate				
11.5796	0.086358622				
Pressure test results					
AF2	AF1	AF	TF(month)	TF(year)	failure rate
28.88254	2.784722	80.42985	643.4388	26.80995	0.0373
liner chemical resistance test results					
AF	TF Op (month)	TF Op (year)	Failure Rate		
162.4221671	32484.43341	88.9984477	0.0112362		

Table 3: *Vessel components accelerator tests and FMEA table failure rates*

Vessel components	Failure rate from FMEA in 10 years	Failure rate from tests in 10 years
Steal head	0.015667846	0.137645403
Nipple	0.004523398	0.137645403
Wind liner	0.0038346	0.11236151
Composite	0.003825354	0
O-rings	0.081855017	0.92082495
Rotary liner	0.004355093	0
Total (rotary)	0.110226709	1.196115756
Total (wind)	0.109706216	1.494975179

Table 4: *Reliability results, failure rates and lifetime of the entire vessel assembly*

	MAX(λ FMEA,λ ALT)	Reliability in 10 years	Lifetime(years)
Total (rotary)	1.196115756	0.302366399	8.360394843
Total(wind)	1.494975179	0.224254171	6.689074266

4- Discussion

While being extremely high, O-ring failure rate is one of the main reasons in increasing failure rate in vessels. The sum of two failure rates in creep accelerator tests and ozone cracking test increases the failure rate in O-rings. This high failure rate made O-rings a critical component for the vessel set, and without having a plan to cope with failure triggers in O-rings, they are potent to reduce the amount of vessel reliability. The method used to calculate vessel reliability and lifetime; especially, in using and combining several diverse accelerator tests results increased the accuracy and precision of the quantitative information.

By studying previous researches, we realized that according to the method and the selected model, they first identified the factors affecting failure rates by using tools such as FTA, FMEA, etc. Second, they designed and performed accelerator tests in accordance with these factors. Then, based on the previous recognition of the system, they assumed the lifetime distribution. Finally, to estimate the related parameters in lifetime distribution, they either used the least square estimation or the maximum probability estimation methods. Then by using these distributions, they obtained the amount of reliability, failure rate, average time to failure, lifetime, etc.

When wind liner was used, vessel reliability was calculated to be 6.6 years, lifetime was 0/22 in a ten-year period, while, when rotary liner was used, vessel reliability and lifetime were calculated to be 8.3 years, and 0.3 in a ten-year period, respectively, which demonstrated a perspective in how to use the components available. It is also important to identify which component plays a significant role in lowering reliability amount.

In this system, the failure rate in O-ring is calculated very high and this can be very effective in modifying the design for future productions in order to produce a product with higher reliability.

By reviewing various articles, we found that in most researches done so far, different accelerator tests were used to calculate reliability and lifetime in other devices, needless to say, only one accelerator test was conducted.

For example, Regattieri et al. 2017 “Reliability assessment of a packaging automatic machine by accelerated life testing approach” presented a way to evaluate the reliability of an automatic packaging machine using both the accelerated life test (ALT), the Weibull distribution method and the maximum probability method [13]. The result of this study indicates that the accelerated life test is effective for predicting the life of the product through a short-time test, which is in line with the results of the current research.

Yunfeng Li 2017 studies titled 'A range for aircraft accelerator life test based on quasi -mechanical analysis' which provided theoretical foundations for selecting work parameters and forming final judgment of failure rate in the accelerator life test. Using accelerator test along with FTA to spot failure, and quasi-dynamic analysis to determine different working conditions, made it possible to determine the equipment lifelong range [14].

Boro et al., also conducted a research in 2017 ' Strength and Reliability Analysis of Metal-Composite Overwrapped Pressure Vessel'. In order to calculate reliability in the composite layers, creep and

fatigue damage were calculated based on experimental models. Especially, in the creep calculation, the use of accelerated life tests is more accurate than experimental diagrams [15]. For the samples used in performing accelerator tests, it would have been better if more samples were used to increase the accuracy of the output information, but due to the limitation in sample availability, 9 samples were used. The purpose of the study was ultimately to calculate the lifetime and reliability of the vessel. Although the mentioned results are appropriate, by modifying the O-ring in vessel, the amount of lifetime and reliability can definitely increase.

5- Conclusion

The study team concluded that when wind liner was used, vessel reliability was calculated to be 6.6 years, and lifetime was 0/22 in a ten-year period, while, when rotary liner was used, vessel reliability and lifetime were calculated to be 8.3 years, and 0.3 in a ten-year period, respectively, which demonstrated a perspective in how to use the components available.
For failure mechanisms, accelerated destruction tests can be used instead of accelerated life tests. Also, in choosing the O-ring for the vessels, O-rings with higher quality and more reliability can be used to increase reliability and lifetime in equipment.

References

[1] Peng, W., et al., Inverse Gaussian process models for degradation analysis: A Bayesian perspective. Reliability Engineering & System Safety, 2014. 130: p. 175-189.

[2] Sanchez, L.M. and R. Pan. Product robust design via accelerated degradation tests. in 2009 Annual Reliability and Maintainability Symposium. 2009. IEEE.

[3] Chen, H.-t. and H.-j. Yuan. Reliability assessment based on proportional degradation hazards model. in 2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management. 2010. IEEE.

[4] Shen, Y., et al. Accelerated degradation testing for systems with multiple performance parameters. in 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering. 2011. IEEE.

[5] Wang, L., et al., A Bayesian reliability evaluation method with integrated accelerated degradation testing and field information. Reliability Engineering & System Safety, 2013. 112: p. 38-47.

[6] Musgrave, G.E., A. Larsen, and T. Sgobba, Safety design for space systems. 2009: Butterworth-Heinemann.

[7] McLaughlan, P.B., S.C. Forth, and L.R. Grimes-Ledesma, Composite overwrapped pressure vessels, a primer. 2011.

[8] Yassine, A., An introduction to modeling and analyzing complex product development processes using the design structure matrix (DSM) method. Urbana, 2004. 51(9): p. 1-17.

[9] Kailash C. Kapur, M.P., Reliability Engineering. 2014: John Wiley & Sons, Inc.

[10] Mikulak, R.J., R. McDermott, and M. Beauregard, The basics of FMEA. 2017: CRC press.

[11] Kim, K.O., Y. Yang, and M.J. Zuo, A new reliability allocation weight for reducing the occurrence of severe failure effects. Reliability Engineering & System Safety, 2013. 117: p. 81-88.

[12] Mc Pehrson, J., Reliability physics and engineering. 2010, London: Springer.

[13] Regattieri, A., et al., Reliability assessment of a packaging automatic machine by accelerated life testing approach. Procedia Manufacturing, 2017. 11: p. 2178-2186.

[14] Wang, L. and Y. Li, Boundary for aviation bearing accelerated life test based on quasi-dynamic analysis. Tribology International, 2017. 116: p. 414-421.

[15] Burov, A., A. Lepikhin, and V. Moskvichev. Strength and reliability analysis of metal-composite overwrapped pressure vessel. in AIP Conference Proceedings. 2017. AIP Publishing LLC.

CERTAIN RESULTS OF ALEPH- FUNCTION BASED ON NATURAL TRANSFORM OF FRACTIONAL ORDER

Farooq Ahmad[1], D.K. Jain[2], Ajjaz Maqbool[3], Aafaq A. Rather[4,*], Maryam Mohiuddin[5], Priya Deshpande[6], Madhulika Mishra[7], Shaikh Sarfaraj[8]

•

[1,3]Department of Mathematics**,** Govt. College for Women Nawakadal, J&K, 190001, India
[2]Department of Engineering Mathematics Computing, Madhav Institute of Technology and Science, Gwalior Madhya Pradesh-474005, India
[4,6,7]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[5]Department of Mathematics, National Institute of Technology, Calicut, Kerala, India
[8]Symbiosis International University, Symbiosis Institute of Technology, Pune-412115, India
[1]sheikhfarooq85@gmail.com, [2]ain_dkj@mitsgwalior.in, [3]ajjazmaqbool013@gmail.com, [4,*]aafaq7741@gmail.com, [5]masmariam7@gmail.com, [6]priyadeshpande06@gmail.com, [7]madhulika1707@gmail.com, [8]sarfarajs@sitpune.edu.in

Abstract

The paper introduces a new type of fractional integral transform called the N-transform of fractional order. This transform is utilized to derive various results for a more generalized function of fractional calculus known as the Aleph-function. The authors present several useful findings and explore the relationship between the N-transform and other existing fractional transforms. Additionally, the paper discusses the relationship between the N-transform of fractional order and other existing fractional transforms. It likely explores how this new transform relates to established transforms in fractional calculus. The authors have also examined special cases or specific examples to further illustrate the applications and properties of the N-transform of fractional order. These cases could involve particular functions or parameter values that offer insight into the behavior of the transform.

Keywords: N-transform of fractional order, **L**-transform of fractional order, S-transform of fractional order Aleph-function

1. Introduction

Our translation of real world problems to mathematical expressions relies on calculus, which in turn relies on the differentiation and integration operations of arbitrary order with a sort of misnomer fractional calculus which is also a natural generalization of calculus and its mathematical history is equally long. It plays a significant role in number of fields such as physics, rheology, quantitative biology, electro-chemistry, scattering theory, diffusion, transport theory, probability, elasticity, control theory, engineering mathematics and many

others. Fractional calculus like many other mathematical disciplines and ideas has its origin in the quest of researchers for to expand its applications to new fields. This freedom of order opens new dimensions and many problems of applied sciences can be tackled in more efficient way by means of fractional calculus.

Laplace and Sumudu transformations are closely linked to the natural transform. The Natural transform, also known as the N-transform, was initially introduced by Khan and Khan [6]; Al-Omari [1]; Belgacem and Silambarasan [3] explored its features. Maxwell's equations were solved using the Natural transform in Belgacem and Silambarasan [11] and [2]. Transform methods for solving partial differential equations discussed by Duffy [5]. Sharma and Shekhawat [8] obtained integral transform and the Solution of Fractional Kinetic Equation Involving Some Special Functions

Belgacem and Silambarasan's [4] works on the Natural transform can be found here [11] for more information. If we assume that the function is fractional derivative and continuous, the Natural transform often works with continuous and continuously differentiable functions. The Natural transform, like the Laplace and Sumudu transforms, does not work since the function is not derivative. In a similar vein, we must establish a new term that we will call fractional Natural transform.

2. Definitions and Preliminaries

2.1 Natural transform

In mathematics, the natural transform is an integral transform similar to the Laplace transform and Sumudu transform, introduced by khan and khan [6]. It converges to both Laplace and Sumudu transform just by changing variables. Given the convergence to the Laplace and Sumudu transforms, the N-transform inherits all the applied aspects of the both transforms. Most recently, Belgacem [11] has renamed it the natural transform and has proposed a detail theory and applications. The natural transform of a function f(t), defined for all real numbers t ≥ 0, is the function R(u, s), defined by:

$$R(u,s) = N[f(t)] = \int_0^\infty e^{-st} f(ut)dt \ , Re(S) > 0, u(-\tau_1, \tau_2) \quad (1)$$

Provided the function $f(t) \in R^2$ *is defined in the set*

$$A = \{ f(t) \mid \exists M, \tau_1, \tau_2 > 0. |f(t)| < M \ e^{\frac{|t|}{\tau_j}} \} \quad (2)$$

Khan [6] showed that the above integral converges to Laplace transform when $u = 1$, and into Spiegel [7] transform for $s = 1$.

2.2 Fractional Natural transform of order α

$$R_\alpha^+(u,s) = N_\alpha^+[f(x)] = \int_0^\infty E_\alpha(-s^\alpha x^\alpha) f(ux)(dx)^\alpha \ , 0 < \alpha \leq 1 \quad (3)$$

or

$$R_\alpha^+(u,s) = \lim_{M\to\infty} \int_0^M E_\alpha(-s^\alpha x^\alpha) f(ux)(dx)^\alpha \quad (4)$$

where s, u ∈ ℂ, and $E_\alpha(x)$ is the Mittag–Leffler function, $E_\alpha(z) = \sum_{n=0}^\infty \frac{x^n}{\alpha n!}$

2.3 Fractional Laplace transform reported by Estrin and Higgins [10]

From the above definition, when u = 1

$$L_\alpha^+(1, s) = L_\alpha^+[f(x)] = \int_0^\infty E_\alpha(-s^\alpha x^\alpha) f(x)(dx)^\alpha \, , \, 0 < \alpha \leq 1 \tag{5}$$

or

$$L_\alpha^+(1, s) = \lim_{M\to\infty} \int_0^M E_\alpha(-s^\alpha x^\alpha) f(x)(dx)^\alpha \tag{6}$$

Where, s ∈ ℂ, and $E_\alpha(x)$ is the Mittag–Leffler function, $E_\alpha(z) = \sum_{n=0}^{\infty} \frac{x^n}{\alpha n!}$

2.4 Fractional Sumudu transform

From the above definition, when S = 1

$$S_\alpha^+(u, 1) = S_\alpha^+[f(x)] = \int_0^\infty E_\alpha(-x^\alpha) f(ux)(dx)^\alpha \, , \, 0 < \alpha \leq 1 \tag{7}$$

or

$$S_\alpha^+(u, 1) = \lim_{M\to\infty} \int_0^M E_\alpha(-x^\alpha) f(ux)(dx)^\alpha \tag{8}$$

where u ∈ ℂ, and $E_\alpha(x)$ is the Mittag–Leffler function , $E_\alpha(z) = \sum_{n=0}^{\infty} \frac{x^n}{\alpha n!}$

3. Aleph-function

The Aleph-function is defined in terms of the Mellin-Barnes type integral in the following manner is

$$\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i} \\ (b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]$$
$$= \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m} \Gamma(b_j - B_j s) \prod_{j=1}^{n} \Gamma(1 - a_j + A_j s)}{\sum_{i=1}^{r} \tau_i \prod_{j=m+1}^{q_i} \Gamma(1 - b_{ji} + B_{ji} s) \prod_{j=n+1}^{p_i} \Gamma[a_{ji} - A_{ji} s]} z^s \mathrm{ds} \tag{9}$$

Lemma 3.1: For instance the fractional natural transform of the $f(x) = x^{n\alpha}$, $n \in N$ then

$$N_\alpha^+[\, x^{n\alpha}] = \int_0^\infty E_\alpha(-s^\alpha x^\alpha)\, (ux)^{n\alpha}(dx)^\alpha = u^{n\alpha} \int_0^\infty E_\alpha(-s^\alpha x^\alpha)\, (x)^{n\alpha}(dx)^\alpha \tag{10}$$

We put $t = xs$. we get

$$N_\alpha^+[\, x^{n\alpha}] = \frac{u^{n\alpha}}{s^{(n+1)\alpha}} \int_0^\infty E_\alpha(-t^\alpha)\, (t)^{n\alpha}(dt)^\alpha \tag{11}$$

or

$$N_\alpha^+[\, x^{n\alpha}] = \frac{(\alpha!)u^{n\alpha}}{s^{(n+1)\alpha}}\ \Gamma_\alpha(\mathrm{n}+1) \tag{12}$$

Note: $\Gamma_\alpha(\mathrm{n}) = \frac{1}{(\alpha!)} \int_0^\infty E_\alpha(-x^\alpha)\, (x)^{(n-1)\alpha}(dx)^\alpha$

Lemma 3.2**:** For instance the fractional Laplace transform of the $f(x) = x^{n\alpha}$, $n \in N$ then

$$L_\alpha^+[\, x^{n\alpha}] = \int_0^\infty E_\alpha(-s^\alpha x^\alpha)\, (x)^{n\alpha}(dx)^\alpha \tag{13}$$

We put $t = xs$. we get

$$L_\alpha^+[\, x^{n\alpha}] = \frac{1}{s^{(n+1)\alpha}} \int_0^\infty E_\alpha(-t^\alpha)\, (t)^{n\alpha}(dt)^\alpha \tag{14}$$

or

$$L_\alpha^+[\, x^{n\alpha}] = \frac{(\alpha!)}{s^{(n+1)\alpha}}\ \Gamma_\alpha(\mathrm{n}+1) \tag{15}$$

Note: $\Gamma_\alpha(\mathrm{n}) = \frac{1}{(\alpha!)} \int_0^\infty E_\alpha(-x^\alpha)\, (x)^{(n-1)\alpha}(dx)^\alpha$

Lemma 3.3: For instance the fractional Sumudu transform of the $f(x) = x^{n\alpha}$, $n \in N$ then

$$S_\alpha^+[\, x^{n\alpha}] = \int_0^\infty E_\alpha(-x^\alpha)\, (ux)^{n\alpha}(dx)^\alpha = u^{n\alpha} \int_0^\infty E_\alpha(-x^\alpha)\, (x)^{n\alpha}(dx)^\alpha \tag{16}$$

We put $t = x$, we get

$$S_\alpha^+[\, x^{n\alpha}] = u^{n\alpha} \int_0^\infty E_\alpha(-t^\alpha)\, (t)^{n\alpha}(dt)^\alpha \tag{17}$$

or

$$S_\alpha^+[\, x^{n\alpha}] = (\alpha!)u^{n\alpha}\ \Gamma_\alpha(\mathrm{n}+1) \tag{18}$$

Note: $\Gamma_\alpha(\mathrm{n}) = \frac{1}{(\alpha!)} \int_0^\infty E_\alpha(-x^\alpha)\, (x)^{(n-1)\alpha}(dx)^\alpha$

4. Some Main Transformations

4.1 Fractional natural transform of order α

In this section, we derived the fractional natural transform of order α in relationship with the known generalized function of fractional calculus known as Aleph-function.

Theorem 4.1.1: Let $N_\alpha^+[f(x)]$, $0<\alpha\leq 1$, be the fractional natural transform of order α associated with Aleph-function . Then there holds the following relationship

$$N_\alpha^+\left\{\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}\right\}=$$

$$\frac{1}{s}\aleph_{p_i,q_i;\tau_i;r}^{m,n+1}\left[\frac{u}{s}\middle|\begin{matrix}(0,1)(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right] \tag{19}$$

Provided the function $f(t)\in R^2$.

Proof: By using the definition of the generalized function of fractional Aleph -function and fractional natural transform of order α we get

$$N_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$N_\alpha^+\left\{\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j-B_jk)\prod_{j=1}^n \Gamma(1-a_j+A_jk)}{\sum_{i=1}^r \tau_i \prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]} z^k \mathrm{dk}\right\} \quad ; \ \mathrm{Re}(\alpha)>0 \tag{20}$$

$$N_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\left\{\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j-B_jk)\prod_{j=1}^n \Gamma(1-a_j+A_jk)}{\sum_{i=1}^r \tau_i \prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\right\}\ N_\alpha^+\{\mathrm{z^k}\} \tag{21}$$

$$N_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\left\{\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j-B_jk)\prod_{j=1}^n \Gamma(1-a_j+A_jk)}{\sum_{i=1}^r \tau_i \prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\right\}\ N_\alpha^+\{\mathrm{z^k}\} \tag{22}$$

By making use of lemma –3.1 in above equation, we get

$$N_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j-B_jk)\prod_{j=1}^n \Gamma(1-a_j+A_jk)}{\sum_{i=1}^r \tau_i \prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\frac{u^k}{s^{(k+1)}}\ \Gamma(\mathrm{k}+1) \tag{23}$$

or

$$N_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{s}\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j-B_jk)\prod_{j=1}^n \Gamma(1-a_j+A_jk)\Gamma(1-0+\mathrm{k})}{\sum_{i=1}^r \tau_i \prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\frac{u^k}{s^k} \tag{24}$$

or

$$N_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{s}\aleph_{p_i,q_i;\tau_i;r}^{m,n+1}\left[\frac{u}{s}\middle|\begin{matrix}(0,1)(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right] \tag{25}$$

This completes proof of theorem.

4.2 Fractional Laplace transform of order α

In this section, we derived the fractional Laplace transform of order α in relationship with the known function of fractional calculus known as Aleph-function.

Theorem 4.2.1: Let $L_\alpha^+[f(x)]$, $0 < \alpha \leq 1$, be the fractional Laplace transform of order α associated with Aleph-function . Then there holds the following relationship

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{s}\,\aleph_{p_i,q_i;\tau_i;r}^{m,n+1}\left[s^{-1}\middle|\begin{matrix}(1,0)(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right] \tag{26}$$

Provided the function $f(t) \in R^2$.

Proof: By using the definition of the generalized function of fractional ML -function and fractional Laplace transform of order α we get

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$L_\alpha^+\left\{\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}z^k\mathrm{dk}\right\} \quad ;\ \mathrm{Re}(\alpha)>0 \tag{27}$$

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\ L_\alpha^+\{z^k\} \tag{28}$$

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\ L_\alpha^+\{z^k\} \tag{29}$$

By making use of lemma –3.2 in above equation, we get

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\frac{1}{s^{(k+1)}}\ \Gamma(\mathrm{k}+1) \tag{30}$$

or

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)\Gamma(\mathrm{k}+1)\Gamma(1-0+\mathrm{k})}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\frac{1}{s^{(k+1)}}\mathrm{dk} \tag{31}$$

$$L_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\middle|\begin{matrix}(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right]\right\}=$$

$$\frac{1}{s}\,\aleph_{p_i,q_i;\tau_i;r}^{m,n+1}\left[s^{-1}\middle|\begin{matrix}(1,0)(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}\\(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}\end{matrix}\right] \tag{32}$$

This completes proof of theorem.

4.3 Fractional Sumudu transform of order α

In this section, we derived the fractional Sumudu transform of order α in relationship with the known function of fractional calculus known as ML-function.

Theorem 4.3.1: Let $S_\alpha^+[f(x)]$, $0 < \alpha \leq 1$, be the fractional Sumudu transform of order α associated with Aleph-function . Then there holds the following relationship

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_{ji},B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=\frac{1}{s}\,\mathrm{E}_\alpha^1\left(\frac{u}{s}\right) \tag{33}$$

Provided the function $f(t)\in R^2$.

Proof: By using the definition of the generalized function of fractional Aleph -function and fractional Sumudu transform of order α we get

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=$$

$$S_\alpha^+\ \frac{1}{2\pi i}\int_L\ \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}z^k\mathrm{dk};\ \ \mathrm{Re}(\alpha)>0 \tag{34}$$

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L\ \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}\ \ S_\alpha^+\{z^k\} \tag{35}$$

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L\ \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}S_\alpha^+\{\mathrm{z^k}\} \tag{36}$$

By making use of lemma –3.3 in above equation, we get

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L\ \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}\mathrm{dk}u^k\ \ \Gamma(\mathrm{k}+1) \tag{37}$$

or

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=$$

$$\frac{1}{2\pi i}\int_L\ \frac{\prod_{j=1}^{m}\Gamma(b_j-B_jk)\prod_{j=1}^{n}\Gamma(1-a_j+A_jk)\Gamma(1-0+\mathrm{k})}{\sum_{i=1}^{r}\tau_i\prod_{j=m+1}^{q_i}\Gamma(1-b_{ji}+B_{ji}k)\prod_{j=n+1}^{p_i}\Gamma[a_{ji}-A_{ji}k]}u^k\mathrm{dk} \tag{38}$$

$$S_\alpha^+\left\{\aleph_{p_i,q_i;\tau_i;r}^{m,n}\left[Z\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right]\right\}=$$

$$\aleph_{p_i,q_i;\tau_i;r}^{m,n+1}\left[u\Big|_{(b_j,B_j)_{1,m}[\tau_i(b_j,B_{ji})]_{m+1,q_i}}^{(0,1)(a_j,A_j)_{1,n}[\tau_i(a_{ji},A_{ji})]_{n+1,p_i}}\right] \tag{39}$$

This completes proof of theorem.

5. Special Cases

In this section, we discuss some of the important special cases of the main results established discussed above, If we take $\propto=\tau_i=1$ in the theorems (4.1.1), (4.2.1) and (4.3.1) we get well known results of ordinary calculus like Natural transform of Saxen's i-function, Laplace transform of Saxen's i-function and finally ordinary Sumudu transform of Saxen's i-function as reported in [9].

6. Conclusion

This paper introduces a novel type of fractional integral transform called the N-transform of fractional order. This transform is proposed as a new addition to the theory of fractional order transforms. The paper emphasizes that the contributions made by this new transform are

believed to be significant and offer valuable insights to the field. Furthermore, the paper suggests that the N-transform of fractional order has potential applications in solving fractional differential and integral equations. By utilizing this model, it may be possible to find solutions or approaches for various equations involving fractional derivatives or integrals.

Conflict of Interest

The authors declare that they have no conflicts of interest and no personal relationships that could have appeared to influence the work.

References

[1] Al-Omari, S. K. Q. (2013). On the application of natural transforms. *International Journal of Pure and Applied Mathematics, 84*, 729–744.

[2] Belgacem, F. B. M., & Silambarasan, R. (2012a). Advances in the natural transform. *AIP Conference Proceedings, 1493*, 106. doi:10.1063/1.4765477.

[3] Belgacem, F. B. M., & Silambarasan, R. (2012b). Theory of natural transform. *Mathematics In Engineering Science and Aerospace, 3*, 105–135.

[4] Belgacem, F. B. M., & Silambarasan, R. (2012c). Maxwell's equations solutions through the natural transform. *Mathematics in Engineering, Science and Aerospace, 3*, 313–323.

[5] Duffy, D. G., (2004). Transform methods for solving partial differential equations, Second edition, Chapman & Hall/CRC, BocaRaton, FL.

[6] Khan, Z. H., & Khan, W. A. (2008). N-transform-properties and applications. *NUST Journal of Engineering Sciences, 1*, 127–133.

[7] Spiegel, M. R., (1965). Schaum's outline of theory and problems of Laplace transforms, McGraw-Hill, New York.

[8] Sharma, S.K., Shekhawat, A.S. Integral Transform and the Solution of Fractional Kinetic Equation Involving Some Special Functions, Volume-55 Number-2, by IJMTT Journal.

[9] Saxena, R.K., Ram, J. and Kumar, D. (2012). Generalized fractional differentiation for Saigo operators involving Aleph-Function, J. Indian Acad. Math., **34**(1), 109-115.

[10] Estrin, T. A. and Higgins, T. J., (1951). The solution of boundary value problems by multiple Laplace transformations, J. Franklin Inst., 252, 153–167.

[11] Belgacem, F. B. M., & Silambarasan, R. (2011). Theoretical investigations of the natural transform. *Progress In Electromagnetics Research Symposium Proceedings* (pp. 12–16).

RELIABILITY TEST PLAN FOR THE POISSON-POWER LINDLEY DISTRIBUTION

Alphonsa George

Dais George

•

Department of Statistics, St.Thomas College, Palai, Arunapuram, Kerala, India
Department of Statistics, Catholicate College, Pathanamthitta, Kerala, India
alphonsageorge95@gmail.com
daissaji@rediffmail.com

Abstract

In this article, we introduce a new member to Poisson-X family namely, the Poisson-power Lindley distribution. The statistical as well as the distributional properties of the new distribution are studied. The flexibility of the distribution is illustrated by means of real data sets. We also introduce a reliability test plan for acceptance or rejection of a lot of products submitted for inspection when lifetimes follow the new distribution. The minimum sample size using binomial and Poisson approximations, operating characteristic values and minimum ratios of the true value and the required value of the parameter with a given producer's risk are also developed with respect to the newly introduced sampling plans. A real data example is also given to illustrate the sampling plan developed.

Keywords: Poisson-X family, Poisson-power Lindley distribution, Reliability test plan, Producers risk, Operating characteristic function

1. Introduction

Recently researchers shows a special attention to Lindley distribution proposed by Lindley [16] by considering its importance in modelling complex lifetime data. Also, it is one of the well known distribution to analyze failure time data with different shapes of hazard rates. The probability density function and the cumulative distribution function of Lindley distribution are respectively given by,

$$g(t) = \frac{\beta^2}{\beta+1}(1+t)e^{-\beta t}, x > 0, \beta > 0 \tag{1}$$

and

$$G(t) = 1 - \frac{\beta+1+\beta t}{\beta+1}e^{-\beta t} x > 0, \beta > 0. \tag{2}$$

Ghitany et al. [14] studied the Lindley distribution and its applications in the contest of reliability studies and shows that its mathematical properties are more flexible than those of the exponential distribution. Ghitany et al. [13] introduced power Lindley distribution, a new generalization of Lindley distribution by considering the power transformation $X = T^{\frac{1}{\alpha}}$ in Lindley distribution. The pdf and cdf of power Lindley distribution are respectively

$$g(x) = \frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha} \quad x > 0, \alpha \quad \beta > 0 \tag{3}$$

and

$$G(x) = 1 - (1 + \frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}} \quad x > 0, \alpha, \beta > 0. \tag{4}$$

Here, we introduce a new distribution based on power Lindley distribution called Poisson-power Lindley, a new member of Poisson-X family. Tahir et al. [21] developed Poisson-X family from T-X family of distributions introduced by Alzaatreh et al.[4] with the cumulative distribution function

$$F(x) = \int_{a}^{W(G(x))} r(t)dt \tag{5}$$

where W(G(x)) satisfies the conditions

- W(G(x)) ∈ [a,b],
- W(G(x)) is differentiable and monotonically non decreasing and
- W(G(x)) → a as x → $-\infty$ and $W(G(x)) \to b$ as x → ∞.

Here r(t) is the pdf of the random variable T∈ [a,b] for -∞ < a < b < ∞.
For a Poisson-X family, the cdf and pdf of T are given respectively by Tahir et al. [21] as

$$R(t) = \frac{1 - e^{-\delta t^m}}{1 - e^{-\delta}} \tag{6}$$

and

$$r(t) = \frac{\delta m}{1 - e^{-}\delta} t^{m-1} e^{-\delta t^m} - \qquad 0 \leq t \leq 1. \tag{7}$$

Let G(x) be the baseline cdf. By substituting G(x) for the upper limit W(G(x)) and r(t) as (7) with δ=1 , we get the cdf and pdf of Poisson-X family respectively from (5) as

$$F(x; c; \xi) = \left(1 - e^{-1}\right)^{-1} \left[1 - e^{-[G(x)]^m}\right] \tag{8}$$

and

$$f(x; c; \xi) = \frac{m}{1 - e^{-1}} g(x; \xi) \left[G(x; \xi)\right]^{m-1} e^{-[G(x;\xi)]^m}. \tag{9}$$

Here after a random variable X with cdf (9) is denoted by $X \sim PX(m; \xi)$.

Lemma 1. If X have a density of the form (9), then the random variable T=G(x) has a pdf of the form (7) with δ=1. The converse is also true, that is, if T has density of the form (7) with δ=1 then the random variable $X = G^{-1}(T)$ has Poisson-X distribution with density (9).

The rest of the paper is unfolded as follows. In section 2, we introduce Poisson-power Lindley (PPL) distribution and study its statistical properties. We explore the feasibility of the new model by means of real data sets in section 3. In section 4, we developed a reliability test plan and the operating characteristic values for the PPL distribution. We illustrate the results with a real data example. Finally, we conclude the work by section 5.

2. Poisson-power Lindley (PPL) Distribution

For a Poisson-X family of distribution, if the parent distribution is power Lindley, we obtain Poisson-power Lindley distribution having cdf and pdf as

$$F(x) = \left(1 - e^{-1}\right)^{-1} \left[1 - e^{-\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^m}\right] \quad x > 0\ \alpha,\ \beta,\ m > 0 \tag{10}$$

and

$$\begin{aligned} f(x) &= \frac{m}{(1 - e^{-1})} \left[\frac{\alpha\beta^2}{\beta+1}(1 + x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}\right] \\ &\quad \left[e^{-\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^m}\right] \left[1 - (1 + \frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m-1} x > 0\ \alpha,\ \beta, m > 0. \end{aligned} \tag{11}$$

The PDF graphs of the PPL distribution for various values of the parameters are given in Figure 1

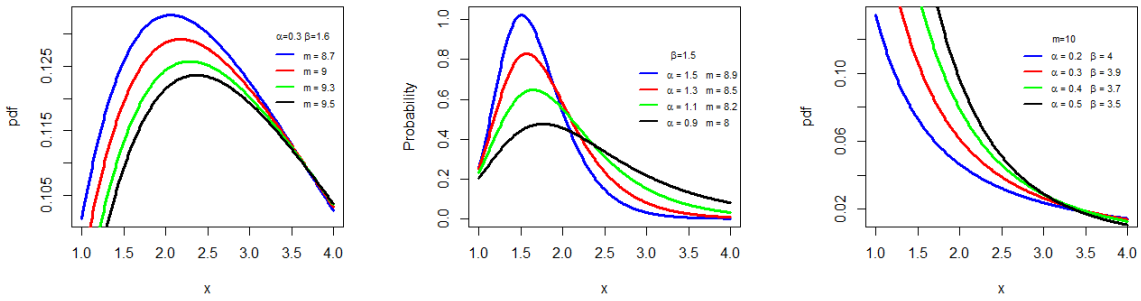


Figure 1: *pdf graphs of Poisson-power Lindley distribution for various values of parameters*

2.1. Properties of Poisson-power Lindley Distribution

The survival function (sf), hazard rate function (hrf), cumulative hazard rate function (chrf), Reversed hazard rate function of Poisson-power Lindley distribution are respectively given by

$$S(x) = 1 - \left(1 - e^{-1}\right)^{-1} \left[1 - e^{-[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}]^{m}}\right], \tag{12}$$

$$h(x) = \frac{m\frac{\alpha\beta^2}{\beta+1}(1+x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m-1}}{1-e^{-(1-[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}]^{m})}}, \tag{13}$$

$$H(x) = -log\left[\frac{e^{-\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m}} - e^{-1}}{1-e^{-1}}\right] \tag{14}$$

and

$$q(x) = \frac{m\frac{\alpha\beta^2}{\beta+1}(1+x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m-1}}{1-e^{-[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}]^{m})}}. \tag{15}$$

The residual life time of the PPL distribution at time t and the corresponding survival function are respectively given by

$$r_{x_t}(x) = \frac{m\frac{\alpha\beta^2}{\beta+1}(1+x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m-1}}{1-e^{-(1-[1-(1+\frac{\beta}{\beta+1}t^{\alpha})e^{-\beta t^{\alpha}}]^{m})}} \tag{16}$$

and

$$R_{x_t}(x) = \frac{e^{-\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m}} - e^{-1}}{e^{-\left[1-(1+\frac{\beta}{\beta+1}t^{\alpha})e^{-\beta t^{\alpha}}\right]^{m}} - e^{-1}}. \tag{17}$$

The past life time and corresponding distribution function of PPL distribution are

$$d_{t_x}(x) = \frac{m\frac{\alpha\beta^2}{\beta+1}(1+x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m-1}}{1-e^{-[1-(1+\frac{\beta}{\beta+1}t^{\alpha})e^{-\beta t^{\alpha}}]^{m})}} \tag{18}$$

and

$$D_{t_x}(x) = \frac{1-e^{-\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m}}}{1-e^{-\left[1-(1+\frac{\beta}{\beta+1}t^{\alpha})e^{-\beta t^{\alpha}}\right]^{m}}}. \tag{19}$$

The hrf graphs of the PPL distribution for different values of the parameters are given in Figure 2

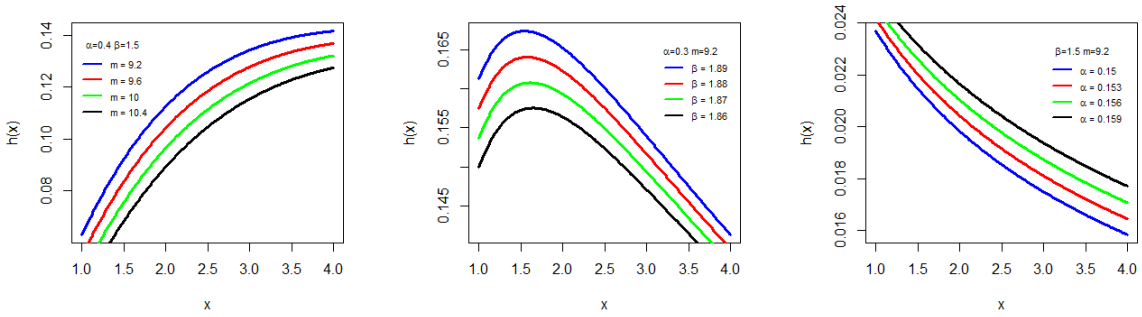


Figure 2: *hrf graphs of Poisson-power Lindley distribution for different values of the parameters*

For fixed values of α and β with different m values, we can see an increasing failure rate here in the hrf graph. Again, for a fixed values of β and m with different α values, we can see a decreasing failure rate. A reverse J-shaped curve can also be seen for fixed values of α and m with different β values.

2.2. Linear Representation

Here we derive some useful expansions for (10) and (11) using the concept of exponentiated distributions. A random variable with an arbitrary baseline cdf G(x) is said to have the exp-G distribution with parameter $\gamma > 0$, if its pdf and cdf are $l_\gamma = \gamma G^{\gamma-1}(x)g(x)$ and $L_\gamma(x) = G^\gamma(x)$, respectively. The exponential function of (10) can be expanded as

$$F(x;c,a,b) = \sum_{i=0}^{\infty} w_{i+1}\left(1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right)^{(i+1)^m} \tag{20}$$

where $w_{i+1} = \frac{((-1)^i)}{[(i+1)!(1-e^{-1})]}$ (for i≥0) thereby the pdf is

$$\begin{aligned} f(x;c,a,b) &= \sum_{i=0}^{\infty} w_{i+1}\left((i+1)m\frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}\right) \\ &\quad \left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^{(i+1)m-1}. \end{aligned} \tag{21}$$

This can also be expressed as

$$f(x;c,a,b) = \sum_{i=0}^{\infty} w_{i+1} l_{(i+1)m}(x) \tag{22}$$

where $l_{(i+1)m}(x)$ is the density function of the exp-power Lindley distribution with power parameter $(i+1)m$.
(22) reveals that the Poisson-power Lindley density can be expressed as a linear representation of exp-power Lindley density.

Lemma 2. If $Y \sim$ power Lindley(α,β), then

$$X = \left[-1-\frac{1}{\beta}-\frac{1}{\beta}W_{-1}\left(\frac{\beta+1}{e^{\beta+1}}\left[1-\left\{-log\left(1-Y\left(1-e^{-1}\right)\right)\right\}^{\frac{1}{m}}\right]\right)\right]^{\frac{1}{\alpha}}$$

follows Poisson-power Lindley (m,α,β). Where $W_{-1}(.)$denotes the negative branch of the Lambert W function.

Proof. Consider the cdf of power Lindley(α,β),

$$F(x) = 1 - (1 + \frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}.$$

If T have the form (7) with $\delta = 1$, then from Lemma 1 we have,

$$\begin{aligned} X &= G^{-1}(T) \\ &= \left[-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}\left[\frac{\beta+1}{e^{\beta+1}}(T-1)\right]\right]^{\frac{1}{\alpha}} \end{aligned} \tag{23}$$

follows Poisson-power Lindley (m,α, β). Here for simulating T, we use R(T) given in (6) with $\delta = 1$, (by inverse probability integral transformation).
We have,

$$T = \left\{-log\left[1 - Y\left(1 - e^{-1}\right)\right]\right\}^{\frac{1}{m}}.$$

and by substituting the value of T in (23), we get

$$X = \left[-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}\left(-\frac{\beta+1}{e^{\beta+1}}\left[1 - \left\{-log\left(1 - Y\left(1 - e^{-1}\right)\right)\right\}^{\frac{1}{m}}\right]\right)\right]^{\frac{1}{\alpha}} \tag{24}$$

which follows Poisson-power Lindley(m,α,β). ■

2.3. Quantile function and Median

The quantile function of Poisson-power Lindley distribution is given by

$$X = \left[-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}\left(-\frac{\beta+1}{e^{\beta+1}}\left[1 - \left\{-log\left(1 - u\left(1 - e^{-1}\right)\right)\right\}^{\frac{1}{m}}\right]\right)\right]^{\frac{1}{\alpha}} \tag{25}$$

Hence the median is,

$$X = \left[-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}\left(-\frac{\beta+1}{e^{\beta+1}}\left[1 - \{0.165\}^{\frac{1}{m}}\right]\right)\right]^{\frac{1}{\alpha}} \tag{26}$$

2.4. Moments

Lemma 3. If X is a random variable from a Poisson-power Lindley (m,α, β) distribution, then the k^{th} moment of X is

$$\mu'_k = E(X^k) = E_t\left[\left(-1 - \frac{1}{\beta} - \frac{1}{\beta}W_{-1}\left(-\frac{\beta+1}{e^{\beta+1}}(1-t)\right)\right)^{\frac{k}{\alpha}}\right]$$

where T follows distribution of the form (6) with $\delta = 1$

Proof.

$$\begin{aligned} E(X^k) \quad = \quad & \int_0^{\infty} \frac{c}{1-e^{-1}}\left[\frac{\alpha\beta^2}{\beta+1}(1+x^{\alpha})x^{\alpha-1}e^{-\beta x^{\alpha}}\right] \\ & \left[1 - (1 + \frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^{m-1} e^{-\left[1-(1+\frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}}\right]^m} \end{aligned}$$

By using Lemma 1). and the transformation

$$\begin{aligned} T &= G(x) \\ &= 1 - (1 + \frac{\beta}{\beta+1}x^{\alpha})e^{-\beta x^{\alpha}} \end{aligned}$$

We have

$$E(X^k) = \int_0^\infty \left(-1-\frac{1}{\beta}-\frac{1}{\beta}W_{-1}\left(-\frac{\beta+1}{e^{\beta+1}}(1-t)\right)\right)^{\frac{k}{\alpha}} \frac{m}{1-e^{-1}} t^{m-1}e^{-t^m}dt$$
$$= E_t\left[\left(-1-\frac{1}{\beta}-\frac{1}{\beta}W_{-1}\left(-\frac{\beta+1}{e^{\beta+1}}(1-t)\right)\right)^{\frac{k}{\alpha}}\right]$$

■

2.5. Order Statistics

Suppose $X_1, ...X_n$ is a random sample from the Poisson-power Lindley distribution. Let $X_{(r)}$ denote the rth order statistics. The pdf of $X_{(r)}$ of PPL distribution can be expressed as

$$\begin{aligned} f_r(x) &= \frac{n!}{(r-1)!\,(n-r)!}\left[\frac{m\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}\right]\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^{m-1} \\ &\quad e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}\sum_{j=0}^{n-r}(-1)^j\binom{n-r}{j}(1-e^{-1})^{-(r+j)} \\ &\quad \left[1-e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}\right]^{r+j-1}. \end{aligned} \tag{27}$$

Now the cdf, $F_n(x)$ of the largest order statistics $X_{(n)}$ is given by,

$$F_n(x) = \left(1-e^{-1}\right)^{-n}\left[1-e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}\right]^n \tag{28}$$

and the corresponding pdf is

$$\begin{aligned} f_n(x) &= nm\left(1-e^{-1}\right)^{-n}\left[1-e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}\right]^{n-1} \\ &\quad \left[\frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}\right]\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^{m-1} \\ &\quad e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}. \end{aligned} \tag{29}$$

Again, the cdf $F_1(x)$ and pdf $f_1(x)$ of the smallest order statistic $X_{(1)}$ is given by,

$$F_1(x) = 1-\left[1-\left(1-e^{-1}\right)^{-1}\left[1-e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}\right]\right]^n \tag{30}$$

and the corresponding pdf is

$$\begin{aligned} f_1(x) &= \frac{nm}{(1-e^{-1})}\left[1-\left(1-e^{-1}\right)^{-1}\left[1-e^{-[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}]^m}\right]\right]^{n-1} \\ &\quad \left[\frac{\alpha\beta^2}{\beta+1}(1+x^\alpha)x^{\alpha-1}e^{-\beta x^\alpha}\right]\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^{m-1} \\ &\quad e^{-\left[1-(1+\frac{\beta}{\beta+1}x^\alpha)e^{-\beta x^\alpha}\right]^m}. \end{aligned} \tag{31}$$

2.6. Parameter Estimation

Here, we use maximum likelihood method of estimation. Let $X_1, X_2...X_n$ be independent and identically distributed Poisson-power Lindley random variables. The log likelihood function is,

$$\begin{aligned} L(\alpha,\beta,m;x) &= n log\left(\frac{m}{1-e^{-1}}\right) + n log\left(\frac{\alpha\beta^2}{\beta+1}\right) + \sum_{i=1}^{n} log(1+x_i^{\alpha}) + (\alpha-1) \\ & \sum_{i=1}^{n} log(x_i) - \beta\sum_{i=1}^{n} x_i^{\alpha} - \sum_{i=1}^{n}\left[1-\left(1+\frac{\beta x_i^{\alpha}}{\beta+1}\right)e^{-\beta x_i^{\alpha}}\right] + (m-1) \\ & \sum_{i=1}^{n} log\left[1-\left(1+\frac{\beta x_i^{\alpha}}{\beta+1}\right)e^{-\beta x_i^{\alpha}}\right]. \end{aligned} \tag{32}$$

The computations were implemented using the nlm function of R software.

3. APPLICATIONS

In this section, we illustrate the flexibility of the Poisson-power Lidley distribution using two real data sets. The model parameters are estimated by the method of maximum likelihood using the nlm function of R software. The K-S statistics and the corresponding p values were also calculated for establishing its goodness of fit.

3.1. Dataset 1 Repair times data

We have considered a data set which represents the maintenance data with 46 observations reported on active repair times (hours) for an airborne communication transceiver given by [17]. The data set is
0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0,1.0, 1.0, 1.0, 1.1 1.3, 1.5, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2, 2.5, 2.7,3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8,9.0, 10.3, 22.0, 24.5.
We plot the histogram of the observed data and the embedded pdf plot of PPL disribution. It is seen that the PPL distribution is a good fit for the observed data. For comparison study we include the embedded pdfs of power Lindley(PL), Poisson Weibull(PW) and Inverse power Lindly(IPL) distributions also in the same graph and it is given in Figure 3

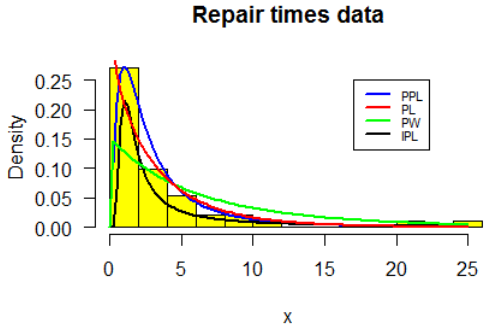


Figure 3: *Fitted pdf plot of Dataset 1*

The numerical values of statistics of the fitted models for Dataset 1 are presented in Table 1. We note that the PPL distribution has the lowest value of -logl, AIC, BIC, K-S and highest p value as compared to power Lindley(PL), Poisson Weibull(PW) and Inverse power Lindly(IPL)

distributions. Therefore PPL distribution provides a better fit for the given Dataset 1 than the other considered distributions.

Table 1: *The MLE, -logl, AIC, BIC, K-S and p-value for the fitted models to the Dataset1*

Distribution	Parameters	-logl	AIC	BIC	K-S	p value
PPL	$\alpha = 0.4218$ $\beta = 1.9947$ m=8.2000	0.0635	6.1270	11.6130	0.28986	0.972
PL	$\alpha = 0.7581$ $\beta = 0.6757$	105.0133	214.0266	217.6839	0.76087	0.2171
PW	$\alpha = 0.4218$ $\beta = 0.1249$ $\gamma = 1.0004$	112.6021	231.2042	236.6901	0.63043	0.2129
IPL	$\alpha = 1.6899$ $\beta = 0.6222$	69.3208	142.6417	146.302	0.5	0.7241

3.2. Dataset 2 Carbon fibers data

The data provides the tensile strength of 69 carbon fibers, measured in GPa, tested under tension at gauge given by [13]
1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.997, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.14, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.57, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.88, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585, 3.585.
The histogram and the embedded pdfs of Poisson-power Lindley (PPL), power Lindly (PL), Poisson Exponential (PE) and Exponentiated power Lindley (EPL) distributions for Dataset 2 are give in Figure 4.

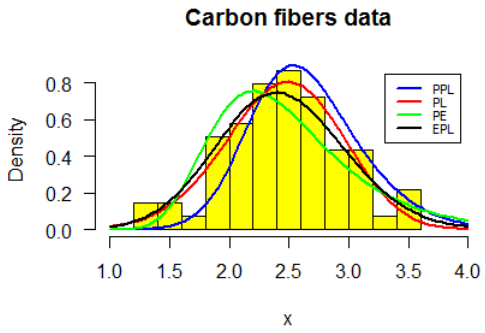


Figure 4: *Fitted pdf plot of Dataset 2*

The MLE, -logl, AIC, BIC, K-S and p-value for the fitted models are presented in Table 2. Here also PPL distribution seems to be the better fit as it has the lowest value of -logl, AIC, BIC, K-S and highest p value when compared to power Lindly (PL), Poisson Exponential (PE), Exponentiated power Lindley (EPL) distributions.

Table 2: *The MLE, -logl, AIC, BIC, K-S and p-value for the fitted models to the Dataset 2*

Distribution	Parameters	-logl	AIC	BIC	K-S	p value
PPL	$\alpha = 2.3845$ $\beta = 0.3095$ m=5.1124	46.67928	99.35855	106.0609	0.3333	0.9068
PL	$\alpha = 3.868$ $\beta = 0.050$	49.06274	102.1255	106.5937	0.5	0.7161
PE	$\theta = 91.2776$ $\lambda = 2.0476$	54.4334	112.8668	117.3351	0.5	0.7161
EPL	$\beta = 2.5554$ $\theta = 0.2487$ $\alpha = 2.2514$	117.7761	241.5522	248.2545	0.42029	0.6901

That is in the case of these two data sets, PPL distribution seems to be the better fit.

4. RELIABILITY TEST PLAN

In Reliability test plan, we want to decide whether to accept or not to accept the lot based on a sample of products taken from the lot. In a life testing experiment, the procedure is to terminate the test by a predetermined time t and note the number of failures. If the number of failures at the end of time t does not exceed a given number c, called acceptance number then we accept the lot with a given probability of at least p^*. But if the number of failures exceeds c before time t then the test is terminated and the lot is rejected. For such truncated life test and the associated decision rule, we are interested in obtaining the smallest sample size to arrive at a decision. Assume that the lifetime of a product follows Poisson-power Lindley distribution. If a scale parameter $\lambda > 0$ is introduced, the cdf of PPL is given by,

$$G(x;m;\xi) = \left(1-e^{-1}\right)^{-1}\left[1-e^{-\left[1-(1+\frac{\beta}{\beta+1}(\frac{x}{\lambda})^{\alpha})e^{-\beta(\frac{x}{\lambda})^{\alpha}}\right]^{m}}\right], x>0, \alpha, \beta, m>0. \tag{33}$$

The average life time depends only on λ if α,β and m are known. Let λ_0 be the required minimum average life time. Then

$$G(t,\alpha,\beta,m,\lambda) \leq G(t,\alpha,\beta,m,\lambda_0) \Leftrightarrow \lambda \geq \lambda_0$$

A sampling plan consists of the following quantities: (1) the number of units n on test; (2) the acceptance number c; (3) the maximum test duration t, and (4) the minimum average lifetime represented by λ_0. The probability of accepting a bad lot, that is the consumer's risk should not exceed the value $1-p^*$ where p^* is a lower bound for the probability that a lot of true value λ below λ_0 is rejected by the sampling plan. For fixed p^* the sampling plan is characterized by $(n, c, \frac{t}{\lambda_0})$. By sufficiently large lots we can apply binomial distribution to find acceptance probability. The problem is to determine the smallest positive integer n for given value of c and $\frac{t}{\lambda_0}$ such that

$$L(P_o) = \sum_{i=0}^{c}\binom{n}{i}p_0^i(1-p_0)^{n-i} \leq 1-p^* \tag{34}$$

where $p_0 = G(t,\alpha,\beta,m,\lambda_0)$. The function $L(p)$ is the operating characteristic function of the sampling plan i.e. the acceptance probability of the lot as a function of the failure probability $p(\lambda) = G(t,\alpha,\beta,m,\lambda)$. The operating characteristic function is an increasing function in λ as the average life time of the product is increasing with λ and thus the failure probability $p(\lambda)$ decreases. Table 3 gives the minimum values of n satisfying (34) for $\alpha = 1$, $\beta = 2$, $m = 2$, p^*=

0.75, 0.90, 0.95, 0.99 and $\frac{t}{\lambda_0}$= 0.628, 0.942, 1.257, 1.571, 2.356, 3.141, 3.927, 4.712.

$$L_1(p_0) = \sum_{i=0}^{c} \frac{\theta^i}{i!} e^{-\theta} \leq 1 - p^*. \tag{35}$$

The minimum values of n satisfying (35) are obtained for the same combination of values of $\alpha, \beta, m, \frac{t}{\lambda_0}$ and p^* when the binomial probability approximated by Poisson probability with parameter $\theta = np_0$ in the case of $p_0 = G(t, \alpha, \beta, m, \lambda_0)$ is very small and n is large, are presented in Table 4. The operating characteristic function of the sampling plan $(n, c, \frac{t}{\lambda_0})$ gives the probability $L(p)$ of accepting the lot with

$$L(P) = \sum_{i=0}^{c} \binom{n}{i} p_0^i (1 - p_0)^{n-i} \tag{36}$$

where $p = G(t, \lambda)$ is considered as a function of λ. The values of n and c are determined by means of operating characteristic function for given values of p^*, $\frac{t}{\lambda_0}$ and considering $p = G\left(\frac{t}{\lambda_0} / \frac{\lambda_0}{\lambda}\right)$ are displayed in Table 5. The producer's risk, probability of rejecting a lot having the quality $\lambda \geq \lambda_0$ is specified by a value say 0.05, it is interested to know that which values of $\frac{\lambda}{\lambda_0}$ will ensure a producer's risk less than or equal to 0.05 for a given sampling plan. The value of $\frac{\lambda}{\lambda_0}$ is the smallest positive number for which the following inequality

$$\sum_{i=0}^{c} \binom{n}{i} p_0^i (1 - p_0)^{n-i} \geq 0.95 \tag{37}$$

holds. For some sampling plan $\left(n, c, \frac{t}{\lambda_0} / \frac{\lambda}{\lambda_0}\right)$ and values of p^*, the minimum values of $\frac{\lambda}{\lambda_0}$ satisfying (37) are given in Table 6 and Table 7 gives the values of L(p) for given values of $\frac{\lambda}{\lambda_0}$

Table 3: *Minimum sample sizes using Binomial probabilities*

		t/λ_0							
p^*	c	0.628	0.942	0.1.257	1.571	2.356	3.141	3.927	4.712
	0	3	2	1	1	1	1	1	1
	1	5	3	3	2	2	2	2	2
	2	8	5	4	4	3	3	3	3
	3	10	7	5	5	4	4	4	4
	4	13	8	7	6	5	5	5	5
0.75	5	15	10	8	7	6	6	6	6
	6	17	11	9	8	7	7	7	7
	7	19	13	11	10	8	8	8	8
	8	22	15	12	11	9	9	9	9
	9	24	16	13	12	10	10	10	10
	10	26	18	14	13	12	11	11	11
	0	4	2	2	2	1	1	1	1
	1	7	4	3	3	2	2	2	2
	2	10	6	5	4	3	3	3	3
	3	13	8	6	5	5	4	4	4
	4	15	10	8	7	6	5	5	5
0.90	5	18	11	9	8	7	6	6	6
	6	20	13	10	9	8	7	7	7
	7	23	15	12	10	9	8	8	8
	8	25	16	13	12	10	9	10	10
	9	28	18	14	13	11	10	10	10
	10	30	20	16	14	12	11	11	11
	0	5	3	2	2	1	1	1	1
	1	9	5	4	3	3	2	2	2
	2	11	7	5	5	4	3	3	3
	3	14	9	7	6	5	4	4	4
	4	17	11	8	7	6	5	5	5
0.95	5	20	13	10	8	7	6	6	6
	6	22	15	11	10	8	7	7	7
	7	25	16	13	11	9	9	8	8
	8	27	18	14	12	10	10	9	9
	9	21	15	13	12	12	11	12	11
	10	32	21	17	15	12	12	11	11
	0	8	4	3	3	2	1	1	1
	1	11	7	5	4	3	2	2	2
	2	15	9	7	6	4	3	3	3
	3	18	11	8	7	5	4	4	4
	4	21	13	10	8	7	5	5	5
0.99	5	24	15	11	10	8	6	6	6
	6	26	17	13	11	9	7	7	7
	7	29	18	14	12	10	9	8	8
	8	32	20	16	13	11	10	9	9
	9	35	22	17	15	12	11	10	10
	10	37	24	19	16	13	12	11	11

Table 4: *Minimum sample sizes using Poisson probabilities*

		t/λ_0							
p^*	c	0.628	0.948	1.257	1.571	2.356	3.141	3.927	4.712
	0	4	3	2	2	2	2	2	2
	1	6	4	4	4	3	3	3	3
	2	9	6	5	5	5	4	4	4
	3	11	8	7	6	6	6	6	6
	4	14	10	8	8	7	7	7	7
0.75	5	16	11	10	9	8	8	8	8
	6	19	13	11	10	9	9	9	9
	7	21	15	12	11	11	10	10	10
	8	24	16	14	13	12	11	11	11
	9	19	15	14	13	13	13	13	13
	10	28	20	16	15	14	14	14	14
	0	5	4	3	3	3	3	3	3
	1	9	6	5	5	4	4	4	4
	2	12	8	7	6	6	6	6	6
	3	15	10	9	8	7	7	7	7
	4	17	12	10	9	9	9	9	8
0.90	5	20	14	12	11	10	10	10	10
	6	23	16	13	12	11	11	11	11
	7	25	18	15	14	13	12	12	12
	8	28	19	16	15	14	14	14	13
	9	31	21	18	16	15	15	15	15
	10	33	23	19	18	16	16	16	16
	0	7	5	4	4	4	4	4	4
	1	11	8	7	6	6	6	5	5
	2	14	10	9	8	7	7	7	7
	3	15	12	10	10	9	9	9	8
	4	17	14	12	11	10	10	10	10
0.95	5	20	16	14	13	12	12	11	11
	6	23	18	16	14	13	13	13	13
	7	25	20	17	16	14	14	14	14
	8	28	22	19	17	16	15	15	15
	9	31	24	20	19	17	17	17	17
	10	33	26	22	20	18	18	18	18
	0	10	7	6	6	5	5	5	5
	1	15	10	9	8	7	7	7	7
	2	18	13	11	10	9	9	9	9
	3	22	15	13	12	11	11	11	11
	4	25	17	15	13	12	12	12	12
0.99	5	28	20	17	15	14	14	14	14
	6	31	22	18	17	15	15	15	15
	7	24	22	20	18	17	17	17	17
	8	37	26	22	20	18	18	18	18
	9	40	28	24	22	20	19	19	19
	10	43	30	25	23	21	21	21	21

Table 5: *OC values for the plan* $(n, c, t/\lambda_0)$ *for given confidence level* $p^\star$, $\alpha = 1$, $m = \beta = 2$

$p^\star$	n	c	$\frac{t}{\lambda_0}$	λ/λ_0					
				2	4	6	8	10	12
	8	2	0.628	0.8249	0.9913	0.9989	0.9997	0.9999	0.9999
	5	2	0.942	0.7877	0.9865	0.9982	0.9996	0.9998	0.9999
	4	2	1.257	0.7263	0.9774	0.9966	0.9992	0.9997	0.9999
	4	2	1.571	0.5416	0.9427	0.9902	0.9976	0.9992	0.9997
	3	2	2.356	0.5095	0.9145	0.9822	0.9951	0.9983	0.9993
0.75	3	2	3.141	0.2885	0.7945	0.9438	0.9822	0.9934	0.9972
	3	2	3.927	0.1537	0.6502	0.8792	0.95589	0.9822	0.9920
	3	2	4.712	0.0793	0.5095	0.7945	0.9145	0.9623	0.9822
	10	2	0.628	0.7168	0.9829	0.9978	0.9995	0.9998	0.9999
	6	2	0.942	0.6772	0.9754	0.9966	0.9992	0.9997	0.9999
	5	2	1.257	0.5498	0.9514	0.9923	0.9982	0.9994	0.9998
	4	2	1.571	0.5416	0.9427	0.9902	0.9976	0.9992	0.9997
	3	2	2.356	0.5095	0.9145	0.9822	0.9951	0.9983	0.9993
0.90	3	2	3.141	0.2885	0.7945	0.9438	0.9822	0.9934	0.9972
	3	2	3.927	0.1537	0.6502	0.8792	0.95589	0.9822	0.9920
	3	2	4.712	0.0793	0.5095	0.7945	0.9145	0.9623	0.9822
	11	2	0.628	0.6606	0.9775	0.9971	0.9993	0.9998	0.9999
	7	2	0.942	0.5669	0.9607	0.9943	0.9987	0.9996	0.9998
	5	2	1.257	0.5498	0.9514	0.9923	0.9982	0.9994	0.9998
	5	2	1.571	0.3345	0.8843	0.9780	0.9944	0.9982	0.9993
	4	2	2.356	0.1985	0.7711	0.9427	0.9831	0.9940	0.9997
0.95	3	2	3.141	0.2885	0.7945	0.9438	0.9822	0.9934	0.9972
	3	2	3.927	0.1537	0.6502	0.8792	0.95589	0.9822	0.9920
	3	2	4.712	0.0793	0.5095	0.7945	0.9145	0.9623	0.9822
	15	2	0.628	0.4485	0.9477	0.9926	0.9984	0.9995	0.9998
	7	2	0.942	0.3732	0.9211	0.9875	0.9971	0.9991	0.9996
	5	2	1.257	0.2725	0.8732	0.9768	0.9942	0.9982	0.9993
	6	2	1.571	0.1930	0.8125	0.9606	0.9895	0.9965	0.9986
	4	2	2.356	0.1985	0.7711	0.9427	0.9831	0.9940	0.9997
0.99	4	2	3.141	0.0595	0.5419	0.8398	0.9427	0.9774	0.9902
	3	2	3.927	0.1537	0.6502	0.8792	0.95589	0.9822	0.9920
	3	2	4.712	0.0793	0.5095	0.7945	0.9145	0.9623	0.9822

Table 6: *Minimum ratio of true mean life to specified mean life for the acceptability of a lot with $\alpha = 0.05$*

		t/λ_0							
p^*	c	0.628	0.942	0.1.257	1.571	2.356	3.141	3.927	4.712
	0	0.31235	0.46621	0.62574	0.78206	1.1728	1.5636	1.9549	2.3456
	1	0.30953	0.46696	0.62487	0.77886	1.1680	1.5572	1.9469	2.3360
	2	0.30830	0.46375	0.61861	0.77314	1.1622	1.5495	1.9372	2.3245
	3	0.30689	0.45844	0.6152	0.7689	1.1508	1.5343	1.9182	2.3017
	4	0.3032	0.4573	0.6121	0.7651	1.1450	1.5266	1.9086	2.2901
0.75	5	0.3029	0.4567	0.6076	0.7619	1.1374	1.5164	1.8959	2.2748
	6	0.2989	0.4534	0.6049	0.7535	1.0983	1.4642	1.8307	2.1966
	7	0.2971	0.4502	0.5996	0.7477	1.1234	1.4977	1.8725	2.2468
	8	0.2961	0.4484	0.5965	0.7442	1.1169	1.4891	1.8617	2.2339
	9	0.2945	0.4440	0.5916	0.7170	1.1098	1.4796	1.8499	2.2197
	10	0.2943	0.4412	0.5931	0.7443	1.1029	1.4743	1.8433	2.2117
	0	0.31261	0.46768	0.62616	0.78188	1.17351	1.56451	1.95602	2.34702
	1	0.31178	0.46738	0.62537	0.78159	1.17103	1.56121	1.95188	2.34206
	2	0.30937	0.46401	0.62008	0.77561	1.16362	1.5133	1.93953	2.32724
	3	0.30850	0.46235	0.61821	0.77131	1.15673	1.54169	1.92748	2.31278
	4	0.30644	0.45863	0.61273	0.76771	1.15097	1.53054	1.91354	2.29605
0.90	5	0.30465	0.45734	0.60887	0.76212	1.13909	1.52488	1.90647	2.28757
	6	0.30254	0.45466	0.60539	0.75557	1.13375	1.50856	1.88606	2.26308
	7	0.30015	0.45209	0.60226	0.75048	1.12580	1.50019	1.87559	2.25052
	8	0.29960	0.44887	0.59802	0.74859	1.12348	1.49781	1.87262	2.24696
	9	0.29697	0.44492	0.59328	0.74248	1.11331	1.48242	1.85338	2.22387
	10	0.29489	0.44304	0.59198	0.73818	1.10732	1.48019	1.85060	2.22053
	0	0.31342	0.46930	0.62630	0.78275	1.17417	1.56540	1.95713	2.34835
	1	0.31301	0.46742	0.62571	0.78190	1.17260	1.56200	1.95288	2.34325
	2	0.31285	0.46638	0.62510	0.78125	1.17059	1.56107	1.95172	2.34186
	3	0.31249	0.46599	0.62489	0.77827	1.16723	1.55472	1.94378	2.33233
	4	0.31058	0.46438	0.62139	0.77449	1.16150	1.55442	1.94339	2.33187
0.95	5	0.30982	0.46382	0.61888	0.77433	1.16073	1.54632	1.93327	2.31973
	6	0.30889	0.46296	0.61848	0.77169	1.15626	1.54544	1.93217	2.31841
	7	0.30835	0.46215	0.61767	0.77075	1.15323	1.53748	1.92512	2.30994
	8	0.30808	0.46042	0.61302	0.76404	1.14812	1.53067	1.91120	2.29324
	9	0.30635	0.45935	0.61247	0.76234	1.14329	1.52422	1.90252	2.28284
	10	0.30533	0.45755	0.61117	0.76219	1.14322	1.52414	1.89925	2.27890
	0	0.31410	0.47158	0.62932	0.78653	1.17517	1.57161	1.96488	2.35766
	1	0.31349	0.46950	0.62815	0.78583	1.17293	1.56519	1.95686	2.3480
	2	0.31322	0.46912	0.62662	0.78525	1.17282	1.56216	1.95308	2.34349
	3	0.31264	0.46901	0.62634	0.78099	1.1697	1.55585	1.94518	2.33402
	4	0.31248	0.46875	0.62437	0.77818	1.16764	1.55522	1.94440	2.33308
0.99	5	0.31131	0.46633	0.62203	0.77685	1.16330	1.55066	1.93870	2.32624
	6	0.31025	0.46618	0.61933	0.77598	1.16289	1.55015	1.93806	2.32547
	7	0.31009	0.46322	0.61857	0.77578	1.15937	1.54816	1.93782	2.32519
	8	0.30864	0.46104	0.616415	0.76888	1.15524	1.54185	1.92134	2.30541
	9	0.30646	0.46035	0.61599	0.76727	1.15464	1.53299	1.91697	2.30017
	10	0.30629	0.45845	0.61446	0.76669	1.14908	1.53029	1.91078	2.29275

4.1. Description of the Tables

Assume that the lifetime distribution is Poisson-power Lindley distribution with $\alpha = 1$, $\beta = 2$, $m = 2$, and that the experimenter irested in establishing that the true unknown average life is at

least 1000 hours with confidence $p^* = 0.75$. It is desired to stop the experiment at t = 628 hours. Then, for an acceptance number $c = 2$, the required n in Table 3 is 8. If, during 628 hours, no more than 2 failures out of 8 are observed, then the experimenter can assert, with a confidence level of 0.75 that the average life is at least 1000 hours. For the same situation we obtained the value of n = 9 from Table 4when the Poisson approximation to binomial probability is used. Comparing with reliability test plan for the two parameter Quasi Lindley distribution[1] and three parameter Lindley distribution [2] the minimum sample size using binomial approximation for the sampling plan $c = 10$, $\frac{t}{\lambda_0} = 0.628$ with confidence level $p^* = 0.75$ are 29 and 125 respectively, whereas for the Poisson-power Lindley distribution it is 26. This indicate that the newly developed reliability test plan gives a propitious improvement in making optimal decisions as compared to the other two distributions. For the sampling plan $(n = 8, c = 2, \frac{t}{\lambda_0} = 0.628)$ and confidence level $p^* = 0.75$ under Poisson power Lindley distribution with $\alpha = 1$,$\beta = 2$,$m = 2$, the values of the operating characteristic function from Table 5 are as follows:

Table 7: *Values of L(p) for various values of* $\frac{\lambda}{\lambda_0}$

$\frac{\lambda}{\lambda_0}$	2	4	6	8	10	12
L(p)	0.8249	0.9913	0.9989	0.9997	0.9999	0.9999

From Table 7 we can find that if the true mean lifetime is twice the required mean lifetime ($\frac{\lambda}{\lambda_0} = 2$) the producer's risk is approximately 0.1751. We can get the values of the ratio $\frac{\lambda}{\lambda_0}$ for various choices of $(c, \frac{t}{\lambda_0})$ in order that the producer's risk may not exceed 0.05, for example if $p^* = 0.75$, $\frac{t}{\lambda_0} = 4.712, c = 2$, Table 6 gives a reading of 2.3245. This means that the product can have an average life of 2.32 times the required average lifetime in order that under the above acceptance sampling plan the product is accepted with probability of at least 0.95. The actual average lifetime necessary to accept 95 percent of the lots is provided in Table 6.

4.2. Real Data Example

Consider the data studied by Bjerkeda [10] which represents the survival times (in hours) of guinea pigs infected with virulent tubercle bacilli, after they were injected with a given dose of tubercule bacilli in a medical experiment. This data can be regarded as an ordered sample as given below:
43, 45, 53, 56,56,57, 58, 66, 67, 73, 74, 79, 80, 80, 81, 81, 81, 82, 83, 83, 84, 88, 89, 91, 91, 92, 92, 97, 99, 99, 100, 100, 101, 102, 102, 102, 103, 104, 107, 108, 109, 113, 114, 118, 121, 123, 126, 128, 137, 138, 139, 144, 145, 147, 156, 162, 174, 178, 179, 184, 191, 198, 211, 214, 243, 249, 329, 380, 403, 511, 522, 598
Let the experimenter is interested to establish the true unknown mean life time is 70 hours with confidence $p^* = 0.75$ and testing time be 44 hours, which gives the ratio $\frac{t}{\lambda_0} = \frac{44}{70} = 0.628$.. Thus, for an acceptance number c=8 and confidence level $p^* = 0.75$, the required sample size n from Table 3 is found to be 22. Thus the sampling plan for the above sample data is $\left(n = 22, c = 8, \frac{t}{\lambda_0} = 0.628\right)$. Based on the data, we have to decide whether to accept the drug or reject it. We accept the drug only if the number of survival before 44 hours is greater than or equal to 14 among the first 22 observations. However, the confidence level is assured by the sampling plan only if the given lifetimes following the Poisson-power Lindley distribution. Thus in order to confirm it, the goodness of fit test for these observations were done and it gives a p-value 0.6889 and K-S statistics as 0.41667. The corresponding fitted embedded graph is given in Figure 5. Since there is only 2 failures before the time 45, we can say that in a testing time of 44 hours, 20 were survived among 22 observations. So we can accept the effectiveness of tubercle bacilli in developing acquired resistance against tuberculosis in guinea pigs according to our

sampling plan.

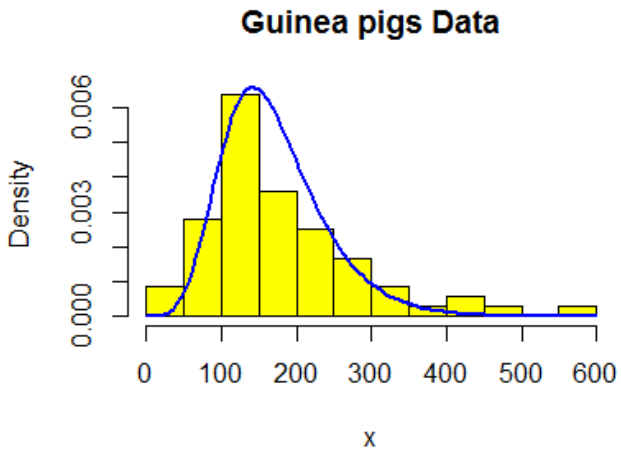


Figure 5: *fitted pdf plot for guinea pigs data*

5. CONCLUSION

In this paper, we introduce a new member of Poisson-X family called Poisson-power Lindley distribution. Some of its structural properties are investigated. The model parameters are estimated by the method of maximum likelihood. The flexibility of the new distribution is illustrated by means of real data sets. It is seen that the Poisson- power Lidley distribution provides a better fit than other compared distributions for these data sets. Also, a reliability test plan is developed when the life time follows the Poisson-power Lindley distribution. We have shown in general that under similar conditions, in order to ensure a specified mean life with a given confidence level, Poisson-power Lindley model results in smaller sample sizes than some other models used in acceptance sampling. In order to verify the applicability of the distribution in reliability test plan a real data analysis is also conducted and it found to be given as a better decision regarding the acceptability of the product.

References

[1] Al-Omari, A. I., and Al-Nasser, A.(2019). A two-parameter quasi Lindley distribution in acceptance sampling plans from truncated life tests,*Pakistan Journal of Statistics and Operation Research*,39–47.

[2] Al-Omari, A. I., Ciavolino, E., and Al-Nasser, A. D. (2019).Economic design of acceptance sampling plans for truncated life tests using three-parameter Lindley distribution, *Journal of Modern Applied Statistical Methods*, **18**,2,16.

[3] Altun, E.(2019). A new model for over- dispersed count data: Poisson quasi- Lindley regression model, *Mathematical Sciences*,1–7.

[4] Alzaatreh, A., Lee, c., Famoye, F. (2013). A new methods for generating families of distributions,*Metron*, **71**,1,63–79.

[5] Ashour, S. K., and Eltehiwy, M. A.(2015). xponentiated power Lindley distribution. *Journal of advanced research*, **6**,6,895–905.

[6] Balakrishnan, N., Buono, F., and Longobardi, M. (2022). On weighted extropies. *Communications in Statistics-Theory and Methods*, **51**, 18, 6250–6267.

[7] Balakrishnan, N., Leiva, V., and L?pez, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum–Saunders distribution. *Communications in Statistics—Simulation and Computation*, **36**, 3, 643–656.

[8] Barco, K. V. P., Mazucheli, J., and Janeiro, V. (2017). The inverse power Lindley distribution. *Communications in Statistics-Simulation and Computation*, **46**, 8, 6308–6323.

[9] Bereta, E. M., Louzanda, F., and Franco, M. A. (2011). The poisson-weibull distribution. *Advances and Applications in Statistics*, **22**, 2, 107–118.

[10] Bjerkedal, T. (1960). Acquisition of Resistance in Guinea Pies infected with Different Doses of Virulent Tubercle Bacilli. *American Journal of Hygiene*, **72**, 1, 130–48.

[11] Cancho, V. G., Louzada-Neto, F., and Barriga, G. D. (2011). The Poisson-exponential lifetime distribution. *Computational Statistics and Data Analysis*, **55**, 1, 677–686.

[12] George, A., and George, D. (2022). The Poisson-Uniform Distribution and its Applications. *International Journal of Statistics and Reliability Engineering*, **9**, 2, 206–212.

[13] Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., and Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference.*Computational Statistics and Data Analysis*, **64**, 20–33.

[14] Ghitany, M. E., Atieh, B., and Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics and computers in simulation*, **78**, 4, 493–506.

[15] Grine, R., and Zeghdoudi, H. (2017). On Poisson quasi-Lindley distribution and its applications.*Journal of modern applied statistical methods*, **16**, 2, 21.

[16] Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 102–107.

[17] Pararai, M., Warahena-Liyanage, G., and Oluyede, B. O. (2015). A new class of generalized power Lindley distribution with applications to lifetime data. *Theoretical mathematics and applications*, **5**, 1, 53.

[18] Rao, G. S. (2013). Acceptance sampling plans from truncated life tests based on the Marshall–Olkin extended exponential distribution for percentiles. *Brazilian Journal of Probability and Statistics*, **27**, 2, 117–132.

[19] Shahbaz, S. H., Khan, K., and Shahbaz, M. Q. (2018). Acceptance sampling plans for finite and infinite lot size under power Lindley distribution. *Symmetry*, **10**, 10, 496.

[20] Tahir, M. H., Cordeiro, G. M., Alzaatreh, A., Mansoor, M., and Zubair, M. (2016). The logistic-X family of distributions and its applications. *Communications in statistics-Theory and methods*, **45**, 24, 7326–7349.

[21] Tahir, M., H., Zubair, M., Cordeiro, Gauss., M., Alzaatreh, A., Mansoon, M. (2016).The Poisson-X family of distributions. *Journal of Statistical Computation and Simulation*, **86**, 4, 2901–2921.

[22] Thomas, S. P., Tomy, L., and Jose, K. K. (2021). Harris Extended Two Parameter Lindley Distribution and Applications in Reliability. *Reliability: Theory and Applications*, **16**, 3 63, 302–321.

[23] Wongrin, W., and Bodhisuwan, W. (2017). Generalized Poisson–Lindley linear model for count data. *Journal of Applied Statistics*, **44**, 15, 2659–2671.

CREATION AND ANALYSIS OF MULTIMODAL EMOTION RECOGNITION CORPUS WITH INDIAN ACTORS

Komal Anadkat[1], Dr. Hiteishi Diwanji[2], Dr. Shahid Modasiya[3] , Mihir Mehta[4]

•

[1]Assistant Professor, Information Technology department, G.E.C., Gandhinagar, Gandhinagar, India. komalanadkat@gecg28.ac.in

[2] Professor, Information Technology department, L.D. College of Engineering, Ahmedabad, India. hiteishi.diwanji@gmail.com

[3]Assistant Professor, Electronics & Communication department, G.E.C., Gandhinagar, Gandhinagar, India. shahid@gecg28.ac.in

[4]Assistant Professor, Computer Engineering department, G.E.C., Gandhinagar, Gandhinagar, India. mihir_mehta@gecg28.ac.in

Abstract

Emotion recognition plays an important role in many real-life application areas of artificial intelligence like human-computer interactions, autism detection, stress and depression detection, measuring mental health, and suicide prevention. Emotion state of a person can be decided by the facial expression, tone of voice, words of speech, and body gestures when they are having a face-to-face conversation. People widely use social media platforms to post their feelings and mood through status. So, the status text can be used to identify the emotional state of a person. Physiological signals (EEG, ECG, and EDA) can identify the emotional state more accurately as people cannot be faked during the data collection but it is difficult to collect data. Many unimodal and multimodal datasets are publicly available but still, there is a strong need to create a multimodal dataset that consists of all the important modalities for the identification of emotional state. In this paper, first, we have reviewed all the available unimodal and multimodal datasets, then in the next section, we discuss the method to prepare the multimodal dataset. The data of four different modalities like facial expressions, audio, social media text, and EEG have been collected from seven different actors of different age groups and of different demographic regions. The dataset is non-spontaneous and contains discrete emotion labels like happy, sad, and angry. The procedure to create a dataset of different modalities include steps like capturing data, pre-processing, feature extraction, and storing to the relevant format. In last, to observe effect of different emotions, analysis of proposed multimodal database is carried out using efficient image, speech and text parameters.

Keywords: Unimodal, Emotion recognition, Feature extraction, Multimodal, EEG, Modality.

I. Introduction

Human emotion, sentiment, and feelings [1], emotion identification, and sentiment analysis all fall under the umbrella category of human-computer interaction. It has guided research on computers' abilities to recognize and convey emotions, respond intelligently to human feelings, and manage and exploit human emotions [2]. Nowadays it is highly desirable to build a system that can recognize and understand the emotions of a person, and respond in a way like a human [3]. For instance, if the driver's emotional state can be monitored and appropriate responses generate in a mart vehicle

system, then observed findings might effectively minimize the risk of accidents [4]. So, many authors have done significant research and concluded that emotion recognition is the key to promoting human-machine interaction, AI, and many other research fields.

Many Datasets are publicly available for emotion recognition, some of them are unimodal like visual, audio, text, and psychological. Some recent advancements have been done in the field of creating multimodal datasets like audio-textual-visual databases and video-physiological databases. People use to express their emotions through facial expression, speech intensity, and social media status, thus multimodal signals. So, many researchers have developed multimodal databases by collecting on-the-spot emotions, gathering online data, or inducing emotions through videos. Most of these datasets are either multi-physical or physiological. Nowadays people use social media platforms to express their feeling mostly. We have observed that social media status can be used to classify emotions effectively.

We have studied all the available unimodal and multimodal emotion recognition datasets but there is no dataset exist which contains all important dimensions for classification.so, In this paper, We have prepared the multimodal emotion recognition database of 7 demographically different actors which consists of their facial expression, audio, social media status, and EEG signal for three classification categories happy, sad, and angry.

II. Existing Databases for Emotion Recognition

The datasets available for the Emotion recognition task are mainly classified into facial expression, audio, Text, and physiological signals. Some available datasets are multimodal which consist of more than one dimension of emotion recognition. Multimodal datasets are required to make a robust model which identifies the human emotions of different categories.

I. Facial expression databases

Many authors have tried to recognize emotions using facial expressions but people can be faked expressing emotion. So, the dataset used to train the model should be sparse and not limited. The early facial expression databases were created using emotions that were purposefully expressed by people in the lab.

JAFFE [5] is a facial expression dataset, which has 213 images of 256*256 pixels for 7 different emotions (6 basic facial expressions and neutral). The Actors are 10 Japanese female models who acted to create the dataset.

Cohn-Kanade (CK+) [6], which is an extension of CK [7], the dataset contains 593 video sequences from 123 different subjects who were instructed to perform 7 facial expressions (anger, contempt, disgust, fear, happiness, sadness, and surprise). This dataset is lab-controlled yet very extensive to provide comparatively good results for emotion recognition.

Oulu-CASIA [8], the dataset consists of 6 expressions (surprise, happiness, sadness, anger, fear, and disgust) from 80 Subjects and includes 2,880 image sequences captured with one of two kinds of imaging systems. Subjects were asked to make a facial expression according to an expression example shown in picture sequences. The imaging hardware works at the rate of 25 frames per second and the image resolution is 320 × 240 pixels.

BP4D [9], the dataset is a well-annotated 3D video database of spontaneous facial expressions, collected from 41 participants (23 women, 18 men), demographically from a different region. Eight tasks were covered with an interview process and a series of activities to elicit eight emotions.

4DFAB 10] , the dataset consists of at least 1,800,000 dynamic high-resolution 3D faces captured from 180 subjects in four different sessions spanning over a five-year period.

So far, the acted databases are listed which were constructed in a particular environment. On the other way, we can create a dataset by collecting images or videos from the internet and it is called In-the-Wild.

FER2013 [11], is a large-scale dataset consists 35,887 gray images with 48×48 pixels, collected automatically through the Google image search API. The 07 emotion categories were collected and labeled 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. SFEW 2.0 [12], is an in-the-wild dataset that is divided into three sets, including Train (891 images) and Val (431 images), labeled as one of six basic expressions (anger, disgust, fear, happiness, sadness, and surprise), as well as the neutral and Test (372 images) without expression labels. EmotioNet [13], dataset consists of one million images with 950,000 automatically annotated AUs and 25,000 manually annotated AUs. AffectNet [14], the dataset contains over 1,000,000 facial images, of which 450,000 images are manually annotated with eight discrete expressions (six basic expressions plus neutral and contempt), and the dimensional intensity of valence and arousal.

II. Speech/Audio databases

There are main two categories of speech datasets: - induced and spontaneous. Induced datasets were created from the professional actor's performance and they have to act in a particular environment and record the dataset which is more authentic.

Berlin Database of Emotional Speech (Emo-DB) [15], the dataset contains about 500 utterances spoken by 10 actors (5 men and 5 women) in a happy, angry, anxious, fearful, bored, and disgusted way.

Belfast Induced Natural Emotion (Belfast) [16], the dataset was recorded from 40 subjects at Queen University in Northern Ireland, UK. Each subject took part in five tests, each of which contained short video recordings 9 (5 to 60 seconds in length) with stereo sound, and related to one of the five emotional tendencies: anger, sadness, happiness, fear, and neutrality.

Ryerson Audio-Visual Database of Emotional Speech and Song (Ravdess), the dataset contains 7356 files, recorded by 24 actors (12 males, 12 females). Each actor vocalized two statements in North American accent. Speech includes calm, happy, sad, angry, fearful, surprise, and disgust expressions, and the song contains calm, happy, sad, angry, and fearful emotions.

III. Textual (Social media) databases

Text data in various levels (e.g., word, phrase, and document) is tagged with emotion or sentiment tags, such as positive, negative, emphatic, general, sad, glad, and so on, in databases for textual emotion analysis.

The multi-domain sentiment (MDS) [17,18], dataset consists of more than 100,000 sentences, which are product reviews acquired from Amazon.com. These sentences are labeled with both two sentiment categories (positive and negative) and five sentiment categories (strong positive, weak positive, neutral, weak negative, and strong negative).

IMDB [19], is a popularly used largest dataset that provides 25,000 highly polar movie reviews for training and 25,000 for testing. The first line in each file contains headers that describe what is in each column.

Stanford sentiment treebank (SST) [20] is the semantic lexical database annotated by Stanford University. It includes a fine-grained emotional label of 215,154 phrases in a parse tree of 11,855 sentences, and it is the first corpus with fully labelled parse trees.

IV. EEG databases

Physiological signals cannot be altered intentionally to hide the emotions which is normally happens with facial expressions, audio and textual emotions. So it is more authentic and reliable. It consists EEG,ECG, EMG and RESP data. EEG is a signal which captures the brain signals so it is highly desirable to use in the task of emotion recognition.

DEAP [21], comprises a 32-channel EEG, a 4-channel EOG, a 4-channel EMG, RESP, plethysmograph, Galvanic Skin Response (GSR), and body temperature, from 32 subjects. Each subject participated in 40 EEG trials, in each of which a specific emotion was elicited by a music video. Immediately after watching each video, subjects were required to rate their truly-felt emotion from five dimensions: valence, arousal, dominance, liking, and familiarity.

SEED [22, 23] is based only on EGG recordings, from 15 subjects. However, it enables repeated experiences to improve data reliability/stability. In their study, participants were asked to experience three EEG recording sessions, with an interval of two weeks between two successive recording sessions. Within each session, each subject was exposed to the same sequence of fifteen movie excerpts, each one approximately four-minute-long, to induce three kinds of emotions: positive, neutral, and negative.

AMIGOS [24] was designed to collect participants' emotions in two social contexts: individual and group. AMIGOS was constructed in 2 experimental settings. First, 40 participants watched 16 short emotional videos. Then, they watched 4 long videos, including a mix of lone and group sessions. These emotions were annotated with not only self-assessment of affective levels but also external assessment of valence and arousal, through GSR and ECG signals.

Wearable devices help to bridge the gap between lab studies and real-life emotions. Wearable Stress and Affect Detection (WESAD) [25] was constructed for stress detection, providing multimodal, high-quality data, including three different affective states (neutral, stress, amusement).

IV. Multimodal databases

Humans mostly express their feelings in through multimodal signals. So, rather than focusing only on a single modal like audio, video, text, or EEG, it is desirable to create a multimodal dataset by collecting spontaneous emotions from available online data.

Interactive Emotional Dyadic Motion Capture (IEMOCAP) [26] is constructed by the Speech Analysis and Interpretation Laboratory. During recording, 10 actors are asked to not only perform selected emotional scripts but also improvised hypothetical scenarios designed to elicit 5 specific types of emotions. The face, head, and hands of actors are marked to provide detailed information about their facial expressions and hand movements while performing.

Harvesting Opinions from the Web database (HOW) [27] is the first publicly available database containing visual, audio, and textual modalities for sentiment analysis. HOW consists of 13 positive, 12 negative, and 22 neutral videos captured from YouTube.

ICT-MMMO (Multimodal Movie Opinion) [28], the dataset consists of 308 YouTube videos and 78 movie review videos from ExpoTV. The dataset has five emotion categories named strongly positive, weakly positive, neutral, strongly negative, and weakly negative.

CMU-MOSEI (Multimodal Opinion Sentiment and Emotion Intensity)[29], is the largest dataset which consists of 23,453 sentences and 3,228 videos from more than 1,000 online YouTube speakers. Each video contains a manual transcription that aligns audio and phoneme grade.

Remote Collaborative and Affective Interactions (RECOLA) [30], the dataset consists multimodal corpus of spontaneous interactions from 46 participants. These participants worked in pairs to discuss a disaster scenario escape plan and came to an agreement via remote video conferencing.

The recordings of the participants' activities were labeled by 6 annotators with two continuous emotional dimensions: arousal and valence, as well as social behavior labels on five dimensions.

III. Dataset Design and Acquisition

Emotion Recognition and prediction is indeed a challenging task, even though many researchers have done significant work in this area. There are many multimodal datasets available for emotion recognition but still there is a need to develop extended multimodal dataset, which include physical and psychological data. In the next section, we discusses the procedure followed in the acquisitions of multimodal facial expression, audio, social media status and EEG data of the subjects for the proposed dataset. The actors depict the wider range 'happy', 'sad', and 'anger' feelings. These discrete feelings of humans are universally accepted as basic human emotions.

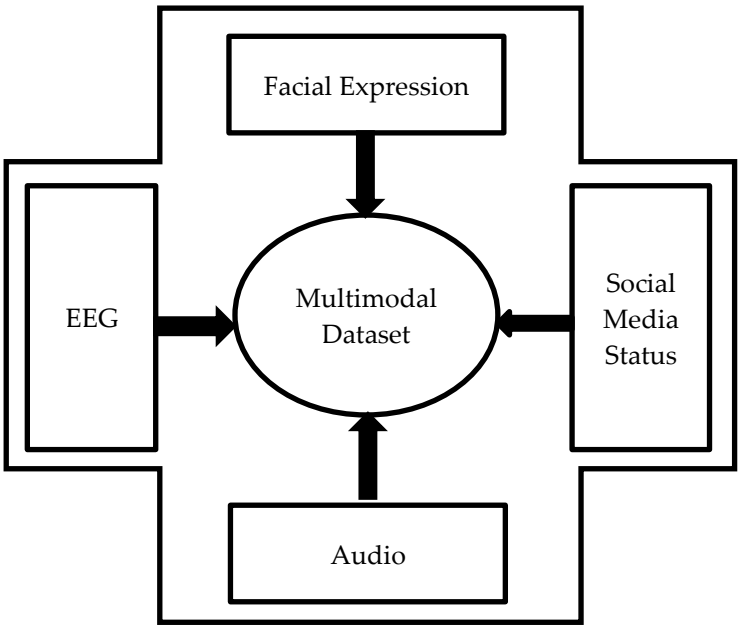


Figure 1: *Multimodal Emotion Recognition dataset dimensions*

I. Actor Details

The 07 Professors of Government Engineering College, Gandhinagar have been participated in the acting sessions to create this dataset. The facial expression, voice, social media status and EEG signals are collected for three basic emotions like happy, sad and angry. Actors' mean age = 39.5 years, age range = 33-45, males = 2, and females = 5. Actors are the basically belongs from Rajasthan (1 actor), North Gujarat (2 actors), South Gujarat (1 actor), Saurashtra (2 actors) and Central Gujarat (1 actor). All actors speak English as a foreign language.

II. Facial Expression Dataset

The facial expression dataset was collected for three different emotions happy, sad and angry by 07 different actors. The FER-13 dataset samples are demonstrated to actors for better understanding of the requested target emotions. The dataset FER-2013 contains 35,887 grayscale images of faces with 48*48 pixels and stored in CSV format. As shown in the figure 2, the images of three different facial

expression happy, sad and angry are captured using RedMe 9 prime phone for all 7 actors. Then, the collected images have been converted into 48*48 grayscale images and the pixel values of these images are stored in .csv file.

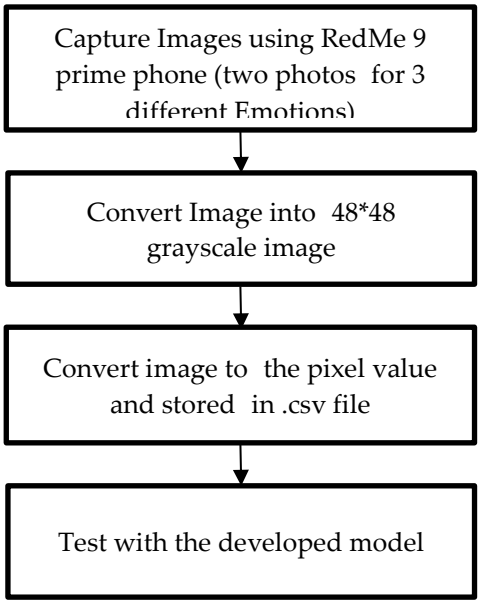


Figure 2: *Procedure in the creation of Facial Expression dataset*

Figure 3: *Actors performed in the proposed Facial expression dataset*

III. Audio Dataset

The Audio dataset was collected for three different emotions happy, sad and angry by 07 different actors. The RAVDESS dataset samples are demonstrated to actors for better understanding of the requested target emotions. The two sentences are given and explained to each actor. They are requested to practice script sentences to evoke the target feelings. The RAVDESS dataset contains 7356 files. The database contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. As shown in figure 4 ,The recording of all actors have been done by RedMe 9 prime phone. Two sentences Dogs are sitting by the door and Kids are talking by the door are recited by the actors in all three emotions. Then recorded audios are converted to .wav format and stereo files are converted to mono files.

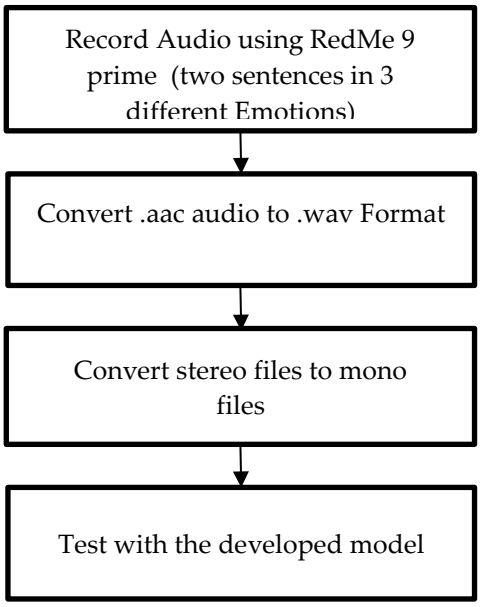


Figure 4: *Procedure in the creation of Audio dataset*

IV. Social Media Status Dataset

The Social media status dataset was collected for three different emotions happy, sad and angry by considering status of 07 different actors. Three basic emotion status sad, happy and angry has been scraped and prepared which is available on kaggle. Every data set contains two columns one the status and another sentiment of that status. So, we pre-process our collected samples accordingly. As shown in Figure 5. ,we have developed a system that could find the emotion label like happy, sad, and anger for any piece of text, especially status and stories on social media with their probability for each emotion using its textual features, Preprocessing techniques, and various machine learning and deep learning algorithms[31]. Samples of collected status from different actors are listed with sentiments in table 1.

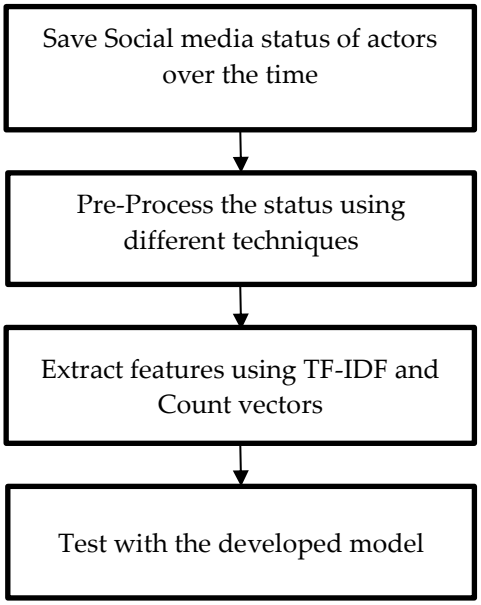


Figure 5: *Procedure in the creation of Social Media Status dataset*

Table 1: *Samples collected from different Actors*

Social Media Status	Sentiment
Happiness is where we find it, but very rarely where we seek it.	Happy
Beware, I'm not in my greatest mood today.	Angry
My silence is just another word for my pain.	Sad
Happiness is when what you think, what you say, and what you do are in harmony.	Happy

V. EEG Dataset

EEG is an electrophysiological monitoring method to record the electrical activity of the brain. We have collected EEG data from 07 people while they are made to concentrate on a particular thought. We have used NeuroMax Nmx-32(channels) series portable device for the collection of data. The subjects were asked to visualize the incident which can make them feel happy, sad, and angry respectively for 60 seconds each. The other activities like eyes blinking, movement, and eyes open are recorded in the data set. The recorded signals contain 32 electrodes data, so the important electrodes for emotion recognition were identified and then prepared data for the analysis task.

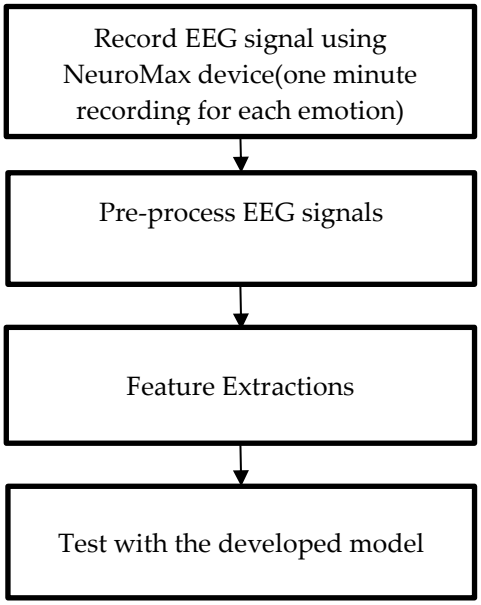


Figure 6: *Procedure in the creation of EEG dataset*

Figure 7: *Actors performed in the proposed EEG dataset*

IV. Experiment Results

The multimodal emotion recognition dataset contains image, audio, social media text and EEG signal. In this paper, different evaluation parameters are used for different modal. Complete evaluation is performed in Python.

I. Facial Expression Dataset Evaluation

In this paper, Blind/reference less image spatial quality evaluator (BRISQUE) assessment method used to evaluate image quality of facial expression dataset, which only uses the image pixels to calculate features. BRISQUE relies on spatial Natural Scene Statistics (NSS) model of locally normalized luminance coefficients in the spatial domain, as well as the model for pairwise products of these coefficients. First, the MSCN coefficients and the pairwise products are calculated and then the distribution is verified using plot which is shown in Figure 8.

Then pre-trained SVR model is used to calculate the quality assessment. However, in order to have good results, we scaled the features to [-1, 1]. The scale used to represent image quality goes from 0 to 100. An image quality of 100 means that the image's quality is very bad. In the case of the analyzed image, we get that it is a good quality image as the image quality score is 3.889.

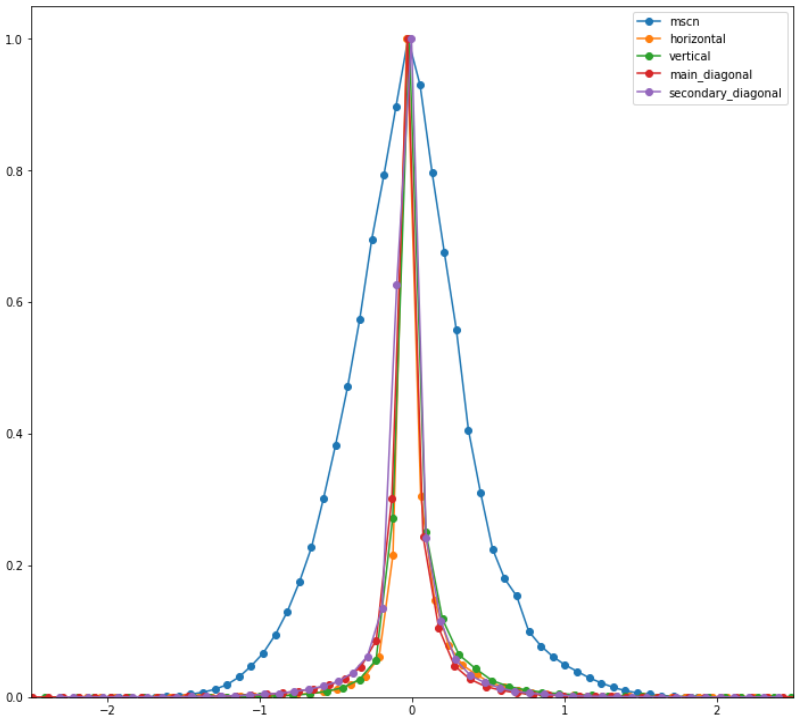


Figure 8. *Plot of image coefficient distribution*

```
%%time

calculate_image_quality_score(brisque_features)
```

```
CPU times: user 10.8 ms, sys: 0 ns, total: 10.8 ms
Wall time: 11.3 ms
3.8893297529106974
```

II. Audio Dataset Evaluation

Speech signal is a denouement of time varying vocal tract system agitated by the time varying excitation source signal. In this paper, the converted audio signal is observed by different parameters shown below. Figure 9 is the Plot of amplitude envelope of a sample waveform. A Mel spectrogram is a spectrogram where the frequencies are converted to the Mel scale is show in in Figure 10. The FFT is computed on overlapping windowed segments of the signal, and we get what is called the spectrogram which is shown in Figure 11.The fast Fourier transform (FFT) is an algorithm that can efficiently compute the Fourier transform which is shown in Figure 12. MFCC and the creation of filter banks are all motivated by the nature of audio signals and impacted by the way in which humans perceive sound .Figure 13 shown the MFCC features of test audio signal. The first horizontal yellow lines below every segment are the fundamental frequency and at their strongest. Above the yellow line are the harmonics that share the same frequency distance between them. The Window Count = 67 and Individual Feature Length = 13 are the MFCC parameters.

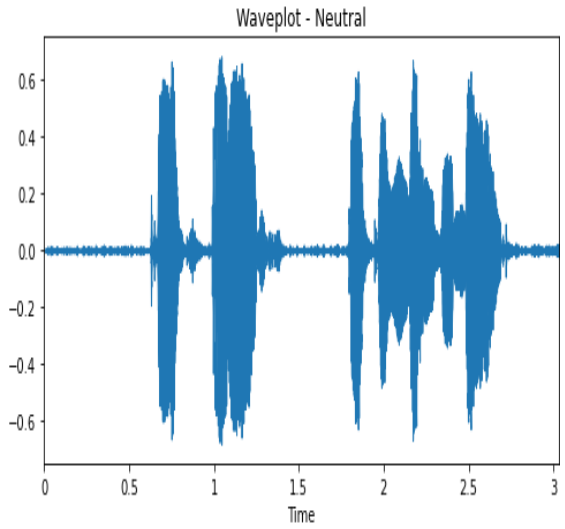


Figure 9: *Plot of Signal Wave plot*

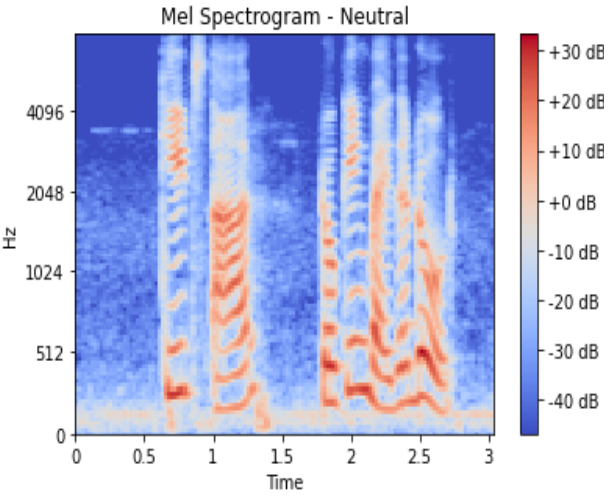


Figure 10: *Plot of Mel Spectrogram*

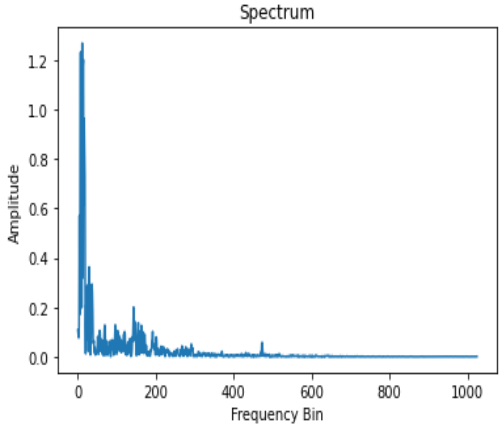


Figure 11: *Plot of Spectrogram*

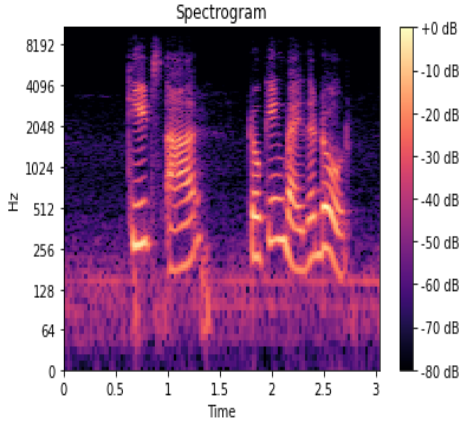


Figure 12 : *Plot of FFT spectrum*

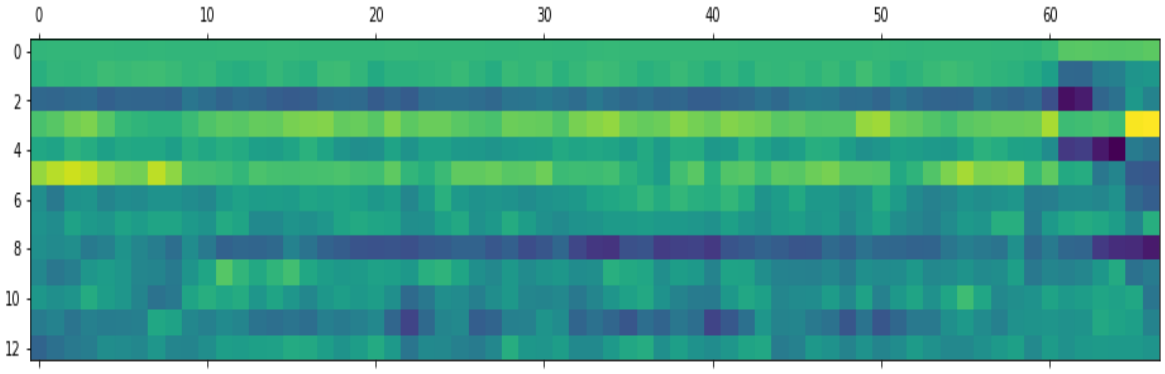


Figure 13: *Plot of MFCC Features*

III. Social Media Status Dataset Evaluation

In this paper, we explored various tools to evaluate social media status dataset to explore and visualize text data efficiently. The number of characters present in each sentence is shown in figure 14 and the histogram shows that status content range from 45 to 130 characters. The average word length of each status is ranges between 3.50 to 5.25 and that is shown in figure 15. The stop words are the words that are most commonly used in any language such as "the"," a"," an" etc. We can evidently see in Figure 16 that stop words such as "you"," the" and "to" dominate in Status contents. Wordcloud is a great way to represent text data. The size and color of each word that appears in the wordcloud indicate its frequency or importance. Figure 17 shows that the terms associated with the emotions are highlighted which indicates that these words occurred frequently in the social media status.

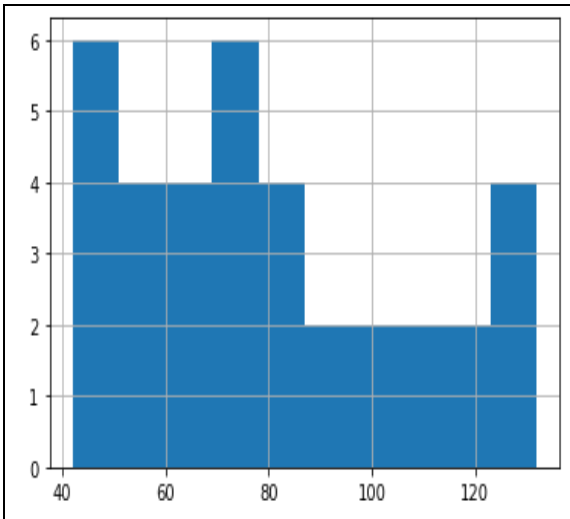


Figure 14: *Plot of number of characters in each sentence*

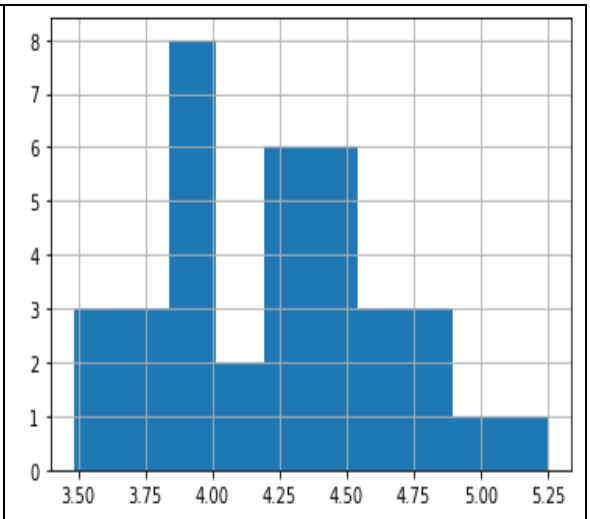


Figure 15 : *Plot of average word length of each status*

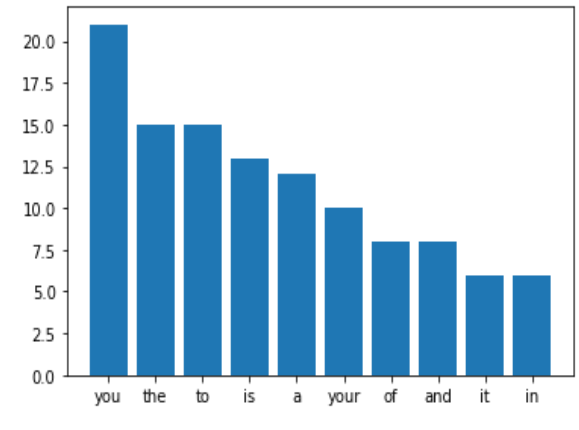


Figure 16 : *Plot of top stop words in corpus*

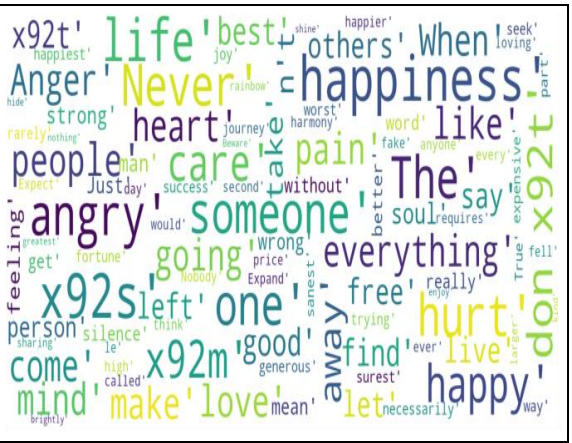


Figure 17: *Plot of Wordcloud*

V. Conclusion and Future Work

As Emotion Recognition is a growing research field in today's era, there are many existing unimodal and multimodal datasets available. The multimodal datasets normally consist of audio-visual, visual-audio-text, audio-visual-physiological data or EEG-EDA-ECG data. Still there is a pressing demand to develop a more extended and new multimodal dataset which consist important modalities like facial expression, audio, social media status and EEG signals. To fulfil this requirement, we have developed a new dataset for multimodal emotion recognition in which 7 different Professors of Government Engineering College, Gandhinagar have acted, for four basic modalities. This dataset has developed for three discrete emotion categories named happy, sad and angry. This multimodal dataset can be used to develop an emotion recognition model which is more robust and more accurate. In the future, the dataset can be tested on a pre-trained unimodal emotion recognition model to check whether the model works well on real-world dataset or not. We will try to include more modality like ECG, EDA ,RSA data to make our dataset more dimensional. The audio data is taken in English language only, so in future we will try to include audio data in other languages like Gujarati and Hindi to make our data set multilingual. In addition to the three basic emotion categories, the dataset can be upgraded for other categories like Surprise,Neutral,fear and disgust.

References

[1] K.S. Fleckenstein (1991). Defining Affect in Relation to Cognition: A Response to Susan McLeod. *J. Adv. Compos. 11 447–53.*

[2] R.W. Picard, E. Vyzas, J. Healey (2001). Toward machine emotional intelligence: analysis of affective physiological state, *IEEE Trans. Pattern Anal. Mach. Intell. 23 1175–1191. https://doi.org/10.1109/34.954607.*

[3] M. Scheutz (2012), The Affect Dilemma for Artificial Agents: Should We Develop Affective Artificial Agents?, *IEEE Trans. Affect. Comput. 3 424–433. https://doi.org/10.1109/T-AFFC.2012.29.*

[4] J.A. Healey, R.W. Picard(2005) , Detecting stress during real-world driving tasks using physiological sensors, *IEEE Trans. Intell. Transp. Syst. 6 156–166. https://doi.org/10.1109/TITS.2005.848368.*

[5] M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba (1998) , Coding facial expressions with Gabor wavelets, in: *Proc. Third IEEE Int. Conf. Autom. Face Gesture Recognit.,: pp. 200–205. https://doi.org/10.1109/AFGR.1998.670949.*

[6] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews(2010) , The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Workshop, IEEE, San Francisco, CA, USA, 2010: pp. 94–101. https://doi.org/10.1109/CVPRW.2010.5543262.*

[7] T. Kanade, J.F. Cohn, Yingli Tian(2000), Comprehensive database for facial expression analysis, in: Proc. Fourth IEEE Int. Conf. Autom. Face Gesture Recognit. *Cat No PR00580: pp. 46–53. https://doi.org/10.1109/AFGR.2000.840611.*

[8] G. Zhao, X. Huang, M. Taini, S.Z. Li, M. Pietikäinen(2011), Facial expression recognition from near-infrared videos, Image Vis. *Comput. 29 607–619. https://doi.org/10.1016/j.imavis.2011.07.002.*

[9] X. Zhang, L. Yin, J.F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, J.M. Girard(2014), BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database, Image Vis. *Comput. 32 (2014) 692–706. https://doi.org/10.1016/j.imavis.2014.06.002*.

[10] S. Cheng, I. Kotsia, M. Pantic, S. Zafeiriou(2018), 4DFAB: A Large Scale 4D Database for Facial Expression Analysis and Biometric Applications, *IEEECVF Conf. Comput. Vis. Pattern Recognit., IEEE, Salt Lake City, UT, USA, 2018: pp. 5117–5126. https://doi.org/10.1109/CVPR.2018.00537.*

[11] I.J. Goodfellow, D. Erhan, P.L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov (2013), Challenges in Representation Learning: A report on three machine learning contests, *ArXiv13070414 Cs Stat. http://arxiv.org/abs/1307.0414.*

[12] A. Dhall, R. Goecke, S. Lucey, T. Gedeon (2011), Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark, *IEEE Int. Conf. Comput. Vis. Workshop ICCV Workshop, 2011: pp. 2106–2112. https://doi.org/10.1109/ICCVW.2011.6130508*.

[13] C.F. Benitez-Quiroz, R. Srinivasan, A.M. Martinez(2016) , EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild, *IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, IEEE, Las Vegas, NV, USA, pp. 5562–5570. https://doi.org/10.1109/CVPR.2016.600.*

[14] A. Mollahosseini, B. Hasani, M.H. Mahoor (2019), AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild, *IEEE Trans. Affect. Comput. 10 18–31. https://doi.org/10.1109/TAFFC.2017.2740923.*

[15] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss(2005), A Database of German Emotional Speech, *4*.

[16] I. Sneddon, M. McRorie, G. McKeown, J. Hanratty(2012), The Belfast Induced Natural Emotion Database, *IEEE Trans. Affect. Comput. 3 32–41. https://doi.org/10.1109/T-AFFC.2011.26.*

[17] J. Blitzer, M. Dredze, F. Pereira(2007), Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification, *ACL 2007. (n.d.) 8*.

[18] M. Dredze, K. Crammer, F. Pereira(2008), Confidence-weighted linear classification, *Proc. 25th Int. Conf. Mach. Learn. - ICML 08, ACM Press, Helsinki, Finland, pp. 264–271. https://doi.org/10.1145/1390156.1390190.*

[19] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, C. Potts(2011), Learning Word Vectors for Sentiment Analysis, *Proc. 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., Association for Computational Linguistics, Portland, Oregon, USA,: pp. 142–150. https://www.aclweb.org/anthology/P11-1015.*

[20] R. Socher, A. Perelygin, J. Wu, J. Chuang, C.D. Manning, A. Ng, C. Potts, Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, *(n.d.) 12.*

[21] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras(2012), DEAP: A Database for Emotion Analysis ;Using Physiological Signals, *IEEE Trans. Affect. Comput. 3 18–31. https://doi.org/10.1109/T-AFFC.2011.15.*

[22] R.-N. Duan, J.-Y. Zhu, B.-L. Lu(2013), Differential entropy feature for EEG-based emotion classification, *6th Int. IEEEEMBS Conf. Neural Eng. NER, 2013: pp. 81–84. https://doi.org/10.1109/NER.2013.6695876*.

[23] W.-L. Zheng, B.-L. Lu (2015) , Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks, *IEEE Trans. Auton. Ment. Dev. 7 162–175. https://doi.org/10.1109/TAMD.2015.2431497.*

[24] J.A. Miranda Correa, M.K. Abadi, N. Sebe, I. Patras (2018), AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups, *IEEE Trans. Affect. Comput. 1–1. https://doi.org/10.1109/TAFFC.2018.2884461.*

[25] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K. Van Laerhoven (2018), Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection, *Proc. 20th ACM Int. Conf. Multimodal Interact., Association for Computing Machinery, Boulder, CO, USA,: pp. 400–408. https://doi.org/10.1145/3242969.3242985.*

[26] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, S.S. Narayanan(2008), IEMOCAP: interactive emotional dyadic motion capture database, Lang. Resour. Eval. *42 335. https://doi.org/10.1007/s10579-008-9076-6*.

[27] L.-P. Morency, R. Mihalcea, P. Doshi (2011), Towards multimodal sentiment analysis: harvesting opinions from the web, *Proc. 13th Int. Conf. Multimodal Interfaces, Association for Computing Machinery, Alicante, Spain, 2011: pp. 169–176. https://doi.org/10.1145/2070481.2070509.*

[28] M. Wollmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, L.-P. Morency(2013), YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context, *IEEE Intell. Syst. 28 46–53. https://doi.org/10.1109/MIS.2013.34*

[29] A. Bagher Zadeh, P.P. Liang, S. Poria, E. Cambria, L.-P. Morency(2018), Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph, *Proc. 56th Annu. Meet. Assoc. Comput. Linguist. Vol. 1 Long Pap., Association for Computational Linguistics, Melbourne, Australia, : pp. 2236–2246. https://doi.org/10.18653/v1/P18-1208*.

[30] F. Ringeval, A. Sonderegger, J. Sauer, D. Lalanne(2013), Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions, *10th IEEE Int. Conf. Workshop Autom. Face Gesture Recognit. FG,: pp. 1–8. https://doi.org/10.1109/FG.2013.6553805.*

[31] Komal Anadkat, Hiteishi Diwanji, Shahid Modasiya (2022). Effect of Preprocessing in Human Emotion Analysis Using Social Media Status Dataset. *RT & A ,No 1(67),volume-17*.

THE EFFICACY OF TRAPEZOIDAL FUZZY NUMBERS AND ITS APPLICATION

Ajjaz Maqbool Dar[1], Aafaq A. Rather[2, *], Rushika Kinjawadekar[3], Abhay Deshpande[4], Maryam Mohiuddin[5], Rashid A. Ganaie[6], Khursheed Ahmad[7]

•

[1]Department of Mathematics, Govt College for Women's Nawakadal Srinagar j&k-190002, India
[2, 4]Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India
[3]Department of Mathematics & Statistics, Faculty of Science & Technology, Vishwakarma University, Pune, India
[5,6]Department of Statistics, Annamalai University, Annamalai nagar, Tamil Nadu, India
[7]Department of Public Administration, Government Degree College Tangmarg, Kashmir, India
[1]ajjazmaqbool013@gmail.com, [2, *]aafaq7741@gmail.com, [3]rushika.kinjawadekar@vupune.ac.in, [4]abhaypdeshpande18@gmail.com, [5]masmariam7@gmail.com, [6]rashidau7745@gmail.com, [7]dkbutt99@gmail.com

Abstract

Numerous fields, including engineering, agriculture, and management sciences, have been using trapezoidal fuzzy numbers. In this study, we first develop Trapezoidal Fuzzy Number (TFN) and then attempt to formulate a model to handle element uncertainty in order to solve a linear programming problem. Making good decisions will only require this type of approximation.

Keywords: Trapezoidal Fuzzy Number, linear programming problem, initial problem and membership function.

1. Introduction

Optimization problem is a one of the most important operation research techniques, and it is used in many areas in agriculture planning, science, Technology and engineering which may be important in both economic and social point of view. We formulate the problem mathematically which may arise in our daily life our aim is to minimize cost and maximize the profit, with certain constrains or restrictions are to be considered. In order to get the best possible result of those problems faced by Agriculturalists to allocate the optimum number of vegetation in their farmhouses. To increase the area under cultivation there are numerous ways to achieving high productivity. If we utilize the resources in a proper way which may be helpful to increase the crop production. Many operation research techniques have been used in planning agriculture activity one of the technique is linear programming. In 1947 George Danzig was given the concept of Linear Programming problem. If we have a limited number of resources, we use linear programming method to optimize the problem. The Zimmermann [1] presented the concept of formulation of fuzzy linear programming problem. Orlovsky [2] made several attempts to investigate the potential of fuzzy set theory as a valuable tool for comprehensive mathematical analysis of practical problems. To address the many kinds of FLP problems, numerous authors utilize various techniques. In almost all areas of decision-making problems, fuzzy approaches

have been developed. Particularly Tamiz [3], and Ross [4]. Delgado and Verdegay [5] construct a broad model of fuzzy linear programming within the fuzzy and fuzzy right side of technical coefficients and also demonstrates that it is possible to solve the dual problem using the same programme. Fung and Hu [6] introduced the fuzzy number-based approach coefficients. Kumar and Rajendra [7] solved a fuzzy linear programming problem with fuzzy variables in parametric form. By utilizing a ranking function and defining a crisp model, Verdegray [8] and Maleki [9] ranking function can be identified in comparisons between fuzzy numbers. In order to determine a workable and ideal solution, we study the linear programming issue in this essay in its conventional form. To solve the linear fuzzy linear programming problem, we utilize the algorithm by trapezoidal fuzzy number is considered.

2. Model Formulation

$$\begin{aligned} &Maxmize\ Z = CY \\ &AY \le B \\ &Y \ge 0 \end{aligned} \tag{1}$$

Where C is vector component, A is coefficient Matrix, B is crisp parameters and Y is decision variable.

2.1 Generalized Trapezoidal Fuzzy Number (TFN)

The generalized Fuzzy Trapezoidal number $\widetilde{T} = (t_1, t_2, t_3, t_4, w)$ is a fuzzy subset of real line R, whose membership function $\mu_{\widetilde{T}}$ satisfies the following postulates:

- $t_1 \le x \le t_2$, is a continuous mapping from R to the closed interval [0, 1]
- $\mu_{\widetilde{T}}(x) = 0,\ -\infty < x \le t_1$
- $\mu_{\widetilde{T}}(x)$ is strictly increasing with constant rate on $t_1 \le x \le t_2$
- $\mu_{\widetilde{T}}(x)$ is strictly decreasing with constant rate on $t_3 \le x \le t_4$
- $\mu_{\widetilde{T}}(x) = 0,\ t_4 \le x < \infty$

Membership function is given by

$$\mu_{\widetilde{T}}(y) = \begin{cases} w(\dfrac{Y - t_1}{t_2 - t_1}), & t_1 \le Y \le t_2 \\ w, & t_2 \le Y \le t_3 \\ w(\dfrac{t_4 - Y}{t_4 - t_3}), & t_3 \le Y \le t_4 \\ 0, & elesewhere \end{cases} \tag{2}$$

Where, $t_1 < t_2 < t_3 < t_4\ and\ w \in (0,1]$

If w=1, the generalized TFN can be written as

$\widetilde{T} = (t_1, t_2, t_3, t_4)$ and the membership function is given by

$$\mu_{\tilde{T}}(y)=\begin{cases} w(\frac{Y-t_1}{t_2-t_1}), & t_1 \le Y \le t_2 \\ 1, & t_2 \le Y \le t_3 \\ w(\frac{t_4-Y}{t_4-t_3}), & t_3 \le Y \le t_4 \\ 0, & elesewhere \end{cases} \tag{3}$$

Now we can take ordered pair of parametric of fuzzy numbers with left-hand alpha- cut and right-hand alpha-cut, which are bounded left non-decreasing and bounded right non-increasing functions over [0,1],

i.e. $\widetilde{T}=\{(t_1+\alpha(t_2-t_1)),t_4+\alpha(t_4-t_3)\}$

The above mathematical model can be formulated as given below

$$\begin{aligned} &Maximize && Z=C_1(y_{L1},y_{R1})+C_2(y_{L2},y_{R2}).....+C_n(y_{Ln},y_{Rn}) \\ &Subject\ to && \\ & && a_{h1}(y_{L1},y_{R1})+a_{h2}(y_{L2},y_{R2})+.....+a_{hn}(y_{Ln},y_{Rn})\le(b_{Ln},b_{Rn}) \\ & && y_{Lj},y_{Rj}\ge 0, \quad for\,all\,h=1,2,3,.....m \ \ and \ \ j=1,2,3,....n. \end{aligned} \tag{4}$$

3. Applications

A 20 hectares of land is under cultivation of three different crops such as wheat, corn and pulses with certain requirement for capital (in euros) and working hours as shown below:

Table 1: *20 hectares of land under cultivation*

Crops per acre	Capital (€)	Workers (hours)
Wheat	50	10
Corn	33	8
Pulses	27	4

In this problem the profit of the above three different crops are wheat €38/ acre, Corn €32/ acre and pulses €28/ acre acres of land. The amount and working hours are respectively €460 and around 52 hours. Now, we decide how many hectares of land are required for each crop in order to maximize the profit. Let y_1 be the cultivated area with wheat, y_2 be the cultivated area with corn and y_3 be the cultivated area with pulses. We can have characterized the rough data by a trapezoidal fuzzy number as: 22 hectares = (18, 22, 24, 25) about €460 = (380, 410, 440, 450); around 55 hours = (46, 48, 50, 55). The problem can be written as:

$$\begin{aligned} &\text{Maximize} && Z=38y_1+32y_2+28y_3 \\ &\text{Subject to} && \\ & && 50y_1+33y_2+27y_3\le(380,410,440,450) \\ & && 10y_1+\ 8y_2+\ 4y_3\ \le(46,48,50,55) \\ & && y_1+\ y_2+\ y_3\ \le(18,22,24,25) \end{aligned} \tag{5}$$

The crisp model of the above problem

$$\begin{aligned} &\text{Maximize} && Z_1=38y_{L1}+32y_{L2}+28y_{L3}\ \ and \\ & && Z_2=38y_{R1}+32y_{R2}+28y_{R3} \\ &\text{Subject to} && \end{aligned}$$

$$
\begin{aligned}
50y_{L1} + 33y_{L2} + 27y_{L3} &\le 380 + 30\alpha \\
50y_{R1} + 33y_{R2} + 27y_{R3} &\le 450 + 10\alpha \\
10y_{L1} + 8y_{L2} + 4y_{L3} &\le 46 + 2\alpha \\
10y_{R1} + 8y_{R2} + 4y_{R3} &\le 55 + 5\alpha \\
y_{L1} + y_{L2} + y_{L3} &\le 18 + 4\alpha \\
y_{R1} + y_{R2} + y_{R3} &\le 25 + \alpha
\end{aligned} \qquad (6)
$$

LINGO 12.0 [10] can be used to acquire the results for the various values of presented in table (2) below. The ideal response to the initial problem is y_{L1} = (0, 0,0,0), y_{L2} = (0, 0, 0, 0), and y_{L3} = (11.62, 11.75, 11.87, 12.0) and is the ideal value Z_1 = (325.50,329,323.50,336).

Table 2: *Cropping combination provides best overall solution*

α	0.25	0.50	0.75	1.0
y_{L1}	0	0	0	0
y_{L2}	0	0	0	0
y_{L3}	11.62	11.75	11.87	12
y_{R1}	0	0	0	0
y_{R2}	0	0	0	0
y_{R3}	14.06	14.37	14.68	15
Z_1	325.50	329	323.50	336
Z_2	393.75	402.50	411.25	420

4. Conclusion

The application of fuzzy linear programming to resolve a production planning problem in agriculture has been covered in this study. The paper finishes by explaining how FLPP is transformed into clear multi-objective linear programming problems and how the farmer achieves the best possible outcomes while using constrained resources. Only the trapezoidal membership function is taken into account in this paper.

References

[1] Zimmermann, H.J., (1991). Fuzzy Set Theory and Its Applications, (2nd rev. ed). *Boston: Kulwer.*

[2] Orlovsky, S. A. (1980). Fuzzy Sets and Systems, 3, 311-321.

[3] Tamiz, M., (1996). Multi-objective programming and goal programming theories and Applications, *Germany: Springer-Verlag.*

[4] Ross, T.J., (1995). Fuzzy logic with engineering Applications, *New York: McGraw*-Hill.

[5] Deldago, M., Verdegay, J.L., Vila, M.A. (1989). A General Model for Fuzzy Linear Programming, *Fuzzy Set and System,* 29: 21-29.

[6] Fang, S.C., Hu, C.F., Wu, S.Y. & Wang, H.F. (1999). Linear Programming with Fuzzy Coefficients in Constraint, *Computers and Mathematics with Applications*, 37: 63-76.

[7] P. Senthilkumar and G. Rajendran, (2010). On the solution of Fuzzy linear programming Problem, *International journal of computational Cognition*, 8(3) 45-47.

[8] Campos, L. and Verdegay, J. L., (1989). Fuzzy Sets and Systems, 32, 1-11.

[9] Maleki, H.R., Tata, M., Mashinchi, M., (2000). Fuzzy Sets and Systems, 9, 21-33.

[10] Lingo 12.0, LINDO *Inc. Ltd*.

RT&A, No 3 (74)
Volume 18, September 2023

A Novel Three - Parameter Version of the Ailamujia Distribution

Idzhar A. Lakibul

•

Department of Mathematics and Statistics
Mindanao State University - Iligan Institute of Technology
Iligan City, Philippines
idzhar.lakibul@g.msuiit.edu.ph

Abstract

In this paper, a novel three - parameter continuous distribution is introduced. This novel distribution is an extented version of the Exponenentiated Ailamujia distribution. This extended version called as Exponentiated Generalized Ailamujia (EGA) distribution. The Exponentiated Generalized class is used to derive the proposed distribution by considering Ailamujia distribution as a baseline distribution. A special case of the EGA distribution called Generalized Ailamujia (GA) distribution is also derived. Properties of the proposed distribution such as moments, mean, variance, harmonic mean, moment generating function, survival function, hazard function, reverse hazard rate, Mills ratio and order statistics are derived. In addition, maximum likelihood approach is used to estimate the proposed distribution parameters. Finally, the proposed distribution is applied to two real datasets and compare with the Exponentiated Ailamujia and the Ailamujia Inverted Weibull distributions. Results reveal that the proposed distribution provides better estimate as compared to the said distributions for the given two real datasets.

Keywords: Ailamujia distribution, Exponentiated Generalized - G Class, Exponentiated Ailamujia distribution.

1. Introduction

Distributions with support on non-negative real numbers are important in modelling lifetime data. There are lifetime distributions which are popular in modelling lifetime data such as Weibull, log-logistic and lognormal distributions. These distributions are widely used in engineering and other related fields. Lv [8] proposed the Ailamujia distribution as an additional lifetime distribution and studied its properties such as mean, variance, median and maximum likelihood estimators. This distribution was studied further its properties such as interval estimation and hypothesis testing [9] and minimax estimation under non-informative prior using the loss functions [7].

There are different extensions of the Ailamujia distribution were considered in the literature. For example, Ahmad [1] introduced the Transmuted Ailamujia distribution and studied its several properties. Other identified extensions are the weighted analogue of Ailamujia distribution [14], the area biased distribution [6], the inverse analogue of Ailamujia distribution [2], the size biased Ailamujia distribution [11], the Power Ailamujia distribution [13] and the Power Ailamujia distribution [5].

Moreover, Rather [10] introduced the extended version of the Ailamujia distribution called Exponentiated Ailamujia distribution and explored some of its structural properties such as moments, reliability analysis and harmonic mean. They used the Exponentiated - G family of distributions in the derivation of the Exponentiated Ailamujia distribution. In addition, they fitted the Exponentiated Ailamujia distribution into two real datasets and they found that the Exponentiated Ailamujia distribution had a better fit compared to the Ailamujia and Lindley

distributions.

In this paper, the main goals are the following: (i) to extend the exponentiated Ailamujia distribution using the exponentiated generalized class; (ii) to derive some properties of the proposed distribution such as moments, mean, variance, harmonic mean, moment generating function, survival function, hazard function, reverse hazard rate, Mills ratio, maximum likelihood estimates and order statistics; and (iii) to apply the proposed distribution into two real datasets and compare with the Exponentiated Ailamujia and Ailamujia Inverted Weibull distributions.

The rest of paper is organized as follows: Exponentiated Generalized Ailamujia distribution is introduced in section 2. In section 3, some statistical properties of the proposed distribution are derived. Order Statistics of the proposed distribution is presented in section 4 while the Maximum likelihood estimates of the proposed distribution parameters are discussed in section 5. In section 6, the application of the proposed distribution is illustrated. Some concluding remarks is presented in section 7.

2. Exponentiated Generalized Ailamujia distribution

This section presents the derivation of the Exponentiated Generalized Ailamujia (EGA) distribution. Let X be a random variable follows an Ailamujia distribution then the cumulative distribution function of the Ailamujia distribution is given by

$$G(x,\theta) = 1 - (1+2\theta x)e^{-2\theta x}, x \geq 0, \theta > 0 \tag{1}$$

with corresponding probability density function given by

$$g(x,\theta) = 4\theta^2 x e^{-2\theta x}.$$

Cordeiro [4] introduced the exponentiated generalized class to extend any univariate continuous distribution into generalized distribution with additional two parameters. The cumulative distribution function of the exponentiated generalized class is given by

$$F(x) = (1-(1-G(x))^a)^b, a, b > 0, \tag{2}$$

where $G(x)$ is any baseline cumulative distribution function. The cumulative distribution function of the Exponentiated Generaized Ailamujia (EGA) distribution is obtained by inserting (1) into (2) and is

$$F(x,\theta,a,b) = (1-(1+2\theta x)^a e^{-2a\theta x})^b, x \geq 0, \theta, a, b > 0 \tag{3}$$

with corresponding probability density function given by

$$f(x,\theta,a,b) = 4ab\theta^2 x(1+2\theta x)^{a-1}e^{-2a\theta x}(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1}. \tag{4}$$

Notethat if $a = 1$ then the EGA distribution reduces to Exponentiated Ailamujia distribution. Then it is Ailamujia distribution if $a = b = 1$. If $b = 1$ then the cumulative distribution function of the EGA distribution reduces to a cumulative distribution function of new special distribution that is given by

$$F(x,\theta,a) = 1 - (1+2\theta x)^a e^{-2a\theta x} \tag{5}$$

with probability density function given by

$$f(x,\theta,a) = 4a\theta^2 x(1+2\theta x)^{a-1}e^{-2a\theta x}.$$

We name the cumulative distribution function (5) as the cumulative distribution function of the Generalized Ailamujia (GA) distribution.

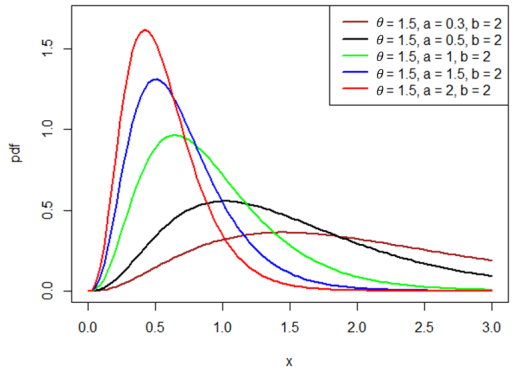


Figure 1: *pdf plots of EGA distribution for* $\theta = 1.5$, $b = 2$ *and varying values of* a

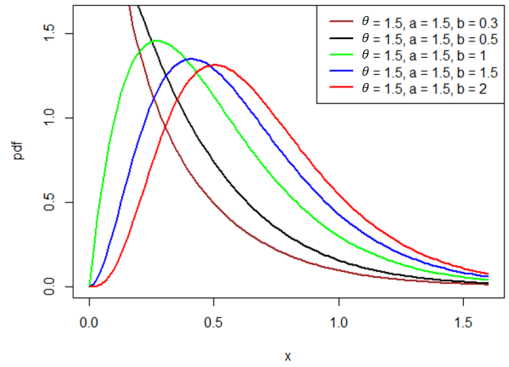


Figure 2: *pdf plots of EGA distribution for* $\theta = 1.5$, $a = 1.5$ *and varying values of* b

Figures 1 and 2 show some possible probability density shapes of the EGA distribution. It reveals that the probability density function of the EGA distribution can model a data with right tailed unimodal and exponential shapes.

3. Statistical Properties

In this section, we derive some properties of the EGA distribution such as moments, mean, variance, moment generating function, harmonic mean, survival fuction, hazard function, reverse hazard rate and Mills ratio.

3.1. Moments

Theorem 1. The *rth* raw moment of EGA with density (4) is

$$\mu'_r = \frac{b}{(2\theta)^r} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-1)^j \binom{a-1}{i} \binom{b-1}{j} \binom{aj}{l} \Gamma(r+l+i+2)}{a^{r+l+i+1}(j+1)^{r+l+i+2}}. \tag{6}$$

The mean μ'_1 and variance σ^2 are respectively, given by

$$\mu'_1 = \frac{b}{2\theta} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-1)^j \binom{a-1}{i} \binom{b-1}{j} \binom{aj}{l} \Gamma(l+i+3)}{a^{l+i+2}(j+1)^{l+i+3}}$$

and

$$\sigma^2 = \frac{b}{4\theta^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+4)}{a^{l+i+3}(j+1)^{l+i+4}} - \frac{b^2}{4\theta^2}\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+3)}{a^{l+i+2}(j+1)^{l+i+3}}\right)^2.$$

Proof. The *rth* raw moment is defined by

$$\begin{aligned}\mu_r' &= E[X^r] \\ &= \int_{-\infty}^{\infty} x^r f(x)dx \\ &= \int_0^{\infty} x^r 4ab\theta^2 x(1+2\theta x)^{a-1}e^{-2a\theta x}(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1}dx.\end{aligned}$$

Using the binomial expansion for $(1+2\theta x)^{a-1}$, we have

$$(1+2\theta x)^{a-1} = \sum_{i=0}^{\infty}\binom{a-1}{i}2^i\theta^i x^i.$$

Again, using the binomial expansion for $(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1}$, we have

$$\begin{aligned}(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1} &= \sum_{j=0}^{\infty}(-1)^j\binom{b-1}{j}(1+2\theta x)^{aj}e^{-2aj\theta x} \\ &= \sum_{j=0}^{\infty}(-1)^j\binom{b-1}{j}e^{-2aj\theta x}\sum_{l=0}^{\infty}\binom{aj}{l}2^l\theta^l x^l \\ &= \sum_{j=0}^{\infty}\sum_{l=0}^{\infty}(-1)^j\binom{b-1}{j}\binom{aj}{l}2^l\theta^l x^l e^{-2aj\theta x}.\end{aligned}$$

Now, μ_r' becomes

$$\mu_r' = 4ab\theta^2\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}(-1)^j B_i B_j B_l 2^{i+l}\theta^{i+l}\int_0^{\infty}x^{r+i+l+1}e^{-2a(j+1)\theta x}dx,$$

where $B_i = \binom{a-1}{i}$, $B_j = \binom{b-1}{j}$ and $B_l = \binom{aj}{l}$. Moreover,

$$\mu_r' = \frac{b}{(2\theta)^r}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j B_i B_j B_l\Gamma(r+l+i+2)}{a^{r+l+i+1}(j+1)^{r+l+i+2}}.$$

The mean μ_1' of EGA is obtained by setting $r=1$ in (6) and is

$$\mu_1' = \frac{b}{2\theta}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+3)}{a^{l+i+2}(j+1)^{l+i+3}}.$$

The μ_2' is derived from (6) by setting $r=2$ and is

$$\mu_2' = \frac{b}{4\theta^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+4)}{a^{l+i+3}(j+1)^{l+i+4}}.$$

The variance σ^2 of EGA is obtained as

$$\begin{aligned}\sigma^2 =&\mu_2' - (\mu_1')^2\\ =&\frac{b}{4\theta^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+4)}{a^{l+i+3}(j+1)^{l+i+4}}\\ &-\left(\frac{b}{2\theta}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+3)}{a^{l+i+2}(j+1)^{l+i+3}}\right)^2\\ =&\frac{b}{4\theta^2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+4)}{a^{l+i+3}(j+1)^{l+i+4}}\\ &-\frac{b^2}{4\theta^2}\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(l+i+3)}{a^{l+i+2}(j+1)^{l+i+3}}\right)^2.\end{aligned}$$

■

3.2. Moment Generating Function

Theorem 2. Let X be a random variable follows EGA distribution then the moment generating function $M_X(t)$ is given by

$$M_x(t) = b\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{t^r(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(r+l+i+2)}{(2\theta)^r a^{r+l+i+1}(j+1)^{r+l+i+2}r!}, \tag{7}$$

where $t \in \mathbb{R}$.

Proof. By definition of moment generating function and equation (6), we have

$$M_X(t) = \mathbb{E}(e^{tX}) = \int_0^{\infty} e^{tx} f(x,\theta,a,b)dx.$$

Recall that $e^{tx} = \sum_{r=0}^{\infty}\frac{t^r}{r!}x^r$ then we have

$$M_X(t) = \int_0^{\infty}\sum_{r=0}^{\infty}\frac{t^r}{r!}x^r f(x,\theta,a,b)dx = \sum_{r=0}^{\infty}\frac{t^r}{r!}x^r\int_0^{\infty} f(x,\theta,a,b)dx = \sum_{r=0}^{\infty}\frac{t^r}{r!}\mu_r'.$$

Thus,

$$M_x(t) = b\sum_{r=0}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{t^r(-1)^j\binom{a-1}{i}\binom{b-1}{j}\binom{aj}{l}\Gamma(r+l+i+2)}{(2\theta)^r a^{r+l+i+1}(j+1)^{r+l+i+2}r!},$$

where $t \in \mathbb{R}$. ■

3.3. Harmonic Mean

Theorem 3. Let X be a random variable follows EGA distribution then the harmonic mean of EGA is

$$H.M = 2\theta b\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-1)^j B_i B_j B_l\Gamma(l+i+1)}{a^{l+i}(j+1)^{l+i+1}}.$$

Proof. The harmonic mean is defined by

$$\begin{aligned} H.M &= \mathbb{E}\left[\frac{1}{X}\right] \\ &= \int_{-\infty}^{\infty} \frac{1}{x} f(x)dx \\ &= \int_{0}^{\infty} \frac{1}{x} 4ab\theta^2 x(1+2\theta x)^{a-1} e^{-2a\theta x}(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1} dx \\ &= \int_{0}^{\infty} 4ab\theta^2 (1+2\theta x)^{a-1} e^{-2a\theta x}(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1} dx. \end{aligned}$$

Using the binomial expansion for $(1+2\theta x)^{a-1}$, we have

$$(1+2\theta x)^{a-1} = \sum_{i=0}^{\infty} \binom{a-1}{i} 2^i \theta^i x^i.$$

Again, using the binomial expansion for $(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1}$, we have

$$\begin{aligned} (1-(1+2\theta x)^a e^{-2a\theta x})^{b-1} &= \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} (1+2\theta x)^{aj} e^{-2aj\theta x} \\ &= \sum_{j=0}^{\infty} (-1)^j \binom{b-1}{j} e^{-2aj\theta x} \sum_{l=0}^{\infty} \binom{aj}{l} 2^l \theta^l x^l \\ &= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} (-1)^j \binom{b-1}{j} \binom{aj}{l} 2^l \theta^l x^l e^{-2aj\theta x}. \end{aligned}$$

Now, $H.M$ becomes

$$H.M = 4ab\theta^2 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} (-1)^j B_i B_j B_l 2^{i+l} \theta^{i+l} \int_0^{\infty} x^{i+l} e^{-2a(j+1)\theta x} dx$$

where $B_i = \binom{a-1}{i}$, $B_j = \binom{b-1}{j}$ and $B_l = \binom{aj}{l}$. Thus,

$$H.M = 2\theta b \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \frac{(-1)^j B_i B_j B_l \Gamma(l+i+1)}{a^{l+i}(j+1)^{l+i+1}}.$$

■

3.4. Reliability Analysis

Let X be a random variable with cdf (3) and pdf (4) then the survival $S(x,\theta,a,b)$ and hazard $h(x,\theta,a,b)$ functions of the EGA distribution are respectively, given by

$$S(x,\theta,a,b) = 1-(1-(1+2\theta x)^a e^{-2a\theta x})^b,\ x \geq 0, \theta, a, b > 0$$

and

$$h(x,\theta,a,b) = \frac{4ab\theta^2 x(1+2\theta x)^{a-1} e^{-2a\theta x}(1-(1+2\theta x)^a e^{-2a\theta x})^{b-1}}{1-(1-(1+2\theta x)^a e^{-2a\theta x})^b}.$$

In addition, the reverse hazard rate $h_r(x,\theta,a,b)$ and the Mills ratio of the EGA distribution are respectively, given by

$$h_r(x,\theta,a,b) = \frac{4ab\theta^2 x(1+2\theta x)^{a-1} e^{-2a\theta x}}{1-(1+2\theta x)^a e^{-2a\theta x}}$$

and

$$MillsRatio = \frac{1-(1+2\theta x)^a e^{-2a\theta x}}{4ab\theta^2 x(1+2\theta x)^{a-1}e^{-2a\theta x}}.$$

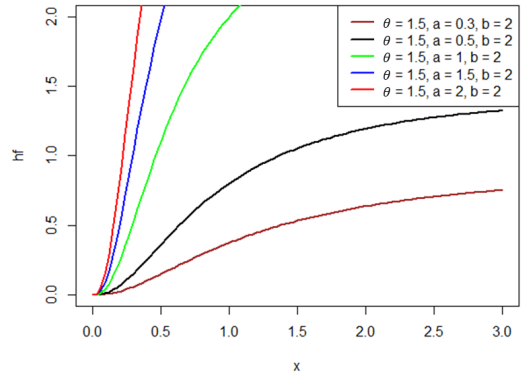


Figure 3: *hf plots of EGA distribution for* $\theta = 1.5$, $b = 2$ *and varying values of* a

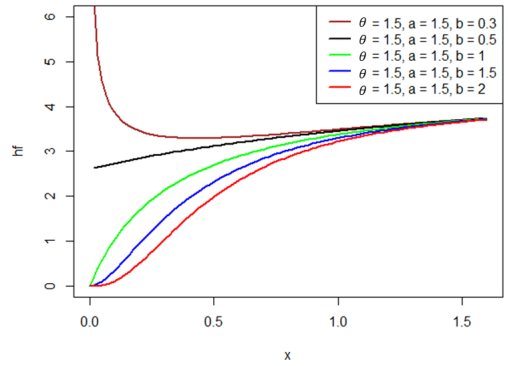


Figure 4: *hf plots of EGA distribution for* $\theta = 1.5$, $a = 1.5$ *and varying values of* b

Figures 3 and 4 present some possible shapes of the hazard function of EGA distribution. It reveals that the hazard function of the EGA distribution can model data with increasing or decreasing hazard rate behaviors.

4. Order Statistics

Let $X_{(1)}$, $X_{(2)}$,..., $X_{(n)}$ be the order statistics of a random sample X_1, X_2,..., X_n drawn from the continuous population with cumulative distribution function $F_X(x)$ and probability density function (pdf) $f_X(x)$, then the pdf of rth order statistics $X_{(r)}$ is given by

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x)[F_X(x)]^{r-1}[1-F_X(x)]^{n-r}. \tag{8}$$

The pdf of rth order statistics $X_{(r)}$ of the EGA distribution is obtained by inserting (3) and (4) into (8) and is

$$f_{X_{(r)}}(x,\theta,a,b) = \frac{4ab\theta^2 n!}{(r-1)!(n-r)!} x z^{a-1} e^{-2a\theta x}\left(1-z^a e^{-2a\theta x}\right)^{br-1}\left[1-\left(1-z^a e^{-2a\theta x}\right)^b\right]^{n-r}, \tag{9}$$

where $z = 1 + 2\theta x$. The pdf of the $1st$ or smallest order statistics of the EGA distribution is derived by setting $r = 1$ in equation (9) and is

$$f_{X_{(1)}}(x, \theta, a, b) = 4abn\theta^2 xz^{a-1}e^{-2a\theta x}\left(1 - z^a e^{-2a\theta x}\right)^{b-1}\left[1 - \left(1 - z^a e^{-2a\theta x}\right)^b\right]^{n-1}.$$

If $r = n$ then the pdf of the nth or largest order statistics of EGA distribution is given by

$$f_{X_{(n)}}(x, \theta, a, b) = 4ab\theta^2 nxz^{a-1}e^{-2a\theta x}\left(1 - z^a e^{-2a\theta x}\right)^{bn-1}.$$

5. Maximum Likelihood Estimation

This section is dedicated to maximum likelihood estimation as an estimation approach for EGA parameters.

Let X_1, X_2,..., X_n be a random sample of size n from EGA distribution then the likelihood function is given by

$$L = \prod_{i=1}^{n} 4ab\theta^2 x_i(1 + 2\theta x_i)^{a-1}e^{-2a\theta x_i}(1 - (1 + 2\theta x_i)^a e^{-2a\theta x_i})^{b-1}.$$

The log-likelihood is

$$l = nlog(4) + nlog(a) + nlog(b) + 2nlog(\theta) + \sum_{i=1}^{n} log(x_i) + a\sum_{i=1}^{n} log(1 + 2\theta x_i)$$

$$- \sum_{i=1}^{n} log(1 + 2\theta x_i) - 2a\theta\sum_{i=1}^{n} x_i + (b-1)\sum_{i=1}^{n} log(1 - (1 + 2\theta x_i)^a e^{-2a\theta x_i}). \quad (10)$$

Taking the the derivative of (10) with respect to parameters θ, a and b then we have the following equations:

$$\frac{\partial l}{\partial \theta} = \frac{2n}{\theta} + 2(a-1)\sum_{i=1}^{n}\frac{x_i}{z_i} - 2a\sum_{i=1}^{n} x_i + 4a\theta(b-1)\sum_{i=1}^{n}\frac{x_i^2 e^{-2a\theta x_i}z_i^{a-1}}{1 - z_i^a e^{-2a\theta x_i}}; \quad (11)$$

$$\frac{\partial l}{\partial a} = \frac{n}{a} + \sum_{i=1}^{n} log(z_i) - 2\theta\sum_{i=1}^{n} x_i + (b-1)\sum_{i=1}^{n}\frac{z_i^a e^{-2\theta a x_i}(2\theta x_i - log(z_i))}{1 - z_i^a e^{-2a\theta x_i}}; \quad (12)$$

$$\frac{\partial l}{\partial b} = \frac{n}{b} + \sum_{i=1}^{n} log(1 - z_i^a e^{-2a\theta x_i}), \quad (13)$$

where $z_i = 1 + 2\theta x_i$. The numerical maximum likelihood estimates of the EGA parameters can be computed by equating (11), (12) and (13) to 0, respectively.

6. Application

In this section, the EGA distribution is applied to two real datasets and compare with the following distributions:

- Exponentiated Ailamujia distribution (EA) [10]

$$f(x, \theta, b) = 4b\theta^2 xe^{-2\theta x}(1 - (1 + 2\theta x)e^{-2\theta x})^{b-1}, x \geq 0, \theta, b > 0.$$

- Ailamujia Inverted Weibull distribution (AIW) [12]

$$f(x, \theta, \alpha) = 4\alpha\theta^2 x^{-2\alpha-1}e^{-2\theta x^{-\alpha}}, x > 0, \theta, \alpha > 0.$$

In this application, we use two datasets from Badar[3]. These sets of data are presented as follow:

Data Set 1. This dataset is related to failure stresses (in GPa) and it is composed of 65 single carbon fibers of lengths 50 mm. The observations are given as follow: 1.339, 1.434, 1.549, 1.574, 1.589, 1.613, 1.746, 1.753, 1.764, 1.807, 1.812, 1.84, 1.852, 1.852, 1.862, 1.864, 1.931, 1.952, 1.974, 2.019, 2.051, 2.055, 2.058, 2.088, 2.125, 2.162, 2.171, 2.172, 2.18, 2.194, 2.211, 2.27, 2.272, 2.28, 2.299, 2.308, 2.335, 2.349, 2.356, 2.386, 2.39, 2.41, 2.43, 2.431, 2.458, 2.471, 2.497, 2.514, 2.558, 2.577, 2.593, 2.601, 2.604, 2.62, 2.633, 2.67, 2.682, 2.699, 2.705, 2.735, 2.785, 3.02, 3.042, 3.116 and 3.174.

Data Set 2. This dataset represents the strength data measured in GPa of 69 single carbon fibres tested under tension at gauge lengths of 20mm. The data is given as follows: 1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.997, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.140, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.570, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.880, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585 and 3.585.

In this analysis, we use the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov - Smirnov (K-S), Anderson - Darling (A) and Cramer - von Mises (W*) statistics for comparison. In addition, a package "fitdistrplus" in R is used to fit the distributions into given datasets.

The ML estimates of the fitted models to both sets of data are presented in Tables 1 and 3. Furthermore, results are given in Table 2 for first set of data and in Table 4 for second set of data. Tables 2 and 4 indicate that the proposed distribution provides better estimate for two given datasets as compared to the EA and AIW distributions since it has a smallest values of AIC, BIC, K-S, A and W*. Moreover, same results are observed from figures 5 and 6.

Table 1: *MLEs of the fitted models for a first set of data.*

Distribution	$\hat{\theta}$	$\hat{\alpha}$	$\hat{a}$	$\hat{b}$
EGA	0.007532943		5404.596	12.49782
EA	1.473919			58.063764
AIW	12.894223	3.542585		

Table 2: *Numerical values of AIC, BIC, K-S, A and W* of the fitted models for a first set of data.*

Distribution	AIC	BIC	$K-S$	A	$W*$
EGA	77.65784	84.181	0.08196969	0.43719505	0.07281970
EA	79.85257	84.20135	0.09598235	0.75832969	0.12159989
AIW	85.65023	89.99900	0.1150352	1.2112933	0.1948092

Table 3: *MLEs of the fitted models for a second set of data.*

Distribution	$\hat{\theta}$	$\hat{\alpha}$	$\hat{a}$	$\hat{b}$
EGA	0.006973733		4720.681000000	9.017470000
EA	1.204724000			31.814881000
AIW	10.924163000	3.000044000		

Table 4: *Numerical values of AIC, BIC, K-S, A and W* of the fitted models for a second set of data.*

Distribution	AIC	BIC	$K-S$	A	$W*$
EGA	107.4429	114.1452	0.06677067	0.45740745	0.06486004
EA	111.7837	116.2520	0.09042448	0.99232483	0.14097939
AIW	122.5224	126.9907	0.1199293	1.9072115	0.2835300

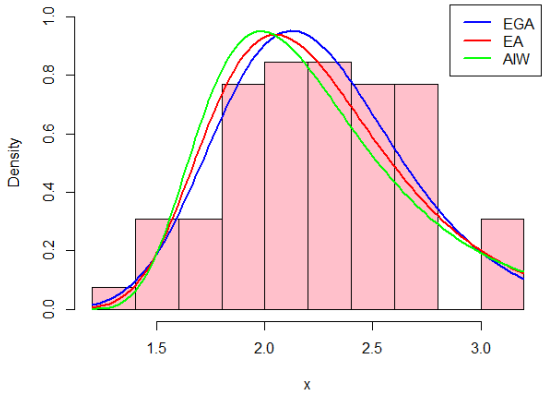


Figure 5: *Estimated pdf of the fitted models for the first set of data*

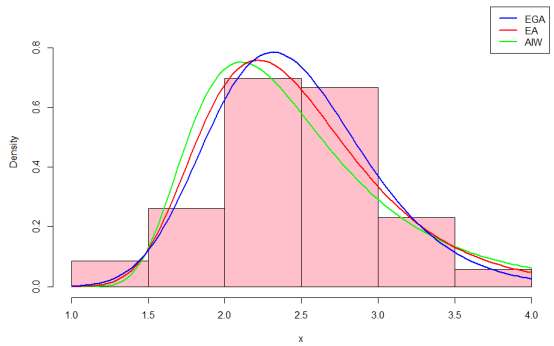


Figure 6: *Estimated pdf of the fitted models for the second set of data*

7. Concluding Remarks

In this paper, a novel generalized version of the Exponentiated Ailamujia distribution called Exponentiated Generalized Ailamujia distribution has been introduced. Some properties of the proposed distribution such as moments, mean, variance, harmonic mean, moment generating function, survival function, hazard function, reverse hazard rate, Mills ratio and order statistics were derived. Maximum likelihood approach was implemented to estimate the proposed distribution parameters. The applicability of the proposed distribution was evaluated by fitting on two real datasets and compared with the EA and AIW distributions. It was found that the proposed distribution provides better estimate for the given datasets compared to the said distributions.

References

[1] Ahmad, A., Ahamad, S. P. and Ahmed, A. (2017). Performance Rating of Transmuted Ailamujia Distribution: An Analytical Approach. *Journal of Applied Information Science*, 5:31–34.

[2] Aijaz, A., Ahmad, A. and Tripathi, R. (2020). Inverse analogue of Ailamujia distribution with statistical properties and applications. *Asian Research Journal of Mathematics*, 16:36–46.

[3] Badar, M. G. and Priest, A. M. (1982) Statistical Aspects of fiber and bundle strength in hybrid composites. In T. Hayashi, K. Kawata, and S. Umekawa (Eds). *Progress in Science and Engineering Composites*, ICCM-IV, Tokyo,1129–1136.

[4] Cordeiro, G. M., Ortega, E.M. M. and da Cunha, D.C. C. (2013). The Exponentiated Generalized Class of Distributions. *Journal of Data Science*, 11:1–27.

[5] Jamal, F., Chesneau, C., Aidi, K. and Ali, A. (2021). Theory and application of the power Ailamujia distribution. *Journal of Mathematical Modeling*, 9:391–413.

[6] Jayakumar B. and Elangovan, R. A. (2019). New generalization of Ailamujia distribution with applications in bladder cancer data. *IJSRMSS*, 6:61–68.

[7] Li, L. P. (2016). Minimax estimation of the parameter of Ailamujia disstribution under different loss functions. *Science Journal of Applied Mathematics and Statistics*, 4:229–235.

[8] Lv, H.Q., Gao, L.H. and Chen, C.L. (2002). Ailamujia distribution and its application in supportability data analysis. *Journal of Academy of Armored Force Engineering*, 16:28–52.

[9] Pan, G.T., Wang, C.L., Huang, Y.B. and Dang, M.T. (2009). The research of interval estimation and hypothetical test of small sample of Ailamujia distribution. *Application of Statitistics and Management*, 28:468–472.

[10] Rather, A., Subramanian, C., Al-Omari, A.I. and Alanzi, A.R.A. (2022). Exponentiated Ailamujia distribution with statistical inference and application of medical data. *Journal of Statistics and Management Systems*, DOI:10.1080/09720510.2021.1966206.

[11] Rather, A., Subramanian, C., Shafi, S., Malik, K. A., Ahmad, P. J., Para, B. A. and Jan, T. A. (2018). New size biased distribution with Applications in Engineering and Medical Science. *IJSRMSS*, 5:75–85.

[12] Smadi, M. M. and Ansari, S. I. (2022). Ailamujia Inverted Weibull Distribution with Application to Lifetime Data. *Pakistan Journal of Statistics*, 38:341–358.

[13] UI Ain, S.Q., Aijaz, A. and Tripathi, R. (2020). A new two parameter Ailamujia distribution with application in Bio-Medicine. *Journal of Xi'an University of Architechture and Technology*, 12:592–604.

[14] Uzma, J., Kawsar, F. and Ahmad, S. P. (2017). On weighted Ailamujia distribution and its applications to lifetime data. *Journal of Statistics Applications and Probability an International Journal*, 6:619–633.

THE EFFICIENCY OF ESTIMATING A POPULATION AVERAGE USING INDEX-TYPE ESTIMATORS IN SEQUENTIAL SAMPLING

Srinivasa Rao Kolli[1] ,U.V.Adinarayana Rao[2], Adilakshmi Siripurapu[3]

Taviti Naidu Gongada[4]

•

Dept. of Operations, GITAM school of Business, GITAM (Deemed to be University), Visakhapatnam, AP, India[1,2,4]

Dept. of Basic Science and Humanities, Vignan's Institute of Information Technology (A), Duvvada, Visakhapatnam, AP, India[3]

srinivaskolli4@gmail.com[1]
auppu@gitam.edu [2]
laxmimaths2008@gmail.com[3]
dr.tavitinaidugongada@gmail.com[4]

Abstract

This paper discusses the difficulty of approximating the population average of a variable y by knowledge about a supplementary variable x in the context of two successive (rotation) sampling occasions. The paper proposes a group of exponential-class estimators that includes the regular balanced estimator, produce-class estimator, and proportion-class estimator and suggests that these estimators are superior to existing estimators. The paragraph also mentions that the paper discusses optimal substitute statements and then the implementation of the recommended estimators, which may be important considerations for practical applications of the proposed methods. Finally, an empirical study is mentioned as supporting evidence for the research.

Keywords: Auxiliary variable, Study variable, Bias, Mean squared error, Successive sampling

1. Introduction

It is general towards procedure the supplementary knowledge at the approximation stand to achieve developed estimations of the population average y of the analyze variate x. Auxiliary information can be valuable in improving estimates of population parameters by incorporating additional relevant data. Ratio, Product, and Regression methods are good examples in this context. This statement is also generally true. Ratio, Product, and Regression methods are commonly used techniques for estimating auxiliary information. When the auxiliary variable x is positively (high) correlated with the study variable y, the Ratio estimation method is quite effective. The Ratio estimation method is effective when the auxiliary variable and the study

variable have a positive correlation. When there is a negative correlation between the auxiliary variable and the study variable, the Product estimation method is commonly employed. The theory of successive sampling has been developed by various researchers, starting with [1] and followed [2], [4], [5] and several others. The mentioned researchers have indeed contributed to developing the theory of successive sampling.

Auxiliary information was utilized by [6] and [7] multiple times to estimate the current population means in successive sampling. Supplementary information implemented by [6] and [7] within the framework of subsequent selection to estimate the population mean on the current occasion. Recent research conducted by [8] where they utilized auxiliary information on both events and proposed various estimators for estimating the population mean on the current occasion within the framework of two-occasion successive (rotation) sampling. In their work on two-occasion subsequent selection, [8] employed auxiliary information to develop estimators designed explicitly for estimating the population mean on the current occasion. These researchers have successfully used supplementary information and developed various estimation methods to evaluate the population mean on the recent occurrence in the context of successive sampling.

2. Introduces the formulation and notation of the proposed estimator

Consider a set $V = \{v_1, v_2, \ldots\ldots v_K\}$ Consisting of K elements. Let $[y, x]$ Represent the study and auxiliary variables, respectively, with $[y_i, x_i]$ denoting the values on the i^{th} unit $V_i\{i = 1,2,3, \ldots\ldots K\}$ of the population. Furthermore, let $[\hat{y}, \hat{x}]$ Represent the population values of $[y, x]$ Respectively. In this scenario, we assume that the population mean$\hat{X}$ Of the auxiliary variable x is already known. Subsequently, to estimate the population mean $\hat{Y}$ Of the study variable y, a simple random sample of size k is selected from the population V without replacement. The Classical Ratio Estimator and the Ordinary Product Estimator are two methods used to estimate the population mean $\hat{Y}$. The Classical Ratio Estimator is defined as $l_F = \tilde{F}\hat{X}$, where $\tilde{F} = \frac{\hat{y}}{\hat{x}}$ $\hat{x} \neq 0$ is the estimate of the Ratio F of the population means. The unweighted sample means y and x,, represented as $\hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i$ and $\hat{x} = \frac{1}{k}\sum_{i=1}^{k} x_i$, respectively, are utilized. The efficiency of the classical ratio estimator relies on a strong positive correlation between the variables y and x. On the other hand, the Ordinary Product Estimator for $\hat{Y}$ is defined as $l_S = \frac{\tilde{S}}{\hat{X}}$ where $\tilde{S} = \hat{y}.\hat{x}$ is the estimate of the Product S of the population means. The Product Estimator is commonly employed when there is an expectation of a strong negative correlation between the two variables.

The research on estimating the population mean $\hat{y}$ has led to the proposal of different estimators by various researchers. The Estimator l_S, credited to [9] and revisited by [10] is one such example. Additionally, [11] suggested Ratio and Product-Type Exponential Estimators for estimating the population mean $\hat{Y}$.

The Ratio-Type Exponential Estimator $l_{Fe} is\ defined\ as\ \hat{y}\, e^{\left[\frac{\hat{X}-\hat{x}}{\hat{X}+\hat{x}}\right]}$, while the Product-Type Exponential Estimator $l_{Se}\ is\ defined\ as\ \hat{y}\, e^{\left[\frac{\hat{x}-\hat{X}}{\hat{X}+\hat{x}}\right]}$ The simple Expansion Estimator l_0, which is used for estimating the population mean $\hat{Y}$ When none of the previously mentioned estimators are suitable. The Expansion Estimator is defined as$l_0 = \hat{y} = \frac{1}{k}\sum_{i=1}^{k} y_i$Which is the unweighted sample mean of y. Overall, the sentence accurately presents the contributions of different researchers in proposing various estimators for the population mean $\hat{y}$.

It discusses the findings of [12] regarding the variability of sample means and the coefficients of variation. The variability of the sample mean $\hat{x}$ is usually less than that of the sample mean $\hat{y}$. If then introduces $D[\hat{x}]$ as the coefficient of variation of $\hat{x}$ and $D[\hat{y}]$ as the coefficient of variation of $\hat{y}$. The sentence further provides the relationships between the squared coefficients of variation and the population coefficients of variation. It states that $D^2[\hat{x}] = \frac{[1-g]}{k} D_x^2$ and $D^2[\hat{y}] = \frac{[1-g]}{k} D_y^2$ where $g = \frac{k}{K}$ is the sampling fraction, $D_x = \frac{T_x}{\hat{X}}$ is the population coefficient of variation for x, and $D_y = \frac{T_y}{\hat{Y}}$ is the population coefficient of variation for y, $T_x^2 = [K-1]^{-1} \sum_{i=1}^{K} \left[x_i - \hat{X}\right]^2$ and $T_y^2 = [K-1]^{-1} \sum_{i=1}^{K} \left[y_i - \hat{Y}\right]^2$ Represent the sums of squared deviations from the respective population means. If $D_x = bD_y$, where b is a constant between 0 and 1, then $D[\hat{x}] = bD[\hat{y}]\,; 0 < b \leq 1$ This implies that the coefficient of variation for $\hat{x}$ is proportional to the coefficient of variation for $\hat{y}$. [10] findings on the variability of sample means and the coefficients of variation explicitly focus on the relationship between $D[\hat{x}]$ and $D[\hat{y}]$ when $D[\hat{x}] = bD[\hat{y}]$.

$$l_F \quad if \ \frac{\gamma}{b} > 0.5,$$
$$\hat{y} \ if -0.5 \leq \frac{\gamma}{b} \leq 0.5,$$
$$l_F \ if \ \frac{\gamma}{b} < -0.5\,,$$

Where, $\gamma = \frac{T_{xy}}{T_x T_y}$ and $T_{xy} = [K-1]^{-1} \sum_{i=1}^{K} \left[x_i - \hat{X}\right] \left[y_i - \hat{Y}\right]$

$$\frac{\gamma}{b} \geq -\frac{1}{b} \ and \ \frac{\gamma}{b} \leq \frac{1}{b} \ as \ |\gamma| \leq 1$$

3.The Suggested Estimator

Motivated by [13] based on that motivation, they derived a modified Exponential-Type Estimator for estimating the population mean $\hat{Y}$ as

$$l_{Me} = \hat{y} e^{\left\{\frac{[\hat{x}+\emptyset\hat{X}]-[X+\emptyset\hat{x}]}{[\hat{x}+\emptyset\hat{X}]+[X+\emptyset\hat{x}]}\right\}}$$
$$= \hat{y}\, e^{\left\{\frac{\hat{x}\,[1-\emptyset]+[\emptyset-1]\hat{X}}{\hat{x}\,[1+\emptyset]+[\emptyset+1]\hat{X}}\right\}}$$
$$= \hat{y}\ \ e^{\left\{\frac{[\emptyset-1]\hat{X}-[\emptyset-1]\hat{x}}{[\emptyset+1]\hat{X}+[\emptyset+1]\hat{x}}\right\}}$$
$$= \hat{y}\, e^{\left\{\frac{[\emptyset-1][\hat{X}-\hat{x}]}{[\emptyset+1][\hat{X}+\hat{x}]}\right\}}$$

Where $\emptyset$ is a scalar used as a design parameter. for $\emptyset = 1 \ \ l_{Me} = \hat{y}\, \emptyset = 0, l_{Me} = \ l_{Pe}$ the value of $\emptyset$ (presumably a variable or parameter) is very large, the proposed estimator

$$\lim_{\emptyset \to \infty} l_{Me} = \lim_{\emptyset \to \infty} \hat{y}\, e^{\left\{\frac{[\emptyset-1][\hat{X}-\hat{x}]}{[\emptyset+1][\hat{X}+\hat{x}]}\right\}}$$
$$\cong l_{Fe} = \hat{y}\, e^{\left\{\frac{[\hat{X}-\hat{x}]}{[\hat{X}+\hat{x}]}\right\}}$$

3.1.Sampling bias and mean squared error of estimator

Sampling bias refers to the systematic error when the sample used for estimation does not represent the studied population. It can occur due to non-random sampling, non-response bias, or selection bias. Understanding and addressing sampling bias is crucial for obtaining reliable and accurate estimators.

$$\hat{y} = \hat{Y}[1+e_0],\ \hat{x} = \hat{X}[1+e_1] \ni H[h_0] = H[h_1] = 0 \ \&$$
$$H[h_0^2] = \frac{[1-g]}{k} D_y^2,\ H[h_1^2] = \frac{[1-g]}{k} D_x^2,\ H[h_0 h_1] = \frac{[1-g]}{k} \gamma D_y D_x$$
$$l_{Mh} = \hat{Y}[1+h_0] e^{\left\{\frac{-[\emptyset-1]h_1}{[\emptyset+1][2+e_1]}\right\}}$$
$$= \hat{Y}[1+h_0] e^{\left\{\frac{[1-\emptyset]}{[1+\emptyset]} \frac{h_1}{2}\left[1+\frac{h_1}{2}\right]^{-1}\right\}}$$
$$= \hat{Y}[1+h_0] e^{\left\{\frac{Hh_1}{2}\left[1+\frac{h_1}{2}\right]^{-1}\right\}} \quad (1)$$

Where $H = \frac{[1-\emptyset]}{[1+\emptyset]}$

$$l_{Mh} \cong \hat{Y}\left[1 + h_0 + \frac{Rh_1}{2} + \frac{Rh_0h_1}{2} + \frac{G[G-2]}{8} h_1^2\right] \quad (2)$$
$$\left[l_{Mh} - \hat{Y}\right] \cong \hat{Y}\left[h_0 + \frac{Rh_1}{2} + \frac{Rh_0h_1}{2} + \frac{R[R-2]}{8} h_1^2\right] \quad (3)$$
$$A[l_{Mh}] = \hat{Y} \frac{[1-g]}{k} \left[\frac{R\gamma D_y D_x}{2} + \frac{R[R-2]D_x^2}{8}\right] \quad (4)$$
$$= A_0 \left[\frac{R}{2}\right]\left[\gamma + \frac{[R-2]}{4b}\right]$$

Where $A_0 = \frac{b(1-g)\hat{Y}C_y^2}{k}$ and $A_0 = \frac{(1-\emptyset)}{(1+\emptyset)}$

$$\left(l_{Me} - \hat{Y}\right)^2 \cong \hat{Y}^2\left[h_0^2 + \frac{R^2h_1^2}{4} + Re_0e_1\right] \quad (5)$$

$$\text{MSE}(l_{Mh}) = \text{MSE}(l_{Mh})_I = \frac{(1-g)}{k} \hat{Y}^2 \left[D_y^2 + R\gamma D_y D_x + \frac{R^2 D_x^2}{4}\right] \quad (6)$$
$$= U_0\left[1 + \frac{R^2b^2}{4} + bR\gamma\right]$$
Where $U_0 = \frac{(1-g)T_y^2}{k} = Var(\hat{y})$
The MSE (l_{Mh}) is minimum when $R_0 = -2\left(\frac{\gamma}{b}\right)$
$$\text{MSE of } (l_{Mh}) = U_0(1 - \gamma^2) \quad (7)$$
$$\widehat{y_a} = \hat{y} + \hat{\beta}\left(\hat{X} - \hat{x}\right) \quad (8)$$

Table 1: PREs Estimator $l_{Mh}$is better than $\hat{y}$, l_F , l_S l_{Fh} *and* l_{Sh}

Estimator	**PREs** $(*\hat{y})$				
	Population				
	1	2	3	4	5
$\hat{y}$	100.0000	100.0000	100.0000	100.0000	100.0000
l_F	66.5810	30.5860	156.3967	31.1061	56.2431
l_S	10.5463	7.6514	25.8171	92.9342	167.5887
l_{Fh}	781.3982	292.0779	197.7846	54.9135	74.5067
l_{Sh}	24.2836	19.0754	47.1121	133.0386	133.064

3.2. Effectiveness Relationship

Scalar R(or∅)" and true optimum value:

$$\text{Var}\quad (\hat{y}) = MSE(\hat{y}) = \left(\frac{1-g}{n}\right) T_y^2 = \left(\frac{1-g}{n}\right) \hat{y}^2 C_y^2 = U_0 \quad (9)$$
$$A(l_F) = A_0(A - \gamma)$$
$$A(l_S) = A_0\gamma$$
$$A(l_{Fe}) = \left(\frac{A_0}{8}\right)(3b - 4\gamma) \quad (10)$$
$$A(l_{Fe}) = \left(\frac{A_0}{8}\right)(4\gamma - b)$$
$$\text{MSE}(l_F) = U_0[1 + b^2 - 2\gamma b]$$

$$\text{MSE}= U_0[1 + b^2 + 2\gamma b]$$

$$\text{MSE}(l_S) = U_0\left[1 + \left(\frac{b}{4}\right)(b - 4\gamma)\right]$$

$$\text{MSE}(l_S) = U_0\left[1 + \left(\frac{b}{4}\right)(b + 4\gamma)\right] \quad (11)$$

Where $A_0 = \frac{b(1-g)}{k}\hat{Y}D_y^2$

$$MSE(\hat{y}) - \min. MSE(l_{Me}) = U_0\,\gamma^2 \geq 0 \quad (12)$$

$$MSE(l_R) - \min. MSE(l_{Me}) = U_0\,(a - \gamma)^2 \geq 0 \quad (13)$$

$$MSE(l_S) - \min. MSE(l_{Me}) = U_0\,(a + \gamma)^2 \geq 0 \quad (14)$$

$$MSE(l_{Re}) - \min. MSE(l_{Me}) = U_0\left(\frac{a}{2} - \gamma\right)^2 \geq 0$$

$$MSE(l_{Re}) - \min. MSE(l_{Me}) = U_0\left(\frac{a}{2} + \gamma\right)^2 \geq 0 \quad (15)$$

- The usual Unbiased Estimator $\hat{y}$ does not require the correlation between the study variable y and the auxiliary variable x to be zero. when $\gamma = 0$
- Usual Ratio Estimator l_R) and mentioning a condition when $b = \gamma$; in this condition, it states that both the Estimators l_S and l_{Me} are equally efficient.
- Ratio-Type Exponential Estimator [11], denoted as l_{Re}, does not hold when $b = 2\gamma$, which represents the scenario where both the estimators l_{Fe} and l_{Me} exhibit equal efficiency.
- Product-Type Exponential Estimator [11], referred to as l_{Fe}, is not applicable when $b = -2\gamma$, which corresponds to the situation where both the estimators l_{Se} and l_{Me} demonstrate equal efficiency.

Table 2: The proposed estimator l_{Mh} is superior to $\hat{y}$, l_F , l_S l_{Fh} *and* l_{Sh} when the value of R falls within a specific range.

Population	**The proposed estimator l_{Mh} Outperforms other estimators when the value of R lies within a specific range.**					**The standard range of R in which $l_{Mh}$is better than $\hat{y}$, l_F , l_S l_{Fh} *and* l_{Sh}**
	$\hat{y}$	l_F	l_S	l_{Fh}	l_{Sh}	
1	(-1.7765,0)	(-2,0.2235)	(-3.7765,2)	(-1,-0.7765)	(-2.7765,1)	(-1,-0.7765)
2	(-1.3669,0)	(-2,0.6331)	(-3.3669,2)	(-1,-0.3669)	(-2.3669,1)	(-1,-0.3669)
3	(-2.5742,0)	(-2,-0.5742)	(-4.5742,2)	(-1.5742, -1)	(-3.5742,1)	(-1.5742,- 0.5740)
4	(0,1.8673)	(-2,3.8673)	(-0.1328,2)	(-1,2.8673)	(0.8673,1)	(0.8673,1)
5	(0,6.3059)	(-2,8.3059)	(2,4.3054)	(-1,7.3059)	(1,5.3059)	(-1,4.3054)

3.3.When the scalar R (or ∅) does not align with its actual optimum value:

$$MSE(l_{Me}) - Variance\,(\hat{y}) = U_0 b^2\left[\frac{R^2}{4} + \frac{R\gamma}{b}\right]$$

$$= U_0 b^2\left[\frac{R^2}{4} - \frac{RR_0}{b}\right]$$

$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0]$$

$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 + R_0^2 - R_0^2\,]$$

$$= \frac{U_0 b^2}{4}[(R - R_0)^2 - R_0^2\,] \quad (16)$$

$$(R - R_0)^2 < R_0^2 \quad \text{i.e } |R - R_0| < |R_0^2|$$

$$min\left(0, -\frac{4\gamma}{b}\right) < R < max.\left(0, -\frac{4\gamma}{b}\right)$$

From eq. (15) and eq. (16), we have,

$$MSE(l_{Me}) - MSE\ (l_F) = U_0 b^2 \left[\frac{R^2}{4} + \frac{R\gamma}{b} - 1 + \frac{2\gamma}{b}\right]$$
$$= U_0 b^2 \left[\frac{R^2}{4} + \frac{R\gamma}{b} - 1 - R_0\right]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 - 4 - 4R_0]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 + R_0^2 - 4 - 4R_0 - R_0^2]$$
$$= \frac{U_0 b^2}{4}[(R - R_0)^2 - (2 + R_0)^2\ ] \quad (17)$$
$$(R - R_0)^2 < (2 + R_0)^2 \text{ i.e., } |R - R_0| < |2 + R_0|$$
$$\min, \left\{-2, 2\left(1 - \frac{2\gamma}{b}\right)\right\} < R < max\left\{-2, 2\left(1 - \frac{2\gamma}{b}\right)\right\}$$

From eq. (16) and eq. (17), we have,

$$MSE(l_{Me}) - MSE\ (l_S) = U_0 b^2 \left[\frac{R^2}{4} + \frac{R\gamma}{b} - 1 - \frac{2\gamma}{b}\right]$$
$$= U_0 b^2 \left[\frac{R^2}{4} - \frac{RR_0}{2} - 1 + R_0\right]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 - 4 + 4R_0]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 + R_0^2 - 4 + 4R_0 - R_0^2]$$
$$= \frac{U_0 b^2}{4}[(R - R_0)^2 - (2 - R_0)^2\ ] \quad (18)$$
$$(R - R_0)^2 < (2 - R_0)^2 \text{ i.e., } |R - R_0| < |2 - R_0|$$
$$\min, \left\{-2, 2\left(1 + \frac{2\gamma}{b}\right)\right\} < R < max\left\{-2, 2\left(1 + \frac{2\gamma}{b}\right)\right\}$$

From eq. (17) and eq. (18), we have,

$$MSE(l_{Me}) - MSE\ (l_{Fe}) = U_0 b^2 \left[\frac{R^2}{4} + \frac{R\gamma}{b} - \frac{1}{4} + \frac{\gamma}{b}\right]$$
$$= U_0 b^2 \left[\frac{R^2}{4} - \frac{RR_0}{2} - \frac{1}{4} - \frac{R_0}{2}\right]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 - 1 - 2R_0]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 + R_0^2 - 1 - 2R_0 - R_0^2]$$
$$= \frac{U_0 b^2}{4}[(R - R_0)^2 - (1 + R_0)^2\ ] \quad (19)$$
$$(R - R_0)^2 < (1 + R_0)^2 \text{ i.e., } |R - R_0| < |1 + R_0|$$
$$\min, \left\{-1, \left(1 - \frac{4\gamma}{b}\right)\right\} < R < max\left\{-1, \left(1 - \frac{4\gamma}{b}\right)\right\}$$

From eq. (18) and eq. (19), we have

$$MSE(l_{Me}) - MSE\ (l_{Fe}) = U_0 b^2 \left[\frac{R^2}{4} + \frac{R\gamma}{b} - \frac{1}{4} - \frac{\gamma}{b}\right]$$
$$= U_0 b^2 \left[\frac{R^2}{4} - \frac{RR_0}{2} - \frac{1}{4} + \frac{R_0}{2}\right]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 - 1 + 2R_0]$$
$$= \frac{U_0 b^2}{4}[R^2 - 2RR_0 + R_0^2 - 1 + 2R_0 - R_0^2]$$
$$= \frac{U_0 b^2}{4}[(R - R_0)^2 - (1 - R_0)^2\ ] \quad (20)$$
$$(R - R_0)^2 < (1 - R_0)^2 \text{ i.e., } |R - R_0| < |1 - R_0|$$
$$\min, \left\{1 - \left(1 + \frac{4\gamma}{b}\right)\right\} < R < max\left\{1 - \left(1 + \frac{4\gamma}{b}\right)\right\}$$

- The usual Unbiased Estimator $\hat{x}$ is applicable when $|R - R_0| < |R_0|$ Which can be equivalently expressed as $\min\left\{\left(0, -\frac{4\gamma}{b}\right)\right\} < R < max\left\{\left(0, -\frac{4\gamma}{b}\right)\right\}$
- The usual Ratio Estimator l_F is valid when $|R - R_0| < |2 + R_0|$, which can be alternately expressed as $\min\left\{2, -2\left(1 - \frac{2\gamma}{b}\right)\right\} < R < max\left\{-2{,}2\ \left(1 - \frac{2\gamma}{b}\right)\right\}$
- The usual Product Estimator l_S is valid when $|R - R_0| < |2 - R_0|$, which can be alternately expressed as $\min\left\{2, -2\left(1 + \frac{2\gamma}{b}\right)\right\} < R < max\left\{-2{,}2\ \left(1 + \frac{2\gamma}{b}\right)\right\}$

- The Ratio-Type Exponential Estimator l_{Fe} introduced by Bah and Tuteja 1991[11] is applicable if $|R - R_0| < |1 + R_0|$, which can be equivalently expressed as $\min\left\{-1, \left(1 - \frac{4\gamma}{b}\right)\right\} < R < max\left\{-1, \left(1 - \frac{4\gamma}{b}\right)\right\}$.
- The Product-Type Exponential Estimator l_{Fe} introduced Bahl and Tuteja 1991[11] if $|R - R_0| < |1 - R_0|$ which can be equivalently expressed as $\min\left\{1 - \left(1 + \frac{4\gamma}{b}\right)\right\} < R < max\left\{1 - \left(1 + \frac{4\gamma}{b}\right)\right\}$

Table 3: PREs Estimator l_{Mh} concerning $\hat{y}$, for different values of R

R	$\emptyset$	**PREs** $(*\hat{y})$				
		Population				
		1	2	3	4	5
-2.0000	-3.0000	66.5810	30.5860	156.3967	31.1061	56.2431
-1.7500	-3.6667	105.4984	45.4209	182.7966	35.5534	60.2317
-1.5000	-5.0000	187.1938	73.6467	202.4393	40.8775	64.5841
-1.2500	-9.0000	383.2829	135.4864	208.2654	47.2636	69.3319
-1.0000	0.0000	781.3982	292.0779	197.7846	54.9135	74.5067
-0.7500	7.0000	738.4366	585.7795	175.3442	64.0167	80.1386
-0.5000	3.0000	353.0569	448.2367	148.3074	74.6863	86.2537
-0.2500	1.6667	174.9985	200.1900	122.3232	86.8380	92.8714
0.0000	1.0000	100.0000	100.0000	100.0000	100.0000	100.0000
0.2500	0.6000	63.7373	57.9872	81.8494	113.0935	107.6315
0.5000	0.3333	43.8933	37.4100	67.4411	124.3406	115.7344
0.7500	0.1429	31.9699	26.0031	56.0835	131.5695	124.2463
1.0000	0.0000	24.2836	19.0754	47.1121	133.0386	133.0649
1.2500	0.1111	19.0533	14.5708	39.9776	128.3594	142.0399
1.5000	-0.2000	15.3392	11.4840	34.2528	118.7285	150.9669
1.7500	0.2727	12.6097	9.2794	29.6138	106.2421	159.5862
2.0000	-0.3333	10.5463	7.6514	25.8171	92.9342	167.5887

4. The precision of first-order approximations to mean squared errors (MSEs).

Once we have compared the MSEs of the Proposed Estimator and other estimators using first-order approximations, our attention turns towards evaluating the accuracy of these approximations by deriving second-order approximations for the MSEs. In this analysis, we assume that $C_x = C_x = C_{,}$, taking a value of 1, and that the sample is drawn from a large Bivariate Normal population. However, it is essential to mention that more complex expressions are derived for other scenarios, as discussed in [13]

$$\vartheta_{30} = \vartheta_{03} = \vartheta_{12} = \vartheta_{21} = 0$$
$$\vartheta_{04} = \vartheta_{40} = 3D^4$$
$$\vartheta_{31} = \vartheta_{13} = 3\gamma D^4$$
$$\vartheta_{22} = (1 + 2\gamma^2)D^4$$

Where $E\left(e_0^i e_1^j\right) \cong \frac{\vartheta_{(i,j)}}{k^a}$ $, i, j = 0,1,2,4$ $and\ a = 2\ for i + j = 4$

$$l_{Me} = \hat{Y}(1 + h_0)e^{\left[\frac{Rh_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right]}$$
$$= \hat{Y}(1 + h_0)\left\{1 + \frac{Rh_1}{2} + \frac{R(R-2)h_1^2}{8} + \frac{R(R^2-6R+6)h_1^3}{48} + \frac{R(R^3-12R^2+36R-24)h_1^4}{384}\right\} - \quad (21)$$
$$b_1 = \frac{R}{2}, b_2 = \frac{R(R-2)}{8}$$
$$b_3 = \frac{R(R^2-6R+6)h_1^3}{48}, \quad b_4 = \frac{R(R^3-12R^2+36R-24)h_1^4}{384}$$

Then $l_{Me} = \hat{Y}(1 + h_0)[1 + bh_1 + b_2h_1^2 + b_3h_1^3 + bh_1^4 + \cdots]$

$$= \hat{Y}[1 + bh_1 + b_1h_0h_1 + bh_1^2 + b_2h_0h_1^2 + bh_1^3 + b_3h_0h_1^3 + b_4h_{1+\cdots\ldots}^4] \quad (22)$$

By disregarding terms of h's with powers higher than four, we obtain the following.

result:

$$l_{Me} \cong \hat{Y}[1 + h_0 + b_1(h_1 + h_0h_1) + b_2(h_1^2 + h_0h_1^2) + b_3(h_1^3 + h_0h_1^3) + b_4h_1^4] \quad (23)$$

By computing the expectation of both sides of equation (23), we derive the bias of l_{Me} as follows:

$$A(l_{Me}) = \frac{\hat{Y}D^2}{k}\left\{(\gamma b_1 + b_2) + \frac{3D^2}{k}(\gamma b_3 + b_4)\right\} \quad (24)$$

By squaring both sides of equation (24) and disregarding terms of $h's$ with powers higher than two, we obtain the following result:

$$\begin{aligned}\left(l_{Me} - \hat{Y}\right)^2 = \hat{Y}^2\{&h_0^2 + 2b_1h_0h_1 + b_1^2h_1^2 + 2(b_1^2 + b_2)h_0h_1^2 + 2b_1h_0^2h_1 + 2b_1b_2h_1^3 + (b_1^2 + \\ &2b_2)h_0^2h_1^2 + 2(b_3 + 2b_1b_2)h_0h_1^3 + (b_2^2 + 2b_1b_3)h_1^4\} \quad (25)\end{aligned}$$

$$\begin{aligned}MSE(l_{Me}) &= \left\{MSE(l_{Me}) + \left(\frac{D^4\hat{Y}^2}{k^2}\right)[(b_1^2 + 2b_2)(1 + 2\gamma^2) + 6(h_3 + 2b_1b_2) + 3(a_2^2 + 2b_1b_2)]\right\} \\ &= MSE(l_{Me}) + \left(\frac{D^4\hat{Y}^2}{64k^2}\right)\{7R^4 + 4R^3(14\gamma - 9) + 4R^2(16\gamma^2 - 36\gamma + 17) - \\ &\quad 16R(4\gamma^2 - 3\gamma + 2)\} \quad (26)\end{aligned}$$

$R \approx R_0 = -2\gamma$(with a =1). Then putting $R = -2\gamma$ in eq. (25), we have,

If a reliable estimate of R_0 is available, denoted as $R \approx R_0 = -2\gamma$ (with a = 1), then substituting $= -2\gamma$ into equation (26), we obtain the following result:

$$\begin{aligned}MSE(l_{Me}) &= MSE(l_{Me}) + \frac{D^4\hat{Y}^2}{4k^2}\gamma(4 + 11\gamma - 10\gamma^2 - 5\gamma^3) \\ &= MSE(l_{Me})\left\{1 + \frac{D^4}{4k}\gamma\frac{(4+15\gamma+5\gamma^2)}{(1+\gamma)}\right\} \\ &= \frac{D^2\hat{Y}^2}{k}\left\{(1 - \gamma^2) + \frac{D^2}{4k}\gamma(1 - \gamma)(4 + 15\gamma + 5\gamma^2)\right\}\end{aligned}$$

$$MSE(l_{Me}) = MSE\left(\hat{y}_{tf}\right) = \frac{D^4\hat{Y}^2}{k}(1 - \gamma^2) \quad (27)$$

$$\begin{aligned}MSE(l_{Re}) &= MSE(l_{Re})\left\{1 + \frac{D^2}{16k}\left(\frac{(143-248\gamma+128\gamma^2)}{(5-4\gamma)}\right)\right\} \\ &= \frac{D^2\hat{Y}^2}{k}\left\{\left(\frac{5}{4} - \gamma\right) + \frac{D^2}{64k}(143 - 248\gamma + 128\gamma^2)\right\}\end{aligned}$$

$$\begin{aligned}MSE(l_{Re}) &= MSE(l_{Re})\left\{1 + \frac{D^2}{16k}\left(\frac{7-40\gamma}{(5+4\gamma)}\right)\right\} \\ &= \frac{D^2\hat{Y}^2}{k}\left\{\left(\frac{5}{4} + \gamma\right) + \frac{D^2}{64k}(7 - 40\gamma)\right\} \quad (28)\end{aligned}$$

$$MSE(l_{Re}) = \frac{D^2\hat{Y}^2}{4k}(5 - 4\gamma)$$

$$MSE(l_{Re}) = \frac{D^2\hat{Y}^2}{4k}(5 + 4\gamma) \quad (29)$$

$$MSE(l_{Re}) = MSE(l_{Me}) = \frac{D^2\hat{Y}^2}{k}\left\{\left(\frac{1}{2} - \gamma\right)^2 + \frac{D^2}{64k}(143 - 312\gamma - 48\gamma^2 + 160\gamma^3 + 80\gamma^4)\right\} \quad (30)$$

$$\left\{\left(\frac{1}{2} - \gamma\right)^2 + \frac{D^2}{64k}(143 - 312\gamma - 48\gamma^2 + 160\gamma^3 + 80\gamma^4)\right\} > 0$$

$$MSE(l_{Re}) - MSE(l_{Me}) = \frac{D^2\hat{Y}^2}{k}\left\{\left(\frac{1}{2} + \gamma\right)^2 + \frac{D^2}{64k}(7 - 104\gamma - 176\gamma^2 + 160\gamma^3 + 80\gamma^4)\right\}$$

$$\left\{\left(\frac{1}{2} + \gamma\right)^2 + \frac{D^2}{64k}(7 - 104\gamma - 176\gamma^2 + 160\gamma^3 + 80\gamma^4)\right\} > 0 \quad (31)$$

$$MSE(\hat{y}_{th}) = D^2\hat{Y}^2(1 - \gamma^2)\left(\frac{1}{k} + \frac{1}{k^2}\right)$$

$$MSE(\hat{y}_{th}) - MSE(l_{Me}) = \frac{D^2\hat{Y}^2}{k}\left\{(1 - \gamma^2) - \frac{D^2}{k}\gamma(1 - \gamma)(4 + 15\gamma + 5\gamma^2)\right\} > 0 \quad (32)$$

$$\text{If } \left\{(1 - \gamma^2) - \frac{D^2}{k}\gamma(1 - \gamma)(4 + 15\gamma + 5\gamma^2)\right\} > 0$$

$D_x = D_x = D$, where b=1 the Mean Squared Error (MSE) of the usual Ratio Estimator $l_{F=\hat{y}}\ \left(\frac{\hat{X}}{\hat{x}}\right)$ and the Product Estimator $l_{S=\hat{y}}\ \left(\frac{\hat{X}}{\hat{x}}\right)$

$$\begin{aligned}MSE(l_F)_2 &= MSE(l_F)_1\left\{1 + \frac{D^2}{k}(6 - 3\gamma)\right\} \\ &= \left\{\frac{D^2\hat{Y}^2}{k}\left(2(1 - \gamma) + \frac{6D^2}{k}(2 - 3\gamma + \gamma^2)\right)\right\}\end{aligned}$$

$$MSE(l_F)_2 = MSE(l_F)_1\left\{1+\frac{D^2}{2k}\frac{(1+2\gamma^2)}{(1+\gamma)}\right\}$$
$$=\left\{\frac{D^2\hat{Y}^2}{k}\left(2(1+\gamma)+\frac{D^2}{k}(1+2\gamma^2)\right)\right\} \quad (33)$$

From eq. (32) and eq. (33), we have

$$MSE(l_F)_2 - MSE(l_{Fe})_2 = \left(\frac{D^2\hat{Y}^2}{k}\right)\left\{\left(\frac{3}{4}-\gamma\right)+\frac{D^2}{k}\left[6(1-\gamma)(2-\gamma)-\frac{1}{64}(143-248\gamma+128\gamma^2)\right]\right\}$$
$$\left\{\left(\frac{3}{4}-\gamma\right)+\frac{D^2}{k}\left[6(1-\gamma)(2-\gamma)-\frac{1}{64}(143-248\gamma+128\gamma^2)\right]\right\}>0 \quad (34)$$

from eq. (33) and eq. (34), we have,

$$MSE(l_S)_2 - MSE(l_{Sh})_2 = \left(\frac{D^2\hat{Y}^2}{k}\right)\left\{\left(\frac{3}{4}+\gamma\right)+\frac{D^2}{64k}(57+168\gamma)\right\}$$
$$\left\{\left(\frac{3}{4}+\gamma\right)+\frac{D^2}{64k}(57+168\gamma)\right\}>0 \quad (35)$$

5.Conclusion

The research findings highlight the significant impact of introducing new product-type and ratio-type estimators on the efficiency of estimating a population average by index-type estimators in the sequential random sample. The findings demonstrate that incorporating these new types of estimators increases the population average approximations' accuracy, thereby enhancing the overall efficiency of the sampling process. These findings have practical implications for researchers and practitioners in various fields, providing valuable insights to optimize their sampling strategies and obtain more reliable population mean estimates. Future research can explore additional aspects of these new estimators and their potential applications in different sampling scenarios to enhance the efficiency of sequential sampling techniques further.

References

[1] R. J. Jessen, Statistical investigation of a sample survey for obtaining form facts. Iowa Agricultural Experiment Station Road Bulletin no. 304, Ames, USA, (1942).

[2] H. D. Patterson, Sampling on successive occasions with partial replacement of units, *Jour. Roy. Statist. Assoc*, B, 12, 241- 255, (1950).

[3] A. R. Eckler, Rotation sampling. *Ann. Math. Statist.* 26, 664-685, (1955)

[4] J.N.K. Rao and J. E. Graham, Rotation design for sampling on repeated occasions. *Jour. Amer. Statist. Assoc.* 59, 492-509, (1964).

[5] Singh, G.N. (2005): On the use of chain type ratio estimator in successive Sampling. Statistics in Transition - new series,7(1), 21- 26.

[6] S. Feng and G. Zou, Sampling rotation method with auxiliary variable. *Commun. Statist. Theo. Meth.* 26(6), 1497-1509, (1997).

[7] R. S. Biradar, and H.P. Singh, Successive sampling using auxiliary information on both occasions. *Calcutta Statist. Assoc. Bull*., 51, 243-251, (2001).

[8] H. P. Singh and G. K. Vishwakarma, A general procedure for estimating population mean in successive sampling. *Commun. Statist. Theo. Meth.*, 38, 293-308, (2009).

[9] Robson, D. S. (1957). Applications of Multipart K-statistics to the Theory of Unbiased Ratio-Type Estimation. *Journal of the American Statistical Association*, 52, 511–522.

[10] Murthy, M. N. (1964). Product method of estimation. *Sankhya*, 26, 69-74.

[11] Bahl, S. and Tuteja, R.K. (1991). Ratio and Product-Type Exponential Estimators. *Journal Information Optimization Science*, 12(1), 159-164.

[12] Murthy, M. N. (1967). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta, India.

[13] Sahai, A. (1979). An efficient variant of the Product and Ratio Estimators. *Statistical Neerlandica*, 32, 27-35.

A TWO NON-IDENTICAL UNIT PARALLEL SYSTEM WITH PRIORITY IN REPAIR

Alka Chaudhary, Shivali Sharma
•
Principal, KLPG College, Meerut, 250002
Department of Statistics, Meerut College, Meerut, 250004
alkachaudhary527@gmail.com, sharma14shivali@gmail.com

Abstract

The paper deals with a system composed of two-non identical units (unit-1 and unit-2). Initially both the units are arranged in parallel configuration. Each unit has two possible modes- Normal (N) and Total Failure (F). The first unit gets priority in repair. System failure occurs when both the units stop functioning. A single repairman is always available with the system to repair a totally failed unit and repair discipline is first come, first served (FCFS).If during the repair of a failed unit the other unit also fails, then the later failed unit waits for repair until the repair of the earlier failed unit is completed. The repair times of both the units are exponential distribution with different parameters. Each repaired unit works as good as new. Using regenerative point technique, various important measures of system effectiveness have been obtained.

Keywords: Transition probabilities, mean sojourn time, reliability, MTSF, availability, expected busy period of repairman, net expected profit.

1. Introduction

Reliability is an important concept in the planning design and operation stages of various complex systems. Reliability is a significant area that is accepting awareness internationally and it is crucial for actual usage and care of any industrial system. It requires technical Knowledge for growing system effectiveness by decreasing the frequency of failure and reducing the worth of maintenance. Chaudhary and Tyagi [3] analyzed a two non-identical unit parallel system with two types of failure. Pundir et al. [7] analyzed a two non-identical unit parallel system with priority in repair. Chaudhary et al. [5] analyzed two non-identical unit warm standby repairable system with two types of failure. Saxena et al. [9] analyzed two unit parallel system with working and rest time of repairman.

A single repairman is always available with the system to repair a totally failed unit and repair discipline is first come, first served (FCFS). Chaudhary and Masih [1, 4] analyzed a two non-identical unit. Saini et al. [8], Chaudhary and Sharma [2] and Dabas et al. [6] analyzed a two non-identical unit system models assuming two modes- Normal mode and total failure mode of each unit and analyzed parallel system with priority in repair. System failure occurs when both the units stop functioning. The first unit gets priority in repair.

By using regenerative point technique, the following measures of system effectiveness are obtained-

i. Transition probabilities and mean sojourn times in various states.
ii. Reliability and mean time to system failure (MTSF).
iii. Point-wise and steady-state availabilities of the system as well as expected up time of the system during time interval (0, t).
iv. Expected busy period of repairman in the repair of unit-1 and unit-2 during time interval (0, t).
v. Net expected profit earned by the system in time interval (0, t).

2. System Description and Assumptions

1. The system comprises of two non-identical units (unit-1 and unit-2). Initially, both the units work in parallel configuration.
2. Each unit of the system has two possible modes-Normal (N) and total failure (F).
3. The first unit gets priority in repair.
4. System failure occurs when both the units stop functioning.
5. A single repairman is always available with the system to repair a totally failed unit and repair discipline is first come, first served (FCFS).
6. If during the repair of a failed unit the other unit also fails, then the later failed unit waits for repair until the repair of the earlier failed unit is completed.
7. The repair time of both the units is exponential distribution with different parameters. Each repaired unit works as good as new.

3. Notations and States of the System

We define the following symbols for generating the various states of the system-

N_{1O}, N_{2O} : Unit-1and Unit-2 is in N-mode and operative in parallel.
F_{1r}, F_{2r} : Unit-1 and Unit-2 is in F-mode and under repair.
F_{2w} : Unit-2 is in F-mode and under waiting for repair.

Considering the above symbols in view of assumptions stated in section-2, the possible states of the system are shown in the transition diagram represented by **Figure. 1**. It is to be noted that the epochs of transitions into the state S_1 from S_2 are non-regenerative, whereas all the other entrance epochs into the states of the systems are regenerative.

The other notations used are defined as follows:

E : Set of regenerative states.
α_1, α_2 : Constant Failure rate of Unit-1 and Unit- 2.
β_1, β_2 : Constant Repair rate of Unit-1 and Unit- 2.
$G(\bullet)$: CDF of time to repair and its repair is continued to state S_1
$H(\bullet)$: General Distribution of Unit-2
$*$: Symbol for Laplace Transform i.e. $g_{ij}^{*}(s) = \int e^{-st} q_{ij}(t)dt$
$\sim$: Symbol for Laplace Stieltjes Transform i.e. $\tilde{Q}_{ij}(s) = \int e^{-st} dQ_{ij}(t)t$
© : Symbol for ordinary convolution i.e. $A(t)©B(t)=\int_0^t A(u)B(t-u)du$

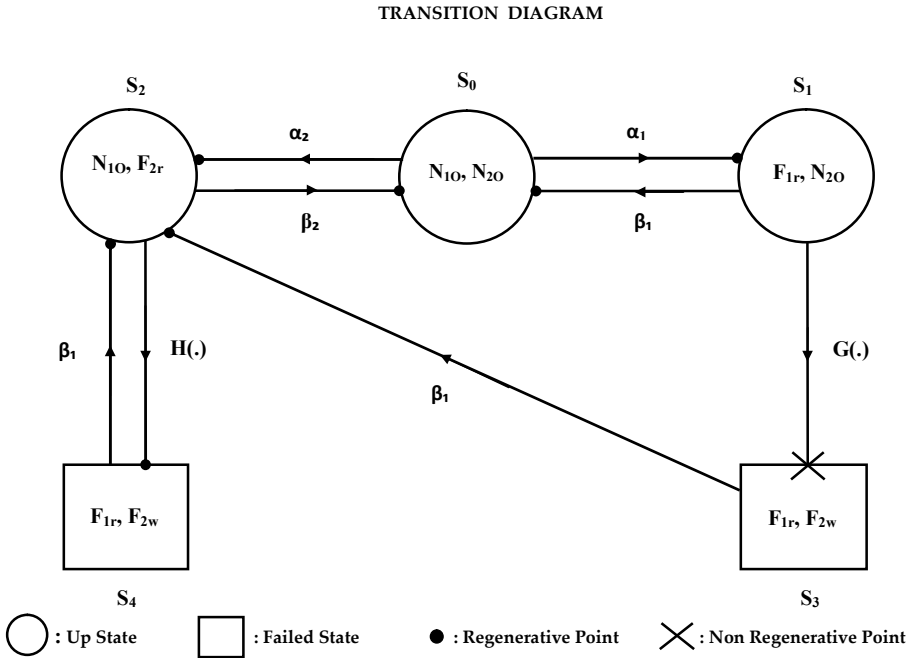


Figure 1: *Exponential Model*

4. Transition Probabilities and Sojourn Times

Let X(t) be the state of the system at epoch t, then $\{X(t); t \geq 0\}$ constitutes a continuous parametric Markov-Chain with state space $E = \{S_0 \text{to} S_4\}$ .The various measures of system effectiveness are obtained in terms of steady-state transition probabilities as follows:

$$p_{01} = \int \alpha_1 e^{-(\alpha_1+\alpha_2)t} dt = \frac{\alpha_1}{\alpha_1+\alpha_2}$$

$$p_{02} = \int \alpha_2 e^{-(\alpha_1+\alpha_2)t} dt = \frac{\alpha_2}{\alpha_1+\alpha_2}$$

$$p_{10} = \int \beta_1 e^{-\beta_1 t} \overline{G}(t) dt = 1 - \widetilde{G}(\beta_1)$$

$$p_{20} = \int \beta_2 e^{-\beta_2 t} \overline{H}(t) dt = 1 - \widetilde{H}(\beta_2)$$

$$p_{24} = \int e^{-\beta_2 t} dH(t) = \widetilde{H}(\beta_2) \qquad p_{42} = \int \beta_1 e^{-\beta_1 t} dt = 1 = p_{32}$$

The two step transition probability (Steady State) is given by

$$p_{12}^{(3)} = \left(1 - e^{-\beta_1 t}\right) \int_0^v e^{-\beta_1 u} dG(u) = \widetilde{G}(\beta_1)$$

It can be easily verified that,

$$p_{01} + p_{02} = 1, \tag{1}$$

$$p_{10} + p_{12}^{(3)} = 1, \tag{2}$$

$$p_{32} = p_{42} = 1, \tag{3}$$

$$p_{20} + p_{24} = 1 \tag{4}$$

[+]The limits of integration are 0 to ∞ whenever they are not mentioned.

5. Mean Sojourn Time

The mean sojourn time ψ_i in state S_i is defined as the expected time taken by the system in state S_i before transiting into any other state. If random variable U_i denotes the sojourn time in state S_i then,

$$\psi_i = \int P[U_i > t]dt$$

Therefore, its values for various regenerative states are as follows-

$$\psi_0 = \int e^{-(\alpha_1+\alpha_2)t}dt = \frac{1}{(\alpha_1+\alpha_2)} \tag{5}$$

So that,

$$\psi_1 = \int e^{-\beta_1 t}\overline{G}(t)\,dt \tag{6}$$

$$\psi_2 = \int e^{-\beta_2 t}\overline{H}(t)\,dt \tag{7}$$

$$\psi_4 = \int e^{-\beta_1 t}dt = \frac{1}{\beta_1} \tag{8}$$

6. Analysis of Characteristics

6.1. Reliability and MTSF

Let $R_i(t)$ be the probability that the system operates during (0, t) given that at t=0 system starts from $S_i \in E$. To obtain it we assume the failed states S_2 and S_4 as absorbing. By simple probabilistic arguments, the value of $R_0(t)$ in terms of its Laplace Transform (L.T.) is given by

$$R_0^*(s) = \frac{Z_0^* + q_{01}^* Z_1^* + q_{02}^* Z_2^*}{1 - q_{01}^* q_{10} - q_{02}^* q_{20}^*} \tag{9}$$

We have omitted the argument's from $q_{ij}^*(s)$ and $Z_i^*(s)$ for brevity. $Z_i^*(s)$; i = 0, 1, 2 are the L. T. of

$$Z_0(t) = e^{-(\alpha_1+\alpha_2)(1-r)t}, \qquad Z_1(t) = \int e^{-\beta_1 t}\overline{G}(t)dt, \qquad Z_2(t) = \int e^{-\beta_2 t}\overline{H}(t)\,dt$$

Taking the Inverse Laplace Transform of (9), one can get the reliability of the system when system initially starts from state S_0 .

The MTSF is given by,

$$E(T_0) = \int R_0(t) = \lim_{s\to 0} R_0^*(s) = \frac{\psi_0 + p_{01}\psi_1 + p_{02}\psi_2}{1 - p_{01}p_{10} - p_{02}p_{20}} \tag{10}$$

6.2. Availability Analysis

Let $A_i(t)$ be the probability that the system is up at epoch t, when initially it starts operation from state $S_i \in E$. Using the regenerative point technique and the tools of Laplace transform, one can obtain the value of $A_0(t)$ in terms of its Laplace transforms i.e. $A_0^*(s)$ given as follows-

$$A_0^*(s) = \frac{N_1(s)}{D_1(s)} \tag{11}$$

Where,

$$N_1(s) = Z_0^*\left[1 - q_{24}^* q_{42}^*\right] + Z_1^* q_{01}^*\left[1 - q_{24}^* q_{42}^*\right] + Z_2^*\left[q_{01}^* q_{12}^{(3)*} + q_{02}^*\right]$$

and

$$D_1(s) = 1 - q_{24}^* q_{42}^* - q_{10}^* q_{01}^* \left(1 - q_{24}^* q_{42}^*\right) - q_{20}^* \left(q_{01}^* q_{12}^{(3)*} + q_{02}^*\right) \tag{12}$$

Where, $Z_i(t)$, i=0,1,2 are same as given in section 6.1.

The steady-state availability of the system is given by

$$A_0 = \lim_{t\to\infty} A_0(t) = \lim_{s\to 0} s\, A_0^*(s) \tag{13}$$

We observe that

$$D_1(0) = 0$$

Therefore, by using L. Hospital's rule the steady state availability is given by

$$A_0 = \lim_{s\to 0} \frac{N_1(s)}{D_1'(s)} = \frac{N_1}{D_1'} \tag{14}$$

Where,

$$N_1 = \psi_0 \left[1 - p_{24} p_{42}\right] + \psi_1 p_{01} \left[1 - p_{24} p_{42}\right] + \psi_2 \left[p_{01} p_{12}^{(3)} + p_{02}\right]$$

and

$$D_1' = \psi_0 p_{20} + \psi_1 p_{01} p_{20} + \psi_2 \left(1 - p_{10} p_{01}\right) + \psi_4 p_{24} \left(1 - p_{10} p_{01}\right) \tag{15}$$

The expected up time of the system in interval (0, t) is given by

$$\mu_{up}(t) = \int_0^t A_0(u)\, du$$

So that, $\mu_{up}^*(s) = \dfrac{A_0^*(s)}{s}$ (16)

6.3. Busy Period Analysis

Let $B_i^1(t)$ and $B_i^2(t)$ be the respective probabilities that the repairman is busy in the repair of unit-1 failed due to first repair with priority of unit-1 and unit-2 failed due to second repair at epoch t, when initially the system starts operation from state $S_i \in E$. Using the regenerative point technique and the tools of L. T., one can obtain the values of above two probabilities in terms of their L. T. i.e. $B_i^{1*}(s)$ and $B_i^{2*}(s)$ as follows-

$$B_i^{1*}(s) = \frac{N_2(s)}{D_1(s)}, \qquad B_i^{2*}(s) = \frac{N_3(s)}{D_1(s)} \tag{17-18}$$

Where,

$$N_2(s) = Z_1^* q_{01}^* \left(1 - q_{42}^* q_{24}^*\right) + Z_4^* \left(q_{01}^* q_{12}^{(3)*} q_{24}^* + q_{02}^* q_{24}^*\right)$$

and

$$N_3(s) = Z_2^* \left(q_{01}^* q_{12}^{(3)*} + q_{02}^*\right)$$

and $D_1(s)$ is same as defined by the expression (12) of section VI(II).

The steady state results for the above two probabilities are given by-

$$B_0^1 = \lim_{s\to 0} s\, B_0^{1*}(s) = N_2 \setminus D_1' \quad \text{and} \quad B_0^2 = \lim_{s\to 0} s\, B_0^{2*}(s) = N_3 \setminus D_1' \tag{19-20}$$

Where,

$$N_2(0) = \psi_1 p_{01} \left(1 - p_{24}\right) + \psi_4 \left(p_{01} p_{12}^{(3)} p_{24} + p_{02} p_{24}\right) \tag{21}$$

$$N_3(0) = \psi_2 \left(p_{01} p_{12}^{(3)} + p_{02}\right) \tag{22}$$

and D_1' is same as given in the expression (15) of section 6.2.

The expected busy period in repair of unit-1 failed due to first repair with priority of unit-1 and unit-2 failed due to second repair during time interval (0, t) are respectively given by-

$$\mu_b^1(t) = \int_0^t B_0^1(u)\,du \qquad \text{and} \qquad \mu_b^2(t) = \int_0^t B_0^2(u)\,du$$

So that,

$$\mu_b^{1*}(s) = \frac{B_0^{1*}(s)}{s} \qquad \text{and} \qquad \mu_b^{2*}(s) = \frac{B_0^{2*}(s)}{s} \tag{23-24}$$

6.4. Profit Function Analysis

The net expected total cost incurred in time interval (0, t) is given by

P (t) = Expected total revenue in (0, t) - Expected cost of repair in (0, t)

$$= K_0\mu_{up}(t) - K_1\mu_b^1(t) - K_2\mu_b^2(t) \tag{25}$$

Where, K_0 is the revenue per- unit up time by the system during its operation. K_1 and K_2 are the amounts paid to the repairman per-unit of time when the system is busy in repair of unit-1 failed due first repair with priority of unit-1 and unit-2 failed due to second repair respectively.

The expected total profit incurred in unit interval of time is $P = K_0A_0 - K_1B_0^1 - K_2B_0^2$

7. Particular Case

Let $G(t) = \lambda e^{-\lambda t}, \qquad H(t) = \mu e^{-\mu t}$

In view of above, the changed values of transition probabilities and mean sojourn times.

$$p_{10} = \frac{\beta_1}{\beta_1 + \lambda}, \qquad p_{20} = \frac{\beta_2}{\beta_2 + \mu}, \qquad p_{24} = \frac{\mu}{\beta_2 + \mu}$$

$$p_{12}^{(3)} = \frac{\lambda}{\beta_1 + \lambda}, \qquad \psi_1 = \frac{1}{\beta_1 + \lambda}, \qquad \psi_2 = \frac{1}{\beta_2 + \mu}$$

8. Graphical Study of Behaviour and Conclusions

For a more clear view of the behaviour of system characteristics with respect to the various parameters involved, we plot curves for MTSF and profit function in Fig. 2 and Fig. 3 w.r.t. α_1 for three different values of failure parameter α_2 =0.1, 0.5, 0.9 and two different values of repair parameter β_1 =0.01, 0.7 while the other parameters are β_2 =0.99, μ = 0.01, λ = 0.06. It is clearly observed from Fig. 2 that MTSF increases uniformly as the value of α_2 and β_1 increase and it decrease with the increase in α_1. Further, to achieve MTSF at least 10 units we conclude for smooth curves that the values of α_1 must be less than 0.18, 0.29 and 0.49 respectively for α_2 =0.1, 0.5, 0.9 when β_1 =0.01. Whereas from dotted curves we conclude that the values of α_1 must be less than 0.15, 0.22 and 0.39 for α_2 =0.1, 0.5, 0.9 when β_1 =0.7.

Similarly, Fig.3 reveals the variations in profit (P) with respect to α_1 for three different values of α_2 = 0.4, 0.6, 0.8 and two different values of β_1 =0.03, 0.2, when the values of other parameters β_2 =0.01, μ = 0.09, λ = 0.6, K_0=80, K_1=125 and K_2=175. Here also the same trends in respect of α_1, α_2 and β_1 are observed in case of MTSF. Moreover, we conclude from the smooth curves that the system is profitable only if α_1 is less than 0.20, 0.39 and 0.79 respectively for α_2 = 0.4, 0.6, 0.8 when

β_1 =0.03. From dotted curves, we conclude that the system is profitable only if α_1 is less than 0.13, 0.27 and 0.58 respectively for α_2 = 0.4, 0.6, 0.8 when β_1 =0.2.

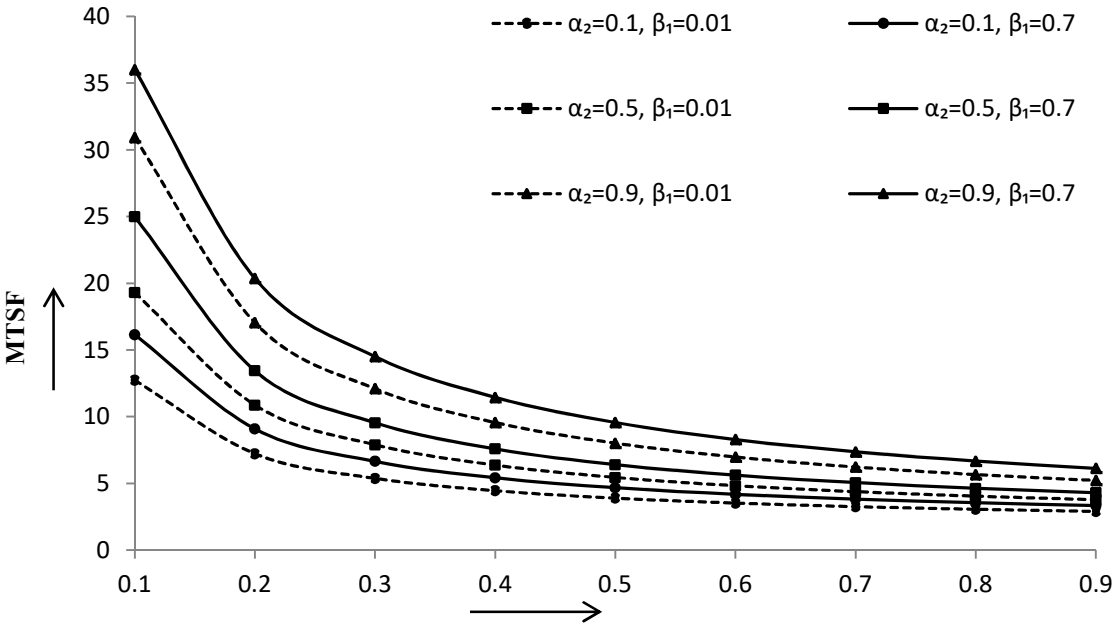


Figure 2: *Behaviour of MTSF w.r.t.* α_1 *for different values of* α_2 *and* β_1

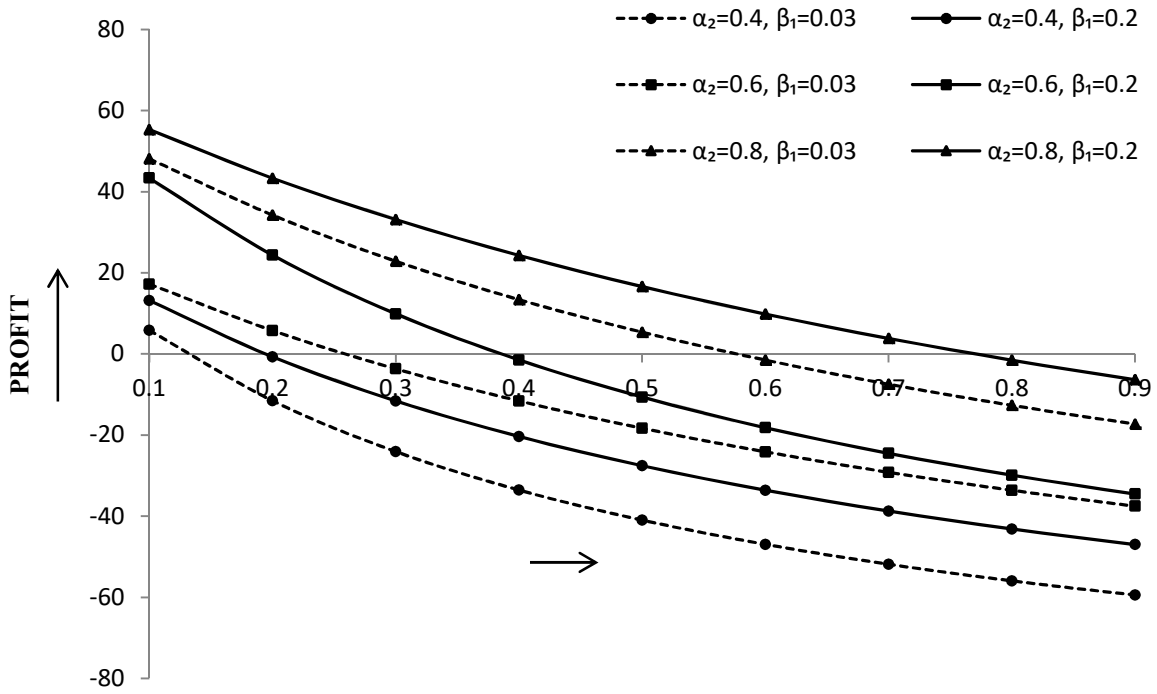


Figure 3: *Behaviour of* ***PROFIT (P)*** *w.r.t.* α_1 *for different values of* α_2 *and* β_1

References

[1] Chaudhary, P. and Masih, S. (2020). Analysis of a standby system with three modes of priority unit and correlated failure and repair times. *Statistics and Reliability Engineering*, 7(1):123-130.

[2] Chaudhary, P. and Sharma, A. (2022). A two non-identical unit parallel system with priority in repair and correlated life times. *Reliability: Theory & Applications*, 1(67):113-122

[3] Chaudhary, P. and Tyagi, L. (2021). A two-non-identical unit parallel system subject to two types of failure and correlated lifetimes. *Reliability: Theory & Applications*, 2(62):247-258.

[4] Chaudhary, P., Masih, S. and Gupta, R. (2022). Parallel system with repair and post repair policies of a failed unit and correlated life times. *Reliability: Theory & Applications*, 3(69):42-51.

[5] Chaudhary, P., Sharma, A. and Gupta, R. (2022). A Discrete Parametric markov chain model of a two non-identical unit warm standby repairable system with two types of failure. *Reliability: Theory & Applications*, 2(68):21-30.

[6] Dabas, N., Rathee, R. and Sheoran, A. (2023). Reliability Analysis of Parallel system using priority to PM over inspection. *Reliability: Theory & Applications*, 1(72):329-339.

[7] Pundir, P.S., Patawa, R. and Gupta, P. K. (2018). Stochastic Outlook of two non-identical unit parallel system with priority in repair. *Cogent Mathematics and Statistics*, 5(1467208): 1-18.

[8] Saini, M., Devi, K. and Kumar, A. (2021). Stochastic Modelling of a two non-identical redundant system with priority in repair activities. *Thailand Statistician*, 19(1):154-161.

[9] Saxena, V., Gupta, R. and Singh, B. (2023). Classical and Bayesian stochastic analysis of a two unit parallel system with working and rest time of repairman. *Reliability: Theory & Applications*, 1(72):43-55.

DEEP LEARNING APPROACH FOR EVENT RECOGNITION IN FIELD HOCKEY VIDEOS

Suhas H. Patel[1*], Dr. Dipesh Kamdar[2*], Dr. D. D. Vyas[3], Dr. Prakash P. Patel[4]
[1]Research Scholar, Gujarat Technological University, Gujarat, India
E-mail: suhas048@gmail.com
[2]Assistant Professor, Electronics and Communication Engg. Department, V.V.P. Engineering College, Rajkot, Gujarat, India
E-mail: kamdardipesh@gmail.com
[3]Dean, Transformative Academics, Atmiya University, Rajkot, Gujarat, India
E-mail: vyasdd@yahoo.com
[4]Principal, V.P.M.P Polytechnic, Gandhinagar, Gujarat, India
E-mail: ecprakashpatel@gmail.com

Abstract

The objectives of this research are to develop a deep learning approach for event recognition in field hockey videos, construct a dataset that includes important activities in field hockey such as goals, penalty corners, and penalty, and evaluate the performance of the approach using the constructed dataset. By achieving these objectives, the research aims to improve the accuracy and effectiveness of event recognition in the fast-paced and complex domain of field hockey videos. The methods employed in this research involve utilizing a pretrained convolutional neural network (CNN) to train a classifier specifically designed for event recognition in field hockey videos. To facilitate this process, a dataset is constructed, consisting of labeled instances of key activities in field hockey, namely goals, penalty corners, and penalty. The performance of the approach is then evaluated using this carefully prepared dataset, providing insights into the effectiveness and accuracy of the proposed method for event recognition in the context of field hockey videos. The findings of this research reveal that the proposed deep learning approach for event recognition in field hockey videos achieves a remarkable accuracy of 99.47%. This high level of accuracy highlights the effectiveness of the approach in accurately identifying and classifying events in field hockey. Furthermore, the results demonstrate the potential of this approach in various field hockey applications, including performance analysis, coaching, and video replay. The accurate recognition of events opens new possibilities for leveraging field hockey videos for enhanced analysis, coaching strategies, and engaging video presentations. The novelty of this research lies in the introduction of a deep learning approach specifically designed for event recognition in field hockey videos. Unlike traditional methods, this approach leverages the power of deep learning, particularly a pretrained CNN, to improve the accuracy of event recognition. Additionally, the construction of a domain-specific dataset addresses the limitation of existing field hockey datasets and enhances the effectiveness of the approach. The remarkable accuracy achieved in event recognition further emphasizes the novelty and potential of this approach in the field of field hockey video analysis.

Keywords: Event recognition, field hockey videos, deep learning, convolutional neural network (CNN), VGG16.

1. INTRODUCTION

Field hockey is a fast-paced and dynamic sport, requiring players to showcase their skills in a highly competitive environment. Event recognition in field hockey videos is a crucial task for extracting valuable insights from the gameplay, enabling performance analysis, coaching, and video replay. Latest computer vision techniques applied in sports video analysis, encompassing player and ball

tracking, trajectory prediction, skill analysis, team strategy assessment, and object detection and classification in sports [1]. Traditional methods for event recognition, such as rule-based systems and motion analysis, often fall short in accurately identifying events due to the sport's complex nature and rapid movements. Some research has prioritized semantic event detection for its capacity to generate insightful outcomes, including pattern recognition and team strategy analysis, while combining video processing, computer vision, and machine learning[2]. This integration offers substantial potential in the sports entertainment domain, enhancing referee decision-making and providing sports fans with improved systems for match analysis. Deep learning techniques have shown significant advancements in event recognition across different sports, offering promising results. One crucial aspect for effective event recognition is the ability of the network to learn high-level features that capture human actions and the contextual scene information[3]. This requires two key factors: Adequate input image size and Network Depth. The input image size should be sufficiently large to enable the network to capture fine-grained details and extract meaningful features related to the sports events. Deep neural networks are essential for learning complex and abstract representations from the input data. A deeper network architecture allows for the extraction of hierarchical features, leading to improved event recognition performance. By incorporating large input image sizes and deep network architectures, deep learning models can effectively learn high-level features that facilitate accurate event recognition in sports. These advancements have contributed to significant improvements in event recognition across various sports domains. Combining convolutional and recurrent neural networks enables the analysis of sports video sequences and yields experimental results[4]. However, sports analytics face challenges and unresolved issues, especially in data collection and labeling, as well as the complexity of recognizing fast actions and analyzing multiple players' involvement in team sports like football and basketball[5]. In this research paper, we present a deep learning-based approach for event recognition in field hockey videos. Our approach leverages the power of pretrained convolutional neural networks (CNNs) to train a classifier capable of accurately identifying different events. By exploiting the knowledge learned from large-scale datasets in related domains, the pretrained CNN captures rich visual representations that are crucial for distinguishing between various field hockey events. One of the challenges we encountered was the lack of existing field hockey datasets suitable for event recognition. To address this, we created our own dataset, specifically designed for field hockey, consisting of three primary activities: goals, penalty corners, and penalty. This dataset enables us to train and evaluate the performance of our deep learning approach in a realistic field hockey scenario. Through extensive evaluations, we demonstrate the effectiveness of our approach in accurately recognizing events in field hockey videos.

Our approach achieves an impressive accuracy of 99.47% on our self-prepared dataset, highlighting its potential for real-world applications in field hockey analysis and coaching. The contributions of this research paper extend beyond event recognition in field hockey. The remaining sections of the paper are structured as follows. Section 2 presents an overview of related work, focusing on event recognition and the application of deep learning in sports. Section 3 outlines the methodology employed for event recognition in field hockey videos. The results and analysis are presented in Section 4. Lastly, Section 5 concludes the paper by summarizing our contributions and highlighting future research directions.

2. RELATED WORK

Event recognition in sports videos has been a topic of extensive research in recent years. Various approaches have been explored to tackle the challenges associated with accurately identifying events in dynamic sporting environments. In this section, we provide an overview of related work in the field of event recognition and highlight the contributions of deep learning methods in sports

analysis.

2.1 Traditional Approaches for Event Recognition

Traditional approaches for event recognition in sports videos often rely on handcrafted features and rule-based systems. These methods involve manually designing features based on domain knowledge and utilizing predefined rules to detect specific events. For field hockey, these rules might include analyzing the positions and movements of players, the trajectory of the ball, or specific gameplay patterns. While these approaches have been effective to some extent, they often struggle to handle the complexity and variability of events in dynamic sports like field hockey. The manual design of features and rules limits their adaptability to different scenarios and may result in suboptimal performance.

In a series of research papers, traditional approaches have been proposed for activity recognition and detection in sport videos. One study introduces Histograms of Oriented Gradients (HOG) for player representation and combines it with a probabilistic framework and multi-class sparse classifier for action recognition [6]. Another paper focuses on evaluating action recognition approaches for fight detection, presenting a new fight dataset, and achieving high accuracy in detecting fights[7]. Hierarchical poselets are introduced in another study, enabling human pose modeling and serving as an intermediate representation for action recognition[8]. A violence detection method utilizing the MoSIFT algorithm and sparse coding is proposed in a different research, addressing the limitations of traditional descriptors and achieving promising results on challenging datasets[9]. Additionally, a novel approach using the Markov Game formalism is presented to value player actions in ice hockey, considering context and lookahead [10]. Lagrangian measures are employed in another paper for violent video detection, outperforming other local features in detecting violence [11]. Lastly, a technique using histogram of oriented gradients and local binary pattern features is presented for accurate recognition of basketball referees' signals in game videos[12]. These studies collectively contribute to the field of video-based activity recognition, offering insights and advancements in various aspects such as player representation, violence detection, pose modeling, action valuation, and gesture recognition in sports videos. Table 1 provides a comprehensive list of traditional sport event detection models.

Table 1: *Traditional event detection models for various sport categories.*

Reference	Problem statement	Proposed method	Sports
[6]	Track and identify the actions of multiple hockey players.	Histograms of Oriented Gradients (HOG), boosted particle filter (BPF)	Ice Hockey
[7]	detection of fights or aggressive behaviors in ice hockey sport videos	Space-Time Interest Points (STIP), Motion Scale-Invariant Feature Transform (MoSIFT)	Ice Hockey
[8]	human parsing and action recognition from static images	hierarchical Poselets	Multiple sports
[9]	detect violence in videos with crowded and non-crowded scenes	MoSIFT and sparse coding.	Ice Hockey
[10]	Assessing player actions in ice hockey.	Markov Game formalism	Ice Hockey

[11]	to detect violent scenes in videos	Lagrangian Scale Invariant Feature Transform (LaSIFT)	Ice Hockey
[12]	recognize the signals of basketball referees from recorded game videos.	histogram of oriented gradients +SVM, local binary pattern features +SVM	Basketball

2.2 Deep Learning in Event Recognition in Sports

Deep learning has emerged as a powerful paradigm for event recognition in sports videos. Convolutional Neural Networks (CNNs) have demonstrated remarkable success in extracting meaningful representations from visual data. CNNs can automatically learn hierarchical features by employing multiple layers of convolutions and nonlinear activations, enabling them to capture complex patterns and spatial dependencies.

In the context of sports event recognition, deep learning models have shown superior performance by leveraging large-scale annotated datasets and pretraining on related tasks. By utilizing pretrained CNNs, such as those trained on ImageNet, the models can capture generic visual representations that are transferable to sports-specific tasks. Fine-tuning or retraining the pretrained models on specific sports datasets further enhances their ability to recognize events accurately.

One paper proposes a novel framework for soccer video event detection, utilizing 3D convolutional networks and shot boundary detection, and introducing temporal action localization and play-break rules [13]. Another study focuses on action recognition in hockey, introducing the ARHN architecture and achieving high accuracy by leveraging pose information[14]. A methodology for fight scene detection in hockey videos is proposed in a different paper, using blur, radon transform, and convolutional neural networks [15]. Furthermore, a 3D CNN-based multilabel deep HAR system is presented for hockey video action recognition, outperforming existing solutions[16]. Another paper introduces a two-stream architecture for hockey action recognition, combining pose estimation and optical flow[17]. A modified 3D ConvNet is proposed for violent video detection, achieving competitive results with improved strategies[18]. An automated activity recognition model for hockey matches using deep learning is presented, achieving a high accuracy of 98% [19]. In cricket, a hybrid deep-neural-network architecture is proposed for shot classification [20]. Puck localization in hockey videos is addressed using a network that incorporates expert annotations and temporal context [21]. Lastly, a deep learning method for event detection in football videos achieves superior precision and recall [22]. These papers collectively contribute to advancing activity recognition in sports videos, offering novel frameworks, architectures, and methodologies for accurate and efficient detection and recognition of various actions and events. Table 2 lists various sport event detection models based on deep learning.

Table 2: *Deep learning-based event detection model for various sport categories.*

Reference	Problem statement	Proposed method	Sports
[13]	Soccer video event detection	3D Convolutional Networks +Deep Feature Distance	Soccer
[14]	Interpreting player actions in ice hockey videos	Action Recognition Hourglass Network (ARHN)	Ice Hockey
[15]	Detecting fight scenes in hockey sport videos	Feed forward neural network and VGG16-Net	Ice Hockey
[16]	Multi-label class-imbalanced action recognition in hockey videos.	3D CNN based multilabel deep HAR system	Ice Hockey
[17]	Action recognition in ice hockey	two-stream architecture	Ice

			Hockey
[18]	violent video detection in ice hockey	Modified 3D ConvNet	Ice Hockey
[19]	Field hockey activity recognition	VGG-16	Field Hockey
[20]	The task involved classifying 10 different cricket batting shots from offline videos.	CNN+GRU	Cricket
[21]	Puck localization and event recognition in broadcast hockey videos,	CNN	Ice Hockey
[22]	Event detection in football videos	InceptionV2, 3DCNN	Soccer

Deep learning methods have been successfully applied to event recognition in various sports, including soccer, basketball, tennis, and cricket. These approaches often involve preprocessing video frames, extracting visual features using pretrained CNNs, and employing classifiers to recognize specific events. In the domain of field hockey, however, there is limited research on event recognition using deep learning methods. Our work aims to bridge this gap by proposing a deep learning-based approach specifically designed for field hockey event recognition. By leveraging the power of pretrained CNNs, we aim to overcome the challenges associated with accurately identifying events in the fast-paced and complex nature of field hockey gameplay.

In summary, while traditional approaches for event recognition in sports videos have been explored, deep learning methods have shown significant promise in improving event recognition accuracy. The utilization of pretrained CNNs and transfer learning techniques enables these models to learn rich visual representations and adapt to specific sports domains. In the case of field hockey, there is a need for further research and development of deep learning-based approaches tailored to the unique characteristics of the sport. Our proposed approach aims to address this gap and contribute to the advancement of event recognition in field hockey videos.

3. METHODOLOGY

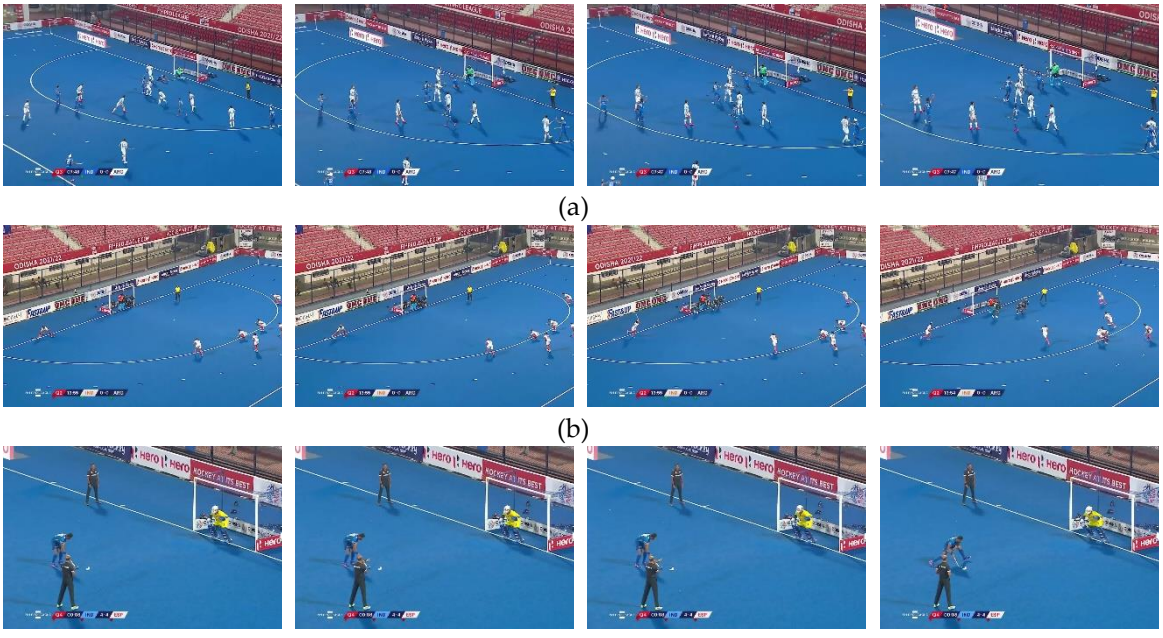
3.1 Field Hockey Dataset

As there is a lack of publicly available field hockey datasets for event recognition, we constructed our own dataset specifically tailored to the sport. We analyzed a collection of 28 highlights videos from the tournaments of the hockey pro league for the years 2021-22 and 2022-23. These videos showcase the most remarkable moments and thrilling gameplay sequences from various field hockey games played between two teams. By carefully analyzing these highlights videos, we were able to identify and extract important events such as goals, penalty corners, and penalty. These events represent significant turning points in the matches and provide valuable data for building a comprehensive hockey event detection dataset. The utilization of highlights videos ensures that the dataset captures the most exciting and impactful moments from the field hockey games. This enables the development and evaluation of event detection models on key events that greatly influence the outcome of matches and attract the attention of viewers. By leveraging the information extracted from these videos, we aim to construct a high-quality field hockey event detection dataset that can be used for training and evaluating event recognition models. To create the ground truth annotations, we meticulously watched and manually labeled the videos. Each event of interest, such as goals, penalty corners, penalty, and other relevant actions, was annotated with their respective start and end timestamps The dataset is designed to be representative of the challenges encountered

in real-world field hockey scenarios. It includes enough positive samples for each event category, ensuring a balanced distribution for training and evaluation purposes. Figure 1 illustrates the sequential video frames used for hockey event recognition.

Table 3: *Hockey Event Recognition Dataset*

Total Images	3035
Classes	3
Unannotated	0
Training Set	2276 (75%)
Testing Set	759 (25%)
Average Image Size	2.07 mp
Median Image Ration	1920x1080
Class Instances	
Goal	1000(32.95%)
Penalty Corner	1017(33.51%)
Penalty	1018(33.54%)



(a) (b) (c)

Figure 1: *Sequential video frames for (a) Goal, (b) Penalty Corner, (c) Penalty*

3.2 Model Architecture

In our approach, we utilize a pretrained convolutional neural network (CNN) for feature extraction, as depicted in Figure 2. The model architecture consists of a CNN that has been pretrained on a large-scale image dataset, such as ImageNet, to learn generic visual representations. This pretrained CNN can capture low-level to high-level visual features, making it well-suited for recognizing complex events in field hockey videos. The process begins with the application of the "flatten" operation, which reshapes the output of the pretrained CNN into a one-dimensional vector. This step enables easier processing and subsequent layers in the model. Following the flattening, a "dense512" layer is introduced. This layer is a fully connected layer with 512 units and applies the

rectified linear unit (ReLU) activation function to introduce non-linearity to the network. To mitigate the risk of overfitting, a "dropout(0.5)" layer is included. During training, this layer randomly sets 50% of the values to 0, effectively disabling certain connections between neurons. By doing so, it helps prevent the model from relying too heavily on specific features and enhances its generalization capability. The subsequent layer in the model is a "dense(3)" layer, which is used for multi-class classification. This layer consists of 3 units, representing the number of classes, and applies the softmax activation function to produce class probabilities. It is through this layer that the model assigns probabilities to each class, indicating the likelihood of the input frame belonging to a particular class. To summarize, the model architecture encompasses a pretrained CNN base, the flattening operation, a dense layer with ReLU activation, dropout regularization, and a dense layer with softmax activation for classification. These operations constitute a deep learning model configuration commonly employed in classification tasks. Moreover, Figure 8 provides an overview of the model architecture and its components. During the inference stage, each video frame is passed through the pretrained CNN, and activations from one of the intermediate layers are extracted. These activations represent the specific visual features learned by the model for that frame. By considering multiple frames within a temporal window, we capture the temporal dynamics of events, enabling a comprehensive understanding of the evolving actions in the field hockey videos.

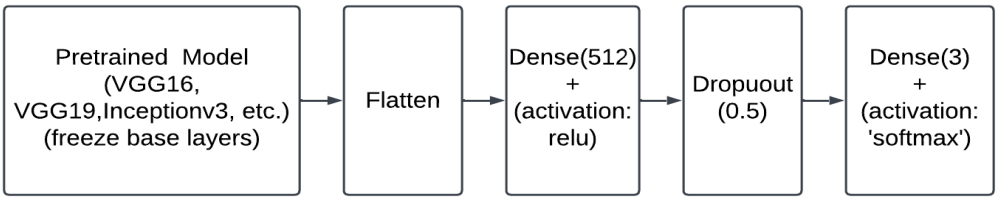


Figure 2: *Overview of different pretrained model architectures*

3.3 Model Training and Evaluation

To train our deep learning-based event recognition system, we split the annotated dataset into training and test sets. The dataset used in this study, as shown in Table 3, consists of a total of 3,035 images that are classified into three distinct classes. All images in the dataset have been fully annotated, ensuring that there are no unannotated instances. The training set comprises 2,276 images, which corresponds to 75% of the total dataset, while the remaining 759 images form the testing set, accounting for 25% of the dataset. On average, the images have a size of approximately 2.07 megapixels. The median image resolution is 1920x1080 pixels, indicating a consistent aspect ratio among the images. The class distribution within the dataset is as follows: The "Goal" class consists of 1,000 instances, representing 32.95% of the dataset. The "Penalty Corner" class comprises 1,017 instances, accounting for 33.51% of the dataset. Lastly, the "Penalty" class contains 1,018 instances, making up 33.54% of the dataset. The images are resized to a dimension of 224x224 pixels. The VGG16 model is pretrained on a dataset like ImageNet to learn features and predict labels. To adapt it for a new task, the top layers are replaced, and the base layers are frozen as a feature extractor. New layers are added on top, such as fully connected and dropout layers. The modified VGG16 model is then trained on the new dataset, optimizing its parameters using the training set. Cross-validation techniques can be used for robustness. The system's performance is evaluated on the test set, measuring event recognition accuracy and other metrics. Results are compared with baseline methods or alternative architectures to assess the effectiveness of the approach. The following pretrained models from Keras were utilized in this study, as shown in Table 4.

The event detection system is implemented on Google Colaboratory, which is a Python 3 environment, utilizing the GPU support provided by Google Compute Engine backend. In this

study, a pre-trained VGG16 based model-1 was utilized for event detection. The input images were resized to 224x224 pixels, and a batch size of 32 was used during training.

Table 4 : *Overview of Pretrained Networks*

Reference	Model	Description
[23]	VGG16	A deep convolutional neural network (CNN) with 16 layers, known for its simplicity and effectiveness.
[23]	VGG19	Like VGG16 but with 19 layers, providing a slightly deeper architecture.
[24]	ResNet50	A deep residual network with 50 layers, designed to address the vanishing gradient problem and enable training of very deep networks.
[25]	InceptionV3	A deep CNN architecture with multiple parallel branches, allowing for efficient feature extraction at different scales
[26]	MobileNet	A lightweight CNN architecture designed for mobile and embedded devices, balancing model size and accuracy.
[27]	DenseNet121	A densely connected CNN architecture that facilitates feature reuse and enables deeper networks without sacrificing performance.
[28]	Xception	An extension of the Inception architecture that replaces standard convolutions with depthwise separable convolutions, resulting in improved performance.

The model is trained for 100 epochs using Stochastic Gradient Descent (SGD) optimizer with specific parameters including a learning rate of 1e-4, momentum of 0.9, and a decay of 1e-4/100. The video frames representing hockey events are used as inputs to the fine-tuned VGG16 based model-1 that was specifically tailored for event detection in this study. Figure 3 illustrates the architecture of the proposed pretrained model framework.

Table 5 displays the modified model details, including the modules, output dimensions, and trainable parameters. The flowchart in Figure 5 illustrates the process of hockey event recognition using a deep learning model. It encompasses preprocessing the input video, training the model, evaluating its performance, and utilizing the model to predict events in a video clip. The output is a labeled video where events are assigned specific labels based on the model's predictions.

We evaluate our event recognition system using standard evaluation metrics, including accuracy, precision, recall, and F1 score, to measure its performance. Accuracy represents the proportion of correctly classified events, while precision, recall, and F1 score assess the system's performance in identifying specific event categories. After training, we assess the system's accuracy, precision, recall, and F1 score on the test set and compare our results with alternative deep learning architectures. Table 6 presents the results of seven different architectures for hockey event detection on dataset, demonstrating the effectiveness and performance improvements of our proposed approach. The experimental results are visualized in Figure 6.

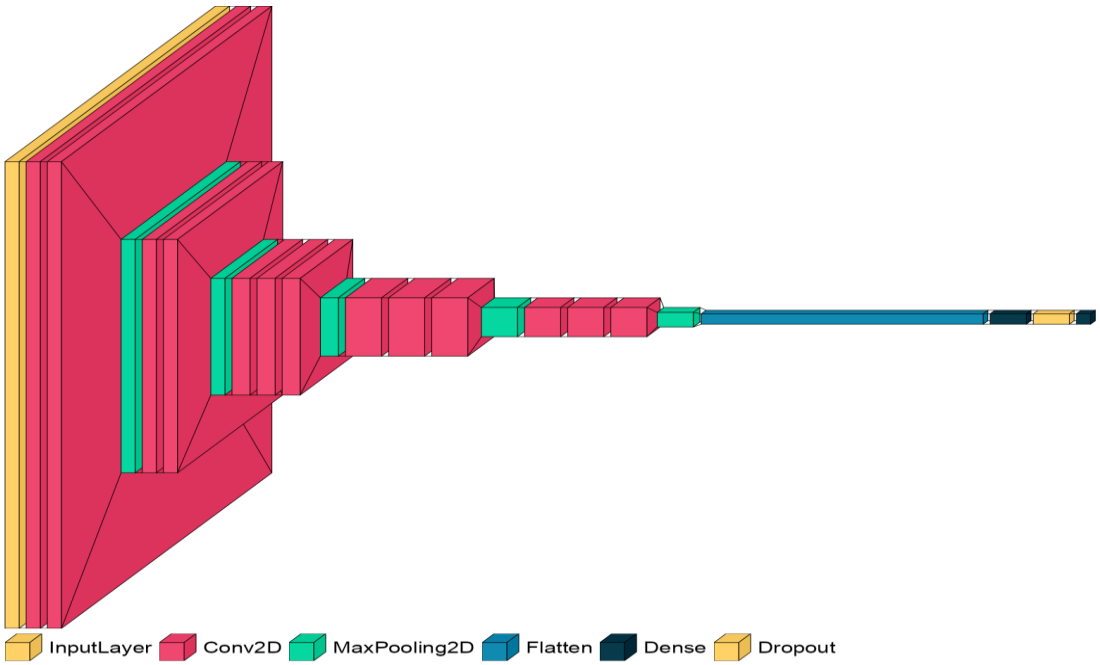


Figure 3: *Model-1 Architecture.*

Table 5: *Model -1 details in terms of module, output dimension and trainable parameters.*

Model: "model-1"

Layer (type)	**Output Shape**	**Param #**
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 512)	12845568
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 3)	1539
Total params: 27,561,795		
Trainable params: 12,847,107		
Non-trainable params: 14,714,688		

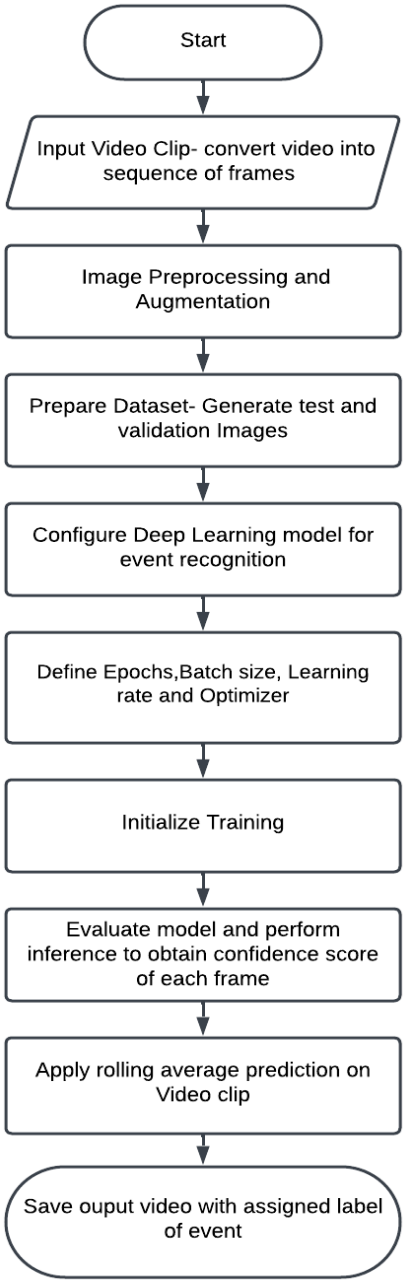


Figure 5 : *Process of hockey event recognition using a deep learning mode*

Table 6 : *Fine-tuned Deep Learning model results.*

Sr no.	Model Name	Pre-trained Network	Trainable Parameters	Precision (%)	Recall (%)	F1 Score (%)	Accuracy (%)
1	Model-1	VGG16	12,847,107	99.33	99.33	99.33	99.47
2	Model-2	VGG19	264,195	97.67	97.67	97.33	97.50
3	Model-3	ResNet50	1,050,627	96.33	96.33	96.33	96.44
4	Model-4	InceptionV3	4,196,355	84.67	83.67	84.00	83.66
5	Model-5	MobileNet	526,339	88.33	87.67	87.67	87.62
6	Model-6	DenseNet12	526,339	86.00	81.67	81.67	81.69
7	Model-7	Xception	1,050,627	76.33	74.00	74.00	74.44

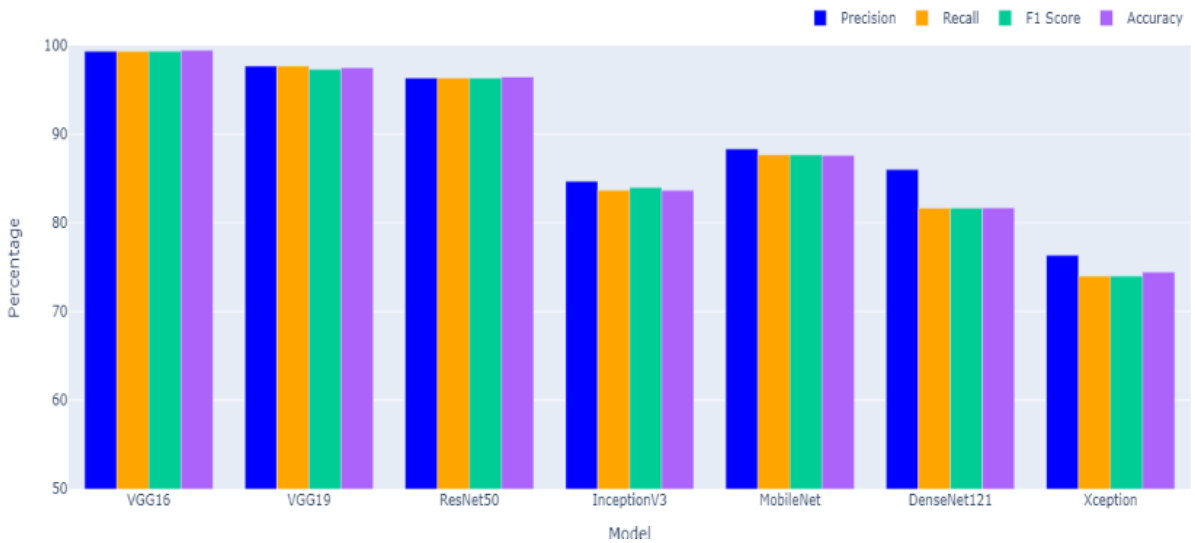


Figure 6: *Comparison of Model-1 with other six models*

Figure 7 illustrates the flowcharts that enable us to utilize the power of convolutional neural networks (CNNs) in analyzing video frames and making predictions for each frame. The flowchart starts with inputting the video frames into the CNN model. As the frames are processed through the model, predictions are generated for each frame, indicating the likelihood of a particular event occurring at that specific moment.

To enhance the stability and reliability of the predictions, we apply a rolling average technique. This involves averaging the recent predictions over a certain period or a specific number of frames. By incorporating information from multiple frames, we can mitigate the impact of temporary variations or noise in individual frame predictions, resulting in a more robust and consistent prediction for the event happening in the video. The rolling average prediction approach helps to smooth out any fluctuations or inconsistencies in the frame-level predictions, providing a more accurate estimation of the event occurring in the video at any given time[29]. This can be particularly beneficial when dealing with real-world scenarios where videos may contain motion blur, camera movement, or other factors that can introduce uncertainties in individual frame predictions. Overall, the flowcharts in figure 7, combined with the rolling average prediction technique, enable us to leverage the CNN's analytical capabilities to make reliable and stable predictions for the events taking place in the video, improving the overall accuracy and effectiveness of our event recognition system.

4. Results and Analysis

4.1 Results

The model-1, with 12,847,107 trainable parameters, demonstrates exceptional performance in image classification. It achieves high precision, recall, F1 score, and accuracy, all at approximately 99.47%. These metrics indicate that the model excels in accurately classifying images with a high level of precision and recall.

Upon analyzing the results, it is evident that the model-1 outperforms other models across all evaluated metrics. It achieves the highest precision, recall, F1 score, and accuracy among the models considered. While model-2 and model-3 also show commendable performance, models such as model-3, model-4, model-5, and model-6 exhibit relatively lower performance in comparison. These findings highlight the effectiveness of the VGG16 based model-1 for image classification tasks, showcasing its superiority in accurately classifying images. Figure 8 illustrates the training and loss

accuracy of the VGG16 based model-1 up to 100 epochs, showcasing the model's learning progression over the training process. Figure 9 displays the confusion matrix of the proposed model-1 for the given dataset, providing insights into the model's performance in terms of classification accuracy. Figure 10 visually presents the output video frames of hockey event recognition. Our event recognition system, evaluated experimentally, achieved an impressive overall accuracy of 99.47% on the field hockey dataset, indicating its effectiveness in accurately recognizing and classifying field hockey events.

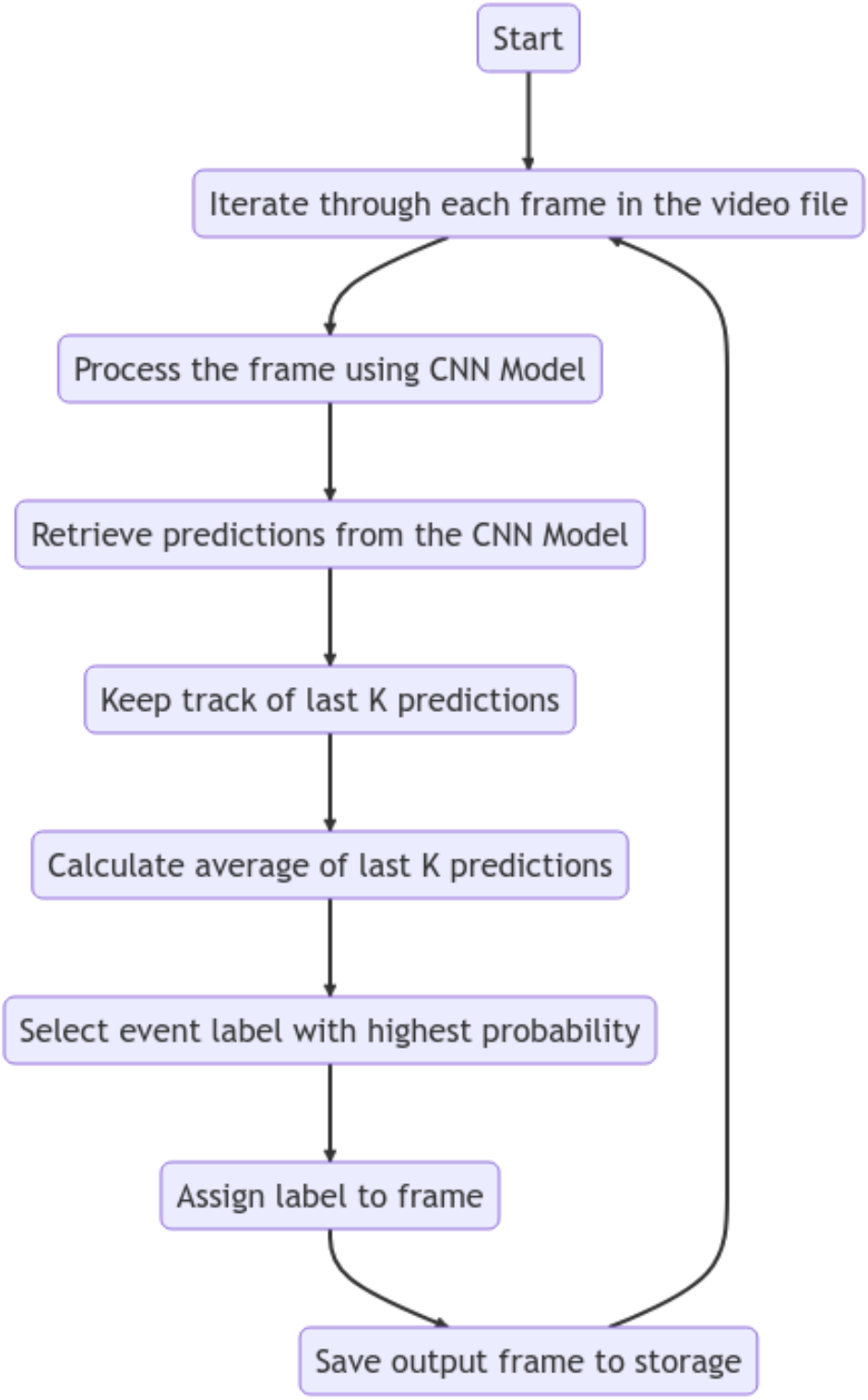


Figure 7 : *Process of rolling average prediction for Event detection*

Table 7: *Hockey event recognition results for model-1.*

Event	Precision	Recall	F1-score	Support
Goal	0.99	0.99	0.99	250
Penalty Corner	1.0	1.0	1.0	254
Penalty	0.99	0.99	0.99	255

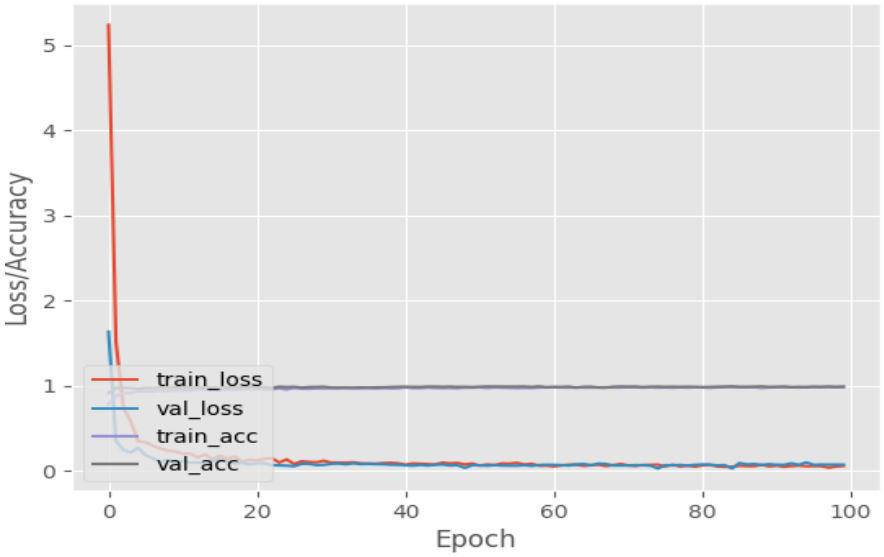


Figure 8: *Training loss and accuracy of proposed model-1.*

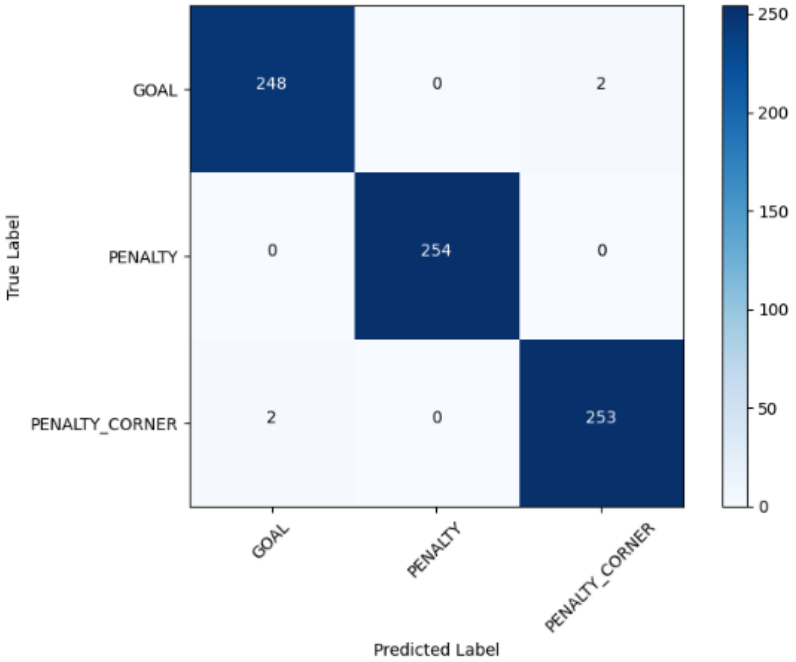


Figure 9: *Confusion Matrix of model-1.*

Table 7 presents the results of event recognition using the model 1. The "Goal" event achieved excellent precision, recall, and F1-score values of 0.99, accurately classifying instances with a support of 250. Similarly, the "Penalty Corner" event demonstrated perfect precision, recall, and F1-score values of 1.0, precisely identifying instances with a support of 254. The "Penalty" event exhibited high precision, recall, and F1-score values of 0.99, correctly identifying instances with a support of 255. These findings highlight the model's strong performance and consistency in accurately

classifying hockey events across all categories.

In addition to achieving impressive results, we conducted comparative evaluations against baseline methods and alternative deep learning architectures. The results consistently demonstrated the superiority of our deep learning-based event recognition system over the baseline methods. This further reinforces the effectiveness of our approach in accurately recognizing field hockey events. In conclusion, our experimental evaluation showcases the effectiveness of our deep learning-based event recognition system for field hockey videos. The achieved accuracy and performance metrics validate its ability to accurately classify events such as goals, penalty corners, and penalty. This highlights the potential of our approach in various applications within the field of hockey, including performance analysis, coaching, and video replay.

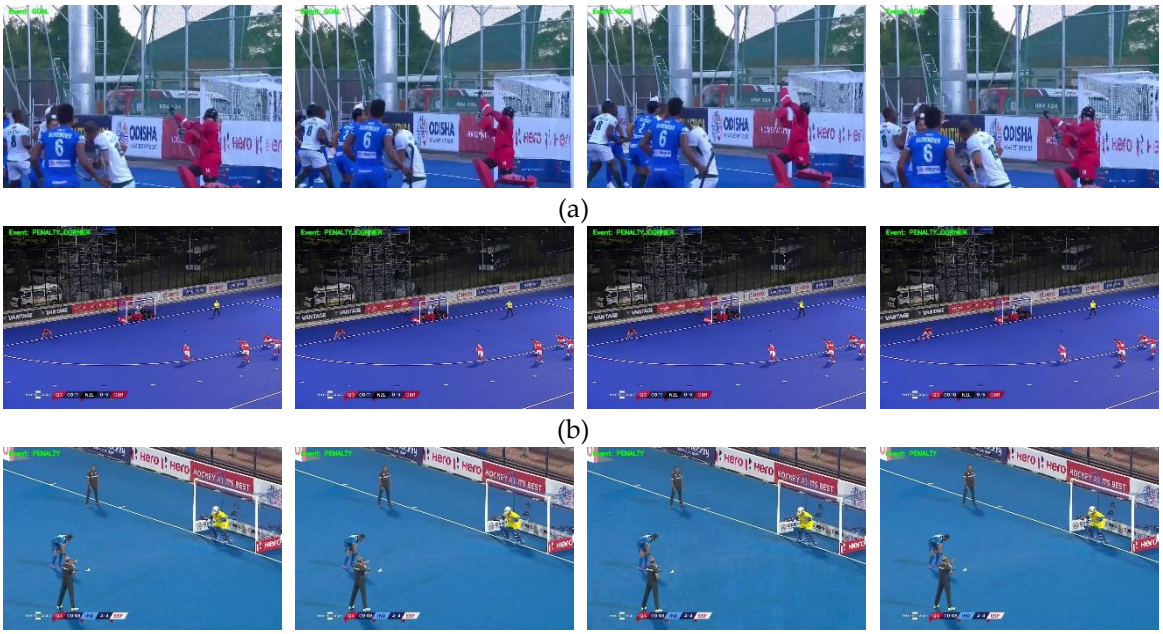


Figure 10. *Hockey event recognition output for model-1 (a) Goal, (b) Penalty Corner, (c) Penalty*

4.2 Discussion of Findings

Our research demonstrates the effectiveness of using a pre-trained deep learning model for recognizing hockey activities in videos. The results indicate a high accuracy of 99.47% in recognizing activities such as goals, penalty corners, and penalty. This highlights the ability of deep learning models to capture and analyze the visual features required for accurate activity recognition in the fast-paced and complex sport of hockey. Utilizing a pre-trained model enables efficient transfer learning, as it leverages knowledge learned from a large-scale dataset. Fine-tuning the pre-trained model on our own dataset yields excellent performance without the need for extensive data collection and training from scratch. Additionally, constructing a domain-specific dataset is crucial for activity recognition. As no existing field hockey datasets were available, we created our own dataset with annotated videos capturing various hockey activities. This dataset serves as a valuable resource for training and evaluating activity recognition models specific to the sport of hockey. In terms of future directions, expanding the dataset to include a larger and more diverse collection of field hockey videos would enhance the models' generalizability. Exploring fine-grained activity recognition, real-time recognition during live matches, and multi-modal fusion can further improve the accuracy and applicability of hockey activity recognition models. Additionally, transferring the developed models and methodologies to other sports with similar characteristics can broaden the

scope of activity recognition research. It is important to consider limitations and challenges in hockey activity recognition using deep learning models, such as dataset bias and occlusion and camera variability. Addressing these challenges and ensuring the robustness and unbiased nature of the models are essential for reliable activity recognition in hockey. Overall, our research contributes to the field of hockey activity recognition, demonstrating the potential of deep learning models in accurately recognizing and classifying hockey activities.

5. CONCLUSION

Our research showcases the exceptional outcomes attained by utilizing a pre-trained deep learning model for hockey activity recognition. Through fine-tuning the model on a meticulously constructed dataset, we accomplished an impressive accuracy rate of 99.47% in accurately classifying various activities such as goals, penalty corners, and penalty. This emphasizes the effectiveness of deep learning models in capturing and analyzing the visual features required for precise activity recognition in the dynamic sport of hockey. The construction of a domain-specific dataset plays a pivotal role in the success of activity recognition models, and our carefully curated dataset of annotated field hockey videos frames serves as a valuable resource for further advancements in this area. The practical implications of our research hold great significance for stakeholders within the hockey domain.

Firstly, accurate activity recognition can provide valuable insights for performance analysis. Coaches and analysts can utilize the recognized activities to evaluate player performance, identify patterns and strategies, and make data-driven decisions to enhance team performance.
Secondly, our approach can support coaching and training activities by automatically identifying and analyzing key events in field hockey videos. Coaches can leverage the insights gained from the system to offer targeted feedback, identify areas for improvement, and develop customized training programs for specific activities.

Lastly, the capability to automatically recognize and classify activities in real-time can enhance the viewing experience for spectators and broadcasters. Instant replays, highlights, and in-depth analysis can be generated using the recognized activities, thereby enriching the storytelling and engagement during hockey matches.

In conclusion, our research demonstrates the effectiveness of utilizing a pre-trained deep learning model for hockey activity recognition. We have presented a comprehensive evaluation of our approach, achieving exceptional accuracy in classifying hockey activities. The construction of a domain-specific dataset further reinforces the reliability and applicability of our findings. Although our research has provided valuable insights and practical implications, there are still avenues for future exploration and improvement. Further work can be conducted to expand the dataset, explore fine-grained activity recognition, enable real-time recognition, and investigate multi-modal fusion approaches.

Overall, our research contributes to the field of hockey activity recognition and lays the groundwork for further advancements in analyzing and comprehending the intricate dynamics of field hockey. We hope that our work serves as an inspiration for future research and applications in this domain, ultimately benefiting players, coaches, analysts, and hockey enthusiasts.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] B. T. Naik, M. F. Hashmi, and N. D. Bokde. (2022). A Comprehensive Review of Computer Vision in Sports: Open Issues. Future Trends and Research Directions, Appl. Sci., 12(9):4429

[2] S. F. De Sousa Júnior, A. De A. Araújo, and D. Menotti. (2011). An overview of automatic event detection in soccer matches. 2011 IEEE Workshop on Applications of Computer Vision(WACV),31–38.

[3] M. A. Russo, A. Filonenko, and K. H. Jo. (2018). Sports Classification in Sequential Frames Using CNN and RNN. 2018 Int. Conf. Inf. Commun. Technol. Robot. (ICT-ROBOT), 1-3.

[4] M. A. Russo, L. Kurnianggoro, and K.-H. Jo. (2019). Classification of sports videos with combination of deep learning models and transfer learning. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1–5.

[5] Wu, Fei and Wang, Qingzhong and Bian, Jiang and Ding, Ning and Lu, Feixiang and Cheng, Jun and Dou, Dejing and Xiong, Haoyi. (2022). A Survey on Video Action Recognition in Sports: Datasets, Methods and Applications. IEEE Transactions on Multimedia,1–25.

[6] W.-L. Lu, K. Okuma, and J. J. Little(2009), Tracking and recognizing actions of multiple hockey players using the boosted particle filter. Image and Vision Computing,27(1):189–205.

[7] E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, and R. Sukthankar.(2011). Violence Detection in Video Using Computer Vision Techniques. Computer Analysis of Images and Patterns,6855: 332–339.

[8] Y. Wang, D. Tran, Z. Liao, and D. Forsyth. (2017), Discriminative Hierarchical Part-Based Models for Human Parsing and Action Recognition. Journal of Machine Learning Research, 13(10): 273–301.

[9] Xu, C. Gong, J. Yang, Q. Wu, and L. Yao. (2014). Violent video detection based on MoSIFT feature and sparse coding. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 3538–3542.

[10] K. Routley and O. Schulte. (2015). A Markov Game model for valuing player actions in ice Hockey. Uncertain. Artif. Intell. - Proc. 31st Conf. UAI 2015, 782–791.

[11] T. Senst, V. Eiselein, and T. Sikora,. (2015). A local feature based on lagrangian measures for violent video classification. IET Semin. Dig.,1-6.

[12] J. Žemgulys, V. Raudonis, R. Maskeliūnas, and R. Damaševičius. (2020). Recognition of basketball referee signals from real-time videos. J. Ambient Intell. Humaniz. Comput., 11(3) :979–991.

[13] Liu, T., Lu, Y., Lei, X., Zhang, L., Wang, H., Huang, W., & Wang, Z. (2017). Soccer Video Event Detection Using 3D Convolutional Networks and Shot Boundary Detection via Deep Feature Distance. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 10635: 440–449.

[14] M. Fani, H. Neher, D. A. Clausi, A. Wong, and J. Zelek. (2017). Hockey Action Recognition via Integrated Stacked Hourglass Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 85–93.

[15] S. Mukherjee, R. Saini, P. Kumar, P. P. Roy, D. P. Dogra, and B.-G. Kim. (2017). Fight Detection in Hockey Videos using Deep Network. Journal of Multimedia Information System,, 4(4):225–232.

[16] K. Sozykin, S. Protasov, A. Khan, R. Hussain, and J. Lee. (2018). Multi-label class-imbalanced action recognition in hockey videos via 3D convolutional neural networks. Proc. - 2018 IEEE/ACIS 19th Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2018, 146–151.

[17] Z. Cai, H. Neher, K. Vats, D. A. Clausi, and J. Zelek. (2019). Temporal hockey action recognition via pose and optical flows. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.,

2543–2552

[18] W. Song, D. Zhang, X. Zhao, J. Yu, R. Zheng, and A. Wang. (2019). A Novel Violent Video Detection Scheme Based on Modified 3D Convolutional Neural Networks. IEEE Access, 7: 39172–39179

[19] K. Rangasamy, M. A. As'ari, N. A. Rahmad, and N. F. Ghazali. (2020). Hockey activity recognition using pre-trained deep learning model. ICT Express, 6(3):170–174.

[20] A. Sen, K. Deb, P. K. Dhar, and T. Koshiba. (2021). Cricshotclassify: An approach to classifying batting shots from cricket videos using a convolutional neural network and gated recurrent unit. Sensors, 21(8):2846.

[21] K. Vats, M. Fani, D. A. Clausi, and J. Zelek. (2021). Puck localization and multi-task event recognition in broadcast hockey videos. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE,4562–4570.

[22] N. Liu, L. Liu, and Z. Sun. (2022). Football Game Video Analysis Method with Deep Learning. Computational Intelligence and Neuroscience, 2022: 1–12.

[23] K. Simonyan and A. Zisserman. (2015). Very deep convolutional networks for large-scale image recognition. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 1–14.

[24] K. He, X. Zhang, S. Ren, and J. Sun. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 770–778.

[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. (2016). Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2818–2826.

[26] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger.(2017). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261–2269.

[28] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions. (2017). 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),1800–1807.

[29] Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J. A., Orts-Escolano, S., Garcia-Rodriguez, J., & Argyros, A. (2022). A Review on Deep Learning Techniques for Video Prediction. IEEE Trans. Pattern Anal. Mach. Intell., 44 (6) :2806–2826.

RT&A, No.3 (74)
Volume 18, September 2023

RELIABILITY AND PROFITABILITY ANALYSIS OF UTENSILS MANUFACTURING INDUSTRY WITH EFFECT OF TEMPERATURE AND PREVENTIVE MAINTENANCE

Manisha

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
manishagaba887@gmail.com

Dalip Singh

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
dsmdur@gmail.com

Kajal Sachdeva*

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
kajal.rs.maths@mdurohtak.ac.in

Sheetal

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
rtksheetal@gmail.com

*Corresponding Author

Abstract

The production stage of the manufacturing process contains numerous subsystems, and the failure of one might have an impact on the entire system. Thus, a manufacturing plant needs to be reliable and well-maintained. This paper examines the profitability and reliability of a production plant for utensils while taking the effect of temperature into account. The plant processes raw materials through several subsystems in series including cutting, pressing, spinning, and polishing & packing. Winter production requires a significant amount of heat which could damage the machinery. As a result, production is low and preventive maintenance is carried out during the winter. For both the summer and winter seasons, many system measures have been assessed. The time distributions have been assumed to be exponential. The model has been analysed using the Markov and Regenerative processes. The production fluctuation between the summer and winter seasons have been illustrated using a numerical example with specific values for the parameters.

Keywords: Utensils Manufacturing System; MTSF; Availability; Regenerative Point Technique, and Preventive Maintenance (PM).

1. Introduction

Companies are constantly adapting their organisational structures and competitive strategies to meet the many markets demands in today's world of global competition. They increase their

capacity, long-term adaptability, and process sensitivity. The focus of entrepreneurial operations and strategies that promote adaptation to actual market needs has been the production system and its internal structures. The assessment and prediction of system reliability concerning the various operational process stages have grown in significance. For complex industrial systems, which typically have different failure modes, it is essential to create effective reliability evaluation tools to ensure appropriate performance under high and ambiguous demands.

Researchers have significantly improved the reliability analysis of manufacturing systems over time. Gupta and Tewari [1] investigated the performance analysis of a thermal power plant, which has four subsystems that operate at full capacity, below capacity, or fail. Rizwan et al. [2,3] provided a case study on a desalination plant that was shut down for annual maintenance during the winter for one month. Rahbi et al. [4] examined the reliability of a roading anode factory in the aluminium industry, where raw materials are routed via eight stations with a mix of series and parallel layouts. Manocha et al. [5] dealt with a system that had one database linked with a hot standby unit. Yusuf et al. [6] looked into the effectiveness of both online and offline preventive maintenance in repairing systems. Goyal et al. [7] studied the physical processing of a sewage treatment plant with five series-connected components. Rizwan et al. [8] examined the three pumps of a system delivering desalinated water to determine which pump was the least effective and required improvement for the system as a whole. Sachdeva et al. [9] analyse the sensitivity and reliability of membrane biofilm fuel cell. From all of these manufacturing plants, we took into account the utensil production plant to analyse reliability and profitability [10] as everyone needs efficient and appropriate utensils. Singh and Mahajan [11] examined the availability of the steel production facility, which includes systems for cutting, pressing, spinning, and polishing. Zaidi and Goya [12] studied the availability analysis of the cutting, furnance, hot-cold rolling, and roller furnisher series systems that make up the sheet formation system of the utensil manufacturing factory. Using a fuzzy technique, Kumar and Kumar [13] calculated the reliability of a production facility for utensils.

The profit analysis of the utensil manufacturing industry with the effect of temperature and preventative maintenance has yet to be examined. So, taking into account the impact of temperature, this article investigates the profitability and reliability of a manufacturing plant for utensils. The facility uses several subsystems to process raw materials, including cutting, pressing, spinning, polishing, and packing. Sheets are cut into circular shapes by cutting machines, and then they are pressed using different dies on a pressing machine according to the size and shape of various types of kitchenware. After that, sheets were sent out for spinning. A polished-ready product has been generated by the last stage of the process. Winter production requires a significant amount of heat, which could damage the machinery. As a result, production is low, and preventive maintenance is carried out during the winter.

The rest of the paper is organised as follow. Different notations, assumptions, and description of the system are included in Section 2. The stochastic model and its state transition probabilities are described in Section 3, along with a number of system metrics as well as a profit analysis of the system. The system measurements acquired using the graphs are examined in Section 4. Section 5 concludes with a few interesting interpretations.

2. System Description, Notations and Assumptions

Utensils manufacturing plants are widely used to produce various kinds of utensils. Utensils plant can have a variety of parts but mainly the plant consists for four subsystems like cutting system, pressing system, spinning system and polishing and packing system. Manufacturing of utensils entails the press or spin forming of metal, which frequently involves complex geometries with straight sides and as well as curves of various radii.

2.1. Description of the System

Sub-system C (Circle Cutting Machine)
As needed, sheets are cut into circular shapes.
Sub-system P (Pressing Machine)
The circle that was cut using a circular saw is now being sent to a pressing machine. Here, it is pressed using various dies in accordance with the size and shape of various types of kitchenware. Due to their shallow depth, some products, including as plates and bojanthal are ready for polishing right away.
Sub-system S (Spinning Machine)
According to their dies, the product created by pressing is sent for spinning. Some goods don't require further annealing before polishing, but others require it because of their deeper shapes. To eliminate contaminants, these items must be subjected to acid cleaning (Acid is a combination of Sulphuric and nitric acid).
Sub-system K (Polishing & Packing)
The final process has produced a product that is polished-ready. This stage involves packing and polishing the final product.

2.2. Notations

$m_1(t), M_1(t)$	probability and cumulative density functions by which the system go for preventive maintenance
$m_2(t), M_2(t)$	probability and cumulative density functions for completion of preventive maintenance time
$w_1(t), W_1(t)$	probability and cumulative density functions for changing the summer to winter season
$w_2(t), W_2(t)$	probability and cumulative density functions for changing the winter to summer season
a_1, a_2, a_3	rate of failure for subsystem C, P, S
b_1, b_2, b_3	rate of repair for subsystem C, P, S
O_{CPSK}	subsystem C, P, S, K operative
$F_C D_{PSK}$	subsystem C under repair and subsystem P, S, K under down state
$F_P D_{CSK}$	subsystem P under repair and subsystem C, S, K under down state
$F_S D_{CPK}$	subsystem S under repair and subsystem C, P, K under down state
$M_K D_{CPS}$	subsystem K under P.M. and subsystem C, P, S under down state
$\odot$	Laplace Stieltjes Convolution
©	Laplace Convolution

For other notations, refer [5].

2.3. Assumptions

- The failure and repair rates are independent and exponential in general.
- None of the sub-systems are experiencing simultaneous failures.
- Subsystem K has never failed.
- The repaired system works just like the new system.
- Subsystems are only repaired when they are in failed state.

3. Analysis of Model

The transition diagram of the system is given in Fig. 1.

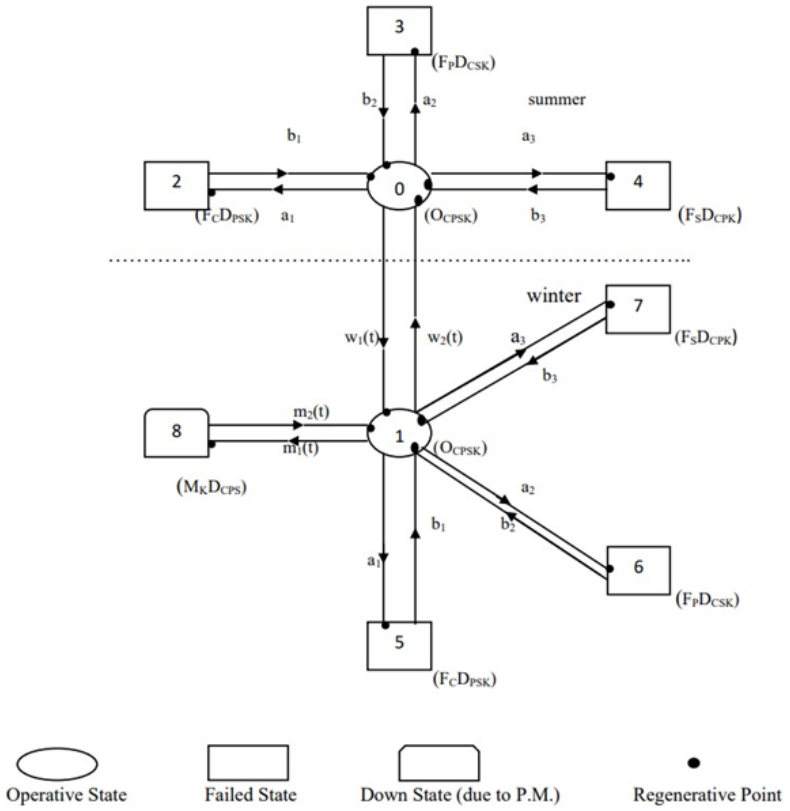


Figure 1: *State Transition Diagram*

Description of the Model and Transition Probabilities

3.1. Description of the model

Various states of the model for the system consisting four subsystems with season wise (summer and winter). The transition between states of system is shown in Fig. 1. States 0, 1, 2, 3, 4, 5, 6, 7, and 8 of the state transition diagrams are regeneration points. States 0 and 1 are the states where four subsystem work and so represents operative state during summer and winter respectively. States 2 and 5 are the states where the sub-system C go in failed states during summer and winter respectively so represents failed state. States 3 and 6 are the states where sub-system P go in failed states during summer and winter respectively so represents failed state. States 4 and 7 are the states where sub-system S go in failed states during summer and winter respectively so represents failed state. States 8 where sub-system K under preventive maintenance.

3.2. State Transition Probabilities

On the basis of state transition diagram, expressions for transition probabilities are given as follows:

$q_{01}(t) = e^{-(a_1+a_2+a_3)t} w_1(t)$ $\qquad$ $q_{02}(t) = a_1 e^{-(a_1+a_2+a_3)t} \overline{W_1}(t)$

$q_{03}(t) = a_2 e^{-(a_1+a_2+a_3)t}\overline{W_1}(t)$ $q_{04}(t) = a_3 e^{-(a_1+a_2+a_3)t}\overline{W_1}(t)$
$q_{10}(t) = e^{-(a_1+a_2+a_3+)t}E_{10}(t)$ $q_{15}(t) = a_1 e^{-(a_1+a_2+a_3)t}E_{15}(t)$
$q_{16}(t) = a_2 e^{-(a_1+a_2+a_3)t}E_{15}(t)$ $q_{17}(t) = a_3 e^{-(a_1+a_2+a_3)t}E_{15}(t)$
$q_{18}(t) = e^{-(a_1+a_2+a_3)t}E_{18}(t)$ $q_{20}(t) = b_1 e^{-b_1 t}$
$q_{30}(t) = b_2 e^{-b_2 t}$ $q_{40}(t) = b_3 e^{-b_3 t}$
$q_{51}(t) = b_1 e^{-b_1 t}$ $q_{61}(t) = b_2 e^{-b_2 t}$
$q_{71}(t) = b_3 e^{-b_3 t}$ $q_{81}(t) = m_2(t)$

where

$E_{10}(t) = \overline{M_1}(t)w_2(t)$ $E_{15}(t) = \overline{M_1}(t)\overline{W_2}(t)$
$E_{18}(t) = m_1(t)\overline{W_2}(t)$

Transition probabilities $p_{ij}(t)$ from state i to state j can be calculated by taking Laplace transform of above obtained values of $q_{ij}(t)$ and then using the following mathematical relationship between p_{ij} and $q^*_{ij}(s)$
$p_{ij} = \lim_{s\to 0} q^*_{ij}(s)$
values of for all required combinations of i and j are obtained and the same are given as follows:

$p_{01} = w^*_1(a_1 + a_2 + a_3)$ $p_{02} = \frac{a_1}{(a_1+a_2+a_3)}[1 - w^*_1(a_1 + a_2 + a_3)]$
$p_{03} = \frac{a_2}{(a_1+a_2+a_3)}[1 - w^*_1(a_1 + a_2 + a_3)]$ $p_{04} = \frac{a_3}{(a_1+a_2+a_3)}[1 - w^*_1(a_1 + a_2 + a_3)]$
$p_{10} = E^*_{10}(a_1 + a_2 + a_3)$ $p_{15} = a_1 E^*_{15}(a_1 + a_2 + a_3)$
$p_{16} = a_2 E^*_{15}(a_1 + a_2 + a_3)$ $p_{17} = a_3 E^*_{15}(a_1 + a_2 + a_3)$
$p_{18} = E^*_{18}(a_1 + a_2 + a_3)$

Mean Sojourn time (μ_i)
If T_i denotes the stay time of the system in state i, then using the following mathematical relationship between μ_i and T_i
$\mu_i = \int_0^\infty P[T_i > t]dt$
values of μ_i for all required values of i are found, and the same are provided as:
$\mu_0 = \int_0^\infty e^{-(a_1+a_2+a_3)t}\overline{W_1}(t)dt$
$\mu_1 = \int_0^\infty e^{-(a_1+a_2)t}\overline{E_{15}}(t)$
$\mu_2 = \mu_5 = \frac{1}{b_1}$
$\mu_3 = \mu_{16} = \frac{1}{b_2}$
$\mu_4 = \mu_7 = \frac{1}{b_3}$
$\mu_8 = \int_0^\infty \overline{M_2}(t)dt$

The unconditional mean time (m_{ij}) which the system under consideration takes to move to state j where counting of the time starts as soon as it enters into state i can be obtained using the following mathematical relationship between m_{ij} and $q_{ij}(t)$
$m_{ij} = \int_0^\infty t q_{ij}(t)dt$,
values of for all required combinations of i and j thus obtained and given as follows:

$\mu_0 = m_{01} + m_{02} + m_{03} + m_{04}$; $\mu_1 = m_{10} + m_{15} + m_{16} + m_{17} + m_{18}$
$\mu_2 = m_{20}$; $\mu_3 = m_{30}$
$\mu_4 = m_{40}$; $\mu_5 = m_{51}$;
$\mu_6 = m_{61}$ $\mu_7 = m_{71}$;
$\mu_8 = m_{81}$

4. System Performance Measures

4.1. Mean Time to System Failure

We retain failed states as absorbing states in order to calculate the system's MTSF. Using recursive relations for $\phi_i(t)$ can be obtained and the same are given as:
$\phi_0(t) = Q_{01}(t) \odot \phi_1(t) + Q_{02}(t) + Q_{03}(t) + Q_{04}(t)$
$\phi_1(t) = Q_{10}(t) \odot \phi_0(t) + Q_{15}(t) + Q_{16}(t) + Q_{17}(t) + Q_{18}(t) \odot \phi_8(t)$
$\phi_8(t) = Q_{81}(t) \odot \phi_1(t)$
By solving these relations for $\phi_0^{**}(s)$ using the Laplace Stieltjes transformation of these relations, we get
$\phi_0^{**}(s) = \frac{N(s)}{D(s)}$,
where
$N(s) = (q_{02}^*(s) + q_{03}^*(s) + q_{04}^*(s))(1 - q_{18}^*(s)q_{81}^*(s)) + q_{01}^*(s)(q_{15}^*(s) + q_{16}^*(s) + q_{17}^*(s))$
$D(s) = 1 - q_{01}^*(s)q_{10}^*(s) - q_{18}^*(s)q_{81}^*(s)$
Using above calculated value of $\phi_0^{**}(s)$, MTSF can be obtained when the system under consideration starts from the state 0 and the same is given as follows:
$T_0 = \lim_{s\to 0} \frac{1-\phi_0^{**}(s)}{s} = \frac{N}{D}$,
where
$N = \mu_0[1 - p_{18}] + \mu_1 p_{01} + \mu_8 p_{01} p_{18}$
$D = 1 - p_{01}p_{10} - p_{18}$

4.2. Availabilities in Summer and Winter

During Summer
To determine the availability in summer $AS_0(t)$ of the system, recursive relations thus obtained using probabilistic arguments, are given as:
$AS_0(t) = M_0(t) + q_{01}(t)\copyright AS_1(t) + q_{02}(t)\copyright AS_2(t) + q_{03}(t)\copyright AS_3(t) + q_{04}(t)\copyright AS_4(t)$
$AS_1(t) = q_{10}(t)\copyright AS_0(t) + q_{15}(t)\copyright AS_5(t) + q_{16}(t)\copyright AS_6(t) + q_{17}(t)\copyright AS_7(t) + q_{18}(t)\copyright AS_8(t)$
$AS_2(t) = q_{20}(t)\copyright AS_0(t)$
$AS_3(t) = q_{30}(t)\copyright AS_0(t)$
$AS_4(t) = q_{40}(t)\copyright AS_0(t)$
$AS_5(t) = q_{51}(t)\copyright AS_1(t)$
$AS_6(t) = q_{61}(t)\copyright AS_1(t)$
$AS_7(t) = q_{71}(t)\copyright AS_1(t)$
$AS_8(t) = q_{81}(t)\copyright AS_1(t)$
where,
$M_0(t) = e^{-(a_1+a_2+a_3)t}\overline{W_1}(t)$
By solving these relations for $AS_0^*(s)$ using the Laplace transform of these relations, we get
$AS_0^*(s) = \frac{N_1(s)}{D_1(s)}$
where,
$N_1(s) = M_0^*(s)[1 - q_{15}^*(s)q_{51}^*(s) - q_{16}^*(s)q_{61}^*(s) - q_{17}^*(s)q_{71}^*(s) - q_{18}^*(s)q_{81}^*(s)]$
$D_1(s) = [q_{02}^*(s)q_{20}^*(s) + q_{03}^*(s)q_{30}^*(s) + q_{04}^*(s)q_{40}^*(s)][q_{15}^*(s)q_{51}^*(s) + q_{16}^*(s)q_{61}^*(s) + q_{17}^*(s)q_{71}^*(s) + q_{18}^*(s)q_{81}^*(s)] - [q_{02}^*(s)q_{20}^*(s) + q_{03}^*(s)q_{30}^*(s) + q_{04}^*(s)q_{40}^*(s) + q_{15}^*(s)q_{51}^*(s) + q_{16}^*(s)q_{61}^*(s) + q_{17}^*(s)q_{71}^*(s) + q_{18}^*(s)q_{81}^*(s)] + 1$
Using above calculated value of $AS_0^*(s)$ availability in summer can be obtained in steady-state and the same is given as follows:
$AS_0 = \lim_{s\to 0} sAS_0^*(s) = \frac{N_1}{D_1}$
where,
$N_1 = \mu_0 p_{10}$
$D_1 = (\mu_0 + \mu_2 p_{02} + \mu_3 p_{03} + \mu_4 p_{04})p_{10} + (\mu_1 + \mu_5 p_{15} + \mu_6 p_{16} + \mu_7 p_{17} + \mu_8 p_{18})p_{01}$

During Winter
Similarly, steady-state availability during winter are given as follows:
$AW_0 = \lim_{s\to 0} sAW_0^*(s) = \frac{N_2}{D_1}$
where,
D_1 already defined and
$N_2 = \mu_1 p_{01}$

4.3. Busy Period Analysis

Busy period of the repairman due to repair in summer
Similarly, steady-state Busy period of the repairman due to repair in summer are given as follows:
$BS_0 = \lim_{s\to 0} sBS_0^*(s) = \frac{N_3}{D_1}$
where,
D_1 already defined and
$N_3 = (\mu_2 p_{02} + \mu_3 p_{03} + \mu_4 p_{04}) p_{10}$
During Winter
Similarly, steady-state Busy period of the repairman due to repair in winter are given as follows:
$BW_0 = \lim_{s\to 0} sBW_0^*(s) = \frac{N_4}{D_1}$
where,
D_1 already defined and
$N_4 = (\mu_5 p_{15} + \mu_6 p_{16} + \mu_7 p_{17}) p_{01}$

4.4. Expected Number of Visits of the Repairman for Repair

During summer
Similarly, steady-state number of visits of the repairman during summer are given as follows:
$VS_0 = \lim_{s\to 0} sVS_0^*(s) = \frac{N_5}{D_1}$
where,
D_1 already defined and
$N_5 = p_{10}(1 - p_{01})$
During Winter
Similarly, steady-state number of visits of the repairman during winter are given as follows:
$VW_0 = \lim_{s\to 0} sVW_0^*(s) = \frac{N_6}{D_1}$
where,
D_1 already defined and
$N_6 = p_{01}(1 - p_{10} - p_{18})$

4.5. Expected Number of Visits of the Repairman for Preventive Maintenance

Similarly, steady-state number of visits of the repairman for preventive maintenance are given as follows:
$PM_0 = \lim_{s\to 0} sPM_0^*(s) = \frac{N_7}{D_1}$
where,
D_1 already defined and
$N_7 = p_{01} p_{18}$

5. Cost-Benefit Analysis

Profit of the system under consideration can be obtained by subtracting the costs due to repair, per visit charges of the repairman for repair in summer and winter and per visit charges of the repairman for preventive maintenance. The same can expressed in terms of the various performance measures obtained through the model developed in this given as follows:
$Profit = CS_0AS_0 + CW_0AS_0 - CS_1BS_0 - CW_1BW_0 - CS_2VS_0 - CW_2VW_0 - C_3PM_0$
where,
CS_0 : revenue during summer, per unit uptime
CW_0 : revenue during winter, per unit uptime
CS_1 : revenue during summer per unit time for repair
CW_1 : revenue during winter per unit time for repair
CS_2 : Cost per visit during summer for repair
CW_2 : Cost per visit during winter for repair
C_3 : Cost per visit for preventive maintenance

6. Numerical Interpretation

Let us assume particular value as:

$w_1(t) = \alpha e^{-\alpha t}$ $\quad$ $w_2(t) = \beta e^{-\beta t}$
$m_1(t) = \gamma e^{-\gamma t}$ $\quad$ $m_2(t) = \delta e^{-\delta t}$
$p_{01} = \frac{\alpha}{a_1+a_2+a_3+\alpha}$ $\quad$ $p_{02} = \frac{a_1}{a_1+a_2+a_3+\alpha}$
$p_{03} = \frac{a_2}{a_1+a_2+a_3+\alpha}$ $\quad$ $p_{04} = \frac{a_3}{a_1+a_2+a_3+\alpha}$
$p_{10} = \frac{\beta}{a_1+a_2+a_3+\gamma+\beta}$ $\quad$ $p_{15} = \frac{a_1}{a_1+a_2+a_3+\gamma+\beta}$
$p_{16} = \frac{a_2}{a_1+a_2+a_3+\gamma+\beta}$ $\quad$ $p_{17} = \frac{a_3}{a_1+a_2+a_3+\gamma+\beta}$
$p_{18} = \frac{\gamma}{a_1+a_2+a_3+\gamma+\beta}$ $\quad$ $p_{20} = p_{30} = p_{40} = p_{51} = p_{61} = p_{71} = p_{81} = 1$
$\mu_0 = \frac{1}{a_1+a_2+a_3+\alpha}$ $\quad$ $\mu_1 = \frac{1}{a_1+a_2+a_3+\beta+\gamma}$
$\mu_2 = \frac{1}{b_1}$ $\quad$ $\mu_3 = \frac{1}{b_2}$
$\mu_4 = \frac{1}{b_3}$ $\quad$ $\mu_5 = \frac{1}{b_1}$
$\mu_6 = \frac{1}{b_1}$ $\quad$ $\mu_7 = \frac{1}{b_3}$
$\mu_8 = \frac{1}{\delta}$

where
$a_1 = 0.635, a_3 = 0.3589, a_2 = 0.781, b_1 = 0.887, b_2 = 0.793, b_3 = 0.821, \alpha = 0.815, \beta = 0.013,$
$\gamma = 0.937, \delta = 0.870, CS_0 = 15000, CS_1 = 1500, CW_0 = 15000, CW_1 = 1600, CS_2 = 1450,$
$CW_2 = 1550, C3 = 1400.$
Various graphs have been plotted but all the graphs have not been shown here to use minimum space and to avoid repetition of similar interpretations. However, the users of such systems may plot graph of their interest as per the requirement and may take important decision regarding profitability of the system. Regarding the availability and nature of MTSF, various rates have been depicted as shown in Fig. 2 and 3 which reveal that MTSF and Availability decreases as failure rates increases. However, their values go in the direction δ and b_2. Some of the plotted graphs are shown as follows:
The MTSF behaviour for varied δ is given in Fig. 2. MTSF decreases as the failure rate value (a_2) rises. Higher values of δ correspond to higher values in it.

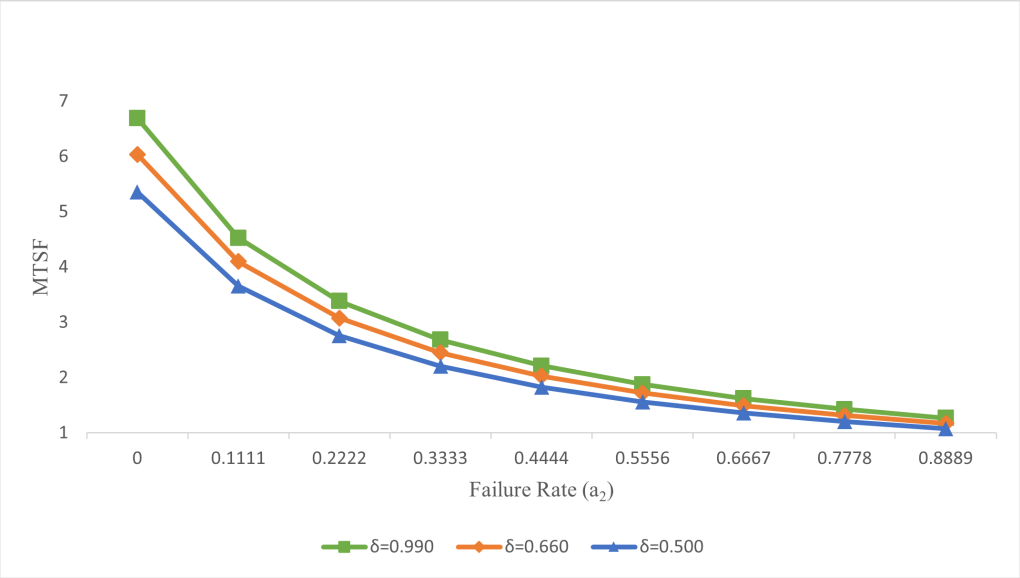


Figure 2: *MTSF versus Failure Rate* (a_2) *for different values of* (δ)

The availability behaviour in the summer and winter w.r.t. failure rate is shown in Fig. 3. Summer availability and winter availability both declines as the value of the failure rate rises (a_2). Also, the system is available more in summer than winter season.

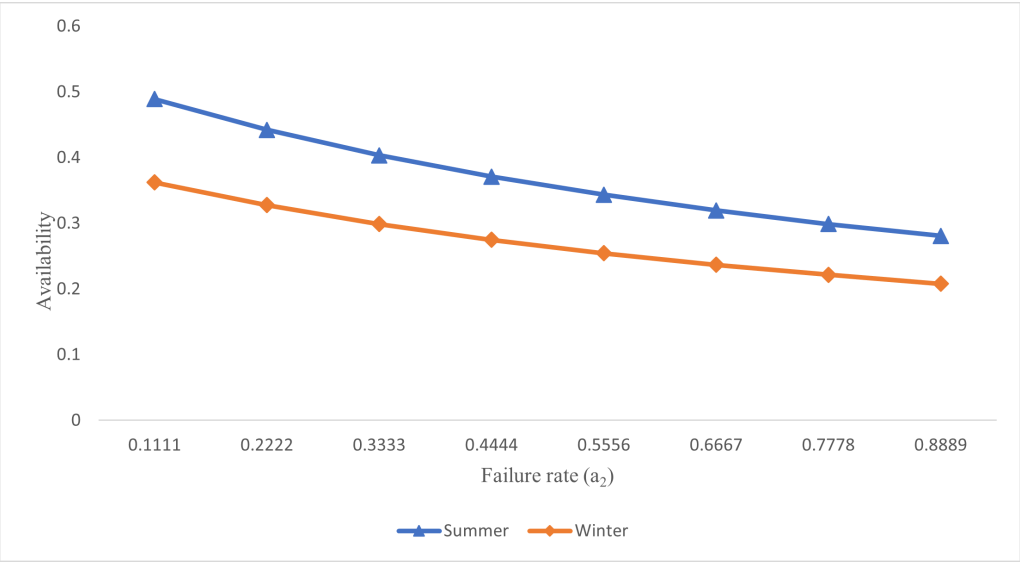


Figure 3: *Availability in Summer and Winter for different values of Failure rate* (a_2)

The way that profit acts in relation to revenue in the summer (CS_0) for various values of the cost paid for repair in the summer (CS_1) is shown in Fig. 4. As revenue values rise in the summer, profit rises as well (CS_0). Additionally, it has been seen that as (CS_1) values rise, the profit falls.

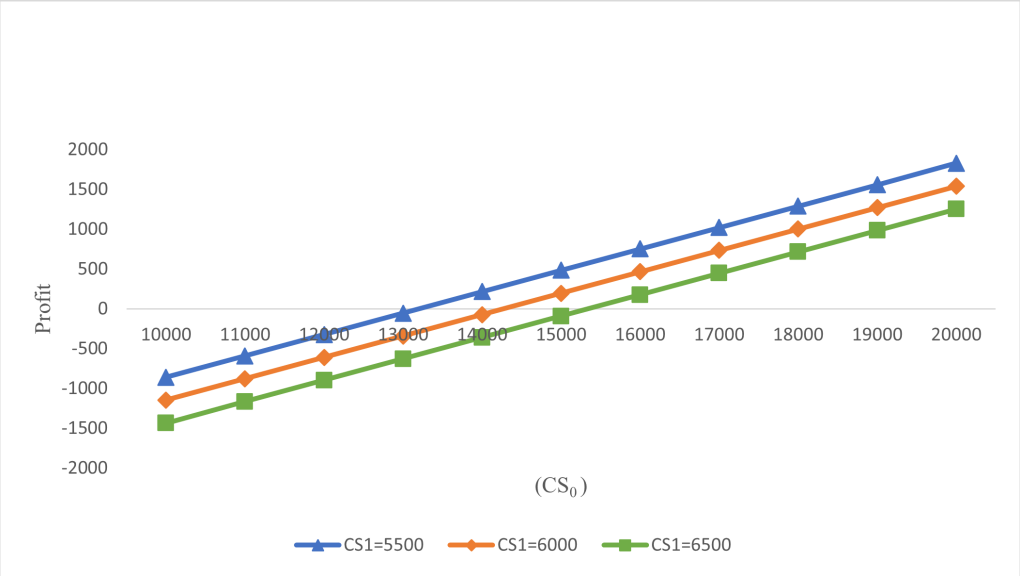


Figure 4: *Profit versus revenue in summer* (CS_0) *for different values of cost paid for repair in summer* (CS_1)

The relationship between profit and revenue in winter (CW_0) for various values of cost paid for repair in winter (CW_1) is shown in Fig. 5. With an increase in winter revenue values, profit rises (CW_0). Additionally, it has been noted that as (CW_1) values rise, the profit falls.

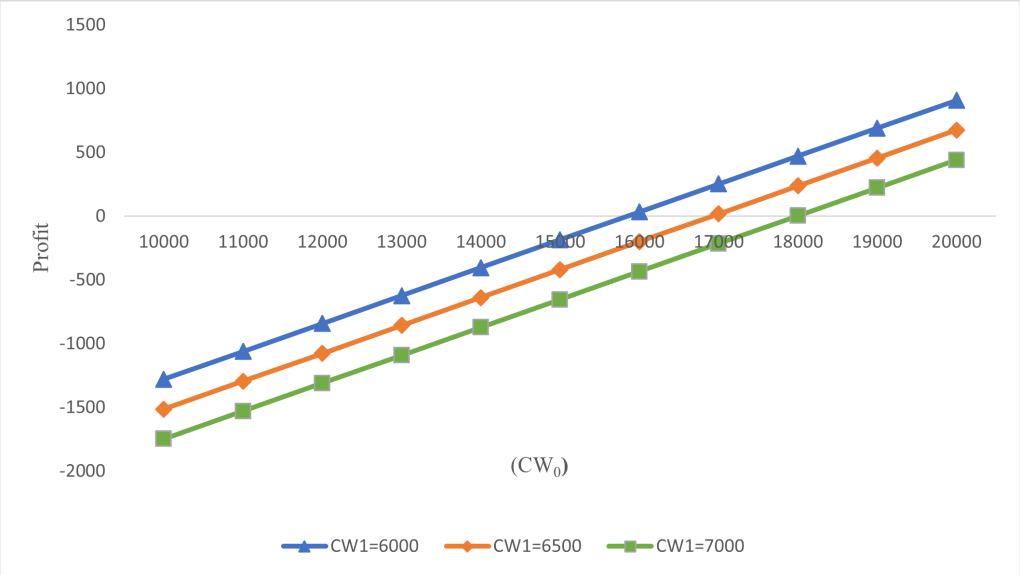


Figure 5: *Profit versus revenue in winter* (CW_0) *for different values of cost paid for repair in winter* (CW_1)

Fig. 6 illustrates the behaviour of profit in relation to revenue during the winter (CW_0) for various costs associated with preventive maintenance (C_3). With an increase in winter revenue values, profit rises (CW_0). Additionally, it has been seen that as (C_3) values rise, the profit falls.

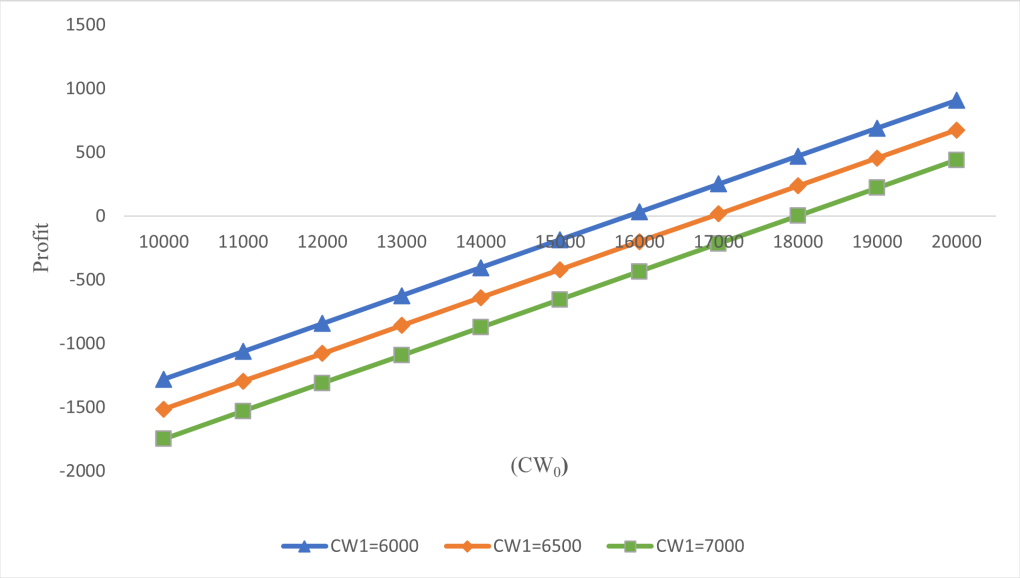


Figure 6: *Profit versus revenue in winter* (CW_0)*for different values of cost paid for Preventive Maintenance* (C_3)

Values of parameters taken and cut-off points obtained from the above figures are tabulated as follows:

Fig	Varied Parameters	Condition	Interpretation
4	CS_1 = 5500	CS_0> 13202.0841	System is profitable
	CS_1 = 6000	CS_0> 14270.9341	System is profitable
	CS_1 = 6500	CS_0> 15339.7841	System is profitable
5	CW_1 =6000	CW_0>15848.2632	System is profitable
	CW_1 =6500	CW_0>16917.2760	System is profitable
	CW_1 =7000	CW_0>17986.2888	System is profitable
6	C_3 = 5000	CW_0>10842.1274	System is profitable
	C_3 = 6000	CW_0>11778.9029	System is profitable
	C_3 = 7000	CW_0>12716.0337	System is profitable

7. Conclusion

In the current study, profitability and reliability of a production plant for utensils is analysed while taking the effect of temperature into account. The system is available more in summer than winter season. The findings for a specific situation demonstrate the relevance of research since cut-off points may be used to set lower and upper limits for a variety of factors. For instance, setting a product's pricing so that the system is profitable depends on the cut-off point for revenue per unit uptime. The cut-off points facilitate many crucial judgments for the profits according to revenue.

Funding

The first author delightedly acknowledges the University Grants Commission (UGC), New Delhi, India for providing financial support.

Disclosure statement

The authors declare that they have no conflict of interest.

References

[1] Gupta, S., and Tewari, P. C. (2009).Simulation modeling and analysis of a complex system of a thermal power plant.*J Ind. Eng. Manag*, 387–406.

[2] Rizwan, S. M., Padmavathi, N., Anita, P. and Taneja, G., (2013).Reliability Analysis of a Seven Unit Desalination Plant with Shutdown During Winter Season and Repair / Maintenance on FCFS Basis.*International Journal of Performability Engineering*, 523–528.

[3] Rizwan, S. M., Thanikal, J. and Torrijos, M., (2014).A general model for reliability analysis of a domestic wastewater treatment plant. *International Journal of Condition Monitoring and Diagnostic Engineering Management* , 3–6.

[4] Al Rahbi, Y., Rizwan, S. M., Alkali, B., Cowell, A. and Taneja, G., (2017).., Reliability analysis of rodding anode plant in aluminum industry. *International Journal of Applied Engineering Research* , 5616–5623.

[5] Manocha, A., Taneja, G., Singh, S. and Rishi R., (2019).Modelling and analysis of two-unit hot standby database system with random inspection of standby unit. *International Journal of Mathematics in Operational Research*, 156–180.

[6] Yusuf, I., Yusuf, B. and Suleiman, K.,(2019). , Reliability assessment of a repairable system under online and offline preventive maintenance*Life Cycle Reliability and Safety Engineering* , 1–16.

[7] Goyal, D., Kumar, A., Saini, M. and Joshi, H., (2019).Reliability, maintainability and sensitivity analysis of physical processing unit of sewage treatment plant.*SN Appl. Sci.* , 1507.

[8] Rizwan, S. M., Sachdeva, K., Alagiriswamy, S. and Al Rahbi, Y., (2023). Performability and Sensitivity Analysis of the Three Pumps of a Desalination Water Pumping Station. *International Journal of Engineering Trends and Technology* , 51(2): 283–292.

[9] Rizwan, S. M., Sachdeva, K., Al Balushi, N., Al Rashdi, S., and Taj, S. Z., (2023) Reliability and Sensitivity Analysis of Membrane Biofilm Fuel Cell. *International Journal of Engineering Trends and Technology* , 71(3): 73–80.

[10] Sachdeva, K., Taneja, G., and Manocha, A., (2022).Sensitivity and economic analysis of an insured system with extended conditional warranty. *Reliability: Theory & Applications*,315–327.

[11] Singh, J., and Mahajan, P., (1999). Reliability of utensils manufacturing plant-a case study. *Opsearch* , 13(24): 260–269.

[12] Zaidi, Z., and Goya, Y. K., (2012). Availability Analysis of Sheet Formation System of the Utensils Industry Mathematical View. *International Transactions in Applied Sciences* , 214–220.

[13] Kumar, K., and Kumar, P., (2010). Mathematical modeling and analysis of stainless steel utensil manufacturing unit using fuzzy reliability. . *International Journal of Engineering Science and Technology* , 2370–2376.

RT&A, No.3 (74)
Volume 18, September 2023

ANALYSIS OF A FLEXIBLE GROUP SERVICE $MAP\backslash PH\backslash 1$ QUEUEING MODEL WITH, IMMEDIATE FEEDBACK, BALKING AND RENEGING

G. Ayyappan, S. Kalaiarasi

•

Department of Mathematics
Puducherry Technological University
Puducherry, India.
ayyappanpec@hotmail.com, kalaiarasi.math@gmail.com

Abstract

Queueing models in which the services are provided in groups (or blocks or batches) have found to be very useful in real-world applications and such queues been extensively analysed in the literature. In this paper we see one such group service queueing model with balking, reneging and immediate feedback. The arrival processes is a Markovian arrival , where, the arriving customer may balk the system while the server is idle and the pool is empty. Customers are provided service in groups of varying size from 1 *to the fixed constant, say, N. The service time of a batch follows the phase type distribution corresponding to the each size of the group. A group's service time is taken as the highest of the service times of each customers who make up the group. The group of customers who are dissatisfied with the service then that group will get the service immediately. Here, the feedback of a group is defined as the average of the feedback of each customers who make up the group. During the admission period the customers may renege. We calculated the steady state probabilities by using the matrix geometric method, then, by using it we computed few performance measures. We have studied the busy period and the distribution of waiting time is derived. Results are illustrated with some graphical representations.*

Keywords: Markovian Arrival Process, Flexible group service, Phase type Distributions, Immediate Feedback, Balking, Reneging.

1. Introduction

Queueing models with a group service play a vital role in many real life and engineering systems. And these queues can be generated physically or simulated by computers. Usually in the group service models, the minimum and maximum size of group are presumed. Bailey introduced the bulk service queueing model with fixed group size in [4]. Chaudry and Templeton in [5] studied bulk service queues in detail. A survey paper on bulk service queueing models by Sasikala and Indhira [6] is noteworthy. And the authors Banerjee.A, Gupta.U, Chakravarthy.S in [7] derived the significant results for queues with group service and many models with real-life applications are presented.

Neuts [8] has introduced the general group service rule, according to which the server will start the service only if "a" or more customers are in the queue and the highest service capacity is "b." At the service completion epoch of a batch, if the number of customers present in the queue is less than $'a'$ then the server has to wait until $'a'$ number of customer arrives. If the number of customers less than or equal to $'b'$ and greater than or equal to $'a'$ then the service commences immediately to existing customers. If the number of customers is greater than $'b'$, then only $'b'$

customers taken in to service. And in the literature very few papers have dealt with group service with non-exponential service times .

The authors Brugno,D'Apice, Dudin, Manzo in [1] have examined a $MAP/PH/1$ queueing model with flexible group service. A predefined integer, let's say N, is typically used in the analysis of group services, and if there are less costumers in the queue than N, service is not initiated. But in this model they predefined the batch size as N, and they assumed that the server's idle time is restricted. Even if there are fewer consumers in the queue than N, service will still start once the idle time runs out. At a service completion moment, if there are N or more than N number of customers, the server provides service for exactly N customers. On the other hand, if the number of customers waiting in the queue is less than N, a admission period starts and its duration follows the PH - distribution. If the number of customers waiting reaches the value N before the admission period expires, the admission period is stopped and the service resumes with N customers. If the admission period expires before the arrival of Nth customer, then the server offers service simultaneously to the group of $'i'$ customers, where $'i'$ ranges between 1 to $N-1$ or if the admission period was over and when there is no one in queue, a new admissions period begins, and the procedure is repeated.

In [1],[7] for a batch service queueing models, a size $'m'$ customer group's service time is assumed to be the highest of $'m'$ identical PH- distributions which in turn a PH- distribution. The batch's service will be completed when the service for the last customer in the batch is completed. This type of batch service models are studied in cloud and grid computing [12].

Ayyappan and Thilagavathy in [9] have studied the $MAP/PH/1$ queueing model with Breakdown, Instantaneous feedback, and server vacation. And Downton in [11] by using random arrivals and a random service time distribution derived the waiting time distribution of bulk service queues.

In this paper we analyse a $MAP/PH/1$ queueing model with flexible group service, balking, reneging and immediate feedback. Here ,the feedback of a group is defined as the average of the feedback of each customers who make up the group.

The article's next sections are organised as follows. The mathematical model is presented with a graphic depiction in section 2. In section 3 we narrated the model and we formulate the QBD matrix. We derive the Ergodicity (stability) condition and the steady state probability vector in section 4. For this model we computed some performance measures in section 5. In section 6 we did busy period analysis. In section 7 waiting time distribution is derived . In section 8 some numerical results with graphical representations are illustrated and the conclusion is given in section 9.

2. The Narration of the Model

In this paper, Markovian arrival process is considered with depiction (D_0, D_1) of order n with the generator matrix $\tilde{D} = D_0 + D_1$. Markovian arrival's fundamental rate is defined as $\lambda = \pi D_1 e_n$, where π is the vector of stationary probability of $\tilde{D}$. Now we assume that the customers will get service in groups of size N, with $N \geq 2$ a fixed integer. If there are $N-1$ customers in the queue and the server is idle an arriving customer will get a immediate service and its PH representation is denoted as $(\beta^{(N)}, S^{(N)})$ of order $M^{(N)}$ with $S_0^{(N)} + S^{(N)}e = 0$ which implies $S_0^{(N)} = -S^{(N)}e$ otherwise based on the sequence of their arrival, the arriving customers are placed in the buffer. And at this moment choosing customers from the buffer at the time a service is complete is defined as follows

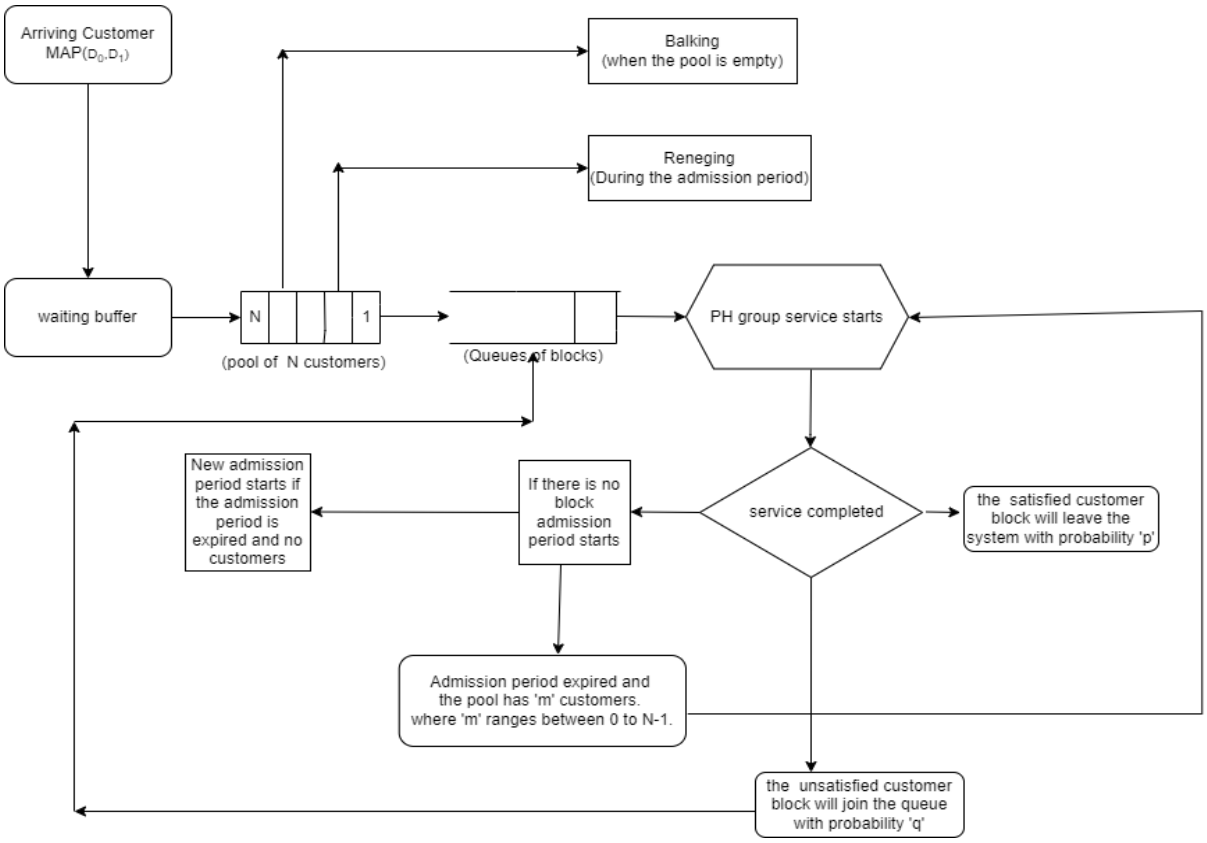


Figure 1: *Diagram illustrating the current model*

- The batch of exactly N consumers (say *Block*) begins service if at this point of time N or more customers are present in the buffer with PH representation $(\beta^{(N)}, S^{(N)})$ of order $M^{(N)}$.
- We refer to a group of consumers as a *pool*, if there are fewer than N customers. Then the so-called admission period begins at this point of time.The admission period follows PH distribution of which the PH representation is denoted by (α, T) of order $M^{(0)}$ with $T_0 + Te = 0$ which implies $T_0 = -Te$. And now
  - When the pool's total customer count equals N, the server starts providing services.
  - or the admission period expires.
  - If the admission period passed and the pool has one or more but fewer than N customers, Then all the customers in the pool strats service immediately and the service of r, customers $1 \leq r \leq N-1$ follows PH distribution of which the PH representation is given by $(\beta^{(r)}, S^{(r)})$ of order $M^{(r)}$ with $S_0^{(r)} + S^{(r)}e = 0$ which implies $S_0^{(r)} = -S^{(r)}e$.
  - A new admission period begins if the admission period passes with the pool empty, and the fundamental rate of admission period is defined as $\eta = [\alpha(-T)^{-1}e]^{-1}$.

Fundamental rate of service to r customers where $1 \leq r \leq N$ is defined as $\gamma_r = [\beta^{(r)}(-S^r)^{-1}e]^{-1}$.

During the server is idle and the pool is empty, with probability b, the customer might quit (balk) the system . During the admission period the customer may renege, which follows exponential distribution with parameter δ. We are dealing the single server queueing model with immediate feedback, this indicates whether a batch of customers is satisfied after receiving

service, they leave the system with probability c otherwise the batch of customers will be receiving feedback service right away with probability d such that $c + d = 1$. Here in the group service, we presume that the group's feedback is positive if the average of the individual feedbacks of the group is positive.

2.1. QBD process of system state

Notations for our model

- $\otimes$ - the matrix Kronecker product.
- $\oplus$ - the matrix Kronecker sum .
- I_{m} - an identity matrix of m- dimension.
- e- a column matrix with each entry is 1 of appropriate dimension.
- $diag\{d_k, k \in 1, ..., n\}$ is the diagonal matrix, whose entries are enclosed in brackets.
- F_k is the row matrix of dimension k each of its entries as 0

Let

$$\xi(t) = \{N(t), I(t), R(t), J(t)^{(R(t))}, V(t) : t \geq 0 \ \}$$

is continuous time Markov chain with state level independent Quasi-Birth-and-Death process , where

- $N(t)$ indicates that how many batches are present in the system at time t, which includes the batch in service,
- $I(t)$ indicates that how many customers are there in the pool at time t , $0 \leq I(t) \leq N-1$,
- $R(t)$ indicates that how many customers are getting the service at time t. Note that $R(t) = 0$ if $N(t) = 0$, as a result the admission period will be ongoing and $1 \ \leq R(t) \leq \ N-1$ if $N(t) \ \geq \ 1$
- $J(t)^{(R(t))}$ indicates the state of the PH process of customer admission if $R(t) = 0$ with $1 \leq J(t)^{(0)} \leq M^{(0)}$ or it indicates the state of the PH process of customer service process if $1 \leq \ R(t) \ \leq \ N$ with $1 \leq J(t)^{(R(t))} \leq M^{(R(t))}$.
- $V(t)$ shows the state of the Markovian arrival process with $1 \ \leq V(t) \leq \ n$.

$\xi(t)$ has the following state space,

$$B \ = \ l(0) \ \cup \ l(k)$$

where,

$$l(0) = \{(0, i, 0, p, s) : 0 \ \leq i \leq \ N-1 \ ; \ 1 \ \leq p \ \leq \ M^{(0)} \ ; 1 \ \leq s \ \leq n \ \}$$

and this can be simply written as

$$l(0) = \{(0, i) : 0 \ \leq i \leq \ N-1 \ \}$$

the PH process of admission period and phases of Markovian arrival are understood.

For $k \ \geq \ 1$,

$$l(k) = \{(k, i, r, p, s) : k \ \geq 1 \ ; \ 0 \ \leq i \leq \ N-1 \ ; \ 1 \ \leq \ r \ \leq \ N \ ; \ 1 \ \leq p \ \leq \ M^{(r)} \ ; 1 \ \leq s \ \leq n \ \}$$

and this can be simply written as

$$l(k) = \{(k,i,r) : k \geq 1\ ;\ 0 \leq i \leq N-1\ ;\ 1 \leq r \leq N\}$$

the PH process service to r number of customers where $1 \leq r \leq N$ and arrival phases are understood.

The QBD process of infinitesimal matrix generation is given by

$$Q = \begin{pmatrix} B_{00} & B_{01} & 0 & 0 & 0 & 0 & \cdots & \cdots \\ B_{10} & A_1 & A_0 & 0 & 0 & 0 & \cdots & \cdots \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & \cdots & \cdots \\ 0 & 0 & 0 & A_2 & A_1 & A_0 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \ddots & \ddots \end{pmatrix}$$

The matrix Q's block matrices are given below

$$B_{00} = \begin{pmatrix} {B_{00}}^{11} & {B_{00}}^{12} & 0 & 0 & 0 & 0 & \cdots & 0 \\ {B_{00}}^{21} & {B_{00}}^{22} & {B_{00}}^{23} & 0 & 0 & 0 & \cdots & 0 \\ 0 & {B_{00}}^{32} & {B_{00}}^{33} & {B_{00}}^{34} & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & {B_{00}}^{i\ i-1} & {B_{00}}^{i\ i} & {B_{00}}^{i\ i+1} & & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & & & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & {B_{00}}^{N-1\ N-2} & {B_{00}}^{N-1\ N-1} & \end{pmatrix} \text{ where,}$$

${B_{00}}^{11} = (T + T_0\alpha) \oplus (D_0 + bD_1)$; ${B_{00}}^{12} = (1-b)D_1 \otimes I_{M^{(0)}}$;
${B_{00}}^{21} = \delta I_n \otimes I_{M^{(0)}}$; ${B_{00}}^{22} = T \oplus (D_0 - \delta I_n)$;
${B_{00}}^{23} = D_1 \otimes I_{M^{(0)}}$; ${B_{00}}^{i\ i-1} = \delta I_n \otimes I_{M^{(0)}}$;
${B_{00}}^{i\ i} = T \oplus (D_0 - \delta I_n)$; ${B_{00}}^{i\ i+1} = D_1 \otimes I_{M^{(0)}}$

$$B_{01} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ (B_{01})_{1,0} & & & \\ (B_{01})_{2,0} & \vdots & \ddots & \vdots \\ \vdots & & & \\ (B_{01})_{N-1,0} & 0 & \cdots & 0 \end{pmatrix}, \text{ where}$$

$(B_{01})_{i,0} = (F_{i-1}, T_0 \otimes \beta^{(m)} \otimes I_n, F_{N-i})$ for $1 \leq i \leq N-2$
and $(B_{01})_{N-1,0} = (F_{N-2}, T_0 \otimes \beta^{(N-1)} \otimes I_n, e_{M^{(0)}} \otimes \beta^{(N)} \otimes D_1)$

$$A_1 = \begin{pmatrix} (A_1)_{0,0} & (A_1)_{0,1} & 0 & 0 & \cdots & 0 \\ 0 & (A_1)_{1,1} & (A_1)_{1,2} & 0 & \cdots & 0 \\ 0 & 0 & (A_1)_{2,2} & (A_1)_{2,3} & \cdots & 0 \\ \vdots & \vdots & \ddots & & \ddots & \vdots \\ \vdots & \vdots & & \ddots & & \\ 0 & \vdots & \cdots & & & (A_1)_{N-1,N-1} \end{pmatrix}, \text{ where}$$

$(A_1)_{i,i} = diag\ \{S^{(r)} + dS_0{}^{(r)}\beta^{(r)} \oplus D_0; 1 \leq r \leq N, 0 \leq i \leq N-1\}$

and $(A_1)_{i,i+1} = diag\ \{I_{M^{(r)}} \otimes D_1; 1 \leq \ \ r \ \ \leq N, 0 \leq \ \ i \ \ \leq N-2\}$

$$A_0 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & & & \\ & \vdots & \ddots & \vdots \\ 0 & & & \\ (A_0)_{N-1,0} & 0 & \cdots & 0 \end{pmatrix}, \text{ where } (A_0)_{N-1,0} = diag\ \{I_{M^{(r)}} \otimes D_1; 1 \leq \ \ r \ \ \leq N\}$$

$$A_2 = \begin{pmatrix} (A_2)_{0,0} & 0 & 0 & \cdots & 0 \\ 0 & (A_2)_{1,1} & 0 & \cdots & 0 \\ 0 & 0 & (A_2)_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & & \cdots & & (A_2)_{N-1,N-1} \end{pmatrix},$$

$$\text{with } (A_2)_{i,i} = \begin{pmatrix} 0 & \cdots & 0 & cS_0{}^{(1)} \otimes \beta^{(N)} \otimes I_n \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & 0 & cS_0{}^{(N)} \otimes \beta^{(N)} \otimes I_n \end{pmatrix}$$

$$B_{1,0} = diag\ \{(B_{1,0})_{i,i}\ ;\ 0 \leq \ \ i \ \ \leq N-1\}, \text{ where } (B_{1,0})_{i,i} = \begin{pmatrix} cS_0{}^{(1)} \otimes \alpha \otimes I_n \\ \vdots \\ cS_0{}^{(N)} \otimes \alpha \otimes I_n \end{pmatrix}.$$

3. Condition for stableness

Let us define the matrix $A = A_0 + A_1 + A_2$, then

$$A = \begin{pmatrix} F & F' & & \\ & \ddots & \ddots & \\ & & F & F' \\ F' & & & F \end{pmatrix}, \ where \ \ F' = \begin{pmatrix} I_{M^{(1)}} \otimes D_1 & & \\ & \ddots & \\ & & I_{M^{(N)}} \otimes D_1 \end{pmatrix},$$

$$and \ \ F = \begin{pmatrix} S^{(1)} + dS_0{}^{(1)}\beta^{(1)} \oplus D_0 & & cS_0{}^{(1)}\beta^{(N)} \otimes I_n \\ & \ddots & \vdots \\ & \ddots & cS_0{}^{(N-1)}\beta^{(N)} \otimes I_n \\ & & S^{(N)} + dS_0{}^{(N)}\beta^{(N)} \oplus D_0 + cS_0{}^{(N)}\beta^{(N)} \otimes I_n \end{pmatrix}$$

It is clear that A is a square matrix which is an irreducible infinitesimal generator matrix whose order is $N\ M^{(1)}\ n + N\ M^{(2)}\ n + \cdots + N\ M^{(N)}\ n$. The steady-state probability vector of A is indicated by z. And the vector z is denoted as $z = (z_0, z_1, z_2, \cdots, z_{N-1})$, where $z_i = (z_i^1, z_i^2, \cdots, z_i^N)$, $0 \leq \ \ i \ \ \leq N-1$ which satisfies $zA = 0$ and $ze = 1$. The QBD structure exists for the Markov process. Also there exists Ergodicity (stability) criteria for our model and that it should satisfy $zA_0e \ < \ zA_2e$, which is the if and only if condition for stability of a *QBD* process. By resolving the following equations, the vector z can be determined.
$z_0^1(S^{(1)} + dS_0{}^{(1)}\beta^{(1)} \oplus D_0) + z_{N-1}^1(I_{M^{(1)}} \otimes D_1) = 0$
$z_0^2(S^{(2)} + dS_0{}^{(2)}\beta^{(2)} \oplus D_0) + z_{N-1}^2(I_{M^{(2)}} \otimes D_1) = 0$

$\vdots$

$$\sum_{i=1}^{N-1} z_0^i cS_0{}^{(i)}\beta^{(i)} \otimes I_n + z_0^N cS_0{}^{(N)}\beta^{(N)} \otimes I_n + (S^{(1)} + dS_0{}^{(1)}\beta^{(1)} \oplus D_0) + z_{N-1}^N (I_{M^{(N)}} \otimes D_1) = 0.$$

Similarly, $for\ \ i\ \ , 0\ \leq i \leq N-2$ we have,
$z_{i+1}^1 (S^{(1)} + dS_0{}^{(1)}\beta^{(1)} \oplus D_0) + z_i^1 (I_{M^{(1)}} \otimes D_1) = 0$
$z_{i+1}^2 (S^{(2)} + dS_0{}^{(2)}\beta^{(2)} \oplus D_0) + z_i^2 (I_{M^{(2)}} \otimes D_1) = 0$

$$\vdots$$

$$\sum_{i=1}^{N-1} z_{i+1}^i cS_0{}^{(i)}\beta^{(i)} \otimes I_n + z_{i+1}^N cS_0{}^{(N)}\beta^{(N)} \otimes I_n + (S^{(1)} + dS_0{}^{(1)}\beta^{(1)} \oplus D_0) + z_i^N (I_{M^{(N)}} \otimes D_1) = 0.$$

Following some algebraic calculation, the stability condition $zA_0e < zA_2e$, which is turns to be

$$\sum_{r=1}^{N} z_{N-1}^r (e_{M^{(r)}} \otimes D_1 e_n) \quad < \quad \sum_{i=0}^{N-1} \sum_{r=1}^{N} z_i^r cS_0{}^{(r)} \otimes e_n.$$

After simplification the Ergodicity condition can be precisely written as $\lambda\gamma_N < N$.

3.1. Study of the Stationary Probability vector

Let x be the Q 's the steady-state probability vector and it is subdivided as $x = (x_0, x_1, x_{2,}, \cdots)$. Note that x_0's dimension is NM^0n and dimension of $x_1, x_2, x_3, \cdots$, are $N(M^{(1)} + M^{(2)} + \cdots + M^{(N)})n$. Then x satisfies the condition $xQ = 0$ and $xe = 1$. Once the stability condition is met, the subvectors of x, except for x_0 and x_1 are provided by the following equation, which corresponds to the various level states.

$$x_j = x_1 R^{j-1}, \ \ j \geq 2$$

where R represents the minimum non-negative solution of the matrix quadratic equation as $R^2A_2 + RA_1 + A_0 = 0$, as defined by Neuts [3]. Due to the stability of our system and the fact that the row sums of the sum of square matrices A_0, A_1, and A_2 is zero, R , the rate matrix is a square matrix with order $N(M^{(1)} + M^{(2)} + \cdots + M^{(N)})n$. The R matrix is derived from the above quadratic equation and also fulfils $RA_2e = A_0e$.
The following equations were solved to obtain the sub vectors x_0 and x_1.

$$x_0B_{00} + x_1B_{10} = 0$$

$$x_0B_{01} + x_1(A_1 + RA_2) = 0$$

conditioned on the normalising state

$$x_0 e_{NM^{(0)}n} + x_1(1-R)^{-1} e_{N\ n(M^{(1)}+M^{(2)}+\cdots+M^{(N)})} = 1.$$

hence, the matrix R could be computed theoriticaly with the reference of Latouche and Ramaswami [2] using necessary steps in the R's Logarithmic reduction algorithm.

4. Performance Measures

- The expected number of customer blocks, including the one receiving service $E_{block} = \sum_{k=1}^{\infty} kx_ke = x_1(1-R)^{-2}$.
- The expected number of blocks of customers excluding the one in service $\tilde{E}_{block} = \sum_{k=1}^{\infty}(k-1)x_ke = E_{block} - 1 + x_0e_0$.

- The expected number of customers in the pool
$E_{pool} = \sum_{i=0}^{N-1} ix_0\tilde{e}_{0i} + \sum_{k=1}^{\infty}\sum_{i=0}^{N-1}\sum_{m=1}^{N} ix_{kim}e = \sum_{i=0}^{N-1} ix_0\tilde{e}_{0i} + \sum_{i=0}^{N-1} ix_1(1-R)^{-1}\tilde{e}_i.$

  where $\tilde{e}_{0i}$ is the column vector of order $NM^{(0)}n$ with $(i(M^{(0)})n\ +1)$ st to $((i+1)M^{(0)}n)$ th entries are 1 and all other entries are zeros.
  and $\tilde{e}_i$ is the column vector of order $(N(M^{(1)}+\cdots+M^{(N)})n$ with $(i(M^{(1)}+\cdots+M^{(N)})n\ +$ 1) st to $((i+1)(M^{(1)}+\cdots+M^{(N)})n)$ th entries are 1 and all other entries are zeros.

- The expected number of customers in the service
$E_{service} = \sum_{k=1}^{\infty}\sum_{i=0}^{N-1}\sum_{m=1}^{N} mx_{kim}e$
$= \sum_{m=1}^{N} m(x_1(1-R)^{-1}e_{0i} + x_1(1-R)^{-1}e_{1m} + \cdots + x_1(1-R)^{-1}e_{N-1m})$
$= \sum_{m=1}^{N} m(\sum_{i=0}^{N-1} x_1(1-R)^{-1}e_{im})$
  where e_{im} are all column vectors of order $(N(M^{(1)}+\cdots+M^{(N)})n$ defined as
  for $m=1$, e_{i1} has $(m(\sum_{k=1}^{N} M^k n)+1)$ st to $(i(\sum_{k=1}^{N} M^k n) + M^{(1)}n)$ th entries are 1 and all other elements are zeros.
  for $2 \le m \le N-1$, e_{im} has $(i(\sum_{k=1}^{N} M^k n) + (\sum_{j=1}^{m-1} M^j n)+1)$ st to $(i(\sum_{k=1}^{N} M^k n) + (\sum_{j=1}^{m} M^j n))$ th entries are 1 and all other entries are zeros.
  and for $m = N$, e_{iN} has $(i(\sum_{k=1}^{N} M^k n) + (\sum_{j=1}^{N-1} M^j n)+1)$ st to $((i+1)(\sum_{k=1}^{N} M^k n))$ th entries are 1 and all other entries are zeros.

- The mean size of the system (mean system size or expected system size) at an arbitrary moment including the customers in service
$E_{system} = \sum_{k=1}^{\infty}\sum_{i=0}^{N-1}\sum_{m=1}^{N}(kN+i+m)x_{kim}e + \sum_{i=0}^{N-1} ix_0\tilde{e}_{0i}$
$= NE_{block} + E_{pool} + E_{service}$

- The mean size of the system at some random time excluding the customers in service
$\tilde{E}_{system} = \sum_{k=1}^{\infty}\sum_{i=0}^{N-1}\sum_{m=1}^{N}((k-1)N+i+m)x_{kim}e + \sum_{i=0}^{N-1} ix_0\tilde{e}_{0i}$
$= N\tilde{E}_{block} + E_{pool} + E_{service}$

- The probability of the server is idle at some random time
$P_{idle} = x_0e_0$

5. Busy Period Analysis

- A busy period is defined as the period of time from when a customer first enters an empty system till the first epoch after that when the system is empty once more. Thus it is the first passing time from level 1 to level 0. A busy cycle is defined as the whole first time at level 0 after visiting a state in any other level at least once.

- We must first define the term "fundamental period" before we can study the busy time. It is the first passing time from level k to level $k-1$ for the QBD process under examination. for $k \ge 2$.

- The case where $k = 0, 1$ correspond to boundary states must be discussed seperately.

- Note that for each level k, for $k \ge 1$, there corresponds $N(M^{(1)}\ +M^{(2)}\ +\cdots+M^{(N)})n$ states. The ordered pair (k, j) represents j th state of level k where the states are ordered in the lexicographic order.

- Let $G_{jj'}(v, x)$ provides the conditional probability of the $Quasi-Birth-Death$ process, this process commences from the state (k,j) at time $t = 0$ accesses the level $k-1$ within the time x. We can alter the v transition move left and get into the the state $(k-1, j')$.

To proceed further, we present the combined transform

$$\tilde{G}_{kk'}(z,s) = \sum_{v=1}^{\infty} z^v \int_0^{\infty} e^{-sx} dG_{kk'}(v,x) \; ; \; |z| \leq 1, Re(s) \geq 0$$

and the matrix is denoted as

$$\tilde{G}(z,s) = \tilde{G}_{kk'}(z,s) \tag{1}$$

then (1) satisfies the equation

$$\tilde{G}(z,s) = z(sI - A_1)^{-1}A_2 + (sI - A_1)^{-1}A_0\tilde{G}^2(z,s)$$

Now $G = G_{kk'} = \tilde{G}(1,0)$ handles the first passage timings, with the exception of the boundary states. Using the result

$$G = -(A_1 + RA_2)^{-1}A_2.$$

G matrix could be determine if R matrix is already known. Otherwise G matrix could be determine using logarithmic reduction algorithm method.
From the above discussions for the boundary levels 1 and 0 we have

$$\tilde{G}^{(1,0)}(z,s) = z(sI - A_1)^{-1}B_{10} + (sI - A_1)^{-1}A_0\tilde{G}(z,s)\tilde{G}^{(1,0)}(z,s),$$
$$\tilde{G}^{(0,0)}(z,s) = (sI - B_{00})^{-1}B_{01}\tilde{G}(z,s)\tilde{G}^{(1,0)}(z,s).$$

Since G, $\tilde{G}^{(1,0)}(1,0)$ and $\tilde{G}^{(0,0)}(1,0)$ are stochastic, using the above matrices we can calculate the below cases.

$$\vec{\mathbb{H}}_1 = -\frac{\partial}{\partial s}\tilde{G}(z,s)|_{z=1,s=0}e = -[A_1 + A_0(I+G)]^{-1}e \tag{2}$$

$$\vec{\mathbb{H}}_2 = -\frac{\partial}{\partial z}\tilde{G}(z,s)|_{z=1,s=0}e = -[A_1 + A_0(I+G)]^{-1}A_2e \tag{3}$$

$$\vec{\mathbb{H}}_1^{(1,0)} = -\frac{\partial}{\partial s}\tilde{G}(z,s)^{(1,0)}|_{z=1,s=0}e = -[A_1 + A_0G]^{-1}(A_0\vec{\mathbb{H}}_1 + e) \tag{4}$$

$$\vec{\mathbb{H}}_2^{(1,0)} = -\frac{\partial}{\partial z}\tilde{G}(z,s)^{(1,0)}|_{z=1,s=0}e = -[A_1 + A_0G]^{-1}(A_0\vec{\mathbb{H}}_2 + B_{10}e) \tag{5}$$

$$\vec{\mathbb{H}}_1^{(0,0)} = -\frac{\partial}{\partial s}\tilde{G}(z,s)^{(0,0)}|_{z=1,s=0}e = -B_{00}^{-1}[B_{01}\vec{\mathbb{H}}_1^{(1,0)} + e] \tag{6}$$

$$\vec{\mathbb{H}}_2^{(0,0)} = -\frac{\partial}{\partial z}\tilde{G}(z,s)^{(0,0)}|_{z=1,s=0}e = -B_{00}^{-1}[B_{01}\vec{\mathbb{H}}_2^{(1,0)}] \tag{7}$$

6. Analysis of Waiting Time Distribution

The first passage time analysis is used in this section to analyse the distribution of a customer's waiting time when they enter the queueing line. Let $W(t)$ be the waiting time distribution function, which takes in to account new customers joining the queue. If there are $N-1$ costumers in line and the server is idle, the arriving customer will receive service right away, otherwise an arrival has to wait. Let $\tilde{\Omega}$ be the state space of the absorption time of a Markov chain,

$$\tilde{\Omega} = (*) \cup \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \cdots\}$$

The absorption state corresponds to the tagged customer will be getting service without waiting. The absorption state is defined as follows

$$(*) = \{(0, N-1)\}.$$

The level state 0 is represented as follows,

$$\bar{0} = \{(0,i); 0 \leq i \leq N-2\}$$

the level state for p where $p \geq 1$ is given by

$$l(p) = \{(p,l,r,k) : p \geq 1\ ;\ 0 \leq l \leq N-1\ ;\ 1 \leq r \leq N\ ;\ 1 \leq k \leq M^{(r)}\}$$

The absorbing Markov chain's transition matrix $\tilde{Q}$ is given by

$$\tilde{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots \\ U_0 & W_0 & 0 & 0 & 0 & 0 & \cdots & \cdots \\ U_1 & W_2 & W_1 & 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & W_3 & W_1 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & 0 & W_3 & W_1 & 0 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \ddots & \ddots \end{pmatrix}$$

entries of the above matrix are as follows,

$$U_0 = [F_{N-2}]^T\ ;\quad U_1 = [[F_{((N-1)N)}]^T, cS_0{}^{(1)} \otimes \alpha, \cdots, cS_0{}^{(N)} \otimes \alpha]$$

$$W_0 = \begin{pmatrix} T+T_0\alpha & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \delta I_{M^{(0)}} & T-\delta I_{M^{(0)}} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \delta I_{M^{(0)}} & T-\delta I_{M^{(0)}} & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & & \vdots & \cdots & \vdots \\ 0 & 0 & \delta I_{M^{(0)}} & T-\delta I_{M^{(0)}} & 0 & & & 0 \\ \vdots & \vdots & \ddots & \ddots & & & & \vdots \\ & & & \ddots & \ddots & \vdots & & \\ 0 & 0 & \cdots & & & \delta I_{M^{(0)}} & T-\delta I_{M^{(0)}} & \end{pmatrix}$$

$$W_2 = \begin{pmatrix} (W_2)_{1,1} & 0 & \cdots & 0 \\ 0 & (W_2)_{2,2} & & \vdots \\ & & \ddots & \\ \vdots & & & (W_2)_{N-2,N-2} \\ 0 & \cdots & & 0 \end{pmatrix}$$

where $(W_2)_{i,i}$ $for\ 0 \leq i \leq N-2$, $(W_2)_{i,i} = \begin{pmatrix} cS_0{}^{(1)} \otimes \alpha \\ \vdots \\ cS_0{}^{(N)} \otimes \alpha \end{pmatrix}$

$$W_1 = \begin{pmatrix} (W_1)_{0,0} & 0 & \cdots & 0 \\ 0 & (W_1)_{1,1} & \cdots & 0 \\ 0 & 0 & (W_1)_{2,2} & \\ \vdots & \vdots & & \ddots \\ 0 & 0 & & (W_1)_{N-1,N-1} \end{pmatrix}, \text{ where}$$

$(W_1)_{i,i} = diag\ \{S^{(r)} + dS_0{}^{(r)}\beta^{(r)}; 1 \leq r \leq N, 0 \leq i \leq N-1\}$

$$W_3 = \begin{pmatrix} (W_3)_{0,0} & 0 & 0 & \cdots & 0 \\ 0 & (W_3)_{1,1} & 0 & \cdots & 0 \\ 0 & 0 & (W_3)_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & & \cdots & & (W_3)_{N-1,N-1} \end{pmatrix},$$

with $(W_3)_{i,i} = \begin{pmatrix} 0 & \cdots & 0 & cS_0{}^{(1)} \otimes \beta^{(N)} \\ \vdots & \ddots & \vdots & \\ 0 & \cdots & 0 & cS_0{}^{(N)} \otimes \beta^{(N)} \end{pmatrix}$.

We begin by calculating the system's state, stationary probability distribution (that is, how many customers were in the system) as observed by the tagged client at the moment of arrival. It is denoted by $y(0) = (y_0(0), y_1(0), y_2(0), \cdots)$ and the system's state conditional probability distribution under tagged customer's arrival can be used to determine it, using $x(0) = (x_0(0), x_1(0), x_2(0), \cdots)$ by the following method

$$y_0(0) = x_0(0)(\frac{D_1 e_n}{\lambda})$$

$$y_j(0) = x_j(0)(I_{N^2(M^{(1)}+M^{(2)}+\cdots+M^{(N)})} \otimes \frac{D_1 e_n}{\lambda})$$

where λ indicates the basic (fundamental) rate of Markovian arrival process. Now define $y(t) = (y_*(t), y_1(t), y_2(t), y_3(t), \cdots)$, where y_0 is of dimension $(1 \times NM^{(0)})$ and $y_i(t)$ for $i \geq 1$ is of dimension $(1 \times (NM^{(1)} + NM^{(2)} + \cdots + NM^{(N)}))$. The elements of $Y_{(t)}$ represents the probability of the CTMC with generator $\tilde{Q}$ is in the respective state of level i at time t. The probability that the tagged customer is in the absorbing state is given by $y_*(t)$. Thus we have $W(t) = y_*(t)$, for all $t \geq 0$. From the differential equation $y'(t) = y(t)\tilde{Q}$ we have,

$$y'_*(t) = \sum_{j=0}^{1} y_i(t) U_j$$

$$y'_0(t) = y_0(t)W_0 + y_1(t)W_2$$
$$y'_i(t) = y_i(t)W_1 + y_{i+1}(t)W_3, \quad for \;\; i \geq 1.$$

The row vector $\psi(s)$ provides the Laplace-Steeltjes transform (LST) of the first passage through level 1. By Neuts,M.F in [3], we get

$$\psi(s) = \sum_{i=1}^{\infty} y_i(0)[(sI - W_1)^{-1} W_3]^{i-1}$$

We use $\varphi(i, s)$ to represent the LST of the absorbing time to the state $\{*\}$ when the process begins at level $i = 0, 1$. Using $\tilde{Q}$ we have,

$$\varphi(0, s) = (sI - W_0)^{-1} U_0 \tag{8}$$

$$\varphi(1, s) = (sI - W_1)^{-1} W_2 \varphi(0, s) + (sI - W_1)^{-1} U_1.$$

Consequently, the LST for the waiting time distribution is

$$\tilde{W}(s) = y_0(0)\varphi(0, s) + \psi(s)\varphi(1, s). \tag{9}$$

7. Numerical Results

In this section, we use graphical representations of the numerical values to investigate the model's nature. Where the numerical values for arrival process, admission period and service process were referred by Chakravarthy in [21].
Numerical values for Markovian arrival process,

- Exponential Arrival (E-A)
$$D_0 = (-1), \quad D_1 = (1)$$
- Erlang Arrival (Er-A)
$$D_0 = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}$$

- Hyper-Exponential Arrival (Hyp-A)

$$D_0 = \begin{pmatrix} -1.90 & 0 \\ 0 & -0.19 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1.710 & 0.190 \\ 0.171 & 0.019 \end{pmatrix}$$

- MAP-Negative Correlation Arrival (MNC-A)

$$D_0 = \begin{pmatrix} -1.00243 & 1.00243 & 0 \\ 0 & -1.00243 & 0 \\ 0 & 0 & -225.797 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.01002 & 0 & 0.99241 \\ 223.539 & 0 & 2.258 \end{pmatrix}$$

- MAP-Positive Correlation Arrival (MPC-A)

$$D_0 = \begin{pmatrix} -1.00243 & 1.00243 & 0 \\ 0 & -1.00243 & 0 \\ 0 & 0 & -225.797 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.99241 & 0 & 0.01002 \\ 2.258 & 0 & 223.539 \end{pmatrix}$$

Numerical values for Phase type admission period.

- Exponential admission period (E-AP)

$$\alpha = (1), \quad T = (-1)$$

- Erlang admission period (Er-AP)

$$\alpha = (-1, \quad 0), \quad T = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}$$

- Hyper-Exponential admission period (Hyp-AP)

$$\alpha = (0.8, \quad 0.2), \quad T = \begin{pmatrix} -2.80 & 0 \\ 0 & -0.28 \end{pmatrix}$$

We assume that the numerical values for Phase type distributions for service times to m customers where $1 \leq m \leq N$ are all of exponential distributions. That is all of $(\beta^m, S^(m))$ are exponential distributions irrespective of size. In all the examples we assumed, the arrival rate $\lambda = 1$, the admission period rate $\eta = 3$, the service rate $\gamma = 6$.
The numerical value of the service time is taken as

- Exponential (E)

$$\beta^m = (1), \quad S^m = (-1) \ \forall \ 1 \ \leq \ m \ \leq \ N$$

Illustrative Example: 8.1.

We have illustrated the effect of the rate of renege in counter to the mean size of the system in the upcoming figures 2 and 3. We assume $\lambda = 1$, $\eta = 3$, $\gamma = 6$, $b = 0.5$, $c = 0.6$, $d = 0.4$ and we amplify the renege rate such that the values leaves the system to be stable. We execute the example for batch size $N = 2, 3, 4$.

In Figure 2 we fixed the arrival to follow exponential distribution and we assume the admission periods to follow exponential, Erlang and hyper-exponential distribution respectively. We observed that by amplifying the renege rate the system size decreases. We also noticed that the system size decreases slowly in exponential and Erlang whereas quickly in hyper-exponential distribution.

In Figure 3 we fixed the arrival to follow Erlang distribution and we assume the admission periods to follow exponential, Erlang and hyper-exponential distribution respectively.

We observed that by amplifying the renege rate the system size decreases. We also noticed that the system size decreases moderately in all three admission periods.

Illustrative Example: 8.2.

We have analysed the hyper-exponential arrival with exponential admission period case in the following figures 4 and 5. We assume $\lambda = 1$, $b = 0.5$, $c = 0.6$, $d = 0.4$ and increase the renege rate, admission period rate and service rate such that the values leaves the system to be stable. We execute the example for batch size $N = 2, 3, 4$.

In Figure 4 we fixed the service rate as $\gamma = 6$ and we amplify both the admission period rate and the renege rate against the mean system size. We observed that by amplifying the renege rate and admission period rate the mean system size decreases gradually.

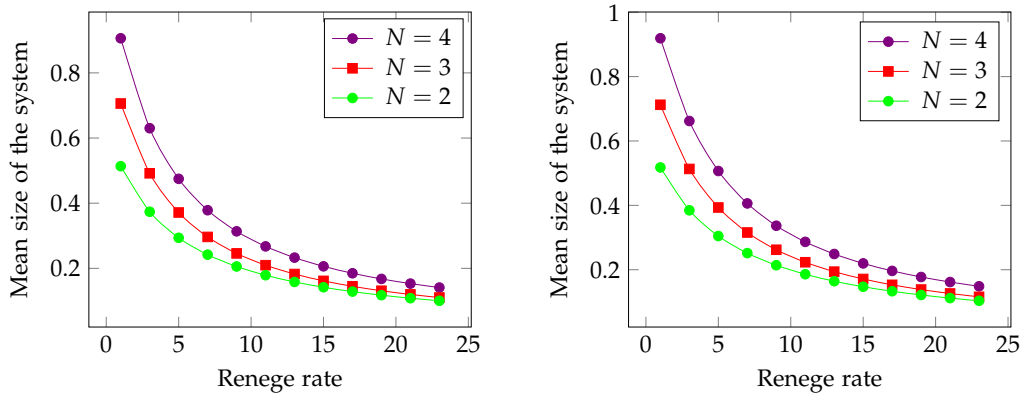
In Figure 5 we fixed the admission period rate as $\eta = 3$ and we amplify both the rate of service and the renege against the mean size of the system. We observed that by amplifying the rate of renege and service the mean size of the system decreases and it falls down rapidly.

Illustrative Example: 8.3.

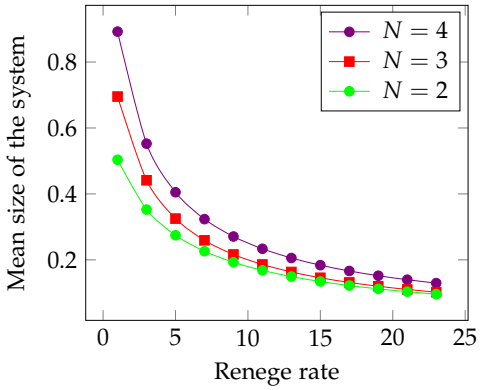
We have analysed the MAP-Positive Correlation Arrival with exponential admission period case in the following figures 6 and 7. We assume $\lambda = 1$, $b = 0.5$, $c = 0.6$, $d = 0.4$ and increase the renege rate, admission period rate and service rate such that the values leaves the system to be stable. We execute the example for batch size $N = 2, 3, 4$.

In Figure 6 we fixed the service rate as $\gamma = 6$ and we amplify both the admission period rate and the renege rate against the mean system size. We observed that by amplifying the renege rate and admission period rate the mean system size decreases slowly.

In Figure 7 we fixed the admission period rate as $\eta = 3$ and we amplify both the rate of service and renege against the mean size of the system. We observed that by amplifying the rate of renege and service, the mean size of the system decreases and it falls down moderately.

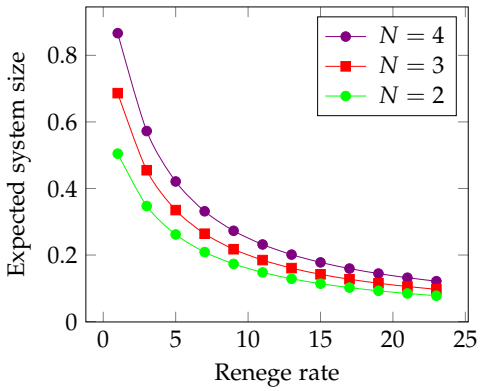


(a) *E-A and E-AP* **(b)** *E-A and Er-AP*

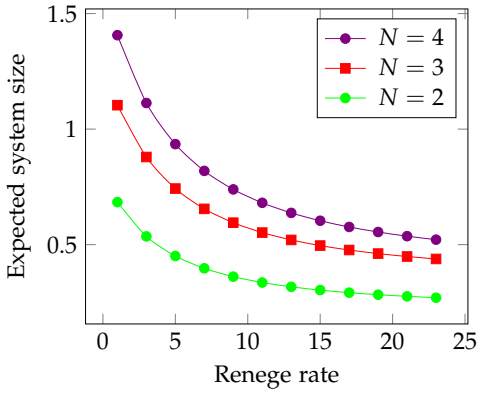


(c) *E-A and Hyp-AP*

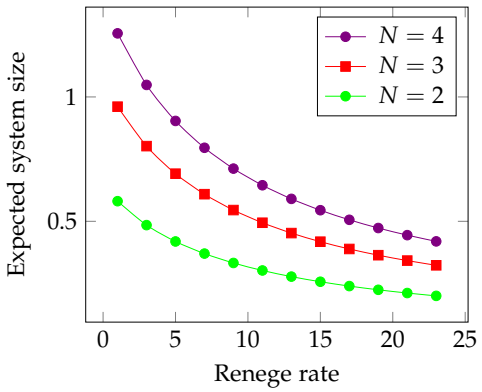
Figure 2: *Renege rate (vs) Expected system size -Exponential Arrival*



(a) *Er-A and E-AP*



(b) *Er-A and Er-AP*



(c) *Er-A and Hyp-AP*

Figure 3: *Renege rate (vs) Expected system size -Erlang Arrival*

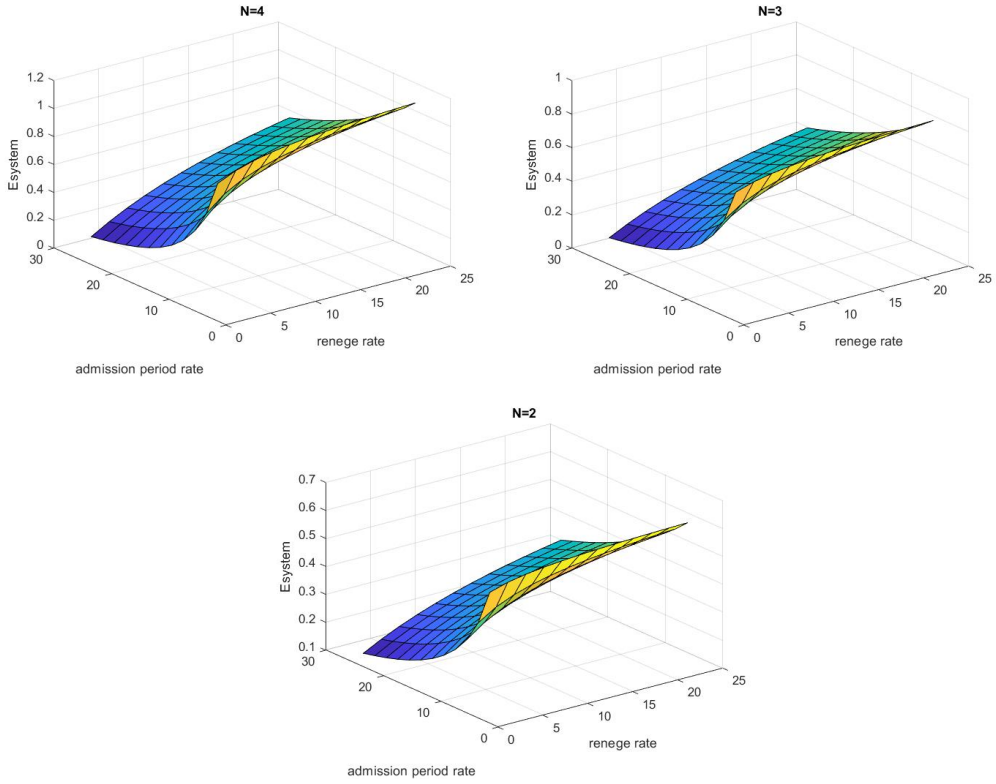


Figure 4: *(Renege rate(δ) and Admission period rate(η) (vs) Esystem)*

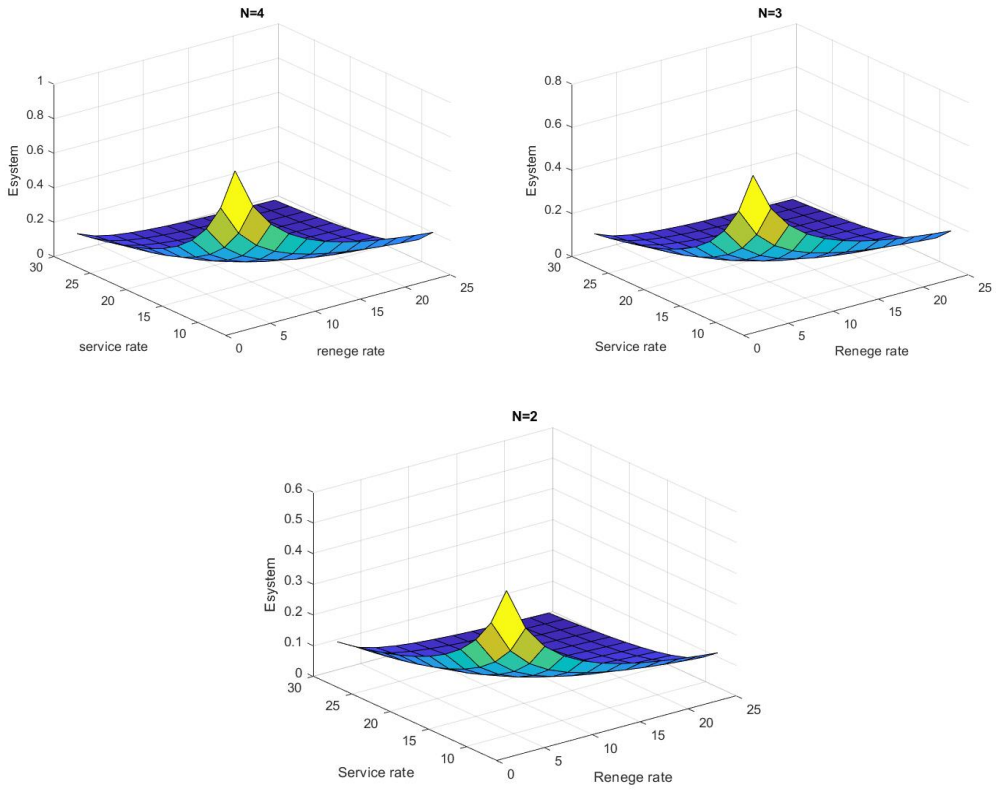


Figure 5: *(Renege rate(δ) and Service rate(γ) (vs) Esystem) [Hyper exponential arrival with Exponential Admission period]*

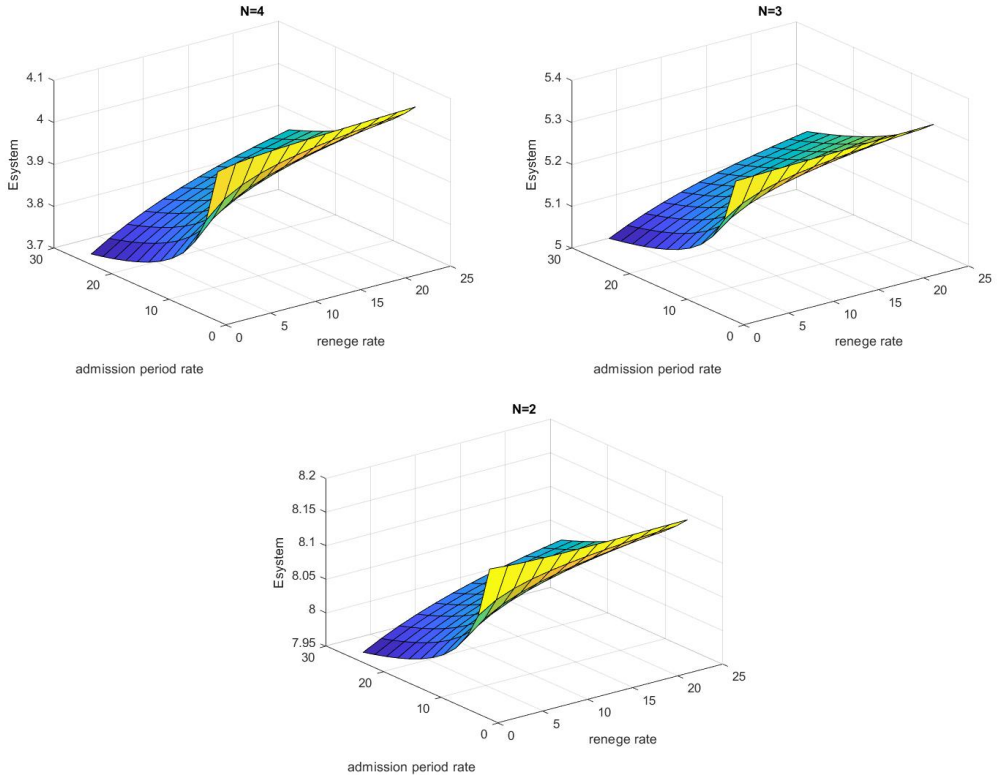


Figure 6: *(Renege rate(δ) and Admission period rate(η) (vs) Esystem)*

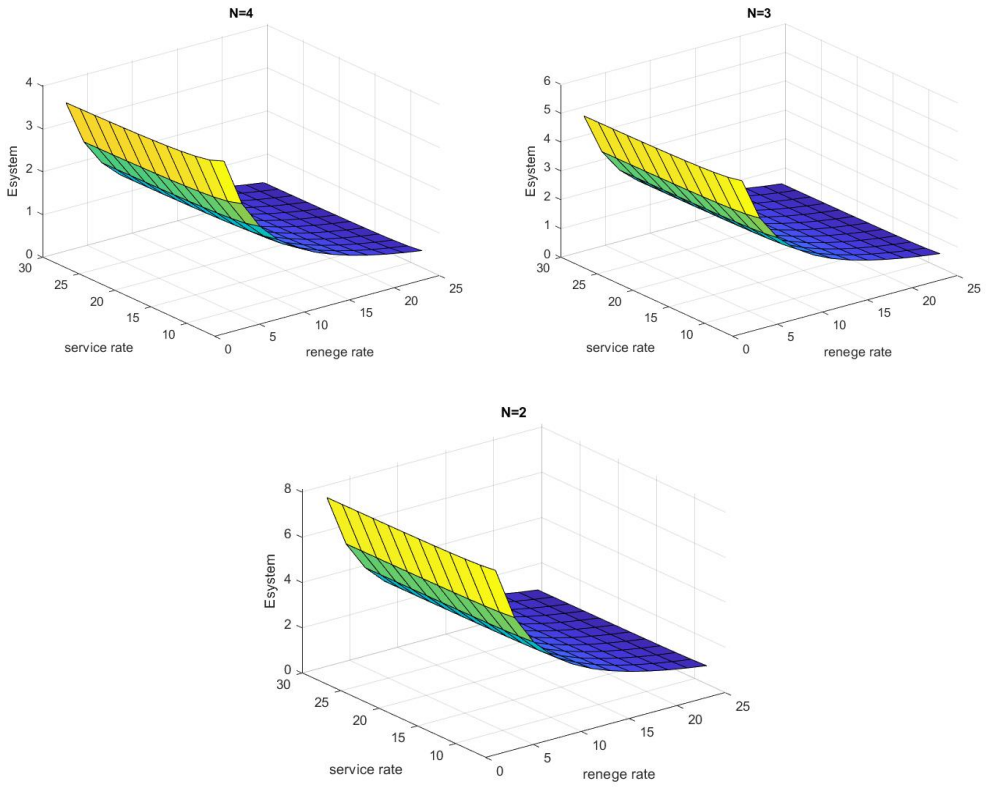


Figure 7: *(Renege rate(δ) and Service rate(γ) (vs) Esystem) [MAP Positive correlation Arrival with Exponential Admission period]*

8. The conclusion

In this paper we studied a group service queueing model with arrivals happen according to a Markovian arrival process in which arrivals may balk or renege the system. The service follows Phase-type distributions in which size of the group may vary and on depending the size of the group, that is, the number of customers getting service, the service time owns different Phase-type distribution representations. If any group of customers would like to receive feedback service, they will receive it immediately. The busy period analysis was done and waiting time distribution was computed. Using the Numerical values of arrival and service times, we compared the mean size of the system counter to renege rate with different batch sizes , which is represented graphically. This model can be extended with various catastrophes on servers, which is currently being probed.

References

[1] Arianna Brugno, Ciro D Apice, Alexander Dudin, Rosanna Manzo, (2017). Analysis of an $MAP/PH/1$ queue with Flexible Group Service. *International Journal of Applied Mathematics and Computer Science*, 27:119-131.

[2] Latouche,G. and Ramaswami,V. (1993). A Logarithmic Reduction Algorithm for Quasi-Birth-Death process. *Journal of Applied Probability*, 30: 650-674

[3] Neuts, M.F. Matrix-Geometric solutions in stochastic models - An Algorithmic Approach: 2nd ed., Dover Publications Inc., New york,1994.

[4] Bailey, N. (1954). On queueing processes with bulk service, *Journal of the Royal Statistical Society* B 16(1): 80-87.

[5] Chaudhry, M.L.; Templeton, J.G.C. A First Course in Bulk Queues: John Wiley and Sons; NY, USA, 1983.

[6] Sasikala, S. and Indhira, K. (2016).Bulk service queueing models-a survey,*International Journal of Pure and Applied Mathematics* 106(6): 43-56.

[7] Banerjee, A., Gupta, U. and Chakravarthy, S. (2015). Analysis of a finite-buffer bulk-service queue under Markovian arrival process with batch-size-dependent service, *Computers and Operations Research*. 60: 138-149.

[8] Neuts, M. (1967).A general class of bulk queues with Poisson input.*The Annals of Mathematical Statistics*. 38(3): 759-770.

[9] Ayyappan.G, Thilagavathy.K.(2020) Analysis of MAP/PH/1 Queueing Model with Breakdown, Instantaneous Feedback and Server Vacation.*International Journal of Application and Allied Mathematics*,15(2).

[10] Ayyappan.G, Gowthami.R.(2021) A MAP/PH/1 Queue with Setup time, Bernoulli vacation, Balking, Reneging, Bernoulli Feedback, Breakdown and Repair *Reliability Theory and Applications* 16:191-221.

[11] Downton, F. (1955).Waiting time in bulk service queues,*Journal of the Royal Statistical Society* B 17(2): 256-261.

[12] Chakravarthy, S.R. Rumyantsev, A.(2018). Efficient redundancy techniques in cloud and desktop grid systems using MAP/G/c-type queues.*Open Eng.* 8, 17-31.

[13] Chakravarthy, S. (2001). The batch Markovian arrival process: A review and future work. *V.R.E.A. Krishnamoorthy and Stochastic Processes, Notable Publications Inc., Branchburg, NJ*. 21-29.

[14] Dudin, A., Manzo, R. and Piscopo, R. (2015). Single server retrial queue with adaptive group admission of customers. *Computers and Operations Research* 61: 89-99.

[15] Brugno, A. Dudin, A. and Manzo, R. (2017). Retrial queue with discipline of adaptive permanent pooling.*Appl. Math. Model*. 50: 1-16.

[16] Chakravarthy, S.R. Maity, A. and Gupta, U.C.(2017). An (s,S) inventory in a queueing system with batch service facility. *Ann. Oper. Res.*. 258: 263-283.

[17] Rakesh Kumar and Bhavneet sing soodan (2019) Transient analysis of a single-server queueing system with correlated inputs and renegings *Reliability Theory and Applications* 14:46-54.

[18] Rakesh Kumar and Bhavneet sing soodan (2022) Modelling and Analysis of Cloud Computing Systems using Queuing Models with Correlated Arrivals and Correlated Reneging *IEEE.* https://doi.org/10.1109/ICCT56057.2022.9976500

[19] D Arienzo, M.P. Dudin, A.N. Dudin, S.A.and Manzo, R. (2020). Analysis of a retrial queue with group service of impatient customers. *J. Ambeint. Intell. Humaniz. Comput*. 11: 2591-2599.

[20] Krishna Kumar,B.,Rukumani,R. and Thangaraj,V.(2008). Analysis of MAP/ PH(1); PH(2)/2 queue with Bernoulli vacations. *Journal of Applied Mathematics and Stochastic Analysis*. 396871:1-20.

[21] Chakravarthy,S.R. Markovian arrival process, Wiley Encyclopaedia of Operation Research and Management Science. https://doi.org/10.1002/9780470400531.eorms0499, 2010.

[22] Ayyappan.G, Meena.S.(2023) Phase type queueing model of server vacation,repair and degrading service with breakdown, starting failure and close-down *Reliability Theory and Applications* 18:464-483.

A STATISTICAL ANALYSIS OF FRACTIONAL FACTORIAL DESIGN USING A FUZZY PROBABILISTIC APPROACH

Sri Devi, P.
•
Research Scholar, Department of Statistics,
Periyar University, Salem – 636011, Tamil Nadu, India.
sridevista@periyaruniversity.ac.in

Pachamuthu, M.
•
Assistant Professor, Department of Statistics,
Periyar University, Salem – 636 011, Tamil Nadu, India.
pachamuthu@periyaruniversity.ac.in

Abstract

In factorial experiments, treatment combinations increase as the number of factors increases. While handling a large number of factors, many difficulties are encountered. Moreover, mechanical errors like mistaken identification of plots, wrong labeling of treatments, etc., may creep in. To overcome these difficulties, only a fraction of treatment combinations can be tested. This technique is known as fractional replication. The design with fractional replication is known as fractional factorial design (FFD). In FFD, the choice of the fraction of treatment depends on what type of information is sacrificed. Usually, the interactions with higher-order are omitted, and all main effects and two-factor interactions are estimated without loss of information. The procedure for the layout of FFD is closely related to the concept of confounding. The analysis of fractional factorials is similar to the analysis of full factors. FFD is used to reduce treatment combinations by a fraction. FFD plays a significant role when the experiment is too large. When compared to classical designs, FFD yields a cost-benefit relationship. Fuzzy theory is used to deal with the imprecise observations in this design. This paper proposes the statistical analysis of fuzzy fractional factorial design with numerical illustration.

Keywords: Fuzzy Fractional Factorial Design, Fuzzy Sets, Trapezoidal Fuzzy Number, $\alpha-$ cut interval method.

1. Introduction

In a complete factorial, each treatment combination is applied to at least one of the experimental units. In some situations, the total number of treatment combinations is too large. Each factor involves two levels. If there are 8 factors, then there are $2^8 = 256$ plots needed for the experiment. As the number of factors increases, treatment combinations also increase. Sometimes, it is difficult to handle such a big experiment practically. Since the time, experimental material, cost, manpower,

etc., also increases, it is impossible to conduct a complete factorial experiment. Typically, the higher-order interactions are not much significant; also, it is difficult to interpret, and these can be used to estimate the error. The total degrees of freedom for 2^8 design is 255 with 8 main effects and 28 two-factor interactions. However, the error degrees of freedom are quite large (219). In handling such a big experiment, the non-experimental error may also lurk in. The higher-order interactions are ignored. The main effects and the lower order interactions information are obtained by a fraction of the complete factorial experiments. This type of experiment is known as the fractional factorial design (FFD). Sometimes, the observations that correspond to FFD will be imprecise. In this case, the fuzzy sets are used to calculate such a design. The fuzzy sets were developed by Lofti A. Zadeh [17] in 1965. Some of the authors who scrutinized to the relevant study are Holland, C.W., [4] outlined the fractional factorial design with its uses in marketing problems. Cotter, S.C., [3] describes the blocking in a fractional factorial experiment by using incomplete block designs without aliasing. Stolle, D.P. [14] suggests that fractional factorial design is the best alternative for factorial designs where the psycholegal researcher examines the main effects of a large number of factors. Ke, W., et.al., [6] propound an efficient method of selecting blocking two-level fractional factorial designs when some two-factor interactions are non-negligible. By screening the important drugs and drug interactions, the sequential usage of two-level and three-level fractional factorial designs was shown, and also Jaynes, J., et.al. [5] provoked the potential optimal drug dosages through contour plots. Parthiban S and Gajivaradhan P [10] have studied the 2^2 factorial experiment using fuzzy environments and compared the result with various tests. Using 3^{5-2} fractional factorial design with resolution III, Zaluski, D., et.al., [16] discern the effect of 5 cultivation factors at 3 levels of intensity under various weather conditions. Using a three-level three-factor factorial design, Anand, R, and Sridhar, V.G [1] focus FSW interlock lap joint of AA7475-T7 to study the correlation among process parameters. In addition to LiDAR range on the robot's navigation time, Mazen, A, et.al., [8] studies the effect of choice for 5 factors of forward and angular velocities using fractional factorial experiment with resolution V. Qamar, S, et.al., [11] determined the effectiveness of four main factors on the extraction of cannabinoids using scCO2 by half-fractional factorial design and identified the highest yield of cannabinoids provided by the extraction of scCO2at high pressure and temperature. In this study, statistical analysis of FFD using TrFNs with $\alpha-$ interval method was proposed through a numerical example.

2. Preliminaries

2.1 Factorial Design

The factorial experiment is defined as the experiment consisting of two or more factors, each with two or more levels. Each factor with the same number of levels is called symmetrical factorial; otherwise, it is called an asymmetrical factorial experiment.

2.2 Fractional Factorial Design (FFD)

When the number of factors is too large, even at two levels, the treatment combinations are also large. While planning such a big experiment, non-experimental types of errors may occur. So therefore, Finney resorted to fractional replicating the more extensive factorial experiments. The information on the main effects and lower order interactions are obtained only from a fraction of the complete factorial experiment called the Fractional Factorial Experiment or Fractional Replicated Design.

2.3 Features of FFD

The features of FFD are: (i) If the fraction of the factorial design s^n (n - number of factors, s -

levels) is of order $\frac{1}{s^m}$. Then, the fractional design is defined as $\frac{1}{s^m}s^n$, where, $m < n$. (ii) If the key block of the factorial experiments with block size s^{n-m-k}, then $\frac{s^{k-1}}{s-1}$ interactions are confounded (iii) The only one block obtained from the defining relation is taken as the defining relation are not estimable (iv) In FFD, confounding is necessary to reduce the block size, and aliases for interactions are confounded.

2.4 Trapezoidal Fuzzy Number (TrFN)

The TrFN is defined as, if a fuzzy set $\tilde{A} = (a_1, a_2, a_3, a_4)$, then its membership function is stated as;

$$\mu_{\tilde{A}}(x) = \begin{cases} 0; & x < a_1 \text{ or } x > a_4 \\ \dfrac{x - a_1}{a_2 - a_1}; & a_1 < x \le a_2 \\ 1; & a_2 < x \le a_3 \\ \dfrac{a_4 - x}{a_4 - a_3}; & a_3 < x \le a_4 \end{cases} \quad (1)$$

where, $a_1 \le a_2 \le a_3 \le a_4$. A TrFN becomes triangular fuzzy number if it satisfies $a_2 = a_3$. In terms of $\alpha -$ cut interval, TrFN is defined as follows:

$$\tilde{A} = [a_1 + (a_2 - a_1)\alpha, a_4 - (a_4 - a_3)\alpha]; \qquad 0 \le \alpha \le 1,$$

where, $a_1 \le a_2 \le a_3 \le a_4$.

3. Methodology

3.1 One-Half Fraction of the 2^5 Design

Consider five factors each at two levels, that is, $2^{5-1} = 32$ treatment combinations. The treatment seems quite large. This leads to a one-half fraction of 2^5 design. We select the 16 treatment combinations as a, b, c, d, e, abc, acd, abd, bcd, abe, ace, bce, ade, bde, cde, $abcde$ as one-half fraction. The signs in the Yates' table [13] is derived by writing down first the five main effects and then forming the interactions of those effects by using + and -. A further process gives the interaction between the five factors. From the Yates' sign table of 2^5, linear combinations for the estimate of the main effects and the interactions for one-half fraction of 2^{5-1} design are

$$l_A = \frac{1}{8}(abc + a - b - bcd - d + abd + acd - c - e + abe + ade - bce + ace - cde - bde + abcde);$$

$$l_B = \frac{1}{8}(b - a - c + abc - d + abd - acd + bcd - e + abe - ace + bde - ade + bce - cde + abcde);$$

Similarly, other effects can be calculated. The formal expressions of the interactions worth nothing. In terms of treatment combinations, the expressions given by the ordinary rules of algebra. Thus, $l_I = l_{ABCDE}$, $l_A = l_{BCDE}$, $l_B = l_{ACDE}$, $l_C = l_{ABDE}$, $l_D = l_{ABCE}$, $l_E = l_{ABCD}$, $l_{AB} = l_{CDE}$, $l_{AC} = l_{BDE}$, $l_{AD} = l_{BCE}$, $l_{AE} = l_{BCD}$, $l_{BC} = l_{ADE}$, $l_{BD} = l_{ACE}$, $l_{BE} = l_{ACD}$, $l_{CD} = l_{ABE}$, $l_{CE} = l_{ABD}$, $l_{DE} = l_{ABC}$, also it is impossible to differentiate A and $BCDE$, B and $ACDE$, and so on. In fact, by estimating main effects and two-factor interactions, actually estimates $A + BCDE$, $B + ACDE$, etc., Two are more effects that have this property are called aliases. The alias structure of this design can be determined

by defining the relation $I = ABCDE$. Multiplying any effect by using the defining relation yields the alias structure of that effect. The alias structure of A is $A.I = A.ABCDE = A^2BCDE$, the square of any column is just identity I, that is, $A + BCDE$. Similarly, the alias structure for other main effects and interaction effects are determined. In this one-half fraction, $I = ABCDE$ is called as the principal fraction. Suppose, considering the other one-half fraction with the minus sign in $ABCDE$ column. The defining relation is $I = -ABCDE$. By this type of fraction, the aliases $A - BCDE$, $B - ACDE$, and so on.

3.2 Statistical Analysis of 2^{5-1} FFD

Consider the Randomised Block Design (RBD) linear model for 2^{5-1} factorial design,

$$y_{ij} = \mu + \tau_i + b_j + e_{ij}; \qquad i = 1,...,t; j = 1,...,r \tag{2}$$

where μ is the general mean effect, τ_i is the fixed effect due to i^{th} treatment, b_j is the fixed effect due to j^{th} replicate, e_{ij} is the random error effect. From the Yates' table of 2^5 design (table 3.1),

$$A = (abc - b - c + a - d + acd + abd - bcd - e + ace + abe - bce + ade - cde - bde + abcde);$$

$$B = (b - c + abc - a - d + abd - acd + bcd - e + bce + abe - ace - ade - cde + bde + abcde);$$

Similarly, other effects can be calculated. Since, the higher order interactions are negligible. The sum of squares sare $SS_A = \frac{[A]^2}{N}$; $SS_B = \frac{[B]^2}{N}$; $SS_C = \frac{[C]^2}{N}$; $SS_D = \frac{[D]^2}{N}$; $SS_E = \frac{[E]^2}{N}$; $SS_{AB} = \frac{[AB]^2}{N}$; $SS_{AC} = \frac{[AC]^2}{N}$; $SS_{AD} = \frac{[AD]^2}{N}$; $SS_{AE} = \frac{[AE]^2}{N}$; $SS_{BC} = \frac{[BC]^2}{N}$; $SS_{BD} = \frac{[BD]^2}{N}$; $SS_{BE} = \frac{[BE]^2}{N}$; $SS_{CD} = \frac{[CD]^2}{N}$; $SS_{CE} = \frac{[CE]^2}{N}$ and $SS_{DE} = \frac{[DE]^2}{N}$ each with 1 degrees of freedom (*df*), where, $N = rt$.

The sum of squares for replications, total and errors are $SS_R = \frac{1}{t}\sum_{j=1}^{r} y_{.j}^{\ 2} - \frac{G^2}{N}$ with $(r-1)$ *df*, $SS_T = \sum\sum y_{ij}^{\ 2} - \frac{G^2}{N}$ with $(rt-1)$ *df* and $SS_{Er} = SS_T - [SS_R + SS_A + SS_B + SS_C + ... + SS_{DE}]$ respectively. All these values will be filled in the ANOVA table (table 3.1).

Table: 3.1 *ANOVA table for* 2^{5-1} *FFD*

SV	df	SS	MSS	F – Ratio
Replicates	$r-1$	SS_R	$\frac{SS_R}{r-1}$	$F_R = \frac{SS_R}{SS_{Er}}$
Treatments	$t-1$	-	-	-
Main effect A	1	SS_A	SS_A	$F_A = \frac{SS_A}{SS_{Er}}$
Main effect B	1	SS_B	SS_B	$F_B = \frac{SS_B}{SS_{Er}}$
Main effect C	1	SS_C	SS_C	$F_C = \frac{SS_C}{SS_{Er}}$
Main effect D	1	SS_D	SS_D	$F_D = \frac{SS_D}{SS_{Er}}$
Main effect E	1	SS_E	SS_E	$F_E = \frac{SS_E}{SS_{Er}}$
Interaction effect AB	1	SS_{AB}	SS_{AB}	$F_{AB} = \frac{SS_{AB}}{SS_{Er}}$
Interaction effect AC	1	SS_{AC}	SS_{AC}	$F_{AC} = \frac{SS_{AC}}{SS_{Er}}$

Interaction effect AD	1	SS_{AD}	SS_{AD}	$F_{AD}=\frac{SS_{AD}}{SS_{Er}}$
Interaction effect AE	1	SS_{AE}	SS_{AE}	$F_{AE}=\frac{SS_{AE}}{SS_{Er}}$
Interaction effect BC	1	SS_{BC}	SS_{BC}	$F_{BC}=\frac{SS_{BC}}{SS_{Er}}$
Interaction effect BD	1	SS_{BD}	SS_{BD}	$F_{BD}=\frac{SS_{BD}}{SS_{Er}}$
Interaction effect BE	1	SS_{BE}	SS_{BE}	$F_{BE}=\frac{SS_{BE}}{SS_{Er}}$
Interaction effect CD	1	SS_{CD}	SS_{CD}	$F_{CD}=\frac{SS_{CD}}{SS_{Er}}$
Interaction effect CE	1	SS_{CE}	SS_{CE}	$F_{CE}=\frac{SS_{CE}}{SS_{Er}}$
Interaction effect DE	1	SS_{DE}	SS_{DE}	$F_{DE}=\frac{SS_{DE}}{SS_{Er}}$
Error	$(r-1)(t-1)$	SS_{Er}	$\frac{SS_{Er}}{(r-1)(t-1)}$	-
Total	$rt-1$	SS_T	-	-

Note: Souces of Variation – SV and Degrees of Freedom – df, Sum of Squares – SS and Mean Sum of Squares - MSS.

Inference: If the calculated value is less than the table value, then there is no significant difference between the replications and factors.

3.3 Statistical Analysis of FFD with α – Interval Method

If the sample observations are in the form of TrFNs, that is, if the yield of a particular plot (cell) receives the value in the form of (a,b,c,d), then it is converted to interval model to analyze factorial model using TrFNs $\alpha-$ cut relation

$$\tilde{y}_{ij}=[a_{ij}+\alpha(b_{ij}-a_{ij}),d_{ij}-\alpha(d_{ij}-c_{ij})];\qquad i=1,...,t;j=1,...,r \tag{3}$$

where, $\tilde{y}_{ij}$ is the observation corresponding to i^{th} treatment and j^{th} replicate; $a_{ij}+\alpha(b_{ij}-a_{ij})$ is the lower level of the observed interval in i^{th} treatment and j^{th} replicate; $d_{ij}-\alpha(d_{ij}-c_{ij})$ is the upper level of the observed interval in i^{th} treatment and j^{th} replicate and now, split this expression into two levels (lower level and upper level) as $\tilde{y}_{ij}{}^{L}=a_{ij}+\alpha(b_{ij}-a_{ij})$ and $\tilde{y}_{ij}{}^{L}=a_{ij}+\alpha(b_{ij}-a_{ij}).$

Hypothesis: The null hypothesis $H_0:\mu_1=\mu_2=...=\mu_i$ against alternative hypothesis $H_1:\mu_1\neq\mu_2\neq...\neq\mu_i$. The crisp hypothesis is then converted into the fuzzy hypothesis for lower and upper-level models $H_0{}^{L},H_0{}^{U}:\mu_1{}^{L},\mu_1{}^{U}=\mu_2{}^{L},\mu_2{}^{U}=...=\mu_i{}^{L},\mu_i{}^{U}$ against $H_1{}^{L},H_1{}^{U}:\mu_1{}^{L},\mu_1{}^{U}\neq\mu_2{}^{L},\mu_2{}^{U}\neq...\neq\mu_i{}^{L},\mu_i{}^{U}$.

Lower Level Model (L.L.M): Let the sum of observations in the i^{th} treatment be $\tilde{y}_{i.}{}^{L}=[a_{i.}+\alpha(b_{i.}-a_{i.})]=T_i^{L}$; the sum of observations in the j^{th} block be $\tilde{y}_{.j}{}^{L}=[a_{.j}+\alpha(b_{.j}-a_{.j})]=R_j^{L}$ where, $i=1,...,t;j=1,...,r.$ Then, the grand total is $G^{L}=\sum\sum y_{ij}$. The sum of squares are $SS_A{}^{L}=\frac{[A]^2}{N}$; $SS_B{}^{L}=\frac{[B]^2}{N}$; and similarly other interactions

are also calculated. $SS_R{}^L = \frac{1}{t}\sum_{j=1}^{r}(y_{.j}{}^L)^2 - \frac{\left(G^L\right)^2}{N}$; $SS_T^L = \sum\sum y_{ij}{}^2 - \frac{\left(G^L\right)^2}{N}$ and

$SS_{Er}^L = SS_T^L - \left[SS_R^L + SS_A^L + SS_B^L + SS_C^L + SS_D^L + SS_E^L + SS_{AB}^L + SS_{AC}^L + SS_{AD}^L + SS_{AE}^L + SS_{BC}^L + SS_{BD}^L + SS_{BE}^L + SS_{CD}^L + SS_{CE}^L + SS_{DE}^L\right]$.

All the calculated values are presented in the ANOVA table as in Table: 3.1.

Upper Level Model (U.L.M): Let the sum of observations in the i^{th} treatment be $\tilde{y}_{i.}{}^U = [a_{i.} + \alpha(b_{i.} - a_{i.})] = T_i^U$; the sum of observations in the j^{th} block be $\tilde{y}_{.j}{}^U = [a_{.j} + \alpha(b_{.j} - a_{.j})] = R_j^U$ where, $i = 1,...,t; j = 1,...,r.$ Then, the grand total is $G^U = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}$. The sum of squares are $SS_A{}^U = \frac{[A]^2}{N}$; $SS_B{}^U = \frac{[B]^2}{N}$; and so on;

$SS_R{}^U = \frac{1}{t}\sum_{j=1}^{r}(y_{.j}{}^U)^2 - \frac{\left(G^U\right)^2}{N}$; $SS_T^U = \sum\sum y_{ij}{}^2 - \frac{\left(G^U\right)^2}{N}$ and

$SS_{Er}^U = SS_T^U - \left[SS_R^U + SS_A^U + SS_B^U + SS_C^U + SS_D^U + SS_E^U + SS_{AB}^U + SS_{AC}^U + SS_{AD}^U + SS_{AE}^U + SS_{BC}^U + SS_{BD}^U + SS_{BE}^U + SS_{CD}^U + SS_{CE}^U + SS_{DE}^U\right]$.

All the calculated values are presented in the ANOVA table as in Table: 3.1.

Decision Rule

Lower-Level Model (L.L.M)

If the calculated value is less than the *F* table value, then the null hypothesis is accepted. That is, the effects due to treatments are equal.

Upper-Level Model (U.L.M)

If the calculated value is less than the *F* table value, then the null hypothesis is accepted. That is, the effects due to treatments are equal.

The partial acceptance of the null hypothesis in lower and upper-level models will be considered as null hypothesis is accepted.

3.4. Advantages of Fuzzy Fractional Factorial Design

- It reduces cost and time when compared to other experimental designs.
- It is used to optimize yield with minimum defects.
- It also reduces the non-experimental type of errors when handling a big experiment.

4. Applications

Example 4.1

The following table 4.1 shows the yield of mustard seeds with five fertilizers by investigating a 2^{5-1} design to improve the yield. The five fertilizers were Farm Yard Manure (FYM) (17 Quintel, 25 Quintel), Nitrogen (120 Kg/Ha, 130 Kg/Ha), Phosphorus (40 Kg/Ha, 50 Kg/Ha), Potassium (60 Kg/Ha, 70 Kg/Ha) and Calcium (10 Kg/Ha, 15 Kg/Ha). Test whether there is a significant difference between the factors A - FYM, B – Nitrogen, C – Phosphorus, D – Potassium and E -Calcium or not?

Table: 4.1 *The yield of mustard seeds with five fertilizers*

Treatment Combination	Response 1	Response 2	Response 3
e	(7,9,11,13)	(8,10,11,13)	(8,9,12,14)
a	(9,11,12,14)	(9,11,13,14)	(8,10,12,13)
b	(31,33,35,36)	(30,33,34,35)	(30,32,33,35)
abe	(53,55,56,58)	(52,54,56,57)	(49,51,52,54)
c	(14,15,16,17)	(16,18,19,21)	(13,16,17,19)
ace	(20,22,23,26)	(24,26,27,29)	(21,23,24,25)

bce	(42,45,46,47)	(41,43,46,47)	(42,44,45,47)
abc	(58,60,61,63)	(54,57,58,60)	(55,58,59,62)
d	(7,8,9,10)	(9,12,13,14)	(5,9,10,12)
ade	(11,13,15,16)	(14,16,17,18)	(10,13,15,17)
bde	(30,31,32,34)	(28,29,30,32)	(31,33,35,36)
abd	(50,51,53,54)	(52,54,55,57)	(48,51,52,55)
cde	(14,16,18,20)	(13,16,17,19)	(16,18,19,21)
acd	(20,22,23,25)	(23,26,27,29)	(21,23,24,26)
bcd	(42,44,45,47)	(39,42,43,45)	(42,45,46,47)
abcde	(63,65,67,69)	(61,64,66,68)	(58,61,63,66)

First, the given trapezoidal fuzzy observations are converted into interval data using (3 and are given in the table 4.2.

Table: 4.2 *The interval observations of the TFN data*

Treatment Combination	Response	Treatment Combination	Response
e	$23+5\alpha, 40-6\alpha$	*d*	$21+8\alpha, 36-4\alpha$
a	$26+6\alpha, 41-4\alpha$	*ade*	$35+7\alpha, 51-4\alpha$
b	$91+7\alpha, 106-4\alpha$	*bde*	$89+4\alpha, 102-5\alpha$
abe	$154+6\alpha, 169-5\alpha$	*abd*	$150+6\alpha, 166-6\alpha$
c	$43+6\alpha, 57-5\alpha$	*cde*	$43+7\alpha, 60-5\alpha$
ace	$68+6\alpha, 80-6\alpha$	*acd*	$64+7\alpha, 80-6\alpha$
bce	$125+7\alpha, 141-4\alpha$	*bcd*	$123+9\alpha, 139-5\alpha$
abc	$167+8\alpha, 185-7\alpha$	*abcde*	$185+8\alpha, 203-7\alpha$

Hypothesis H_0^L, H_0^U : There is no significant difference between the factors A (FYM), B (Nitrogen), C (Phosphorus), D (Potassium) and E (Calcium).

Here, the lower-level and upper-level models are calculated separately per the methodology constructed.

Lower-Level Model (L.L.M.)

The effects and sum of squares for the main effects (A, B, C, D, E) and the two-factor interactions ($AB, AC, AD, AE, BC, BD, BE, CD, CE, DE$) of the L.L.M. is calculated and given in the table 4.3.

Table: 4.3 *Effects and sum of squares of the main effects and interactions L.L.M.*

Variable	Estimated Effect	Sum of Squares
A	$\frac{1}{8}\left[70756+532\alpha+\alpha^2\right]$	$\frac{1}{48}\left[70756+532\alpha+\alpha^2\right]$
B	$\frac{1}{8}\left[617796+4716\alpha+9\alpha^2\right]$	$\frac{1}{48}\left[617796+4716\alpha+9\alpha^2\right]$
C	$\frac{1}{8}\left[41616+3672\alpha+81\alpha^2\right]$	$\frac{1}{48}\left[41616+3672\alpha+81\alpha^2\right]$
D	$\frac{1}{8}\left[1444+380\alpha+25\alpha^2\right]$	$\frac{1}{48}\left[1444+380\alpha+25\alpha^2\right]$
E	$\frac{1}{8}\left[144-168\alpha+49\alpha^2\right]$	$\frac{1}{48}\left[144-168\alpha+49\alpha^2\right]$
AB	$\frac{1}{8}\left[36100+380\alpha+\alpha^2\right]$	$\frac{1}{48}\left[36100+380\alpha+\alpha^2\right]$

AC	$\frac{1}{8}\left[256+32\alpha+\alpha^2\right]$	$\frac{1}{48}\left[256+32\alpha+\alpha^2\right]$
AD	$\frac{1}{8}\left[2500-100\alpha+\alpha^2\right]$	$\frac{1}{48}\left[2500-100\alpha+\alpha^2\right]$
AE	$\frac{1}{8}\left[64+112\alpha+49\alpha^2\right]$	$\frac{1}{48}\left[64+112\alpha+49\alpha^2\right]$
BC	$\frac{1}{8}\left[784+504\alpha+81\alpha^2\right]$	$\frac{1}{48}\left[784+504\alpha+81\alpha^2\right]$
BD	$\frac{1}{8}\left[324+252\alpha+49\alpha^2\right]$	$\frac{1}{48}\left[324+252\alpha+49\alpha^2\right]$
BE	$\frac{1}{8}\left[1024-192\alpha+9\alpha^2\right]$	$\frac{1}{48}\left[1024-192\alpha+9\alpha^2\right]$
CD	$\frac{1}{8}\left[1296+216\alpha+9\alpha^2\right]$	$\frac{1}{48}\left[1296+216\alpha+9\alpha^2\right]$
CE	$\frac{1}{8}\left[196-84\alpha+9\alpha^2\right]$	$\frac{1}{48}\left[196-84\alpha+9\alpha^2\right]$
DE	$\frac{1}{8}\left[576+48\alpha+\alpha^2\right]$	$\frac{1}{48}\left[576+48\alpha+\alpha^2\right]$

By screening experiments, the factors with larger effects are considered and their sum of squares is presented in the ANOVA table (table 4.4).

Table: 4.4 *ANOVA for L.L.M.*

SV	df	SS	MSS	F – Ratio
Replications	2	$\frac{1}{48}[402-102\alpha+200\alpha^2]$	$\frac{1}{2(48)}[402-102\alpha+200\alpha^2]$	$\frac{41(402-102\alpha+200\alpha^2)}{2(-12287+32215866\alpha+883\alpha^2)}$ $<1\forall 0\le\alpha\le 1$
Main effect A	1	$\frac{1}{48}\left[70756+532\alpha+\alpha^2\right]$	$\frac{1}{48}\left[70756+532\alpha+\alpha^2\right]$	$\frac{41(70756+532\alpha+\alpha^2)}{(-12287+32215866\alpha+883\alpha^2)}$ $<1\forall 0\le\alpha\le 1$
Main effect B	1	$\frac{1}{48}\left[617796+4716\alpha+9\alpha^2\right]$	$\frac{1}{48}\left[617796+4716\alpha+9\alpha^2\right]$	$\frac{41(617796+4716\alpha+9\alpha^2)}{(-12287+32215866\alpha+883\alpha^2)}$ $<1\forall 0.8\le\alpha\le 1$
Main effect C	1	$\frac{1}{48}\left[41616+3672\alpha+81\alpha^2\right]$	$\frac{1}{48}\left[41616+3672\alpha+81\alpha^2\right]$	$\frac{41(9409+194\alpha+\alpha^2)}{(-12287+32215866\alpha+883\alpha^2)}$ $<1\forall 0\le\alpha\le 1$
Interaction effect AB	1	$\frac{1}{48}\left[36100+380\alpha+\alpha^2\right]$	$\frac{1}{48}\left[36100+380\alpha+\alpha^2\right]$	$\frac{41(36100+380\alpha+\alpha^2)}{(-12287+32215866\alpha+883\alpha^2)}$ $<1\forall 0\le\alpha\le 1$
Error	41	$\frac{1}{48}\left[-12287+32215866\alpha+883\alpha^2\right]$	$\frac{1}{41(48)}\left[-12287+32215866\alpha+883\alpha^2\right]$	-
Total	47	$\frac{1}{48}\left[754383-32827242\alpha+1175\alpha^2\right]$	-	-

Note: Souces of Variation – SV and Degrees of Freedom – df, Sum of Squares – SS and Mean Sum of Squares - MSS.

Inference: For replications, the table value is $F_t(2,41)=3.23$ and for the treatments, the table value is $F_t(1,41)=4.08$. When comparing the calculated values with these table values, it is less. Therefore, there is no significant difference between the factors A (FYM), B (Nitrogen), C (Phosphorus), D (Potassium) and E (Calcium).

Upper-Level Model (U.L.M.)

The effects and the sum of squares of the main effects and the two-factor interactions of the U.L.M. is given in table 4.5.

Table: 4.5 *Effects and the sum of squares of the main effects and interactions U.L.M.*

Variable	Estimated Effect	Sum of Squares
A	$\frac{1}{8}\left[85849-3516\alpha+36\alpha^2\right]$	$\frac{1}{48}\left[85849-3516\alpha+36\alpha^2\right]$
B	$\frac{1}{8}\left[585225-3060\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[585225-3060\alpha+4\alpha^2\right]$
C	$\frac{1}{8}\left[54289-2796\alpha+36\alpha^2\right]$	$\frac{1}{48}\left[54289-2796\alpha+36\alpha^2\right]$
D	$\frac{1}{8}\left[361-76\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[361-76\alpha+4\alpha^2\right]$
E	$\frac{1}{8}[1225]$	$\frac{1}{48}[1225]$
AB	$\frac{1}{8}\left[31329-2832\alpha+64\alpha^2\right]$	$\frac{1}{48}\left[31329-2832\alpha+64\alpha^2\right]$
AC	$\frac{1}{8}\left[81-144\alpha+64\alpha^2\right]$	$\frac{1}{48}\left[81-144\alpha+64\alpha^2\right]$
AD	$\frac{1}{8}[961]$	$\frac{1}{48}[961]$
AE	$\frac{1}{8}\left[729+108\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[729+108\alpha+4\alpha^2\right]$
BC	$\frac{1}{8}[289]$	$\frac{1}{48}[289]$
BD	$\frac{1}{8}\left[1+8\alpha+16\alpha^2\right]$	$\frac{1}{48}\left[1+8\alpha+16\alpha^2\right]$
BE	$\frac{1}{8}\left[9+12\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[9+12\alpha+4\alpha^2\right]$
CD	$\frac{1}{8}[14400]$	$\frac{1}{48}[14400]$
CE	$\frac{1}{8}\left[121+44\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[121+44\alpha+4\alpha^2\right]$
DE	$\frac{1}{8}\left[2209-188\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[2209-188\alpha+4\alpha^2\right]$

By screening experiments, the factors with more significant effects are considered, and their sum of squares is presented in the ANOVA table (table 4.6).

Table: 4.6 *ANOVA for U.L.M.*

SV	df	SS	MSS	F – Ratio
Replications	2	$\frac{1}{48}\left[146+108\alpha+24\alpha^2\right]$	$\frac{1}{2(48)}\left[146+108\alpha+24\alpha^2\right]$	$\frac{15(146+108\alpha+24\alpha^2)}{(4871-2820\alpha+652\alpha^2)}$ $<1\forall 0\le\alpha\le 0.6$
Main effect A	1	$\frac{1}{48}\left[85849-3516\alpha+36\alpha^2\right]$	$\frac{1}{48}\left[85849-3516\alpha+36\alpha^2\right]$	$\frac{30(10201-606\alpha+9\alpha^2)}{(4871-2820\alpha+652\alpha^2)}$ $0\le\alpha\le 1$
Main effect B	1	$\frac{1}{48}\left[585225-3060\alpha+4\alpha^2\right]$	$\frac{1}{48}\left[585225-3060\alpha+4\alpha^2\right]$	$\frac{30(585225-3060\alpha+4\alpha^2)}{(4871-2820\alpha+652\alpha^2)}$ $0\le\alpha\le 1$
Main effect C	1	$\frac{1}{48}\left[54289-2796\alpha+36\alpha^2\right]$	$\frac{1}{48}\left[54289-2796\alpha+36\alpha^2\right]$	$\frac{30(54289-2796\alpha+36\alpha^2)}{(4871-2820\alpha+652\alpha^2)}$ $0\le\alpha\le 1$
Interaction effect AB	1	$\frac{1}{48}\left[31329-2832\alpha+64\alpha^2\right]$	$\frac{1}{48}\left[31329-2832\alpha+64\alpha^2\right]$	$\frac{30(31329-2832\alpha+64\alpha^2)}{(4871-2820\alpha+652\alpha^2)}$ $0\le\alpha\le 1$
Interaction effect CD	1	$\frac{1}{48}[14400]$	$\frac{1}{48}[14400]$	$\frac{30(14400)}{(4871-2820\alpha+652\alpha^2)}$

$0 \le \alpha \le 1$

Error	30	$\frac{1}{48}\left[4871-2820\alpha+652\alpha^2\right]$	$\frac{1}{30(48)}\left[4871-2820\alpha+652\alpha^2\right]$	-
Total	47	$\frac{1}{48}\left[766367-12120\alpha+816\alpha^2\right]$	-	-

Note: Souces of Variation – SV and Degrees of Freedom – df, Sum of Squares – SS and Mean Sum of Squares - MSS.

Inference: For replications, the table value is $F_t(2,30)=3.32,$ and for the treatments, the table value is $F_t(1,30)=4.17.$ When comparing the calculated values with these table values, it is high. Therefore, there is a significant difference between the factors A (FYM), B (Nitrogen), C (Phosphorus), D (Potassium) and E (Calcium).

4. Conclusion

This paper proposes a new method of FFD using trapezoidal fuzzy numbers. This method is used to deal with imprecise observations. When compared to classical designs, fuzzy FFD yields a cost-benefit relationship. In the yield of mustard seeds with five-fertilizers FYM (17 Quintel, 25 Quintel), Nitrogen (120 Kg/Ha, 130 Kg/Ha), Phosphorus (40 Kg/Ha, 50 Kg/Ha), Potassium (60 Kg/Ha, 70 Kg/Ha) and Calcium (10 Kg/Ha, 15 Kg/Ha), the hypothesis for the L.L.M. is accepted and U.L.M. is rejected. But it is concluded that there is no significant difference between the factors A (FYM), B (Nitrogen), C (Phosphorus), D (Potassium) and E (Calcium), where the hypothesis is partially accepted. From the applications given, it is proved that the factors can be tested at different values of α. This method can be applied in the agricultural field, the engineering field, the medical field and so on. In the future, this work could be extended to one quarter fraction, asymmetrical factorial experiments and some special types of designs.

References

[1] Anand, R. and Sridhar, V.G. (2021). Microstructure and mechanical properties of interlock friction stir weld lap joint AA7475-T7 using fractional factorial design. *Journal of Mechanical Engineering Science*, 236(1): 318-329.

[2] Box, G.E., Hunter, J.S. and Hunter, W.G. Statistics for Experimenters Design, innovation and Discovery, Wiley Series in Probability and Statistics, 2nd ed., 2005.

[3] Cotter, S.C. (1978). Block Designs for Fractional Factorial Experiments with Replication. *Journal of Royal Statistical Society. Series B (Methodological)*, 40(1): 71-78.

[4] Holland, C.W. and Cravens, D.W. (1973). Fractional Factorial Experimental Designs in Marketing Research. *Journal of Marketing Research*, 10(3): 270–276.

[5] Jaynes, J., Ding, X., Xu, H., Wong, W.K., and Ho, C-M. (2013). Application of fractional factorial designs to study drug combinations. *Statistics in Medicine*, 32(2): 1-12.

[6] Ke, W., Ren, C., and Lu, H. (2007). Selection of blocked two-level fractional factorial designs for agricultural experiments. *19th Annual Conference Proceedings: Conference on Applied Statistics in Agriculture*, 62-73.

[7] Klir, G.J and Yuan, B. Fuzzy Sets and Fuzzy Logic: Theory and Applications, Pearson India Education Services Pvt. Ltd. 2015.

[8] Mazen, A., Faled, M., and Krishnan, M. (2022). Tuning of robot navigation performance using factorial design. *Journal of Intelligent and Robotic Systems*, 150(50): 1-14.

[9] Montgomery, D.C. Design and Analysis of Experiments, Wiley India Pvt. Ltd., 8th edition, 2017.

[10] Parthiban, S and Gajivaradhan, P. (2016). A comparative study of statistical hypothesis test

for 2^2 factorial experiment under fuzzy environments. *Bulletin of Mathematics and Statistics Research*, 4(1): 46-70.

[11] Qamar, S., Torres, Y.J.M., Parekh, H.S and Falconer, J.R. (2022). Fractional factorial design study for the extraction of cannabinoids from CBD-dominant cannabis flowers by supercritical carbon dioxide. *Processes*, 10(93): 1-15.

[12] Robert W. Mee. Resolution V Fractional Factorial Designs, A Comprehensive Guide to Fractional Two-Level Experimentation. Springer, New York, 2009.

[13] Sri Devi, P and Pachamuthu, M. (2022). A Statistical analysis of total confounding for factorial experiments using fuzzy approach. International Journal of Agricultural and Statistical Sciences, 18(2): 653-665.

[14] Stolle, D.P., Robbennolt, J.K., Patry, M. and Penrod, S.D. (2002). Fractional factorial designs for legal psychology. *Behavioral Sciences and the Law*, 20: 5-17.

[15] Zadeh. L.A. (1965). Fuzzy sets. *Information and Control*, 8: 338-353.

[16] Zaluski, D., Dubis, B., Budzynski, W., and Jankowski, K. (2016). Applicability of the 3^{5-2} fractional factorial design in determining the effects of cultivation factors on hullness oat. *Agronomy Journal*, 108(1): 205-218.

ON THE PROPERTIES OF GENERALIZED RAYLEIGH DISTRIBUTION WITH APPLICATIONS

Sule Omeiza Bashiru[1], Ibrahim Ismaila Itopa[2] & Alhaji Modu Isa[3]

•

[1,2]Department of Mathematical Sciences, Prince Abubakar Audu University, Anyigba, Kogi State, Nigeria
[3]Department of Mathematics and Computer Science, Borno State University, Maiduguri, Nigeria
Email: [1]bash0140@gmail.com, [2]tripplei1975@yahoo.com, [3]alhajimoduisa@bosu.edu.ng

Abstract

In this study, a new three-parameter lifetime distribution called the generalized Rayleigh distribution was introduced. The new model is an extension of classical Rayleigh distribution. An extension of density of the generalized Rayleigh distribution was derived from which some of the statistical and mathematical properties were derived. Some mathematical properties of the distribution were presented such as moments, moment generating function, quantile function, survival function, hazard function, reversed hazard function and odd function. The distribution of order statistic was obtained in which the maximum and minimum order statistics were derived. Estimation of the parameters by maximum likelihood method was discussed. Two real-life application of the distribution was presented and the analysis showed the fit and flexibility of the generalized Rayleigh distribution over odd Lindley Rayleigh distribution and Rayleigh distribution. The analysis showed that the generalized Rayleigh distribution is more effective and robust in fitting the data sets.

Keywords: flexibility, transect line, myelogeneous leukemia, odd Lindley Rayleigh, quadratic rank transmutation technique

I. Introduction

Over a long period of time, probability distributions have been established through mathematical and statistical study. The assumed probability model or distributions have a significant impact on the effectiveness of the processes used in a statistical study. As a result, numerous common probability distributions and pertinent statistical techniques are described in the literature. However, there are still a number of issues where the real data set does not fit into any of the conventional or classical probability models. In literature numerous generalized distributions have been developed with common feature of having more parameters. Induction of parameters in existing distribution improves the goodness of fit of the distribution under study and tail properties of a distribution increases.

The Rayleigh distribution is one of the most widely applied probability distributions. The Rayleigh distribution is a special case of Weibull distribution, which was first described in [1]. In areas including project effort loading modeling, survival and reliability analysis, communication theory, physical sciences, technology, diagnostic imaging, applied statistics, and clinical research, the Rayleigh distribution is crucial for modeling and interpreting life-time data. Numerous researchers have developed significant expansions to the Rayleigh distribution in light of its significance and the need to provide this distribution more flexibility.

Numerous expansions of Raleigh distribution have been produced as a result of the significance of Raleigh distribution in numerous disciplines. The generalized Raleigh distribution was proposed

by [2], and its unknown parameters were determined using several estimation techniques. The novel generalization of Rayleigh distribution was developed by [3] by utilizing the conservability approach. The Bayes estimators for the parameter of the Rayleigh distribution using square error and LINEX loss functions were derived by [4]. The transmuted Rayleigh distribution was developed by [5] using the quadratic rank transmutation technique. The transmuted generalized Rayleigh distribution was proposed by [6]. The Weibull Rayleigh distribution was studied by [7]. The parameters estimation of exponentiated Rayleigh based on type II censored data was deliberated by [8]. A new distribution named as Rayleigh–Rayleigh distribution was derived by [9] and motivated by the transformed transformer technique by [10]. In contrast to the Lindley distribution, the Rayleigh distribution, and other generalizations of the Rayleigh distribution, an extension of the Rayleigh distribution was developed by [11] with two parameters having greater flexibility. An extension of the exponentiated Rayleigh distribution known as the Gompertz-exponentiated Rayleigh distribution was proposed by [12] by utilizing the Transformed-Transformer family of distributions' methodology.

This paper proposes a new distribution that generalizes the Rayleigh distribution using the family of distribution proposed by [13]. This motivation behind this work is to improve the flexibility of the Rayleigh distribution to fit varieties of real life data sets arising from different disciplines including unimodal and bimodal shapes. Also, to make the kurtosis more flexible compared to the baseline Rayleigh model, to produce skewness for symmetrical distributions using type I half-logistic family of distributions derived by [13] and bathtub shapes.

II. Methods

2.1 Generalized Rayleigh (GRa) Distribution

In this section, a new continuous probability distribution function known as generalized Rayleigh distribution is derived. Also, some plots of its pdf, cdf, survival function and hazard rate function (hrf) are shown in order to assess the shape of the new distribution in fitting different kinds of data.

Recently, [13] developed the Type I Half-logistic family of distributions with cdf and pdf given as

$$F(x;\lambda,\alpha,\boldsymbol{\beta})=\frac{1-\left[1-\left[H(x;\boldsymbol{\beta})\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[H(x;\boldsymbol{\beta})\right]^{\alpha}\right]^{\lambda}} \tag{1}$$

$$f(x;\lambda,\alpha,\boldsymbol{\beta})=\frac{2\lambda\alpha h(x;\boldsymbol{\beta})\left[H(x;\boldsymbol{\beta})\right]^{\alpha-1}\left[1-\left[H(x;\boldsymbol{\beta})\right]^{\alpha}\right]^{\lambda-1}}{\left[1+\left[1-\left[H(x;\boldsymbol{\beta})\right]^{\alpha}\right]^{\lambda}\right]^{2}} \tag{2}$$

where $\boldsymbol{\beta}$ is the vector of parameters of the baseline distribution.

where $G(x;\boldsymbol{\beta})$ is the cumulative distribution function (cdf) of the baseline distribution with vector of parameter $\boldsymbol{\beta}$.

For $x\geq 0,\alpha,\lambda,\boldsymbol{\beta}\geq 0$, where equations (1) and (2) are the cumulative distribution function and probability density function (pdf) of the family of distributions.

The cdf and pdf of the Rayleigh (Ra) distribution are given respectively as

$$H(x;\theta)=1-e^{-\left(\theta/2\right)x^{2}} \tag{3}$$

$$h(x;\theta)=\theta x e^{-\left(\theta/2\right)x^{2}} \tag{4}$$

To obtain the cdf of the new model, equation (3) is inserted into equation (1) as

$$F(x;\lambda,\alpha,\theta)=\frac{1-\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}} \tag{5}$$

On differentiating equation (5) with respect to x, the pdf of the GRa distribution is obtained which is given as

$$f(x;\lambda,\alpha,\theta)=\frac{2\lambda\alpha\theta x e^{-\left(\theta/2\right)x^2}\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha-1}\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda-1}}{\left[1+\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}\right]^{2}} \tag{6}$$

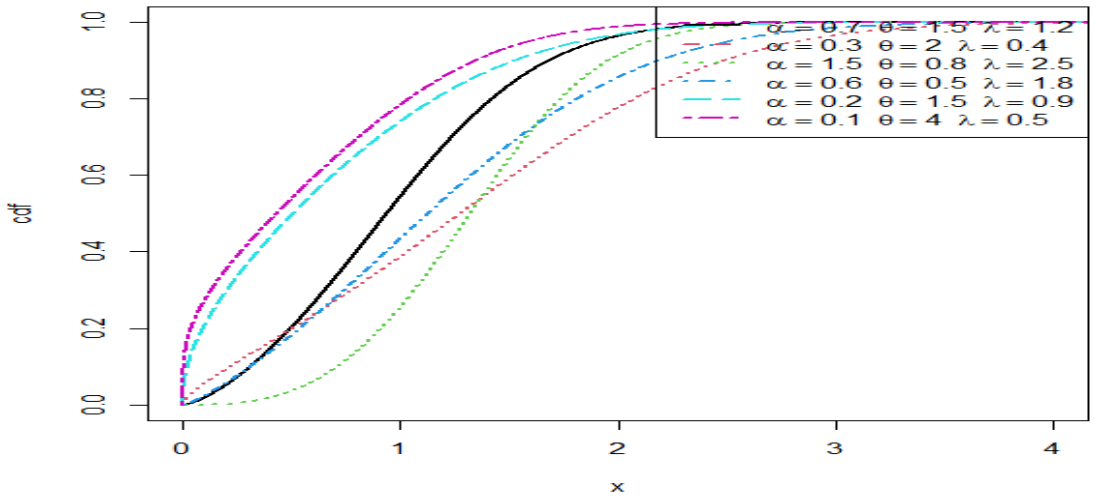


Figure 1: *Plots of cdf of the GRa distribution for different parameter values*

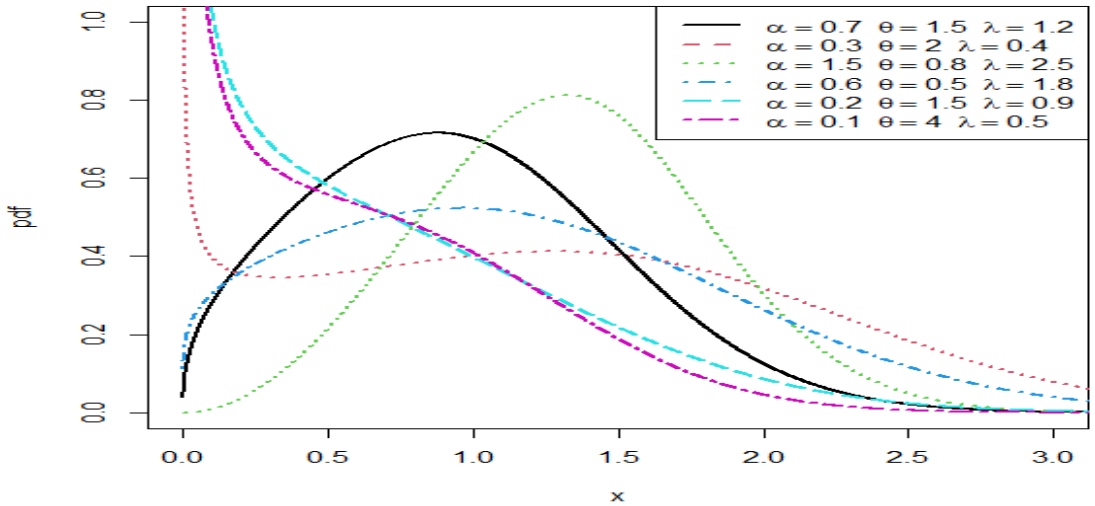


Figure 2: *Plots of pdf of the GRa distribution for different parameter values*

where $x \geq 0$, $\theta > 0$ is the scale parameter and $\alpha, \lambda > 0$ are the shape parameters respectively.

2.1.1 Expansion of density

In this section the pdf in equation (6) is expanded using binomial expansion.
Expanding the last term in equation (6), we have

$$f(x;\lambda,\alpha,\theta)=2\lambda\alpha\theta x e^{-(\theta/2)x^2}\left[1-e^{-(\theta/2)x^2}\right]^{\alpha-1}\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda-1}\left[1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^{-2}$$

$$\left[1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^{-2}=\sum_{i=1}^{\infty}(-1)^i\binom{1+i}{i}\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda i}$$

$$\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda(i+1)-1}=\sum_{j=1}^{\infty}(-1)^j\binom{\lambda(i+1)-1}{j}\left[1-e^{-(\theta/2)x^2}\right]^{\alpha j}$$

$$\left[1-e^{-(\theta/2)x^2}\right]^{\alpha(j+1)-1}=\sum_{k=0}^{\infty}(-1)^k\binom{\alpha(j+1)-1}{k}\left[e^{-(\theta/2)x^2}\right]^{k}$$

$$f(x;\lambda,\alpha,\theta)=2\lambda\alpha\theta x\sum_{i,j,k=0}^{\infty}(-1)^{i+j+k}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\left[e^{-(\theta/2)x^2}\right]^{k+1} \tag{7}$$

Equation (7) is the expansion of equation (6) which will be used to derive some of the properties of the distribution.
Also, equation (5) is expanded as

$$\left[F(x;\lambda,\alpha,\theta)\right]^h=\left[1-\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^h\left[1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^{-h}$$

$$\left[1-\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^h=\sum_{m=0}^{h}(-1)^m\binom{h}{m}\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda m}$$

$$\left[1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^{-h}=\sum_{p=0}^{h}(-1)^p\binom{h+p-1}{p}\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda p}$$

$$\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda(p+m)}=\sum_{z=0}^{\infty}(-1)^z\binom{\lambda(p+m)}{z}\left[1-e^{-(\theta/2)x^2}\right]^{\alpha z}$$

$$\left[1-e^{-(\theta/2)x^2}\right]^{\alpha z}=\sum_{q=0}^{\infty}(-1)^q\binom{\alpha z}{q}\left[e^{-(\theta/2)x^2}\right]^{q}$$

$$\left[F(x;\lambda,\alpha,\theta)\right]^h = \sum_{m,p=0}^{h}\sum_{z,q=0}^{\infty}(-1)^{m+p+z+q}\binom{h}{m}\binom{h+p-1}{p}\binom{\lambda(p+m)}{z}\binom{\alpha z}{q}\left[e^{-(\theta/2)x^2}\right]^q \tag{8}$$

Equation (8) is the expansion of equation (5) which will be used to derive some of the properties of the distribution.

2.1.2 Properties of the Generalized Rayleigh distribution

In this section, some of the mathematical and statistical properties of GRa distribution such as the quantile function, moments, moment generating function, reliability measure, odds function, reversed hazard function and order statistics are derived.

2.1.2.1 Moments

$$E\left(X^r\right) = \int_0^{\infty} x^r f\left(x\right)dx \tag{9}$$

$$= 2\lambda\alpha\theta\sum_{i,j,k=0}^{\infty}(-1)^{i+j+k}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\int_0^{\infty} x^{r+1}\left[e^{-(\theta/2)x^2}\right]^{k+1}dx$$

$$\int_0^{\infty} x^{r+1}\left[e^{-(\theta/2)x^2}\right]^{k+1}dx = \left[\frac{2}{\theta(k+1)}\right]^{\frac{r}{2}}\Gamma\left(1+\frac{r}{2}\right)$$

$$E\left(X^r\right) = 2\lambda\alpha(k+1)\sum_{i,j,k=0}^{\infty}(-1)^{i+j+k}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\left[\frac{2}{\theta(k+1)}\right]^{\frac{r}{2}}\Gamma\left(1+\frac{r}{2}\right) \tag{10}$$

Equation (10) is the moments of GRa distribution. To obtain the mean, we set *r = 1* in equation (10).

2.1.2.2 Moment generating function (mgf)

$$M_{(x)}(t) = \int_0^{\infty} e^{tx} f\left(x\right)dx \tag{11}$$

since the series expansion for e^{tx} is given as

$$e^{tx} = \sum_{w=0}^{\infty}\frac{(tx)^w}{w!}$$

Then, following the method of moments, the mgf is obtained as follows

$$M_{(x)}(t) = 2\lambda\alpha(k+1)\sum_{w=0}^{\infty}\frac{t^w}{w!}\sum_{i,j,k=0}^{\infty}(-1)^{i+j+k}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\left[\frac{2}{\theta(k+1)}\right]^{\frac{w}{2}}\Gamma\left(1+\frac{w}{2}\right) \tag{12}$$

2.1.2.3 Quantile function

Quantile function has a significant position in probability theory and it is the inverse of the cdf. The quantile function is obtained using

$$Q(u) = F^{-1}(u) \tag{13}$$

Using the inverse of equation (5), we have the quantile function of GRa distribution given as

$$x = Q(u) = \frac{1}{\theta}\left[-\log\left[1-\left[1-\left[\frac{1-u}{u+1}\right]^{\frac{1}{\lambda}}\right]^{\frac{1}{\alpha}}\right]^{\frac{1}{2}}\right] \tag{14}$$

The median is obtained by setting $u = 0.5$ in equation (14) given as

$$x_{median} = Q(0.5) = \frac{1}{\theta}\left[-\log\left[1-\left[1-\left[\frac{0.5}{1.5}\right]^{\frac{1}{\lambda}}\right]^{\frac{1}{\alpha}}\right]^{\frac{1}{2}}\right] \tag{15}$$

2.1.2.4 Hazard function

Hazard function is given as

$$\pi(x;\alpha,\lambda,\theta) = \frac{f(x;\alpha,\lambda,\theta)}{S(x;\alpha,\lambda,\theta)} \tag{16}$$

The hazard function of the GRa distribution is given as

$$\pi(x;\alpha,\lambda,\theta) = \frac{\left[\dfrac{2\lambda\alpha\theta x e^{-(\theta/2)x^2}\left[1-e^{-(\theta/2)x^2}\right]^{\alpha-1}\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda-1}}{\left[1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}\right]^{2}}\right]}{1-\left[\dfrac{1-\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}}\right]} \tag{17}$$

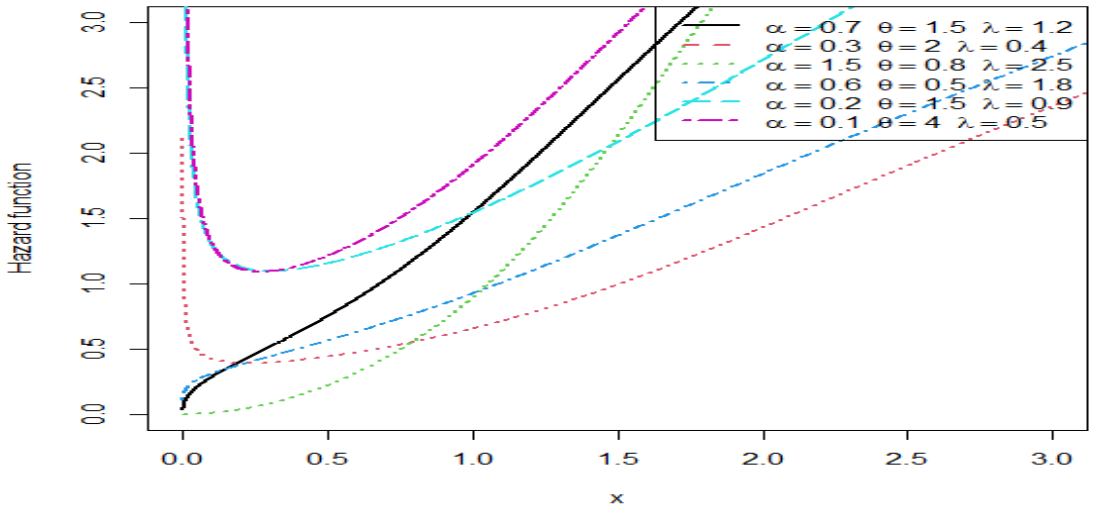


Figure 3: *Plots of hazard function of the GRa distribution for different parameter values*

2.1.2.5 Survival function

The reliability function is also known as survival function, which is the probability of an item not failing prior to some time. It can be defined as

$$S(x;\alpha,\lambda,\theta)=1-F(x;\alpha,\lambda,\theta) \tag{18}$$

The survival function of the GRa distribution is given as

$$S(x;\alpha,\lambda,\theta)=1-\left[\frac{1-\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-(\theta/2)x^2}\right]^{\alpha}\right]^{\lambda}}\right] \tag{19}$$

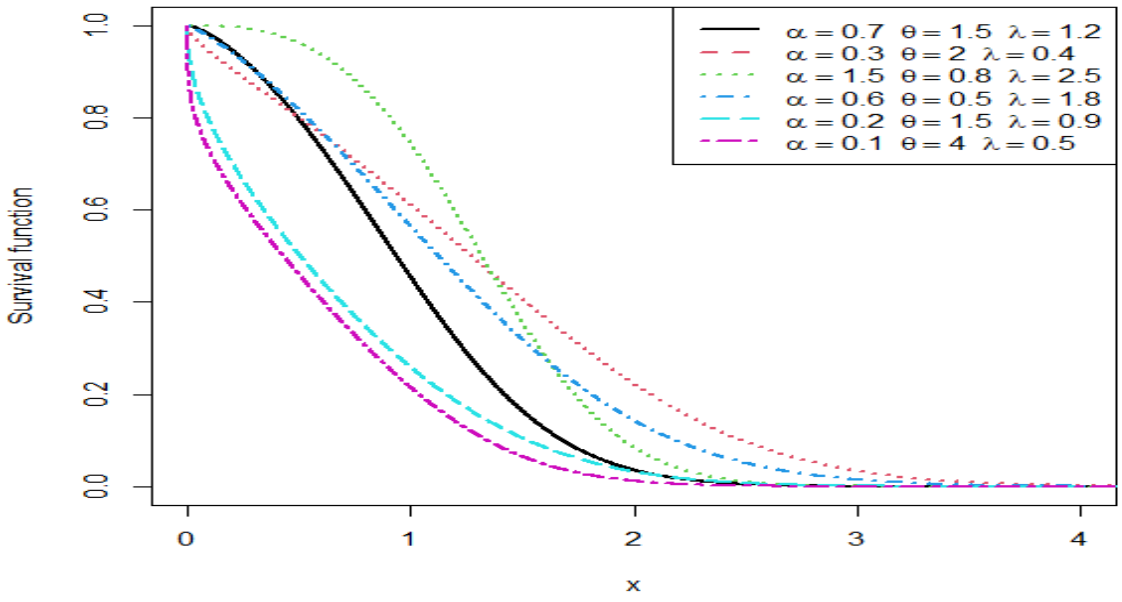


Figure 4: *Plots of survival function of the GRa distribution for different parameter values*

2.1.2.6 Reversed hazard function

Reversed hazard function of a random variable x is given as

$$\Re(x;\alpha,\lambda,\theta)=\frac{f(x;\alpha,\lambda,\theta)}{F(x;\alpha,\lambda,\theta)} \tag{20}$$

The reverse hazard rate function of the GRa distribution is given as

$$\Re(x;\alpha,\lambda,\theta)=\frac{\left[\frac{2\lambda\alpha\theta x e^{-\left(\theta/2\right)x^2}\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha-1}\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda-1}}{\left[1+\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}\right]^2}\right]}{\left[\frac{1-\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}\right]} \tag{21}$$

2.1.2.7 Odds function

The odds function of the GRa distribution is given as

$$\prod(x;\alpha,\lambda,\theta)=\frac{\left[\frac{1-\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}\right]}{1-\left[\frac{1-\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}{1+\left[1-\left[1-e^{-\left(\theta/2\right)x^2}\right]^{\alpha}\right]^{\lambda}}\right]} \tag{22}$$

2.2 Order Statistics

Let $X_1,X_2,...,X_n$ be n independent random variable from the GRa distribution and let $X_{(1)}\le X_{(2)}\le ...\le X_{(n)}$ be their corresponding order statistic. Let $F_{r:n}(x)$ and $f_{r:n}(x)$, $r=1,2,3,...n$ denote the cdf and pdf of the r^{th} order statistics $X_{r:n}$ respectively. The pdf of the r^{th} order statistics of $X_{r:n}$ is given as

$$f_{r:n}(x)=\frac{f(x)}{B(r,n-r+1)}\sum_{v=0}^{n-r}(-1)^v\binom{n-r}{v}F(x)^{v+r-1} \tag{23}$$

The pdf of rth order statistic for the new distribution is obtained by replacing h with v+r-1 in

equation (8) as

$$f_{r:n}(x)=2\lambda\alpha\theta x\frac{1}{B(r,n-r+1)}\sum_{v=0}^{n-r}\sum_{i,j,k=0}^{\infty}\sum_{m,p=0}^{v+r-1}\sum_{z,q=0}^{\infty}(-1)^{i+j+k+m+p+z+q+v}\binom{n-r}{v}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\binom{v+r-1}{m}\binom{v+r+p-2}{p}\binom{\lambda(p+m)}{z}\binom{\alpha z}{q}\left[e^{-(\theta/2)x^2}\right]^{k+q+1} \tag{24}$$

The pdf of minimum order statistic of the GRa distribution is obtained by setting r=1

$$f_{1:n}(x)=2\lambda\alpha\theta xn\sum_{v=0}^{n-1}\sum_{i,j,k=0}^{\infty}\sum_{m,p=0}^{v}\sum_{z,q=0}^{\infty}(-1)^{i+j+k+m+p+z+q+v}\binom{n-1}{v}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\binom{v}{m}\binom{v+p-1}{p}\binom{\lambda(p+m)}{z}\binom{\alpha z}{q}\left[e^{-(\theta/2)x^2}\right]^{k+q+1} \tag{25}$$

Also, the pdf of maximum order statistic of the distribution is obtained by setting r = n

$$f_{n:n}(x)=2\lambda\alpha\theta xn\sum_{m,p=0}^{v+n-1}\sum_{i,j,k=0}^{\infty}\sum_{z,q=0}^{\infty}(-1)^{i+j+k+m+p+z+q+v}\binom{1+i}{i}\binom{\lambda(i+1)-1}{j}\binom{\alpha(j+1)-1}{k}\binom{v+n-1}{m}\binom{v+n+p-2}{p}\binom{\lambda(p+m)}{z}\binom{\alpha z}{q}\left[e^{-(\theta/2)x^2}\right]^{k+q+1} \tag{26}$$

2.3 Estimation method

The method of maximum likelihood estimation (MLE) is used in this section to estimate the parameters of the GRa distribution. For a random sample, $X_1,X_2,...,X_n$ of size n from the GRa distribution (α,θ,λ), the log-likelihood function $L(\alpha,\theta,\lambda)$ of (6) is given as

$$\log(L)=n\log(2)+n\log(\lambda)+n\log(\alpha)+n\log(\theta)-\frac{\theta}{2}\sum_{i=1}^{n}x_i^2+(\alpha-1)\sum_{i=1}^{n}\log\left[1-e^{-(\theta/2)x_i^2}\right]+(\lambda-1)\sum_{i=1}^{n}\log\left[1-\left[1-e^{-(\theta/2)x_i^2}\right]^{\alpha}\right]-2\sum_{i=1}^{n}\log\left[1+\left[1-\left[1-e^{-(\theta/2)x_i^2}\right]^{\alpha}\right]^{\lambda}\right] \tag{27}$$

Differentiating the log-likelihood with respect to λ,α,θ and equating the result to zero, we have

$$\frac{\partial L}{\partial\lambda}=\frac{n}{\lambda}+\sum_{i=1}^{n}\log\left[1-\left[1-e^{-(\theta/2)x_i^2}\right]^{\alpha}\right]-2\sum_{i=1}^{n}\frac{\left[1-\left[1-e^{-(\theta/2)x_i^2}\right]^{\alpha}\right]^{\lambda}\log\left[1-e^{-(\theta/2)x_i^2}\right]^{\alpha}}{1+\left[1-\left[1-e^{-(\theta/2)x_i^2}\right]^{\alpha}\right]^{\lambda}}=0 \tag{28}$$

$$\frac{\partial L}{\partial \alpha}=\frac{n}{\alpha}+\sum_{i=1}^{n}\log\left[1-e^{-\left(\theta/2\right)x_i^2}\right]-\left(\lambda-1\right)\sum_{i=1}^{n}\frac{\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}\log\left[1-e^{-\left(\theta/2\right)x_i^2}\right]}{1-\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}}$$

$$+2\lambda\sum_{i=1}^{n}\frac{\left[1-\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}\right]^{\lambda-1}\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}\log\left[1-e^{-\left(\theta/2\right)x_i^2}\right]}{1+\left[1-\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}\right]^{\lambda}}=0 \tag{29}$$

$$\frac{\partial L}{\partial \theta}=\frac{n}{\theta}-\frac{1}{2}\sum_{i=1}^{n}x_i^2-\left(\alpha-1\right)\sum_{i=1}^{n}\frac{e^{-\left(\theta/2\right)x_i^2}\left(\theta/2\right)x_i^2\left(1/2\right)x_i^2}{1-e^{-\left(\theta/2\right)x_i^2}}+\left(\lambda-1\right)\sum_{i=1}^{n}\frac{\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha-1}e^{-\left(\theta/2\right)x_i^2}\left(\theta/2\right)x_i^2\left(1/2\right)x_i^2}{1-\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}}$$

$$-2\alpha\lambda\sum_{i=1}^{n}\frac{\left[1-\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}\right]^{\lambda-1}\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha-1}e^{-\left(\theta/2\right)x_i^2}\left(\theta/2\right)x_i^2\left(1/2\right)x_i^2}{1+\left[1-\left[1-e^{-\left(\theta/2\right)x_i^2}\right]^{\alpha}\right]^{\lambda}}=0 \tag{30}$$

Now, equations (28), (29) and (30) do not have a simple analytical form and are therefore not tractable. As a result, we have to resort to non-linear estimation of the parameters using iterative method.

II. Results

3.1 Applications

In this section, we present two applications of GRa distribution using different data sets from different fields to demonstrate the flexibility of the distribution in modeling real-life data sets. The data are fitted to the GRa distribution and two other distributions as comparators such as Odd Lindley Rayleigh (OLRa) distribution by [11] and Rayleigh (Ra) distribution. This is done to test the new distribution's flexibility against the comparators. *Adequacy Model* which is a package in R-software, is used to produce the results of the analysis. Using the Akaike information criterion (AIC) and Bayesian information criterion (BIC), respectively, the performance of the distribution was compared to other existing distributions that were consistent with the baseline distribution in terms of providing good parametric fit to the data sets.

$$AIC=-2ll+2k \tag{31}$$

$$BIC=-2ll+k\log(n) \tag{32}$$

The model selection is carried out using the AIC and the BIC. Where ll denotes the log-likelihood function evaluated at the maximum likelihood estimates, k is the number of parameters, and n is the sample size from the data.

The model with minimum value of AIC and BIC is chosen as the best model to fit the data set. The comparators presented are odd Lindley Rayleigh (OLRa) distribution and Rayleigh (Ra) distribution.

The first data set, taken from [14] and also reported in [15], shows the locations of the 68 stakes found while walking L = 1000 m and looking w = 20 m on either side of the transect line. The dimensions are:

2.0, 0.5, 10.4, 3.6, 0.9, 1.0, 3.4, 2.9, 8.2, 6.5, 5.7, 3.0, 4.0, 0.1, 11.8, 14.2, 2.4, 1.6, 13.3, 6.5, 8.3, 4.9, 1.5, 18.6, 0.4, 0.4, 0.2, 11.6, 3.2, 7.1, 10.7, 3.9, 6.1, 6.4, 3.8, 15.2, 3.5, 3.1, 7.9, 18.2, 10.1, 4.4, 1.3, 13.7, 6.3, 3.6, 9.0, 7.7, 4.9, 9.1, 3.3, 8.5, 6.1, 0.4, 9.3, 0.5, 1.2, 1.7, 4.5, 3.1, 3.1, 6.6, 4.4, 5.0, 3.2, 7.7, 18.2, 4.1.

The second data set represents the survival times (in weeks) of 33 patients suffering from acute myelogeneous leukemia. These data have been studied by [16]. The data are:

65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65, 56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43.

Table 1: *The MLEs and Information Criteria of the models based on the first data set*

Model	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\theta}$	$-l$	**AIC**	**BIC**
GRa	0.4774	6.3091	0.0018	185.7156	377.4313	384.0898
OLRa	1.2608	-	0.0114	364.0566	732.1132	736.5522
Ra	-	-	0.0362	202.4675	406.1545	409.1545

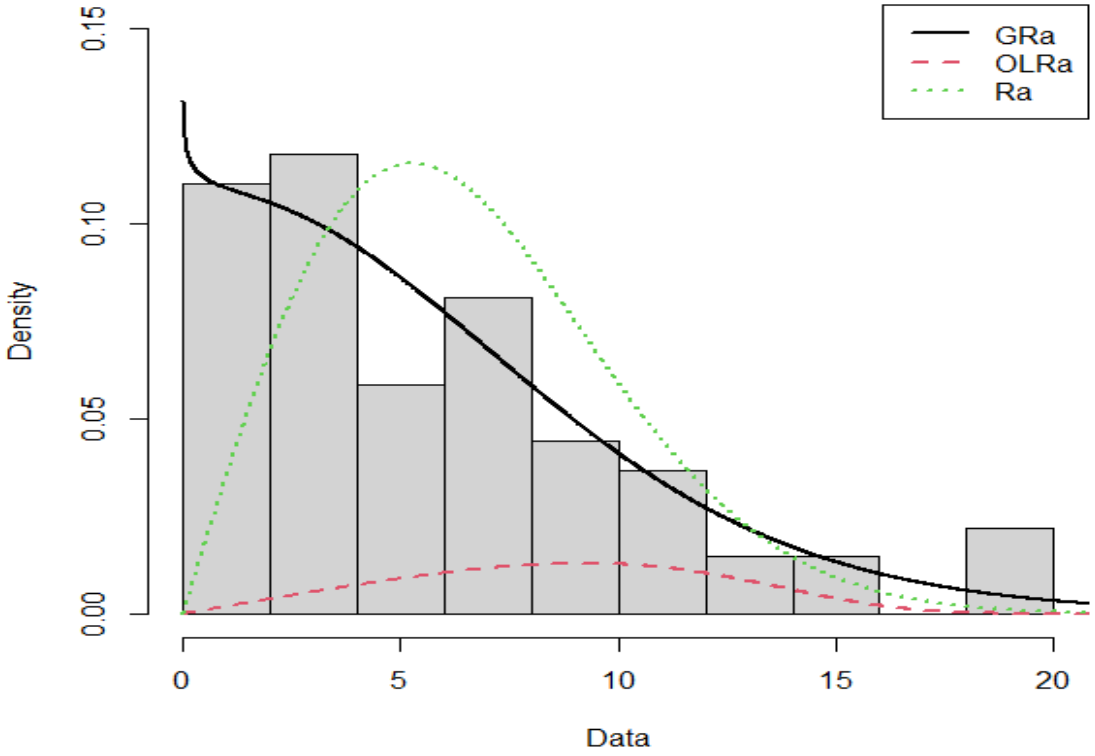


Figure 5: *Histogram and fitted pdfs for the GRa, OLRa and Ra models to the first data set*

Table 2: *The MLEs and Information Criteria of the models based on the second data set*

Model	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\theta}$	$-l$	**AIC**	**BIC**
GRa	0.2293	0.6271	0.0005	158.1467	322.2934	326.7829
OLRa	1.2857	-	0.0001	270.2955	544.5910	547.5840
Ra	-	-	0.0005	188.6356	379.2713	380.7678

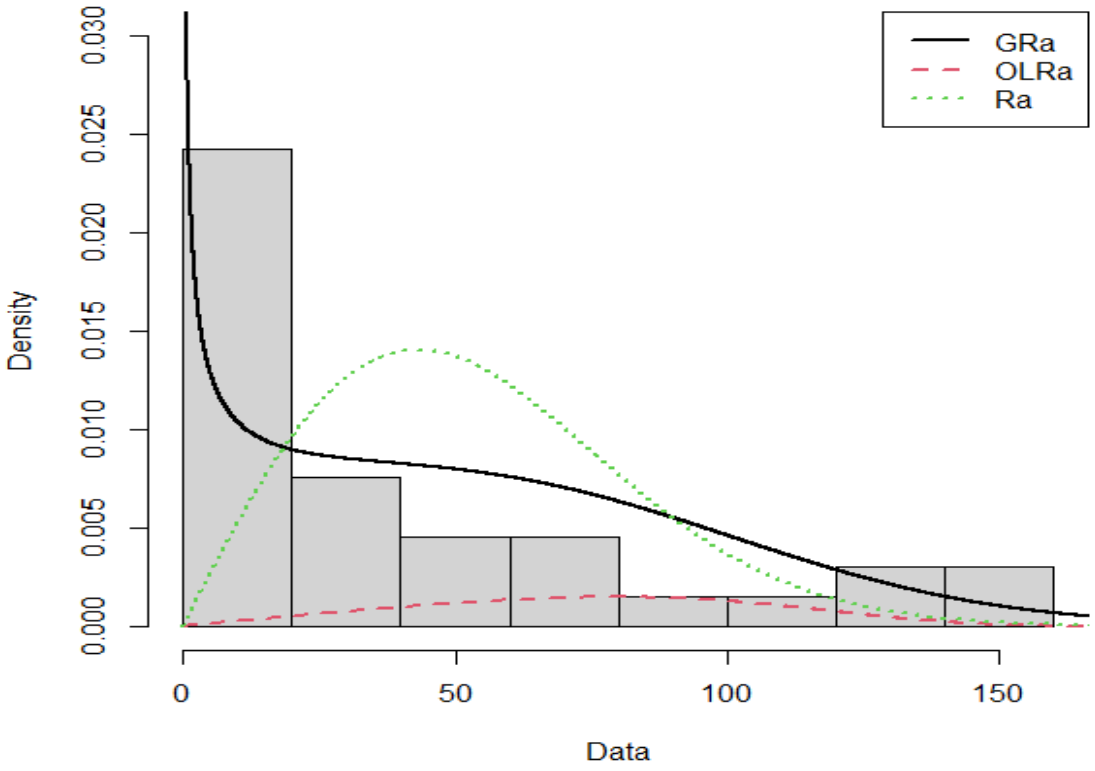


Figure 6: *Histogram and fitted pdfs for the GRa, OLRa and Ra models to the second data set*

IV. Discussion

The estimated values for each parameter and the models' goodness of fits are shown in Tables 1 and 2. AIC and BIC are two metrics for goodness of fits. The model performs better when the AIC and BIC values are lower. Tables 1 and 2 show that the GRa distribution has the lowest AIC and BIC, respectively and this makes the GRa model more adaptable and suitable for handling the data sets.

The new model's forms, fit, and adaptability in connection to the data sets under consideration are shown in Figures 5 and Figure 6. The black line, which represents the GRa model, more closely matched the data's pattern than the competitors. The histogram and fitted plots make it clear that the black line, which represents the GRa distribution, matches the two data sets under consideration better.

This study generalized the Rayleigh distribution by deriving a new continuous distribution known as the generalized Rayleigh distribution. Some of the statistical and mathematical properties of the GRa are obtained such as the survival function, hazard rate function, quantile function, inverted hazard function, odds function, and order statistics from the new distribution. Plotting the pdf and hazard rate function graphs revealed the contours of the suggested distribution. It was discovered that the hazard function is shaped like bathtub. *AdequacyModel* package in R was used to estimate the model parameters using the maximum likelihood method. The generalized Rayleigh distribution and its comparators considered were applied to two real life data sets, and the outcomes are shown in Tables 1 and 2. The findings demonstrated that the GRa distribution is much more potent, robust and superior at fitting the two data sets under consideration. The density graphs in figures 5 and 6 for the two data sets further show how adaptable and robust the new model is.

References

[1] Lord Rayleigh, F. R. S. (1880). On the Resultant of a Large Number of Vibrations of the same Pitch of Arbitrary Phase. *The London, Edinburgh and Dublin Philosophical Magazine and journal of Science*, 10, 73-78

[2] Kundu, D., and Raqab, M. Z. (2005). Generalized Rayleigh distribution: Different methods of estimations. *Computational Statistics and Data Analysis*, 49, 187–200.

[3] Voda, V. G. (2007). A new generalization of Rayleigh distribution. *Reliability: Theory and Applications*, 2(2): 1-10

[4] Dey, S. (2009). Comparison of Bayes estimators of the parameter and reliability function for Rayleigh distribution under different loss functions. *Malaysian Journal of Mathematical Sciences*, 3(2): 249-266.

[5] Merovci, F. (2013). Transmuted Rayleigh distribution. *Austrian Journal of Statistics*, 42, 21–31.

[6] Merovci, F. (2014). Transmuted generalized Rayleigh distribution. *Journal of Statistics Applications and Probability*, 3(1): 9-20.

[7] Merovci, F. and Elbatal, I. (2015). Weibull Rayleigh Distribution: Theory and Applications, *Applied Mathematics and Information Sciences*, 9(4): 2127-2137.

[8] Mahmoud, M., and Ghazal, M. (2017). Estimations from the exponentiated Rayleigh distribution based on generalized Type-II hybrid censored data. *Journal of the Egyptian Mathematical Society*, 25, 71–78.

[9] Ateeq, K., Qasim, T. B. and Alvi, A. R. (2019). An extension of Rayleigh distribution and applications, *Cogent Mathematics and Statistics*, 6:1.

[10] Alzaatreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions, *Metron*, 71(1), 63-79.

[11] Ieren, T. G., Abdulkadir, S. S. and Issa, A. A. (2020). Odd Lindley-Rayleigh Distribution: Its Properties and Applications to Simulated and Real Life Datasets, *Journal of Advances in Mathematics and Computer Science*, 35(1): 63-88.

[12] Abdulsalam, H. A., Yahaya, A. and Dikko, H. G. (2021). On the properties and applications of a new extension of exponentiated Rayleigh distribution, *FUDMA Journal of Sciences*, 5(2): 377-398.

[13] Bello, O. A., Doguwa, S. I., Yahaya, A. and Jibril, H. M. (2020) A type I half Logistic exponentiated-G family of distributions: Properties and application, *Communication in Physical Sciences*, 7(3): 147 – 163.

[14] Patil, G. P., Patil, G. P., and Rao, C. R. (1994). Handbook of statistics 12: *Environmental statistics*

[15] Almetwally, E. M and Meraou, M. A. (2022). Application of Environmental Data with New Extension of Nadarajah-Haghighi Distribution, *Computational Journal of Mathematical and Statistical Sciences*, 1(1): 26-41.

[16] Feigl, P., and Zelen, M. (1965). Estimation of exponential probabilities with concomitant information, *Biometrics*; 21:826-38.

RT&A, No.3 (74)
Volume 18, September 2023

HALF CAUCHY - EXPONENTIAL DISTRIBUTION: ESTIMATION AND APPLICATIONS

K.Jayakumar

•

Department of Statistics, University of Calicut, Kerala -673 635, India
jkumar19@rediffmail.com

Fasna.K

•

Department of Statistics, University of Calicut, Kerala -673 635, India
fasna.asc@gmail.com

Abstract

In this paper, we introduce a new two-parameter distribution called the new Half Cauchy - exponential distribution (HCE) for modeling lifetime data. The structural properties of the new distribution are discussed. Expressions for the quantiles, mode, mean deviation, and distribution of order statistics are derived. The model parameters of HCE distribution are estimated by the method of maximum likelihood, method of least square, method of Cramer-von-Mises, and Anderson-Darling methods. The existence and uniqueness of maximum likelihood estimates are proved. The importance of the new distribution is proved empirically by real-life data set.

Keywords: Half-Cauchy distribution, Method of least-squares, Method of Cramer-von-Mises, Maximum likelihood estimation, T-X family

1. Introduction

The Cauchy distribution named after Augustin Cauchy is a continuous probability distribution and is also known as the Lorentz distribution or Breit-Wigner distribution. It is also the distribution of the ratio of two independent normally distributed random variables with zero mean. Cauchy distribution is unimodal, symmetric, and bell-shaped with much heavier tails than normal distribution. It is also used for the analysis when outliers are presented in the data. Cauchy distribution has received applications in many areas including physics, mathematics, econometrics, engineering, spectroscopy, biological analysis, clinical trials, stochastic modeling of decreasing failure rate life components, queueing theory, and reliability. For more details and discussion, the reader is referred to [20], [3], and [8].
For the Cauchy distribution, the finite moments of order greater than or equal to one do not exist and hence the central limit theorem does not hold. Further, the maximum likelihood estimation of its parameters is not ideal because of no closed-form solution of the likelihood equations. The use of Edgeworth expansion to construct an accurate approximation to the sampling distribution of the maximum likelihood estimator of parameters of Cauchy distribution suggested by [1]. The method of moments is also not possible for this distribution.
Because of these facts, the Cauchy distribution serves as a counterexample for some well-accepted results and concepts in Statistics. This also makes the choice of the Cauchy distribution as an unrealistic model. That is why, modification of Cauchy distribution have been suggested in the

literature to overcome the problem of the moments and other useful properties.
The Half-Cauchy (HC) distribution is the folded standard Cauchy distribution around the origin so that positive values are observed.
A random variable X has the HC distribution with scale parameter $\sigma > 0$, if its cumulative distribution function (cdf) is given by

$$R(x) = \frac{2}{\pi} \arctan\left(\frac{x}{\sigma}\right), \qquad x > 0 \tag{1}$$

The probability density function (pdf) corresponding to 1 is

$$r(x) = \frac{2}{\pi\sigma}\left[1 + \left(\frac{x}{\sigma}\right)^2\right]^{-1}. \tag{2}$$

Although some applications of the half Cauchy distribution exist in the literature, the fact that the finite moments of order greater than or equal to one do not exist, the central limit theorem does not hold. This fact reduces the applicability of this distribution in modeling real-life scenarios. As a heavy-tailed distribution, the HC distribution has been used as an alternative to exponential distribution to model dispersal distances by [4], as the former predicts more frequent long-distance dispersed events than the later. The HC distribution to model ringing data on two species of tits (Parus caeruleus and Parus major) in Britain and Ireland used by [6].

In the real situation, we come across non-normal data sets frequently. One usual way of dealing with non-normal data is to find a suitable transformation that makes the data more normal-like and to apply standard normal-based methods to the transformed data. Finding a suitable transformation can be difficult with data and it is often preferable to work with data without changing the original scale as the easy way of interpretation. These difficulties motivated for more-flexible parametric families of distributions to model non-normal data.

Our focus in this article is on continuous non-normal data. Because real data often deviate from normality in the tails or exhibit asymmetry in the distribution, there has been a growing interest in distributions with additional parameters regulating asymmetry and tails directly. Traditionally, log-normal or gamma distributions are used to model positively skewed data. As a viable and flexible alternative, in this study, we propose Half-Cauchy Exponential (HCE) distribution. In a number of domains such as medical applications, atmospheric sciences, microbiology, environmental science, and reliability theory among others, data are positive and right skewed. The suitable models used by researchers and practitioners to deal with this kind of data are usually parametric distributions such as log-normal, gamma, and Weibull. However, these distributions are not always enough to reach a good fit of the data. This has motivated the interest in the development of more flexible and better-adapted distributions, which have been generated using different strategies as the combination of known distributions.

In the last two decades, there has been an increased interest in defining new generators for univariate continuous distributions to model data in several areas such as engineering, actuarial, medical sciences, biological studies, demography, economics, finance, and insurance. However, in many applied areas like lifetime analysis, finance, and insurance, there is a clear need for extended forms of these distributions, that is, new distributions which are more flexible to model real data. The addition of parameters has been proved useful in exploring skewness and tail properties, and for improving the goodness-of-fit. Thus motivated in to introduce an extended form of HC distribution.
In this paper, we propose a new lifetime model using the technique introduced by [19].
A family of distributions generated by gamma random variables have introduced by [12]. This family of distributions has its cumulative distribution function (cdf) as

$$G(x) = \int_a^{-\ln[1-F(x)]} r(t)dt. \tag{3}$$

Using similiar approach [19] introduced a new family of distributions with cdf given by

$$G(x) = 1 - \int_a^{-\ln[F(x)]} r(t)dt. \tag{4}$$

In this paper the T-X family defined by [19] is used to create the half-Cauchy X family of distribution. Let T be a random variable having HC distribution with pdf , $r(t) = \frac{2}{\pi\sigma}\left[1 + (\frac{t}{\sigma})^2\right]^{-1}, t > 0$. Then, the pdf of the Half Cauchy - X family of distributions from equation (4) is

$$g(x) = \frac{2f(x)}{\pi\sigma F(x)}\left[1 + \left(\frac{-\ln(F(x))}{\sigma}\right)^2\right]^{-1}. \tag{5}$$

The cdf corresponding to (6) is given by

$$G(x) = 1 - \frac{2}{\pi}\arctan\left[\frac{-\ln(F(x))}{\sigma}\right]. \tag{6}$$

One of the main benefits of the Half Cauchy - X family is its ability of fitting skewed data that cannot be properly fitted by existing distributions.

The paper is organized as follows. In Section 2, we proposed Half Cauchy-Exponential (HCE) model, and discuss the shape of the density function and distribution function of the model. We derive the quantiles, mode, and Mean deviation. Analytical shapes of the reliability functions of the model under study and pdf of order statistics and their moments are derived in Section 3. In Section 4, the method of maximum likelihood estimation(MLE), method of least square (LSE), method of Cramer-von-Mises (CVME), and Anderson-Darling methods (ADE) are discussed. We explore the usefulness of the proposed distribution by means of real data set and estimation techniques are applied to calculate the model parameters in Section 5. In Section 6, concluding remarks are presented.

2. Half Cauchy-Exponential Distribution

In this section, we consider the case where f follows exponential distribution with parameter $\theta > 0$ and the cdf and pdf are respectively $F(x) = 1 - e^{-\theta x}$ and $f(x) = \theta e^{-\theta x}$; $x > 0$,$\theta > 0$.
The cdf and pdf of this new distribution are respectively, given by

$$G(x) = 1 - \frac{2}{\pi}\arctan\left(\frac{-\ln(1 - e^{-\theta x})}{\sigma}\right), \quad x > 0,\ \theta > 0,\ \sigma > 0 \tag{7}$$

and

$$g(x) = \frac{2\theta}{\pi\sigma}\frac{e^{-\theta x}}{1 - e^{-\theta x}}\left[1 + \left(\frac{-\ln(1 - e^{-\theta x})}{\sigma}\right)^2\right]^{-1}. \tag{8}$$

We call this new distribution Half Cauchy Exponential (HCE) distribution with parameters θ and σ. Evidently, the density function (8) does not involve any complicated function. Also, there is no functional relationship between the parameters. We denote the random variable X having pdf (8) as $HCE(\theta, \sigma)$.
The pdf plots of $HCE(\theta, \sigma)$ for various values of the parameters are presented in Figure 1. From the figure, it can be seen that the HCE distribution is well-suited for modelling right-skewed data.
The cdf plots of $HCE(\theta, \sigma)$ for various choices of the values of parameters are presented in Figure 2.

3. Properties of the Half Cauchy - Exponential Distribution

Lemma 1. The q^{th} quantile x_q of the HCE random variable is given by

$$x_q = -\frac{1}{\theta}\ln\left(1 - exp^{-\sigma\tan\left(\frac{\pi(1-q)}{2}\right)}\right). \tag{9}$$

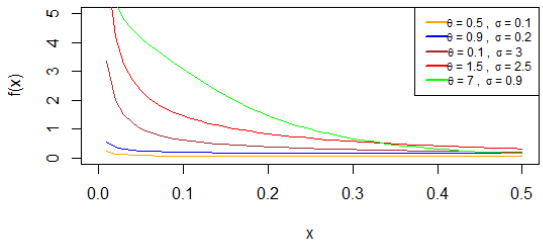


Figure 1: *Plots of the pdf of $HCE(\theta, \sigma)$ distribution.*

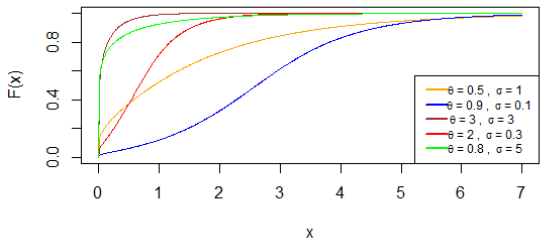


Figure 2: *Plots of the cdf of $HCE(\theta, \sigma)$ distribution.*

Proof. The q^{th} quantile x_q of the HCE random variable is defined as

$$q = P(X \leq x_q) = G(x_q), \quad x_q > 0$$

Using the distribution function of the HCE distribution, we have

$$q = G(x_q) = 1 - \frac{2}{\pi} \arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)$$

That is,

$$\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right) = \frac{\pi(1-q)}{2}$$

Hence

$$x_q = -\frac{1}{\theta} \ln\left(1 - exp^{-\sigma \tan\left(\frac{\pi(1-q)}{2}\right)}\right).$$

This completes the proof. ■

Using the usual inverse transformation method, random numbers can be sampled from the proposed model. Let U be a random number drawn from a uniform distribution on (0, 1). Then a random number X following $HCE(\theta, \sigma)$ distribution is obtained by the equation (9) .
In particular, the median is given by,

$$x_{0.5} = -\frac{1}{\theta} \ln(1 - exp^{-\sigma}). \tag{10}$$

Theorem 1. The mode of the $HCE(\theta, \sigma)$ is the solution of the equation $k(x) = 0$, where

$$k(x) = 2e^{-\theta x} \ln(1-e^{-\theta x}) - \sigma^2 \left[1 + \left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right].$$

Proof. The critical point of the HCE density function are the roots of the equation:

$$\frac{\partial \log(g(x))}{\partial x} = 0$$

That is

$$\frac{\partial \log(g(x))}{\partial x} = -\theta - \frac{\theta e^{-\theta x}}{1-e^{-\theta x}} + \frac{2\theta e^{-\theta x}\ln(1-e^{-\theta x})}{\sigma(1-e^{-\theta x})\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right]}. \tag{11}$$

The critical values of (11) are the solution of $k(x) = 0$.
Hence the proof. ■

3.1. Mean Deviation

The mean deviation about the median can be used as a measure of the degree of scattering in a population. Let M be the median of the HCE distribution given by (10).
The mean deviation about the median can be calculated as

$$\delta(X) = E|X - M| = \int_{-\infty}^{\infty} |x - M| g(x) dx,$$

Hence we obtain the following equation $\delta = \mu - 2J(M)$ where J(q) is

$$J(q) = \frac{2\theta}{\pi\sigma}\int_{-\infty}^{q} x\frac{e^{-\theta x}}{1-e^{-\theta x}}\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right]^{-1} dx. \tag{12}$$

One can easily compute this integral numerically in software such as MATLAB, Mathcad, R, and others and hence obtain the mean deviation about the median as desired.

3.2. Stochastic Ordering

Stochastic orders have been used during the last forty years, at an accelerated rate, in many diverse areas of probability and statistics. Such areas include reliability theory, survival analysis, queueing theory, biology, economics, insurance, and actuarial science (see, [5]). Let X and Y be two random variables having cdf's F and G respectively, and denote by $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$ their respective survival functions, with corresponding pdf's f,g. The random variable X is said to be smaller than Y in the:

1. stochastic order (denoted as $X \leq_{st} Y$) if $\bar{F}(x) \leq \bar{G}(x)$ for all x;
2. likelihood ratio order (denoted as $X \leq_{lr} Y$) if $\frac{f(x)}{g(x)}$ is decreasing in $x \geq 0$;
3. hazard rate order (denoted as $X \leq_{hr} Y$)if $\frac{\bar{F}(x)}{\bar{G}(x)}$is decreasing in $x \geq 0$;
4. reversed hazard rate order (denoted as $X \leq_{rhr} Y$) if $\frac{F(x)}{G(x)}$ is decreasing in $x \geq 0$.

The four stochastic orders defined above are related to each other, have the following implications (see, [5]):

$$X \leq_{rhr} Y \Leftarrow X \leq_{lr} Y \Rightarrow X \leq_{hr} Y \Rightarrow X \leq_{st} Y. \tag{13}$$

The HCE is ordered with respect to the strongest likelihood ratio ordering as shown in the following theorem. It shows the flexibility of the two-parameter HCE distribution.

Theorem 2. Let $X \sim HCE(\theta_1, \sigma_1)$ and $Y \sim HCE(\theta_1, \sigma_1)$. If $\theta_1 = \theta_2 = \theta$ and $\sigma_1 < \sigma_2$; then $X \leq_{lr} Y$ hence $X \leq_{rhr} Y$,$X \leq_{hr} Y$ and $X \leq_{st} Y$.

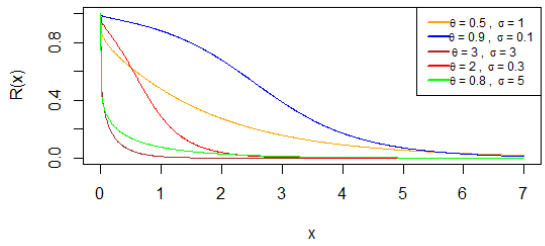


Figure 3: *Plots of reliability function of the $HCE(\theta,\sigma)$ distribution.*

Proof. The likelihood ratio is

$$\frac{g_X(x)}{g_Y(x)} = \frac{\sigma_2}{\sigma_1}\frac{\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma_2}\right)^2\right]}{\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma_1}\right)^2\right]}$$

Thus,

$$\frac{d}{dx}\left[\frac{g_X(x)}{g_Y(x)}\right] = \frac{2\theta e^{-\theta x}}{1-e^{-\theta x}}\left[\frac{1}{{\sigma_2}^2+-\ln(1-e^{-\theta x})} - \frac{1}{{\sigma_1}^2+-\ln(1-e^{-\theta x})}\right]$$

Now, if $\theta_1 = \theta_2 = \theta$ and $\sigma_1 < \sigma_2$, then $\frac{d}{dx}\left[\frac{g_X(y)}{g_Y(y)}\right] < 0$, which implies that $X \leq_{lr} Y$ hence $X \leq_{rhr} Y$,$X \leq_{hr} Y$ and $X \leq_{st} Y$. ■

Lemma 2. If a random variable Y follows the standard exponential distribution, then $X = -\frac{1}{\theta}\ln\left(1-exp^{-\sigma\tan\left(\frac{\pi(1-e^{-y})}{2}\right)}\right) \sim HCE(\theta,\sigma)$.

3.3. Reliability Analysis

The reliability function is the characterization of an explanatory that maps a set of events, usually associated with the failure of some system onto time. It is the probability that the system will survive beyond a specified time, which is defined by $R(t) = 1 - G(t)$. The Reliability function of $HCE(\theta,\sigma)$ is given by,

$$R(t) = \frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta t})}{\sigma}\right). \tag{14}$$

The reliability behaviour of $HCE(\theta,\sigma)$ for various choices of the values of the parameters is presented in Figure 3. The other characteristic of interest of a random variable is the hazard rate function defined by

$$h(t) = \frac{g(t)}{1-G(t)}$$

The hazard rate function of $HCE(\theta,\sigma)$ is given by,

$$h(t) = \frac{\frac{2\theta}{\pi\sigma}\frac{e^{-\theta t}}{1-e^{-\theta t}}\left[1+\left(\frac{-\ln(1-e^{-\theta t})}{\sigma}\right)^2\right]^{-1}}{\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta t})}{\sigma}\right)}. \tag{15}$$

The behaviour of the hazard rate function of $HCE(\theta,\sigma)$ for various choices of the values of the parameters is presented in Figure 4. The cumulative hazard rate function of HCE distribution,

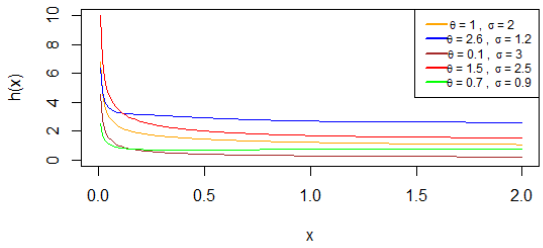


Figure 4: *Plots of hazard rate function of the $HCE(\theta,\sigma)$ distribution.*

$H(t)$ is given by,

$$H(t) = -\ln R(t)$$
$$= -\ln \frac{2}{\pi} \arctan\left(\frac{-\ln(1-e^{-\theta t})}{\sigma}\right). \tag{16}$$

It is important to know that the units for H(t) are the cumulative probability of failure per unit of time, distance, or cycles.

3.4. Order Statistics

Let $X_1, X_2, ..., X_n$ be a random sample from $HCE(\theta,\sigma)$. Also, let $X_{(1)}, X_{(2)}, ..., X_{(n)}$, denote the corresponding order statistics. Then the pdf and cdf of k^{th} order statistics, are given by

$$f_X(x) = \frac{n!}{(k-1)!(n-k)!}[G(x)]^{k-1}\,[1-G(x)]^{n-k}\,g(x)$$

$$= \frac{n!}{(k-1)!(n-k)!}\left[1-\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)\right]^{k-1}\left[\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)\right]^{n-k}$$
$$\frac{2\theta}{\pi\sigma}\frac{e^{-\theta x}}{1-e^{-\theta x}}\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right]^{-1} \tag{17}$$

and

$$F_X(x) = \sum_{j=k}^{n}\binom{n}{j}[G(x)]^j[1-G(x)]^{n-j}$$

$$= \sum_{j=k}^{n}\binom{n}{j}\frac{n!}{(k-1)!(n-k)!}\left[1-\left[\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)\right]\right]^{j}\left[\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)\right]^{n-j} \tag{18}$$

respectively.
The pdf of the minimum is,

$$f_{X_{(1)}}(x) = n\frac{2\theta}{\pi\sigma}\frac{e^{-\theta x}}{1-e^{-\theta x}}\left[\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)\right]^{n-1}\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right]^{-1} \tag{19}$$

and the pdf of the maximum is,

$$f_{X_{(n)}}(x) = n\frac{2\theta}{\pi\sigma}\frac{e^{-\theta x}}{1-e^{-\theta x}}\left[1-\left[\frac{2}{\pi}\arctan\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)\right]\right]^{n-1}\left[1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right]^{-1}. \tag{20}$$

4. Parameter Estimation

In this section, we describe the maximum likelihood estimation procedure, method of least squares, method of Cramer-von-Mises, Anderson-Darling methods, and Method of maximum product of spacings to estimate the parameters θ and σ, in the HCE distribution. We assume throughout that $x_1, x_2, \ldots, x_n$ is a random sample of size n from the HCE distribution both parameters θ and σ are unknown.

4.1. Maximum Likelihood Estimation(MLE)

Here, we consider the estimation of the unknown parameters of the new distribution by the maximum likelihood method. Consider a random sample $(x_1, x_2, \ldots, x_n)$ of size n, from the $HCE(\theta, \sigma)$ distribution. Then, the log likelihood function is given by,

$$\log L = n\log(2\theta) - n\log(\pi\sigma) - \theta\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln(1-e^{-\theta x_i}) - \sum_{i=1}^{n}\left(1+\left(\frac{-\ln(1-e^{-\theta x})}{\sigma}\right)^2\right)$$

The likelihood equations are,

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\frac{x_i e^{-\theta}}{(1-e^{-\theta x_i})} - 2\sum_{i=1}^{n}\frac{x_i e^{-\theta x_i}}{\sigma^2(1-e^{-\theta x_i})}\frac{\ln(1-e^{-\theta x})}{(1+\frac{-\ln(1-e^{-\theta x})}{\sigma})^2}$$

$$= 0, \tag{21}$$

and

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} - 2\sum_{i=1}^{n}\frac{\ln(1-e^{-\theta x})(1-e^{-\theta x_i})}{\sigma^3(1+\frac{-\ln(1-e^{-\theta x})}{\sigma})^2}$$

$$= 0. \tag{22}$$

These equations do not have explicit solutions and they have to be obtained numerically using statistical software like the *nlm* package in R programming.
If the parameter vector of $HCE(\theta, \sigma)$ be $\Theta = (\theta, \sigma)$ and the associated MLE for Θ is $\hat{\Theta} = (\hat{\theta}, \hat{\sigma})$, then the resulting asymptotic normality is $(\hat{\Theta} - \Theta) \to N(0, (I(\Theta))^{-1})$. Where the observed Fisher's information matrix $(I(\Theta))$ is given by,

$$I(\Theta) \approx \begin{bmatrix} -E(\frac{\partial^2 \log L}{\partial \theta^2}) & -E(\frac{\partial^2 \log L}{\partial \theta \partial \sigma}) \\ -E(\frac{\partial^2 \log L}{\partial \theta \partial \sigma}) & -E(\frac{\partial^2 \log L}{\partial \sigma^2}) \end{bmatrix},$$

and hence the variance covariance matrix would be $I^{-1}(\Theta)$.
As a result of MLEs' asymptotic normality, we may construct approximate $100(1-\alpha)\%$ confidence intervals for θ and σ of $HCE(\theta, \sigma)$ as below:

$$\hat{\theta} \pm Z_{\frac{\alpha}{2}} SE(\hat{\theta}), \hat{\sigma} \pm Z_{\frac{\alpha}{2}} SE(\hat{\sigma})$$

Theorem 3. Let $g_1(\theta; \sigma, x)$ denote the function on the right-hand side (RHS) of equation (21), where σ is the true value of the parameter. Then there exists a unique solution for $g_1(\theta; \sigma, x) = 0$, for $\hat{\theta} \epsilon (0, \infty)$.

Proof. We have

$$g_1(\theta; \sigma, x) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\frac{x_i e^{-\theta}}{(1-e^{-\theta x_i})} - 2\sum_{i=1}^{n}\frac{x_i e^{-\theta x_i}}{\sigma^2(1-e^{-\theta x_i})}\frac{\ln(1-e^{-\theta x})}{(1+\frac{-\ln(1-e^{-\theta x})}{\sigma})^2}$$

Now

$$\lim_{\theta\to 0} g_1(\theta;\sigma,x) = \infty,$$

On the other hand

$$\lim_{\theta\to \infty} g_1(\theta;\sigma,,x) < 0.$$

Therefore there exist at least one root, say $\hat{\theta}\epsilon(0,\infty)$ such that $g_1(\theta;\sigma,x) = 0$
To show uniqueness, the first derivative of $g_1(\theta;\sigma,x) = 0$ is

$$\frac{\partial g_1(\theta;\sigma,x)}{\partial\theta} < 0.$$

Hence there exist a solution for $g_1(\theta;\sigma,,x) = 0$, and root, $\hat{\theta}$ is unique. ■

Theorem 4. Let $g_2(\sigma,\theta,x) = 0$ denote the function on the right hand side (RHS) of equation (22), where θ is the true value of the parameter. Then there exists a unique solution for $g_2(\sigma,\theta,x) = 0$, for $\hat{\sigma}\epsilon(0,\infty)$.

Proof. We have

$$g_2(\sigma,\theta,x) = -\frac{n}{\sigma} - 2\sum_{i=1}^{n}\frac{\ln(1-e^{-\theta x})(1-e^{-\theta x_i})}{\sigma^3(1+\frac{-\ln(1-e^{-\theta x})}{\sigma})^2}.$$

Now

$$\lim_{\sigma\to 0} g_2(\sigma,\theta,x) = -\infty,$$

On the other hand

$$\lim_{\sigma\to \infty} g_2(\sigma,\theta,x) > 0.$$

Therefore there exist atleast one root, say $\hat{\sigma}\epsilon(0,\infty)$ such that $g_2(\sigma,\theta,x) = 0$
To show uniqueness, the first derivative of $g_2(\sigma,\theta,x) = 0$ is

$$\frac{\partial g_2(\sigma,\theta,x)}{\partial\sigma} < 0.$$

Hence there exist a solution for $g_2(\sigma,\theta,x) = 0$, and root, $\hat{\sigma}$ is unique. ■

4.2. Method of Cramer-von Mises

Cramer-von-Mises type minimum distance estimators are based on minimizing the distance between the theoretical and empirical cumulative distribution functions. In [7] provided empirical evidence that the bias of these estimators is smaller than the bias of other minimum distance estimators. The Cramer-von-Mises estimators, $\hat{\theta}_{CME}$ and $\hat{\sigma}_{CME}$ are the values of θ and σ minimizing

$$C(\theta,\sigma) = \frac{1}{12n} + \sum_{i=1}^{n}\left[G(t_i \mid \theta,\sigma) - \frac{2i-1}{2n}\right]^2.$$

Differentiating the above equation partially, with respect to the parameters θ and σ respectively and equating them to zero, we get the normal equations. Since the normal equations are non-linear, we can use iterative method to obtain the solution.

4.3. Method of Anderson-Darling

The method of Anderson-Darling test was developed by [9] as an alternative to statistical tests for detecting sample distributions departure from normality.

The Anderson-Darling estimators $\hat{\theta}_{ADE}$ and $\hat{\sigma}_{ADE}$ are the values of θ and σ minimizes

$$A(\theta,\sigma) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\{\log G(t_i \mid \theta,\sigma,\theta) - \log \bar{G}(t_{n+1-i} \mid \theta,\sigma)\}.$$

Differentiating the above equation partially, with respect to the parameters θ and σ respectively and them equating to zero, we get the normal equations. Since the normal equations are non-linear, we can use the iterative method to obtain the solution.

4.4. Method of Least-Square Estimation

The least-square estimators were proposed by [2] to estimate the parameters of Beta distributions. Here, we apply the same technique for the HCE distribution. The least-square estimators of the unknown parameters θ and σ of HCE distribution can be obtained by minimizing

$$\sum_{i=1}^{n}\left[G(t_i \mid \theta,\sigma) - \frac{i}{n+1}\right]^2.$$

with respect to unknown parameters θ and σ.

5. Applications

In this section, we have taken two real life data set to illustrate the importance of the proposed distribution. For each data set, we estimate the unknown parameters of each distribution by the maximum-likelihood method, method of least squares, method of Cramer-von-Mises, Anderson-Darling methods, and Method of maximum product of spacings.With these obtained estimates, we obtain the values of the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as well as Kolmogorov-Smirnov statistic and the corresponding p-value. Here, $AIC = -2\ln(L) + 2k$ and $BIC = -2\ln(L) + k\ln(n)$; where L is the likelihood function evaluated at the maximum likelihood estimates, k is the number of parameters and n is the sample size. The K-S distance $D_n = sup_x|F(x) - F_n(x)|$,where,$F_n(x)$ is the empirical distribution. Kolmogorov-Smirnov (K-S) statistics is computed to compare the fitted models.
The required computations are carried out in the R-language introduced by [10].

5.1. Data set I

We consider the corona-virus cases distribution among the fifteen countries viz.,France, Italy, Spain, US, Germony, UK, Turkey, Iran, Russia, China, Brazil, Canada, Belgium, Netherlands and Switzerland. Data has taken from a website and URL is https://www.worldometers.info/coronavirus/coronavirus-cases/.
Data is given in percentage and the observations are:
5.37,6.56,7.61,32.83,5.24,5.06,3.65,3.03 2.89,2.74,2.10,1.57,1.55,1.27,0.97.
The data is skewed-to-the right with skewness =3.0901 and kurtosis =8.4119
The descriptive statistics of the above data set are given in Table 1. The MLEs for θ and σ are listed in Table 2 along with their standard errors (S.E.). The values in Table 3 show that the HCE distribution leads to a better fit for the other three models. Based on the values of the AIC and BIC criteria as well as the value of the KS-statistic and the corresponding p-value, we observe that the HCE distribution provides the best fit for these data among all the models considered. Table 3, it has been observed that the proposed model is best fit as compared to xgamma distribution

Table 1: *The descriptive statistics of Data set.*

Min	1st Q	Median	Mean	3rd Q	Max
0.970	1.835	3.030	5.496	5.305	32.830

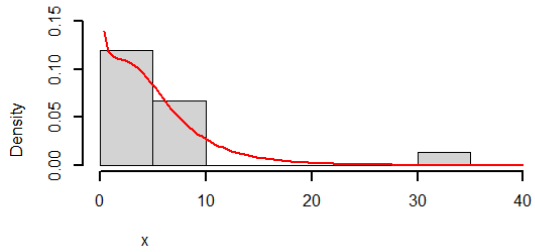
Table 2: *S.E., MLE for θ and σ*

Parameters	MLE	S.E.
θ	0.2359	0.0799
σ	0.5618	0.5618

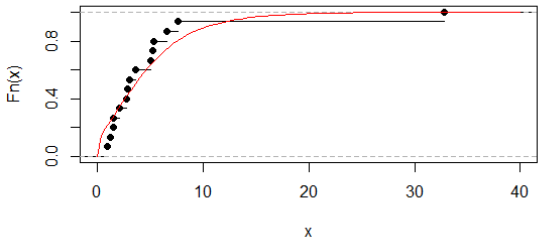
(XGD) by [11], Akash distribution (AKD) by [21], and exponential power distribution (EPD) by [17].

Figure 5 shows the fitted density curves, Empirical and the fitted cumulative distribution functions for the Data set I.

The goodness-of-fit of the CVME, MLE, LSE, and ADE methods are observed by the test statistic



(a) Fitted density curves for the data set I



(b) Empirical and the fitted cumulative distribution functions for the data set I

Figure 5: *Histogram with fitted pdf's (left) and Empirical cdf with fitted cdf's (right) for the data set I.*

values and their p-values for CVM (Cramer-Von Mises), KS (Kolmogorov-Simnorov), and AD (Anderson-Darling) for the dataset I which are displayed in Table 4.

Figure 6 shows fitted distribution's histogram and the density function having CVME, MLE, LSE, and ADE for the data set I of HCE distribution.

5.2. Data set II

The following data comes from a 59-conductor accelerated life test by [16]. Atomic movement in the circuit's conductors can create failures in microcircuits, which is known as electromigration.

Table 3: *Goodness of fit for various models fitted for the Data set I.*

Model	-LL	AIC	BIC	K-S(p-value)
XGD	-43.555	89.111	89.819	0.2501(0.2585)
AKD	-44.565	91.131	91.839	0.2687(0.1905)
EPD	-42.940	89.881	91.297	0.2471(0.2709)
HCE	-42.435	88.871	90.286	0.2166(0.4224)

Table 4: *Statistics values and their associated p-values for the dataset I.*

Estimation method	Estimates	K-S(p-value)	CVM(p-value)	AD(p-value)
CVME	0.4402 0.2360	0.1398(0.8924)	0.0446(0.9144)	0.9312(0.3937)
MLE	0.2359 0.5618	0.2166(0.4224)	0.1210(0.4962)	1.0133(0.349)
ADE	0.3529 0.3219	0.1619(0.7698)	0.0570(0.84)	0.8506(0.4438)
LSE	0.3918 0.2866	0.1553(0.8101)	0.0480(0.8951)	0.8731(0.4292)

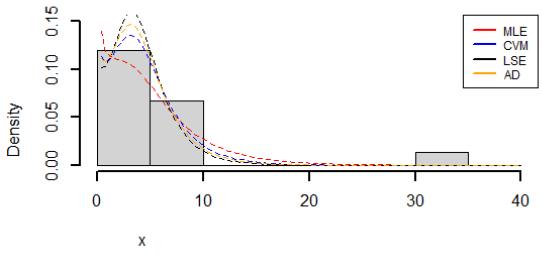


Figure 6: *fitted distribution's histogram and the density function having CVME, MLE, LSE, and ADE for the data set I.*

There are no censored observations, and the failure times are in hours.
5.923, 4.288, 6.522, 4.137, 6.071, 7.495, 6.573, 6.538, 5.589, 6.087, 5.807, 6.725, 8.532, 9.663, 6.545, 10.491, 7.543, 6.956, 6.492, 5.459, 8.120, 4.706, 8.687, 2.997, 8.591, 6.129, 11.038, 5.381, 10.092, 7.496, 4.531, 7.974, 8.799, 7.683, 7.224, 7.365, 6.923, 5.640, 5.434, 7.937, 6.515, 6.476, 6.369, 7.024, 8.336, 9.218, 7.945, 6.869, 6.352, 4.700, 6.948, 9.254, 5.009, 7.489, 7.398, 6.033,7.459, 9.289 , 6.958.
The data is skewed-to-the right with skewness =0.1932 and kurtosis =0.0874
The descriptive statistics of the above data set are given in Table 5. The MLEs for θ and σ are listed in Table 6 along with their standard errors (S.E.). The values in Table 7 show that the *HCE* distribution leads to a better fit for the other five models. Based on the values of the AIC and BIC criteria as well as the value of the KS-statistic and the corresponding p-value, we observe that the HCE distribution provides the best fit for these data among all the models considered. Table 7, it has been observed that the proposed model is best fit as compared to Lindley-Exponential (LE) model by [13], generalized exponential (GE) model by [14], modified Weibull (MW) model by [15], exponential power (EP) model by [17], and Weibull extension (WE) model by [18].

Figure 7 shows the fitted density curves, Empirical and the fitted cumulative distribution functions for the Data set II.

The goodness-of-fit of the CVME, MLE, LSE, and ADE methods are observed by the test statistics values and their p-values for CVM (Cramer-Von Mises), KS (Kolmogorov-Simnorov), and AD (Anderson-Darling) for the dataset II which are displayed in Table 8.

Table 5: *The descriptive statistics of Data set.*

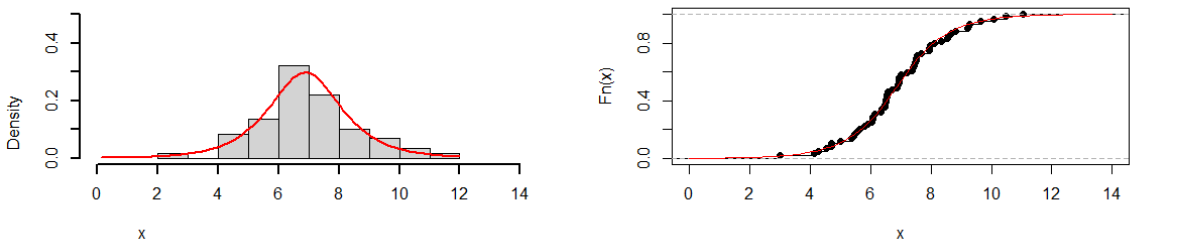
Min	1st Q	Median	Mean	3rd Q	Max
2.997	6.052	6.923	6.980	7.941	11.038

Table 6: *S.E., MLE for θ and σ*

Parameters	MLE	S.E.
θ	0.9352	0.1117
σ	0.0015	0.0012

Table 7: *Goodness of fit for various models fitted for the Data set.*

Model	-LL	AIC	BIC	K-S	p-value
EP	−116.5015	237.0029	237.2098	0.1042	0.5103
LE	−114.9528	233.9055	234.1198	0.1042	0.5099
GE	−114.9473	233.8946	234.1098	0.1365	0.2021
WE	−113.6745	233.3491	233.7855	0.1067	0.4796
MW	−112.5218	231.0435	231.4799	0.0914	0.6738
HCE	−111.7792	227.5584	231.713	0.05806	0.9819



(a) Fitted density curves for the data set II

(b) Empirical and the fitted cumulative distribution functions for the data set II

Figure 7: *Histogram with fitted pdf's (left) and Empirical cdf with fitted cdf's (right) for the data set II.*

Figure 8 shows fitted distribution's histogram and the density function having CVME, MLE, LSE, and ADE for the data set II of HCE distribution.

6. Concluding remarks

In this article, we have introduced and studied a new family of distributions called the Half Cauchy exponential(*HCE*) distribution. we have provided explicit expressions for the quantiles, hazard rates, mean deviation about median, the stochastic ordering and order statistics. The model parameters are estimated by maximum likelihood, least-squares, Cramer-von Mises, and Anderson-Darling. Our formulas related to the *HCE* model are manageable, and with the use of modern computer resources with analytic and numerical capabilities, may turn into adequate tools for a certain purpose of statisticians.
The applicability of the model is demonstrated by using two real data set. From Tables 3 and 7, we observed a better performance of our distribution than the existing distributions. Because HCE distribution has the least test statistic value and the highest p value, we can deduce that it has a considerably better fit than the other distributions studied. Based on these findings, the newly suggested model can be considered as a more efficient, flexible, and therefore may be an alternative to other distributions for modeling positive real data sets. Our proposed model may attract wider applications in survival analysis for modeling positive real data sets. Estimation of

Table 8: *Statistics values and their associated p-values for the dataset II.*

Estimation method	Estimates	K-S(p-value)	CVM(p-value)	AD(p-value)
CVME	0.9055 0.0019	0.0563(0.9867)	0.0233(0.9932)	0.1667(0.997)
MLE	0.9352 0.0015	0.0580(0.9819)	0.0246(0.9908)	0.1798(0.995)
ADE	0.9057 0.0018	0.05598(0.9876)	0.0232(0.9932)	0.1672(0.9969)
LSE	0.8797 0.0022	0.0633(0.9596)	0.0279(0.9832)	0.1800(0.995)

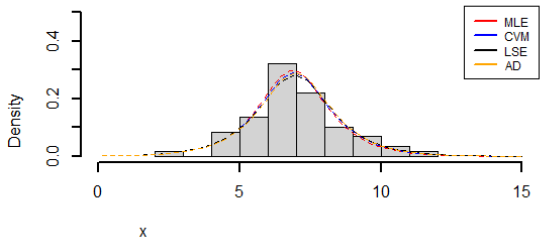


Figure 8: *fitted distribution's histogram and the density function having CVME, MLE, LSE, and ADE for the data set II.*

the model parameters under the Bayesian paradigm is currently underway.

References

[1] Vrbik and Jan. (2011). Sampling distribution of ML estimators: Cauchy example, *Mathematica Journal*, 13:13–19,

[2] Swain, James J and Venkatraman, Sekhar and Wilson, James R. (1988) Least-squares estimation of distribution functions in Johnson's translation system, *Journal of Statistical Computation and Simulation*, 29:271–297.

[3] Krishnamoorthy and Kalimuthu. (2006). Handbook of statistical distributions with applications, *Chapman and Hall/CRC*

[4] Shaw, MW. (1995). Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance, *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 259:243–248.

[5] Shaked, Moshe and Shanthikumar, J George. (2007). Stochastic orders, *Springer*

[6] Paradis, Emmanuel and Baillie, Stephen R and Sutherland, William J.(2002) Modeling large-scale dispersal distances, *Ecological Modelling*, 151:279–292.

[7] Macdonald, PDM. (1971). Comments and queries comment on an estimation procedure for mixtures of distributions by choi and bulgren, *Journal of the Royal Statistical Society: Series B (Methodological)*, 33:326–329.

[8] Forbes, Catherine and Evans, Merran and Hastings, Nicholas and Peacock, Brian. (2011). Statistical distributions, author=Forbes, Catherine and Evans, Merran and Hastings, Nicholas and Peacock, Brian, *Wiley New York*,4.

[9] Anderson, Theodore W and Darling, Donald A. (1952). Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes, *The annals of mathematical statistics*, 193–212.

[10] Team, R Develpment Core.(2009) A language and environment for statistical computing, *http://www. R-project. org*.

[11] Sen, Subhradev and Maiti, Sudhansu S and Chandra, N. (2016). The xgamma distribution: statistical properties and application, *Journal of Modern Applied Statistical Methods*, 15:38

[12] Zografos, Konstantinos and Balakrishnan, Narayanaswamy. (2009). On families of beta-and generalized gamma-generated distributions and associated inference, *Statistical methodology*, 6:344–362.

[13] Bhati, Deepesh and Malik, Mohd Aamir and Vaman, HJ. (2015). Lindley–exponential distribution: properties and applications, *Metron*, 73:335–357.

[14] Gupta, Rameshwar D and Kundu, Debasis. (1999). Theory & methods: Generalized exponential distributions, *Australian & New Zealand Journal of Statistics*, 41:173–188.

[15] Lai, CD and Xie, Min and Murthy, DNP. (2003). A modified Weibull distribution, *IEEE Transactions on reliability*, 52:33–37.

[16] Nelson, Wayne and Doganaksoy, Necip. (1995). Statistical analysis of life or strength data from specimens of various sizes using the power-(log) normal model, *Recent Advances in Life-Testing and Reliability*, 377–408.

[17] Smith, Robert M and Bain, Lee J. (1975). An exponential power life-testing distribution, *Communications in Statistics-Theory and Methods*, 4:469–481,

[18] Tang, Y and Xie, M and Goh, TN. (2003). Statistical analysis of a Weibull extension model, *Communications in Statistics-Theory and Methods*, 32:913–928.

[19] Ristić, Miroslav M and Balakrishnan, Narayanaswamy. (2012). The gamma-exponentiated exponential distribution, *Journal of statistical computation and simulation*, 82:1191–1206.

[20] Johnson, Norman L and Kotz, Samuel and Balakrishnan, Narayanaswamy. (1995). Continuous univariate distributions, volume 2, *John wiley & sons*

[21] Shanker, Rama. (2015). Akash distribution and its applications, *International Journal of Probability and Statistics*, 4:65–75.

ANALYSIS OF ENCOURAGED ARRIVAL MULTIPLE WORKING VACATION QUEUING MODEL UNDER THE STEADY STATE CONDITION

Ismailkhan Enayathulla Khan, Rajendran Paramasivam*

•

[1,2]Department of Mathematics, School of Advanced Sciences,
Vellore Institute of Technology, Vellore, India - 632014.
Ismailkhan.e@vit.ac.in
Correspondence email: prajendran@vit.ac.in

Abstract

Businesses typically entice customers with alluring offers and discounts. Encouraged arrivals is the name given to these curious clients. In certain situations, the service offered by queuing models, notably in transportation networks, enables the simultaneous serving of several consumers. In general, closed-form solutions to bulk service queuing models with idle servers are difficult to find. By coordinating the operations at each workstation using the Chapman-Kolmogorov research technique, the main objective of this study is to assess the performance of the car assembly line in order to reduce waiting times. The server is in a busy state, is idle, is regularly busy, and is in a busy state when it breaks down. Performance metrics are being tracked using a multiple working vacation approach. In this study, analysis of encouraged arrival multiple working vacation queuing model under the steady state condition. In this model, we included encouraged arrival. By resolving difference equations and Chapman Kolmogorov balancing equations, the steady state queue size problem is found. Additionally, the server is in the busy, idle state, regular, and breakdown busy states, and performance metrics are conducted. The server was sent for repair and is now completely repaired to avoid the crash at any time. After that, the server continues to offer the service. It is evidently identified that the efficiency level increased while the encouraged arrival is incorporated. The main contribution of this paper is to show the server is in the busy, idle state, regular, and breakdown busy states, and performance metrics efficient level increases. It is found that they offer more efficient results when compared with the Poisson process method

Keywords: Encouraged arrival, steady state, multiple working vacations, single server, queue size.

I. Introduction

There are times when the service provided in queuing models, particularly in transportation systems, allows for the simultaneous serving of many customers. In general, bulk service queuing models with idle servers are challenging to solve in closed form in [1]. A generic class of bulk queues with encouraged input was researched in [12].

A study on the examination of a GI/M/I- queue with several vacations is described in [2]. The Markovian M/ (q, **β**)/1 queuing model while taking several working vacations was also examined in [4]. An M/M/1 lines with working vacations (M/M/1/WV) model was investigated in [15]. A well-considered time-dependent bulk queuing service solution issues with queuing in [3].

Analyses the best management strategy for a heterogeneous M/M/1- queue with server downtime [5]. An analysis of the N-Policy and the M/M/1 queue with numerous working vacations is found in [10]. The queuing procedure using bulk services is examined in [19]. A Study on the Analysis of the M/G/1 Queue's Queue Length Distribution with Working Vacations [6].

Stochastic models with matrix geometric solutions were examined in [18]. GI/Geo/1 queue

with several Vacations were investigated in [9]. Working vacation queue and matrix analysis a study was done in [7]. The optimal operation of a Markovian queuing system with a transportable and unreliable server was examined in [14].

The M/M/1-queue with a single working vacation was investigated in [21]. The M/M/1 queue with a single working holiday and setup times were investigated in [22]. A matrix analytical approach for the examination of the M/G/1 queue with exponential working vacations was looked at in [8 and 16]. Recent advancements in the queuing and bulk models were examined in [17]. An M/M/1/N queue system with encouraged arrivals was investigated in [23].

A finite and infinite M/H2/1 queuing system with a detachable, unreliable server was examined in [20]. The M/M(a,b)/1/MWV/Br Model [25] was investigated. Research on an M/G/1 queue with numerous working vacations may be found in [11]. Reducing wait times in an M/M/1/N encouraged arrival line by providing feedback, balking at unpaid customers, and sustaining those customers was investigated in [24]. A break-down-prone portable service station with M/Ek/1 queuing system optimization was examined in [13].

II. Model Recitation

In this model recitation provided by

- This method is encouraged with parameters λ^* (1+ϑ). In the manner of General-Bulk Service-Rule (G-B-S-R), the server handles the customers in batches.
- As a result of this rule, the server only begins to provide service when at least a "q" customer is present.
- The server serves the first β – customers, leaving the others in the line.
- Each batch of units must have a certain minimum and maximum number of "q" units to be used.
- The assumption is that the batch size α(q≤α≤β) will have an accelerated distribution of the parameter and will be an independent random variable with an identical distribution.
- When the server breaks down with the parameter $(1 - e^{-bt})$.
- If the queue- size reaches minimum "q" imagine a situation where a server starts offering service while on vacation at a different rate from the standard one.
- The server was sent for repair and is now completely repaired to avoid the crash at any time. After that, the server continues to offer the service.

Let $K_A(t)$ represent the number of customers waiting in line at time t, and C(t) represent one of zero, one, two, or three depending on whether the server is idle, busy, in a typical busy state, or experiencing a breakdown respectively.

Let $M_k(t) = \text{Prob}\{K_A(t) = k, C(t) = 0\}; 0 \le k \le q - 1,$

$A_k(t) = \text{Prob}\{K_A(t) = k, C(t) = 1\}; k \ge 0,$

$P_k(t) = \text{Prob}\{K_A(t) = k, C(t) = 2\}; k \ge 0,$

$\Upsilon_k(t) = \text{Prob}\{K_A(t) = k, C(t) = 3\}; k \ge 0,$

The queue-size and the system-size are equal for C(t) = 0.

$$M_k = \lim_{t.\to\infty} M_k(t); A_k = \lim_{t.\to\infty} Q_k(t); P_k = \lim_{t.\to\infty} P_k(t); \Upsilon_k = \lim_{t.\to\infty} \Upsilon_k(t);$$

As an outcome, the Chapman-Kolmogorov equations that satisfy the condition are as follows:

$$\lambda * (1+\vartheta)\ M_0 = \mu . P_0 + \mu_u\ A_0 \quad (1)$$

$$\lambda * (1+\vartheta)\ M_k = \lambda * (1+\vartheta)\ M_{k-1} + \mu\ P_k + \mu_u\ A_k; (1 \le k < q-1), \quad (2)$$

$$(\lambda * (1+\vartheta) + \chi + \mu_u)\ A_0 = \lambda * (1+\vartheta)\ M_{q-1} + \mu \sum_{k=q}^{\beta} A_k, \quad (3)$$

$$(\lambda * (1+\vartheta) + \chi + \mu_u)\ A_k = \lambda * (1+\vartheta)\ A_{k-1} + \mu_u\ A_{k+\beta}; (k \ge 1) \quad (4)$$

$$(\lambda * (1+\vartheta) + \mu + s)\ P_0 = \mu \sum_{k=q}^{\beta} P_k + \chi A_0 + b\Upsilon_0, \quad (5)$$

$$(\lambda * (1+\vartheta) + \mu + s) P_k = \lambda * (1+\vartheta) P_{k-1} + \mu P_{k+\beta} + \chi A_k + b\Upsilon_k; (k \ge 1), \quad (6)$$

$$(\lambda * (1+\vartheta) + b)\Upsilon_0 = sP_0, \quad (7)$$

$$(\lambda * (1+\vartheta) + b)\Upsilon_k = sP_k + \lambda * (1+\vartheta) \sum_{n=1}^{k} \Upsilon_{k-n} h_n; (k \ge 1). \quad (8)$$

The forward shifting operator Expectation on $P_k$and A_k are introduced as follows and will be used to solve the steady-state equations:

$Exp(P_k) = P_{k+1}; Exp(A_k) = A_{K+1}; Exp(\Upsilon_k) = \Upsilon_{k+1}; (k \geq 0).$

Therefore, the homogeneous differential equation is given by equation (4).

$(\lambda * (1+\vartheta) + \chi + \mu_u)A_k = \lambda * (1+\vartheta)A_{k-1} + \mu_u A_{k+\beta}; (k \geq 1),$

$\left[\mu_u Exp^{\beta+1} - Exp(\lambda * (1+\vartheta) + \chi + \mu_u) + \lambda * (1+\vartheta)\right]A_k = 0; (k \geq 0),$ (9)

The difference characteristic equation is provided by

$g(o) = \left[\mu_u o^{\beta+1} - (\lambda * (1+\vartheta) + \chi + \mu_u)o + \lambda(1+\vartheta)\right] = 0$

by taking $e(o) = (\lambda * (1+\vartheta) + \chi + \mu_u)o; h(o) = \mu_u o^{\beta+1} + \lambda * (1+\vartheta)$

It is found that |h (o)| <|e (o)| on |o| = 1.

The solution of the homogeneous differential equation is given by,

$A_k = m_u^k A_0 ; (k \geq 0)$ (10)

Additionally, equation (6) will be expressed as,

$(\lambda * (1+\vartheta) + \mu + s)P_k = \lambda(1+\vartheta)p_{k-1} + \mu P_{k+\beta} + \chi A_k + b\Upsilon_k; (k \geq 1),$

$\left[\mu Exp^{\beta+1} - Exp(\lambda * (1+\vartheta) + \mu + s) + \lambda * (1+\vartheta)\right]P_k = -\chi A_{k+1} - bExp(\Upsilon_k); (k \geq 0).$ (11)

By applying, Rouche's theorem, we discover that the equation,

$\mu o^{\beta+1} - (\lambda * (1+\vartheta) + \mu + s)o + \lambda * (1+\vartheta) = 0$ has unique root with $|m| < 1$provided by $\frac{\lambda(1+\vartheta)}{\beta\mu}$<1.

Equation (8) will be expressed as follows:

$(\lambda * (1+\vartheta) + b)\Upsilon_k = sP_k + \lambda * (1+\vartheta)\Upsilon_{k-1}; (k \geq 1)$

$\Upsilon_k = \frac{sExp(P_k)}{(\lambda*(1+\vartheta)+b)Exp - \lambda*(1+\vartheta)}.$ (12)

By substituting (11) to (12),

We obtain,

$\left[\mu Exp^{\beta+1} - Exp(\lambda * (1+\vartheta) + \mu + s) + \lambda * (1+\vartheta)\right]P_k = -\chi m_u^{k+1}A_0 - b\left[\frac{sExp^2(P_k)}{(\lambda*(1+\vartheta)+b)Exp - \lambda*(1+\vartheta)}\right].$ (13)

Consequently, the non-homogeneous difference equation (13) has the following solution:

$$P_k = \left[Qm^k - \frac{\chi m_u^{k+1}}{\mu m_u^{\beta+1} - (\lambda * (1+\vartheta) + \mu + s)m_u + \lambda * (1+\vartheta) + \frac{sbm_u^2}{(\lambda * (1+\vartheta) + b)m_u - \lambda * (1+\vartheta)}}\right]A_0$$

(i.e)$P_k = (Q\, m^k + \Upsilon\, m_u^k).$ (14)

The sequence $M_k (0 \leq k \leq q-1)$ for the condition equation (1) & (2) adding, we get

$\lambda * (1+\vartheta)\sum_{n=0}^{k} M_k = \lambda * (1+\vartheta)\sum_{n=0}^{k-1} M_N + \mu \sum_{N=0}^{k} P_N + \mu_u \sum_{N=0}^{k} A_n,$

$$\lambda * (1+\vartheta)\, M_k = \mu \sum_{n=0}^{k} P_N + \mu_u \sum_{N=0}^{k} A_N.$$

It is discovered by the interchange A_k and P_k in equations (9) and (14) that

$\lambda * (1+\vartheta)\, M_k = [\mu(Qm^k + \Upsilon m_u^k) + \mu_u(m_u^k)]A_0,$

$$M_k = \left[\frac{\mu}{\lambda * (1+\vartheta)}\left(\frac{Q(1-m^{k+1})}{(1-m)} + \frac{\Upsilon(1-m_u^{k+1})}{1-m_u}\right) + \frac{\mu_u}{\lambda * (1+\vartheta)}\left(\frac{1-m_u^{k+1}}{1-m_u}\right)\right]A_0.$$

Since Q and A_0 are unknowns, the steady-state queue size probabilities are also unknown.

To determine Q, we now take into account the equation (5),

$$(\lambda * (1+\vartheta) + \mu + s)P_0 = \mu \sum_{k=q}^{\beta} P_k + \chi A_0 + b\Upsilon_0$$

Equations (14) and (12)'s P_k will be substituted to determine that

$(\lambda * (1+\vartheta) + \mu + s)(Q + \Upsilon)A_0 = \mu\left(Q\left[\frac{m^q - m^{\beta+1}}{1-m}\right] + \Upsilon\left[\frac{m_u^q - m_u^{\beta+1}}{1-m_u}\right]\right)A_0 + \chi A_0 +$

$b\left(\frac{sP_1}{(\lambda*(1+\vartheta)+b)m_u - \lambda*(1+\vartheta))}\right).$

It is expressed as,

$$Q\frac{\mu(1-m^q)}{(1-m)}=\frac{\chi}{(1-m_u)}-\frac{\Upsilon\mu\left(1-m_u^q\right)}{(1-m_u)}+b\left(\frac{sp_1}{(\lambda*\ (1+\vartheta)+b)(Exp-\lambda*\ (1+\vartheta))}\right)$$

$$Q=\left[\frac{\mu(1-m^q)}{(1-m)}-\frac{sbm}{\lambda*(1+\vartheta)(m-1)+bm}\right]^{-1}\left[\frac{\chi}{(1-m_u)}-\frac{\Upsilon\mu(1-m_u^q)}{(1-m_u)}+\frac{sb\Upsilon m_u}{\lambda*(1+\vartheta)(m_u-1)+bm_u}\right].$$

As an outcome, the steady-state queue-size probabilities A_0 and are provided by,

$$A_k=m_u^kA_0;(k\geq 0), \tag{15}$$

$$P_k=(Qm^k+\Upsilon m_u^k)A_0;(k\geq 0), \tag{16}$$

$$Q=\left[\frac{\mu(1-m^q)}{(1-m)}-\frac{sbm}{\lambda*(1+\vartheta)(m-1)+\Upsilon m}\right]^{-1}\left[\frac{\chi}{(1-m_u)}-\frac{\Upsilon\mu(1-m_u^q)}{(1-m_u)}+\frac{sb\Upsilon m_u}{\lambda*(1+\vartheta)(m_u-1)+bm_u}\right].$$

Where,

$$\Upsilon=\frac{\chi m_u(\lambda(1+\vartheta)\,(m_u-1)+bm_u)}{\left[\lambda*(1+\vartheta)\,(m_u-1)+\mu m_u\left(1-m_u^{\beta}\right)+sm_u\right][\lambda*(1+\vartheta)(m_u-1)+bm_u]-sbm_u^2}$$

$$M_k=\left[\frac{\mu}{\lambda*(1+\vartheta)}\left(\frac{Q(1-m^{k+1})}{(1-m)}+\frac{\Upsilon(1-m_u^{k+1})}{(1-m_u)}\right)+\frac{\mu_u}{\lambda*(1+\vartheta)}\left(\frac{1-m_u^{k+1}}{1-m_u}\right)\right]A;(0\leq k\leq q-1). \tag{17}$$

Using the normalizing condition, the expression for A_0 is determined as follows:

$\sum_{k=0}^{\infty}A_k+\sum_{k=0}^{\infty}P_k+\sum_{k=0}^{q-1}M_k+\sum_{k=0}^{\infty}\Upsilon_k=1.$

Which follows that,

$$A_0^{-1}=E(m_u,\mu_u)+QE(m,\mu)+\Upsilon E(m_u,\mu)+QS(m,s)+\Upsilon S(m_u,s).$$

Where $E(\pi,\delta)=\frac{1}{(1-\pi)}\left[1+\frac{\delta}{\lambda*(1+\vartheta)}\left(q-\frac{\pi(1-\pi^q)}{(1-\pi)}\right)\right]$ and $S(\pi,\delta)=\frac{1}{(1-\pi)}\left[\frac{\pi\delta}{\lambda*(1+\vartheta)(\pi-1)+b\pi}\right]$.

Thus, the value A_0^{-1} is evaluated.

III. Evaluating Performance

In this section, we have performance metrics of M/(q, **β**)/1/MWV/ϒm computed.

The expected queue- length (ι_a) is,

$$\iota_a=\sum_{k=1}^{\infty}k(A_k+P_k+\Upsilon_k)+\sum_{k=1}^{q-1}kM_k. \tag{18}$$

By Substituting the values of A_k , P_k, $\Upsilon_k$and M_k from (15) to (17), we get

$$\iota_a=\sum_{k=1}^{\infty}k\left(m_u^k+(Qm^k+\Upsilon m_u^k)\right)+\sum_{k=1}^{q-1}k\left[\frac{\mu}{\lambda*(1+\vartheta)}\left(\frac{Q(1-m^{k+1})}{(1-m)}+\frac{\Upsilon(1-m_u^{k+1})}{(1-m_u)}\right)+\frac{\mu_u}{(\lambda*(1+\vartheta)}\left(\frac{(1-m_u^{k+1})}{(1-m_u)}\right)\right]+$$

$$\sum_{k=1}^{\infty}k\left(\frac{smQm^k}{(\lambda*(1+\vartheta)+b)m-\lambda(1+\vartheta)}+\frac{sm_u\Upsilon m_u^k}{(\lambda*(1+\vartheta)+b)m_u-\lambda(1+\vartheta)}\right).$$

Moreover, ι_a can be simplified as,

$$\iota_a=QG(m,\mu)+\Upsilon G(m_u,\mu)+G(m_u,\mu_u)+QC(m,s)+\Upsilon C(m_u,s). \tag{19}$$

Where $G(\pi,\delta)=\frac{\pi}{(1-\pi)^2}+\frac{\delta}{\lambda(1+\vartheta)(1-\pi)}\left\{\frac{q(q-1)}{2}+\frac{q\pi^{q+1}(1-\pi)-\pi^2(1-\pi^2)}{(1-\pi)^2}\right\}$,

$C(\pi,\delta)=\frac{\pi}{(1-\pi)^2}\left[\frac{\pi\delta}{\lambda(1+\vartheta)(\pi-1)+b\pi}\right]$.

Now P_u, $P_{by}(busy)$, $P_{ie}(idle)$ and $P_{\beta m}$ are given by,

$P_u=\sum_{k=0}^{\infty}A_k=\frac{A_0}{(1-m_u)}$,

$$P_{by}=\sum_{k=0}^{\infty}P_k=\sum_{k=0}^{\infty}(Qm^k+\Upsilon m_u^k)A_0=\left[\frac{Q}{(1-m)}+\frac{\Upsilon}{(1-m_u)}\right]A_0,$$

$P_{ie}=\sum_{k=0}^{q-1}M_k$,

$P_{\beta m}=\sum_{k=0}^{\infty}\Upsilon_k=\frac{Qsm}{(\lambda*\ (1+\vartheta)+b)m-\lambda*\ (1+\vartheta)}\left[\frac{m}{(1-m)^2}\right]+\frac{\Upsilon sm_u}{(\lambda*\ (1+\vartheta)+b)m_u-\lambda*\ (1+\vartheta)}\left[\frac{m_u}{(1-m_u)^2}\right]$.

IV. Conclusion

The analysis of encouraged arrival multiple working vacation queuing model under the steady state condition is discussed in this study in the context of the server being busy, idle state, regular, breakdown busy–state. Using Chapman Kolmogorov balancing equations, the total probability generating function was determined for this model. This model is more effective than the comparative Poisson arrival model [25].In the future to be included EASTA property for this encouraged arrival multiple vacation queuing model.

References

[1]. Afthab Begum, M. I, queuing models with bulk service with a vacation, Ph.D. Thesis awarded by Bharathiar University, Coimbatore 1996.

[2]. Baba, Y. (2005). Analysis of a GI/M/I queue with multiple working vacations. *Operations Research Letters*, 33:201-205.

[3]. Jaishwal, N. K. (1960). Time dependent solution of the bulk queuing service queuing problems. *Operations Research Society*, 8:773-778.

[4]. Julia Rose Mary, K. and Afthab Begum, M. I . (2009). Closed form Analytical solution of the General Bulk service queuing model M/M(a,b)/1 under working vacation, International conference on Mathematical and Computational models, PSG College of Tech. 92-100.

[5]. Ke, J. C. and Pearn, W. L. (2004). Optimal management policy for heterogeneous arrival queuing systems with server breakdowns and vacations. *Quality Technology & Quantitative management*, 1:149-162.

[6]. Kim, J. Choi, D. and Chae, K. (2003). Analysis of queue length distribution of the M/G/1 queue with working vacations, International Conference on Statistics and Related fields.

[7]. Tian, N. Li, J. and Zhang, G. (2009). Matrix analytic method and working vacation queue-A survey. *International Journal of Information Management Sciences*, 20:603-633.

[8]. Neuts, M. F. Matrix analytic methods in queuing theory in advance queuing, 1995.

[9]. Li, J. and Tian, N. (2007). The M/M/1 queue with working vacations and vacation interruptions. *Journal of Systems Science and Systems Engineering*, 16:121-127.

[10]. Zhang, Z. and Xu, X. (2008). Analysis for the M/M/1 queue with multiple working vacations and N- Policy. *Information and Management services*, 19(3):495-506.

[11]. Wu, D. and Takagi, H. (2006). An M/G/1 queue with multiple working vacations. *Performance Evalution*, 63:654-681.

[12]. Neuts, M. F. (1967). A general class of bulk queues with Poisson input. *Annals of Mathematical Statistics*, 38:759-770.

[13]. Wang, K. H. (1997). Optimal control of an $M/E_k/1$ queuing system with removable service station subject to breakdowns. *Journal of Operational Research Society*, 48:936-942.

[14]. Wang, K. H. (1995). Optimal Operation of a Markovian queuing system with a removable and non-reliable server. *Microelectronics Reliability*. 35:1131-1136.

[15]. Servi, L. D. and Finn, S. G. (2002). M/M/1 queues with working vacations (M/M/1/WV), *Performance Evaluation*, 50:41-52.

[16]. Li, J. Tian, N. and Zhang, Z. D. (2009). Analysis of M/G/1 queue with exponential working vacations-A matrix analytic approach. *Queuingsystems*, 61:139-166.

[17]. Medhi, J. Recent developments in bulk and queuing models, John Wiley Eastern Limited, New Delhi, 1984.

[18]. Neuts, M. F. Matrix geometric solutions in stochastic models, Johns Hopkins university press baltmore, 1981.

[19]. Bailey, N. T. J. (1954). On queuing process with bulk service, *Journal of the Royal Statistical Society*, 16:80-87.

[20]. Wang, K. H. Chang, K. W. and Sivazlian, B. D. (1999). Optimal control of a removable and non-reliable server in an infinite and a finite $M/H_2/1$ queuing system. *Applied Mathematical Modelling*, 23:651-666.

[21]. Tian, N. Zhao, X. and Wang, K. (2008). The M/M/1 queue with single working vacation. *International Journal of Information Management Sciences*, 19:621-634.

[22]. Xu, Z. Zhang, Z. and Tian, N. (2009). The M/M/1 queue with single working vacation and setup times. *International Journal of Operational Research*, 6:420-434.

[23]. Som, B.K. and Seth, S. (2017). An M/M/1/N Queuing system with Encouraged Arrivals, *Global Journal of Pure and Applied Mathematics*,13:3443-3453.

[24]. Khan,I.E. Paramasivam, R. (2022). Reduction in Waiting Time in an M/M/1/N Encouraged Arrival Queue with Feedback, Balking and Maintaining of Reneged Customers. 14:1743.

[25]. Rajalakshmi, R. Pavithra, J. and Julia Rose Mary, K. (2016). Steady State Analysis of M/M(a,b)/1/MWV/Br Queuing Model. *International Journal of Innovative Research in Science, Engineering and Technology*, 5(3):1-6.

TECHNICAL DEVICE WEAR-OUT PERIOD INFLUENCE ON QUANTITATIVE RISK ASSESSMENT RESULTS

Yuriy D. Kuznetsov[1], Evgeniy Yu. Kolesnikov[2]

•

Peter the Great Saint-Petersburg Polytechnic University, Saint-Petersburg, Russia
[1]yura.kuz30@gmail.com
[2]key3108@yandex.ru

Abstract

Hazardous production facilities contain numerous technical devices, the reliability assessment of which is a part of quantitative risk assessment. The paper considers the pressure valve as a safety system element of equipment operating under excessive pressure and evaluates its reliability (survival function value) during the operational period. Valve reliability during the wear-out period has been modeled to assess wear-out period influence on this element failure probability. Modeling was carried out by approximating the failure rate tabular values obtained based on statistical data. Approximation was carried out by: a second-degree polynomial, the Weibull distribution law and a power function. Comparison of the obtained quantitative estimates with the element failure probability, calculated without taking into account the wear-out period, showed necessity of wear-out period influence consideration in risk assessment procedure.

Keywords: reliability, quantitative risk assessment, modeling, hazardous production facility, pressure valve

1. Introduction

One of the main tools for performing accident quantitative risk assessment (QRA) at hazardous production facilities (HPF) is logical-probabilistic modeling (LPM), in particular, fault tree analysis (FTA), [1, 2].

HPF are complex technical systems (TS) consisting of technological blocks, technical devices, and elements. HPF safety is ensured by specialized safety units / elements, used for technological process deviation prevention from escalating into an accident.

As is well known, there are no absolutely reliable technical devices. Any device (technical system) failure probability is calculated based on its components (elements) failure probabilities which are usually estimated by consideration of failure probability dependence from operating time described by exponential law. In this case, element failure rate is postulated by a constant, time-independent value.

The choice of described mathematical model for estimating element failure probability is based on:

- its mathematical simplicity.
- the fact that the longest period in the element life cycle is the useful life period which is characterized by failure rate approximate constancy, [1].

Described model does not consider elements wear-out period, characterized with failure rate significant increase. Neglect of this circumstance can result in inadequate technical device failure probability quantitative estimates, and, consequently, in distorted results of various accident scenarios probability assessment obtained from event tree analysis (ETA). ETA results are sensitive

to the accuracy of initial data, including the probability of both the initiating event and the safety device's conditional failure probabilities.

At present time, there is not yet a single generally accepted approach to technical devices reliability assessment in the wear-out period. However, many researchers have repeatedly drawn attention to the need for wear-out period consideration in technical device reliability assessment [3, 4, 5, 6]. The prevailing opinion is that the Weibull distribution is the most suitable for elements reliability assessment in the wear-out period [4, 5, 7].

Besides, there is no consensus among researchers regarding technical devices' imperfect maintenance impact on the reliability/failure rate during the wear-out period. The most common are two concepts: PAR (Proportional Age Reduction) and PAS (Proportional Age Setback) [4, 6]. Described concepts implementation allows for assess maintenance impact on the technical device aging process, varying it from completely ignoring such an influence (BAO, Bad As Old, postulating that the degree of device degradation does not decrease after maintenance); to completely eliminating degradation during maintenance (GAN, Good As New).

The choice of the most suitable mathematical model for describing technical device wear-out period is carried out on the basis of the Akaike information criterion (AIC), which allows choosing among the models under consideration the one that has the least number of parameters and will have the best approximation to the available data. The authors of [4] argue that for the given purposes, the most suitable model is the Weibull distribution with maintenance effectiveness equal to 1 (GAN). Moreover, the authors believe that usage of the PAR/PAS concept with a different maintenance efficiency will result in significant failure rate value overestimation when predicting the state of the technical device, [4]. This conclusion is based on the technical device's statistical data acquired from 17 years of observance comparison with their mathematical assessment.

Yet another approaches for reliability assessment in the wear-out period exist. Research, [5], in addition to the widely recognized wear-out period modeling approach of the device by the Weibull distribution, proposes the usage of power distribution. The type of power distribution proposed by its author allows assessment of technical devices' reliability throughout their entire life cycle and various types of failure rate functions usage.

The authors of [3, 8] consider the elements' wear-out period consideration problem from a more general perspective, using algorithms [3], and complexes of techniques [8] to influence the assessment of the wear-out process on technical devices' overall performance.

The authors of this paper point out that none of the analyzed studies carried out a prediction of the technical device survival function values for its operational period. The articles reviewed describe only possible approaches to taking into account the device wear-out period influence on the process of its operation and predicting the failure rate magnitude. This results in the impossibility of different wear-out period reliability assessment approach comparisons. Due to quantitative assessment absence wear-out period impact on the value of survival function is unclear. Discovered uncertainty in technical system reliability assessment throughout its entire life cycle (including the wear-out period) leads to an increase of accident risk indicators value uncertainty.

This paper is dedicated to the demonstration of the fact that neglect of technical system elements wear-out period in its reliability assessment leads to an underestimation of their failure probability, which results in an underestimation of the magnitude of accident risk indicators (potential territorial, individual, etc.).

The purpose of this article is to substantiate the need for technical system element wear-out period consideration in the domain of a quantitative risk assessment.

2. Methods

In this paper authors have considered vessel operating under excessive pressure as an example of technical device and its safety valve as a technical device element. Reliability assessment has been conducted for safety valve (pressure valve) with operational period equal to 30 years [9]. For further

analysis, it is necessary to describe the dependence of the failure rate [10] as a function of time (operating time). In this paper following assumptions are made: total duration of burn-in period and useful life period is 27 years, wear-out period duration is 3 years (figure 1).

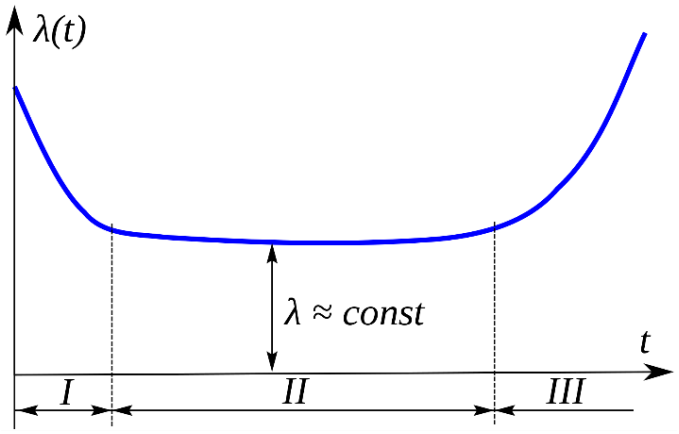


Figure 1: *The bathtub curve: I – burn-in period, II – useful life period, III – wear-out period*

Consider the valve failure rate value at three points in time during the wear-out period:

- at the beginning of wear-out period – λ_{min};
- at the middle of wear-out period – λ_{mean};
- at the end of wear-out period – λ_{max} (concurrently – at the time of decommissioning of the valve due to its reaching the limit state).

As reference values of the failure rate during the wear-out period statistical data was used [11]. it is assumed that during useful life period valve failure rate was equal to ⊚min. The relation between the failure rates values and time points is shown in Table 1.

Table 1: *Tabular failure rate function*

Designation	Time point t, h	Value of λ(t), h^{-1}
λ_{min}	236520	$0{,}122 * 10^{-6}$
λ_{mean}	249660	$5{,}6 * 10^{-6}$
λ_{max}	262800	$32{,}5 * 10^{-6}$

The time point equal to 236520 hours, i.e. 27 years from the beginning of the device operation, is taken as the beginning of the wear-out period. The middle of the wear-out period corresponds to a time point equal to 249660 hours, i.e. 28 years from the beginning of device operation. The end of the wear-out period (device operation end) corresponds to a time point equal to 262800 hours, i.e. 30 years from the beginning of device operation. In order to quantify the valve wear-out period effect on its failure probability, different failure rate functions were proposed.

As it well-known, unrepairable element failure probability depends on its failure rate in a following way:

$$Q(t) = 1 - e^{-\int_0^t \lambda(\tau)d\tau} = e^{-\left(\int_0^{t_1} \lambda_2(\tau)d\tau + \int_{t_1}^{t_2} \lambda_3(\tau)d\tau\right)} \quad (1)$$

where λ_2 – failure rate, h^{-1}, of an element during the useful life period, can be equated to λ_{min};
λ_3 – failure rate, h^{-1}, of an element during the wear-out period;
t_1 – wear-out period beginning, h, equals to 236520 h.

2. 1. Approximation by a second degree polynomial

Considering the data given in Table 1 as a tabular function λ(t), it is possible to approximate it [12] with a second degree polynomial represented by a following expression (2):

$$\lambda_{app}(t) = 6{,}2 \cdot 10^{-14} \cdot t^2 - 2{,}97 \cdot 10^{-8} \cdot t + 3{,}56 \cdot 10^{-3} \quad (2)$$

The approximation parameters are given in Table 2.

Table 2: *Second degree polynomial approximation accuracy*

Parameter	Value
Correlation coefficient	1
Coefficient of determination	1
Average approximation relative error	0%

Therefore, the survival function of a process described by exponential law with failure rate (2) for wear-out period and constant failure rate for burn-in period and useful life period, can be represented by following expression (3):

$$P_{app}(t) = e^{-\left(\int_0^{t_1} \lambda_{min} dt + \int_{t_1}^{t} \lambda_{app}(t) dt\right)} \quad (3)$$

Where $P_{app}(t)$ – the probability of failure-free operation based on a wear-out period failure rate approximation by second degree polynomial;

λ_{min} – useful life period failure rate, h^{-1};

t_1 – time point corresponding to the end of the useful life period and the beginning of the wear-out period. It is assumed to be equal to 236520 hours;

t – time, measured in hours, $t > t_1$;

$\lambda_{app}(t)$ – the failure rate value during the wear-out period at time point t, obtained from the approximation of the tabular function by a polynomial of the second degree (2), h^{-1}.

2. 2. Calculation of failure rate function λ(t) on the basis of Weibull distribution law

The probability of failure-free operation of the element during the wear-out period in this case [7] will have the form (4):

$$P_{wb}(t) = e^{-\left(\int_0^{t_1} \lambda_{min} dt + (\beta t)^{\alpha} - (\beta t_1)^{\alpha}\right)} \quad (4)$$

Where $P_{wb}(t)$ – the probability of failure-free operation under the assumption that the failure rate obeys the Weibull distribution law;

β – the rate parameter of Weibull distribution;

λ_{min} – useful life period failure rate, h^{-1};

t_1 – time point corresponding to the end of the useful life period and the beginning of the wear-out period. It is assumed to be equal to 236520 hours;

t – time, measured in hours, $t > t_1$

α – the shape parameter of Weibull distribution.

Thus, in order to find the element failure-free operation probability value, Weibull distribution parameters (α and β) values must be determined. For this purpose, it is necessary to use the failure rate function for Weibull law (5):

$$\lambda_{wb}(t) = \alpha \cdot \beta^{\alpha} \cdot t^{\alpha-1} \quad (5)$$

Where β – the rate parameter of Weibull distribution;

α – the shape parameter of Weibull distribution.

Equating the values of this function to the available table values, a system (6) of two nonlinear equations for the failure rate was obtained:

$$\begin{cases} \lambda_{max} = \beta^{\alpha} {t_{max}}^{\alpha-1} \\ \lambda_{min} = \beta^{\alpha} {t_{min}}^{\alpha-1} \end{cases} \quad (6)$$

Its solution gives an expression for determining the parameters α:

$$\alpha = \frac{\ln\left(\frac{\lambda_{max}}{\lambda_{min}}\right)}{\ln\left(\frac{t_{max}}{t_{min}}\right)} + 1 \quad (7)$$

and β:

$$\beta = \left(\frac{\lambda_{max}}{{t_{max}}^{\alpha-1}}\right)^{\frac{1}{\alpha}} \quad (8)$$

Acquired values of the parameters (α and β) are shown in Table 3.

Table 3: *Weibull law parameter values*

Parameter	Value
α	54,82
β	$3{,}68 \cdot 10^{-6}$

2. 3. Power function approximation

Consider the following function describing failure rate in wear-out period:

$$\lambda_{deg}(t) = \lambda_0 + \zeta \cdot \mathrm{t}^{\gamma} \quad (9)$$

where λ_0 – failure rate at the beginning of wear-out process, h^{-1};

γ and ζ – function parameters.

For parameters γ and ζ value determination equality of the function $\lambda_{\mathrm{deg}}(\mathrm{t})$ to the failure rate values based on statistical data (Table 1) at two time points is required:

$$\begin{cases} \lambda_{max} = \lambda_0 + \zeta \cdot {t_{max}}^{\gamma} \\ \lambda_{mean} = \lambda_0 + \zeta \cdot {t_{mean}}^{\gamma} \end{cases} \quad (10)$$

Parameter γ can be determined from the following expression:

$$\gamma = \frac{\ln\left(\frac{\lambda_{max} - \lambda_0}{\lambda_{mean} - \lambda_0}\right)}{\ln\left(\frac{t_{max}}{t_{mean}}\right)} \quad (11)$$

Substituting the obtained value of the parameter γ into any of the equations (10), calculation of parameter ζ is possible:

$$\zeta = \frac{\lambda_{max} - \lambda_0}{{t_{max}}^{\gamma}} \quad (12)$$

For the case under consideration, the values of the parameters γ and ζ obtained by us are shown in Table 4.

Table 4: *Weibull law parameter values*

Parameter	Value
γ	34,61
ζ	$8{,}7 \cdot 10^{-193}$

3. Results

As a result of this research, five functions that characterize the failure probability during the wear-out period were determined and compared:

- $Qp_{min}(t)$ – a function describing the simulation of the failure probability at a constant failure rate λ_{min} in a process obeying an exponential law. Thus, the effect of the wear-out process on the failure probability is completely ignored.
- $Qp_{mean}(t)$ – a function describing the simulation of the failure probability at a constant failure rate λ_{mean} in a process obeying an exponential law. Wear-out process is taken into account by changing the parameter in the process, obeying the exponential law by the average failure rate value during wear-out period.
- $Q_{wb}(t)$ – a function describing the simulation of the failure probability with a failure rate varying according to Weibull law with parameters α, β. Choice of this distribution allows accounting of failure rate growth due to aging.
- $Q_{app}(t)$ – a function describing the simulation of the failure probability with a failure rate varying according to the law $\lambda_{app}(t)$. Wear-out process is taken into account by approximating statistical data on the failure rate during the wear-out period.
- $Q_{deg}(t)$ – a function describing the simulation of the failure probability with a failure rate varying according to the law $\lambda_d(t)$. The approximating function is obtained based on the assumption that the number of failed elements of the same type obeys the normal distribution law.

Failure probability of the pressure relief valve is calculated according to equation (1), the results are presented in Figure 2.

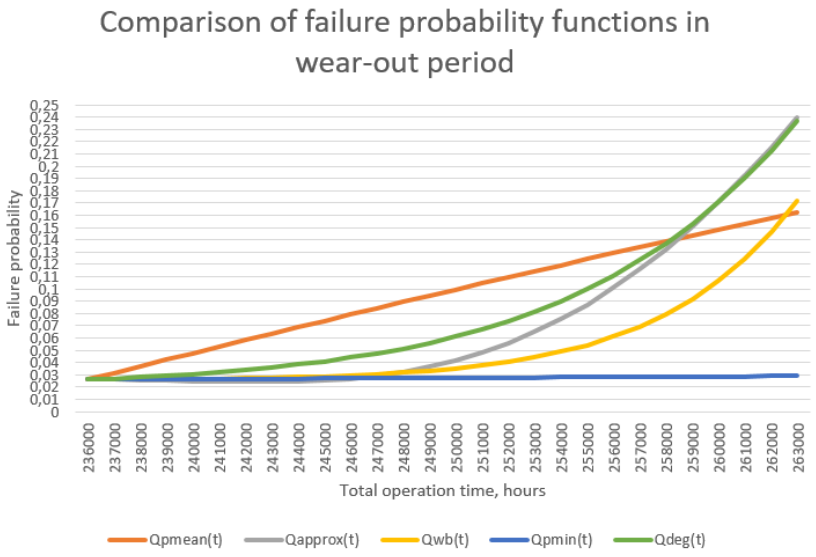


Figure 2: *Comparison of failure probability functions in wear-out period*

Values of the considered failure probability functions at the end of device operation period are given in Table 5.

Table 5: *Valve failure probability at the end of its operation period*

Type of function	Function value at time point t = 263000 h
Qpmin(t)	0,0290264
Qpmean(t)	0,16275
Qwb(t)	0,171749
Qapprox(t)	0,239714
Qdeg(t)	0,236512

4. Discussion

The analysis of Figure 2 shows that the graph of the function $Q_{app}(t)$ decreases in some area, which is determined by the type of the approximating function. Since this contradicts the meaning of the concept of the failure probability, which is a non-decreasing function of operating time, further consideration of this function is pointless.

From the data given in Table 5, it follows that the most conservative estimate of the options considered is the $Q_{deg}(t)$ function. Moreover, it becomes most conservative estimate only at the final stage of the wear-out period. As follows from the graph shown in Figure 2, up to a certain point, the most conservative estimate is $Qp_{mean}(t)$. Assessment of failure probability provided by this type of function is rather rough, but at the same time calculation of this function value is rather simple.

It is also important that the complete disregard of the wear-out period in QRA (which is a fairly common practice) leads to an underestimation of the device failure probability (compared with other methods of assessment).

In this work, a quantitative assessment of pressure relief valve failure probability during wear-out process was obtained, justifying the need to take this period into account when conducting the QRA. This study can be considered as the first stage of assessing the impact of taking into account the wear-out period of technical devices on the risk indicators assessment. Obtained estimates clearly show that accounting of wear-out process in reliability assessment will result in accident risk magnitude increase. Yet it remains unclear how strong is impact on accident risk magnitude. Next studies will be aimed on evaluation of technical systems elements wear-out process impact on the subsequent stages of QRA.

References

[1]. Akimov V. A., Lapin V. L., Popov V. M., Puchkov V. A., Tomakov V. I., Faleev M. I., *Nadezhnost' tehnicheskih sistem i tehnogennyj risk* [Reliability of technical systems and technogenic risk]. Moscow, JSC Financial Publishing House «Delovoj ekspress», 2002, 386 p.

[2]. Order of the Federal Service for Environmental, Technological and Nuclear Supervision dated 03.11.2022 No. 387 "On Approval of the Safety Manual "Methodological foundations of hazard analysis and Accident Risk Assessment at Hazardous Production facilities". Available at: https://www.consultant.ru/document/cons_doc_LAW_433652/ (Accessed: 06 May 2023).

[3]. Monika Sandelic, Saeed Peyghami, Ariya Sangwongwanich, Frede Blaabjerg, Reliability aspects in microgrid design and planning: Status and power electronics-induced challenges, Renewable and Sustainable Energy Reviews,Volume 159, 2022, 112127, ISSN 1364-0321.

[4]. I. Martón, A.I. Sánchez, S. Carlos, R. Mullor, S. Martorell, Prognosis of wear-out effect on of safety equipment reliability for nuclear power plants long-term safe operation, Reliability Engineering & System Safety, Volume 233, 2023, 109121, ISSN 0951-8320.

[5]. Volodarskij, V. A. *O raspredelenijah dlja opisanija otkazov tehnicheskih ustrojstv* [About distributions for describing failures of technical devices]. *Metody menedzhmenta kachestva* [Quality management methods]. – 2014. – № 4. – pp. 50-56. – EDN SCYUZF.

[6]. Martorell, Sebastian & Martorell, Pablo & Sánchez, Ana & Mullor, Ruben & Martón, Isabel. (2017). Parameter Estimation of a Reliability Model of Demand-Caused and Standby-Related Failures of Safety Components Exposed to Degradation by Demand Stress and Ageing That Undergo Imperfect Maintenance. Mathematical Problems in Engineering. 2017. 1-11. 10.1155/2017/7042453.

[7]. Rausand, Marvin. System reliability theory: models, statistical methods, and applications / Marvin Rausand, Arnljot Hoyland. - 2nd ed. – 636 p.

[8]. Ostrejkovskij, V. A., Sorochkin A. V., *Modeli i metody statisticheskoj teorii nadezhnosti v razvitii koncepcii urovnej opisanija starenija oborudovanija slozhnyh sistem s dlitel'nymi srokami aktivnogo sushhestvovanija* [Models and methods of the statistical theory of reliability in the development of the concept of levels of description of equipment aging of complex systems with long periods of active existence]. *Nadezhnost' i kachestvo slozhnyh system* [Reliability and quality of complex systems]. – 2022. – № 3(39). – pp. 5-15. – DOI 10.21685/2307-4205-2022-3-1. – EDN XEJWAI.

[9]. Filar. Klapany izbytochnogo davlenija [Filar. Overpressure valves]. Available at: https://filar.ru/svedenia2.html (Accessed 11 March 2023).

[10]. Andreev A. V., Yakovlev V. V., Korotkaja T. Yu., *Teoreticheskie osnovy nadezhnosti tehnicheskih sistem: uchebnoe posobie* [Theoretical foundations of reliability of technical systems: textbook]. Saint-Petersburg: Polytechnic University Publishing House, 2018. – 164 p.

[11]. National Standard GOST 12.1.004-91. Interstate Standard. The system of occupational safety standards. Fire safety. General requirements. Available at: https://docs.cntd.ru/document/9051953 (Accessed 06 May 2023);

[12]. PLANETCALC. Online calculators: Available at: https://planetcalc.ru/5992/ (Accessed 06 May 2023).

BEHAVIOR ANALYSIS OF WASHING UNIT IN A PAPER PLANT EMPLOYING FUZZY APPROACH

Mamta[1], Seema Sharma [2, *]

•

[1]Research Scholar, Department of Mathematics and Statistics, Gurukul Kangri (Deemed to be University) Haridwar, Uttarakhand, India-249404
[2]Professor, Department of Mathematics and Statistics, Gurukul Kangri (Deemed to be University) Haridwar, Uttarakhand, India-249404

[1]e-mail: rs.mamta@gkv.ac.in, [2]e-mail: seema@gkv.ac.in
*Corresponding Author

Abstract

Aim. *The purpose of this research is to employ a fuzzy approach to assess the system behavior of the washing unit in a paper plant using vague, uncertain and inaccurate data. The washing unit is the main operational part of a paper plant for which analysis of system behavior is important to choose an appropriate maintenance strategy. The analysis has been carried out for washing unit of a paper plant situated in northern India.* ***Methods****. The proposed approach comprises qualitative and quantitative analysis. In qualitative analysis, the basic arrangement of the washing unit is modelled by Petri Net model. In quantitative analysis, the fuzzy λ-τ approach has been used for analyzing the systems' failure behavior more accurately. Uncertainties in failure/repair data of every subsystem/component of the washing unit are quantified using trapezoidal fuzzy numbers.* ***Results.*** *To assess the performance and failure dynamic behavior of the washing unit quantitatively, six reliability parameters including failure rate, repair time, mean time between failure, expected number of failures, reliability and availability at three different spread levels have been evaluated employing trapezoidal fuzzy numbers. The fuzzified values of these reliability parameters of washing unit have been defuzzified employing center of area defuzzification technique. Further, crisp values and defuzzified values of these parameters using triangular fuzzy numbers have also been obtained. The results obtained by the proposed methodology have been compared with those obtained by fuzzy λ-τ approach based on triangular fuzzy numbers. The information/results obtained through the fuzzy λ-τ approach with trapezoidal fuzzy number are conservative in nature, therefore, these results may be used by system specialist/system analysts for the future plan of implementation.* ***Conclusion.*** *Using this approach, six reliability parameters are evaluated and the trend (increase or decrease) of these reliability parameters is examined for performance analysis of washing unit in a paper plant. Based on these investigations, suitable maintenance policy can be established that will assist maintenance manager/system analysts/engineers in improving system performance by implementing appropriate preventive maintenance procedures. As a result, it will help in achieving a long time system availability and maximizing overall productivity of the paper plant. The implications of this fuzzy reliability approach to industry maintenance and operation planning are quite beneficial.*

Keywords: paper plant, uncertain data, petri net, fuzzy methodology, reliability, trapezoidal fuzzy number

1. Introduction

Reliability analysis plays a vital role in successful functioning of a repairable industrial system. Most of the repairable industrial systems consist of various subsystems. Every subsystem is comprised of several sophisticated components. It is almost impossible to completely avoid the failure in an industrial system but it can be reduced by implementing appropriate maintenance policies. System availability is also identified as an important factor of performance for such systems. To achieve the objective of improving the availability/reliability with low-cost inputs, it is required that every component or subsystem must operate adequately and provide excellent performance. The behavior of a system under specified operating conditions can be used to design its components to minimize the failure and to plan the preventive or scheduled maintenance of the system. However, today, the behavior analysis of repairable industrial system has become a great challenge for an expert/system analyst due to technological advancements and growing complexities of components/subsystems of the system.

A number of researchers evaluated the performance of numerous operational structures in different process plants, namely thermal power plant, sugar industry, paper plant, chemical industry and urea plant. The probability of survival of an industrial system depends upon all of its basic components. Thus, the behavior of these components will assist in the analysis of its overall performance. The behavior and performance of such systems are often assessed using reliability, system availability and other reliability parameters. Many different techniques such as the Markovian approach, reliability block diagrams (RBDs), fault tree analysis (FTA), Petri Nets (PNs) and others are widely used for reliability analysis of repairable industrial systems [1-8]. However, nowadays, among these techniques, FTA and PN [5-10] are being preferred by the researchers to analyze the failure of repairable systems.

To deal with the uncertainties in data, Knezevic and Odoom [11] developed the λ-τ approach by using the fuzzy theory and the PN model. In their methodology, triangular fuzzy numbers (TFNs) have been employed to address the vagueness in the failure/repair information. Sharma et al. [12] used fuzzy λ-τ approach for behavior analysis of a large industrial system employing FTA and PNs. Sharma et al. [13] applied FTA and the fuzzy λ-τ approach to examine the reliability of complex robotic unit. Sharma et al. [14] analyzed the system behavior and performance for two grinding machines employing fuzzy λ-τ approach with PN. Garg et al. [15] analyzed the system reliability parameters of screening unit of a paper plant using a fuzzy approach. Verma et al. [16] examined the vague reliability of the combustion unit employing vague λ-τ approach with PN. Panchal et al. [17] employed a fuzzy approach for RAM and risk analysis of a chemical industry.

To fuzzify clinical information/data, Princy and Dhenakaran [18] compared trapezoidal fuzzy number (TrFN) and TFN and discovered that the classification performance of TrFN is better than that of TFN. Most recently, Sharma and Sushma [19, 20] used TrFNs to estimate the performance and behavior of coal handling unit and water circulation unit, respectively, in a thermal power plant employing fuzzy approach. Further, Sharma and Mamta [21] performed reliability analysis of the feeding system of a paper plant under fuzzy environment employing TrFN.

The major purpose of this study is to quantify the uncertainties of failure/repair data of washing unit of a paper plant using TrFN for analyzing the systems' behavior more accurately. The fuzzy λ-τ approach for behavior analysis is used together with PN modelling. Further, the obtained

results from the fuzzy λ-τ approach with TrFN are compared with the results obtained using TFN. The acquired information will assist the managers in planning preventive or scheduled maintenance policies, to attain maximum system availability.

The structure of remaining part of this paper is as follows. Some important concepts of fuzzy set theory relevant to this study have been presented in section 2. The proposed methodology has been described in Section 3. Section 4 deals with the implementation of proposed methodology to perform fuzzy reliability analysis of considered unit. Lastly, section 5 presents the conclusions drawn.

2. Fuzzy Set Theory

The important concepts of fuzzy set theory used in the present study are given as [22]:

2.1. Fuzzy Set

A fuzzy set $\tilde{A}$ defined on universal set X is described by

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) : x \in X\} \ , \tag{1}$$

where, $\mu_{\tilde{A}}(x) \in [0,1]$ denotes the degree of membership for element x.

2.2. Fuzzy Number

A fuzzy number is a convex, normal fuzzy set defined on real line with bounded support.

2.3. Trapezoidal Fuzzy Number

A TrFN $\tilde{A}$ is a fuzzy number described as (t_1, t_2, t_3, t_4) with membership function as

$$\mu_{\tilde{A}}(x) = \begin{cases} \frac{(x-t_1)}{(t_2-t_1)}, & t_1 \le x \le t_2 \\ 1, & t_2 \le x \le t_3 \\ \frac{(t_4-x)}{(t_4-t_3)}, & t_3 \le x \le t_4 \\ 0, & \text{otherwise} \end{cases} . \tag{2}$$

The α-cut of TrFN $\tilde{A}$ given by

$$\tilde{A}^{\alpha} = [t_1^{\alpha}, \ t_4^{\alpha}] = [t_1 + (t_2 - t_1)\alpha \ , t_4 - (t_4 - t_3)\alpha] , \quad \alpha \in [0,1] \ , \tag{3}$$

is represented in Figure 1.

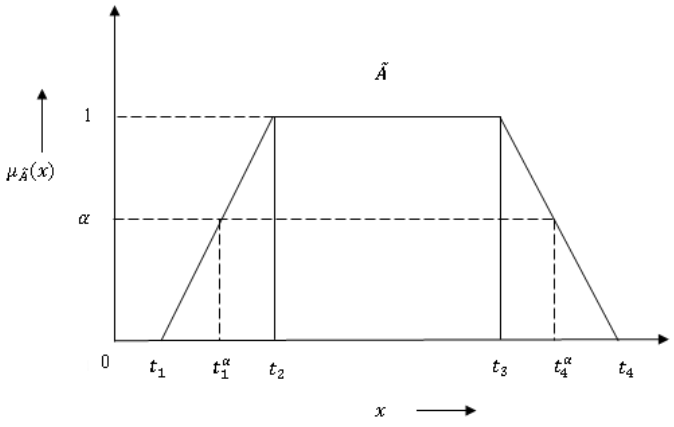


Figure 1: *α-cut for TrFN $\tilde{A}$*

3. Proposed Methodology

The behavior of a repairable system in an ambiguous environment can be analyzed using fuzzy λ-τ approach [11]. This approach uses failure and repair data of subsystems and components of considered system. This approach handles the ambiguity and vagueness in failure and repair data and hence is preferred over other approaches. The stepwise procedure of proposed methodology is as follows:

Step 1: Collection of information of various subsystems and components of washing unit.

Step 2: Construction of PN model.

Step 3: Collection of input data related to failure and repair of various subsystems and components.

Step 4: Employing TrFN for fuzzification of input data collected in step 3.

Step 5: Calculation of various reliability parameters of washing unit at different spread levels using fuzzy λ-τ approach.

Step 6: Defuzzification of fuzzified reliability parameters.

Step 7: Behavior analysis of washing unit.

4. System Description

There are several working units serving different purposes in a paper plant. In this paper, one of the main functional unit i.e. the washing unit of a paper plant situated in northern India has been studied. Washing unit of the paper plant washes the wood pulp coming out of the pulping unit with water to remove any chemicals or black liquor and prepares the fine fiber from the pulp. The schematic diagram of washing unit [23] is depicted in Figure 2.

The washing unit consists of four main subsystems, which are as follows:

- **Filter [S_1]**: This subsystem has a single filter unit, which extracts black liquor from prepared wood pulp.
- **Cleaner [S_2]**: This subsystem consists of three components, connected in parallel arrangement. These components are utilized for cleaning the pulp using centrifugal motion. Failure of one component reduces system efficiency and paper quality.
- **Screener [S_3]**: This subsystem consists of two components, which are connected in series. The wood pulp is strained by these components to separate large, unprepared and irregular fibers from it. The failure of any one component will result in the unit failure completely.

Decker [S_4]: This subsystem consists two components connected in parallel arrangement. These components remove the chemicals/darkness from prepared wood pulp. The failure of both components will result in the poor quality of paper.

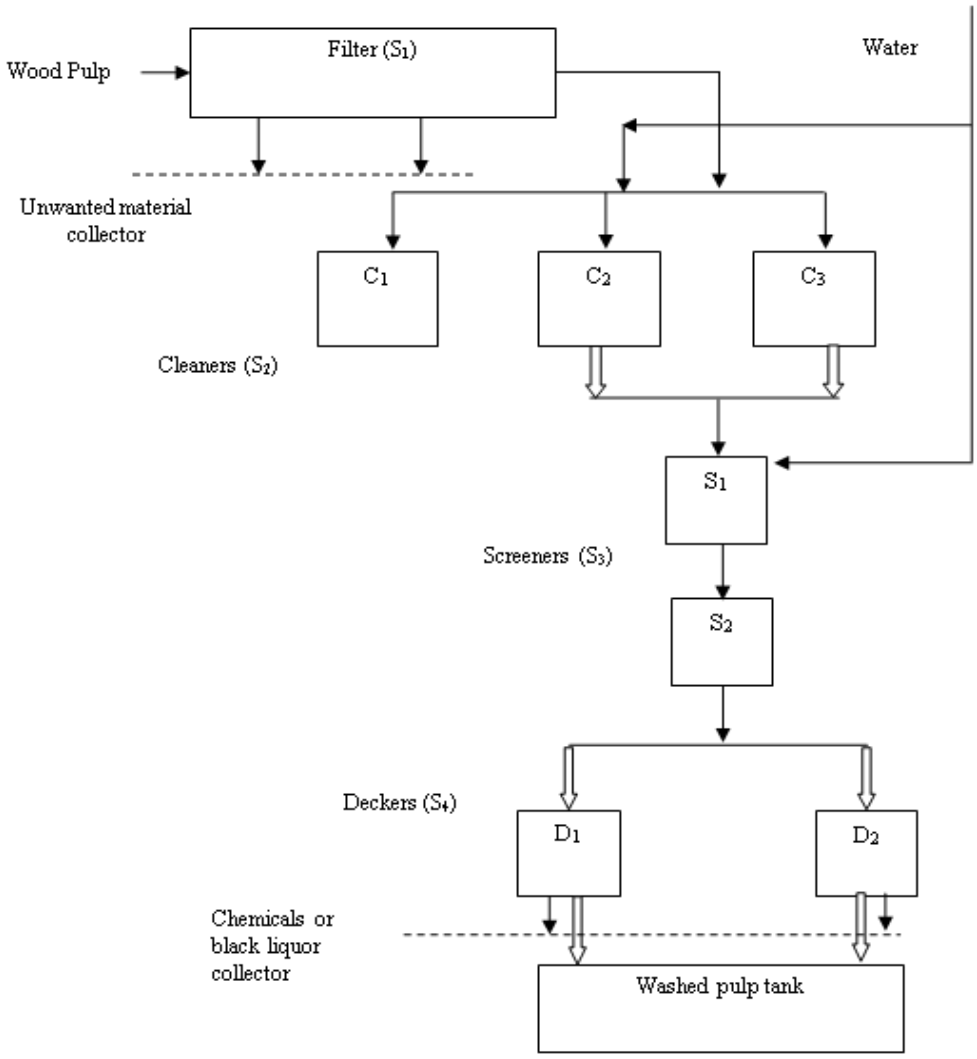


Figure 2: *The washing unit*

4.1. Reliability Analysis

The basic arrangement of the washing unit is modelled by PN model, which is depicted by constituent structure in series/parallel combination with OR/AND transitions, shown in Figure 3. The steps associated with the fuzzy λ-τ approach are presented in following subsections.

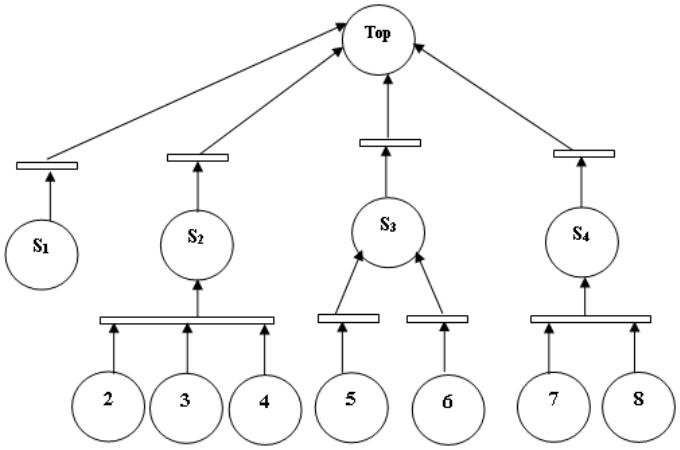


Figure 3: *PN model of washing unit*

4.1.1. Data Collection

The data regarding the failure rate (λ_k) and repair time (τ_k) for each component of various subsystems acquired from the maintenance logbook/historical system records and analyzed by maintenance professionals [23], is represented in Table 1.

Table 1: *Data for λ_k and τ_k*

Subsystems	λ_k (Failures / h)	τ_k (h)
Filter (S_1) (k=1)	1×10^{-3}	3
Cleaner (S_2) (k=2,3,4)	3×10^{-3}	2
Screener (S_3) (k=5,6)	5×10^{-3}	3
Decker (S_4) (k=7,8)	5×10^{-3}	3

4.1.2. Fuzzification of Data

The data collected from various resources is inaccurate, imprecise and ambiguous as it is acquired under different types of operating and environmental conditions. As a consequence, uncertainties in failure/repair data of each component/subsystem of washing unit are quantified by TrFNs at distinct spreads (±15%, ±25% and ±40%). Figure 4 depicts the failure rate (λ_1) and repair time (τ_1) of the filter subsystem (S_1) as TrFNs at ±15% spread.

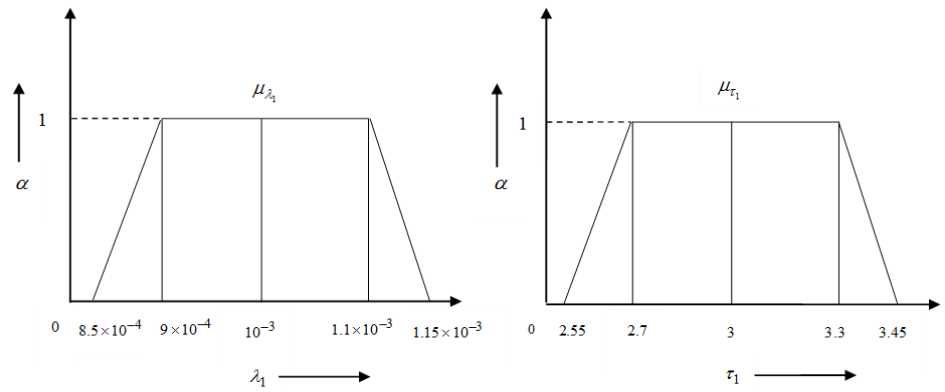


Figure 4: *TrFNs for λ_1 and τ_1 of filter subsystem at ±15% spread*

4.1.3. Reliability Parameters Estimation

After acquiring fuzzified values for λ and τ for all basic components of the washing unit, the fuzzy values for λ and τ of the top position in the PN model of the washing unit are computed by the interval expressions of AND/OR transitions given in equations (4-7). The interval expressions of TrFNs have been calculated by applying interval arithmetic operations on the respective basic expressions of λ and τ for AND/OR gates demonstrated in Table 2 together with the extension principle and alpha-cut.

Table 2: *Basic expressions of λ and τ*

Logic gate	$\boldsymbol{\lambda_{OR}}$	$\boldsymbol{\tau_{OR}}$	$\boldsymbol{\lambda_{AND}}$	$\boldsymbol{\tau_{AND}}$
n-input gate expression	$\sum_{k=1}^{n}\lambda_k$	$\frac{\sum_{k=1}^{n}\lambda_k\ \tau_k}{\sum_{k=1}^{n}\lambda_k}$	$\prod_{l=1}^{n}\lambda_l\left\{\sum_{k=1}^{n}\prod_{\substack{l=1\\k\neq l}}^{n}\tau_l\right\}$	$\frac{\prod_{k=1}^{n}\tau_k}{\sum_{l=1}^{n}\left\{\prod_{\substack{k=1\\k\neq l}}^{n}\tau_k\right\}}$

Interval expression AND transition

$$\lambda^{\alpha}=\begin{bmatrix}\prod_{k=1}^{n}\{(\lambda_{k2}-\lambda_{k1})\alpha+\lambda_{k1}\}\,.\sum_{l=1}^{n}\left\{\prod_{\substack{k=1\\k\neq l}}^{n}\{(\tau_{k2}-\tau_{k1})\alpha+\tau_{k1}\}\right\}, \\ \prod_{k=1}^{n}\{\lambda_{k4}-(\lambda_{k4}-\lambda_{k3})\alpha\}\,.\ \sum_{l=1}^{n}\left\{\prod_{\substack{k=1\\k\neq l}}^{n}\{\tau_{k4}-(\tau_{k4}-\tau_{k3})\alpha\}\right\}\end{bmatrix}, \tag{4}$$

$$\tau^{\alpha}=\left[\frac{\prod_{k=1}^{n}\{(\tau_{k2}-\tau_{k1})\alpha+\tau_{k1}\}}{\sum_{l=1}^{n}\left[\prod_{\substack{k=1\\k\neq l}}^{n}\{\tau_{k4}-(\tau_{k4}-\tau_{k3})\alpha\}\right]},\frac{\prod_{k=1}^{n}\{\tau_{k4}-(\tau_{k4}-\tau_{k3})\alpha\}}{\sum_{l=1}^{n}\left[\prod_{\substack{k=1\\k\neq l}}^{n}\{(\tau_{k2}-\tau_{k1})\alpha+\tau_{k1}\}\right]}\right]. \tag{5}$$

Interval expression OR transition

$$\lambda^{\alpha}=\left[\sum_{k=1}^{n}\{(\lambda_{k2}-\lambda_{k1})\alpha+\lambda_{k1}\}\ ,\ \sum_{k=1}^{n}\{\lambda_{k4}-(\lambda_{k4}-\lambda_{k3})\alpha\}\right], \tag{6}$$

$$\tau^{\alpha}=\left[\frac{\sum_{k=1}^{n}[\{(\lambda_{k2}-\lambda_{k1})\alpha+\lambda_{k1}\}\,.\,\{(\tau_{k2}-\tau_{k1})\alpha+\tau_{k1}\}]}{\sum_{k=1}^{n}\{\lambda_{k4}-(\lambda_{k4}-\lambda_{k3})\alpha\}},\frac{\sum_{k=1}^{n}[\{\lambda_{k4}-(\lambda_{k4}-\lambda_{k3})\alpha\}\,.\,\{\tau_{k4}-(\tau_{k4}-\tau_{k3})\alpha\}]}{\sum_{k=1}^{n}\{(\lambda_{k2}-\lambda_{k1})\alpha+\lambda_{k1}\}}\right]. \tag{7}$$

To assess the behavior of the washing unit quantitatively, the reliability parameters including, failure rate (λ), repair time (τ), mean time between failure (MTBF), expected number of failures (ENOF), reliability and availability are estimated utilizing the terms in Table 3 at spread levels ±15%, ±25% and ±40% for various degrees of membership.

Table 3: *Reliability parameters*

Reliability parameters	**Expressions**
MTTR	$\tau=\frac{1}{\mu}$
MTTF	$\frac{1}{\lambda}$
MTBF	$\tau+\frac{1}{\lambda}$
ENOF	$\frac{\lambda\,\mu\,t}{(\lambda+\mu)}+\frac{\lambda^2}{(\lambda+\mu)^2}[1-e^{-(\lambda+\mu)t}]$
Reliability	$e^{-\lambda t}$
Availability	$\frac{1}{(\lambda+\mu)}\left[\mu+\lambda e^{-(\lambda+\mu)t}\right]$

The variations of reliability parameters of washing unit at 15% spread for TrFN and TFN are depicted in Figures 5 (a-f).

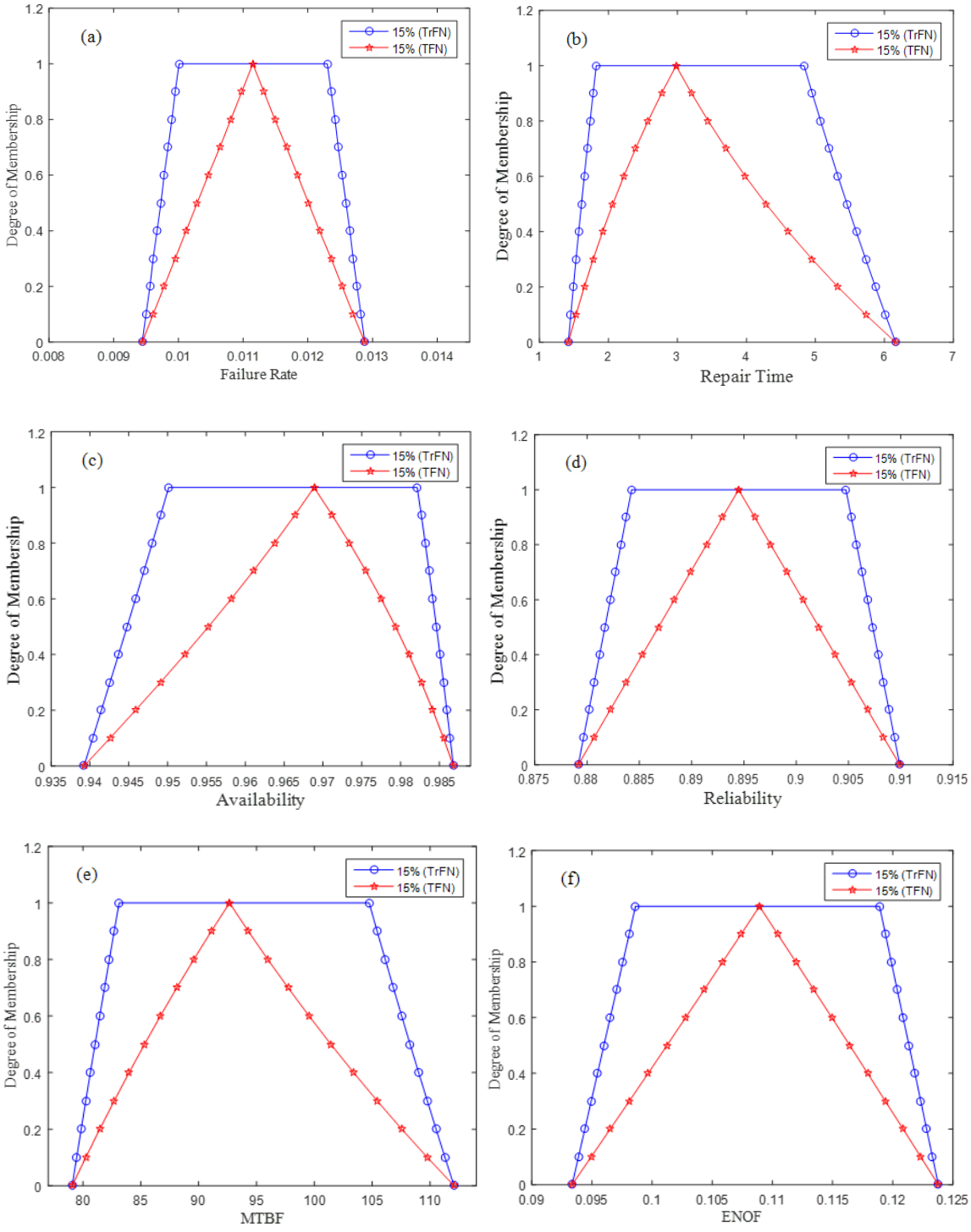


Figure 5: *Fuzzy reliability parameters using TrFN and TFN at ±15% spread*

4.1.4. Defuzzification

The estimated fuzzy values of different reliability parameters should be defuzzified in order to implement maintenance activities. Among various defuzzification techniques, such as the bisector, weighted average, center of area, centroid, middle of max, etc. the center of area technique has been chosen here for defuzzification of reliability parameters, as this technique is easy to implement. The defuzzified value ($\tilde{x}$) for fuzzy set $\tilde{A}$ is described as

$$\tilde{x} = \frac{\int_{x_1}^{x_2} x\mu_{\tilde{A}}(x)dx}{\int_{x_1}^{x_2} \mu_{\tilde{A}}(x)dx} \tag{8}$$

where, $\mu_{\tilde{A}}(x)$ is the membership function of $\tilde{A}$ described on [x_1, x_2]. The crisp and defuzzified results of reliability parameters for washing unit with three considered spreads are demonstrated in Table 4.

Table 4: *Crisp and defuzzified results*

Reliability parameters	Crisp results	Fuzzy Numbers	Defuzzified results		
			±15%	±25%	±40%
Failure Rate	0.011150	Trapezoidal	0.011159	0.011167	0.011194
		Triangular	0.011157	0.011167	0.011197
Repair Time	2.979753	Trapezoidal	3.577073	4.573931	8.204359
		Triangular	3.520117	4.631914	8.406379
Reliability	0.894518	Trapezoidal	0.894518	0.894562	0.894673
		Triangular	0.894530	0.894567	0.894687
Availability	0.968846	Trapezoidal	0.964473	0.959521	0.950455
		Triangular	0.964950	0.959287	0.949457
MTBF (×10²)	0.926632	Trapezoidal	0.948037	0.981285	1.087302
		Triangular	0.946023	0.983636	1.096407
ENOF	0.108919	Trapezoidal	0.108659	0.108340	0.107706
		Triangular	0.108673	0.108323	0.107643

Table 5: *Percent changes in defuzzified values for TrFN and TFN*

Change in spread	Fuzzy Numbers	Percent changes in defuzzified values					
		Failure rate	Repair time	Reliability	Availability	MTBF	ENOF
15% to 25%	TrFN	0.07 (↑)	27.87 (↑)	0.005 (↑)	0.51 (↓)	3.51 (↑)	0.29 (↓)
15% to 25%	TFN	0.09 (↑)	31.58 (↑)	0.004 (↑)	0.59 (↓)	3.98 (↑)	0.32 (↓)
25% to 40%	TrFN	0.24 (↑)	79.37 (↑)	0.012 (↑)	0.95 (↓)	10.80 (↑)	0.59 (↓)
25% to 40%	TFN	0.27 (↑)	81.49 (↑)	0.013 (↑)	1.03 (↓)	11.47 (↑)	0.63 (↓)

4.2. Behavioral Study

The fluctuations of fuzzy reliability parameters at 15% spread have been presented graphically in Figures 5 (a-f). The obtained membership curves are deformed trapeziums, with parabolic non parallel sides since fuzzy mathematics transforms left and right sides of membership curves of TrFN into curved ones [24]. The crisp and defuzzified results of reliability parameters for three distinct spread levels, obtained using TrFN and TFN both, in fuzzy λ-τ approach have been presented in Table 4. When the spread level increases, an increasing trend is observed in defuzzified values of parameters failure rate, repair time, reliability and MTBF, while, a decreasing trend is observed in values of system availability and ENOF in case of TrFN and TFN both. Also, the results obtained by the fuzzy λ-τ approach using TrFN follow the same pattern (increase or decrease) as those obtained by using TFN.

Table 5 presents the percent increase or decrease in values of different reliability parameters with spread level increase for both TrFN and TFN. It is observed that for spread increase from 15% to 25%, in case of TrFN, the failure rate of the system increases by 0.07%, while, for TFN, it increases by 0.09%. For a spread change from 25% to 40%, the failure rate increases by 0.24% for TrFN, while, for TFN, it increases by 0.27%. Similarly, for other reliability parameters, for the spread expansion from 15% to 25%, in case of TrFN, the repair time increases by 27.87% and MTBF increases by 3.51%, while, system availability decreases by 0.51% and ENOF decreases by 0.29%. On the other hand, in case of TFN, repair time increases by 31.58% and MTBF increases by 3.98%, while system availability and ENOF decrease by 0.59% and 0.32%, respectively. A very marginal and almost same increase is observed in system reliability for both TrFN and TFN. Similar inferences are drawn for spread expansion from 25% to 40% from Table 5. It is clear from these observations that when uncertainty level increases in terms of spread increase, the trend of reliability parameters (increase or decrease) remains almost the same for both TrFN and TFN. Therefore, the results obtained through the fuzzy λ-τ approach with TrFN are conservative in nature, which might be useful for system specialist/analysts for the future plan of implementation. Therefore, instead of crisp results, the maintenance plan should be based on defuzzified results.

5. Conclusion

The behavior of the washing unit in a paper plant has been analyzed in this study using the fuzzy λ-τ approach with TrFN. Using the fuzzy approach, some reliability parameters involving failure rate, repair time, MTBF, ENOF, reliability and availability are estimated and the trend of reliability parameters is evaluated for performance and behavior study of washing unit. Further, the obtained results by the fuzzy λ-τ approach with TrFN are compared with the results of the fuzzy λ-τ approach based on TFN. The implications of this fuzzy reliability approach to industry maintenance and operation planning are quite significant. Based on these investigations, suitable maintenance policy can be established that will assist maintenance managers in improving system performance by implementing appropriate preventive maintenance procedures. As a result, it will help in achieving a long time system availability and maximizing overall productivity of the paper plant.

Declaration of Conflicting Interests

The authors declare that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Kumar, D., Pandey, P. C. and Singh, J. (1991). Process design for a crystallization system in the urea fertilizer industry. *Microelectronics Reliability*, 31(5):855-859.

[2] Kumar, D., Singh, J. and Pandey, P. C. (1993). Operational behaviour and profit function for a bleaching and screening system in the paper industry. *Microelectronics Reliability*, 33(8):1101-1105.

[3] Arora, N. and Kumar, D. (2000). System analysis and maintenance management for the coal handling system in a paper plant. *International Journal of Management and Systems*, 16(2):137-56.

[4] Gupta, P., Lal, A. K., Sharma, R. K. and Singh, J. (2005). Numerical analysis of reliability and availability of the serial processes in butter-oil processing plant. *International Journal of Quality & Reliability Management*, 22(3):303-316.

[5] Sharma, R. K., Kumar, D. and Kumar, P. (2007). FM – a pragmatic tool to model, analyse and predict complex behaviour of industrial systems. *International Journal for Computer-Aided Engineering and Software*, 24(4):319-346.

[6] Sharma, R. K., Kumar, D. and Kumar, P. (2007). Modeling system behavior for risk and reliability analysis using KBARM. *Quality & Reliability Engineering International*, 23(8):973-998.

[7] Sachdeva, A., Kumar, D. and Kumar, P. (2008). Reliability analysis of pulping system using Petri nets. *International Journal of Quality & Reliability Management*, 25:860-877.

[8] Arabian-Hoseynabadi, H., Oraee, H. and Tavner, P. J. (2010). Failure modes and effects analysis (FMEA) for wind turbines. *International Journal of Electrical Power & Energy Systems*, 32(7):817-824.

[9] Garg, H. (2014). Performance and behavior analysis of repairable industrial systems using Vague Lambda–Tau methodology. *Applied Soft Computing*, *22*:323-338.

[10] Panchal, D. and Kumar, D. (2015). Stochastic behaviour analysis of power generating unit in thermal power plant using fuzzy methodology. *Opsearch*, 53(1):16-40.

[11] Knezevic, J. and Odoom, E. R. (2001). Reliability modeling of repairable systems using Petri nets and fuzzy Lambda-Tau methodology. *Reliability Engineering & System Safety*, 73(1):1-17.

[12] Sharma, R. K., Kumar, D. and Kumar, P. (2007). FM – a pragmatic tool to model, analyse and predict complex behaviour of industrial systems. *International Journal for Computer-Aided Engineering and Software*, 24(4):319-346.

[13] Sharma, S. P., Sukavanam, N., Kumar, N. and Kumar, A. (2008). Performance analysis of a complex robotic system using fault tree and fuzzy methodology. Proceedings of NSC08, Electrical Engineering Department, IIT Roorkee, India, 874-878.

[14] Sharma, S. P., Garg, H. and Kumar, A. (2009). Fuzzy system reliability of two grinding machine using fuzzy lambda-tau methodology. 4th International Conference on Quality, Reliability and Infocom Technology (ICQRIT), University of Delhi, Delhi, 68-69.

[15] Garg, H., Rani, M. and Sharma, S. P. (2012). Fuzzy RAM analysis of the screening unit in a paper industry by utilizing uncertain data. *Journal of Quality and Reliability Engineering.*

[16] Verma, M., Kumar, A. and Singh, Y. (2013). Vague reliability assessment of combustion system using Petri nets and vague lambda-tau methodology. *Engineering Computations*, 30(5):665-681.

[17] Panchal, D., Singh, A. K., Chatterjee, P., Zavadskas, E. K. and Keshavarz-Ghorabaee, M. (2019). A new fuzzy methodology-based structured framework for RAM and risk analysis. *Applied Soft Computing Journal*, 74:242–254.

[18] Princy, S. and Dhenakaran, S. S. (2016). Comparison of triangular and trapezoidal fuzzy membership function. *Journal of Computer Science Engineering*, 2(8):46-51.

[19] Sharma, S. and Sushma, (2022). Behaviour Analysis of Coal Handling System of a Thermal Power Plant using Fuzzy Methodology. *Palestine Journal of Mathematics*, 11:202-212.

[20] Sharma, S. and Sushma, (2023). Performance and Behavior Analysis of water circulation System of a Thermal Power Plant. *Reliability: Theory & Applications*, 18(1):316-328.

[21] Sharma, S. and Mamta, (2022). Behavior Analysis of Feeding Unit of a Paper industry in Fuzzy Environment. *International Journal of Reliability, Quality and Safety Engineering*, 30(1):2250027.

[22] Zimmermann, H. Fuzzy Set Theory and Its Applications, third ed., Kluwer Academic Publishers, New York, 2001.

[23] Sharma, R. K. (2006). Analysis, design and optimization of QRM aspects in production systems. PhD thesis, University of Roorkee, India.

[24] Bai, X. and Asgarpoor, S. (2004). Fuzzy based approaches to substation reliability evaluation. *Electric Power Systems Research*, 69(2-3):197-204.

THE CONTINUOUS BERNOULLI-GENERATED FAMILY OF DISTRIBUTIONS: THEORY AND APPLICATIONS

Ngozi O. Ubaka[1]
Friday Ewere[2]

[1]Department of Statistics, Federal University of Oyo-Ekiti, Ekiti State, Nigeria.
[2]Department of Statistics, University of Benin, Benin City, Edo State, Nigeria.

obiaderi.odafi@fuoye.edu.ng[1]
ewere.friday@uniben.edu[2]

Abstract

The continuous Bernoulli distribution is a one-parameter probability distribution which is useful in analysis on machine learning. A handful of studies has been done to generalize the continuous Bernoulli distribution. In this paper, we introduced a wider extension of the continuous Bernoulli distribution by considering its distribution function as a generator. We referred to the proposed family as the continuous Bernoulli-generated family of distributions. Basic statistical treatments of the proposed family such as the density and cumulative distribution functions, survival and hazard rate functions, quantile, moments, moment generating function, and Renyi entropy are derived. The method of maximum likelihood is employed to estimate the unknown parameters of the family and the asymptotic behaviour of the parameter estimates is investigated via Monte Carlo simulation study. The waiting time (in minutes) of 100 Bank customers and the tensile strength measured in GPa, of 69 carbon fibers data sets formed the basis for real-life data fittings. Results obtained from the fitting of the two data sets when compared with some existing non-nested models revealed that the fittings were in favor of the continuous-Bernoulli Weibull distribution over the rest competing distributions.

Keywords: Continuous Bernoulli Distribution; Moments; Quantile; Monte Carlo Simulation Study

1. INTRODUCTION

The cumulative distribution function (cdf) of the one-parameter continuous Bernoulli distribution has been defined by [13] as

$$F(x,\lambda)=\begin{cases}\dfrac{\lambda^{x}\left(1-\lambda\right)^{1-x}+\lambda-1}{2\lambda-1}, & \lambda\neq 1/2,\ 0<x<1,\\ x, & \lambda=1/2\end{cases} \tag{1}$$

with the probability density function (pdf) associated to (1) obtained as

$$f(x,\lambda)=\begin{cases} C_\lambda \lambda^x (1-\lambda)^{1-x}, & \lambda \neq 1/2,\ 0<x<1, \\ 1, & \lambda = 1/2 \end{cases} \tag{2}$$

where the normalizing constant C_λ is defined as

$$C_\lambda=\begin{cases} \dfrac{2\tanh^{-1}(1-2\lambda)}{1-2\lambda}, & \lambda \neq 1/2, \\ 2, & \lambda = 1/2 \end{cases} \tag{3}$$

and $2\tanh^{-1}(1-2\lambda)=\ln(1-\lambda)-\ln(\lambda)$, using the relation $\tanh^{-1}(x)=\frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$.

We denote a random variable *X* following the continuous Bernoulli distribution as $X \sim CB(\lambda)$. The continuous Bernoulli distribution has special application in machine learning. Particularly, in simulating the pixel intensities of natural images in deep learning and computer vision, mostly in the development of variational autoencoders. Similar to the one-parameter Topp-Leone and power distributions, the $CB(\lambda)$ distribution is also a one-parameter distribution with support on a unit-interval.

In the theory of statistical analysis of lifetime data, bounded distributions have found a wide variety of applications ranging from the field of engineering, actuarial sciences, economics, biological sciences, etc. Particularly, when the data are recorded in rates, percentages and proportions. Over the years, the beta and Kumaraswamy distributions are the topmost bounded distributions to be reckon with in regards to fitting [0,1]-valued data sets, until the advent of several methodologies in developing unit-interval distributions. Notable among these distributions are the log-Lindley distribution proposed by [10], unit-logistic distribution developed by [14], log-Xgamma distribution introduced by [2], Marshall-Olkin Topp-Leone distribution developed by [17], unit-Burr XII distribution studied by [11], Marshall-Olkin extended unit-Gompertz distribution studied by [15], transmuted Marshall-Olkin extended Topp-Leone Distribution introduced by [18], Kumaraswamy unit-Gompertz distribution proposed by [1], etc. It is noteworthy to mention that the power continuous Bernoulli distribution due to [3] and transmuted continuous Bernoulli distribution due to [4], apparently the only extensions of the classical continuous Bernoulli distribution belong to this list. The goal of this paper is to develop a novel family of distributions based on the continuous Bernoulli distribution, which is hoped to birth more tractable and flexible lifetime distributions in analyzing real data sets.

The rest of the paper is organized in the following sections. Section 2 is devoted to model formulation. Section 3 provides some sub-models from the proposed family of distributions. General mathematical treatments for the proposed family of distributions, the parameter estimation as well as the investigation of the asymptotic behaviour of the parameter estimates of the model via a Monte Carlo simulation are discussed in Section 4. Section 5 provides the applicability of the proposed family of distributions in real-life data fitting. Section 6 concludes the paper.

2. MODEL FORMULATION

Suppose a random variable *T* follows a known probability distribution with pdf $f(t)$, [20] adopted the beta-generated technique developed by [6] to introduce the Topp-Leone-generated family of

distributions with cdf defined by

$$F(x,\alpha,\xi) = 2\alpha\int_0^{G(x,\xi)}(1-t)\big(t(2-t)\big)^{\alpha-1}\,dt, \quad 0<t<1,\ \alpha>0, \\ = G(x,\xi)^{\alpha}\left(2-G(x,\xi)\right)^{\alpha}, \tag{4}$$

and the associated pdf obtained as

$$f(x,\alpha,\xi) = 2\alpha g(x,\xi)G(x,\xi)^{\alpha-1}\left(1-G(x,\xi)\right)\left(2-G(x,\xi)\right)^{\alpha-1}. \tag{5}$$

As an alternative to the technique in (4), [5] introduced the so-called type II Topp-Leone generated (TIITL-G) family of distributions based on the methodology of [19] who introduced an alternative gamma-generator reported in [22]. The cdf and pdf of TIITL-G family are, respectively, defined by

$$F(x,\alpha,\xi) = 1-2a\int_0^{1-G(x,\xi)}t^{a-1}(1-t)(2-t)^{\alpha-1}\,dt, \quad 0<t<1,\ \alpha>0, \\ = 1-\left(1-G^2(x,\xi)\right)^{\alpha}, \tag{6}$$

and

$$f(x,\alpha,\xi) = 2\alpha g(x,\xi)G(x,\xi)\left(1-G^2(x,\xi)\right)^{\alpha-1}. \tag{7}$$

Motivated by the simplicity of the technique in (6) and using the $CB(\lambda)$ distribution defined in (3) as the generator, we develop a novel class of distributions with the cdf defined by

$$F(t,\lambda,\xi) = \begin{cases} \dfrac{\lambda^{1-G(t,\xi)}\left(1-\lambda\right)^{G(t,\xi)}-\lambda}{1-2\lambda}, & \lambda\neq 1/2,\ \ 0<t<1, \\ G(t,\xi), & \lambda = 1/2 \end{cases} \tag{8}$$

The pdf corresponding to (8) is obtained as

$$f(t,\lambda,\xi) = \begin{cases} C_\lambda\, g(t,\xi)\lambda^{1-G(t,\xi)}\left(1-\lambda\right)^{G(t,\xi)}, & \lambda\neq 1/2,\ \ 0<t<1, \\ g(t,\xi), & \lambda = 1/2 \end{cases} \tag{9}$$

A random variable *T* having the cdf and pdf defined in (8) and (9), respectively, is said to follow the continuous Bernoulli-generated ($CB(\lambda,\xi)-G$) family of distributions.

The survival and hazard rate functions of $CB(\lambda,\xi)-G$ family of distributions are defined in (10) and (11), respectively, as

$$S(t,\lambda,\xi) = \begin{cases} \dfrac{\lambda^{1-G(t,\xi)}\left(1-\lambda\right)^{G(t,\xi)}+\lambda-1}{2\lambda-1}, & \lambda\neq 1/2,\ \ 0<t<1, \\ 1-G(t,\xi), & \lambda = 1/2 \end{cases} \tag{10}$$

and

$$h(t,\lambda,\xi) = \begin{cases} \dfrac{C_\lambda^{*}\, g(t,\xi)\lambda^{1-G(t,\xi)}\left(1-\lambda\right)^{G(t,\xi)}}{\lambda^{1-G(t,\xi)}\left(1-\lambda\right)^{G(t,\xi)}+\lambda-1}, & \lambda\neq 1/2,\ \ C_\lambda^{*} = (2\lambda-1)C_\lambda, \\ \dfrac{g(t,\xi)}{1-G(t,\xi)}, & \lambda = 1/2 \end{cases} \tag{11}$$

Furthermore, the quantile function of the $CB(\lambda,\xi)-G$ family of distributions is obtained as

$$Q_T(u) = G^{-1}\left[\frac{\ln\left[(1-2\lambda)u+\lambda\right]-\ln[\lambda]}{2\tanh^{-1}(1-2\lambda)}\right], \quad 0<u<1. \tag{12}$$

Whereas substituting $u=0.5$ in (12), the median of the $CB(\lambda,\xi)-G$ family of distributions is obtained as

$$Q_T(0.5) = G^{-1}\left[-\frac{\ln[2]+\ln[\lambda]}{2\tanh^{-1}(1-2\lambda)}\right]. \tag{13}$$

The utility of (12) is in generating random numbers from the $CB(\lambda,\xi)-G$ family of distributions, where *u* is generated from the uniform distribution satisfying $0<u<1$.

3. SUB-MODELS OF THE $CB(\lambda,\xi)-G$ FAMILY OF DISTRIBUTIONS

This section is concerned with the formulation of tractable models from the $CB(\lambda,\xi)-G$ family of distributions based on the Weibull, Topp-Leone, Kumaraswamy and Burr XII distributions as the baseline distribution in (8).

3.1 The continuous Bernoulli Weibull $CBW(\lambda,\alpha,\beta)$ distribution

Let *T* be a random variable following the Weibull distribution with cdf, $G(t,\alpha,\beta)=1-e^{-\beta t^{\alpha}}$ and pdf, $g(t,\alpha,\beta)=\alpha\beta t^{\alpha-1}e^{-\beta t^{\alpha}}$, $t>0,\ \alpha,\beta>0$. We defined the cdf and pdf of the $CBW(\lambda,\alpha,\beta)$ distribution, respectively, as follows

$$F(t,\lambda,\alpha,\beta)=\begin{cases}\dfrac{\lambda^{e^{-\beta t^{\alpha}}}(1-\lambda)^{1-e^{-\beta t^{\alpha}}}-\lambda}{1-2\lambda}, & \lambda\neq 1/2,\ \alpha,\beta>0,\ t>0.\\ 1-e^{-\beta t^{\alpha}}, & \lambda=1/2,\ \alpha,\beta>0.\end{cases} \tag{14}$$

and

$$f(t,\lambda,\alpha,\beta)=\begin{cases}C_{\lambda}^{w}t^{\alpha-1}e^{-\beta t^{\alpha}}\lambda^{e^{-\beta t^{\alpha}}}(1-\lambda)^{1-e^{-\beta t^{\alpha}}}, & \lambda\neq 1/2,\ \alpha,\beta>0,\ t>0,\ C_{\lambda}^{w}=\alpha\beta C_{\lambda}.\\ \alpha\beta t^{\alpha-1}e^{-\beta t^{\alpha}}, & \lambda=1/2,\ \alpha,\beta>0\end{cases} \tag{15}$$

3.2 The continuous Bernoulli Topp-Leone $CBTL(\lambda,\alpha)$ distribution

The one-parameter Topp-Leone distribution is defined by the density function

$$g(t,\alpha)=2\alpha(1-t)\left[t(2-t)\right]^{\alpha-1}, \quad \alpha\neq 1,\ \alpha>0,\ 0<t<1, \tag{16}$$

and the associated cdf is given by

$$G(t,\alpha)=\left[t(2-t)\right]^{\alpha}, \quad \alpha\neq 1,\ \alpha>0,\ 0<t<1, \tag{17}$$

By inserting the pdf and cdf in (16) and (17) into (8) and (9), we defined the cdf and pdf of the $CBTL(\lambda,\alpha)$ distribution, respectively, as

$$F(t,\lambda,\alpha)=\begin{cases}\dfrac{\lambda^{1-\left[t(2-t)\right]^{\alpha}}\left(1-\lambda\right)^{\left[t(2-t)\right]^{\alpha}}-\lambda}{1-2\lambda}, & \lambda\neq 1/2,\ \ \alpha>0,\ 0<t<1,\\ \left[t\left(2-t\right)\right]^{\alpha}, & \lambda=1/2,\ \ \alpha>0.\end{cases} \tag{18}$$

and

$$f(t,\lambda,\alpha)=\begin{cases}C_{\lambda}^{TL}\left(1-t\right)\left[t\left(2-t\right)\right]^{\alpha-1}\lambda^{1-\left[t(2-t)\right]^{\alpha}}\left(1-\lambda\right)^{\left[t(2-t)\right]^{\alpha}},\ \lambda\neq 1/2,\ \alpha>0,\ 0<t<1,\ C_{\lambda}^{TL}=2\alpha C_{\lambda}.\\ 2\alpha\left(1-t\right)\left[t\left(2-t\right)\right]^{\alpha-1}, \qquad \lambda=1/2,\ \ \alpha>0\end{cases} \tag{19}$$

3.3 The continuous Bernoulli Kumaraswamy $CBK(\lambda,\alpha,\beta)$ distribution

The Kumaraswamy distribution developed by [12] is a bounded distribution with 2 shape parameters having the cdf, $G(t)=1-\left(1-t^{\alpha}\right)^{\beta}$ and pdf, $g(t)=\alpha\beta t^{\alpha-1}\left(1-t^{\alpha}\right)^{\beta-1}$, $\alpha,\beta>0$.
By this information, the cdf and pdf of the $CBK(\lambda,\alpha,\beta)$ distribution is defined, respectively, as

$$F(t,\lambda,\alpha,\beta)=\begin{cases}\dfrac{\lambda^{\left(1-t^{\alpha}\right)^{\beta}}\left(1-\lambda\right)^{1-\left(1-t^{\alpha}\right)^{\beta}}-\lambda}{1-2\lambda}, & \lambda\neq 1/2,\ \ \alpha,\beta>0,\ 0<t<1,\\ 1-\left(1-t^{\alpha}\right)^{\beta}, & \lambda=1/2,\ \ \alpha,\beta>0.\end{cases} \tag{20}$$

and

$$f(t,\lambda,\alpha,\beta)=\begin{cases}C_{\lambda}^{k}t^{\alpha-1}\left(1-t^{\alpha}\right)^{\beta-1}\lambda^{\left(1-t^{\alpha}\right)^{\beta}}\left(1-\lambda\right)^{1-\left(1-t^{\alpha}\right)^{\beta}}, & \lambda\neq 1/2,\ \ \alpha,\beta>0,\ 0<t<1,\ C_{\lambda}^{k}=\alpha\beta C_{\lambda}.\\ \alpha\beta t^{\alpha-1}\left(1-t^{\alpha}\right)^{\beta-1}, & \lambda=1/2,\ \ \alpha,\beta>0\end{cases} \tag{21}$$

3.4 The continuous Bernoulli Burr XII $CBBXII(\lambda,\alpha,\beta)$ distribution

A random variable T is said to follow the two-parameter Burr XII distribution, if the density function of T is defined by

$$g(t,\alpha,\beta)=\alpha\beta t^{\alpha-1}\left(1+t^{\alpha}\right)^{-(\beta+1)}, \quad \alpha,\beta>0,\ \ t>0, \tag{22}$$

and the corresponding cdf is given by

$$G(t,\alpha,\beta)=1-\left(1+t^{\alpha}\right)^{-\beta}, \quad \alpha,\beta>0,\ \ t>0, \tag{23}$$

By inserting (22) and (23) into (8) and (9), we defined the cdf and pdf of the $CBBXII(\lambda,\alpha,\beta)$ distribution, respectively, as follows

$$F(t,\lambda,\alpha,\beta)=\begin{cases}\dfrac{\lambda^{\left(1+t^{\alpha}\right)^{-\beta}}\left(1-\lambda\right)^{1-\left(1+t^{\alpha}\right)^{-\beta}}-\lambda}{1-2\lambda}, & \lambda\neq 1/2,\ \ \alpha,\beta>0,\ t>0,\\ 1-\left(1+t^{\alpha}\right)^{-\beta}, & \lambda=1/2,\ \ \alpha,\beta>0.\end{cases} \tag{24}$$

and

$$f(t,\lambda,\alpha,\beta)=\begin{cases}C_{\lambda}^{k}t^{\alpha-1}\left(1+t^{\alpha}\right)^{-(\beta+1)}\lambda^{\left(1+t^{\alpha}\right)^{-\beta}}\left(1-\lambda\right)^{1-\left(1+t^{\alpha}\right)^{-\beta}}, & \lambda\neq 1/2,\ \ \alpha,\beta>0,\ t>0,\ C_{\lambda}^{B}=\alpha\beta C_{\lambda}.\\ \alpha\beta t^{\alpha-1}\left(1+t^{\alpha}\right)^{-(\beta+1)}, & \lambda=1/2,\ \ \alpha,\beta>0\end{cases} \tag{25}$$

4. MATHEMATICAL PROPERTIES OF THE $CB(\lambda,\xi)-G$ FAMILY OF DISTRIBUTIONS

In this section, the mathematical properties of the $CB(\lambda,\xi)-G$ family of distributions such as the r^{th} non-central moments, moment generating function (mgf) and Renyi entropy are discussed. The method of maximum likelihood estimation is employed to estimate the model parameters and the asymptotic behaviour of the parameter estimates are investigated through a Monte Carlo simulation study.

4.1 The r^{th} non-central moments

Let T be a random variable having the density function of the $CB(\lambda,\xi)-G$ family of distributions, then the r^{th} non-central moments of T is defined by

$$E\left[T^{r}\right]=\upsilon_{r}^{'}=\int_{-\infty}^{\infty}t^{r}\,f\left(t,\lambda,\xi\right)dt,\qquad r=1,2,3,4,\ldots \tag{26}$$

$$=C_{\lambda}\int_{-\infty}^{\infty}t^{r}\,g(t,\xi)\,\lambda^{1-G(t,\xi)}\left(1-\lambda\right)^{G(t,\xi)}\,dt.$$

Evaluating (26) yields the following results

$$\begin{aligned}E\left[T^{r}\right]&=C_{\lambda}\int_{-\infty}^{\infty}t^{r}\,g(t,\xi)\exp\left(\left(G(t,\xi)\right)\ln\left(\lambda\right)+\left(1-G(t,\xi)\right)\ln\left(1-\lambda\right)\right)dt,\\&=\lambda\,C_{\lambda}\int_{-\infty}^{\infty}t^{r}\,g(t,\xi)\exp\left(G(t,\xi)\left[\ln\left(1-\lambda\right)-\ln\left(\lambda\right)\right]\right)dt,\\&=\lambda\,C_{\lambda}\int_{-\infty}^{\infty}t^{r}\,g(t,\xi)\exp\left(G(t,\xi)\left[2\tanh^{-1}\left(1-2\lambda\right)\right]\right)dt.\end{aligned} \tag{27}$$

Applying the Maclaurin's series expansion of the exponential function,

$$e^{G(t,\xi)\left[2\tanh^{-1}(1-2\lambda)\right]}=\sum_{n=0}^{\infty}\frac{\left[2\tanh^{-1}\left(1-2\lambda\right)\right]^{n}}{n!}\left[G(t,\xi)\right]^{n},$$

so that (27) now becomes,

$$E\left[T^{r}\right]=\lambda\,C_{\lambda}\sum_{n=0}^{\infty}\frac{\left[2\tanh^{-1}\left(1-2\lambda\right)\right]^{n}}{n!}\int_{-\infty}^{\infty}t^{r}\,g(t,\xi)\left[G(t,\xi)\right]^{n}\,dt,$$

$$= \lambda C_\lambda \sum_{n=0}^{\infty} \frac{\left[2\tanh^{-1}(1-2\lambda)\right]^n}{n!(n+1)} \int_{-\infty}^{\infty} t^r h_{n+1}(t,\xi)dt,$$

$$= \lambda C_\lambda \sum_{n=0}^{\infty} \frac{\left[2\tanh^{-1}(1-2\lambda)\right]^n}{n!(n+1)} E\left[Y_{n+1}^r\right]. \quad (28)$$

Where $h_{n+1}(t,\xi) = (n+1)\, g(t,\xi)\, [G(t,\xi)]^n$ and $E\left[Y_{n+1}^r\right]$ are, respectively, the density function and r^{th} non-central moments of the exp-G family of distributions with power parameter $(n+1)$.

Thus, we can express the r^{th} non-central moments of the $CB(\lambda,\xi)-G$ family of distributions as a linear combination of the r^{th} non-central moments of the exp-G family of distributions with power parameter $(n+1)$.

For the purpose of numerical computation, we consider the two-parameter Weibull distribution as the baseline distribution. Hence, we compute the first four raw moments, variance, measures of skewness and kurtosis of the continuous Bernoulli Weibull $CBW(\lambda,\alpha,\beta)$ distribution in Table 1.

Table 1: *The Moments of the* $CBW(\lambda,\alpha,\beta)$ *distribution for selected values of the Parameters*

λ	β	α	$\upsilon_1^{'}$	$\upsilon_2^{'}$	$\upsilon_3^{'}$	$\upsilon_4^{'}$	σ^2	S	K
0.4	0.5	3	1.1721	1.5417	2.2068	3.3774	0.1679	0.0905	2.7315
		5	1.0822	1.2283	1.4485	1.7636	0.0571	-0.3259	2.9750
		7	1.0524	1.1367	1.2549	1.4119	0.0292	-0.5465	3.4994
	3.0	3	0.6450	0.4669	0.3678	0.3098	0.0509	0.0889	2.7397
		5	0.7563	0.5999	0.4944	0.4206	0.0279	-0.3265	2.8830
		7	0.8147	0.6812	0.5822	0.5072	0.0175	-0.5310	3.6310
0.8	0.5	3	0.9696	1.0912	1.3758	1.9007	0.1511	0.4223	2.9993
		5	0.9616	0.9824	1.0551	1.1827	0.0577	-0.0428	2.9123
		7	0.9659	0.9640	0.9896	1.0415	0.0310	-0.2721	3.2045
	3.0	3	0.5336	0.3305	0.2293	0.1743	0.0458	0.4181	2.9976
		5	0.6720	0.4798	0.3601	0.2821	0.0282	-0.0523	3.0015
		7	0.7478	0.5778	0.4592	0.3741	0.0186	-0.2719	3.0701

Information from Table 1 shows that the CBW distribution exhibits a left-skewed, right-skewed, platykurtic and leptokurtic properties which are essential in modeling heavy-tailed distributions.

4.2 The moment generating function

The moment generating function (mgf) of a random variable T with density function $f(t)$ is defined by

$$M_T(q) = E\left[e^{qt}\right] = \int_{-\infty}^{\infty} e^{qt} f(t)\, dt, \quad (29)$$

Using similar approach in (29), we defined the mgf of the $CB(\lambda,\xi)-G$ family of distributions as

$$M_T(q) = \lambda C_\lambda \sum_{n=0}^{\infty}\sum_{p=0}^{\infty} \frac{\left[2\tanh^{-1}(1-2\lambda)\right]^n q^p}{n!(n+1)\,p!} E\left[Y_{n+1}^p\right]. \tag{30}$$

Since, $e^{qt} = \sum_{p=0}^{\infty} \frac{(qt)^p}{p!}$.

4.3 The Renyi entropy

An entropy of a random variable say T, measures the degree of randomness associated with the random variable T. The Renyi entropy of T is defined by [18] as

$$\tau_R(\gamma) = \frac{1}{1-\gamma}\log\int_{-\infty}^{\infty} f^\gamma(t)\,dt, \qquad \gamma > 0, \quad \gamma \neq 1. \tag{31}$$

By substituting (9) into (31), we defined the Renyi entropy of a random variable T following the $CB(\lambda,\xi)-G$ family of distributions as follows

$$\begin{aligned}
\tau_R(\gamma) &= \frac{1}{1-\gamma}\log\left[(C_\lambda)^\gamma \int_{-\infty}^{\infty} g^\gamma(t,\xi)\,\lambda^{\gamma(1-G(t,\xi))}(1-\lambda)^{\gamma G(t,\xi)}\,dt\right], \\
&= \frac{1}{1-\gamma}\log\left[(C_\lambda)^\gamma\,\lambda^\gamma \int_{-\infty}^{\infty} g^\gamma(t,\xi)\exp\left(\gamma G(t,\xi)\left[\ln(1-\lambda)-\ln(\lambda)\right]\right)dt\right], \\
&= \frac{1}{1-\gamma}\log\left[(C_\lambda)^\gamma\,\lambda^\gamma \int_{-\infty}^{\infty} g^\gamma(t,\xi)\exp\left(\gamma G(t,\xi)\left[2\tanh^{-1}(1-2\lambda)\right]\right)dt\right].
\end{aligned} \tag{32}$$

Again, applying the Maclaurin's series expansion of the exponential function,

$$e^{\gamma G(t,\xi)\left[2\tanh^{-1}(1-2\lambda)\right]} = \sum_{n=0}^{\infty} \frac{\gamma^n\left[2\tanh^{-1}(1-2\lambda)\right]^n}{n!}\left[G(t,\xi)\right]^n,$$

so that (32) now becomes,

$$\tau_R(\gamma) = \frac{1}{1-\gamma}\log\left[(C_\lambda)^\gamma\,\lambda^\gamma \sum_{n=0}^{\infty} \frac{\gamma^n\left[2\tanh^{-1}(1-2\lambda)\right]^n}{n!}\int_{-\infty}^{\infty} g^\gamma(t,\xi)\left[G(t,\xi)\right]^n dt\right]. \tag{33}$$

Two major properties of the Renyi entropy of a random variable T were identify by [9]. These include

(i) The Renyi entropy of T can assume a negative value;

(ii) For any $\gamma_1 < \gamma_2, R_{\gamma_2} \leq R_{\gamma_1}$ and equality holds if and only if T is a uniform random variable.

Again, we compute the Renyi entropy of the $CBW(\lambda,\alpha,\beta)$ distribution for selected values of the parameters as shown in Table 2.

Table 2: *Numerical computation of the Renyi entropy of the* $CBW(\lambda,\alpha,\beta)$ *distribution (* $\lambda = 0.8$ *)*

i	γ_i	$\alpha=0.9,\ \beta=0.5$	$\alpha=0.9,\ \beta=3.0$	$\alpha=1.5,\ \beta=3.0$	$\alpha=1.5,\ \beta=0.5$
1	0.1	3.5600	1.5691	0.8868	2.0813
2	0.3	2.4724	0.4815	0.3213	1.5158
3	0.5	1.9849	-0.0060	0.0923	1.2869
4	0.7	1.6766	-0.3142	-0.0433	1.1513
5	0.9	1.4573	-0.5336	-0.1356	1.0589
6	2	0.8522	-1.1387	-0.3746	0.8199
7	4	0.4451	-1.5458	-0.5180	0.6765
8	6	0.2343	-1.7565	-0.5793	0.6152
9	8	0.0647	-1.9262	-0.6147	0.5799

The result in Table 2 validates the aforementioned properties of the Renyi entropy as suggested by [9].

4.4 Parameter estimation

4.4.1 Maximum likelihood estimation

The maximum likelihood estimation method is employed to estimate the parameters of the $CB(\lambda,\xi)-G$ family of distributions. Suppose $(t_1,t_2,....t_n)$ are random samples of size n from the $CB(\lambda,\xi)-G$ family of distributions, then the likelihood function is obtained as

$$L(t,\varphi) = \prod_{i=1}^{n}\left[g(t,\xi)\,\lambda^{1-G(t,\xi)}\,(1-\lambda)^{G(t,\xi)}\right], \quad \varphi=(\lambda,\xi)^T. \tag{34}$$

By taking the natural logarithm of both sides of (34), the log-likelihood function is obtained as

$$\ell(t,\varphi) = \sum_{i=1}^{n}\ln\left[g(t_i,\xi)\right]+\ln\left[\lambda\right]\sum_{i=1}^{n}\left(1-G(t_i,\xi)\right)+\ln\left[1-\lambda\right]\sum_{i=1}^{n}G(t_i,\xi). \tag{35}$$

The maximum likelihood estimate, say $\hat{\varphi}=\left(\hat{\lambda},\hat{\xi}\right)^T$ is obtained by differentiating the log-likelihood function in (35) with respect to the parameters and equating the corresponding function to zero as shown below

$$\frac{\partial\ell(t,\varphi)}{\partial\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n}\left(1-G\left(t_i,\xi\right)\right)-\frac{1}{1-\lambda}\sum_{i=1}^{n}G\left(t_i,\xi\right)=0$$

Further simplification yields,

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\left(1-G\left(t_i,\xi\right)\right),$$

$$\frac{\partial\ell(t,\varphi)}{\partial\xi} = \sum_{i=1}^{n}\frac{g^{'}(t_i,\xi)}{g(t_i,\xi)}+\ln\left(1-\lambda\right)\sum_{i=1}^{n}g\left(t_i,\xi\right)-\ln\left(\lambda\right)\sum_{i=1}^{n}g\left(t_i,\xi\right)=0.$$

Where $g^{'}(t_i,\xi)=\frac{\partial g(t_i,\xi)}{\partial\xi_i}$ and $\partial\xi_j$ is the j^{th} element of the vector of parameter ξ.

It is clear from these expressions that the parameters $\hat{\lambda}$ can be solved analytically, whereas the parameter(s) $\hat{\xi}_j$ may require the use of software program such as ***R*** program for estimation.

4.4.2 Simulation study

In this subsection, we investigate the asymptotic behaviour of the parameter estimates of the $CBW(\lambda,\alpha,\beta)$ distribution. Random samples of size $n=(15,25,50,75,100)$ are generated from the $CBW(\lambda,\alpha,\beta)$ distribution at randomly fixed values of the parameters. A Monte Carlo simulation is repeated 1000 times and the following quantities are computed:

i) $\text{bias}=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\varphi}_i-\bar{\varphi}\right)$,

ii) root mean square error (RMSE) $= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\varphi}_i - \bar{\varphi}\right)^2}$.

iii) Coverage Probability of the 95% confidence interval of the estimates $\hat{\varphi}_i$ given by

$$CP\left(\hat{\varphi}\right) = \frac{1}{N}\sum_{i=1}^{N} I\left(\hat{\varphi}_i - Z_{\delta/2}\sqrt{\text{var}\left(\hat{\varphi}\right)} < \varphi_0 < \hat{\varphi}_i + Z_{\delta/2}\sqrt{\text{var}\left(\hat{\varphi}\right)}\right).$$

Where $I(.)$ is an indicator function and $\left(\hat{\varphi}\right)$ is the standard error of the estimate φ_i.

Table 3: *Simulation results for bias, RMSE and CP of parameter estimates of* $CBW\left(\lambda,\alpha,\beta\right)$ *distribution*

Parameters	n	Bias			RMSE			CP		
		α	β	λ	α	β	λ	α	β	λ
$\alpha=0.3$ $\beta=0.6$ $\lambda=0.8$	15	0.0042	0.3614	-0.2437	0.0752	0.5992	0.3598	0.986	0.988	0.908
	25	-0.0215	0.3395	-0.2522	0.0623	0.5566	0.3558	0.958	0.972	0.888
	50	-0.0578	0.3019	-0.2781	0.0527	0.4956	0.3441	0.948	0.970	0.864
	75	-0.0704	0.2741	-0.2996	0.0477	0.4877	0.3253	0.938	0.940	0.878
	100	-0.0961	0.2210	-0.3323	0.0421	0.4231	0.2926	0.958	0.964	0.910
$\alpha=0.5$ $\beta=0.3$ $\lambda=0.6$	15	0.0324	0.1972	-0.1020	0.1422	0.2625	0.2880	0.978	0.958	0.918
	25	0.0093	0.1887	-0.1074	0.1057	0.2472	0.2808	0.988	0.986	0.890
	50	-0.0154	0.1628	-0.1158	0.0832	0.2404	0.2749	0.964	0.978	0.876
	75	-0.0184	0.1017	-0.1356	0.0828	0.2361	0.2741	0.942	0.966	0.872
	100	-0.0209	0.0772	-0.1648	0.0724	0.2227	0.2578	0.944	0.952	0.878
$\alpha=0.9$ $\beta=3.0$ $\lambda=0.4$	15	0.1085	0.3271	0.0496	0.3043	1.2171	0.2746	0.956	0.998	0.914
	25	0.0599	0.1131	0.0401	0.2177	0.9368	0.2645	0.964	0.990	0.904
	50	0.0174	0.1082	0.0192	0.1920	0.8197	0.2632	0.926	0.956	0.852
	75	0.0026	0.0824	0.0186	0.1619	0.7159	0.2586	0.914	0.942	0.824
	100	-0.0043	0.0531	0.0079	0.1615	0.6676	0.2499	0.904	0.940	0.814
$\alpha=0.9$ $\beta=0.6$ $\lambda=0.4$	15	0.0932	0.0618	0.0485	0.2758	0.3681	0.2862	0.978	0.940	0.910
	25	0.0439	0.0527	0.0468	0.2190	0.3658	0.2812	0.966	0.938	0.858
	50	0.0266	0.0523	0.0293	0.1871	0.3382	0.2702	0.938	0.928	0.818
	75	0.0082	0.0470	0.0256	0.1551	0.3023	0.2532	0.950	0.938	0.844
	100	0.0073	0.0452	0.0180	0.1524	0.3007	0.2521	0.922	0.912	0.828

From Table 3, we observe that the bias and root mean square errors of the parameter estimates decrease as the sample size n increases. Moreover, the coverage probability of the parameter estimates approaches the nominal level of 95% confidence interval.

5. REAL-LIFE DATA FITTINGS

The applicability of the proposed family of distributions is investigated in this section. To achieve this, two data sets including the waiting time (in minutes) of 100 Bank customers and the tensile strength measured in GPa, of 69 carbon fibers data sets are employed for data fittings. Some well-known non-nested models such as the Kumaraswamy Weibull ($KW\left(\lambda,\alpha,\beta\right)$), Kumaraswamy inverse Weibull ($KIW\left(\lambda,\alpha,\beta\right)$), Topp-Leone inverse Weibull ($TLIW\left(\lambda,\alpha,\beta\right)$), transmuted Weibull ($TW\left(\lambda,\alpha,\beta\right)$) and the two-parameter Weibull distributions are employed alongside with the proposed continuous-Bernoulli Weibull ($CBW\left(\lambda,\alpha,\beta\right)$) distribution to fit the two data sets. The data sets for the analysis are given below.

Data set 1: The first data set represents the waiting time (in minutes) of 100 Bank customers reported in [16]. The data set was first used by [8] to illustrate the flexibility of the Lindley distribution over the exponential distribution in data fittings. The data are given as follows: 0.8, 0.8, 1.3, 1.5, 1.8, 1.9 ,1.9, 2.1, 2.6, 2.7,2.9, 3.1, 3.2, 3.3 ,3.5, 3.6, 4.0, 4.1, 4.2, 4.2,4.3, 4.3, 4.4, 4.4, 4.6, 4.7, 4.7, 4.8, 4.9, 4.9,5.0, 5.3, 5.5, 5.7, 5.7, 6.1, 6.2, 6.2, 6.2, 6.3,6.7, 6.9, 7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8.0,8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6,9.7, 9.8, 10.7, 10.9, 11.0, 11.0, 11.1, 11.2, 11.2, 11.5,11.9, 12.4, 12.5, 12.9, 13.0, 13.1, 13.3, 13.6, 13.7, 13.9,14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0,19.9, 20.6, 21.3, 21.4, 21.9, 23.0, 27.0, 31.6, 33.1, 38.5.

Data set 2: The second data set comprises of the tensile strength measured in GPa, of 69 carbon fibers tested under tension at gauge length of 20mm reported in [21]. This data set was also employed by [7] to demonstrate the applicability of the power Lindley distribution. The data are represented as follows: 1.312, 1.314, 1.479, 1.552, 1.700, 1.803, 1.861, 1.865, 1.944, 1.958, 1.966, 1.997, 2.006, 2.021, 2.027, 2.055, 2.063, 2.098, 2.14, 2.179, 2.224, 2.240, 2.253, 2.270, 2.272, 2.274, 2.301, 2.301, 2.359, 2.382, 2.382, 2.426, 2.434, 2.435, 2.478, 2.490, 2.511, 2.514, 2.535, 2.554, 2.566, 2.57, 2.586, 2.629, 2.633, 2.642, 2.648, 2.684, 2.697, 2.726, 2.770, 2.773, 2.800, 2.809, 2.818, 2.821, 2.848, 2.88, 2.954, 3.012, 3.067, 3.084, 3.090, 3.096, 3.128, 3.233, 3.433, 3.585, 3.585.

Some popularly used model selection criteria such as the maximized log-likelihood (*LL*), Akaike Information Criteria (AIC), and some goodness of fit test statistics such as the Komolgorov-Smirnov (K-S), Crammer von Mises (W^*) and Anderson Darling (A^*) test statistics with their corresponding *p-value* are considered to access the appropriate model for analyzing the two data sets. Tables 4 and 5 present the summary statistics for the fit of the distributions for the two data sets, respectively.

Table 4: *Summary statistics for the waiting time data set*

Models	*Estimates*	*LL*	*AIC*	*K-S* *(p-value)*	W^* *(p-value)*	A^* *(p-value)*
CBW	$\alpha=1.7229$	-317.3098	640.6196	0.0423	0.0248	0.1682
	$\beta=0.0071$			(0.994)	(0.9904)	0.9968)
	$\lambda=0.9356$					
KW	$\alpha=1.3727$	-317.6755	641.3510	0.0508	0.0414	0.2578
	$\beta=0.2015$			(0.9587)	(0.9263)	(0.9660)
	$\lambda=1.3379$					
KIW	$\alpha=2.6384$	-332.9531	671.9062	0.1099	0.4051	2.6255
	$\beta=1.1424$			(0.1785)	(0.0698)	(0.0427)
	$\lambda=-1.5224$					
TLIW	$\alpha=0.5235$	-327.1056	641.2112	0.0891	0.2449	1.6727
	$\beta=12.5524$			(0.4044)	(0.1951)	(0.1402)
	$\lambda=0.9569$					
TW	$\alpha=1.5692$	-317.8896	641.7791	0.0481	0.0384	0.2599
	$\beta=0.0157$			(0.9746)	(0.9420)	(0.9648)
	$\lambda=0.6181$					
Weibull	$\alpha=1.4584$	-318.7307	641.4614	0.0577	0.0609	0.4051
	$\beta=0.0305$			(0.8929)	(0.8095)	(0.8433)

Table 5: *Summary statistics for tensile strength data set*

Models	*Estimates*	*LL*	*AIC*	*K-S* *(p-value)*	W^* *(p-value)*	A^* *(p-value)*
CBW	$\alpha = 2.7806$	-49.0740	104.1481	0.0400	0.0142	0.1210
	$\beta = 0.1778$			(0.9999)	(0.9998)	(0.9998)
	$\lambda = 0.0026$					
KW	$\alpha = 3.9464$	-49.9210	105.8421	0.0675	0.0581	0.3901
	$\beta = 0.1690$			(0.9112)	(0.8276)	(0.8580)
	$\lambda = -0.1312$					
KIW	$\alpha = 4.2588$	-56.2704	118.5408	0.1061	0.1995	1.3439
	$\beta = 2.8719$			(0.4193)	(0.2688)	(0.2185)
	$\lambda = -3.7556$					
TLIW	$\alpha = 0.5468$	-58.0304	122.0608	0.1176	0.2617	1.7344
	$\beta = 34.8898$			(0.2960)	(0.1741)	(0.1294)
	$\lambda = 3.4115$					
TW	$\alpha = 5.9303$	-49.1325	104.2650	0.0433	0.0191	0.1714
	$\beta = 0.0021$			(0.9995)	(0.9979)	(0.9963)
	$\lambda = 0.6363$					
Weibull	$\alpha = 5.5045$	-49.5961	104.1923	0.0560	0.0343	0.2739
	$\beta = 0.0046$			(0.9819)	(0.9611)	(0.9563)

From Tables 4 and 5, based on the conditions to measure superiority of models, the continuous-Bernoulli $CBW(\lambda, \alpha, \beta)$ distribution having the maximized log-likelihood value, least value in terms of the *AIC*, *K-S*, W^* and A^* test statistics with the corresponding highest *p-value*, outperforms the competitor distributions in analyzing the two data sets, and thus becomes the most appropriate model in fitting the data sets.

6. CONCLUSION

In this paper, we have developed a new class of probability distributions based on the continuous Bernoulli distribution. The proposed family is called the continuous Bernoulli-generated family of distributions. Mathematical derivation of some basic properties of the proposed family such as the density and cumulative distribution functions, survival and hazard rate functions, quantile, moments, moment generating function, and Renyi entropy were obtained. The method of maximum likelihood was employed to estimate the unknown parameters of the family and the asymptotic behaviour of the parameter estimates was investigated via Monte Carlo simulation study. Two real-life data sets including the waiting time (in minutes) of 100 Bank customers and the tensile strength measured in GPa, of 69 carbon fibers data sets were employed to illustrate the applicability of the proposed family. Existing non-nested models such as the Kumaraswamy Weibull, Kumaraswamy inverse Weibull, Topp-Leone inverse Weibull, transmuted Weibull and the two-parameter Weibull distributions were employed alongside the proposed continuous-Bernoulli Weibull distribution to

fit the two data sets. Results obtained from the fitting of the two data sets when compared using some model selection criteria and goodness of fit test statistics, revealed that the fittings were in favor of the continuous-Bernoulli Weibull distribution over the rest competing distributions.

References

[1] Akata, I. U., Opone, F. C. and Osagiede, F.E.U. (2023). The Kumaraswamy Unit-Gompertz Distribution and its Application to Lifetime Dataset. *Earthline Journal of Mathematical Sciences,* 11(1): 1-22.

[2] Altun, E. (2018). The log-xgamma distribution with inference and application. *Journal de la Société de Statistique de Paris,* 159(3): 40-55.

[3] Chesneau, C. and Opone, F. C. (2022). The power continuous Bernoulli distribution: Theory and Applications. *Reliability: Theory & Application,* 17(4): 232-248.

[4] Chesneau, C., Opone, F., and Ubaka, N. (2022). Theory and applications of the transmuted continuous Bernoulli distribution. *Earthline Journal of Mathematical Sciences.* 10(2): 385-407.

[5] Elgarhy, M., Nasir, M. A., Farrukh Jamal, F. and Ozel, G. (2018). The type II Topp-Leone generated family of distributions: Properties and applications, *Journal of Statistics and Management Systems*, 21: 1529-1551.

[6] Eugene, N., Lee, C. and Famoye, F. (2002). The beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods.* 31: 497-512.

[7] Ghitany, M., Al-Mutairi, D., Balakrishnan, N. and Al-Enezi, I. (2013). Power Lindley distribution and associated inference. *Computational Statistics and Data Analysis.* 64: 20-33.

[8] Ghitany, M., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its applications. *Mathematics and Computers in Simulation*. 78: 493-506.

[9] Golshani, L. and Pasha, E. (2010). Renyi entropy rate for Gaussian processes. *Information Sciences*, 180: 1486-1491.

[10] Gómez-Déniz, E., Sordo, M. A. and Calderín-Ojeda, E. (2014). The Log-Lindley distribution as an alternative to the beta regression model with applications in insurance. *Insurance: Mathematics and Economics,* 54(1): 49-57.

[11] Korkmaz, M. and Chesneau, C. (2021). On the unit Burr-XII distribution with the quantile regression modeling and applications. *Computational and Applied Mathematics,* 40(1): 1-26.

[12] Kumaraswamy, P. (1980). A Generalized Probability Density Function for Doubly Bounded Random Process. *Journal of Hydrology*. 46: 79-88.

[13] Loaiza-Ganem, G. and Cunningham, J.P. (2019). The continuous bernoulli: fixing a pervasive error in variational autoencoders. *In Advances in Neural Information Processing Systems*, 13266-13276.

[14] Menezes, A. F. B., Mazucheli, J. and Dey, S. (2018). The unit-logistic distribution: different methods of estimation. *Pesquisa Operacional*, 38(3): 555-578

[15] Opone, F. C., Akata, I. U. and Altun, E. (2022). The Marshall-Olkin Extended Unit-Gompertz Distribution: Its Properties, Regression Model and Applications. *Statistica*, 82(2): 97-118.

[16] Opone, F. C. and Ekhosuehi, N. (2018). Methods of Estimating the Parameters of the Quasi Lindley Distribution. *Statistica*, 78(2): 183-193.

[17] Opone, F. C. and Iwerumor, B. N. (2021). A New Marshall-Olkin Extended Family of Distributions with Bounded Support. *Gazi University Journal of Science*, *34(3): 899-914.*

[18] Opone, F. C. and Osemwenkhae, J. E. (2022). The transmuted Marshall-Olkin extended Topp-Leone Distribution. *Earthline Journal of Mathematical Sciences, 9(2): 179-199.*

[18] Rényi, A. (1961). On measure of entropy and information. Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability 1, University of California Press, Berkeley, Vol. 4, 1 January 1961, 547-561

[19] Ristic, M. M., & Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation, 82 (6): 1191-1206*.

[20] Sangsanit Y. and Bodhisuwan, W. (2016). The Topp-Leone generator of distributions: properties and inferences. *Songklanakarin Journal of Science and Technology,* 38: 537-548.

[21] Tuoyo, D. O, Opone, F. C. and N. Ekhosuehi, N. (2021). The Topp-Leone Weibull distribution: its properties and application, *Earthline Journal of Mathematical Sciences,* 7(2): 381-401.

[22] Zografos, K. and Balakrishnan, N. (2009). On Families of Beta-G and Generalized Gamma-generated Distribution and Associate Inference. *Statistical Methodology,* 6: 344-362.

A NEW EXPONENTIAL TYPE RATIO ESTIMATOR FOR THE POPULATION MEAN IN SYSTEMATIC SAMPLING

Ayed AL e'damat[1] and Khalid Ul Islam Rather[2]*

•

[1]*Department of Mathematics, AL Hussain Bin Talal University, ma'an, Jordan*
[2]*Division of Statistics and Computer Science, SKUAST-Jammu, India*[2]
ayed.h.aledamat@ahu.edu.jo
khalidstat34@gmail.com

Abstract

Utilizing auxiliary information effectively in sample surveys can enhance the accuracy of estimations by capitalizing on the relationship between the main variable under study and the auxiliary variable. Estimators such as ratio, product, exponential, and regression estimators are frequently employed either during the estimation process, the design phase, or both. In everyday situations, it is common to incorporate information from one or two auxiliary variables to improve the precision of estimators. Auxiliary information has been in practice in sampling theory since the advent of modern sample surveys. Information on auxiliary variable having high correlation with the variable under study is quite useful in improving the sampling design. Cochran (1940) used the highly positively correlated study and auxiliary variable to propound the ratio estimator. Product estimator requires a high negative correlation between study and auxiliary variable. By reviewing the literature, it is concluded that applying the auxiliary information enhances the efficiencies of the estimators for estimating any parameter under consideration. So it is well established fact that the use of auxiliary variable technique improves the estimation process for target population. It is also noticed that ratio method of estimation is relatively simple and one of the commonly used methods of estimation. Due to limitations in terms of time and cost, sample surveys are often preferred over census surveys for collecting primary data. In these sample surveys, the ratio, product, and regression estimators are frequently employed to estimate the mean or other parameters of interest for the variable under study. To assess their efficiency, these estimators are compared based on their approximate mean squared errors. In this paper we proposed an exponential ratio type estimator for the estimation of finite population mean under systematic sampling. The mean square error of the proposed estimator is computed up to the first order of approximation and we find proposed estimator is efficient as compared to other existing estimators. Furthermore this theoretical result is supported by numerical examples as well.

Keywords: Systematic sampling, exponential ratio type estimator, mean square error, efficiency.

1. Introduction

In the literature of survey sampling, it is well known that the efficiencies of the estimators of the population parameters of the study variable can be increased by the use of auxiliary information related to auxiliary variable x, which is highly correlated with the study variable y. Auxiliary information may be efficiently utilized either at planning stage or at design stage to arrive at an improved estimator compared to those estimators, not utilizing auxiliary information. A simple technique of utilizing the known information of the population parameters of the auxiliary variables is through ratio, product, and regression method of estimations using different probability sampling designs such as simple random sampling, stratified random sampling, cluster sampling, systematic sampling, and double sampling. In the present paper we will use knowledge of the auxiliary variables under the framework of systematic sampling. Due to its simplicity,

systematic sampling is the most commonly used probability design in survey of finite populations; see W. G. Madow and L. H. Madow [23]. Apart from its simplicity, systematic sampling provides estimators which are more efficient than simple random sampling or stratified random Sampling for certain types of population; see Cochran [24], Gautschi [25], and Hajeck [9]. later on the problem of estimating the population mean using information on auxiliary variable has also been discussed by various authors including Quenouille [15], Hansen et al. [12], Swain [1], Singh [14], Shukla [16], Srivastava and Jhajj [21], Kushwaha and Singh [10], Bahl and Tuteja [20], Banarasi et al. [2], H. P. Singh and R. Singh [5], Kadilar and Cingi [3], Koyuncu and Kadilar [17], Singh et al. [8], Singh and Solanki [6], Singh and Jatwa [7], Tailor et al. [22], Khan and Singh [13], and Khan and Abdullah [11], R. Singh et al. [18], R. Singh et al. [19], D. S. Robson [4].

Consider a finite population $U = U_1, U_2, U_3 \dots \dots \dots U_N$ of size N units numbered from 1 to N in some order .A sample of size n is taken size n units is taken at random from the first k units and every kth subsequent unit; then, = nk where n and k are positive integers; thus, there will be k samples (clusters) each of size n and observe the study variate y and auxiliary variate x for each and every unit selected in the sample

Let $(y_{ij,}x_{ij})\ for\ i = 1.2 \dots \dots k\ ,\quad j = 1{,}2 \dots \dots \dots n$ denote the value of jth unit in the ith sample. Then, the systematic sample means are defined as follows:

$\bar{y}_{st} = t_0 = 1/n \sum_{j=1}^{n} y_{ij,}$ and $\bar{x}_{st} = t_0 = 1/n \sum_{j=1}^{n} y_{ij,}$ are the unbiased estimators of the population means

$\bar{Y} = 1/n \sum_{j=1}^{n} y_{ij,}$ and $\bar{X} = 1/n \sum_{j=1}^{n} x_{ij,} of\ \ y\ \ on\ x$

To obtain the properties of the estimators up to first order of approximation, we use the following errors terms:

$$e_0 = \bar{y}_{sys} - \bar{Y}/\bar{Y}\ , e_1 = \bar{x}_{sys} - \bar{X}/\bar{X}\ , e_2 = \bar{z}_{sys} - \bar{Z}/\bar{Z}\ , Such\ that\ E(e_1) = 0\ for\ i = 0{,}1\ and\ 2$$

and

$$\rho_{yx} = \frac{s_{yx}}{s_y s_x}$$

$$\rho_{yz} = \frac{s_{yz}}{s_y s_z}$$

$$\rho_{xz} = \frac{s_{xz}}{s_x s_z}$$

$$k = \frac{\rho_{yx} C_y}{C_x}$$

$$k^* = \frac{\rho_{yz} C_y}{C_z}$$

$$\rho^*{}_y = \{1 + (n-1)\rho_y\}$$

$$\rho^*{}_x = \{1 + (n-1)\rho_x\}$$

$$\rho^*{}_z = \{1 + (n-1)\rho_z\}$$

$$\rho^{**} = \frac{\rho^*{}_y}{\rho^*{}_x}$$

$$\rho_2{}^{**} = \frac{\rho^*{}_y}{\rho^*{}_z}$$

$$\rho_1{}^{**} = \frac{\rho^*{}_y}{\rho^*{}_z}$$

Where

ρ_y, ρ_x, ρ_z are intra class correlation among the pair of units for the variables y,x and z.

2. Estimators in Literature:

In this part, we consider some estimators of the finite population mean in the sampling literature. The variance and MSE's of all the estimators computed here are obtained in the first order of approximation.

The variance of the unbiased estimator for population mean is

$$var(t_0) = \lambda \bar{Y}^2 \varphi_0 \tag{1.1}$$

Swain [14] and Shukla [16] suggested the classical ratio and product estimators for finite population mean by are given by

$$t_1 = \bar{y}_{sy}\left(\frac{\bar{X}}{\bar{x}_{sy}}\right) \tag{1.2}$$

$$t_2 = \bar{y}_{sy} exp\left(\frac{\bar{z}_{sy}}{\bar{z}}\right) \tag{1.3}$$

The mean square errors of the estimators to the first order of approximation are given as follows:

$$MSE(t_1) = \lambda \bar{Y}^2\left[\varphi_0 + \varphi_2\left(1 - 2k\sqrt{\rho^{**}}\right)\right] \tag{1.4}$$

$$\mathrm{MSE}(t_2) = \lambda \bar{Y}^2\left[\varphi_0 + \varphi_3\left(1 + 2k\sqrt{\rho_2{}^{**}}\right)\right] \tag{1.5}$$

Singh et al. [20] utilizing the known knowledge of the auxiliary variable, suggested the following ratio and product type exponential estimators:

$$t_3 = \bar{y}_{sy} exp\left(\frac{\bar{X} - \bar{x}_{sy}}{\bar{X} + \bar{x}_{sy}}\right) \tag{1.6}$$

$$t_4 = \bar{y}_{sy} exp\left(\frac{\bar{x}_{sy} - \bar{X}}{\bar{x}_{sy} + \bar{X}}\right) \tag{1.7}$$

$$\mathrm{MSE}(t_3) = \lambda \bar{Y}^2\left[\varphi_0 + \frac{\varphi_2}{4}\left(1 - 4k\sqrt{\rho^{**}}\right)\right] \tag{1.8}$$

$$\mathrm{MSE}(t_4) = \lambda \bar{Y}^2\left[\varphi_0 + \frac{\varphi_2}{4}\left(1 + 4k\sqrt{\rho^{**}}\right)\right] \tag{1.9}$$

Tailor et al. [25] define the following ratio-cum product estimator for the population mean $\bar{Y}$:

$$t_5 = \bar{y}_{sy}\left(\frac{\bar{X}}{\bar{x}_{sy}}\right)\left(\frac{\bar{z}_{sy}}{\bar{Z}}\right) \tag{2.0}$$

The mean square error of the estimator t_6 up to first order of approximation, is given by

$$\text{MSE}(t_5) = \lambda\bar{Y}^2\left[\varphi_0 + \varphi_2\left(1 - 2k\sqrt{\rho^{**}}\right) + \varphi_3\left(1 - 2k^{**}\sqrt{{\rho_1}^{**}}\right) + 2\varphi_4\sqrt{{\rho_y}^*{\rho_z}^*}\right] \tag{2.1}$$

Where,

$\varphi_0 = {\rho_y}^*{C_y}^2$

$\varphi_1 = 2{C_x}^2\sqrt{{\rho_y}^*{\rho_x}^*}$

$\varphi_2 = {\rho_x}^*{C_x}^2$

$\varphi_3 = {\rho_z}^*{C_z}^2$

$\varphi_4 = k^*{C^2}_z$

3. Proposed estimator

In this section, we have proposed an exponential ratio type estimator for population mean of the study variable y under systematic sampling as given by:

$$t_{RK} = \bar{\text{y}}_{\text{sy}}\left(\frac{\bar{\text{x}}_{\text{sy}}}{\bar{\text{X}}}\right)^{\alpha}\exp\left(\frac{\bar{\text{X}} - \bar{\text{x}}_{\text{sy}}}{\bar{\text{X}} + \bar{\text{x}}_{\text{sy}}}\right) \tag{2.2}$$

The first order of approximation of the above error terms is given by

$E({e_0}^2) = \lambda{\rho_y}^*{C_y}^2$,

$E({e_1}^2) = {\rho_x}^*{C_x}^2$

$E(e_0e_1) = \lambda{C^2}_x\sqrt{{\rho_y}^*{\rho_x}^*}$

Where, $\lambda = \left(\frac{N-1}{nN}\right)$

Expressing (2.2) in terms of e's

$$t_{RK} = \bar{Y}(1+\varepsilon_0)(1+\varepsilon_1)^{\alpha}exp\left(\frac{\bar{X} - \bar{X}(1+\varepsilon_1)}{\bar{X} + \bar{X}(1+\varepsilon_1)}\right)$$

$$t_{RK} = \bar{Y}(1+\varepsilon_0)(1+\varepsilon_1)^{\alpha}exp\left[\frac{\varepsilon_1}{2}\left(1+\frac{\varepsilon_1}{2}\right)^{-1}\right] \tag{2.3}$$

$$t_{RK} = \bar{Y}(1+\varepsilon_0)(1+\varepsilon_1)^{\alpha}exp\left[\frac{\varepsilon_1}{2}\left(1+\frac{\varepsilon_1}{2}+\frac{{\varepsilon_1}^2}{4}+\cdots\right)\right]$$

$$t_{RK} = \bar{Y}(1+\varepsilon_0)\left(1+\alpha\varepsilon_1+\frac{\alpha(\alpha-1)}{2}{\varepsilon_1}^2+\cdots\right)exp\left[\frac{\varepsilon_1}{2}\left(1+\frac{\varepsilon_1}{2}+\frac{{\varepsilon_1}^2}{4}+\cdots\right)\right] \tag{2.4}$$

From (2.4)

$$t_{RK} - \bar{Y} \cong \bar{Y}\left[\alpha\varepsilon_1 + \frac{\alpha(\alpha-1)}{2}{\varepsilon_1}^2 + \frac{\varepsilon_1}{2} + \frac{\alpha{\varepsilon_1}^2}{2} + \frac{3{\varepsilon_1}^2}{8} + \varepsilon_0 + \alpha\varepsilon_1\varepsilon_0 + \frac{\varepsilon_0\varepsilon_1}{2}\right] \tag{2.5}$$

Squaring (2.5) and then taking expectation on both sides, the MSE of the estimator $\bar{y}_{RK}$ is

$$MSE(t_{RK}) = \lambda\bar{Y}^2\left[\varphi_0 + \frac{K\varphi_1}{2} + \alpha K\varphi_1 + (\alpha^2+\alpha)\frac{\varphi_2}{2} + \frac{\varphi_2}{8}\right] \quad (2.6)$$

Obtain the optimum α to minimize $MSE(\bar{y}_{RK})$. Differentiating $MSE(\bar{y}_{RK})$ w.r.t α and equating the derivative to zero. Optimum value of α is given by:

$$\alpha = -\frac{(K\varphi_1 + \varphi_2)}{2\varphi_2}$$

Substituting the value of α_{opt} in (2.6), we get the minimum value of $\bar{y}_{RK}$ as:

$$MSE_{min}(t_{RK}) = \lambda\bar{Y}^2\varphi_0\left[1 - {\rho_{yx}}^2\right] \quad (2.7)$$

It follows from (2.7) that the proposed estimator $\bar{y}_{RK}$ at its optimum condition is equal efficient as that of the usual linear regression estimator.

4. Efficiency Comparisons

In this section, the MSE of traditional estimatorst_0,t_1, t_2, t_3, $t_4$and $t_5$are compared with the MSE of proposed estimator $\bar{y}_{RK}$.

From (1.1) - (2.0) and (2.1)

$$[var(t_0) - MSE_{min}(\bar{y}_{RK})] > 0$$

$$\left[\lambda\bar{Y}^2\varphi_0{\rho_{yx}}^2\right] > 0 \quad (2.8)$$

$$[MSE(t_1) - MSE_{min}(\bar{y}_{RK})] > 0$$

$$\lambda\bar{Y}^2\left[\varphi_0\left(1 - 2k\sqrt{\rho^{**}}\right)\right] - \left[\lambda\bar{Y}^2\varphi_0{\rho_{yx}}^2\right] > 0 \quad (2.9)$$

$$[MSE(t_2) - MSE_{min}(\bar{y}_{RK})] > 0$$

$$\lambda\bar{Y}^2\left[\varphi_0\left(1 + 2k\sqrt{\rho^{**}}\right)\right] - \left[\lambda\bar{Y}^2\varphi_0{\rho_{yx}}^2\right] > 0 \quad (3.0)$$

$$[MSE(t_3) - MSE_{min}(\bar{y}_{RK})] > 0$$

$$\lambda\bar{Y}^2\left[\frac{\varphi_0}{4}\left(1 - 4k\sqrt{\rho^{**}}\right)\right] - \left[\lambda\bar{Y}^2\varphi_0{\rho_{yx}}^2\right] > 0 \quad (3.1)$$

$$[MSE(t_4) - MSE_{min}(\bar{y}_{RK})] > 0$$

$$\lambda\bar{Y}^2\left[\frac{\varphi_0}{4}\left(1 + 4k\sqrt{\rho^{**}}\right)\right] - \left[\lambda\bar{Y}^2\varphi_0{\rho_{yx}}^2\right] > 0 \quad (3.2)$$

$$[MSE(t_5) - MSE_{min}(\bar{y}_{RK})] > 0$$

$$\lambda\bar{Y}^2\left[\varphi_0\left(1 - 2k\sqrt{\rho^{**}}\right) + \varphi_3\left(1 - 2k^{**}\sqrt{{\rho_1}^{**}}\right) + 2\varphi_4\sqrt{{\rho_y}^*{\rho_z}^*}\right] - \left[\lambda\bar{Y}^2\varphi_0{\rho_{yx}}^2\right] > 0 \quad (3.3)$$

5. Empirical Study

To examine the merits of the proposed estimator over the other existing estimators at optimum conditions, we have considered natural population data sets from the literature. The sources of population are given as follows.

(Source: Tailor et al. [20]). Consider Population

$N = 15$, $n = 3, \bar{X} = 44.47, \bar{Y} = 80, \bar{Z} = 48.40, C_y = 0.56, C_x = 0.28, C_z = 0.43$

$S_y^2 = 2000, S_x^2 = 149.55, S_z^2 = 427.83, S_{yx} = 538.57, S_{yz} = -902.86, S_{xz} = -241.06,$

$\rho_{yx} = 0.9848, \rho_{yz} = -0.9760 \rho_{xz} = -0.9530, \rho_y = 0.6652, \rho_x = 0.707, \rho_z = 0.5487.$

In order to find (PREs) of the estimator, we use the following formula and $PRE(t_\alpha, t_0) = MSE(t_0)/MSE(t_\alpha) \times 100$, For $\alpha = 0,1,2,3,4,5$ and RK

Table 1: *The percent relative efficiency of different estimators with respect to* t_0.

Population		
Estimator	MSE(t_α)	$PRE(t_\alpha, t_0)$
t_0	1455.08	100.00
t_1	373.32	389.62
t_2	768.06	189.45
t_3	820.09	177.43
t_4	1044.42	139.32
t_5	187.08	777.79
t_{RK}	43.88	3316.04

Fig 1: Estimator V/s MSE(t_α)

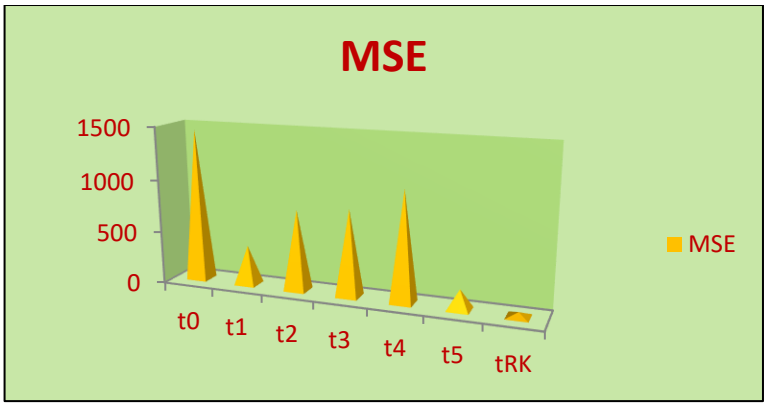
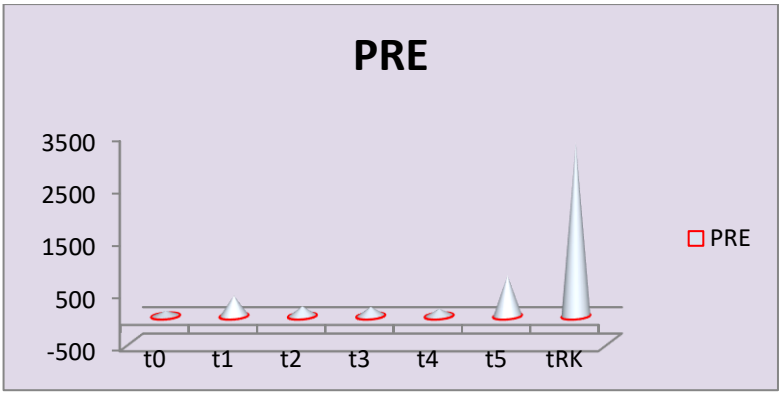


Fig 2: Estimator V/s $PRE(t_\alpha, t_0)$



6. Conclusion

In this article, an exponential ratio type estimator has been proposed under systematic design. The mathematical form of the estimator has been derived and its condition of efficiencies has been formulated with respect to some existing estimators from literature. The properties of the proposed estimator are obtained up to first order of approximation .it has been seen that the suggested estimator performed better than the existing estimator both theoretically and empirically.

References

[1] A. K. P. C. Swain, "The use of systematic sampling ratio estimate," Journal of the Indian Statistical Association, vol. 2, pp. 160–164, 1964.

[2] Banarasi, S. N. S. Kushwaha, and K. S. Kushwaha, "A class of ratio, product and difference (R.P.D.) estimators in systematic sampling," Microelectronics Reliability, vol. 33, no. 4, pp. 455–457, 1993.

[3] C. Kadilar, and H. Cingi, "An improvement in estimating the population mean by using the correlation coefficient," Hacettepe Journal of Mathematics and Statistics, vol. 35, no. 1, pp. 103–109, 2006.

[4] D. S. Robson, "Application of multivariate polykays to the theory of unbiased ratio-type estimation," The Journal of the American Statistical Association, vol. 59, pp. 1225–1226, 1957.

[5] H. P. Singh and R. Singh, "Almost unbiased ratio and product type estimators in systematic sampling," Questiio, vol. 22, no. 3, pp. 403–416, 1998.

[6] H. P. Singh and R. S. Solanki, "An efficient class of estimators for the population mean using auxiliary information in systematic sampling," Journal of Statistical Theory and Practice, vol. 6, no. 2, pp. 274–285, 2012.

[7] H. P. Singh and N. K. Jatwa, "A class of exponential type estimators in systematic sampling," Economic Quality Control, vol. 27, no. 2, pp. 195–208, 2012.

[8] H. P. Singh, R. Tailor, and N. K. Jatwa, "Modified ratio and product estimators for population mean in systematic sampling," Journal of Modern Applied Statistical Methods, vol. 10, no. 2, pp. 424–435, 2011.

[9] J.Hajeck, "Optimum strategy and other problems in probability sampling," Casopis pro Pestovani Matematiky, vol. 84, pp. 387– 423, 1959.

[10] K. S. Kushwaha and H. P. Singh, "Class of almost unbiased ratio and product estimators in systematic sampling," Journal of the Indian Society of Agricultural Statistics, vol. 41, no. 2, pp. 193– 205, vi, 1989.

[11] M. Khan and H. Abdullah, "A note on a differencetype estimator for population mean under twophase sampling design," SpringerPlus, vol. 5, article 723, 7 pages, 2016.

[12] M. H. Hansen, W. N. Hurwitz, and M. Gurney, "Problems and methods of the sample survey of business," Journal of the American Statistical Association, vol. 41, no. 234, pp. 173–189, 1946.

[13] M. Khan and R. Singh, "Estimation of population mean in chain ratio-type estimator under systematic sampling," Journal of Probability and Statistics, vol. 2015, Article ID248374, 2 pages, 2015.

[14] M. P. Singh, "Ratio-cum-product method of estimation," Metrika, vol. 12, pp. 34–42, 1967.

[15] M. H. Quenouille, "Notes on bias in estimation," Biometrika, vol. 43, pp. 353–360, 1956.

[16] N.D. Shukla, "Systematic sampling and product method of estimation," in Proceeding of the all India Seminar on Demography and Statistics, Varanasi, India, 1971.

[17] N. Koyuncu and C. Kadilar, "Family of estimators of population mean using two auxiliary variables in stratified random sampling," Communications in Statistics: Theory and Methods, vol. 38, no. 13–15, pp. 2398 2417, 2009.

[18] R. Singh, S.Malik, M. K. Chaudhary, H. Verma, and A. A. Adewara, "A general family of ratio-type estimators in systematicsampling," Journal of Reliability and Statistical Studies, vol. 5, no.1, pp. 73–82, 2012.

[19] R. Singh, S. Malik, and V. K. Singh, "An improved estimator in systematic sampling," Journal of Scientific Research, vol. 56, pp. 177–182, 2012.

[20] S. Bahl and R. K. Tuteja, "Ratio and product type exponential estimators," Journal of Information&Optimization Sciences, vol. 12, no. 1, pp. 159–164, 1991.

[21] S. K. Srivastava and H. S. Jhajj, "A class of estimators of the population mean using multi-auxiliary information," Calcutta Statistical Association Bulletin, vol. 32, no. 125-126, pp. 47–56, 1983.

[22] T. Tailor, N. Jatwa, and H. P. Singh, "A ratio-cum-product estimator of finite population mean in systematic sampling," Statistics in Transition, vol. 14, no. 3, pp. 391–398, 2013.

[23] W. G. Madow and L. H. Madow, "On the theory of systematic sampling. I," Annals of Mathematical Statistics, vol. 15, pp. 1–24,1944.

[24] W. G. Cochran, "Relative accuracy of systematic and stratified random samples for a certain class of populations," The Annals of Mathematical Statistics, vol. 17, pp. 164–177, 1946.

[25] W. Gautschi, "Some remarks on systematic sampling," Annals of Mathematical Statistics, vol. 28, pp. 385–394, 1957.

FAILURE RATE ESTIMATION BY WEIBULL DISTRIBUTION IN A STOCHASTIC ENVIRONMENT: APPLICATION TO THE HEMODIALYSIS MACHINE

Sofiene Fenina, Souheyl Jendoubi, Faker Bouchoucha,

Laboratoire de recherche LR-18ES45 Physique, Mathématique, Modélisation Quantique et Conception Mécanique, Institut Préparatoire aux Etudes d'Ingénieurs de Nabeul (IPEIN), université de Carthage, Tunisie, Campus Mrezgua, 8000 Nabeul.

E-mail address of the corresponding author: fakersbouchoucha@yahoo.fr

sofienefenina@yahoo.fr, jendoubi1souheyl@gmail.com

Abstract

This paper presented a study of the failure rate by introducing the effect of influencing variables. These variables have a random effect which depends on the external environment of the system. There are a multitude of variables and their modeling is difficult. The perturbation, to the failure rate, caused by external factors, has a direct impact on the time scale by the acceleration (or deceleration) of the degradation of the system. Therefore, the adopted methodology consists in introducing a perturbation on the Weibull parameters and studying its effect on the failure rate. Weibull parameters are considered random variables with a Gaussian distribution. The failure rate formulation in a random environment is offered through Weibull distribution. A case study of the hemodialysis machine is offered to illustrate the proposed approach and validate the results. The simulations presented show the failure rate statistics for different configurations of the Weibull distribution. The validation of the results was done using Monte Carlo simulations.

Keywords: Failure rate, Uncertain Factor, Gaussian perturbation, Weibull distribution

1. Introduction

Reliability is the probability that a system will perform a required function for a given period of time, under specified operating conditions [1, 2]. The conditions are all external constraints, whether mechanical, chemical, atmospheric, human, others. Dynamic reliability expands the static concept of failure by considering it dynamically over time [3]. This development may be due to variations in parameters influencing dysfunctional behavior (dynamic process, aging, etc.), modification of the system structure or environmental constraints [4].

The reliability and the functioning safety of the medical material are essential to ensure that a machine functions in accord with constructor instructions and assurances the patients and operators security. Damage of medical equipment may touch the healthcare services efficiency and causes severe harm to the patients and troubles the environment [5]. Scientific research plays an important role in today's society faced with major global challenges such as climate change, eradicating poverty and improving the quality of healthcare.

There are several methods and procedures employed to develop the reliability of the medical equipment. Dhillon [6] defined the maintenance of medical equipment as all actions necessary for retaining in, or returning to, a specified functioning condition. The aim of the study presented by Bahreini [7] is the extraction of the factors affecting the medical equipment reliability. The objective of the work presented by Khalaf [8] is the development of a mathematical approach that studies the influence of maintenance on the survival probability of medical equipment based on operating history and useful life.

Failure is a partial or total loss of the properties of an element which significantly decrease and leads to the total loss of its operating capacity. This failure may be due to its design, manufacture, installation, or even maintenance [9, 10, 11, 12]. Any production systems are subject to aging and wear [13]. These physical phenomena cause the failure, which has a significant impact on the cost of operating the system or on security.

In the literature, several authors have presented numerous classifications of failures. For example, Rausand [14] classified failures by cause, time, detectability, and degree. Deloux [15] categorized failures according to cause, on the one hand, and the impact on system performance, on the other. The classification of failures by cause differentiates between random and systematic failures. Classification of failures based on their impact on system performance distinguishes intermittent failures from extensive failures.

It is found that the failure rate in the different reliability databases vary significantly [16]. The causes of these discrepancies are numerous, for example, because of the characteristics of the equipment, the operating conditions, and the operating environment, the lack of precision in the information supplied and the increasing complexity of a reliable evaluation of equipment comparable to that of a component.

Stochastic degradation models are mathematical models that describe the degradation of the system over time. Degradation models were proposed as a tool to describe the state of a production system, to constitute a maintenance policy, to measure system availability and to obtain optimal maintenance periods [17].

The mechanisms of component degradation (operating conditions, fatigue, vibrations and other stresses, etc.) lead to time-dependent modeling of failure rates. Many studies show the impact that aging mechanical systems have on reliability as demonstrated [18]. Mechanical components are characterized by multiple, often complex degradation mechanisms of various origins (cracking, creep, wear, fatigue cracking, etc.) [19, 20]. These degradation modes include several parameters like material and dimensional characteristics, external stresses, etc. [21].

In some studies [22], two main types of models associated with the effects of aging factors are identified: physical models and empirical models. Booher [23] destined three models of degradation: shock model, wear model and hybrid model which combines the two processes (shock and wear). Degradation models may also be classified, according to Deloux [15], in two categories: discrete degradation models and continuous degradation models.

Influence factors are either internal or external factors that affect the reliability of the system. The influence can be positive by causing a reduction in the number of failures or, on the contrary, have negative effects on reliability. Depending on the type of system studied (human, electric/ electronic, mechanical), observed factors are generally different (human or organizational factors, system intrinsic or extrinsic factors). A classification of these factors, based on the life stages of the system under consideration, was proposed by Brissaud [24]: design factors, manufacturing factors, system installation factors, factors which influence system usage and maintenance factors. We can add to this list human and organizational factors that generally have a broader impact on the system [25].

The influence coefficients are calculated using physical relations, which are functions of many parameters like temperature, pressure, dimensions, properties of fluids and materials, etc. For many electrical and electronic components, failure rates, constant over time, are expressed analytically based on defined parameters. Couallier [26] developed a model adapted to the aeronautical maintenance data of on-board systems, some components of which are reliable according to the condition of other components. Without a priori knowledge of the physical relationships which link influencing factors to failure rates, statistical methods try to express correction coefficients. Ouakki [27] presented (figure 1) a cause-effect diagram (Ishikawa diagram) in order to better describe the different factors that cause the reliability varies.

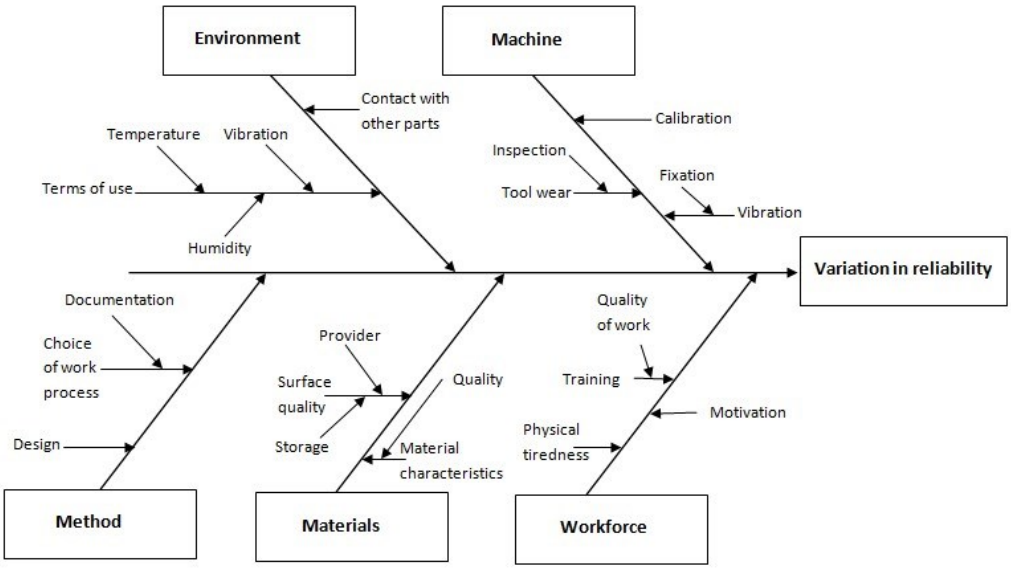


Figure 1: *Cause and effect diagram of the reliability*

The failure rate of mechanical systems is not constant and continually changes over time as a result of degradation phenomena such as wear, aging, etc. Any system is related to its external environment, the consideration of external influence variables allows a more robust modeling of the failure rate of the system. Several researchers proposed a failure rate model taking into account deterioration over time and influence factors where a Cox model is integrated to study the effect of the stress [24, 28, 29]. This is a semi-parametric model which describes the failure rate as a function of the basic failure rate of the system, which depends only on the time, and the influence function which depends only on the state of the influencing factors, it is independent of time. Several researchers have adopted the Cox model to estimate the influence function [24, 29, 30, 31, 32, 33]. The construction of this model requires the collection of input data, the preparation of a coding for the states of the influence factors, the determination of the parameters of the influence function, the determination of the parameters of the base failure rate and finally the synthesis of the results. It should be noted that in this model, the measures are always subject to a certain degree of uncertainty, such that the states of the influence factors can vary over time. The effectiveness of this model is closely associated with the quantity and quality of information available for the study. An analysis of the material is then required to select the influencing factors to consider.

In engineering studies, the distribution that best characterizes a set of data should be chosen [34]. In industry area, the Weibull distribution is one of the most used probability density functions. According to Lyonnet [35], the Weibull model is the best appropriate when carrying out reliability analysis for mechanical components. The main advantage of this distribution is its ability to account for small samples of failure data. Nevertheless the graphical method is

recommended in case of little model size data in terms of the estimation precision and accuracy [36]. The flexibility in fitting different failure modes and in simulating many other statistical distributions is one of the important attractions of the mentioned distribution [37]. The Weibull probability analysis is widely employed for studying the life data and can be applied to several situations. Weibull distribution is used in various domains [38] such as aerospace, electronics, materials, automotive industries and civil aviation [39]. This statistical approach can be an important way to analysis the reliability of semi-conductors, ball bearings, engines, spot weldings, biological organisms… [40]. Ahsan [41] studied the reliability of gas turbine using three parameters Weibull distribution based on historical data.

This paper presented a study of failure rate in stochastic environment by introducing the effect of influential variables. These variables have a random effect that depends on the external environment of the system. The external factors, which can influence the reliability of a system, are much diversified and their modeling is difficult. An approach to characterize the failure rate is proposed, that takes into account the consequences of these factors without modeling them. Our methodology consists in introducing a perturbation on the Weibull parameters and studying its effect on the failure rate. This perturbation is the result of influence variables. Weibull parameters are considered random variables with a Gaussian distribution. The formulation of the failure rate in a stochastic environment is developed through the Weibull distribution. A case study is offered to illustrate the proposed approach and validate the results. The simulations presented show the failure rate statistics for different configurations of the Weibull distribution.

2. Failure rate estimation by Weibull distribution in stochastic environment

The two-parameter Weibull distribution is a continuous probability distribution widely used for analyzing reliability and lifetime data [42]. The Weibull distribution is characterized by two parameters β and η, where β is the shape parameter and η is the scale parameter.

The failure rate is expressed through the Weibull distribution by the following function:

$$\lambda(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} \quad (1)$$

The Weibull distribution is characterized by two parameters (β, η):

β: is the shape parameter, (β >0). This parameter gives indications on the failure mode and on the evolution of the failure rate over time.

η: is the scale parameter, Which specifies the order of magnitude of the average lifespan.

The reliability analysis is based on a deterministic approach. In fact, all reliability parameters, which are uncertain, are described by unfavorable characteristic values. This often leads to unwarranted modeling and dimensioning. In that sense, behavioral prediction should preferably be in terms of probabilities. Uncertainties are related to variability in physical and geometric parameters, to fluctuations in load conditions, to stress boundary conditions and also to physical laws and simplifying assumptions used in the modeling process [21]. It is therefore the analysis of reliability by probabilistic approaches, taking into account the dispersion of the variables described by probabilistic distributions. Several factors contribute to the degradation of the component or entity. Consequently, the lifetimes must be explained by the different variables (degradation factors) contributing to the failure.

Several researchers [22, 29] have considered that the influencing variables mainly affect the weibull parameters. Their effect is to slow down or accelerate the degradation. The influence of these variables is random and hardly modelable. In this work, we will introduce a perturbation in the weibull parameters due to the influencing variables that contribute to the degradation and our objective is to estimate the effect of this perturbation on the failure rate. The shape and scale

parameters are modeled by random variables which follow Gaussian distributions and written in the following form:

$$\tilde{\beta} = \beta + \sigma_{\beta}\varepsilon \quad (2)$$

$$\tilde{\eta} = \eta + \sigma_{\eta}\varepsilon \quad (3)$$

Where

- $\tilde{\beta}$, $\tilde{\eta}$: Random Weibull parameters following a Gaussian distribution,
- β, η: The mean of the random weibull parameters,
- σ_{β}, σ_{η}: The standard deviation of the weibull parameters (disturbance around the mean),
- ε: Reduced centered Gaussian variable,

The deterministic weibull parameters are represented by their means β and η. The perturbation caused by several external influencing factors is characterized by their standard deviation σ_{β} and σ_{η}. Hence, our study consists of estimating the effect of this perturbation on the failure rate using the Weibull distribution. The random failure rate is expressed as a random function following a Gaussian distribution as follow:

$$\tilde{\lambda}(t) = \lambda(t) + \sigma_{\lambda}(t)\varepsilon \quad (4)$$

Where λ and σ_{λ} are the mean and the standard deviation of the failure rate, respectively.

According to Weibull distribution, the random failure rate is offered by the following expression:

$$\tilde{\lambda}(t) = \frac{\tilde{\beta}}{\tilde{\eta}}\left(\frac{t}{\tilde{\eta}}\right)^{\tilde{\beta}-1} \quad (5)$$

In order to formulate the mean and the standard deviation of the failure rate, we can write:

$$log\left(\tilde{\lambda}(t)\right) = log\left(\tilde{\beta}\right) - log(\tilde{\eta}) + \left(\tilde{\beta} - 1\right)[log(t) - log(\tilde{\eta})] \quad (6)$$

$$log\left(\tilde{\lambda}(t)\right) = log\left(\tilde{\beta}\right) - \tilde{\beta}\, log(\tilde{\eta}) + \left(\tilde{\beta} - 1\right) log(t) \quad (7)$$

To linearize the failure rate equation, we will use the first-order Taylor series expansion of log ($\tilde{\lambda}$(t)), log ($\tilde{\beta}$) and log ($\tilde{\eta}$) in the vicinity of their mean, we obtain:

$$log\left(\tilde{\lambda}(t)\right) = log\left(\lambda(t)\right) + \frac{\sigma_{\lambda}(t)}{\lambda(t)}\varepsilon \quad (8)$$

$$log\left(\tilde{\beta}\right) = log(\beta) + \frac{\sigma_{\beta}}{\beta}\varepsilon \quad (9)$$

$$log(\tilde{\eta}) = log(\eta) + \frac{\sigma_{\eta}}{\eta}\varepsilon \quad (10)$$

Introducing equations (8), (9) and (10) in equation (7), we can write:

$$log\left(\lambda(t)\right) + \frac{\sigma_{\lambda}(t)}{\lambda(t)}\varepsilon = [log(\beta) - \beta\, log(\eta) + (\beta - 1)\, log(t)] + \left[\frac{\sigma_{\beta}}{\beta} - \frac{\beta}{\eta}\sigma_{\eta} - \sigma_{\beta}\, log(\eta) + \sigma_{\beta}\, log(t)\right]\varepsilon \quad (11)$$

The identification of the different terms in equation (11) leads to extract the mean and the standard deviation of the failure rate as:

$$log\left(\lambda(t)\right) = log(\beta) - \beta\, log(\eta) + (\beta - 1)\, log(t) \quad (12)$$

And

$$\frac{\sigma_{\lambda}(t)}{\lambda(t)} = \frac{\sigma_{\beta}}{\beta} - \frac{\beta}{\eta}\sigma_{\eta} - \sigma_{\beta}\, log(\eta) + \sigma_{\beta}\, log(t) \quad (13)$$

Finally, we obtain:

$$\lambda(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \quad (14)$$

$$\sigma_{\lambda}(t) = \lambda(t)\left[\frac{\sigma_{\beta}}{\beta} - \frac{\beta}{\eta}\sigma_{\eta} - \sigma_{\beta}\, log(\eta) + \sigma_{\beta}\, log(t)\right] \quad (15)$$

3. Numerical results and discussion

3.1. Case study: Weibull distribution analysis of the Hemodialysis machine

Hemodialysis machines are one of the important medical equipment which is directly responsible for the patient's life, used to treat kidney failure. The reliability of hemodialysis machines is very important for nephrologists to guarantee not only patient safety but also efficiency and continuity of treatment. Weibull distribution is very flexible and can, through an appropriate choice of parameters, model many types of failure rate behaviors.

In this paragraph, the failure rate of a group of 3 hemodialysis machines (M1, M2 and M3) will be studied. The data of the failure history of the 3 machines was collected during the period from 2013 to 2022. The data of the failure history of the 3 devices was collected and summarized in table 1.

Table 1: *The failure history (TBF) of hemodialysis machines.*

Failure number	M1		M2		M3	
	Date of failure	TBF	Date of failure	TBF	Date of failure	TBF
1	29/01/2014	1085	05/06/2015	2450	29/05/2013	413
2	04/01/2016	1904	29/05/2017	2040.5	12/07/2013	101.5
3	19/10/2016	609	07/08/2017	192.5	01/09/2014	1130.5
4	24/10/2016	10.5	28/02/2018	549.5	05/06/2015	763
5	29/05/2017	605.5	03/08/2018	311.5	09/11/2015	427
6	06/03/2020	2792	06/03/2020	584.5	07/08/2017	1750
7	12/11/2020	392	03/08/2020	280	31/10/2018	1228.5
8					09/08/2019	780.5
9					06/03/2020	560
10					29/07/2020	255.5
11					07/01/2022	1456

The shape and scale parameter will be extracted in order to characterize the failure rate behavior. The cumulative density function (CDF) is formulated using Weibull distribution in the following form:

$$F(t) = 1 - \exp\left(-\frac{t}{\eta}\right)^{\beta} \tag{16}$$

In order to extract the weibull parameters (β and η), we use the linearization of the cumulative density function given in equation (16) as follow:

$$F(t) - 1 = -\exp\left(-\frac{t}{\eta}\right)^{\beta} \tag{17}$$

$$\ln\big(1 - F(t)\big) = -\left(\frac{t}{\eta}\right)^{\beta} \tag{18}$$

$$\ln\big(-\ln\big(1 - F(t)\big)\big) = \ln\left(\frac{t}{\eta}\right)^{\beta} \tag{19}$$

$$\ln\left(\ln\left(\frac{1}{1-F(t)}\right)\right) = \beta \ln(t) - \beta \ln(\eta) \tag{20}$$

The CFD equation can be written in the following linear form with a slop of β and an intercept of $\beta \ln(\eta)$:

$$y = \beta x - \beta \ln(\eta) \tag{21}$$

Where $y = \ln\left(\ln\left(\frac{1}{1-F(t)}\right)\right)$ and $x = \ln(t)$

Failure times of Hemodialysis Machine were collected and arranged in ascending order to

calculate the cumulative density function (CFD), denoted by $F(t_i)$, by using the Bernard's formula to assign median ranks to each failure given as:

$$F(t_i) = \frac{i-0.3}{n+0.4} \quad (22)$$

Where i is the order of failure and n is the total number of data

The necessary calculation steps performed to determine the Weibull parameters (β, η) are summarized in table 2.

The linear form of the cumulative density function given in equation (21) is presented in figure 2 for the 3 hemodialysis machines. The shape and scale parameters can be calculated from the linear equation offerred in figure 2. Table 3 shows the Weibull parameters β and η for the 3 studied devices (M1, M2 and M3).

Table 2: *The necessary calculation steps*

Failure number	M1		M2		M3	
	TBF_i	$F(ti)$	TBF_i	$F(ti)$	TBF_i	$F(ti)$
1	10.5	0,09459459	192,5	0,09459459	101,5	0,061403
2	392	0,22972973	280	0,22972973	255,5	0,149122
3	605.5	0,36486486	311,5	0,36486486	413	0,236842
4	609	0,5	549,5	0,5	427	0,324561
5	1085	0,63513514	584,5	0,63513514	560	0,412280
6	1904	0,77027027	2040,5	0,77027027	763	0,5
7	2792	0,90540541	2450	0,90540541	780,5	0,587719
8					1130,5	0,675438
9					1228,5	0,763157
10					1456	0,850877
11					1750	0,938596

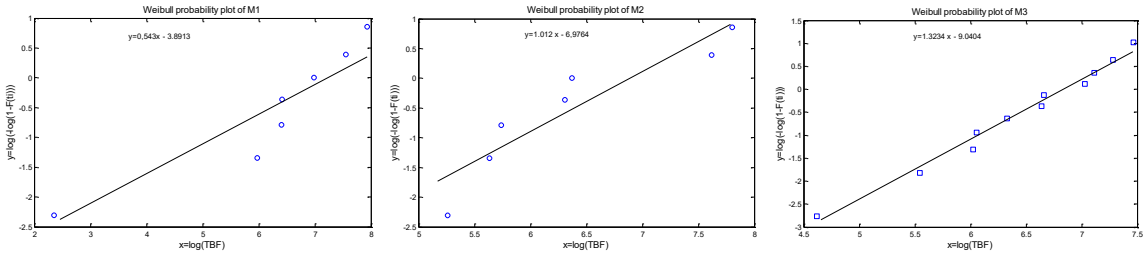


Figure 2: *Weibull probability plot of hemodialysis machines*

Table 3: *The shape and scale parameters of hemodialysis machines*

Machine	β	η
M1	0,543	1294
M2	1.012	986
M3	1,323	928

The uncertainties in the determination of the Weibull parameters lead to introduce a perturbation in the shape and scale parameters. In this part, the effect of this perturbation on the behavior of the failure rate will be studied. The mean and the standard deviation of the failure rate will be presented in many configurations. The shape and scale parameters, according to the equations (2) and (3), follow a Gaussian distribution as $\tilde{\beta} = \beta + \sigma_\beta \varepsilon$ and $\tilde{\eta} = \eta + \sigma_\eta \varepsilon$.

The simulations are performed for the 3 hemodialysis machines. Figure 3 presents the mean of the failure rate $\lambda(t)$ expressed using the Weibull distribution by the equation (14). According to this study, the failure rate $\lambda(t)$ of M3 increases significantly with time. M3 is in the aging life ($\beta>1$) and must be supervised continuously. The failure rate of M2 is constant during the time, the machine is in the useful life ($\beta=1$). The machine M1, which is in the early-life ($\beta<1$), have a decreasing failure rate.

Figure 4 presents the standard deviation of the failure rate for the 3 hemodialysis machines. The simulations were done, according to the equation (15), with a standard deviation introduced in the shape and scale parameters equivalent to: $\sigma_\beta = 10\%\beta$ and $\sigma_\eta = 10\%\eta$. The curves presented in figure 4 illustrate the evolution of the uncertainties in the failure rate for the different phases of the lifetime of the hemodialysis machine. The standard deviation for aging life (M3) is more significant and increases during this mature phase.

To highlight these interpretations, the correlation between the standard deviation and the mean of the failure rate for the 3 hemodialysis machines is studied. According to figure 5, the standard deviation evolves linearly according to the mean, which makes it possible to estimate the variation of the failure rate through the evaluation of the mean. In the aging phase, the effect of the perturbation introduced in the Weibull parameters is more serious and affect perilously the failure rate.

The uncertainty, introduced in the shape parameter, is related to the influencing factors which are ambiguous and described by unfavorable characteristic values. Many case studies are presented in order to illustrate the behavior of the failure rate following different level of uncertainty introduced in the shape parameter. Figure 6 shows the influence of the standard deviation of the shape parameter on the failure rate in the case of the aging life.

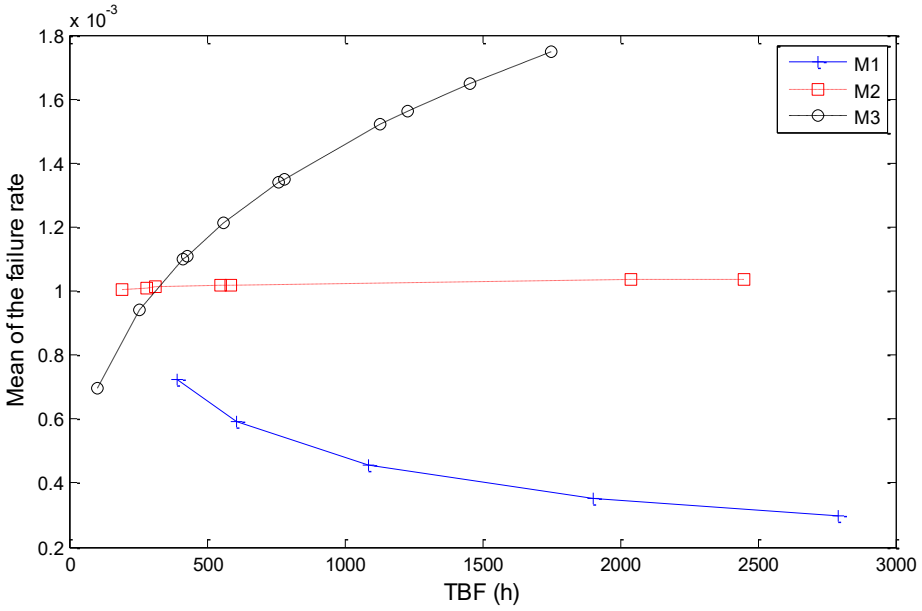


Figure 3: *The mean of the failure rate for the hemodialysis machines*

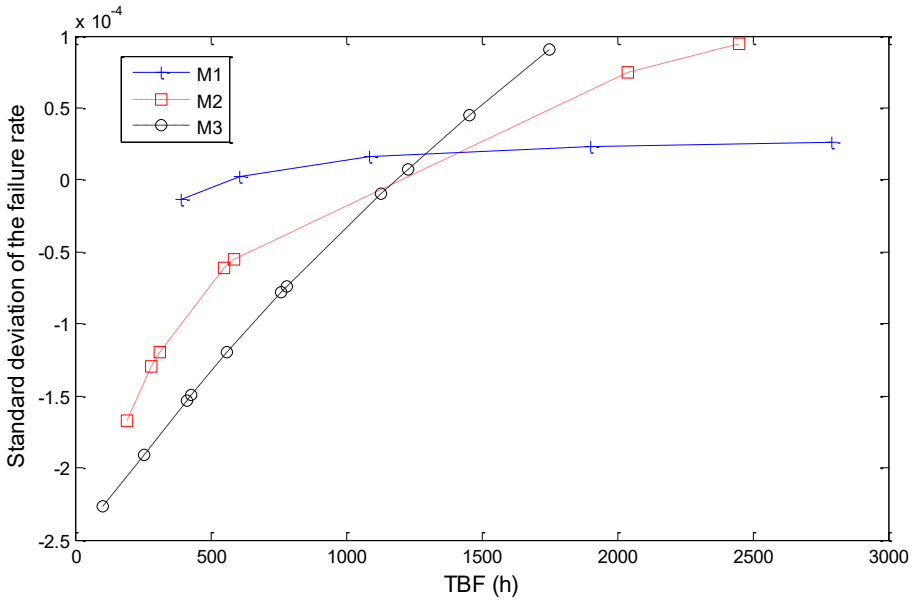


Figure 4: *The standard deviation of the failure rate for the hemodialysis machines*

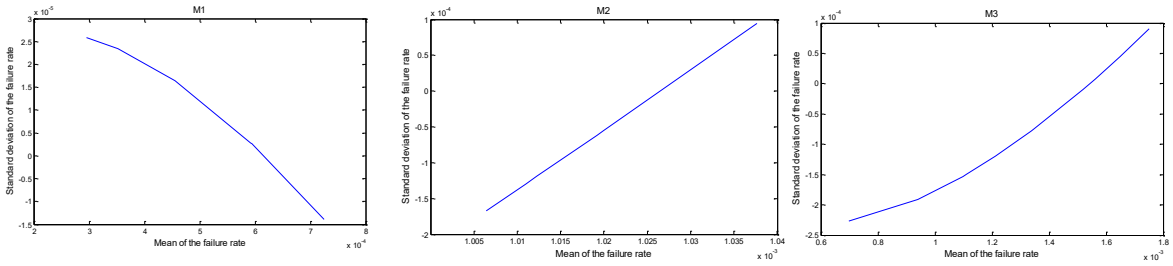


Figure 5: *Correlation between the standard deviation and the mean*

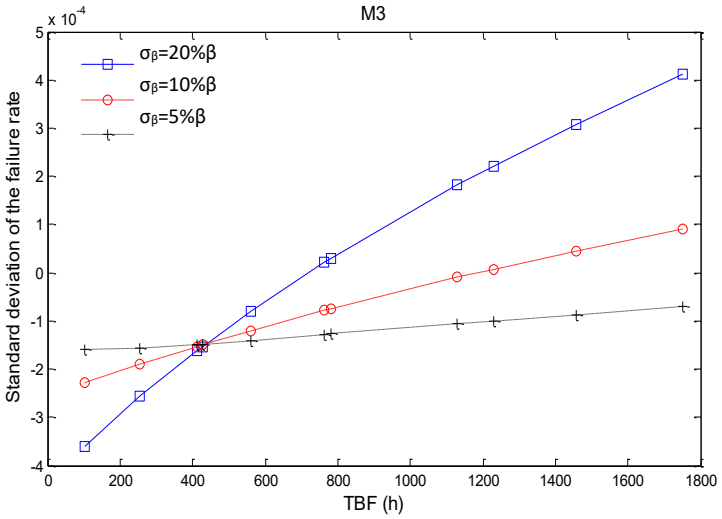


Figure 6: *Influence of the standard deviation of the shape parameter on the failure rate*

In conclusion, the results provide evidence of the propagation of errors that can affect a system and their significant influences on the failure rate. The presence of influencing factors can affect the determination of Weibull parameters and then the failure rate will be significantly affected, in particular, at aging phase. To avoid the inconvenient impact of the uncertainties in the failures on the dialysis system, it is recommended to upgrade the operation management. Moreover, the maintenance strategy must initially focus on the M3 and then at M2, which are on aging phase.

3.2. Validation of the results

The objective of this part is the validation of the results obtained by our proposed method in which the statistics of the failure rate were estimated using the Weibull distribution with uncertain parameters that follow a Gaussian distribution because of various external factors.

Validation will be ensured by comparing the obtained results with Monte Carlo simulations. The Monte Carlo (MC) method can be used as a reference for statistical methods. It consists of producing a Gaussian distribution of time and simulating several draws of the failure rate using the Weibull distribution and deducing the mean by:

$$\lambda(t) = \frac{1}{N}\sum_{i=1}^{N}\lambda_i(t) \tag{23}$$

Where $\lambda_i(t)$ is the failure rate for the i^{th} draw at time (t) and N is the total number of draws.

The standard deviation of the failure rate is given by:

$$\sigma_\lambda(t) = \sqrt{\frac{\sum_{i=1}^{N}(\lambda_i(t)-\lambda(t))^2}{N}} \tag{24}$$

During the MC simulations, we performed N=10000 draws to ensure the convergence of the results.

Figure 7 and Figure 8 illustrate the validation of the mean and standard deviation of the failure rate by MC simulations. The simulations were prepared in the case of aging life which presents the most critical case.

To further our research, the probability function of the failure rate was estimated following a Gaussian perturbation introduced into the Weibull parameters. This study is accomplished for the aging life (M3) and at a time corresponding to 1500 hours. According to figure 9, the probability function of the failure rate follows a Gaussian distribution. This numerical result confirms the analytical formulations proposed in this work. The MC simulations show the validity of the results.

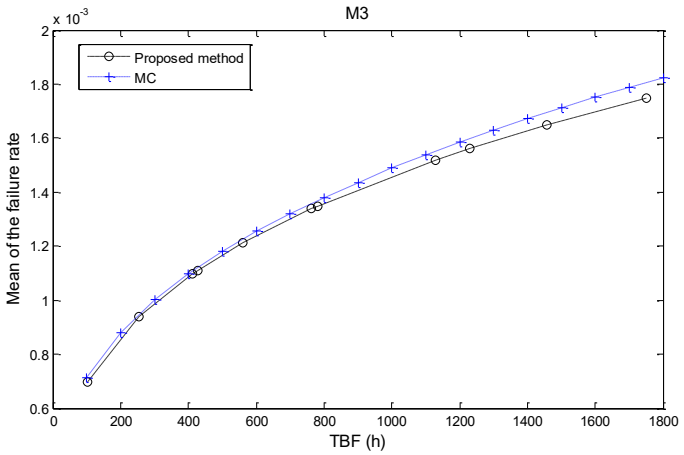


Figure 7: *Validation of the mean of the failure rate for the hemodialysis machine*

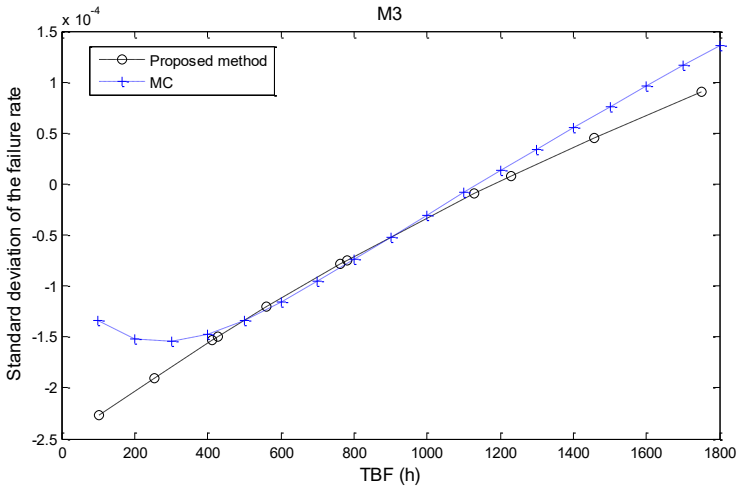


Figure 8: *Validation of the standard deviation of the failure rate for the hemodialysis machine*

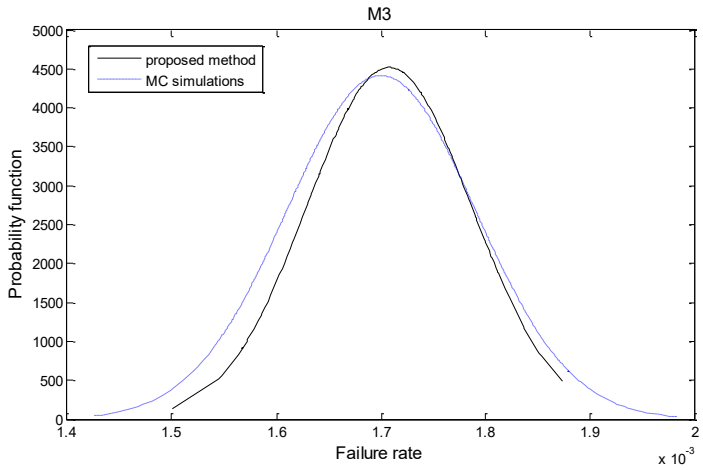


Figure 9: *Probability function of the failure rate*

4. Conclusion and perspectives

In this work, we studied the failure rate by introducing the effect of influencing variables. The effect of these variables is random and depends on the external environment of the system and directly affects the failure rate by accelerating (or decelerating) the degradation of the system.

The methodology adopted consists in introducing a perturbation on the Weibull parameters and studying its effect on the failure rate. Weibull parameters are considered random with a Gaussian distribution. The formulation of the failure rate in a stochastic environment is detailed using Weibull distribution. The simulations presented show the statistics of the failure rate for several configurations of the Weibull distribution. A case study is offered to illustrate the proposed approach and validate the results. The validation of the results was accomplished through Monte Carlo simulations.

This work can be extended along several lines of research such as the study of the reliability of a system in a stochastic environment taking into account external factors, as well as the study of availability in the presence of uncertain external influencing factors. The development of the

failure rate following a second-order perturbation introduced in the Weibull parameters can be the subject of an in-depth and precise study.

Acknowledgments

The authors dedicate this work to the soul of Mr. Rabii Ben Kdissa that god took him for his participation in the realization of this article.

Statements and Declarations

The authors declare no conflict of interest, no funding was received.

References

[1] Dhillon, B. (2006). Maintainability and Reliability for Engineers. CRC Press, Taylor & Francis Group, ISBN-13: 978-0-8493-7243-8.

[2] Lyonnet, P. (1992). La maintenance. Mathématiques et méthodes. 3ème édition revue et augmentée, Tec Doc éditions, ISBN: 2-85206-745-5.

[3] Bertsche, B. (2008). Maintenance and Reliability, Reliability in Automotive and Mechanical Engineering. Springer, Berlin, Heidelberg, ISBN: 978-3-540-34282-3 https://doi.org/10.1007/978-3-540-34282-3_10

[4] Devooght, J., Lewins, J., & Becker, M. (2002). Dynamic Reliability Advances in Nuclear Science and Technology. 25 Springer, Boston, MA, https://doi.org/10.1007/0-306-47812-97

[5] Zamzam, A.H., Abdul Wahab, A.K., Azizan, M.M., Satapathy, S.C., Lai, K.W., & Hasikin, K. (2021). A Systematic Review of Medical Equipment Reliability Assessment in Improving the Quality of Healthcare Services. Front. Public Health, 9:753951. https://doi.org/10.3389/fpubh.2021.753951.

[6] Dhillon, B.S. (2011). Medical Equipment reliability: A review, analysis methods and improvement strategies. International Journal of Reliability, Quality and Safety Engineering, 18(4), 391–403, https://doi.org/10.1142/S0218539311004317

[7] Bahreini, R., Doshmangir, L., & Imani, A. (2018). Factors Affecting Medical Equipment Maintenance Management: A Systematic Review. Journal of Clinical and Diagnostic Research, 12 (4), IC01-IC07.

[8] Khalaf, A.B., & Hamam, Y. (2013). The effect of maintenance on the survival of medical equipment. Journal of Engineering, Design and Technology , 11(2), 142-157. https://doi.org/10.1108/JEDT-06-2011-0033

[9] Procaccia, H., & Piepszownik, L. (1992). Fiabilité des équipements et théorie de la décision statistique fréquentielle et bayésienne. Ed. Ayrolle, ISBN-13: 978-2212016314

[10] Arunraj, N.S., & Maiti, J. (2007). Risk-based maintenance-Techniques and applications. Journal of Hazardous Materials, 142(3), 653-661. ISSN 0304-3894, https://doi.org/10.1016/j.jhazmat.2006.06.069

[11] Monchy, F. (2010). Maintenance: Méthodes et organisations. 3ème édition, Dunod, Paris ISBN 978-2-10-055061-6

[12] Brissaud, F., Charpentier, D., Fouladirad, M., Barros, A., & Bérenguer, C. (2010). Failure rate evaluation with influencing factors. Journal of Loss Prevention in the Process Industries, 23(2), 187-193 ISSN 0950-4230, https://doi.org/10.1016/j.jlp.2009.07.013

[13] Ciriaco, V.F., & Richard, M.F. (1989). A survey of preventive maintenance models for stochastically deteriorating single-unit systems. Naval Research Logistics 36, 419-446 https://doi.org/10.1002/1520-6750(198908)36:4<419::AID-NAV3220360407>3.0.CO;2-5

[14] Rausand, M., & Hoyland, A. (2004). System Reliability Theory: Models, Statistical Methods, and Applications. Second édition, John Wilev & Sons. Inc.

[15] Deloux, E. (2008). Politiques de maintenance conditionnelle pour un système à dégradation continue soumis à un environnement stressant. Sciences de l'ingénieur [physics], Université de Nantes https://tel.archives-ouvertes.fr/tel-00348191

[16] Aupied, J. (1993). Retour d'expérience appliqué à la sûreté de fonctionnement des matériels en exploitation. Eyrolles ISBN-13:978-2-212-01638-3.

[17] Gölbaş, O., & Demirel, N. (2014). Stochastic Models in Preventive Maintenance Policies. Advanced Materials Research, 1016, 802–806, Trans Tech Publications, Ltd. https://doi.org/10.4028/www.scientific.net/amr.1016.802

[18] Alfares, H. (1999). A simulation model for determining inspection frequency. Computers & Industrial Engineering, 36(3), 685-696.

[19] Lalanne, C. (1999). Vibrations et Chocs Mecaniques T4: Dommage Par Fatigue. Hermes Sciences Publications, ISBN-13:978-2746200388.

[20] Habchi, G., & Barthod, C. (2016). An overall methodology for reliability prediction of mechatronic systems design with industrial application. Reliability Engineering & System Safety, 155, 236-254, https://doi.org/10.1016/j.ress.2016.06.013

[21] Tebbi, O. (2005). Estimation des lois de fiabilité en mécanique par les essaies accélérés. PhD thesis, Autre. Université d'Angers, Français. NNT: tel-00009407.

[22] Salameh, F. (2016). Méthodes de modélisation statistique de la durée de vie des composants en génie électrique. PhD thesis , Institut National Polytechnique de Toulouse, http://www.theses.fr/2016INPT0082/document

[23] Booher, T.B. (2006). Optimal Periodic Inspection of a Stochastically Degrading System. Theses and Dissertations 3291, https://scholar.afit.edu/etd/3291

[24] Brissaud, F., Lanternier, B., Charpentier, D., & Lyonnet, P. (2007). Modélisation des taux de défaillance en mécanique - Combinaison d'une loi de Weibull et d'un modèle de Cox pour la modélisation des taux de défaillance en fonction du temps et des facteurs d'influence. Performance et Nouvelles Technologies en Maintenance, PENTOM 2007, Mons Belgique. https://hal.archives-ouvertes.fr/hal-00196765

[25] Aven, T., Sklet, S., & Vinnem, J.E. (2006). Barrier and operational risk analysis of hydrocarbon releases (BORA-Release): Part I. Method description. Journal of Hazardous Materials, 137(2), 681-691. ISSN 0304-3894, https://doi.org/10.1016/j.jhazmat.2006.03.049

[26] Couallier, V., Denis, L., & Bayle, F. (2016). Calcul de MTBF des systèmes aéronautiques inspectés périodiquement soumis à rupture aléatoire de systèmes de régulation. 20ème Congrès de maîtrise des risques et de sûreté de fonctionnement, Saint-Malo, France, Institut pour la Maîtrise des risques https://doi.org/10.4267/2042/61795

[27] Ouakki, W. (2011). Prise en compte de la fiabilité en conception. Mémoire de maîtrise en génie mécanique, Université Laval, http://hdl.handle.net/20.500.11794/23131

[28] Lanternier, B., Charpentier, D., & Lyonnet, P. (2006). Modélisation de taux de défaillance en mécanique. Colloque de maîtrise des risques et sûreté de fonctionnement « Risques et performances », Lille, France, pp 4, ineris-00973237, https://hal-ineris.archives-ouvertes.fr/ineris-00973237

[29] Gouno, E., & Guérineau, L. (2015). Failure Rate Estimation in a Dynamic Environment. Economic Quality Control, 30(1), https://doi.org/10.1515/eqc-2015-6001

[30] Cox, D.R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological), 34(2), 187–220 [Royal Statistical Society, Wiley] http://www.jstor.org/stable/2985181.

[31] Ansell, J.I., & Phillips, M.J. (1994). Practical Methods for Reliability Data Analysis. Oxford Statistical Science Series, 14 Clarendon Press, 1st edition, December 8 ISBN-13: 978-0198536642.

[32] Carroll, K.J. (2003). On the use and utility of the Weibull model in the analysis of survival data. Control Clinical Trials, 24(6), 682-701. doi: 10.1016/s0197-2456(03)00072-2. PMID: 14662274

[33] Øien, K. (2001). A framework for the establishment of organizational risk indicators, Reliability Engineering & System Safety, 74(2), 147-167. ISSN 0951-8320, https://doi.org/10.1016/S0951-8320(01)00068-0

[34] O'Connor, D.T., & Kleyner, A. (2012). Practical reliability engineering. John Wiley and Sons, Chichester, UK, 5th edition.

[35] Lyonnet, P. (1991). Tools of Total Quality, An Introduction to statistical Process Control. Chapman and Hall, London.

[36] Yong, T. (2004). Extended weibull distributions in reliability engineering. PhD thesis, National University of Singapore.

[37] Lihou, D.A., & Spence, G.D. (1988). Proper use of data with the Weibull distribution. Journal of Loss Prevention in the Process Industries, 1(2), 110-113, ISSN 0950-4230, https://doi.org/10.1016/0950-4230(88)80021-4.

[38] Luko, S.N. (1999). A review of the Weibull Distribution and Selected Engineering Applications. SAE technical paper series. DOI: https://doi.org/10.4271/1999-01-2859

[39] Fidanoglu, M., Ungor, U., Ozkol, I., & Komurgoz, G. (2017). Application of Weibull Distribution Method for Aircraft Component Life Estimation in Civil Aviation Sector. Journal of Traffic and Logistics Engineering, 5(1).

[40] Rausand, M., Barros, A., & Hoyland, A. (2020). System Reliability Theory Models, Statistical Methods, and Applications. Third edition Wiley, ISBN: 9781119373957

[41] Ahsan, S., Lemma, T.A., & Gebremariam, M.A. (2019). Reliability analysis of gas turbine engine by means of bathtub-shaped failure rate distribution. Proc Safety Prog. e12115. https://doi.org/10.1002/prs.12115.

[42] Pasha, G.R., Khan, M.S., & Pasha, A.H. (2006). Empirical Analysis of The Weibull Distribution for Failure Data. Journal of Statistics, 13(1), 33-45.

ROBUST CLASSIFICATION USING MINIMUM REGULARIZED COVARIANCE DETERMINANT ESTIMATOR

R Muthukrishnan and Surabhi S Nair
•
Department of Statistics, Bharathiar University, Coimbatore 641046, Tamil Nadu, India
muthukrishnan1970@gmail.com, surabhinair93@gmail.com

Abstract

The association between a categorical variable and a group of interconnected factors is the main objective of the classification procedure. The linear discriminant analysis (LDA) aims to provide a method for classifying populations and dividing up forthcoming observations among the groups that have already been identified. Under the suppositions of normality and homoscedasticity, the LDA produces the best discriminant rule for two or more groups. Outliers have a significant impact on the parameters of the LDA, mean, and covariance matrix. Robust methods are resistant to outliers. This paper explores the robust methods, namely the Minimum Covariance Determinant (MCD) estimator and Minimum Regularized Covariance Determinant (MRCD) estimators in the context of discriminant analysis under real environments. The MCD technique is used to estimate the location and dispersion matrix using the subset of the given size that has the lowest sample covariance determinant. Its fundamental problem is that it doesn't provide a reliable result when the features/dimension is greater than the size of the subset. As a result, the MRCD method is employed and the efficiency is studied by computing the Apparent Error Rate (AER). In this paper, an attempt has been made to review the existing theory and methods of RLDA.

Keywords: classification, linear discriminant analysis, robust linear discriminant analysis, minimum covariance determinant estimator, minimum regularized covariance determinant estimators

I. Introduction

The traditional Linear Discriminant Analysis (LDA) is a frequently used multi-dimensional classification approach to classify new observations according to one of the aforementioned categories, Elizabeth and Andres [4]. The population means (μ_1, μ_2) with homoscedasticity assumptions $(\Sigma_1 = \Sigma_2 = \Sigma)$ is used for establishing traditional LDA. The Traditional Linear Discriminant Rule will be built using predicted mean vector and scatter matrices while real population parameters are typically unachievable. In particular, the traditional mean is very susceptible to anomalies. Merely just one anomaly can affect the accuracy of the covariance and alter the location estimation, Erceg-hurn et al., [5]. Thus, an immense misclassification rate will be caused by the affected mean and covariances Sajobi et al., [15].

Researchers are looking for solutions in Robust Linear Discriminant Analysis (RLDA) to address this sensitivity issue in LDA. Also can build robust discriminant models with low classification error rates by replacing the classical estimators with robust estimators such as M-

estimators, Minimum Covariance Determinant (MCD) estimators, Hubert and Driessen [9], Minimum Volume Ellipsoid (MVE) estimators, Choral and Rousseeuw [2], and S-estimators, He and Fung [8], Croux and Dehon [3], Minimum Regularized Covariance Determinants (MRCD), Boudt and Peter Rousseeuw [1].

This paper mainly compares the robust estimators such as MCD and MRCD in RLDA with traditional LDA. The Apparent Error Rate (AER) is used to calculate how effective certain strategies are. The rest of this paper is structured as follows. Section 2 describes traditional LDA and robust linear discriminant analysis based on MCD and MRCD estimators. The results and discussions based on the real data study will be given in section 3. The conclusion will be provided in the last section.

II. Classification Methods

The traditional classification method, namely, Linear Discriminant Analysis (LDA)), the Robust Linear Discriminant Analysis using Minimum Covariance Determinant (MCD) estimator, and Minimum Regularized Covariance Determinant (MRCD) estimators are briefly discussed in this section.

Linear Discriminant Analysis (LDA)

Fisher [6] introduced the linear discriminant for two classes and C.R. Rao [11] later generalized it for many classes. Linear discriminant analysis (LDA) is a group of multivariate statistical techniques used to identify a linear combination of features that characterize or distinguish two or more classes of objects or events. Hastie et. al. [7], Sharipah et. al. [17]. Classification using a linear function is known as discriminant analysis. The discriminant analysis aims to divide the sample variables into two or more categories. This is accomplished with the use of a linear combination of explanatory factors or forecasting variables. Choosing a group for an object is based on the fundamental tenet that there should be as little chance of misclassification as possible.
On account of the two-group discriminant model, grouping an entity into one of two groups, g_1 or g_2, is the main objective. It is speculated that the explanatory variables will exhibit a multi-variate normal distribution.

$$f(y_1, y_2, \dots, y_p / x = i) = N(Y; \mu, \Sigma)), i = 1,2, \dots, p \tag{1}$$

Let m_i be the number of observations where $x = i, i = 1,2$ and (y_i, x_i) are selected at random using sampling. Alternately, X might be fixed in a way that m_i inspections would be sampled for $X = i$. For each category, sample statistics are computed. The estimates for the sample men $\overline{Y_i}$ and the sample covariance matrix S_i are μ and $\Sigma_i, i = 1,2$ respectively.
Let,

$$V = b_1 Y_1 + b_2 Y_2 + \cdots + b_h Y_h$$
$$S = \left(\frac{(m_1 m_2)}{m_1 + m_2}\right)(\overline{Y_1} - \overline{Y_2})(\overline{Y_1} - \overline{Y_2})'$$

The pooled within the covariance matrix is

$$U_b = \left(\frac{(m_1 m_2)}{m_1 + m_2}\right)\left\{\frac{b'(\overline{Y_1} - \overline{Y_2})(\overline{Y_1} - \overline{Y_2})'b}{b'Sb}\right\} \tag{2}$$

This is the generalized eigenvalue problem given by

$$Ab = cSb;\ A = \left(\frac{(m_1 m_2)}{m_1 + m_2}\right)(\overline{Y_1} - \overline{Y_2})(\overline{Y_1} - \overline{Y_2})'$$

The clarification b is proportional to $S^{-1}(\overline{Y_1} - \overline{Y_2})$. so the sample discriminant variable, $V = (\overline{Y_1} - \overline{Y_2})S^{-1}Y$. It is the linear combination of the original observation which has the largest ratio of the between-groups to the within-group variation.

When X has p categories, then p groups (g_1, g_2… g_p) have been established for the p group situations. Let m_i be the number of clarifications in the i[th] group, $m = \sum_i^p m_i$. The sample mean and covariance matrix is given by $\overline{Y_i}$ and S_i. The matrix U represents the pooled within-group covariance.

$$U = \frac{1}{(m-p)} \sum_{i=1}^{p} (m_i - 1) S_i.$$

The between-group covariance matrix is

$$A = \frac{1}{p-1} \sum_i^p m_i \, (\overline{Y_i} - \bar{Y})(\overline{Y_i} - \bar{Y})' \tag{3}$$

The traditional method relies heavily on the sample mean vector and covariance matrix, which are vulnerable to out-of-the-ordinary data. Additionally, when used to datasets with smashed model assumptions, the LDA model may yield erroneous outcomes. A robust approach can be used to solve this problem.

Robust Linear Discriminant Analysis (RLDA)

RLDA has been developed as a modified version of traditional LDA, especially for handling non-ideal data such as outliers, high dimensions, and multicollinearity, Sharipah et. al. [17]. In robust methods, classical mean vectors and covariance matrices are replaced by robust counterparts, Peter J. Rousseeuw, Mia Hubert [14], and Muthukrishnan et. al. [10]. The robust estimators used here in the robust linear discriminant analysis are MCD and MRCD.

Minimum Covariance Determinant Estimator (MCD)

Rousseeuw [12] developed the Minimum Covariance Determinant (MCD) Estimator to estimate the mean vector and covariance matrix as well as to identify outliers. The lowest determinant-containing subset of h observations with respect to their covariance matrix is sought after. The location estimate is the mean value of that subgroup according to this estimator, and the scatter estimate is a multiple of its scatter matrix.

$$M_X(H) = h^{-1} X_H' I_h \tag{4}$$

$$S_X(H) = (h-1)^{-1} (X_H - M_X(H))'(X_H - M_X(H)) \tag{5}$$

After that, the MCD method seeks to minimize the determinant of $S_X(H)$ for all $H \in \mathcal{H}_h$.

$$H_{MCD} = \underset{H \in \mathcal{H}_h}{argmin} \left(det(S_X(H))^{1/p} \right) \tag{6}$$

For statistical considerations, eqn (3) takes the p^{th} root of the determinant. The geometric average of its eigenvalues is the p^{th} root of the determinant of the scatter matrix. It is referred to as the standardized generalized variance by Sen Gupta (1987).

The mean of the h-subset is used to define the MCD estimate of location M_{MCD}, while the MCD scatter estimate is expressed as a multiple of the sample scatter matrix, and is given by

$$M_{MCD} = M_X(H_{MCD}) \tag{7}$$
$$S_{MCD} = C_\alpha S_X(H_{MCD}) \tag{8}$$

where C_α is a consistency factor that is based on the trimming percentage = $(n-h)/n$ and is similar to the one provided by Croux and Haesbroeck [3]. Its fundamental flaw is that it gives unreliable results when the dimension is greater than the size of the subset. In high dimensions, it is necessary to modify MCD, as the existing MCD methods are slow and less robust in that situation.

Minimum Regularized Covariance Determinant Estimator (MRCD)

The MRCD estimator was proposed by Boudt et. al. [1]. To guarantee that the MRCD scatter estimator is scale equivariant and location unvarying, as is common in the literature, first, standardize the variables. The use of a trustworthy univariate location and scale estimate is required for standardization. For this, the median of each subset is calculated and placed in a location vector called m_x. Additionally, each variable's scale using the Qn estimator of Rousseeuw and Croux (1993) is calculated, then insert these scales into d_x, the diagonal matrix. The standardized observation is given by

$$Z_i = d_x^{-1}(x_i - m_x) \tag{9}$$

The regularized scatter matrix of the standardized observation is

$$\mathcal{S}(H) = \rho T + (1-\rho)C_\alpha S_Z(H)$$

where $S_Z(H)$ is defined in (5), however, in the case of Z, c is the same consistency parameter as in (8).

Let A be the diagonal matrix containing eigenvalues of T, and the orthogonal matrix Q contains the relevant eigenvectors. Utilizing the spectral decomposition $T = QAQ'$ will be practical.

Now,

$$\mathcal{S}(H) = QA^{1/2}[\rho I + (1-\rho)C_\alpha S_W(H)]AA^{1/2}Q' \tag{10}$$

where W is the $n \times p$ matrix consisting the transformed standardized observations

$$w_i = A^{-1/2}Q'Z_i, \text{ and}$$

$$S_W(H) = A^{-1/2}Q'S_Z QA^{-1/2}$$

The MRCD subset is given by

$$H_{MRCD} = \underset{H\in\mathcal{H}_h}{argmin}\left(det(\rho I + (1-\rho)C_\alpha S_W(H))^{1/p}\right) \tag{11}$$

The MRCD location and scatter estimations of the initial data matrix X are defined as follows

$$M_{MRCD} = m_X + d_x M_Z(H_{MRCD})$$

$$S_{MRCD} = C_\alpha S_X(H_{MCD})$$

III. Experimental Results

Table 1: *Apparent Error Rate under Classical and Robust Methods*

Dataset	LDA	RLDA	
		MCD	MRCD
Hemophilia	0.146	0.146	0.133
	(0.150)	(0.146)	(0.125)
Anorexia	0.513	0.486	0.388
	(0.528)	(0.457)	(0.371)

(.) without outliers

Table 2. *Classification Matrix of Hemophilia Data under Classical and Robust Methods*

Methods	LDA	RLDA	
		MCD	MRCD
With outliers	$\begin{pmatrix} 38 & 7 \\ 4 & 26 \end{pmatrix}$	$\begin{pmatrix} 38 & 7 \\ 4 & 26 \end{pmatrix}$	$\begin{pmatrix} 39 & 6 \\ 4 & 26 \end{pmatrix}$
Without outliers	$\begin{pmatrix} 37 & 7 \\ 4 & 25 \end{pmatrix}$	$\begin{pmatrix} 38 & 7 \\ 4 & 26 \end{pmatrix}$	$\begin{pmatrix} 38 & 76 \\ 3 & 26 \end{pmatrix}$

Table 3. *Classification Matrix of Anorexia Data under Classical and Robust Methods*

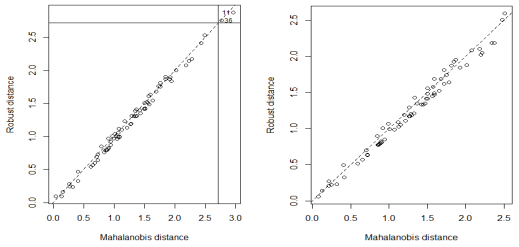
Methods	LDA	RLDA	
		MCD	MRCD
With outlier	$\begin{pmatrix} 11 & 10 & 8 \\ 9 & 17 & 0 \\ 6 & 4 & 7 \end{pmatrix}$	$\begin{pmatrix} 16 & 5 & 8 \\ 16 & 8 & 2 \\ 0 & 4 & 13 \end{pmatrix}$	$\begin{pmatrix} 15 & 5 & 8 \\ 0 & 16 & 4 \\ 0 & 4 & 13 \end{pmatrix}$
Without outlier	$\begin{pmatrix} 10 & 2 & 6 \\ 9 & 17 & 0 \\ 6 & 4 & 6 \end{pmatrix}$	$\begin{pmatrix} 16 & 5 & 7 \\ 14 & 10 & 2 \\ 0 & 4 & 12 \end{pmatrix}$	$\begin{pmatrix} 15 & 6 & 7 \\ 5 & 17 & 4 \\ 0 & 4 & 12 \end{pmatrix}$

This section examined the effectiveness of traditional and robust methods of discriminant analysis techniques in terms of classification problems. For the study, the two actual data sets were taken into account. The first one is the hemophilia data set from the R package named rrcov. There are two assessed factors in the hemophilia data. AHF action and AHV antigen upon 75 women, divided into two groups, namely, compulsory carriers, which includes 45 data points, and the normal group, which includes 30 data points.

The second one is the anorexia data on weight change from the R package named *MASS,* and is divided into three groups, each of which has two variables and a set of 72 occurrences; Information on young female anorexic patients' weight changes. Prewt (patient weight prior to study times) and Postwt (patient weight following study times) are the two variables used to classify the three groups into Cognitive Behavioral Treatment (CBT), Control (Cont), and FT family treatment (FT).

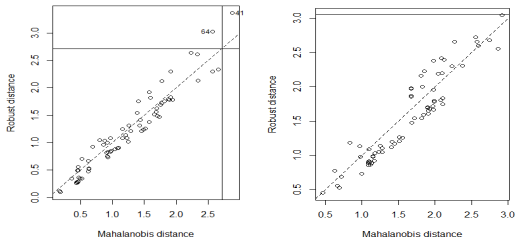
These data sets experienced classification analysis using classic LDA and alternative RLDA algorithms under with/without outliers. Distance-distance plots were used to identify the anomalies (Figure 1, Figure 2).). On the basis of their Classification matrix and Apparent Error Rate (AER), the classification criteria are assessed. The classification matrix is just a table where the rows represent the dependent categories that were observed and the columns represent the expected dependent categories. All examples will fall on the diagonal if the prediction is perfect. The proportion of correctly classified cases is represented by the diagonal cases. The results achieved under various methods are concluded in the form of Apparent Error Rate (Table 1) and classification matrix (Table 2 and Table 3). Robust classification procedure using MRCD estimator gives less Apparent Error Rate and more classification accuracy when compared with other classification procedures.

The result reveals that robust procedures provide better results when compared with the traditional method, Linear Discriminant Analysis. Further, it is observed that MRCD estimator based RLDA outperforms over MCD estimator.



with outliers without outliers

Figure 1*: Distance-Distance Plot (Haemophilia dataset)*



with outliers without outliers

Figure 2: *Distance-Distance Plot (Anorexia dataset)*

IV. Conclusion

Classification analysis is one of the key concepts in the context of statistical learning. This study explores classification analysis using traditional and various approaches of robust linear discriminant analysis. Conventional methods should work reasonably well if certain assumptions are true, however, they may not be trustworthy if one or more of these assumptions are erroneous. Both sample mean vector and covariance matrix are extremely susceptible to anomalies. As a result, when the data contains anomalies, the traditional LDA fails to generate reliable results. For non-normal conditions, a robust alternative is required to improve accuracy even when the data somewhat depart from the model assumptions. When robust estimators such as MCD and MRCD are used in LDA, the analysis performs well compared to traditional methods. Robust methods perform well even when the model assumptions are not met. The study came to the conclusion that robust classification using MRCD estimators offers greater accuracy, followed by other methods. The study could be expanded by using appropriate robust estimators in RLDA to further improve accuracy.

References

[1] Boudt K., Rousseeuw P.J. and Vanduffel S. (2020). The minimum regularized covariance determinant estimator. *Statistics and Computing*, 30:113–128.

[2] Choral C. Y. and Rousseeuw P. J. (1992). Integrating a high-breakdown option into discriminant analysis in exploration geochemistry, *Journal of Geochemical Exploration*, 43:191-203.

[3] Croux C. and Dehon C. (2001). Robust linear discriminant analysis using S-estimators, *Canadian Journal of Statistics*, 29:473-493.

[4] Elizabeth A. M. and Andres M. A. (2016). Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals. *Computational Statistics and Data Analysis*, *70*:67 – 87.

[5] Erceg-Hurn D. M., Wilcox R. R. and Keselman H. J. (2013). Robust statistical estimation. *In T. D. Little (Ed). The Oxford Handbook of Quantitative Methods: Foundations, Oxford University Press*, 1: 388-406.

[6] Fisher R. A. (1936). The use of multiple Measurements in Taxonomic problem. *The Annals of Eugenics*, 7:179-188.

[7] Hastie T. Tibshirani R. and Friedman J. H. (2009). Elements of statistical learning. *Springer, New York*.

[8] He X. and Fung W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis*, 72:151-162.

[9] Hubert M. and Driessen K. (2004). Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, 45:301-320.

[10] Muthukrishnan R. and Udaya Prakash N. (2019). Performance of Classification Techniques along with Support Vector Machine. *International Journal of Innovative Technology and Exploring Engineering*, 9:2278-3075.

[11] Rao C. R. (1948). Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation. *Proceedings of the Cambridge Philosophical Society*, 44:50-57.

[12] Rousseeuw P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880.

[13] Rousseeuw P. J. Croux C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88:1273–1283.

[14] Rousseeuw P. J. and Mia Hubert (2017). Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery*, .8, 10.1002/widm.1236.

[15] Sajobi T. T., Lix L. M., Dansu B. M., Laverty W. and Lix L. (2012). Discriminant Analysis for Repeated Measures Data: Effects of Mean and Covariance Misspecification on Bias and Error in Discriminant Function Coefficients. *Journal of Modern Applied Statistical Methods*, 56:2782-2794.

[16] SenGupta A. (1987). Tests for standardized generalized variances of multivariate normal populations of possibly different dimensions. *Journal of Multivariate Analysis*, 23:209–219.

[17] Sharipah Soaad S. Y., Lim Y. F., Hazlina Ali and Zurni Omar. (2016). Robust Linear Discriminant Analysis. *Journal of Mathematics and Statistics*, 12:312-316.

[18] Todorov V. (2007). Robust selection of variables in linear discriminant analysis. *Statistical Methods and Applications*, 5:395-407.

DESIGN OF MULTIPLE DEPENDENT STATE SAMPLING PLAN USING ZECH DISTRIBUTION WITH APPLICATION TO REAL LIFE DATA

[1] Sunday J. Adeyeye, [2] Ademola J. Adewara, [3] Rao S. Gadde, [4] Samuel K. Adekeye, [5] Adedayo F. Adedotun, [6] Lawrence O. Aako

[1] Department of Statistics, University of Lagos, Nigeria.

[2] Distance Learning Institute, University of Lagos, Nigeria.

[3] Department of Statistics, The University of Dodoma, P.O. Box 259, Dodoma, Tanzania.

[4] Department of Mathematics and Statistics, Redeemer's University, Ede, Osun State, Nigeria.

[5] Department of Mathematics, Covenant University, Ota-Ogun State, Nigeria.

[6] Department of Mathematics and Statistics, The Federal Polytechnic, Ilaro, Ogun State

Emails: [1*] jadeyeyesunday@yahoo.com, [2] jadewara@unilag.edu.ng, [3] gaddesrao@gmail.com, [4] adekeyek@run.edu.ng, [5] adedayo.adedotun@covenantuniversity.edu.ng, [6] olubisi.aako@federalpolyilaro.edu.ng

Abstract

In this work, a multiple dependent state sampling plan, which is an inspection procedure that determines whether an attribute is conforming or non-conforming to a specific requirement, in which the decision criterion for each lot dictates whether to accept the lot; reject the lot; or conditionally accept or reject the lot based on the disposition of future related lots, is introduced. This plan has some advantages over other acceptance sampling plans, like increased efficiency, improved ability to discriminate between acceptable and non – acceptable lots or batches, flexibility in designing the sampling process, and improved cost-effectiveness. To reject a lot, the plan made use of the properties of the sampled current and preceding lots. The study aims to reduce the average sampling number by using a non-linear optimization problem that is subjected to some constraints. With regards to a life test that is truncated in time, the product's median life was used for the proposed sampling plan assuming that the lifetime of the product follows Zech distribution. The usage of median life was necessitated because Zech distribution is an asymmetric distribution with longer tail to the right. Two points on the operating characteristic curve were used for the proposed sampling plan and the following parameters were found; number of preceding lots which is required for deciding if the current lot should be accepted or rejected, the size of the sample, rejection number, and acceptance number. For different shape parameters, we constructed tables for various combinations of consumers' and producers' risks. A real example was provided which showed that a multiple dependent state sampling plan is a good sampling plan for fitting the datasets. Comparing the proposed plan with a single sampling plan, the results reveal that the proposed plan is more effective at securing the consumer and the producer with less inspection. The approach introduced in this study provides an ample opportunity for the manufacturers to reduce the cost and time of inspection by having the sample size reduced without compromising the decision-making accuracy. By implementing the findings of this study, the consumers are confident that their hard-earned money is not used to purchase sub-standard goods.

Keywords: Acceptance Sampling, Zech Distribution, MDSSP, Single Sampling Plan, Operating Characteristic Curve, Average Sampling Number.

1. Introduction

There are growing concerns all over the world about the quality of products in the market this day. It is no longer a hearsay that there is a sharp decline in the quality of some essential commodities from what it used to be some decades ago. The reasons for this could be attributed to a surge in population leading to increase in demand which forces the producers not to take necessary precautions in ascertaining the quality of their products before being rolled to the markets. Some producers are too mindful of the profits thereby doing everything possible to cut the cost of production to the barest minimum. In all of this, the consumers are at the receiving end. In recent times, there are lots of building collapse all over the world which claimed so many lives. Many of these buildings failed integrity tests as a result of usage of poor quality materials when erecting those structures. Several patients have been sent to the early graves due to drugs/vaccines of poor qualities administered on them. It is therefore imperative that products of good qualities are produced by the producers while products of poor qualities are rejected by the consumers. To do this, the two areas of quality control used in monitoring production processes are acceptance sampling and quality control. Acceptance sampling helps to decide whether a lot will be accepted or rejected at the minimum inspection cost in terms of time and money while satisfying not only the producer's but also the consumer's risks at the same time. Producer's risk denoted as α is the risk incurred by the producer when good lots are rejected by the consumer while the consumer's risk, β is the risk incurred by the consumer for accepting a bad lot from the producer.

Limiting quality control (LQL) is the quality level attached to consumer's risk while acceptable quality control (AQL) is the quality level attached to producer's risk. The field of acceptance sampling has gained a lot of patronage in the literature. There are several types of acceptance sampling, of which, the simplest and easiest for practical implementation is the single sampling acceptance plan. Others are double acceptance plan, group acceptance plan, sequential acceptance plan, multiple dependent state sampling plan, modified multiple dependent plan, adjusted multiple dependent plan, rank sampling plan, to mention but a few. Kantam et al. [1] proposed an economic reliability test plan using Log –logistic distribution, Tzong – Ru and Shou – Jye Wu [2] developed acceptance sampling based on truncated life tests for generalized Rayleigh distribution, Syed et al. [3] proposed Mean ranked acceptance sampling plan under exponential distribution, Wenhao and Shangli [4] developed acceptance sampling plans based on truncated life tests for Gompertz distribution, Rao et al. [5] developed acceptance sampling plan for Marshall – Olkin extended Lomax distribution, Al-Omari et al. [6] developed a single sampling plan for the three – parameter Lindley distribution. Al – Omari [7] considered acceptance sampling plans based on truncated life tests for Sushila distribution, Mahdy and Basma [8] proposed double acceptance sampling plan using new distributions, Ramasamy and Sutharani [9] designed double acceptance sampling plans based on truncated life tests in Rayleigh distribution using minimum angle method. For an exponentiated Frechet distribution with known shape parameters, a double acceptance sampling plan was devised by Babu et al. [10]. Wortham and Baker [11] proposed a multiple deferred state sampling inspection. It is a modification to the chain sampling plan and also intends to supplement existing plans like dependent stage sampling, chain sampling, exponential smoothed sampling and fixed deferred state sampling. It is perhaps, very useful where the lots are serially submitted for inspection. The main aim of acceptance sampling is to reduce the sample size and this can be achieved by implementing multiple dependent state sampling plan. The reason for this is because the results of samples drawn by the experimenter from both the current and successive lots is used to make decision concerning the disposition of the current lot. Several works have been done on multiple deferred state sampling plans (MDSSP), among whom are Davood et al. [12] who designed the Bayesian multiple dependent (deferred) sampling plan based on the process capability index. Yan et al. [13] developed a multiple dependent state sampling plan based on the coefficient of variation. The design of multiple deferred state sampling plans for exponentiated half-logistic distribution was proposed by Rao et al. [14], a multiple dependent state sampling plan in a Fuzzy environment was developed by Afshari et al. [15]. Balamulari and Jun. [16] proposed a multiple deferred state sampling plans for lot acceptance based on measurement data. Aslam et al. [17] developed plans for sampling several dependent state variables that take process loss into account, Balamurali et al. [18] designed multiple deferred state sampling plan for generalized inverted exponential distribution. Recently, some scholars extended the

multiple dependent state sampling plan, among whom are: Nadi et al. [19] developed a group multiple dependent state sampling plan using truncated life tests for the Weibull distribution, Aslam et al. [20] proposed generalized multiple dependent state sampling state sampling plans for various life distributions, Cheng et al. [21] proposed adjustable variables multiple dependent state sampling plan based on a process capability index, Aslam et al. [22] developed modified multiple dependent state sampling plan. In this work, we developed a multiple dependent state sampling plan using the median life of the product when the test is truncated at a pre – determined time t, assuming that the lifetime of the product follows Zech distribution. Similar approach can be seen in Rao et al. [23]. To showcase the performance of the proposed plan on Zech distribution, a comparison was made between multiple dependent state sampling plans and single sampling plan to reveal which is better. The MDSSP's operating procedure for Zech distribution is detailed in section 4.

2. Zech Distribution

Zech distribution proposed by Adeyeye, et al. [24] is a heavy – tailed, three – parameter distribution derived by finding the inverse of Gompertz Inverse-Exponential distribution. The cumulative distribution function (cdf) of Zech distribution is given as follows:

$$G(t) = e^{\frac{\gamma}{\delta}\left\{1-\left[1-e^{-\theta t}\right]^{-\delta}\right\}} \;;\; t > 0, \gamma > 0, \delta > 0, \theta > 0 \tag{1}$$

Its probability density function (pdf) is given as follows:

$$g(t) = \gamma\theta e^{-\theta t}\left[1-e^{-\theta t}\right]^{-\delta-1} e^{\frac{\gamma}{\delta}\left\{1-\left[1-e^{-\theta t}\right]^{-\delta}\right\}};\quad t > 0, \gamma > 0, \delta > 0, \theta > 0 \tag{2}$$

The shape parameters are γ and δ while the scale parameter is θ. Assuming γ and δ are known, the cdf depends on θx and the qth quantile of the products.

The qth quantile of a product's lifetime which follows Zech distribution is given by (3)

$$t_q = -\frac{1}{\theta}\left\{\ln\left[1-\left(1-\frac{\delta}{\gamma}ln\, q\right)^{-\frac{1}{\delta}}\right]\right\} = \frac{\xi_q}{\theta} \;\; where \;\; \xi_q = -\left\{\ln\left[1-\left(1-\frac{\delta}{\gamma}ln\, q\right)^{-\frac{1}{\delta}}\right]\right\} \tag{3}$$

The median is derived when $q = 0.5$ and is given by (4)

$$x_{0.5} = -\frac{1}{\theta}\left\{\ln\left[1-\left(1-\frac{\delta}{\gamma}ln(0.5)\right)^{-\frac{1}{\delta}}\right]\right\} \tag{4}$$

Before the experiment time t_0, the probability of failure of products under the Zech distribution is given by

$$p = e^{\frac{\gamma}{\delta}\left\{1-\left[1-e^{-\theta t_0}\right]^{-\delta}\right\}} \tag{5}$$

The experiment is stopped at the time t_q^0 indicated by $t_0 = at_q^0$, where t_0 is the termination time. The scale parameter θ can be expressed as

$$\theta = \frac{\xi_q}{t_q} \tag{6}$$

By substituting θ in equation (5), the probability of failure of the item is obtained as follows

$$p = exp\left(\frac{\gamma}{\delta}\left\{1-\left[1-exp\left(\frac{-a\xi_q}{(t_q/t_q^0)}\right)\right]^{-\delta}\right\}\right) \tag{7}$$

Then in expanded form, equation (7) becomes

$$p = ex\, p\left(\frac{\gamma}{\delta}\left\{1-\left[1-exp\left(\left(aln\left\{\left[1-\left(1-\frac{\delta}{\gamma}lnq\right)^{-\frac{1}{\delta}}\right]\right\}\right)^{-\delta}\left(\frac{t_q}{t_q^0}\right)^{\delta}\right)\right]\right\}\right) \tag{8}$$

The quantile ratio is $\left(\frac{t_q}{t_q^0}\right)$, when it is greater than one, the failure probability in (8) is regarded as acceptable quality level (AQL) (p_1). It is regarded as limiting quality level (LQL) (p_2) when it is equal to 1.

3. Multiple Dependent State Sampling Plan for Zech Distribution

This section is sub–divided into two:

(i) The procedures of operation of MDSSP on Zech distribution.
(ii) Determination of Average Sampling Number.

3.1 Procedures of Operation: this is carried out as itemized in the steps below:

Step 1: A random sample, of size n units will be selected from the current lot. Having fixed a specified time t_0, all units are subjected to a life test at the same time.

Step 2: If d is at most c_1, the present lot is accepted and rejected if d is greater than c_2. The tests ends here but if none of these conditions is met, go to step 3.

Step 3: If d surpasses c_1 and is not greater than c_2 the decision to accept the current lot is made , i.e. $c_1 < d \leq c_2$ provided the succeeding m (preceding m) lots will be accepted with condition $d \leq c_1$.

The four parameters characterizing the proposed MDSSP are: n, m, c_1 and c_2 which represent the sample size, number of preceding lots the experimenter needs to make a good decision, maximum number of failed items for unrestricted (unconditional) acceptance and maximum number of failed items for restricted (conditional) acceptance respectively.

It is important to note that single sampling plan (SSP) can be generalized to form an attribute MDSSP; this is so because it reduces to SSP when either $m \to \infty$ or $c_1 = c_2 = c$

According to Rao [3], the operating characteristic (OC) function of the proposed plan is denoted by

$$P_a(p) = p(d \leq c_1) + p(c_1 < d \leq c_2)\big(p(d \leq c_1)\big)^m \tag{9}$$

To accept a lot, equation (10) provides the probability of failure p if a binomial distribution is considered.

$$P_a(p) = \sum_{d=0}^{c_1} \binom{n}{d} p^d (1-p)^{n-d} + \sum_{d=c_1+1}^{c_2} \binom{n}{d} p^d (1-p)^{n-d} \left(\sum_{d=0}^{c_1} \binom{n}{d} p^d (1-p)^{n-d} \right)^m \tag{10}$$

3.2 Determination of ASN: The main aim of any sampling plan is to reduce the average sampling number (ASN). This will of course, reduce the inspection time and cost. It will also be of immense benefit to the producers in case of destructive sampling. In this study, we seek to reduce the ASN of the suggested MDSSP for the Zech distribution under truncated life testing. This is achieved by using a non-linear optimization problem in which the objective function is the minimization of ASN at p subject to some constraints. The optimization problem is as follows:

Minimize Average Sample Number, ASN $(P_1) = n$

Subject to

$$P_a(P_1) \geq 1 - \alpha, \tag{11}$$

$$P_a(P_2) \leq \beta \tag{12}$$

$$n > 1, m > 1, c_2 > c_1 \geq 0,$$

Failure at the producer's risk has a chance of p_1, while failure at the consumer's risk has a probability of p_2. The ratio $\frac{t_q}{t_q^0}$ is known as the quality level or true life quantile ratio, and it aids the producer in ensuring that the quality of his products is good and acceptable. Equations (13) and (14) are respectively, the probabilities of the acceptance of the lot under the modified sampling scheme at acceptable quality level (AQL) and limiting quality level (LQL).

$$P_a(p_1) = \sum_{d=0}^{c_1} \binom{n}{d} p_1^d (1-p_1)^{n-d} + \sum_{d=c_1+1}^{c_2} \binom{n}{d} p_1^d (1-p_1)^{n-d} \left(\sum_{d=0}^{c_1} \binom{n}{d} p_1^d (1-p_1)^{n-d} \right)^m \tag{13}$$

$$P_a(p_2) = \sum_{d=0}^{c_1} \binom{n}{d} p_2^d (1-p_2)^{n-d} + \sum_{d=c_1+1}^{c_2} \binom{n}{d} p_2^d (1-p_2)^{n-d} \left(\sum_{d=0}^{c_1} \binom{n}{d} p_2^d (1-p_2)^{n-d} \right)^m \tag{14}$$

The median ratio, $\frac{t_q}{t_q^0}$ whose values are 2, 4, 6, 8 and 10 is considered at the risk of the producer. This will make it more likely that people will accept the high quality product while the mean ratio $\frac{t_q}{t_q^0} = 1$ is considered at the consumer' risk to ensure that the product of poor quality is rejected.

4. Simulation Studies

The optimal parameters of the proposed plan for Zech distribution with $\gamma = 0.5, \delta = 0.5$; $\gamma = 1.0, \delta = 1.0$ and $\gamma = 1.5, \delta = 1.5$ are presented in the following tables (Table 1-3). Values assumed for consumer's risks $\beta = 0.25, 0.10, 0.05, 0.01$ while the producer's risk was kept at $\alpha = 0.05$ at 50th percentile. The values considered for termination ratio are *a*= 0.5, 0.7 and 1.0

Results of simulation studies observed from Tables 1-3 when parametric combinations are fixed are as follows:

i. Inverse relationship is observed between sample size and consumers risk. In all the Tables 1-3 sample size increases when consumers risk decreases.
ii. As the termination ratio a rises from 0.5 to 1.0, the sample size drops.
iii. Probability of lot acceptance increases along with quantile ratio. As $\frac{t_q}{t_q^0}$ approaches 10, probability of lot acceptance also increases and approximates to almost 1.

Table 1: *The developed MDSSP's ideal (optimal) parameters for the Zech distribution with* $\gamma = 0.5, \delta = 0.5$

β	$\frac{t_q}{t_q^0}$	a=0.5					a=0.7					a=1.0				
		n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$
0.25	2	18	3	7	2	0.9551	16	4	14	2	0.9508	16	6	8	2	0.9563
	4	5	0	4	2	0.9531	7	1	2	1	0.9870	5	1	2	1	0.9758
	6	5	0	4	2	0.9921	3	0	2	3	0.9798	3	0	2	1	0.9708
	8	5	0	4	2	0.9984	3	0	2	3	0.9945	3	0	2	1	0.9902
	10	5	0	4	2	0.9996	3	0	2	3	0.9983	3	0	2	1	0.9963
0.10	2	33	5	8	1	0.9547	29	7	17	2	0.9534	26	9	19	2	0.9513
	4	13	1	2	1	0.9868	9	1	8	2	0.9751	7	1	3	1	0.9574
	6	8	0	1	1	0.9859	5	0	4	2	0.9642	7	1	3	1	0.9957
	8	8	0	1	1	0.9970	5	0	4	2	0.9900	4	0	1	1	0.9773
	10	8	0	1	1	0.9993	5	0	4	2	0.9969	4	0	1	1	0.9913
0.05	2	41	6	10	1	0.9528	38	9	19	2	0.9542	34	11	18	1	0.9521
	4	16	1	3	1	0.9880	11	1	10	2	0.9522	11	2	5	1	0.9820
	6	10	0	1	1	0.9785	7	0	2	1	0.9638	8	1	3	1	0.9929
	8	10	0	1	1	0.9954	7	0	2	1	0.9900	5	0	2	1	0.9742
	10	10	0	1	1	0.9989	7	0	2	1	0.9969	5	0	2	1	0.9901
0.01	2	64	9	16	1	0.9510	55	12	18	1	0.9505	50	16	22	1	0.9525
	4	21	1	11	2	0.9528	19	2	4	1	0.9782	17	3	8	1	0.9824
	6	15	0	3	1	0.9678	15	1	4	1	0.9950	11	1	5	1	0.9805
	8	15	0	3	1	0.9930	10	0	2	1	0.9804	7	0	2	1	0.9524
	10	15	0	3	1	0.9983	10	0	2	1	0.9938	7	0	2	1	0.9813

Table 2: *The developed MDSSP's ideal (optimal) parameters for the Zech distribution with* $\gamma = 1.0, \delta = 1.0$

β	$\frac{t_q}{t_q^0}$	a=0.5					a=0.7					a=1.0				
		n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$
0.25	2	8	0	1	2	0.9581	8	1	2	2	0.9736	7	2	3	2	0.9661
	4	8	0	1	2	1.0000	4	0	1	2	0.9997	3	0	2	1	0.9971
	6	8	0	1	2	1.0000	4	0	1	2	1.0000	3	0	2	1	1.0000
	8	8	0	1	2	1.0000	4	0	1	2	1.0000	3	0	2	1	1.0000
	10	8	0	1	2	1.0000	4	0	1	2	1.0000	3	0	2	1	1.0000
0.10	2	22	1	5	2	0.9928	11	1	4	2	0.9529	12	3	5	1	0.9755
	4	13	0	10	2	1.0000	7	0	1	1	0.9995	4	0	1	1	0.9931

	6	13	0	10	2	1.0000	7	0	1	1	1.0000	4	0	1	1	0.9999
	8	13	0	10	2	1.0000	7	0	1	1	1.0000	4	0	1	1	1.0000
	10	13	0	10	2	1.0000	7	0	1	1	1.0000	4	0	1	1	1.0000
0.05	2	27	1	2	1	0.9804	18	2	3	1	0.9640	13	3	12	2	0.9502
	4	17	0	1	1	1.0000	8	0	7	2	0.9991	5	0	2	1	0.9921
	6	17	0	1	1	1.0000	8	0	7	2	1.0000	5	0	2	1	0.9999
	8	17	0	1	1	1.0000	8	0	7	2	1.0000	5	0	2	1	1.0000
	10	17	0	1	1	1.0000	8	0	7	2	1.0000	5	0	2	1	1.0000
0.01	2	37	1	3	1	0.9742	24	2	5	1	0.9546	22	5	8		0.9702
	4	25	0	10	2	0.9999	13	0	3	1	0.9988	7	0	2	1	0.9850
	6	25	0	10	2	1.0000	13	0	3	1	1.0000	7	0	2	1	0.9998
	8	25	0	10	2	1.0000	13	0	3	1	1.0000	7	0	2	1	1.0000
	10	25	0	10	2	1.0000	13	0	3	1	1.0000	7	0	2	1	1.0000

Table 3: *The developed MDSSP's ideal parameters for the Zech distribution with* $\gamma = 1.5, \delta = 1.5$.

β	$\frac{t_q}{t_q^0}$	a=0.5					a=0.7					a=1.0				
		n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$
0.25	2	13	0	10	3	0.9988	5	0	4	2	0.2381	5	1	2	1	0.9829
	4	13	0	10	3	1.0000	5	0	4	2	0.2381	3	0	2	1	1.0000
	6	13	0	10	3	1.0000	5	0	4	2	0.2381	3	0	2	1	1.0000
	8	13	0	10	3	1.0000	5	0	4	2	0.2381	3	0	2	1	1.0000
	10	13	0	10	3	1.0000	5	0	4	2	0.2381	3	0	2	1	1.0000
0.10	2	21	0	1	2	0.9974	8	0	1	1	0.0978	7	1	3	1	0.9705
	4	21	0	1	2	1.0000	8	0	1	1	0.0978	4	0	1	1	0.9999
	6	21	0	1	2	1.0000	8	0	1	1	0.0978	4	0	1	1	1.0000
	8	21	0	1	2	1.0000	8	0	1	1	0.0978	4	0	1	1	1.0000
	10	21	0	1	2	1.0000	8	0	1	1	0.0978	4	0	1	1	1.0000
0.05	2	27	0	1	2	0.9957	10	0	1	1	0.0485	8	1	3	1	0.9537
	4	27	0	1	2	1.0000	10	0	1	1	0.0485	5	0	2	1	0.9999
	6	27	0	1	2	1.0000	10	0	1	1	0.0485	5	0	2	1	1.0000
	8	27	0	1	2	1.0000	10	0	1	1	0.0485	5	0	2	1	1.0000
	10	27	0	1	2	1.0000	10	0	1	1	0.0485	5	0	2	1	1.0000
0.01	2	42	0	1	1	0.9938	22	1	3	1	0.0095	14	2	6	1	0.9682
	4	42	0	1	1	1.0000	15	0	1	1	0.0090	7	0	2	1	0.9999
	6	42	0	1	1	1.0000	15	0	1	1	0.0090	7	0	2	1	1.0000
	8	42	0	1	1	1.0000	15	0	1	1	0.0090	7	0	2	1	1.0000
	10	42	0	1	1	1.0000	15	0	1	1	0.0090	7	0	2	1	1.0000

5. Application of the Proposed Sampling Plan to Cancer Data

In this section, the application of multiple dependent state sampling plan to real life data on survival time for 44 patients diagnosed by Head and Neck cancer disease is considered. The data consists of 44 observations. It has been used by Efron [25] for logistic regression and recently analyzed by Sule et al. [26] in Topp Leone Kumaraswamy–G family of distributions. The data are as follows:
12.20, 23.56, 23.74, 25.87, 31.98, 37, 41.35, 47.38, 55.46, 58.36, 63.47, 68.46, 78.26, 74.47, 81.43, 84, 92, 94, 110, 112, 119, 127, 130, 133, 140, 146, 155, 159, 173, 179, 194,195, 209, 249, 281, 319, 339, 432, 469, 519, 633, 725, 817, 1776

The application is done by first checking if the data sets fit the distribution well. Zech distribution, being a heavily right – tailed distribution, the histogram of the data drawn depicts that the data is also positively skewed. The descriptive statistics of the data like minimum,1st Quartile, median, mean, 3rd quartile, maximum, coefficient of skewness, coefficient of kurtosis and standard deviation are shown in the Table 4. Also, Estimates of the parameters, standard errors, negative log-likelihood value, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) together with the goodness of fit tests, viz., Kolmogorov Smirnov Statistic (K –S), and P-value are given in Table 5.

Table 4 : *Summary of head and neck cancer data.*

Min	1st Quartile	Median	Mean	3rd Quartile	Max.	Standard Deviation	Skewness	Kurtosis
12.20	67.21	128.50	223.48	219.00	1776.00	305.4282	3.38382	16.5596

Table 4 above shows that the data is asymmetric with longer tail to the right.

Table 5 : *Goodness of fit tests of Zech distribution using head and neck cancer data.*

Distribution	Estimates	Std Error	-LL	AIC	BIC	KS	P - value
Zech	$\hat{\gamma} = 0.2731$ $\hat{\delta} = 0.7736$ $\hat{\theta} = 0.0023$	0.08364 0.1735 0.0001944	-277.5201	561.0402	566.3928	0.074069	0.9546

Table 5 shows the estimates of parameters, standard errors and goodness of fit tests of Zech distribution for the head and neck cancer data. The parameters are estimated via the method of maximum likelihood. Empirical and theoretical pdf's and cdf's, Quantile – Quantile plot, and Probability-Probability plot are shown in Figure 1 to showcase the goodness of fit of Zech distribution. Judging from the plots in Figure 1, it is not out of place to say that Zech distribution yields a good fit for the head and neck cancer data.

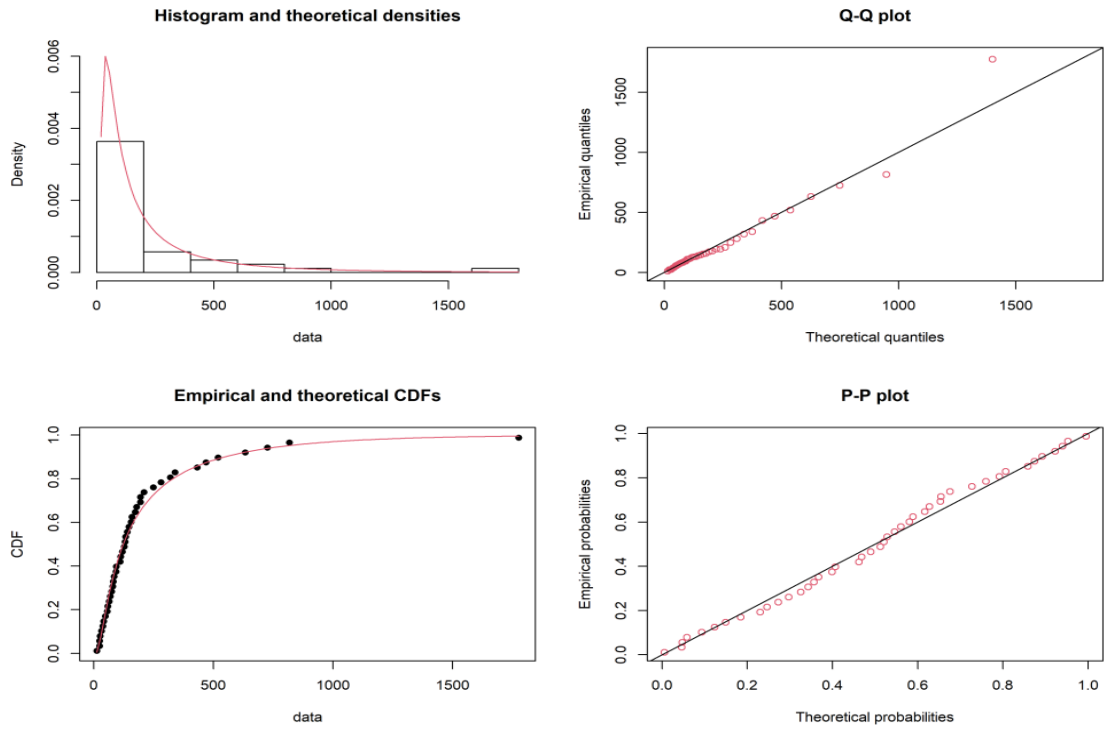


Figure 1: *Empirical and theoretical PDF's, empirical and theoretical CDF'S, Q- Q plots and P-P plots for Zech distribution using Head and Neck cancer data.*

For illustration, the medical practitioner would prefer to employ the developed sampling plan to implement the percentile of the median life of the product where the product's lifetime follows Zech distribution with the shape parameters are $\hat{\gamma} =$ 0.2731 and $\hat{\delta} = 0.7736$. The medical practitioners suggest that given median survival time is 45 days whereas the medical practitioners expected that the median survival time 90 days. The risk of the consumer is 0.05 provided the actual median survival time 45 days while that of the producer is 0.10 given that the true median survival time 90 days. The ideal (optimal) parameters chosen from Table 6 under these restrictions (constraints) are n=13, c_1=3, c_2=6, and m=2 with values of $\hat{\gamma} =$ 0.2731 and $\hat{\delta} = 0.7736$, t_{q_0} =45, α =0.05, β=0.25, t_q / t_{q_0}=2 at a=0.7. The MDS sampling plans are illustrated as follows:

A sample of 13 patients' survival time of Head and Neck cancer disease will be chosen at random from the group of Head and Neck cancer disease patients and their survival time is 45 days. If the survival time before 45 days is 3 patients then a group of the population will be allowed (accepted) and the group of the population will be denied (rejected) if it is greater than 6 patients in a group. In the event that a group of patients' survival time for Head and Neck cancer disease is between 3 and 6, the choice of the group of the population will be delayed until the two preceding groups of the population have been tested. For the purpose of this real world illustration, 7 people in the community with Head and Neck cancer disease survived before the survival time of 45 days. Hence, disregard the survival time of the Head and Neck cancer disease patients in a group of the population. So, doctors could advise the government or general public that the median survival time of people with Head and Neck cancer disease in a particular population group is at an undesirable level.

Table 6: *Optimal parameters of the proposed MDS plan for Zech Distribution with* $\hat{\gamma}$= 0.2731 *and* $\hat{\delta}$=0.7736

β	$\frac{t_q}{t_q^0}$	a=0.5					a=0.7					a=1.0				
		n	c_1	c_2	m	$P_a(p_1)$	n	c_1	c_2	m	$P_a(p_1)$	N	c_1	c_2	m	$P_a(p_1)$
0.25	2	15	2	12	2	0.9683	13	3	6	2	0.9635	14	5	13	2	0.9670
	4	5	0	1	2	0.9915	3	0	2	4	0.9629	5	1	2	1	0.9912
	6	5	0	1	2	0.9997	3	0	2	4	0.9969	3	0	2	1	0.9916
	8	5	0	1	2	1.0000	3	0	2	4	0.9997	3	0	2	1	0.9986
	10	5	0	1	2	1.0000	3	0	2	4	1.0000	3	0	2	1	0.9998
0.10	2	24	3	8	2	0.9620	23	5	15	2	0.9612	21	7	11	2	0.9519
	4	8	0	7	2	0.9827	6	0	2	1	0.9603	7	1	3	1	0.9856
	6	8	0	7	2	0.9994	5	0	1	2	0.9948	4	0	1	1	0.9807
	8	8	0	7	2	1.0000	5	0	1	2	0.9995	4	0	1	1	0.9967
	10	8	0	7	2	1.0000	5	0	1	2	1.0000	4	0	1	1	0.9994
0.05	2	33	4	14	2	0.9607	29	6	16	2	0.9505	29	9	15	1	0.9571
	4	10	0	3	2	0.9739	11	1	2	1	0.9909	8	1	3	1	0.9768
	6	10	0	3	2	0.9990	7	0	1	1	0.9939	5	0	2	1	0.9780
	8	10	0	3	2	1.0000	7	0	1	1	0.9994	5	0	2	1	0.9963
	10	10	0	3	2	1.0000	7	0	1	1	0.9999	5	0	2	1	0.9993
0.01	2	47	5	9	1	0.9522	45	9	19	2	0.9501	43	13	20	1	0.9535
	4	16	0	2	1	0.9663	15	1	3	1	0.9885	14	2	6	1	0.9877
	6	16	0	2	1	0.9987	10	0	1	1	0.9879	7	0	2	1	0.9592
	8	16	0	2	1	0.9999	10	0	1	1	0.9989	7	0	2	1	0.9929
	10	16	0	2	1	1.0000	10	0	1	1	0.9999	7	0	2	1	0.9987

Excerpts from Table 6 are as follows:

i. As consumer's risk decreases, sample size increases

ii. As the termination ratio rises from 0.5 to 1.0, the sample size falls.

iii. As the quantile ratio rises, the likelihood that the lot will be accepted increases. As the quantile ratio approaches 10, probability of lot acceptance also increases and almost approximates to 1

6. Comparative Study

To determine the efficiency of MDSSP over Single sampling plan, a comparison study was made between the two plans when the underlying distribution of data follows Zech distribution. From Table 7, the quantile ratios 2, 4, 6, 8 and 10 were considered for each of the consumer's risks $\beta = 0.25, 0.10, 0.05, 0.01$ while the producer's risk was kept at $\alpha = 0.05$. The sample size n was compared to the probability of acceptance $P_a(p_1)$. The results reveal that the sample size of the single sampling plan is higher than that of the developed MDSSP. For plan ratio 2 under the consumer's risk, $\beta = 0.01$, the parameters for the MDSSP are: $n = 42, c_1 = 0, c_2 = 1$ and $m = 1$. For the single sampling plan, the parameters of the plan are $n = 61$ and $c = 1$ with corresponding probabilities of 0.9938 and 0.9956 for MDSSP and SSP respectively when the termination ratio is 0.5

Also, for plan ratio 2 under the consumer's risk,, $\beta = 0.01$, the plan parameters for MDSSP are: $n = 14, c_1 = 2, c_2 = 6$ and $m = 1$. For the single sampling plan, the plan parameters are $n = 19$ and $c = 4$ with corresponding probabilities of 0.9682 and 0.9569 for MDSSP and SSP respectively when the termination ratio is 1.0

It shows that the proposed MDSSP is more efficient than the existing SSP for Zech distribution. The operating characteristic curve in Figure 2 also emphasizes the MDSSP is more efficient than the existing SSP.

Table 7: *Comparison of optimal parameters of the proposed MDSSP and SSP for Zech distribution with $\gamma = 1.5, \delta = 1.5$.*

		a = 0.5								a = 1.0							
		MDSSP					SSP			MDSSP					SSP		
β	$\frac{t_q}{t_q^0}$	n	c_1	c_2	m	$P_a(p_1)$	n	c	$P_a(p_1)$	n	c_1	c_2	M	$P_a(p_1)$	n	c	$P_a(p_1)$
0.25	2	13	0	10	3	0.9988	13	0	0.9795	5	1	2	1	0.9829	7	2	0.9703
	4	13	0	10	3	1.0000	13	0	1.0000	3	0	2	1	1.0000	2	0	0.9968
	6	13	0	10	3	1.0000	13	0	1.0000	3	0	2	1	1.0000	2	0	1.0000
	8	13	0	10	3	1.0000	13	0	1.0000	3	0	2	1	1.0000	2	0	1.0000
	10	13	0	10	3	1.0000	13	0	1.0000	3	0	2	1	1.0000	2	0	1.0000
0.10	2	21	0	1	2	0.9974	21	0	0.9671	7	1	3	1	0.9705	12	3	0.9694
	4	21	0	1	2	1.0000	21	0	1.0000	4	0	1	1	0.9999	4	0	0.9936
	6	21	0	1	2	1.0000	21	0	1.0000	4	0	1	1	1.0000	4	0	1.0000
	8	21	0	1	2	1.0000	21	0	1.0000	4	0	1	1	1.0000	4	0	1.0000
	10	21	0	1	2	1.0000	21	0	1.0000	4	0	1	1	1.0000	4	0	1.0000
0.05	2	27	0	1	2	0.9957	27	0	0.9578	8	1	3	1	0.9537	13	3	0.9594
	4	27	0	1	2	1.0000	27	0	1.0000	5	0	2	1	0.9999	5	0	0.9921
	6	27	0	1	2	1.0000	27	0	1.0000	5	0	2	1	1.0000	5	0	1.0000
	8	27	0	1	2	1.0000	27	0	1.0000	5	0	2	1	1.0000	5	0	1.0000
	10	27	0	1	2	1.0000	27	0	1.0000	5	0	2	1	1.0000	5	0	1.0000
0.01	2	42	0	1	1	0.9938	61	1	0.9956	14	2	6	1	0.9682	19	4	0.9569
	4	42	0	1	1	1.0000	42	0	1.0000	7	0	2	1	0.9999	7	0	0.9889
	6	42	0	1	1	1.0000	42	0	1.0000	7	0	2	1	1.0000	7	0	0.9999
	8	42	0	1	1	1.0000	42	0	1.0000	7	0	2	1	1.0000	7	0	1.0000
	10	42	0	1	1	1.0000	42	0	1.0000	7	0	2	1	1.0000	7	0	1.0000

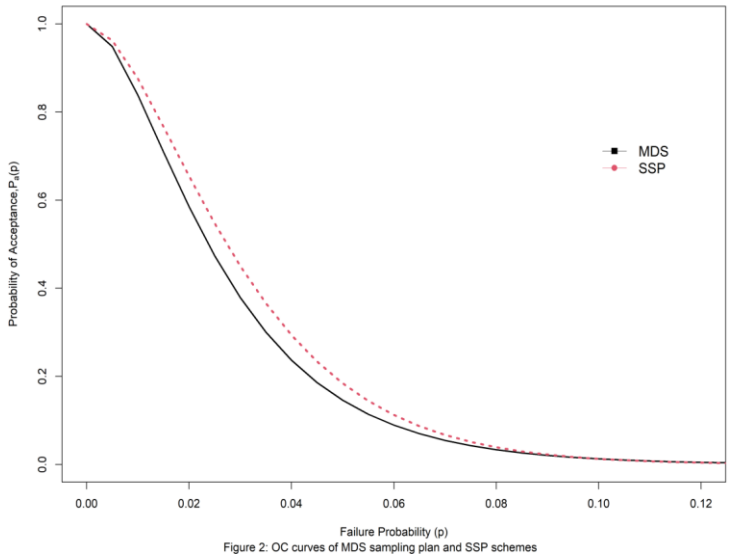


Figure 2: *The operating characteristic curve for multiple dependent state sampling plan and single sampling plan. MDSSP is more accurate than SSP, with steeper slope.*

7. Conclusions

The assumption used in this work to build the multiple deferred state sampling plan is that the product's lifespan will follow Zech distribution when the lifetime tests are truncated. By simultaneously addressing both the producer's and the consumer's risks, the optimal parameters for the proposed MDSSP were established. Application of the proposed sampling plan to real life example shows that it is suitable for the data used. The operating characteristic curve for the proposed sampling plan and single sampling plan (SSP) reveals that MDSSP is more accurate than SSP with steeper slope.

References

[1] R.R.L. Kantam, Rao, G.S. and B. Sriram. "An Economic Reliability Test Plan: Log–logistic distribution". Journal of Applied Statistics. Vol. 33, No. 3, 2006, 291 – 296.

[2] Tzong – Ru and Shou – Jye Wu. "Acceptance sampling based on truncated life tests for generalized Rayleigh distribution". Journal of Applied Statistics. Vol. 33, No. 6, 595 – 600

[3] Syed Adil Hussain, Ishfaq Ahmad, Aamir Saghir, Muhammad Aslam and Ibrahim M. Almanjahie. "Mean ranked acceptance sampling plan under exponential distribution". Ain Shams Engineering Journal 12, 2021, 4125–4131

[4] Wenhao Gui and Shangli Zhang. "Acceptance sampling plans based on truncated life tests for Gompertz distribution". Journal of Industial Mathematics. Volume 2014, Article ID: 391728, 7 pages, http://dx.doi.org/10.1155/2014/391728

[5] Rao, G.S., M.E. Ghitany and R R.L. Kantam. Acceptance sampling plan for Marshall – Olkin extended Lomax distribution. International Journal of Applied Mathematics. Volume 21, No. 2, 2008, 315 – 325.

[6] Al-Omari, A.I., Ciavolino, C., and Al-Nasser, A.D. "Economic design of acceptance sampling plans for truncated life tests using three-parameter Lindley distribution". Journal of Modern Applied Statistical Methods, 18(2), eP2746, 2019, doi: 10.22237/jmasm/1604189220

[7] Amer Ibrahim Al-Omari. "Acceptance sampling plans based on truncated life tests for Sushila distribution". J. Math. Fund. Sci., Vol. 50, No. 1, 72 – 83.

[8] Mervat Mahdy and Basma Hamed. "New distributions in designing double acceptance sampling plan with application". Pak.j.stat.oper.res. Vol. XIV, No.2, 2018, pp: 333 – 346

[9] A. R. Sudamani Ramasamy and R. Sutharani. "Designing double acceptance sampling plans based on truncated life tests in Rayleigh distribution using minimum angle method". American Journal of Mathematics and Statistics, 3(4): 227 – 236, 2013, DOI: 10.5923/j.ajms.20130304.07

[10] M. Sridhar Babu, Rao, G.S. and K. Rosaiah. "Double acceptance sampling plan for exponentiated Frechet distribution with known shape parameters". Mathematical Problems in Engineering, Volume 2021, Article ID: 7308454, 9 pages, 2021, https://doi.org/10.1155/2021/7308454

[11]. Wortham, A. and Maker, R. C. "Multiple deferred state sampling inspection". International Journal of Production Research. 14: 719 – 731, 1976

[12] Davood Shishebori, Mohammad Saber Fallah Nezhad and Sina Seifi."Design of Bayesian MDS sampling plan based on the process capability index". International Journal of Industrial and Manufacturing Engineering. Vol: 11, No: 10, 2017.

[13] Aijun Yan, Sanyang Liu and Xiaojuan Dong."Designing a multiple dependent state sampling plan based on the coefficient of variation". *SpringerPlus 5:1447*, 2016, DOI 10.1186/s40064-016-3087-3

[14] Rao, G.S., K. Rosaiah and C.H. Ramesh Naidu. "Design of multiple deferred state sampling plans for exponentiated half logistic distribution". Cogent Mathematics & Statistics, 7:1, 1857915, 2020, DOI: 10.1080/25742558.2020.1857915.

[15] R. Afshari, and B. Sadeghpour Gildeh. "Designing a multiple deferred state sampling plan in a Fuzzy environment". American Journal of Mathematical Management Sciences, 36:4, 2017, pages 328 – 345.

[16] S. Balamurali and C.H. Jun. ''Multiple dependent state sampling plans for lot acceptance based on measurement data''. European Journal of Operation Research, vol. 180, no. 3, 2007, pp. 1221 – 1230.

[17] Aslam, M., Yen, C.H., Chang, C.H., and Jun, C.H. "Multiple dependent state variable sampling plans with process loss consideration". The International Journal of Advanced Manufacturing Technology, 71 (5 – 8), 013 – 5574 – 9, 1337 – 1343, 2014, https://doi.org/10.1007/s00170

[18] S. Balamurali, P. Jeyadurga and M. Usha. ''Designing of multiple deferred state sampling plan for generalized inverted exponential distribution'', *Sequential Analysis*, vol. 36, 2017, no. 1, pp. 76 – 86.

[19] A. A. Nadi and B.S. Gildeh . ''A group multiple dependent state sampling plan using truncated life tests for the weibull distribution'', Quality Engineeering, vol. 31. No. 4, pp. 1 – 11, 2019.

[20] M. Aslam, G.S. Rao and M. Albassam. ''Time truncated life tests using the generalized multiple dependent state sampling state sampling plans for various life distributions'', Statistical Quality Technologies. ICSA Book Series in Statistics, pp. 153 – 182, 2019, Springer, Cham, Switzerland.

[21] To-Cheng Wang, Ming – Hung Shu and Bi – Min Hsu. "Adjustable variables multiple dependent state sampling plans based on a process capability index". Journal of the Operational Research Society 0:0, pages 1 – 14, 2021.

[22] Muhammed Aslam, P. Jayedurga, S. Balamurali, Muhammed Azam and Ali Al-Marshadi. "Economic determination of modified multiple dependent state sampling plan under some lifetime distributions." Journal of Mathematics, Article ID: 7470196, 13 pages, 2021, https://doi.org/10.1155/2021/7470196

[23] Rao, G.S., Arnold K. Fulment, and Joseohat K. Peter. "Design of multiple dependent state sampling plan application for covid–19 data using exponentiated weibull distribution". Complexity, Vol: 2021, Article ID 2795078, 1-10.

[24]. S. Adeyeye, A. Adewara, E. Akarawak, A. Ogunsanya and A. Jamal. "Zech distribution: Derivation, Properties and Applications". RT&A, No. 2 (68), 2022, Volume 17.

[25] Efron B. "Logistic regression, survival analysis, and the Kaplan – Meier curve". Journal of the American Statistical Association. 1988; 83(402): 414 – 425. https://doi.org/10.1080/01621459.1988.10478612

[26] Sule I, Doguwa S.I., Isah A, and Jibril H.M. "Topp Leone Kumaraswamy-G Family of Distributions with Applications to Cancer Disease Data". Journal of Biostatistics and Epidemiology. 2020; 6(1):37–48.

MULTIBAND COMPACT MICROSTRIP PATCH ANTENNA FOR WIRELESS COMMUNICATION APPLICATIONS

Dr.Shahid Modasiya [1], Dr.Balvant Makwana [2] and Anil Poriya[3]

•

[1]Government Engineering College, Gandhinagar, Gujarat, India.
[2,3]Government Engineering College, Rajkot, Gujarat, India
[1]shahid@gecg28.ac.in, [2] balvantmakwana@iitgn.ac.in

Abstract

In this paper, a multiband compact microstrip patch antenna for different communication frequencies has been presented. The proposed design of the microstrip patch antenna consists of a slotted patch, a quarter-wave feed line, and a ground with a cross-edge slot. The antenna can operate from 2.1 GHz to 3.4 GHz with a bandwidth of 1.3 GHz; this band corresponds to applications such as Mobile WiMax (2110 MHz-2200 MHz, 2300 MHz-2400 MHz, 2500 MHz-2690 MHz), Bluetooth (2400 MHz-2497 MHz), and RFiD. (2400 MHz -2483 MHz). The higher band, 4.7 GHz to 7.4 GHz, covers C-band, WLAN, and sub-6GHz 5G applications and has a gain factor of about 2.15 dB. The antenna is fabricated, and measurements of the radiation pattern and return loss are made. The comparison of observed results with those from simulations reveals excellent symmetry. Furthermore, the 70× 40 mm2 size of the proposed antenna makes it appropriate for use in lower 5G bands.

Keywords: Microstrip antenna, Multi band, patch antenna, RFiD, Sub- 6 GHz, Wi-Max, WLAN.

1. INTRODUCTION

Multi-band microstrip patch antennas (MPAs) have drawn a lot of interest in light of the quick growth of mobile communication systems because to its many benefits, including low profile, lightweight, cheap cost, superior performance, ease of manufacturing, low cost, and multi-band operation. As a result, they have long had a central role in research on antennas and wireless propagation. They have an intrinsic constraint of a small impedance bandwidth (almost 5%) when constructed on thin substrates.

The demand for mobile terminals with a variety of functionalities has surged due to the rapid development of wireless communications. Nowadays, it's practically a must for any Smartphone to integrate a variety of services, such as coverage of several cellular frequencies, the ability to use navigation and positioning, and the availability of Wi-Fi. Thus, in order to support all of these services, mobile terminals must operate at several frequencies. Moreover, portable gadgets must be lightweight, small, cheap, and compact in order to be mobile and flexible.

Antennas are a crucial component of any communication equipment that uses wireless technology and are essential to ensuring that the device complies with operational frequency, size,

weight, and cost requirements. The mobile terminal becomes larger when each service has its own specialized antenna, which also increases the weight and price of the device.

A highly desirable solution to this issue has been the use of a single low-profile antenna that can operate in several bands at the many needed communication frequencies [1]-[4]. The universal mobile telecommunications system (UMTS), long-term evolution (LTE), Bluetooth, Worldwide interoperability for microwave access (WiMAX), and wireless local area network (WLAN) bands can all operate simultaneously on a multi-band microstrip antenna. In recent years, various designs of multi-band patch antennas have been proposed in the literature. Liu, Xiao et al [5]. Presents a novel multiband patch antenna design that can operate at three different frequencies. The antenna is composed of a rectangular patch with two inverted L-shaped slots, a T-shaped feed line, and a ground plane. Islam, Misran et el. [6] presents a compact multiband patch antenna with a novel E-shaped feed line. The antenna operates at three different frequencies, and the bandwidths are increased by using a slotted ground plane. Malhat, Safwat et el. [7] presents a compact multiband patch antenna with a T-shaped feed line. The antenna is designed to operate at four different frequencies, and the bandwidths are enhanced by using a slotted ground plane. Jiang and Xu [8] presents a multiband patch antenna with a slit ring resonator. The antenna is designed to operate at four different frequencies, and the bandwidths are increased by using a meandered ground plane. Islam and Ali [9] present the design and analysis of a dual-band patch antenna for WLAN applications. Abbosh and Bialkowski [10] proposes a multiband patch antenna design that can operate in the frequency bands of 900 MHz, 1.8 GHz, and 2.4 GHz. Das, Bahera and Patnaik [11] proposes a multiband microstrip patch antenna design that can operate in the frequency bands of 2.4 GHz, 5.2 GHz, and 5.8 GHz for wireless communication systems. Vishwakarma and Srivastava [12] presents a multiband stacked patch antenna design that can operate in the frequency bands of 2.4 GHz, 3.5 GHz, and 5.8 GHz for WLAN and WiMAX applications. Samsuzzaman and Islam [13] proposes a multiband fractal patch antenna design that can operate in the frequency bands of 1.8 GHz, 2.4 GHz, and 5.8 GHz for wireless communication systems. Singh, Kumar, and Gupta [14] proposes a multiband printed monopole antenna design that can operate in the frequency bands of 2.4 GHz, 5.2 GHz, and 5.8 GHz for wireless communication systems.

So, the multi-band patch antennas have been widely studied and proposed for various applications such as WLAN, Wi-Fi, and WiMAX. The design of these antennas is challenging due to the requirement of good impedance matching, radiation pattern, and gain over multiple frequency bands. However, the research conducted in this area has shown that it is possible to design compact, lightweight, and low-profile multi-band patch antennas with good performance characteristics, this inspires us to further study and make an antenna that can cover most of the mobile communications bands.

2. DESIGN METHODOLOGY OF THE ANTENNA

This section describes the process that demonstrates how the suggested antenna is created step-by-step to obtain the appropriate frequency band. To get multiband characteristics, four distinct micro strip antennas were constructed and simulated using the ANSYS HFSS software.

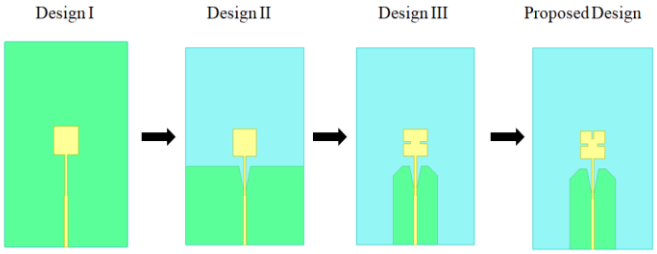


Figure 1: *Different phases of the proposed antenna geometry*

2.1 Basic Rectangular Patch Antenna Design (Design I)

The antenna was designed using the middle frequency of the UWB band, 6 GHz, as the resonant frequency. The model of antenna has been developed using the resonant frequency (f), dielectric constant (ε_r), and dielectric height (h) formulas from [15].The equations provided in [16] have been taken into consideration when designing the classic Rectangular patch antenna (RPA) with dielectric constant (ε_r), resonant frequency (f), and dielectric height (h). Equations (1) to (3) have been used to calculate the length of patch (L_p), width of the patch (W_p) and width of the feed line (W_f). The antenna has been designed and modeled in FEM-based electromagnetic software named Ansys HFSS after all the parameters have been calculated. Some of the parameter values have been tuned to ensure the intended performances. The optimum parameter value for the traditional RPA is displayed in table 1 and the model using the parameters obtained is illustrated in Fig. 1 (Design I).

$$\frac{c}{2f\sqrt{\frac{\varepsilon_r+1}{2}}} \tag{1}$$

$$= \frac{(\varepsilon_r+1)}{2} + \frac{(\varepsilon_r-1)}{2}\left[1+12\frac{h}{Wp}\right]^{\frac{1}{2}} \tag{2}$$

$$\frac{0.412h\left(\varepsilon_{reff}+0.3\right)\left(\frac{W_p}{h}+0.264\right)}{\left(\varepsilon_{reff}-0.258\right)\left(\frac{W_p}{h}+0.8\right)} \tag{3}$$

$$= \frac{c}{2f\sqrt{\varepsilon_{reff}}} \tag{4}$$

$$L_{eff} + 2\Delta L \tag{5}$$

For a rectangular Microstrip patch antenna, the resonance frequency for any TMmn mode is given as [17]:

$$f = \frac{c}{\sqrt{\varepsilon_{reff}}}\left[\left(\frac{m}{L_p}\right)^2 + \left(\frac{n}{W_p}\right)^2\right]^{\frac{1}{2}} \tag{6}$$

Where,

h = dielectric substrate height
W_p = patch width
ε_r = substrate dielectric constant
ε_{reff} = dielectric constant effective

The S11 [dB] of the design I is as illustrated in Fig. 2. It shows the antenna does not radiate in the given frequency range. So more modification are desired in the design and geometry of antenna.

2.2 Rectangular Patch Antenna with Partial Ground (Design II)

In this iteration, more than half of the ground is removed while keeping the patch size constant, and simulation is performed, but the result shows no improvement. To improve the performance, a triangular slot is removed along the microstrip feed line. The antenna began to resonate around 4 GHz and 5.5-6.5 GHz as a result of this modification. Many applications use frequencies ranging from 2.1 GHz to 3 GHz, so this is one of the approximate desired bands. Because the desired frequency band is not obtained in this design, hence, additional modifications are required.

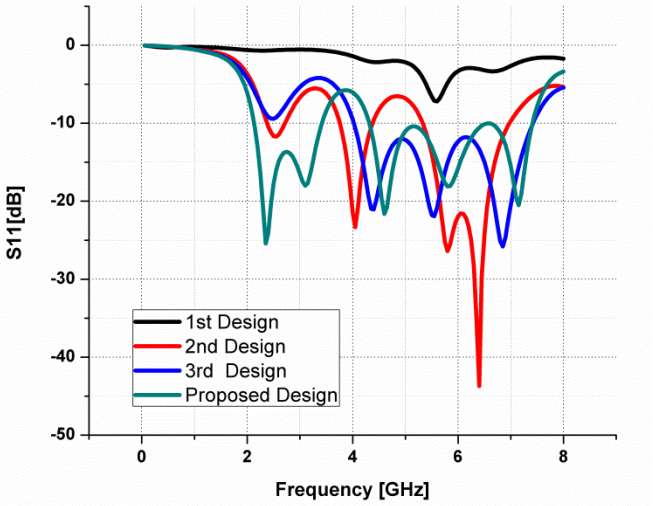


Figure 2: *Comparison of S11 (dB) for the different design under study*

2.3 Slot loaded Patch antenna with small partial ground (Design III)

To achieve the desired frequency band, the ground is reduced further while leaving the triangular slot along the micro stripe line intact. This step yields superior results than design II as shown in Fig. 2. Two 1mm by 3mm slots are introduced along the width of the rectangular patch to improve the design even further. Because of the slots and the removal of the ground, the antenna begins to radiate in a wide band ranging from 4 GHz to 8 GHz. Still, the desired lower band operation is not observed, so more design changes are required.

2.4 Slot loaded (Notched) Patch along width and length with partial ground (Proposed design).

The rejection of a single band frequency was achieved by eliminating a slot from the radiating patch. Fig.4 depicts the suggested patch antenna geometry, whereas Fig. 3. Depicts the S11. Which shows that the slot has enhanced antenna bandwidth as well. The slot dimensions can be changed to control the notched band (length and width). It also shows that the rejected band has a bandwidth ranging from 3.4 GHz to 4.3 GHz. The band-notched feature is obtained by using the stopband's centre frequency, which specifies the dimension of the slot and may be represented as [18],

$$F_n = \frac{c}{4\,L_{slot}\sqrt{\frac{\varepsilon_r+1}{2}}} \tag{7}$$

$$L_{slot} = l_s + w_s \tag{8}$$

Where,
l_s = slot length
w_s = Slot width

The notched can be controlled by changing the value of l_s and w_s.

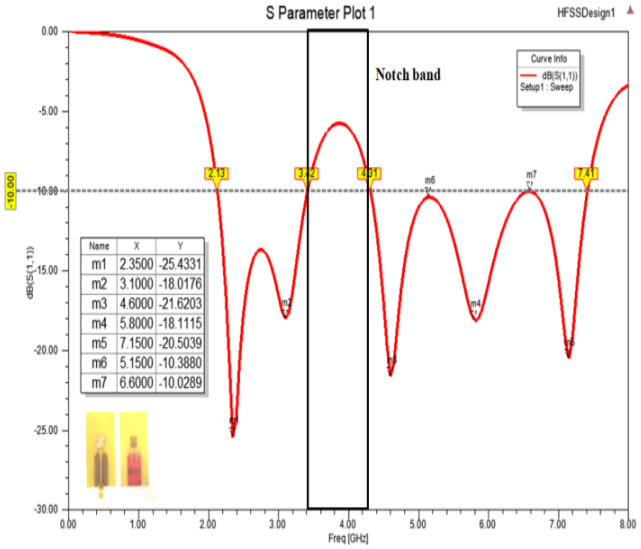


Figure 3: *S11 (dB) graph for proposed design demonstrating notch band.*

Table 1: *Physical Parameters and dimensions of proposed design*

Parameter	Dimensions (mm)
Patch width (W)	11
Patch Length (L)	10.3
Width of Feed line (W_f)	1.34
Length of Feed line(L_f)	36
Ground Length (G_l)	48
Ground width (G_w)	38
Slot width (w_s)	1
Slot Length (l_s)	3
Length of ground Arm (Lg1)	4
Length of ground slot (Lg2)	12
Half Length of ground (Lg3)	9.5

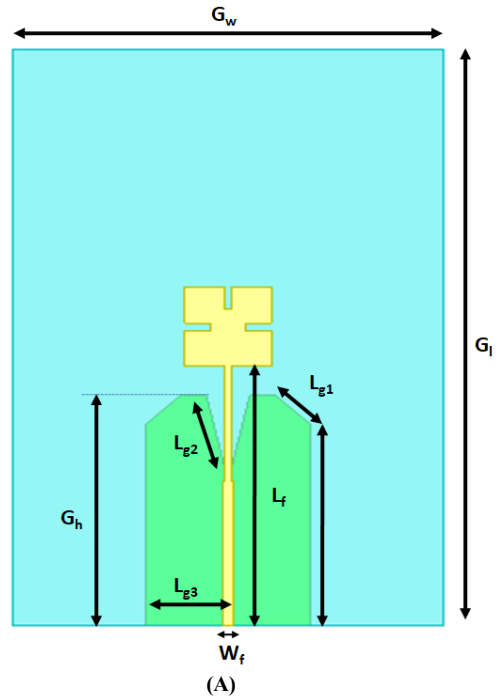


Figure 4.: *Design and parameter of the for proposed antenna design (A)Top view showing parameters of ground and substrate geometry (B) Top View Showing parameters of radiating patch.*

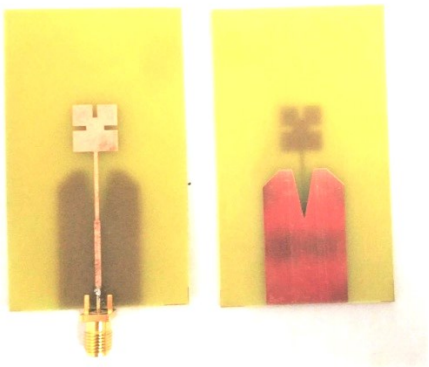


Figure 5: *Fabricated design of proposed antenna with top and bottom view.*

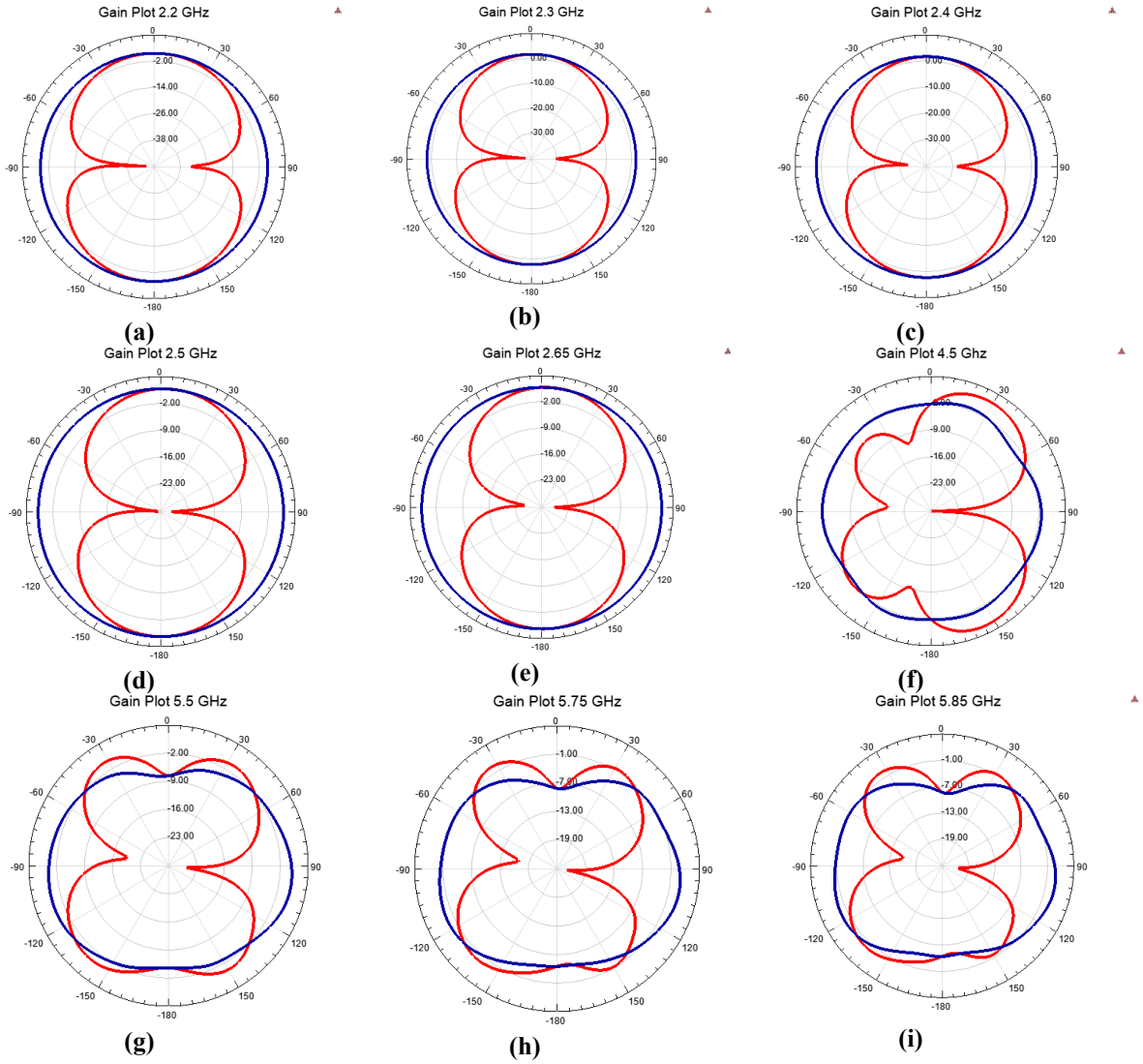


Figure 6: *Radiation pattern of proposed antenna at various frequencies (a)2.2 GHz (b) 2.3 GHz (c) 2.4 GHz (d) 2.5 GHZ (e)2.65 GHz (f)4.5 GHZ(g)5.5 GHz (h)5.75 GHz (i) 5.85 GHz.*

3. RESULTS AND DISCUSSION

After getting results in desired bands it was decided to fabricate the proposed design. The antenna is fabricated on FR4 substrate having loss tangent of 0.001 and permittivity of 4.4. The thickness of FR4 used is 1.6 mm. the fabricated antenna is as shown in Fig 5.

3.1 Comparison of Simulation and Measured S_{11}

The S11 plot of the all four simulated design was compared in Fig 2. It shows that the proposed antenna resonates from 2.13 GHz to 3.42 GHz, having minimum s11 at 2.35 GHz, which is -25.4 dB and at 3.15 GHz it is -17 dB. Further it resonates from 4.32 GHz to 7.41 GHz having three minima at 4.65 GHz, 5.85 GHz and 7.15 GHz.

The comparison of measured result and simulated results are shown in Fig 6. The results shows excellent match at every band. Slight mismatch and slight shift in frequency is observed in

measured result, which may be due to the fabrication error and mismatch at excitation port.

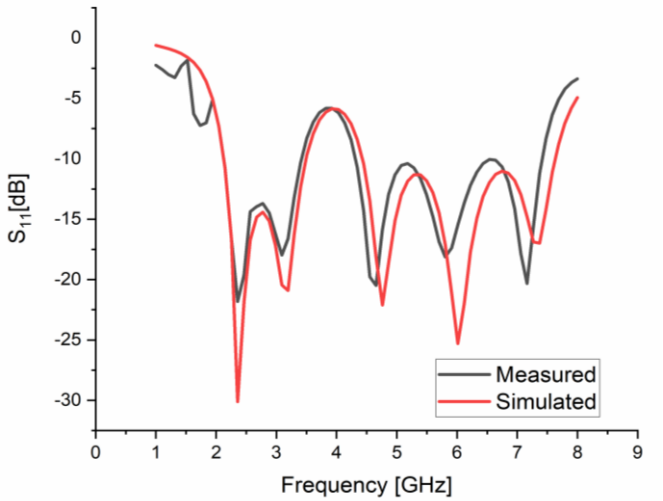


Figure 7: *Comparison of S11 [dB] parameter of proposed antenna with simulated one*

3.2 Radiation Pattern Comparison

The proposed antenna's radiation pattern has been analyzed at various frequencies, ranging from 2.2 GHz to 5.85 GHz, as shown in Fig. X. The results indicate that the radiation pattern remains almost constant in the lower band, 2.1 GHz to 3.1 GHz, with a bi-directional pattern in the elevation plane and an omni-directional pattern in the azimuth plane. This pattern is essential for mobile communication applications. In the higher band, the radiation pattern slightly deteriorates in the azimuth plane from perfect omni-directionality. However, there is no significant variation in the radiation pattern beyond the 5 GHz band, and it remains approximately omni-directional in the azimuth plane.

3.3 Comparison with existing literature

Table 2 provides a comparison of the proposed antenna and a few different UWB antennas. In comparison to the references, it is concluded that the suggested antenna has a compact structure and good notched-band properties.

Table 2: *Comparison of proposed antenna with existing literature*

Ref.	Freq. band (GHz)	Notch band	Application	Dimension	Notch band
[15]	3.1-10.6	3.2-3.8 4.8-6.2	WiMAX, WLAN	30 *30	Dual band
[18]	2.4-5.9	4.2- 4.8	Wifi, WLAN, Wimax	45* 40	Single Band
[19]	2-10	5.10-6.10	UWB	47*37	Single band
[20]	3-11	5-6	UWB	50*41	Single band
[21]	3.5-7.5	5.05-6.17	UWB	49*53	Single band

[22]	3.1-10.6	5-6	UWB	40*35	Single band
This work	2.13-7.14	3.42-4.31	Wifi, Bluetooth, WLAN, Wimax	48*38	Single Band

4. CONCLUSION

In conclusion, the analysis and design of a multiband compact microstrip patch antenna that is appropriate for various wireless communication systems and can function in a wide range of frequencies have been described. Through extensive simulation, optimization and measurement, we have determined that the proposed design provides the best performance, with a gain of 2.15 dB and multi-band operation at 2.11-3.43GHz and 4.30-7.32GHz. The fabricated antenna was tested in ELARC lab V.V. Nagar-Anand, and the simulated and measured results showed excellent agreement. Compared to other multiband and wideband antennas, the proposed antenna is cost-effective, compact, and covers almost all the applications of S-band and C-band. Furthermore, it is suitable for sub-6 GHz 5G applications. Overall, the proposed antenna design offers a promising solution for wireless communication systems operating in a wide range of frequencies.

References

[1] Ali, M., Hayes, G., Hwang, H.-S., & Sadler, R. (2003). Design of a multiband internal antenna for third-generation mobile phone handsets. *IEEE Transactions on Antennas and Propagation*, 51(7):1452-1461.

[2] Liao, W. J., Chang, S. H., & Li, L. K. (2010). A compact planar multi-band antenna for integrated mobile devices. *Progress In Electromagnetics Research*, 109:1-16.

[3] Falade, O., Ur Rehman, M., Gao, Y., Chen, X., & Parini, C. (2012). Single feed stacked patch circular polarized antenna for triple-band GPS receivers. *IEEE Transactions on Antennas and Propagation*, 60(10): 4479-4484.

[4] Zhang, T., Li, R., Jin, G., Wei, G., & Tentzeris, M. (2011). A novel multiband planar antenna for GSM/UMTS/LTE/Zigbee/RFID mobile devices. *IEEE Transactions on Antennas and Propagation*, 59(11):4209-4214.

[5] Liu, Y., Xiao, S., & Yin, Y. (2009). Design and analysis of a novel multiband patch antenna. *Progress in Electromagnetics Research* 93:61-74.

[6] Islam, S. S., Islam, M. T., & Misran, N. (2013). A compact multiband patch antenna with a novel E-shaped feed line. *Microwave and Optical Technology Letters*, 55(3): 530-533.

[7] El-Azem Malhat, H. A., Safwat, A. M. E., & Ibrahem, S. M. M. (2014). A compact multiband patch antenna using a T-shaped feed line. *International Journal of Electronics and Communications*, 68(5):390-396.

[8] Jiang, J., & Xu, Y. (2013). Design of a multiband patch antenna with a slit ring resonator. *IEEE Antennas and Wireless Propagation Letters*, 12:1398-1401.

[9] Islam, T., & Ali, M. (2011). Design and analysis of a compact dual-band patch antenna for WLAN applications. *IEEE Transactions on Antennas and Propagation*, 59(5):1702-1706.

[10] Abbosh, M., & Bialkowski, M. E. (2008). Design and optimization of a multiband patch antenna for wireless communication systems. *IEEE Transactions on Antennas and Propagation*, 56(7), 1868-1876.

[11] Das, S., Behera, S. K., & Patnaik, A. (2014). Multiband microstrip patch antenna for

wireless communication systems. *International Journal of Electronics and Communication Engineering & Technology*, 5(3):66-71.

[12] Vishwakarma, R. K., & Srivastava, D. K. (2014). Design and analysis of multiband stacked patch antenna for WLAN and WiMAX applications. *International Journal of Antennas and Propagation*, 2014, Article ID 739139

[13] Samsuzzaman, S., & Islam, M. T. (2012). Design and analysis of a multiband fractal patch antenna for wireless communication systems. *Progress In Electromagnetics Research*, 132:523-536.

[14] Singh, S. P., Kumar, P., & Gupta, K. C. (2011). Design and analysis of a multiband printed monopole antenna for wireless communication systems. *Microwave and Optical Technology Letters*, 53(11):2639-2643.

[15] Zheng, Z.-A., & Chu, Q.-X. (2009). A compact dual-band printed monopole antenna for WLAN applications. *Progress In Electromagnetics Research*, 11:83-91.

[16] Islam, M. S., Ibrahimy, M. I., Motakabber, S. M. A., Hossain, A. K. M. Z., & Azam, S. M. K. (2019). Design of a Compact UWB Antenna with Dual Notched Bands for WLAN and WiMAX Applications. *Bulletin of Electrical Engineering and Informatics*, 8(2):586-596.

[17] Ye, L.-H., & Chu, Q.-X. (2010). Dual-band monopole antenna with a parasitic strip for *WLAN applications. Electronics Letters*, 46(5): 325-326.

[18] Sun, J.-S., & Huang, S.-Y. (2013). Broadband printed planar monopole antenna for wireless terminal devices applications. Microwave and Optical Technology Letters: 55(1), 79-82.

[19] Zhang, K., Li, Y., & Long, Y. (2010). Band-Notched UWB Printed Monopole Antenna with a Novel Segmented Circular Patch. *IEEE Antennas and Wireless Propagation Letters*, 9:1209-1212.

[20] Ahmed, O., & Sebak, A.-R. (2008). A Printed Monopole Antenna With Two Steps and a Circular Slot for UWB Applications. *IEEE Antennas and Wireless Propagation Letters*, 7:411-413.

[21] Jang, J.-W., & Hwang, H.-Y. (2009). An Improved Band-Rejection UWB Antenna With Resonant Patches and a Slot. *IEEE Antennas and Wireless Propagation Letters*, 8:299-302.

[22] Komjani, M. Y. N. (2011). A Compact Band-Notched UWB Planar Monopole Antenna with Parasitic Elements. *Progress In Electromagnetics Research Letters*, 24:129-138. doi:10.2528/PIERL11043002

RT&A, No.3 (74)
Volume 18, September 2023

DETECTION OF CERVICAL CANCER RISK FACTORS IN VENEZUELA USING DECISION TREE ALGORITHM

*[1]Oladapo M. Oladoja & [2]Taiwo M. Adegoke

•

*[1,2]Department of Mathematics and Statistics, First Technical University, Ibadan, Nigeria
*[1]oladapo.oladoja@tech-u.edu.ng, [2]taiwo.adegoke@tech-u.edu.ng,

Abstract

Cervical cancer, a threat to female existence is one of major cancer affecting women in the developing countries of the world. Several factors are responsible which humans didn™t take cognizance of. These factors are numerous and can at times be difficult to explain using linear regression because it can™t handle many dummy variables that are not necessary to create qualitative predictors. This study uses decision trees to classify and identify the major risk factors causing cervical cancer in women depending on their age since it closely mirrors human decision making than the classical regression approach. A regression tree was constructed from the training data using recursive binary splitting. There was a minimum number of observations required for each terminal node before it stopped. Then cost complexity pruning to the large tree in order to obtain a sequence of best sub trees was applied. By using decision trees as building blocks, we can construct more powerful predictions for decision trees, bagging, random forests, and boosting. 858 cervical cancer patients were observed using 34 risk factor attributes from University Hospital of Caracas, Venezuela. Using classification trees, 14.22% of errors are produced during training. Based on the test data set, 91.5% of the predictions are correct. Based on the data set's pruned data, 91.75% of the observations can be classified correctly. Test predictions generated by this model are within 67 years of the true median age of patients, based on regression trees. Bagging and Random forest show improvement on the regression trees by setting a reduced mean square error. There are four most significant variables among all trees examined by the random forest, including age at first sexual intercourse, number of pregnancies, number of sexual partners, and hormonal contraceptives. The same goes for boosting, as a result of the relative influence statistics.

Keywords: Classification, Pruning, Bagging, Random Forest, Boosting.

1. Introduction

While routine cervical screenings, additional preventive treatments, such as HPV vaccination and safe sex practices, can often help women avoid developing cervical cancer, it is still a severe health risk for them. It's critical for women to monitor their cervical health and to consult with their doctor if they have any worries or queries. The bottom portion of the uterus that links to the vagina, known as the cervix, is typically referred to as the "cervical" in reference to females. The cervix is crucial to the health of female reproductive organs because it supports the baby throughout pregnancy and makes menstruation and birthing easier. A cervical screening test, sometimes referred to as a Pap smear or Pap test, can be used to look at the cells that line the cervix. A small sample of cervix cells is taken for this test in order to look for any abnormal cells that could point to the presence of cervical cancer or other illnesses. Although frequent cervical screenings and other preventive measures like HPV vaccination and safe sex practices can typically help avoid cervical cancer, it remains a severe health problem for women. Women should keep track of their cervical health often and consult with their doctor if they have any worries or inquiries. Hull *et al.* (2020) [1] discussed the difficulties that LMIC healthcare systems encounter

in delivering cervical cancer screening and treatment. These difficulties include a lack of funding, poor infrastructure, and a paucity of qualified healthcare professionals. Moreover, barriers related to culture and society may restrict women's access to programs for cervical cancer prevention. The authors also emphasize how collaborations with international organizations help LMICs' efforts to prevent cervical cancer. They contend that in order to address the difficulties faced by women in these contexts and to lessen the incidence of cervical cancer worldwide, a concerted, international effort is required. Although the study describes the difficulties experienced in LMICs, it offers little advice on how to overcome these difficulties and enhance cervical cancer outcomes in these environments.

According to the cancer statistics, 311,000 women died of cervical cancer in 2018, which was an increase from 570,000 cases in 2017 (Arbyn *et al.*, 2020 [2]). In order to reduce the burden of cervical cancer and achieve global cervical cancer elimination goals, Arbyn *et al.* (2020) [2] argue that investments in preventative and control programs should be increased, especially in low- and middle-income countries. Despite its importance in reducing cervical cancer mortality, this study provides no information on cervical cancer screening in different regions.

Human papillomavirus infection with high-risk strains that continues over time is the main risk factor for cervical cancer (HPV). Smoking, immunosuppression, having several sexual partners, and beginning sexual activity at a young age are other risk factors. A limited resource and infrastructure make cervical cancer screening programs difficult in low- and middle-income countries, according to Zhang *et al.* (2020) [3]. Women's willingness to participate in screening programs may also be affected by social and cultural factors. There is no discussion of the advantages and disadvantages of different cervical cancer screening methods in the article, which provides a general overview of cervical cancer screening methods. Additionally, no detailed guidance is provided on how to implement the screening program. Despite the fact that a substantial share of cervical cancer cases and deaths occur in low- and middle-income countries, the article does not provide a comprehensive analysis of the unique challenges that must be overcome to prevent and treat cervical cancer in these countries.

Human papillomavirus infection with high-risk strains that continues over time is the main risk factor for cervical cancer (HPV). Smoking, immunosuppression, having several sexual partners, and beginning sexual activity at a young age are other risk factors. A limited resource and infrastructure make cervical cancer screening programs difficult in low- and middle-income countries, according to Zhang *et al.* (2020) [3]. Women's willingness to participate in screening programs may also be affected by social and cultural factors. There is no discussion of the advantages and disadvantages of different cervical cancer screening methods in the article, which provides a general overview of cervical cancer screening methods. Additionally, no detailed guidance is provided on how to implement the screening program. Despite the fact that a substantial share of cervical cancer cases and deaths occur in low- and middle-income countries, the article does not provide a comprehensive analysis of the unique challenges that must be overcome to prevent and treat cervical cancer in these countries.

Cervical cancer rates are six times higher among HIV-positive women than among HIV-negative women, and this group accounts for a substantial portion of cervical cancer cases around the world. Based on systematic reviews and meta-analyses, Stelzle *et al.* (2021) [4] assessed the robustness of their findings by conducting sensitivity analyses. They address the significance of their findings for efforts to prevent and control cervical cancer, including the requirement for women who are HIV-positive to have more access to cervical cancer screening and treatment. The study primarily uses information from studies that provide adjusted risk estimates for HIV-associated cervical cancer, which could lead to bias if confounding factor adjustments are insufficient or wrong.

The most frequent cancer that claims the lives of women in developing countries is cervical cancer. New technologies have been developed to enable cervical cancer screening and treatment that is speedier, more reasonably priced, and more sensitive. The use of self-sampling and home-based testing to increase screening uptake as well as the incorporation of biomarkers and personalized screening approaches to increase screening accuracy and decrease the number of

pointless procedures are just a few of the potential future directions for cervical cancer screening discussed by Bedell *et al.* (2020) [5]. The incidence and death differences associated with cervical cancer by race and ethnicity are briefly mentioned in the article, but there is no in-depth discussion of the social and economic variables that contribute to these differences.

A major public health concern worldwide, cervical cancer is the ninth most prevalent disease in terms of new cases. Age-specific incidence and mortality trends presented by Sayo *et al.* (2022) [6] indicate that the decline in cervical cancer incidence and mortality has been greatest among women in their 20s and 30s. The Japanese Cancer Registry, one of the most trustworthy and complete cancer registration systems in the world, provided the data used in this study. The report cites regional and socioeconomic differences in cervical cancer incidence and mortality, but it doesn't go into great detail about how these differences are caused by social and economic variables. The COVID-19 pandemic, which may have altered incidence and mortality rates in recent years, may have had an impact on cervical cancer screening and treatment in Japan, but this is not taken into consideration in the study.

According to research by Rim *et al.* (2022) [7], cervical cancer is more prevalent in Uzbekistan while breast cancer is more prevalent in Korea. Also, compared to Korea, Uzbekistan has a higher mortality rate for cervical and breast cancer. The study stresses the need for enhanced screening programs in Uzbekistan and emphasizes the value of cervical cancer screening in lowering mortality rates. The study only looks at two nations, therefore its conclusions might not apply to other nations with diverse healthcare systems, risk factors, and cultural backgrounds.

Zhao *et al.* (2022) [8] identify several risk factors for cervical cancer in ethnic minorities in Yunnan Province, including advanced age, low education level, young age at first sexual experience, multiple sexual partners, smoking, misunderstanding of cervical cancer prevention, and non-participation in screening for cervical cancer. The study underlines the need for improved screening programs to raise awareness and participation, and emphasizes the significance of cervical cancer screening in lowering the incidence and mortality rates among ethnic minorities in Yunnan Province. The study uses a small sample size, which could restrict how broadly the results can be applied to other populations. The relationships between risk factors and cervical cancer may not be accurate because the study did not account for all relevant confounding variables.

Cervical cancer in women has been the subject of numerous studies, but only a small number of them used machine learning techniques. This study target at determining cervical risk factors of patients at a Venezuelan University Hospital of Caracas. 34 separate risk variables were used to monitor 858 patients.

2. Methodology

A good method for predicting a response from a single predictor variable is simple linear regression. In reality, we frequently use multiple predictors. Giving each predictor in a single model a unique slope coefficient enables us to achieve this. Assume that there are p different predictors in general. The multiple linear regression model then adopts the following form.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon \tag{1}$$

where ϵ is independently and identically distributed with mean zero and variance σ^2 The regression technique may result in accurate predictions on the training set, but it is likely to over-fit the data and perform poorly on the test set. This is because a tree that could be too complex. A smaller tree with fewer splits can lead to lower variance and better interpretation at the cost of some bias. An alternative to the method mentioned above is to only build the tree if the decrease in RSS brought on by each split is greater than a predetermined (high) cutoff point. This strategy will result in smaller trees, however it is too naive because a split that at first glance seems pointless in the tree may be followed by a really good split.

2.0.1 Algorithm 1: Building a Regression Tree

1. Using the training data, construct a huge tree using recursive binary splitting. Only stop when each terminal node has fewer than a predetermined number of observations.
2. Create a list of the top subtrees that are functions of $alpha$ by applying cost complexity pruning to the massive tree.
3. Choose alpha using K-fold cross-validation. Namely, create K folds from the training observations. When $k = 1, ..., K$:
   - Repeat steps 1 and 2 for every fold of the training data other than the kth fold.
   - Analyze the mean squared prediction error as a function of α for the data in the kth fold that were excluded.

   Calculate the results for each value of alpha, then choose the one with the lowest average error.
4. The Step 2 subtree that corresponds to the chosen alpha value should be returned.

2.1. Bagging, Random Forests and Boosting

2.1.1 Bagging

Bagging can enhance predictions for various regression techniques, but decision trees benefit the most from its use. We merely create B regression trees using B bootstrapped training sets, average the outcomes, and apply bagging to the regression trees. These trees are not manicured and are grown deeply. As a result, each tree has a high variation but a low bias. These B trees are averaged to lessen volatility. Bagging has been proven to dramatically improve accuracy by combining hundreds or even thousands of trees into a single step. We create B different bootstrapped training data sets, train our algorithm on the bth bootstrapped training set, and then use that data to generate B additional bootstrapped training data sets in order to obtain and eventually average all the predictions.

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \tag{2}$$

It's referred to as bagging.

2.1.2 Random Forests

Random forests provide an improvement over bagged trees by randomly making minor adjustments to the trees' decorations. We build a number of bagging-like decision tree forests using bootstrapped training samples. However, each time a split in a tree is taken into account when building these decision trees, a random sample of m predictors is chosen as split candidates from the whole collection of p predictors. One of those million dollar forecasters is restricted to the split. Even the majority of the potential predictors at each branch in the tree cannot be considered by the algorithm when building a random forest.

2.1.3 Boosting

Similar to other tree-growing techniques, boosting grows the trees in a certain order utilizing data from earlier trees. With boosting, each tree is fitted to a modified version of the original data set rather than using bootstrap sampling.

2.1.4 Algorithm 2: Boosting for Regression Trees

1. Set $\hat{f}(x) = 0$ and $r_i = y_i$ for all i in the training set.
2. For $b = 1, 2, ..., B$, repeat:
   - Fit a tree $\hat{f}^b$ with d splits (d+1 terminal nodes) to the training data (X, r).
   - Update $\hat{f}$ by adding in a shrunken version of the new tree:
   $$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x) \tag{3}$$
   - Update the residuals,
   $$r_i \leftarrow r_i - \lambda \hat{f}^b(x) \tag{4}$$
3. Output the boosted model,
$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}^b(x) \tag{5}$$

3. Results and Discussion

Both regression and classification problems can be solved using decision trees. We first evaluate the data set using classification trees. Age is a continuous variable in these data, therefore we start by re-coding it as a binary variable that has a value of Yes if the age variable is more than 20, and a value of No otherwise. 14.22% of training errors are made. A tree that offers a good fit to the (training) data will have a minimal deviation. Just divide the deviation by $n - |T_0|$ to get the reported residual mean deviance, which in this instance is $858 - 9 = 849$.

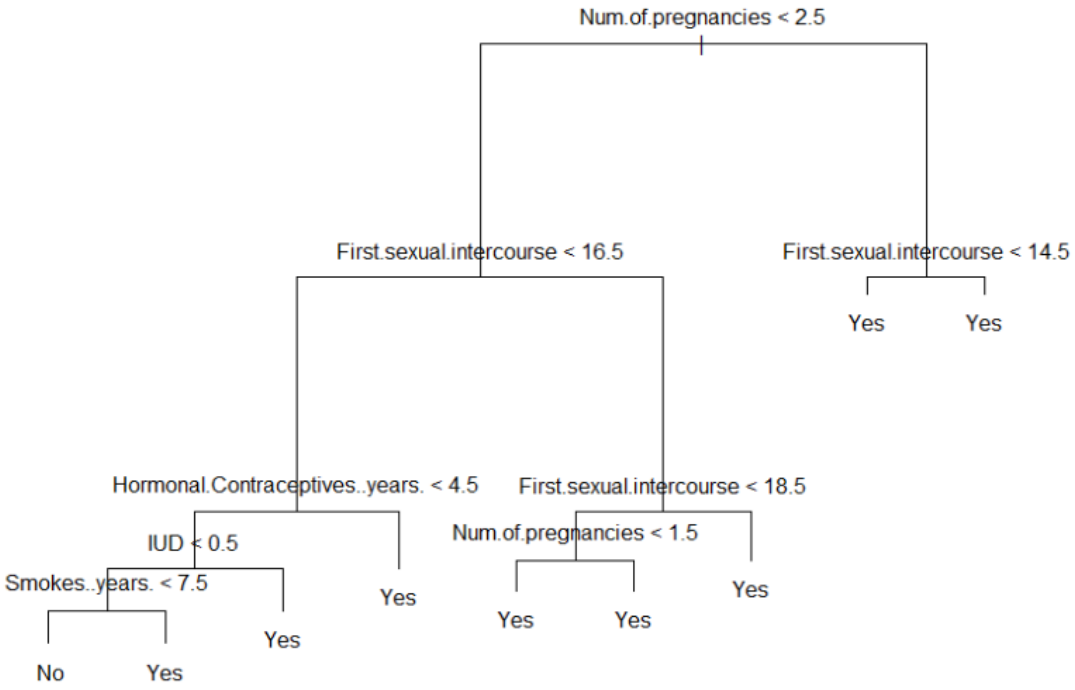


Figure 1: *Regression Trees for Cervical Cancer in Venezuela*

According to Figure 1, which illustrates that the number of pregnancies is a significant risk factor component in determining the patients having cervical cancer depending on their age. The number of pregnancies appears to be the most significant indication of the cervical cancer of patients. Asterisks are used to denote branches that lead to terminal nodes. Instead of just estimating the training error, we must estimate the test error in order to assess a classification tree's performance on this data. The observations are divided into a training set and a test set, the tree is built using the training set, and its performance is assessed using the test set. In the test data set, this method yields accurate predictions for about 90.5% of the locations.

Table 1: *Machine Learning Accuracy Classification*

Tree.Pred	Age	High Test
	No	Yes
No	82	57
Yes	34	285

Now that the pruning procedure has been completed, 91.75% of the test observations are correctly classified, increasing both the interpretability of the tree and the classification accuracy. We get a larger pruned tree with lower classification accuracy if we increase the value of best.

A regression tree is used to fit the data set in this case. We initially create a training set before adjusting the tree to the training data. The findings demonstrate that the tree was constructed using only five of the factors. The total squared errors for the tree constitute the deviation in the context of a regression tree. In accordance with the cross-validation results, we make predictions on the test set using the unpruned tree. As shown in Table 2, we apply bagging and random forests on the data. Random forests beat bagging and regression trees in this case, according to the test set MSE, in terms of outcomes.

Table 2: *Mean Square Error (MSE) for the Test Set*

Algorithm	MSE
Regression Tree	46.05472
Bagging	38.3288
Random Forest	36.65896

There are two reported measurements of varied importance. The %IncMSE is based on the average decline in prediction accuracy on out-of-bag samples when a particular variable is left out of the model. The IncNodePurity is an average over all trees measurement of the total reduction in node impurity caused by splits over that variable (this was plotted in Figure 2).

Table 3: *Importance of Variables*

Variables	%IncMSE	IncNodePurity
Number.of.sexual.partners	13.935063364	1979.451415
First.sexual.intercourse	36.935258441	6489.721033
Num.of.pregnancies	40.324446650	8720.821614
Smokes	1.658543003	188.868735
Smokes..years.	7.417078338	1725.845382
Smokes..packs.year.	3.645987343	630.046003
Hormonal.Contraceptives	5.145181475	448.446348
Hormonal.Contraceptives..years.	19.550078277	3031.607582
IUD	10.366088887	1300.537849
IUD..years.	12.001987891	1696.846107
STDs	1.008508253	148.194058
STDs..number.	2.751653121	171.491002

The training RSS is used to assess node impurity in regression trees, whereas the deviation is used to measure it in classification trees. The findings show that the number of sexual partners, the first sexual encounter, and the number of pregnancies are by far the three most crucial variables across all of the trees taken into account in the random forest.

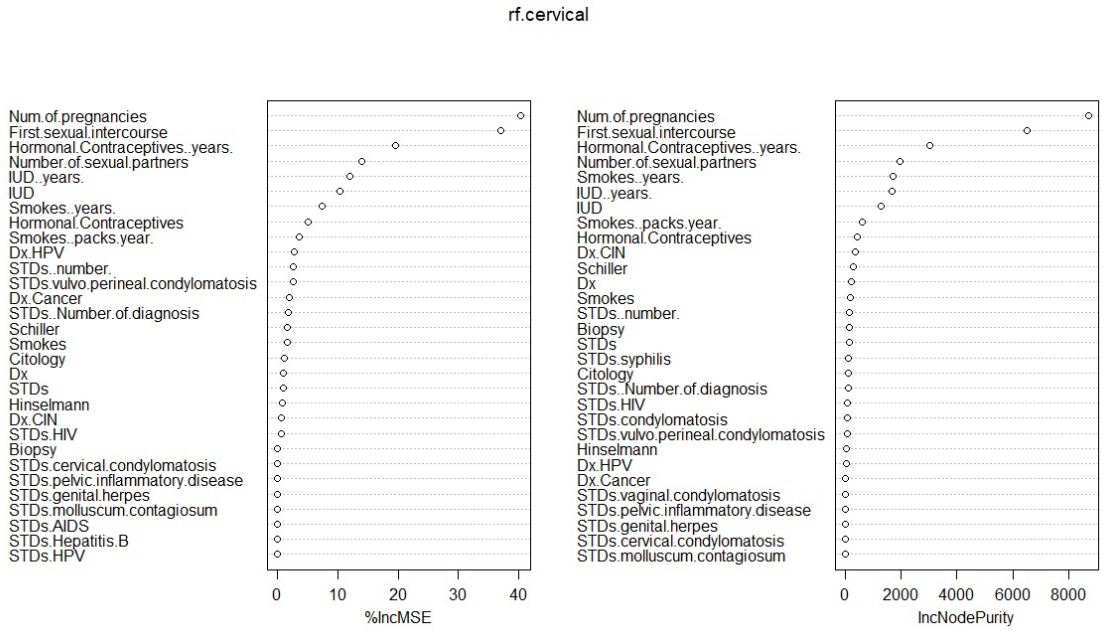


Figure 2: *Plot of Importance Measures*

Relative influence statistics are provided by boosted regression trees along with a relative influence plot as shown in Table 4. We can observe that the most significant factors are clearly the first sexual encounter, number of pregnancies, use of hormonal contraceptives throughout time, and number of sexual partners. Plots of the partial dependence between these four variables are also possible as shown in Figure 3. After integrating out the other factors, these charts show the marginal impact of the chosen variables on the response. Now, we employ the boosted model

Table 4: *Relative Influence Statistics*

Variables	Rel.inf
First.sexual.intercourse	31.900870295
Num.of.pregnancies	23.007299808
Hormonal.Contraceptives..years.	15.871948215
Number.of.sexual.partners	10.889051795
Smokes..years.	3.356270114
IUD..years.	3.217962490
Hormonal.Contraceptives	2.848841347
Smokes..packs.year.	2.728422340
Schiller	2.150597021
IUD	1.786724939
STDs	0.552829213
Smokes	0.509029318
Biopsy	0.391805326
Citology	0.241696894
Dx	0.159287734
STDs..number.	0.147483560
STDs..Number.of.diagnosis	0.136255508
STDs.condylomatosis	0.076823947
Hinselmann	0.018564423
STDs.vulvo.perineal.condylomatosis	0.008235713

to forecast the test set's age at risk for developing cervical cancer. The test MSE obtained is 36.99571, which is higher than that for bagging and comparable to the test MSE for random forests. In boosting, smaller trees are frequently sufficient since the growth of a specific tree takes into account the other trees that have already been developed, as opposed to random forests. Interpretability can also be improved by utilizing smaller trees; for example, using stumps results

in an additive model.

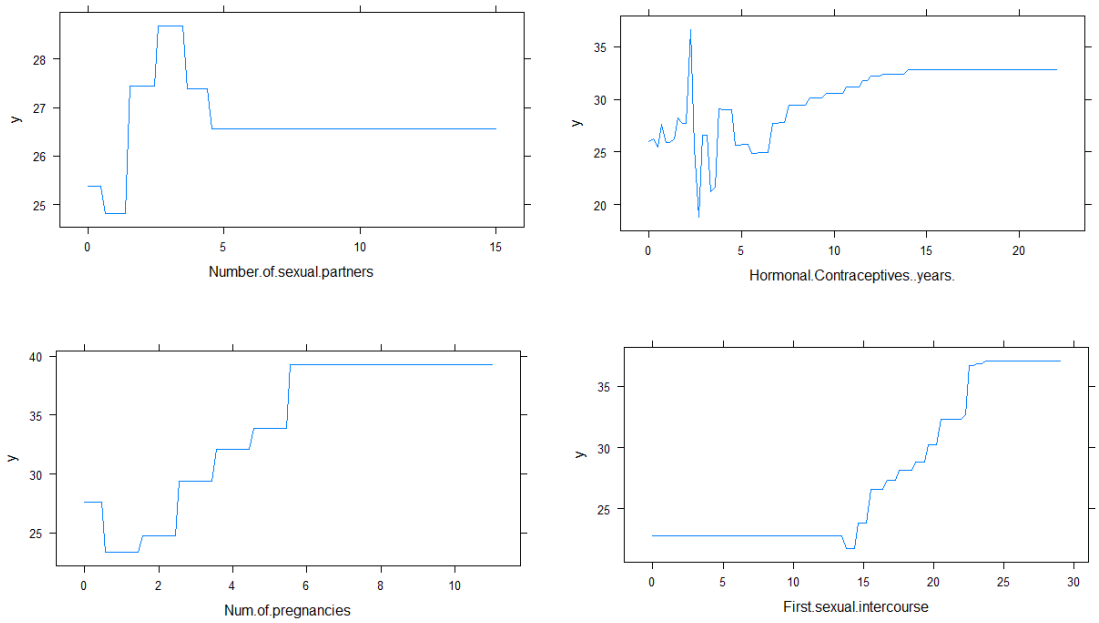


Figure 3: *Plot of Partial Dependence*

4. Conclusion

Regression and classification problems can be solved using decision trees. However, tree-based methods cannot compete with the most appropriate supervised learning approaches in terms of their ease and effectiveness for analysis. In this study, four machine learning algorithms were used namely, regression trees, bagging, random forests and boosting in determining the risk factors of cervical cancer patients due to their age. Random forest and boosting outperformed bagging and regression trees due to their mean square errors. It was discovered that number of pregnancies, first sexual intercourse, number of sexual partners and hormonal contraceptives are the most important risk factors to determine cervical cancer in women depending on their age. Concerned authorities in Venezuela need to take cognizance of the four variables in order to curb cervical cancer.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Hull, R., Mbele, M., Makhafola, T. Hicks, C., Wang, S., Reis, R.M., Mehrotra, R., Mkhize Kwitshana, Z. Kibiki, G., Bates, D.O.and Dlamini, Z. (2020). Cervical Cancer in Low and Middle Income Countries (Review). *Oncology of Letters* 20(3); 2058-2074. https://doi.org/10.3892/ol.2020.11754.

[2] Arbyn, M., Weiderpass, E., Bruni, L., de Sanjosre, S., Saraiya, M., Ferlay, J., and Bray, F. (2020). Estimates of Incidence and Mortality of Cervical Cancer in 2018: A Worldwide Analysis. *The Lancet Global Health*, 8(2), e191-e203. https://doi.org/10.1016/S2214-109X(19)30482-6

[3] Zhang, S., Xu, H., Zhang, L., and Qiao, Y. (2020). Cervical Cancer: Epidemiology, Risk Factors and Screening. *Chinese Journal of Cancer Research*, 32(6), 720-728. https://doi.org/10.21147/j.issn.1000-9604.2020.06.05

[4] Stelzle, D., Tanaka, L. F., Lee, K. K., Ibrahim Khalil, A., Baussano, I., Shah, A. S. V., McAllister, D. A., Gottlieb, S. L., Klug, S. J., Winkler, A. S., Bray, F., Baggaley, R., Clifford, G. M., Broutet, N., and Dalal, S. (2021). Estimates of the Global Burden of Cervical Cancer Associated with HIV. *The Lancet Global Health*, 9(2), e161-e169. https://doi.org/10.1016/S2214-109X(20)30459-9

[5] Bedell, S.L, Goldstein, L.S., Goldstein, A.R. and Goldstein, A.T.(2020). Cervical Cancer Screening: Past, Present, and Future. *Sexual Medicine Reviews*, 8(1), 28–37, https://doi.org/10.1016/j.sxmr.2019.09.005

[6] Sayo, T., Matthew, P. and Kota, K. (2022). Trends in Cervical Cancer Incidence and Mortality of Young and Middle Adults in Japan. *Cancer Science*, 113(5),1801-1807. DOI:10.1111/cas.15320

[7] Rim, C.H., Lee, W.J., Musaev, B., Volichevich, T.Y., Pazlitdinovich, Z.Y., Lee, H.Y., Nigmatovich, T.M. and Rim, J.S. (2022). Comparison of Breast Cancer and Cervical Cancer in Uzbekistan and Korea: The First Report of The Uzbekistan–Korea Oncology Consortium. *Medicina*, 58, 1428. https://doi.org/10.3390/medicina58101428

[8] Zhao, M., Luo, L., Jia, Y., and Chen, B. (2022). Risk factors of Cervical Cancer Among Ethnic Minorities in Yunnan Province, China: A case–control study. *European Journal of Cancer Prevention*, 31(3), 287-292. https://doi.org/10.1097/CEJ.0000000000000704

[9] Obisesan, K.O. and Oladoja, O.M. (2022). On Normal Process of Diffusion Equation in Monitoring Carbon Monoxide Concentrations in Nigeria. *International Journal of Statistical Distributions and Applications*. 8(2)

RT&A, No.3 (74)
Volume 18, September 2023

ASSESSMENT OF WATER QUALITY USING MULTIVARIATE TECHNIQUES

*[1]Olamiji T. Onafowokan, [2]Kazeem O. Obisesan & [3]Oladapo M. Oladoja

•

*[1,2,3]Department of Statistics, University of Ibadan, Ibadan, Nigeria
[2,3]Department of Mathematics and Statistics, First Technical University, Ibadan, Nigeria
*[1]olamijionafowokan17@gmail.com, [2]obidairo@gmail.com,
[3]oladapo.oladoja@tech-u.edu.ng

Abstract

When deciding if water is suitable for a particular usage, its quality"which includes its chemical, physical, and biological characteristics"is referred to. The quality of the water is influenced by many natural and human influences. Despite being in equilibrium, the natural ecosystem and water quality would certainly be disturbed by any large changes in the water quality. In order to assess the levels of water pollution in the Asejire and Eleyele reservoirs, this study conducted a Physico-chemical analysis of the two reservoirs. It also used multivariate techniques to identify the causes of water pollution in the two reservoirs under investigation, used a generalized linear model to analyze the variability in turbidity levels, and suggested regulatory solutions to address water pollution in the two reservoirs under study. In Ibadan, which has a population of about four million, the two main sources of pipe-borne water are the Eleyele and Asejire reservoirs. Between January 2003 and August 2019, water samples were taken from both locations and analyzed for 13 Physico-chemical parameters using the Principal Component Analysis and Cluster Analysis for feature extraction and finally a Generalized Linear model for prediction. Basic Tables and descriptive plots, Principal Component Analysis, Factor Analysis, and Generalized Linear Models were employed. Results: In the Asejire and Eleyele reservoirs, respectively, the PCA yields 5 significant main components explaining 76.56% and 60.97% of the variance, while the FA yields 5 significant major components explaining 94.90% and 79.97%. A generalized linear model (GLM) was used to study the variability in turbidity level, and the results indicate that two parameters"Iron and Silicon"in the Asejire reservoir are crucial for understanding turbidity variation and four"Colour, Alkaline, Silica, and Solids"contribute significantly to turbidity in the water level in the Eleyele Reservoir. With the exception of dissolved oxygen from either reservoir (Eleyele or Asejire) and iron from Eleyele Reservoir, many metrics in Asejire are within SON and WHO acceptable limits. This suggests that the water in the Eleyele reservoir is more contaminated than the Asejire reservoir.

Keywords: Turbidity, Reservoir, Pollution, Contaminated, Principal Component Analysis, Factor, Generalized Linear Model, Physico-chemical.

1. Introduction

Water is required for the development of water transportation, food production, industrial operations, and the development of renewable energy sources, and it is commonly associated with economic progress and human health. As a result, it is critical for human growth as a whole. The survival of living things, the health of ecosystems, the sustainability of human settlements, and economic progress all rely on having access to safe, clean, and sufficient water. However, as the human population has expanded and the demand for additional water has increased, industrial activity, agricultural activity, and climate change have all had a negative impact on water supply and quality (Oketola [1]). However, all human-caused activities have been shown to

have a deleterious impact on water quality in both groundwater and surface water. Domestic use, agricultural operations, and industrial activities are the three main contributors to water contamination, according to the United Nations (2002) [2]. For example, excessive fertilizer use in agricultural activities has been shown to harm human health (Oketola [1]; UNEP [2]). While agriculture is the largest consumer of water, it also contributes significantly to water contamination. In essence, it pollutes water by releasing wastes (bacteria and viruses) from farms into waterways, posing a significant threat to water quality and harming people and wildlife (Nancy [3]).

An indicator of the quality of water is its chemical, physical, and biological characteristics, commonly as a measure of its suitability for a particular purpose. Water quality, according to World Health Organization in 2018 (WHO [4]), can be described by a variety of criteria that limit water use; it is a phrase used to demonstrate the suitability of water to validate certain purposes or processes. The researchers also stated that human and natural factors affect water quality. In regard to water availability and quality, geological, hydrological, and climatic factors are most important (Boyacioglu [5]). However, even when water is available in sufficient amounts, its poor quality restricts the uses that may be made. There is a greater need to maximize the use of limited water resources when available quantities are low. Even though the natural ecosystem and the water quality are in harmony, any large changes to the water quality are typically harmful to the environment.

2. Parameters of Water Quality

The qualities of water include turbidity, which is a fluid's cloudiness or haziness caused by countless small particles that are normally invisible to the human eye, similar to smoke in the air. The measuring of turbidity is an essential water quality test. For aesthetic reasons, water color is primarily an issue for water quality. The perception that colored water is unfit for consumption exists even when it is completely safe for ingestion by the general public. Although iron is a glossy, ductile, malleable, silver-gray metal (group VIII of the periodic table), color can also signal the presence of organic materials like algae or humic chemicals. It is known to exist in four different crystalline forms. To describe how acidic or alkaline a solution is, the pH scale is utilized. It is scored on a scale from 0 to 14. The term pH is made up of the letters "p," which in mathematics represent for negative logarithm, and "H," which chemically stands for hydrogen. The pH range that works best, which is normally between 6.5 and 9.5, is affected by the make-up of the water as well as the materials used to construct the distribution system. Similar to animals and people who live on land, aquatic animals need oxygen to survive. Oxygen from the atmosphere that has dissolved in river and lake water is absorbed by fish and other aquatic species.

Water flowing over rocks in creeks and rivers can introduce oxygen to the water. Since it enhances the taste of the water, a high dissolved oxygen water supply for the community is advantageous. However, high dissolved oxygen concentrations speed up the corrosion of water pipes. Healthy water typically has dissolved oxygen concentrations between 80 and 120 percent and over 6.5-8 mg/L. Total solids, or "TS," is the abbreviation for the sum of the dissolved, colloidal, and suspended solids in a sample of water. This consists of dissolved salts like sodium chloride, or NaCl, as well as solid particles like plankton and silt. An excessive amount of total solids in rivers and streams is a problem that arises rather frequently. The most common contaminant in the investigated streams and rivers is siltation, one of the major contributors to total solids, according to the Environmental Protection Agency's National Water Quality Inventory. The dry weight of non-dissolved suspended particles in a water sample that can be filtered out and measured using a filtering system is known as total suspended solids (TSS). It is a metric for measuring water quality that may be applied to any sort of water or body of water, including ocean water and wastewater that has undergone wastewater treatment. TSS was formerly known as non-filterable residue (NFR), but it was changed to TSS due to misunderstandings in other scientific fields. The National Drinking Water Quality Standard (NDWQS) has established a maximum recommended TSS limit of 25 mg/L.

A silicon and oxygen molecule, or silicon dioxide, is known as silica $(Si0_2)$. It is a tough, glassy mineral that can be found in sand, quartz, sandstone, and granite, among other forms. It is also present in the skeletons of both plants and animals. The majority of water supplies will contain some silica because it is the second most common element on Earth after oxygen. In every system of natural water, some is dissolved. A chemical indicator of water's capacity to neutralize acids is alkalinity. Alkalinity gauges a water's resistance to pH changes brought on by the addition of acids or bases. Strong bases (such OH-) may occasionally play a role in severe situations, but weak acid salts are primarily responsible for naturally alkaline water.

The two main causes of alkalinity in natural rivers are the partitioning of CO_2 from the atmosphere and the weathering of carbonate minerals in rocks and soil. Weak acid salts that may be present in trace amounts include borate, silicates, ammonia, phosphates, and organic bases produced from naturally existing organic compounds. The amount of calcium (Ca+2) ions in water expressed as calcium carbonate $(CaCO_3)$ is referred to as the calcium hardness. Calcium must be present in the pool water at a specified amount. Calcium and other minerals are dissolved from plaster pool surfaces and metal equipment parts when the calcium level in the water is too low (soft water). Calcium carbonate scale can develop on pool surfaces and recirculation equipment, especially heat-exchanging surfaces, when there is an excessive amount of calcium present (hard water, supersaturated). It is recommended to maintain calcium hardness levels between 150 and 1000 ppm. The optimal range for calcium hardness is 200 to 400 ppm.

Several research have used multivariate statistics to better understand natural and anthropogenic water contamination causes (Praus, 2007 [6]; Boyacioglu, 2008 [5]; Pejman *et al.*, 2009 [7]; Koklu *et al.*, 2010 [8]). However, because the majority of these research were conducted in Europe and Asia, there is a spatial-temporal variation that is heavily influenced by seasonality fluctuations.

Obisesan and Christopher (2018) [9] used statistical approaches such as principal component analysis and the General Linear Model to assess water pollution in the Asejire and Eleyele reservoirs, however the dataset was limited in scope (2003-2007). As a result, this analysis expanded on Obisesan and Christopher's (2018) work by extending the dataset from 2003 to 2019. This is to critically evaluate and statistically analyze changes in reservoir (Water) quality over time, as well as to statistically determine the influence of various statistical models on water quality metrics.

The two main reservoirs that provide water to the Ibadan Metropolis are Asejire and Eleyele, and they are the focus of this research. The measurements from these reservoirs were obtained from the Water Corporation of Oyo State in Ibadan, Nigeria, and are included in the data collection. The fact that these data are expansions of those from Obisesan and Christopher should be emphasized (2018). The 13 physicochemical parameters assessed monthly from January 2003 to August 2019 were turbidity, color, pH, dissolved oxygen, alkalinity, total hardness, calcium hardness, iron, silica, total solids, dissolved solids, and total suspended solids. These characteristics are necessary to determine the severity of the effects of water contamination. The aim of this study is to investigate water quality in Asejire and Eleyele Reservoirs using multivariate techniques.

3. Methodology

3.1. Principal Component Analysis

A large number of (potentially) linked variables are reduced to a smaller number of uncorrelated variables known as principal components through the mathematical process of principle component analysis (PCA). Each following component takes into account as much of the remaining variability as is practical whereas the first principal component takes into account as much of the data variability as is practicable.

The main goal of PCA is to transform a set of correlated qualities into a more manageable set of uncorrelated characteristics that account for the majority of the variation in the original attributes.

The sample data matrix of the n samples that were submitted to the n distinct characterization processes can be represented as matrix X.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \tag{1}$$

The data in matrix X are mean-centered in order to generate the deviation matrix D. To accomplish this, the data mean is removed from each data point. Mean centering eliminates measurement bias.

$$D = \begin{bmatrix} x_{11} - \bar{X}_1 & \dots & x_{1n} - \bar{X}_n \\ x_{21} - \bar{X}_2 & \dots & x_{2n} - \bar{X}_n \\ \vdots & \vdots & \ddots \quad \vdots \\ x_{m1} - \bar{X}_m & \dots & x_{mn} - \bar{X}_n \end{bmatrix} \tag{2}$$

The covariance matrix of the data set S, is constructed by,

$$S = \frac{D \cdot D^T}{n} \tag{3}$$

Resulting

$$S = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix} \tag{4}$$

Where,

$$C_{ij} = 1/n\{(x_i - \bar{X}_i)(x_j - \bar{X}_j)\} \quad (i, j = 1, 2, ..., k) \tag{5}$$

The variance and covariance of the covariance matrix are real numbers. As a result, the covariance or variances cannot be compared when the variables in the covariance matrix are not measured in the same units. bigger values for the variables in the measurements will result in bigger variances, whereas lower values for the variables in the measurements will result in lower variances. By dividing each matrix member by its standard deviation, standardize the data to prevent the scale dependency of the covariance matrix.

Normalized matrix element C_{ij}

$$C_{ij} = \frac{c_{ij}}{\sqrt{var(i)var(j)}} \quad (i, j = 1, 2, ..., k) \tag{6}$$

Variance of *ith* element is given as (i). The maximum variation the *ith* and *jth* variable can have is (i) and $Var(j)$ respectively. Therefore, the correlation between *ith* and *jth* variable, C_{ij}, can never exceed $\sqrt{var(i)var(j)}$ resulting the maximum value a covariance matrix element to one. For two variables that are uncorrelated, the covariance is zero $(C_{ij} = C_{ji} = 0)$. Correlation matrix is symmetric due to the fact, $C_{ij} = C_{ji}$ and it is always real and positive definite.

PCA's main goal is to reduce the size of the data set while preserving as much variance as feasible from the original dataset. The covariance matrix describes the spread (variance) and orientation (covariance) of the data collection. As a result, a normally distributed K dimensional data set may be completely explained by a $K \times K$ covariance matrix along with the variable mean values.

The process of converting a square matrix into a diagonal matrix, which shares the same fundamental characteristics as the initial square matrix, is known as matrix diagonalization. A matrix is diagonalized when its original variables are replaced with a certain set of new variables, at which point the matrix assumes its canonical form. In other words, it's the same as figuring out

a square matrix's eigenvalues. These eigenvalues will be the diagonal components of the resulting diagonal matrix. The new set of variables created by diagonalization, known as eigenvectors, correspond to the diagonal matrix.

We must diagonalize the correlation matrix, S, in order to describe it with directions and magnitudes (vectors).

$$S\overrightarrow{v_1} = \gamma_1 \overrightarrow{v_1} \quad (i = 1, 2, ..., k) \tag{7}$$

where γ_1 is the eigenvalue and $\overrightarrow{v_1}$ is the corresponding eigenvector of the correlation matrix, S.

$$S\overrightarrow{v_1} - \gamma_1 \overrightarrow{v_1} = 0 \tag{8}$$

$$(S - \gamma_1 I)\overrightarrow{v_1} = 0 \tag{9}$$

Where I is the identity matrix of the same dimensions as S. If v is not a null vector, then the equation above can only be define if $(S - \gamma_1 I)$ is not invertible. If a square matrix is not invertible then its determinant is zero.

$$Det(S - \gamma_1 I) = 0 \tag{10}$$

Solving the above equation gives a set of k eigenvalues and their corresponding orthogonal eigenvectors.

3.2. Factor Analysis

A statistical method known as factor analysis is used to translate variance among related, observable variables into a potentially more manageable set of unobservable variables known as factors. Alternatively, it is possible that changes in three or four of these observable variables, for instance, largely reflect changes in fewer of these unobservable variables. Such combined changes in response to unobservable hidden factors are what factor analysis searches for. The potential components are combined linearly to represent the observed variables using "error" terms. Later, the number of variables in a dataset can be reduced using the understanding of the relationships between observed variables that has been learned.

Suppose there exists a set of random variables, $x_1, ..., x_p$ and means $\mu_1, ..., \mu_p$. Suppose for some unknown constants I_{ij} and k unobserved random variables F_j where $i \in 1, ..., p$ and $j \in 1, ..., k$ where $k < p$, we have

$$x_i - \mu_i = L_{i1}F_1 + ... + L_{ik}F_k + \varepsilon_i$$

Here, the ε_i are independently distributed error terms with zero mean and finite variance, which may not be the same for all i. Let $Var(\varepsilon_i) = \psi$, so that we have,

$$Cov(\varepsilon) = Diag(\psi_1, ..., \psi_p) = \bar{\psi} \quad \text{and} \quad E(\varepsilon) = 0$$

In matrix terms, we have

$$x - \mu = LF + \varepsilon$$

Any solution of the above set of equations following the constraints for F is defined as the factors, and L as the loading matrix.

3.3. Generalized Linear Model

In a general linear model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + +\beta_q x_{qi} + \varepsilon \tag{11}$$

The response $y_i, i = 1, ..., n$ is modeled by a linear function of explanatory variables $x_j, j = 1, ..., q$ plus an error term.

Here general refers to the dependence on potentially more than one explanatory variable, versus the simple linear model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon \tag{12}$$

For the general linear model with $\varepsilon \sim N(0, \sigma^2)$, we have the linear predictor

$$\vartheta = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_q x_{qi} \tag{13}$$

the link function

$$g(\mu_i) = \mu_i \tag{14}$$

and the variance function

$$var(\mu_i) = 1 \tag{15}$$

4. Results and Discussion

4.1. Exploratory Data Analysis

Table 1: *Descriptive Statistics for Asejire and Eleyele Reservoir*

	ASEJIRE RESERVOIR				ELEYELE RESERVOIR				LIMITS	
PARAMETERS	MIN	MAX	MEAN±SD	MED	MIN	MAX	MEAN±SD	MED	SON	WHO
Tur	0.0000	4.000	0.9278±1.2781	0.08	0.000	32.900	4.420±4.460	3.000	5	5
Col	4.00	7.00	5.01±0.2239	5.00	5.000	30.000	6.085±3.545	5.000	15	15
PH	6.400	8.800	7.476±0.3824	7.400	6.000	8.000	7.010±0.298	7.000	6.5-8.5	6.5-9.5
DO	3.500	15.700	7.638±1.4735	7.400	1.000	54.000	8.925±5.900	7.550	5	5
Alk	22.00	88.00	49.30±11.807	50.00	8.000	157.00	63.08±23.458	60.00	-	100
TH	36.00	104.00	60.83±11.169	58.00	64.00	148.00	94.65±12.947	94.00	150	100
CaH	9.00	80.00	41.89±10.899	41.00	30.00	108.00	62.88±13.769	63.50	-	75
Cl	10.40	67.00	23.51±8.4506	20.75	8.78	90.50	33.87±8.740	34.75	250	250
Fe	0.00	0.2800	0.08157±0.1125	0.00	0.00	3.20	2.27±0.471	2.30	0.3	0.3
Si	3.00	14.00	6.22±3.2534	4.00	4.00	17.00	7.51±4.199	5.00	-	-
Sol	40.0	1402.0	166.7±141.45	137.0	178.0	365.0	246.2±22.9526	245.0	-	-
DS	24.00	682.00	100.91±72.676	88.00	134.0	245.0	171.4±11.052	172.0	500	1000
SS	5.00	720.00	68.33±76.629	44.00	36.00	92.00	72.44±8.116	74.00	-	500

Asejire Reservoir: With high mean concentrations of 166.7 mg/L and 100.91 mg/L, respectively, Sol and DS are clearly the prominent parameters in Table 1. This demonstrates the shared origin of these variables. The typical PH value is 7:476 LU, which is little above neutral. Tur, Col, DO, Alk, TH, CaH, Cl, Fe, Si, and SS had average concentrations of 0.93, 5.01, 7.64, 49.30, 60.83, 41.89, 23.51, 0.08, 6.22, and 68.33 mg/L, respectively.

Eleyele Reservoir: According to Table 1, the dominating parameters in the Eleyele reservoir Sol are DS and TH, with mean concentrations of 246.2 mg/L, 171.4 mg/L, and 94.65 mg/L, respectively. This further demonstrates the human origin of these variables (Mustapha and Abdu 2012; Awoyemi et al. 2014). The average pH level is 7.010 LU, which is just above neutral. Tur, Col, DO, Alk, CaH, Cl, Fe, Si, and SS had average concentrations of 4.42, 6.09, 8.93, 63.08, 62.88, 33.87, 2.27, 77.51, and 72.44 mg/L, respectively.

Thus, in both the Asejire and Eleyele reservoirs, Total Solids and Dissolved Solids are the dominant parameters with high mean concentrations. Additionally, Asejire and Eleyele have dissolved oxygen concentrations that are 7.64 and 8.93 mg/L, respectively, above the allowable limit. Iron (Fe) levels in Eleyele Reservoir are also higher than allowed. This can be viewed in Figure 1 below.

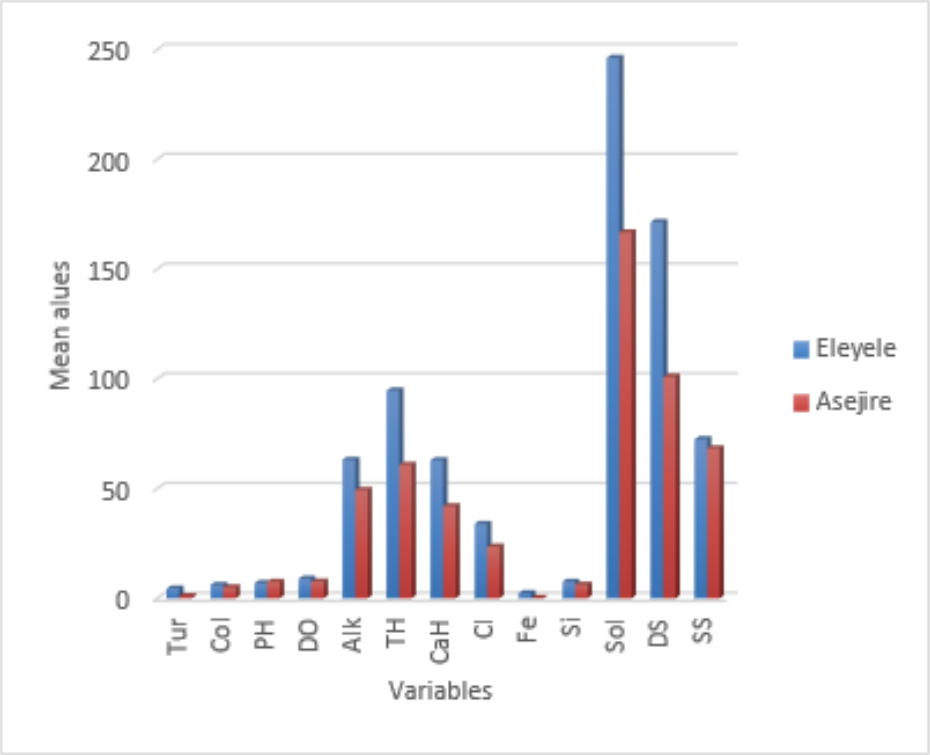


Figure 1: *Mean Plot Variable Values for Asejire and Eleyele Reservoirs*

4.2. Principal Component Analysis

Table 2: *Eigenvalues, Percentage Variance and Percentage Cumulative Variance for Both Reservoir PCs*

	ELEYELE RESERVOIR			ASEJIRE RESERVOIR		
PCs	Eigenvalue	%Variance	%Cum. variance	Eigenvalue	%Variance	%Cum. variance
PC1	2.2649278	17.422522	17.42252	3.47776690	26.7520530	26.75205
PC2	1.7912737	13.779028	31.20155	2.78797470	21.4459593	48.19801
PC3	1.4394865	11.072973	42.27452	1.52965004	11.7665388	59.96455
PC4	1.2299078	9.460829	51.73535	1.16678960	8.9753046	68.93986
PC5	1.2005036	9.234643	60.97000	0.99114982	7.6242294	76.56409
PC6	0.9619500	7.399615	68.36961	0.79671430	6.1285716	82.69266
PC7	0.9576967	7.366898	75.73651	0.77404385	5.9541834	88.64684
PC8	0.8029249	6.176345	81.91285	0.46483935	3.5756873	92.22253
PC9	0.6192486	4.763451	86.67630	0.31852158	2.4501660	94.67269
PC10	0.5870515	4.515781	91.19209	0.29688666	2.2837435	96.95644
PC11	0.5053452	3.887271	95.07936	0.19322174	1.4863211	98.44276
PC12	0.3389613	2.607395	97.68675	0.13872619	1.0671245	99.50988
PC13	0.3007223	2.313248	100.00000	0.06371526	0.4901174	100.00000

The water sample of dairy waste was subjected to a main component analysis, as indicated in Table 2. It contains loading for the component matrix that has been rotated, eigenvalues for each component, variance percentages, and cumulative variance percentages explained by each component. 13 physico-chemical parameters are taken into account during PCA, and the findings are summarized in a table. It shows that the first five principal components together account for 76.56% of the total variance in the dataset, where the first principal component accounts for 26.75%, the second for 21.44%, the third for 11.77%, the fourth for 8.98%, and the third for 7.62% of the total variance. To accurately assess the clustering behavior, PCA is used. Principal

components (PC) are only extracted as components when Eigen values are greater than one. Factor loadings are used to represent PCs with Eigen values greater than the unit value. Factor loading is divided into three categories: strong, moderate, and mild. It ranges from 0.75 to 0.5. High factor loadings are present in the following principal components: PC1: Alk (0.7122), PC2: Sol (0.9026), PC3: DS (0.8397), PC4: SS (0.8313), and PC5: Col (0.7975). These factors have high factor loadings, indicating that they are the main polluters among other parameters and have high concentrations. PC3 and PC4 showed loadings ranging from strong to moderate.

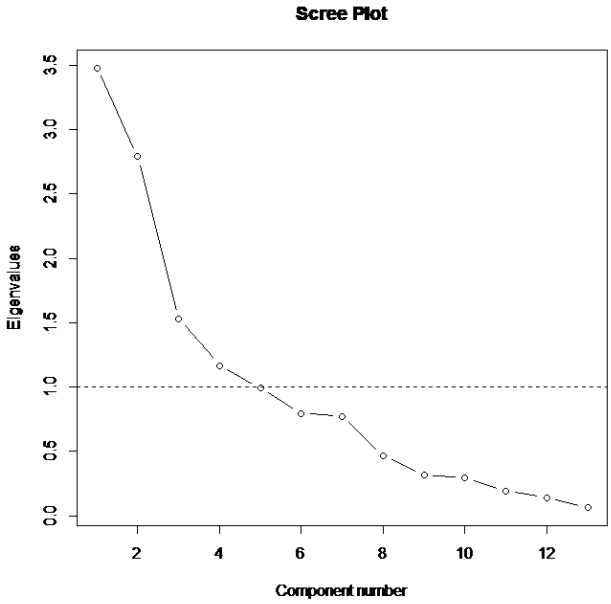


Figure 2: *A Scree Plot of the Principal Components (Eleyele Reservoir)*

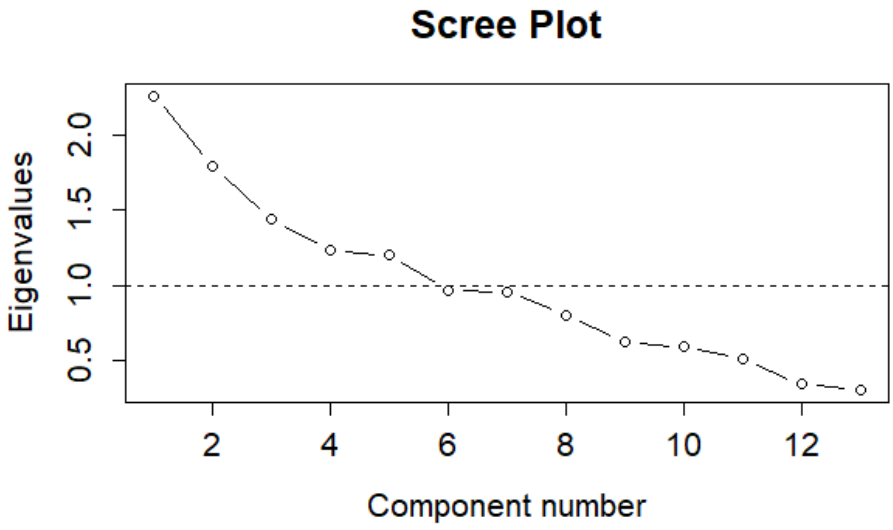


Figure 3: *A Scree Plot of the Principal Components (Asejire Reservoir)*

Table 3: *Factor Analysis for the Asejire Reservoir*

Parameter	PA1	PA2	PA3	PA4	PA5	h2	u2
Tur	-0.88	-0.04	-0.11	0.07	-0.05	0.790	0.21
Col	-0.06	0.00	-0.08	0.14	0.01	0.030	0.97
PH	-0.01	-0.08	0.09	0.65	-0.03	0.432	0.57
DO	0.05	-0.08	0.11	0.28	-0.04	0.099	0.90
Alk	0.55	0.01	0.31	0.29	0.23	0.528	0.47
TH	0.13	-0.05	0.84	-0.02	0.10	0.731	0.27
CaH	0.12	-0.09	0.78	0.17	0.13	0.681	0.32
Cl	0.21	0.09	0.19	-0.06	0.61	0.467	0.53
Fe	-0.93	-0.07	-0.08	0.09	-0.05	0.890	0.11
Si	-0.82	-0.03	-0.06	0.03	-0.15	0.700	0.30
Sol	0.10	1.01	-0.07	-0.11	-0.02	1.050	-0.05
DS	-0.13	0.85	-0.02	-0.03	0.02	0.739	0.26
SS	0.21	0.84	-0.08	-0.15	0.14	0.800	0.20
Eigenvalue	7.7661e-02	3.284e-02	8.9677e-03	5.7938e-03	4.265e-03		
% of variance	56.90	24.06	6.57	4.24	3.12		
Cum. % of variance	56.90	80.96	87.53	91.78	94.90		

4.3. Factor Analysis

The Factor Analysis (FA) helped to identify and extract the variables affecting water quality. Table 12 shows that FA identified latent components that accounted for 94.90% of the variation. Alkaline had a positive loading of 0.55 and Turbidity, Iron, and Silicon all exhibited very substantial negative loadings of -0.88, -0.93, and -0.82, respectively. PA1 explained 56.90% of the total variation. This could be considered anthropogenic input. PA2 demonstrated high positive loadings on suspended particles (0.84) and dissolved solids (0.85) and explained 24.06 percent of the overall variation. With substantial positive loadings on Total Hardness (0.84) and Calcium Hardness (0.78), PA3 explained 6.57 percent of the total variance. With mildly positive loadings on PH (0.65), PA4 explained 4.24 percent of the overall variance. With moderately positive loadings on chlorine (0.61), PA5 explained 3.12% of the overall variation.

Testing the idea that five criteria are adequate. The null model has 78 degrees of freedom and an objective function of 7.14 with a 1383.64 Chi Square. The objective function was 0.16, and the model has 23 degrees of freedom. Root mean square residuals (RMSR) are equal to 0.01. Root mean square of the residuals corrected for degree of freedom is 0.03.

The Factor Analysis (FA) helped to identify and extract the variables affecting water quality. FA identified latent components in table 13 above that accounted for 79.97% of the overall variance. With positive loadings on Turbidity (0.55) and Silicon (0.92), PA1 accounted for 33.60% of the total variance, while PA2 explained 18.89%, PA3 explained 12.95%, PA4 explained 8.28% of the total variance with positive loadings on Total Solids (0.59) and Dissolved Solids (0.53), and PA5 accounted for 6.25% of the total variance.

The five factors are adequate theory is put to the test. The null model has 78 degrees of freedom, an objective function of 2.12, and a Chi Square of 411.88. The model has 23 degrees of freedom, and its objective function was 0.26. The residuals' root mean square root (RMSR) is 0.04. The residuals' df-corrected root mean square is 0.07, their harmonic number is 200, and their empirical chi square is 39.9 with a probability of less than 0.016.

Table 4: *Factor Analysis for the Eleyele Reservoir*

Parameter	PA1	PA2	PA3	PA4	PA5	h2	u2
Tur	0.55	-0.23	-0.02	0.27	0.06	0.431	0.57
Col	0.34	-0.01	-0.04	0.10	-0.04	0.129	0.87
PH	0.43	0.15	0.04	-0.05	-0.09	0.223	0.78
DO	0.32	0.17	0.48	0.02	0.03	0.361	0.64
Alk	-0.07	1.01	0.00	-0.01	0.12	1.040	-0.04
TH	-0.05	0.41	-0.57	0.08	0.00	0.498	0.50
CaH	0.12	0.04	-0.65	0.06	0.05	0.448	0.55
Cl	0.15	-0.15	-0.28	0.23	-0.04	0.174	0.83
Fe	-0.03	-0.05	0.05	0.04	-0.28	0.088	0.91
Si	0.92	-0.17	-0.03	-0.05	-0.02	0.879	0.12
Sol	0.04	0.00	-0.09	0.59	0.13	0.378	0.62
DS	0.10	0.07	-0.02	0.53	-0.42	0.469	0.53
SS	-0.17	0.00	0.09	0.14	0.48	0.282	0.72
Eigenvalue	1.211e-02	6.809e-03	4.667e-03	2.986e-03	2.253e-03		
% of variance	33.60	18.89	12.95	8.28	6.25		
Cum. % of variance	33.60	52.49	65.44	73.72	79.97		

4.4. Generalized Linear Model

The Generalized Linear Model (GLM) estimates of the relationship between Turbidity and other water variables are presented in Table 5 below. It is clear that Silicon (Si) and Iron (Fe) contribute greatly to the water's level of turbidity in Asejire. We could also find that the water variables taken into account in this investigation could explain 72.73% of the turbidity.

Table 5: *GLM estimates and SD for Asejire Reservoir*

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	-1.1636528	1.4200685	-0.819	0.41358
Col	0.2974706	0.2144660	1.387	0.16708
PH	-0.0559984	0.1340955	-0.418	0.67672
DO	0.0405164	0.0334850	1.210	0.22781
Alk	-0.0042912	0.0053421	-0.803	0.42284
TH	0.0029505	0.0060358	0.489	0.62553
CaH	-0.0075631	0.0064229	-1.178	0.24048
Cl	-0.0008144	0.0063186	-0.129	0.89759
Fe	7.8805683	0.6996219	11.264	< 2e-16 ***
Si	0.0694667	0.0243463	2.853	0.00481 **
Sol	-0.0004600	0.0010578	-0.435	0.66419
DS	0.0000971	0.0013802	0.070	0.94398
SS	0.0010290	0.0014926	0.689	0.49145
(Dispersion parameter for Gaussian family taken to be 0.4454064)				
Null deviance: 325.061 on 199 degrees of freedom				
Residual deviance: 83.291 on 187 degrees of freedom				
AIC: 420.38				
Multiple R-squared: 0.7438, Adjusted R-squared: 0.7273				

Only Iron and Silicon contributes greatly to turbidity in the water level. The fitted regression equation for turbidity level is given by

$$Tur = -1.1636528 + 7.8805683 * Fe + 0.0694667 * Si$$

Each one's impact on turbidity is explained by other factors. Turbidity will decrease at

1.1636528 with zero contribution from all independent factors. When all other factors are held equal, a rise in turbidity of 7.881 units is projected for every unit increase in iron (Fe). Similarly, assuming all variables remain constant, it is expected that for every unit increase in silicon (Si), turbidity level will increase by 0.069 units.

The Generalized Least Squares (GLS) estimates of the regression between Turbidity and other water variables are presented in Table 6 below. It is clear that the turbidity level of the Eleyele water level is substantially influenced by color, alkalinity, silicon (Si), and solids. We could also find that the water variables taken into account in this investigation might explain 37.56% of the turbidity.

Table 6: *GLM estimates and SD for Asejire Reservoir*

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	-6.873238	8.399258	-0.818	0.4142
Col	0.176279	0.078312	2.251	0.0256 *
PH	-1.239459	0.999494	-1.240	0.2165
DO	0.093316	0.051428	1.814	0.0712 .
Alk	-0.034439	0.014186	-2.428	0.0161 *
TH	0.032033	0.025496	1.256	0.2105
CaH	-0.007554	0.021783	-0.347	0.7291
Cl	0.048267	0.032100	1.504	0.1344
Fe	0.449737	0.583198	0.771	0.4416
Si	0.480872	0.080525	5.972	1.16e-08 ***
Sol	0.026909	0.012235	2.199	0.0291 *
DS	0.013312	0.025536	0.521	0.6028
SS	0.034759	0.033860	1.027	0.3060
(Dispersion parameter for Gaussian family taken to be 13.21585)				
Null deviance: 3957.9 on 199 degrees of freedom				
Residual deviance: 2471.4 on 187 degrees of freedom				
AIC: 1098.4				
Multiple R-squared: 0.3756, Adjusted R-squared: 0.3355				

Colour, Alkaline, Silica and Solids contributes greatly to turbidity in the water level. The fitted regression equation for turbidity level is given by

$$Tur = -6.873238 + 0.176279 * Col - 0.034439 * Alk + 0.480872 * Si + 0.026909 * Sol$$

Each one's impact on turbidity is explained by other factors. Turbidity will be decreasing at 6.87323 if none of the independent variables contribute. Keeping all variables equal, it is expected that for every unit increase in Color (Col), there will be a 0.1763 unit rise in Turbidity level. Similar to this, it is predicted that for every unit increase in silicon (Si), there will be a 0.4809 unit increase in turbidity level, for every unit increase in alkaline (Alk), there will be a 0.4809 unit decrease in turbidity level, and for every unit increase in solids (Sol), there will be a 0.0269 unit increase in turbidity level, all other variables being held constant.

5. Conclusion

In this study, a variety of multivariate exploratory techniques were used to assess changes in the quality of the surface water in the reservoirs of Eleyele and Asejire. This study's primary goal is to determine the levels of turbidity and water pollution in the Asejire and Eleyele Reservoirs utilizing multivariate and generalized linear model techniques. According to the descriptive statistics, all of the parameters with the exception of dissolved oxygen from both reservoirs (Eleyele and Asejire) and iron from Eleyele Reservoir were within the acceptable ranges.

Since the goal of cluster analysis is to make heterogeneous groups homogeneous, the cluster means were divided into five groups. This revealed that all of the groups were largely homogenous, with the exception of Total Solids, Dissolved Solids, and Suspended Solids, whose cluster means are clearly heterogeneous. The groups were similarly homogeneous for Asejire Reservoir and for Eleyele Reservoir, but Alkaline and Turbidity's cluster means

The PC findings from the PCA were those with eigenvalues greater than 1, according to established measures. Results indicate that 5 of the 13 PCs chosen to effectively explain the variance in the data were taken into consideration. The PCs accounted for 76.56% of the total variance in the Asejire Reservoir and 60.97% of the total variance in the Eleyele Reservoir. According to the PCA results, Tur, Si, TH, CaH, and Alk were the main pollutants among other parameters in Asejire Reservoir, while Tur, Si, TH, and CaH were the major pollutants among other parameters in Eleyele Reservoir. This indicates that these parameters are high in concentration and are the major pollutants among other parameters in Asejire Reservoir.

To further simplify the data structure produced by the PCA, factor analysis was used to remove the contribution of less significant variables. It discovered latent components that, for Asejire and Eleyele Reservoirs, respectively, accounted for 94.90% and 79.97% of all variance. The findings make it abundantly evident that the main pollutants in Asejire Reservoir were Alk, DS, SS, TH, CaH, PH, and Cl, whereas the main pollutants in Eleyele Reservoir were Tur, Si, Sol, and DS.

According to the GLM results, only Iron (Fe) and Silicon (Si) have a significant impact on water turbidity; other physico-chemical factors have a less significant role in describing turbidity variance in the Asejire Reservoir. Additionally, turbidity in the water level is greatly influenced by color (Col), alkalinity (Alk), silica (Si), and total solids (Sol); other physico-chemical parameters are less significant in explaining turbidity variation in Eleyele Reservoir.

Intensive farming practices, home wastes, animal waste, organic wastes, inorganic wastes, and industrial areas close to the river channel all contribute to the general pollution of both reservoirs, which is caused by anthropogenic influence and industrial activity.

In a nutshell, this study shows the value of multivariate statistical techniques for complex data analysis and visualization in water quality assessment, pollution source/factor identification, and realization of temporal/spatial variations in water quality for effective river water quality management.

Through a careful examination of the environment, efforts should be made to reduce anthropogenic influence in the reservoir (Obisesan and Oladoja [10]). Human and industrial activity in the city must be carefully controlled, and human activity near the reservoir's route must be limited. Continual public education regarding the effects of water pollution and the implementation of environmental/water management laws is also necessary.

Acknowledgements

Acknowledgements goes to Oyo State Water Corporation for providing pollution data for this study.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Oketola A., Adekolurejo S., and Osibanjo, O. (2010). Water Quality Assessment of River Ogun Using Multivariate Statistical Techniques. *Journal of Environmental Protection*, 4 (5): 466-479.

[2] UNEP (United Nations Environment Programme) (2002). Convention on Biological Diversity. United Nations. *http://www.Biodiv.Org/Convention/Articles.Asp*

[3] Nancy, D. (2009). Water Quality: Frequently Asked Questions. Florida Brooks National Marine Sanctuary, Key West, FL.

[4] WHO, (2018). Global Nutrition Report. *https://globalnutritionreport.org/0b98fb*

[5] Boyacioglu, H. (2008). Water Pollution Sources Assessment by Multivariate Statistical Methods in the Tahtali Basin, Turkey. *Environ Geol.*, 54, 275-282.

[6] Praus, P. (2007). Urban Water Quality Evaluation Using Multivariate Analysis. *Acta Montanistica Slovaca Ro?n?k*, 12(2): 150-158.

[7] Pejman A. H., Bidhendi G. R. N., Karbassi A. R., Mehrdadi N. and Bidhendi M. E. (2009). Evaluation of Spatial and Seasonal Variations in Surface Water Quality Using Multivariate Statistical Techniques. *International Journal of Environmental Science and Technology*, 6(3): 467-476.

[8] Koklu R., Sengorur B. and Topal B. (2010). Water Quality Assessment Using Multivariate Statistical Methods — A Case Study: Melen River System (Turkey). *Water Resource Management*, 24: 959-978.

[9] Obisesan K. O. and Christopher P. (2018). Statistical Models for Evaluating Water Pollution: The Case of Asejire and Eleyele Reservoirs in Nigeria. *Journal of Environmental Statistics* Volume 8, Issue 5.

[10] Obisesan, K.O. and Oladoja, O.M. (2022). On Normal Process of Diffusion Equation in Monitoring Carbon Monoxide Concentrations in Nigeria. *International Journal of Statistical Distributions and Applications*. 8(2)

ON SOME STATISTICAL PROPERTIES AND APPLICATIONS OF THREE-PARAMETER SUJATHA DISTRIBUTION

Hosenur Rahman Prodhani*
•
Department of Statistics, Assam University, Silchar, India
hosenur72@gmail.com

Rama Shanker
•
Department of Statistics, Assam University, Silchar, India
shankerrama2009@gmail.com

*Corresponding Author

Abstract

In this paper some important statistical properties of three-parameter Sujatha distribution including descriptive measures based on moments, reliability properties, mean deviations, stochastic ordering and Bonferroni and Lorenz curves have been discussed. The estimation of parameters using maximum likelihood estimation has been discussed. Finally, the goodness of fit of the distribution has been presented for two real lifetime datasets and compared with several one and two-parameter well-known lifetime distributions.

Keywords: Sujatha distribution, Extended Sujatha distributions, Statistical Properties, Estimation, Applications.

1. Introduction

Due to stochastic nature of lifetime data, it is really very challenging to search a suitable distribution to model lifetime data. The search for a suitable distribution for modeling of lifetime data is very challenging because the lifetime data are stochastic in nature. The analysis and modeling of lifetime data are essential in almost every field of knowledge including medical science, engineering, physical sciences, finance, insurance, demography, social sciences, literature, etc. and during recent eras several researchers in mathematics and statistics tried to introduce lifetime distributions. Recently, Sharma et al. [1] studied comparative study of several one parameter lifetime distributions and observed that there are some datasets which are extreme skewed to the right where these distributions were not giving well fit.

Recently, Nwikpe and Iwok [2] proposed a three-parameter generalization of Sujatha distribution (ATPSD) and studied few of its properties including hazard rate function, moment generating function, moments about origin, distribution of order statistics, and application on a dataset. The

probability density function (pdf) and the cumulative distribution function (cdf) of ATPSD are given by

$$f\left(x;\theta,\lambda,\alpha\right)=\frac{\theta^2}{2\left(\theta^2+\lambda+\alpha\right)}\left(2\theta+2\lambda x+\theta\alpha x^2\right)e^{-\theta x};\ x>0,\left(\theta,\lambda,\alpha\right)>0 \tag{1}$$

$$F\left(x;\theta,\lambda,\alpha\right)=1-\left[1+\frac{\theta^2\alpha x^2+2\theta\left(\lambda+\alpha\right)x}{2\left(\theta^2+\lambda+\alpha\right)}\right]e^{-\theta x};\ x>0,\left(\theta,\lambda,\alpha\right)>0 \tag{2}$$

The survival function of ATPSD is given by

$$S\left(x;\theta,\lambda,\alpha\right)=1-F\left(x;\theta,\lambda,\alpha\right)=\frac{\left\{\theta^2\alpha x^2+2\theta\left(\lambda+\alpha\right)x+2\left(\theta^2+\lambda+\alpha\right)\right\}e^{-\theta x}}{2\left(\theta^2+\lambda+\alpha\right)};\ x>0,\theta>0,\alpha>0 \tag{3}$$

It has been observed that there are several interesting properties of ATPSD including central moments and moments based descriptive measures, reliability properties, mean deviations, stochastic ordering and Bonferroni and Lorenz curves have not been studied. In this paper an attempt has been made to discuss these statistical properties of ATPSD and propose some areas of applications.

The distributions which are particular case of ATPSD are summarized in table 1 along with its introducers.

Table 1: *Some particular distributions of ATPSD*

Parameter values(distributions)	**Pdf of distribution**	**Introducer**
$\left(\alpha=2,\lambda=\theta\right)$ Sujatha distribution	$f\left(x,\theta\right)=\frac{\theta^3}{\theta^2+\theta+2}\left(1+x+x^2\right)e^{-\theta x}$	Shanker [3]
$\left(\alpha=0,\lambda=\theta\right)$ Lindley distribution	$f\left(x,\theta\right)=\frac{\theta^2}{\theta+1}\left(1+x\right)e^{-\theta x}$	Lindley [4]
$\left(\alpha=2,\lambda=0\right)$ Akash distribution	$f\left(x,\theta\right)=\frac{\theta^3}{\theta^2+2}\left(1+x^2\right)e^{-\theta x}$	Shanker [5]
$\left(\alpha=0,\lambda=1\right)$ Shanker distribution	$f\left(x,\theta\right)=\frac{\theta^2}{\theta^2+1}\left(\theta+x\right)e^{-\theta x}$	Shanker [6]
$\alpha=0$ (Quasi Lindley distribution)	$f\left(x,\theta\right)=\frac{\theta}{\alpha+1}\left(\alpha+\theta x\right)e^{-\theta x}$	Shanker and Mishra [7]
$\left(\alpha=0,\lambda=0\right)$ Exponential distribution	$f\left(x,\theta\right)=\theta e^{-\theta x}$	

The behavior of the pdf and the cdf of ATPSD are presented in figures 1 and 2 respectively.

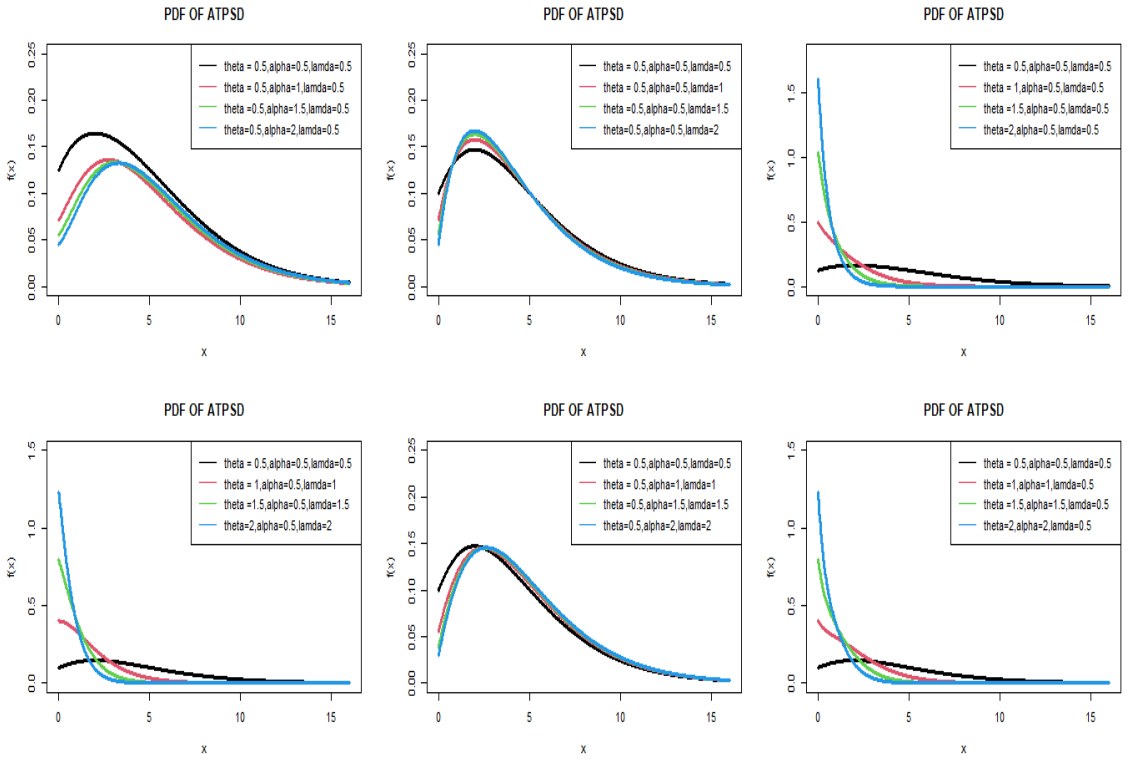


Figure 1: *The graphs of the pdf of ATPSD for different values of* θ, α *and* λ

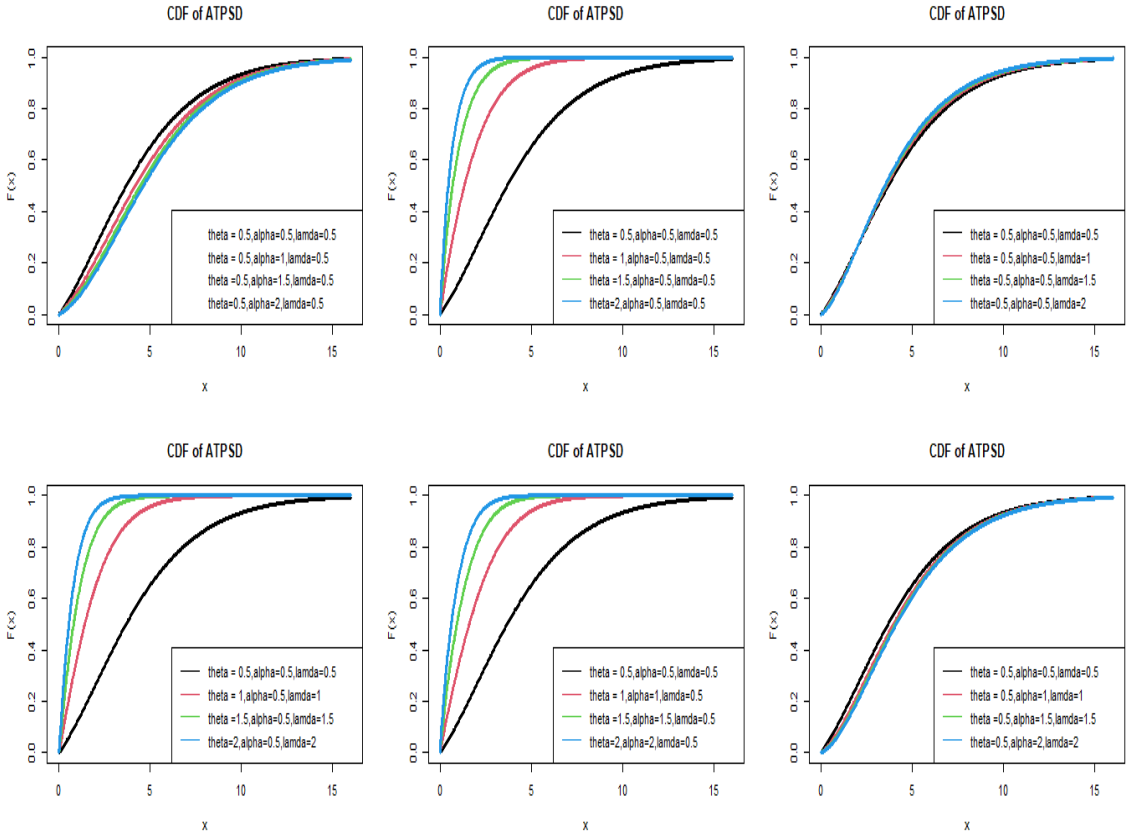


Figure 2: *The graphs of the cdf of ATPSD for different values of* θ, α *and* λ

2. Descriptive Properties Based On moments

The r th moment about origin, μ_r', of ATPSD is given by

$$\mu_r' = \frac{r!\left\{2\theta^2 + 2(r+1)\lambda + (r+1)(r+2)\alpha\right\}}{2\theta^r\left(\theta^2+\lambda+\alpha\right)}; r = 1, 2, 3, \cdots \tag{4}$$

Thus, the first four moments about origin are obtained as

$$\mu_1' = \frac{\theta^2+2\lambda+3\alpha}{\theta\left(\theta^2+\lambda+\alpha\right)}, \mu_2' = \frac{2\left(\theta^2+3\lambda+6\alpha\right)}{\theta^2\left(\theta^2+\lambda+\alpha\right)}, \mu_3' = \frac{6\left(\theta^2+4\lambda+10\alpha\right)}{\theta^3\left(\theta^2+\lambda+\alpha\right)}, \mu_4' = \frac{24\left(\theta^2+5\lambda+15\alpha\right)}{\theta^4\left(\theta^2+\lambda+\alpha\right)}.$$

Now using the relationship between moments about mean and the moments about origin, the moments about the mean of ATPSD are obtained as

$$\mu_2 = \frac{\theta^4+(8\alpha+4\lambda)\theta^2+\left(2\lambda^2+3\alpha^2+6\lambda\alpha\right)}{\theta^2\left(\theta^2+\lambda+\alpha\right)^2}$$

$$\mu_3 = \frac{2\left\{\theta^6+(15\alpha+6\lambda)\theta^4+\left(9\alpha^2+6\lambda^2+21\lambda\alpha\right)\theta^2+\left(2\lambda^3+3\alpha^3+9\lambda^2\alpha+9\lambda\alpha^2\right)\right\}}{\theta^3\left(\theta^2+\lambda+\alpha\right)^3}$$

$$\mu_4 = \frac{3\left\{\begin{array}{l}3\theta^8+(64\alpha+24\lambda)\theta^6+\left(102\alpha^2+172\lambda\alpha+44\lambda^2\right)\theta^4\\+\left(72\alpha^3+32\lambda^3+160\lambda^2\alpha+192\lambda\alpha^2\right)\theta^2+\left(8\lambda^4+15\alpha^4+48\lambda^3\alpha+60\lambda\alpha^3+84\lambda^2\alpha^2\right)\end{array}\right\}}{\theta^4\left(\theta^2+\lambda+\alpha\right)^4}$$

The coefficients of variation (C.V), skewness $\left(\sqrt{\beta_1}\right)$, kurtosis $\left(\beta_2\right)$ and index of dispersion $\left(\gamma\right)$ of ATPSD are thus obtained as

$$C.V = \frac{\sqrt{\mu_2}}{\mu_1'} = \frac{\sqrt{\theta^4+(8\alpha+4\lambda)\theta^2+\left(2\lambda^2+3\alpha^2+6\lambda\alpha\right)}}{\theta^2+2\lambda+3\alpha}$$

$$\sqrt{\beta_1} = \frac{\mu_3}{\left(\mu_2\right)^{3/2}} = \frac{2\left\{\theta^6+(15\alpha+6\lambda)\theta^4+\left(9\alpha^2+6\lambda^2+21\lambda\alpha\right)\theta^2+\left(2\lambda^3+3\alpha^3+9\lambda^2\alpha+9\lambda\alpha^2\right)\right\}}{\left\{\theta^4+(8\alpha+4\lambda)\theta^2+\left(2\lambda^2+3\alpha^2+6\lambda\alpha\right)\right\}^{3/2}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\left\{\begin{array}{l}3\theta^8+(64\alpha+24\lambda)\theta^6+\left(102\alpha^2+172\lambda\alpha+44\lambda^2\right)\theta^4\\+\left(72\alpha^3+32\lambda^3+160\lambda^2\alpha+192\lambda\alpha^2\right)\theta^2+\left(8\lambda^4+15\alpha^4+48\lambda^3\alpha+60\lambda\alpha^3+84\lambda^2\alpha^2\right)\end{array}\right\}}{\left\{\theta^4+(8\alpha+4\lambda)\theta^2+\left(2\lambda^2+3\alpha^2+6\lambda\alpha\right)\right\}^2}$$

$$\gamma = \frac{\sigma^2}{\mu_1'} = \frac{\theta^4 + 4(2\alpha + \lambda)\theta^2 + (2\lambda^2 + 3\alpha^2 + 6\alpha\lambda)}{\theta(\theta^2 + \lambda + \alpha)(\theta^2 + 2\lambda + 3\alpha)}$$

The nature of the coefficient of variation, skewness, kurtosis and index of dispersion of ATPSD are shown graphically in figure 3.

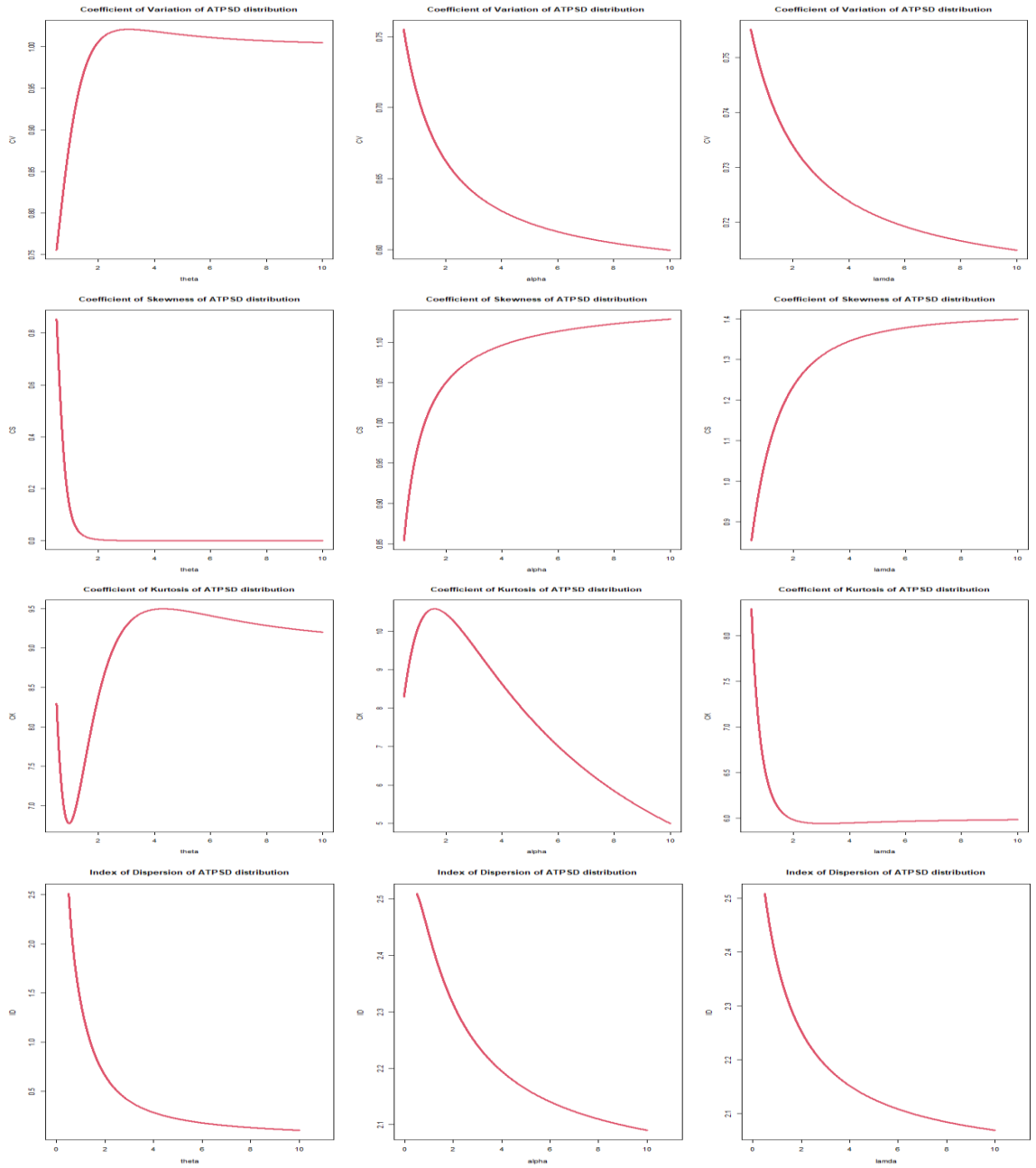


Figure 3: *The nature of the coefficient of variation, skewness, kurtosis and index of dispersion of ATPSD*

When α and λ are fixed and θ increases, the value of the CV is increases till $\theta \leq 3$ and when $\theta > 3$, then CV starts decreasing slowly increasing values of θ. When θ and λ is fixed, then CV decreases for increasing values of α. Similarly, for, fixed values of θ and α and increasing values of λ, CV decreases.

When α and λ are fixed and θ increases, skewness decreases speedily till $\theta \le 1$ and when $\theta > 1$, it becomes constant. When θ and λ is fixed, then skewness decreases for increasing values of α. Similarly, for, fixed values of θ and α and increasing values of λ, skewness increases.

For fixed values of (α, λ) and increasing values of θ, the kurtosis is decreasing, increasing and again decreasing. For fixed values of (θ, λ) and increasing values of α, the kurtosis is increasing and then decreasing. And for fixed values of (θ, α) and increasing values of λ, the kurtosis is decreasing speedily till $\lambda \le 2$ and for $\lambda > 2$, it starts increasing very slowly.

For the nature of index of dispersion, it is always decreasing for increasing values of one parameter and fixed values of another two-parameter.

3. Reliability Properties

3.1. Hazard Rate Function

The hazard rate function of ATPSD is obtained as

$$h(x;\theta,\lambda,\alpha) = \frac{f(x;\theta,\lambda,\alpha)}{S(x;\theta,\lambda,\alpha)} = \frac{\theta^2\left(2\theta + 2\lambda x + \theta\alpha x^2\right)}{\theta^2\alpha x^2 + 2\theta(\lambda+\alpha)x + 2\left(\theta^2+\lambda+\alpha\right)}; x > 0, (\theta,\lambda,\alpha) > 0 \qquad (5)$$

It can be easily verified that $h(0;\theta,\lambda,\alpha) = f(0;\theta,\lambda,\alpha) = \frac{f(x;\theta,\lambda,\alpha)}{S(x;\theta,\lambda,\alpha)} = \frac{\theta^3}{\left(\theta^2+\lambda+\alpha\right)}$.

The behaviors of the hazard rate function of ATPSD are explained in the following figure 4.

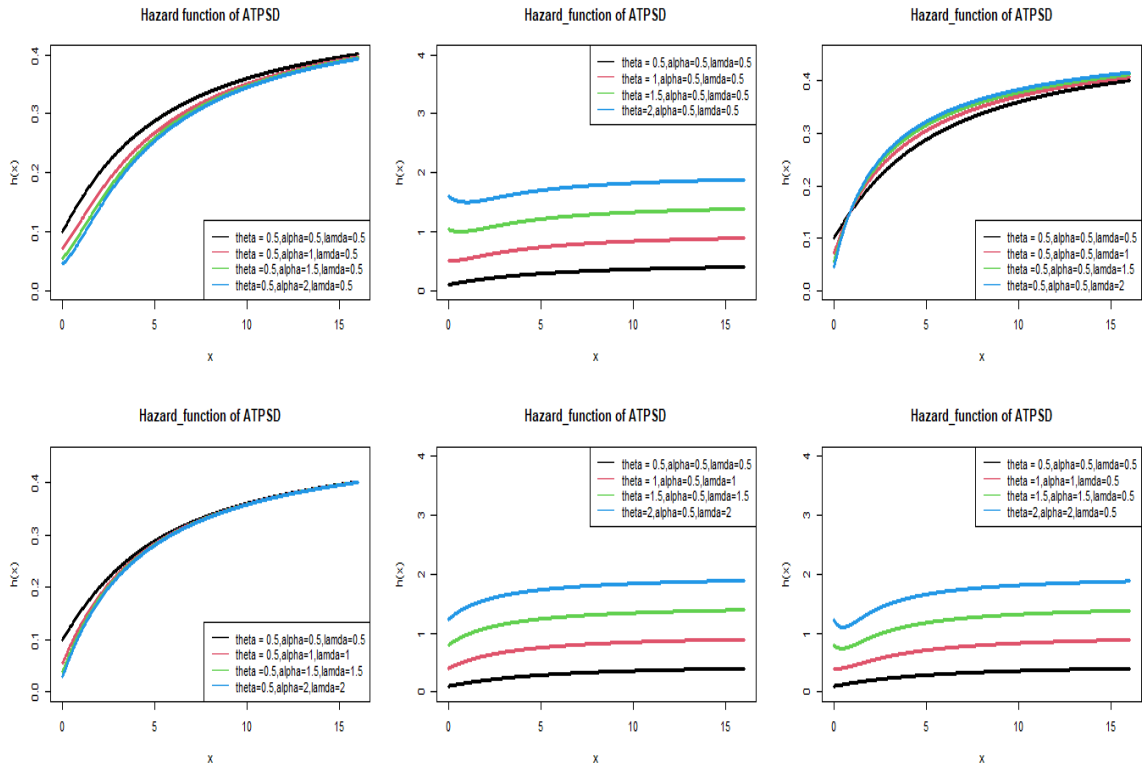


Figure 4: *The graphs of the hazard rate function of ATPSD for different values of* θ, α *and* λ

3.2. Mean Residual Life Function

The mean residual life function of ATPSD can be obtained as

$$m(x;\theta,\lambda,\alpha)=\frac{1}{1-F(x;\theta,\lambda,\alpha)}\int_x^\infty\left[1-F(t;\theta,\lambda,\alpha)\right]dt$$

$$=\frac{1}{\left\{\theta^2\alpha x^2+2\theta(\lambda+\alpha)x+2\left(\theta^2+\lambda+\alpha\right)\right\}e^{-\theta x}}\int_x^\infty\left[\theta^2\alpha t^2+2\theta(\lambda+\alpha)t+2\left(\theta^2+\lambda+\alpha\right)\right]e^{-\theta t}dt$$

$$=\frac{\alpha\left(\theta^2x^2+2\theta x+2\right)+2(\lambda+\alpha)(\theta x+1)+2\left(\theta^2+\lambda+\alpha\right)}{\theta\left[\theta^2\alpha x^2+2\theta(\lambda+\alpha)x+2\left(\theta^2+\lambda+\alpha\right)\right]} \tag{6}$$

It can be easily verified that $m(0;\theta,\lambda,\alpha)=\dfrac{\theta^2+2\lambda+3\alpha}{\theta\left(\theta^2+\lambda+\alpha\right)}=\mu_1'$. The behavior of mean residual life function is explained in the following figure 5.

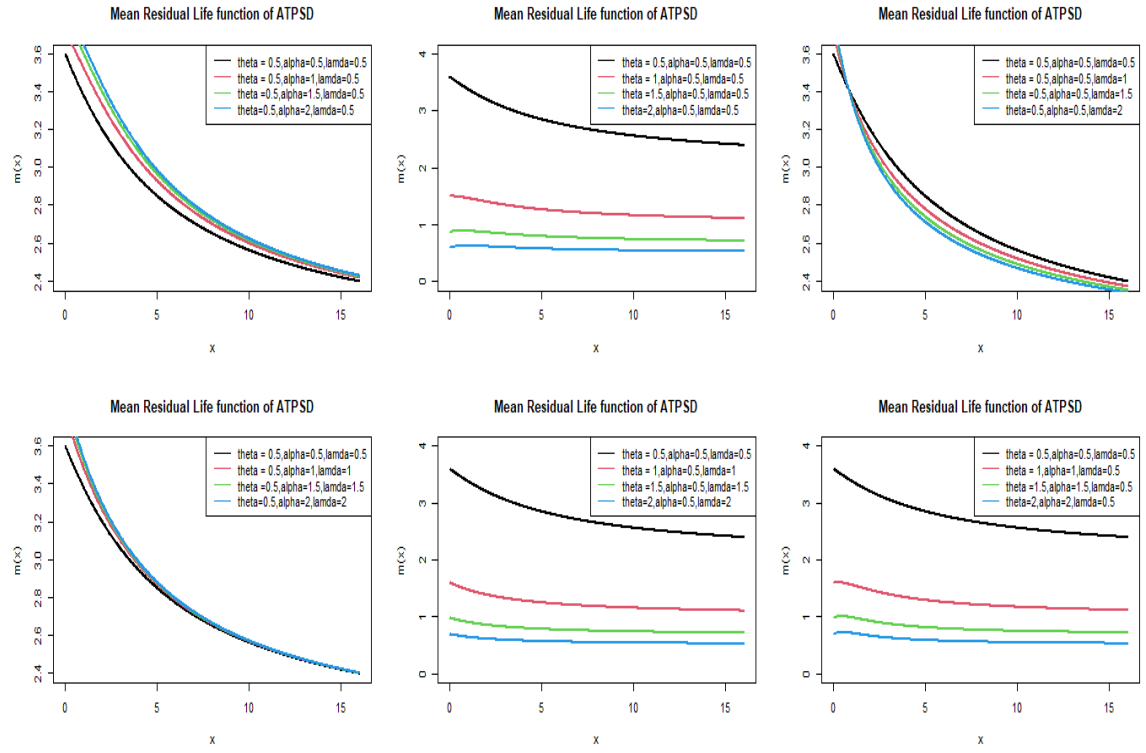


Figure 5: *The graphs of the mean residual life function of ATPSD for different values of* θ,α *and* λ

4. Stochastic Ordering

In probability theory and statistics, a stochastic order quantifies the concept of one random variable being bigger than another. A random variable X is said to be smaller than a random variable Y in the:

i. Stochastic order $\left(X\leq_{st}Y\right)$ if $F_X(x)\geq F_Y(y)$ for all x

ii. Hazard rate order $\left(X \leq_{hr} Y\right)$ if $h_X(x) \geq h_Y(y)$ for all x

iii. Mean residual life order $\left(X \leq_{mrl} Y\right)$ if $m_X(x) \geq m_Y(y)$ for all x

iv. Likelihood ratio order $\left(X \leq_{lr} Y\right)$ if $\frac{f_X(x)}{f_Y(y)}$ decrease in x

The following results due to Shaked and Shantikumar [8] are well known for establishing stochastic ordering of distributions

$$X <_{lr} Y \Rightarrow X <_{hr} Y \Rightarrow X <_{mrl} Y$$
$$\Downarrow$$
$$X <_{st} Y$$

Theorem: Let $X \sim$ ATPSD $\left(\theta_1, \lambda_1, \alpha_1\right)$ and $Y \sim$ ATPSD $\left(\theta_2, \lambda_2, \alpha_2\right)$. If $\lambda_1 = \lambda_2, \alpha_1 = \alpha_2$ and $\theta_1 > \theta_2$ or $\theta_1 = \theta_2, \alpha_1 = \alpha_2$ and $\lambda_1 < \lambda_2$ or $\theta_1 = \theta_2, \lambda_1 = \lambda_2$ and $\alpha_1 < \alpha_2$ then $X <_{lr} Y$ hence $X <_{hr} Y$, $X <_{mrl} Y$ and $X <_{st} Y$.

Proof: We have

$$\frac{f_X(x)}{f_Y(x)} = \frac{\theta_1^3(\theta_2^2 + \lambda_2 + \alpha_2)}{\theta_2^3(\theta_1^2 + \lambda_1 + \alpha_1)}\left(\frac{\theta_1\alpha_1 x^2 + 2\lambda_1 x + 2\theta_1}{\theta_2\alpha_2 x^2 + 2\lambda_2 x + 2\theta_2}\right)e^{-\left(\theta_1 - \theta_2\right)x}; x > 0$$

Now

$$\log\left(\frac{f_X(x)}{f_Y(x)}\right) = \log\left(\frac{\theta_1^3(\theta_2^2 + \lambda_2 + \alpha_2)}{\theta_2^3(\theta_1^2 + \lambda_1 + \alpha_1)}\right) + \log\left(\frac{\theta_1\alpha_1 x^2 + 2\lambda_1 x + 2\theta_1}{\theta_2\alpha_2 x^2 + 2\lambda_2 x + 2\theta_2}\right) - \left(\theta_1 - \theta_2\right)x$$

Therefore

$$\frac{d}{dx}\left(\log\frac{f_X(x)}{f_Y(x)}\right) = \frac{2\left(\alpha_1\theta_1\lambda_2 - \alpha_2\theta_2\lambda_1\right)x^2 + 4\theta_1\theta_2(\alpha_1 - \alpha_2)x + 4(\lambda_1\theta_2 - \lambda_2\theta_1)}{\left(\theta_1\alpha_1 x^2 + 2\lambda_1 x + 2\theta_1\right)\left(\theta_2\alpha_2 x^2 + 2\lambda_2 x + 2\theta_2\right)} - \left(\theta_1 - \theta_2\right)$$

Thus, if $\lambda_1 = \lambda_2, \alpha_1 = \alpha_2$ and $\theta_1 > \theta_2$ or $\theta_1 = \theta_2, \alpha_1 = \alpha_2$ and $\lambda_1 < \lambda_2$ or $\theta_1 = \theta_2, \lambda_1 = \lambda_2$ and $\alpha_1 < \alpha_2$, $\frac{d}{dx}\left(\log\frac{f_X(x)}{f_Y(x)}\right) < 0$. This means that $X <_{lr} Y$ and hence $X <_{hr} Y$, $X <_{mrl} Y$ and $X <_{st} Y$.

5. Deviation from Mean and Median

The amount of dispersion in a population is an evidently measured to some extent by the totality of deviations from the mean and median. These are known as the mean deviation about the mean and mean deviation about median and are defined by

$$\delta_1(x) = \int_0^\infty |x - \mu| f(x)dx \text{ and } \delta_2(x) = \int_0^\infty |x - M| f(x)dx \text{ respectively,}$$

where $\mu = E(X)$ and $M = Median(X)$.

The measures $\delta_1(x)$ and $\delta_2(x)$ can be calculated using the following relationships

$$\delta_1(x) = 2\mu F(\mu) - 2\int_0^{\mu} xf(x)dx$$

and $$\delta_2(x) = -\mu + 2\int_M^{\infty} xf(x)dx$$

Thus, the mean deviation about the mean $\delta_1(x)$, and the mean deviation about the median $\delta_2(x)$ of ATPSD are obtained as

$$\delta_1(x) = \frac{\left[\alpha\theta^2\mu^2 + 2\theta\lambda\mu + 4\alpha\theta\mu + 2\theta^2 + 4\lambda + 6\alpha\right]e^{-\theta\mu}}{\theta(\theta^2 + \lambda + \alpha)} \tag{7}$$

$$\delta_2(x) = \frac{2\left[\alpha\theta^3M^3 + \theta^2(2\lambda + 3\alpha)\mathrm{M}^2 + (2\theta^3 + 4\theta\lambda + 6\alpha\theta)M + 2\theta^2 + 4\lambda + 6\alpha\right]e^{-\theta M}}{2\theta(\theta^2 + \lambda + \alpha)} - \mu \tag{8}$$

6. Bonferroni and Lorenz Curves and Indices

The Bonferroni and Lorenz curves by Bonferroni [9] and Bonferroni and Gini indices have wide applications in economics to study income and poverty, but it also used in other fields like reliability, demography, insurance and medicine. The Bonferroni and Lorenz curves are defined as

$$\mathrm{B(p)} = \frac{1}{p\mu}\int_0^q x\,f(x)\,dx = \frac{1}{p\mu}\left[\int_0^{\infty} x\,f(x)dx - \int_q^{\infty} x\,f(x)\,dx\right] = \frac{1}{p\mu}\left[\mu - \int_q^{\infty} x\,f(x)\,dx\right]$$

and $$\mathrm{L(p)} = \frac{1}{\mu}\int_0^q x\,f(x)\,dx = \frac{1}{\mu}\left[\int_0^{\infty} x\,f(x)\,dx - \int_q^{\infty} x\,f(x)\,dx\right] = \frac{1}{\mu}\left[\mu - \int_q^{\infty} x\,f(x)\,dx\right]$$

respectively.

The Bonferroni and Gini indices are obtained as

$B = 1 - \int_0^1 B(p)\,dp$ and $G = 1 - 2\int_0^1 \mathrm{L}(p)\,dp$, respectively.

Using pdf of ATPSD and little algebraic simplification, we get

$$B(p) = \frac{1}{p}\left[1 - \frac{\left[\alpha\theta^3q^3 + \theta^2(2\lambda + 3\alpha)\mathrm{q}^2 + (2\theta^3 + 4\theta\lambda + 6\alpha\theta)\mathrm{q} + 2\theta^2 + 4\lambda + 6\alpha\right]e^{-\theta q}}{2(\theta^2 + 2\lambda + 3\alpha)}\right] \tag{9}$$

$$\mathrm{L}(p) = 1 - \frac{\left[\alpha\theta^3q^3 + \theta^2(2\lambda + 3\alpha)\mathrm{q}^2 + (2\theta^3 + 4\theta\lambda + 6\alpha\theta)\mathrm{q} + 2\theta^2 + 4\lambda + 6\alpha\right]e^{-\theta q}}{2(\theta^2 + 2\lambda + 3\alpha)} \tag{10}$$

Now using equations and after some simple algebraic simplifications, the Bonferroni and Gini indices of ATPSD are obtained as

$$B = 1 - \frac{\left[\alpha\theta^3q^3 + \theta^2(2\lambda + 3\alpha)\mathrm{q}^2 + (2\theta^3 + 4\theta\lambda + 6\alpha\theta)\mathrm{q} + 2\theta^2 + 4\lambda + 6\alpha\right]e^{-\theta q}}{2(\theta^2 + 2\lambda + 3\alpha)} \tag{11}$$

$$G = \frac{2\left[\alpha\theta^3q^3 + \theta^2(2\lambda + 3\alpha)\mathrm{q}^2 + (2\theta^3 + 4\theta\lambda + 6\alpha\theta)\mathrm{q} + 2\theta^2 + 4\lambda + 6\alpha\right]e^{-\theta q}}{2(\theta^2 + 2\lambda + 3\alpha)} - 1. \tag{12}$$

7. Maximum Likelihood Estimation

Let $\left(x_1, x_2, x_3, ..., x_n\right)$ be a random sample of size n from ATPSD$\left(\theta, \lambda, \alpha\right)$. Then the likelihood function is given by

$$L = \left(\frac{\theta^2}{2\left(\theta^2 + \lambda + \alpha\right)}\right)^n \prod_{i=1}^{n}\left(2\theta + 2\lambda x_i + \theta\alpha\, x_i^2\right)e^{-n\,\theta\,\bar{x}}$$, where $\bar{x}$ is the sample mean.

The log-likelihood function is thus obtained as

$$\log L = n\left[2\log\theta - \log 2 - \log\left(\theta^2 + \lambda + \alpha\right)\right] + \sum_{i=1}^{n}\log\left(2\theta + 2\lambda x_i + \theta\alpha\, x_i^2\right) - n\theta\bar{x}\,. \qquad (13)$$

The maximum likelihood estimates $\left(\hat{\theta}, \hat{\lambda}, \hat{\alpha}\right)$ of parameters $\left(\theta, \lambda, \alpha\right)$ are the solution of the following log-likelihood equations

$$\frac{\partial \log L}{\partial \theta} = \frac{2n}{\theta} + \frac{2n\theta}{\theta^2 + \lambda + \alpha} + \sum_{i=1}^{n}\frac{2 + \alpha x_i^2}{2\theta + 2\lambda x_i + \theta\alpha x_i^2} - n\,\bar{x} = 0$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{-n}{\theta^2 + \lambda + \alpha} + \sum_{i=1}^{n}\frac{2x_i}{2\theta + 2\lambda x_i + \theta\alpha x_i^2} - n\,\bar{x} = 0$$

$$\frac{\partial \log L}{\partial \alpha} = \frac{-n}{\theta^2 + \lambda + \alpha} + \sum_{i=1}^{n}\frac{\theta x_i^2}{2\theta + 2\lambda x_i + \theta\alpha x_i^2} = 0$$

These three log-likelihood equations do not seem to be solved directly. We have to use Fisher's scoring method for solving these three log-likelihood equations. We have

$$\frac{\partial^2 \log L}{\partial \theta^2} = \frac{-2n}{\theta^2} + \frac{2n\left(\theta^2 - \lambda - \alpha\right)}{\left(\theta^2 + \lambda + \alpha\right)^2} - \sum_{i=1}^{n}\frac{\left(2 + \alpha x_i^2\right)}{\left(2\theta + 2\lambda x_i + \theta\alpha x_i^2\right)^2}$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \frac{n}{\left(\theta^2 + \lambda + \alpha\right)^2} - \sum_{i=1}^{n}\frac{4x_i^2}{\left(2\theta + 2\lambda x_i + \theta\alpha x_i^2\right)^2}$$

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \frac{n}{\left(\theta^2 + \lambda + \alpha\right)^2} - \sum_{i=1}^{n}\frac{\theta^2 x_i^4}{\left(2\theta + 2\lambda x_i + \theta\alpha x_i^2\right)^2}$$

$$\frac{\partial^2 \log L}{\partial \theta \partial \lambda} = \frac{2n\theta}{\left(\theta^2 + \lambda + \alpha\right)^2} - \sum_{i=1}^{n}\frac{4x_i + 2\alpha x_i^3}{\left(2\theta + 2\lambda x_i + \theta\alpha x_i^2\right)^2}$$

$$\frac{\partial^2 \log L}{\partial \theta \partial \alpha} = \frac{2n\theta}{\left(\theta^2 + \lambda + \alpha\right)^2} + \sum_{i=1}^{n}\frac{2\lambda x_i^2}{\left(2\theta + 2\lambda x_i + \theta\alpha x_i^2\right)^2}$$

$$\frac{\partial^2 \log L}{\partial \theta \partial \alpha} = \frac{-12n\theta}{\left(\alpha\theta^2 + 6\right)^2} - \sum_{i=1}^{n}\frac{x_i^3}{\left(\alpha + \theta x_i^3\right)^2} = \frac{\partial^2 \log L}{\partial \alpha \partial \theta}\,.$$

$$\frac{\partial^2 \log L}{\partial\lambda\,\partial\alpha} = \frac{n}{\left(\theta^2+\lambda+\alpha\right)^2} - \sum_{i=1}^{n} \frac{2\theta x_i^3}{\left(2\theta + 2\lambda x_i + \theta\alpha x_i^2\right)^2} = \frac{\partial^2 \log L}{\partial\alpha\,\partial\lambda}$$

The following equations can be solved for MLEs $\left(\hat{\theta},\hat{\lambda},\hat{\alpha}\right)$ of $\left(\theta,\lambda,\alpha\right)$ for ATPSD

$$\begin{bmatrix} \frac{\partial^2 \log L}{\partial\theta^2} & \frac{\partial^2 \log L}{\partial\theta\,\partial\lambda} & \frac{\partial^2 \log L}{\partial\theta\,\partial\alpha} \\ \frac{\partial^2 \log L}{\partial\lambda\,\partial\theta} & \frac{\partial^2 \log L}{\partial\lambda^2} & \frac{\partial^2 \log L}{\partial\lambda\,\partial\alpha} \\ \frac{\partial^2 \log L}{\partial\alpha\,\partial\theta} & \frac{\partial^2 \log L}{\partial\alpha\,\partial\lambda} & \frac{\partial^2 \log L}{\partial\alpha^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\lambda}=\lambda_0 \\ \hat{\alpha}=\alpha_0}} \begin{bmatrix} \hat{\theta}-\theta_0 \\ \hat{\lambda}-\lambda_0 \\ \hat{\alpha}-\alpha_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \log L}{\partial\theta} \\ \frac{\partial \log L}{\partial\lambda} \\ \frac{\partial \log L}{\partial\alpha} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\lambda}=\lambda_0 \\ \hat{\alpha}=\alpha_0}} \tag{14}$$

where $\left(\theta_0,\lambda_0,\alpha_0\right)$ are the initial values of $\left(\theta,\lambda,\alpha\right)$ respectively. These equations are solved iteratively till sufficiently close values of $\left(\hat{\theta},\hat{\lambda},\hat{\alpha}\right)$ are obtained.

8. Applications to Lifetime Data

The following real lifetime datasets have been considered for testing the goodness of fit of ATPSD over other one parameter and two-parameter lifetime distributions.

Data Set 1: The real data discussed by Almongy et al [10] that represents a COVID 19 mortality rate data belongs to Mexico of 108 days that is recorded from 4 March to 20 July 2020. This data formed of rough mortality rate. The data are as follows:
8.826, 6.105, 10.383, 7.267, 13.220, 6.015, 10.855, 6.122, 10.685, 10.035, 5.242, 7.630, 14.604, 7.903, 6.327, 9.391, 14.962, 4.730, 3.215, 16.498, 11.665, 9.284, 12.878, 6.656, 3.440, 5.854, 8.813, 10.043, 7.260, 5.985, 4.424, 4.344, 5.143, 9.935, 7.840, 9.550, 6.968, 6.370, 3.537, 3.286, 10.158,8.108,6.697, 7.151, 6.560, 2.988, 3.336, 6.814, 8.325, 7.854, 8.551, 3.228, 3.499, 3.751, 7.486, 6.625, 6.140, 4.909, 4.661, 1.867, 2.838, 5.392, 12.042, 8.696, 6.412, 3.395, 1.815, 3.327, 5.406, 6.182,4.949, 4.089, 3.359, 2.070, 3.298, 5.317, 5.442, 4.557, 4.292, 2.500, 6.535, 4.648, 4.697, 5.459, 4.120, 3.922, 3.219, 1.402, 2.438, 3.257, 3.632, 3.233, 3.027, 2.352, 1.205, 2.077, 3.778, 3.218, 2.926, 2.601, 2.065, 1.041, 1.800, 3.029, 2.058, 2.326, 2.506, 1.923.

Data set-2: The following bi-modal dataset, discussed by Ghitany et al. [11], is obtained from the banking sector discusses the waiting time (in minutes) before the customer received service in a bank. The values are:
0.8, 0.8, 1.3, 1.5, 1.8, 1.9, 1.9, 2.1, 2.6, 2.7, 2.9, 3.1, 3.2, 3.3, 3.5, 3.6, 4.0, 4.1, 4.2, 4.2, 4.3,4.3, 4.4, 4.4, 4.6, 4.7, 4.7,4.8, 4.9, 4.9, 5.0, 5.3, 5.5, 5.7, 5.7, 6.1, 6.2, 6.2, 6.2, 6.3, 6.7, 6.9,7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8.0, 8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6, 9.7, 9.8, 10.7,10.9, 11.0, 11.0, 11.1 ,11.2, 11.2, 11.5,11.9, 12.4, 12.5, 12.9, 13.0, 13.1,13.3, 13.6, 13.7,13.9,14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0, 19.9, 20.6, 21.3, 21.4, 21.9,23.0,27.0, 31.6, 33.1, 38.5.

In order to compare lifetime distributions, values of $-2\log L$, AIC (Akaike Information Criterion), AICC (Akaike Information Criterion Corrected), K-S Statistics (Kolmogorov-Smirnov Statistics)

and the corresponding probability value (p-value) for the above data set has been computed. The formulae for computing AIC, AICC and K-S Statistics are as follows:

$$AIC = -2logL + 2k \text{ , } AICC = AIC + \frac{2k(k+1)}{n-k-1} \text{ , } D = \underset{x}{Sup} \mid F_n(x) - F_0(x) \mid$$

where $k =$ number of parameter, $n =$ sample size

The distribution corresponding to the lower values of $-2\log L$, AIC, AICC, and K-S Statistics is the best fit distribution. These statistical values for the two datasets have been computed and presented in tables 2 and 3 respectively. It is obvious from the goodness of fit of distributions given in tables 2 and 3 that ATPSD gives much closure fit as compared to other one parameter and two-parameter distributions and hence it can be considered as a suitable model for the given dataset.

Table 2: *ML estimates, -2logL , AIC, AICC, K-S value and p-value of the distribution for the data set-1*

Distributions	MLE			-2logL	AIC	AICC	K-S	p-value
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\lambda}$					
ATPSD	0.5209	2277.6180	0.1000	**533.29**	**539.29**	**539.52**	**0.0584**	**0.9246**
TPSD	0.4867	0.0100	…	536.45	540.45	540.56	0.0682	0.8029
NTPSD	0.4842	0.0100	…	542.75	546.75	546.86	0.0869	0.5153
ANTPSD	0.4869	931.2583		536.37	540.37	540.48	0.0684	0.7975
QSD	0.4825	0.1000	…	537.97	541.97	542.08	0.0671	0.8205
NQSD	0.4868	137.8985	…	536.39	540.39	540.50	0.0626	0.8758
WSD	0.9828	3.7285	…	510.77	514.77	515.47	0.0845	0.5561
PSD	0.3491	1.1648	…	537.06	541.06	541.76	0.0937	0.4454
Sujatha	0.4631	…	…	543.36	545.36	545.39	0.0950	0.3961

Table 3: *ML estimates, -2logL , AIC, AICC, K-S value and p-value of the distribution for the data set-2*

Distributions	MLE			-2logL	AIC	AICC	K-S	p-value
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\lambda}$					
ATPSD	0.2025	0.1270	108.8797	**634.60**	**640.6**	**640.85**	**0.0564**	**0.9539**
TPSD	0.2769	2.4379	…	639.25	643.25	644.05	0.0764	0.7146
NTPSD	0.2316	20.3400	…	635.03	639.03	639.15	0.0699	0.8086
ANTPSD	0.2769	0.4102	…	639.25	643.25	644.05	0.0750	0.7361
QSD	0.2769	0.6752	…	639.25	643.25	644.05	0.0897	0.5133
NQSD	0.2769	0.1136	…	639.25	643.25	644.05	0.0908	0.4959
WSD	0.1958	0.0100	…	602.44	606.44	607.24	0.1079	0.2627
PSD	0.3571	0.9012	…	636.48	640.48	641.28	0.0682	0.8206
Sujatha	0.2846	…	…	639.63	641.63	641.88	0.0949	0.4447

The fitted plot of the considered distributions of the data set-1 and data set-2 are presented in figure 6.

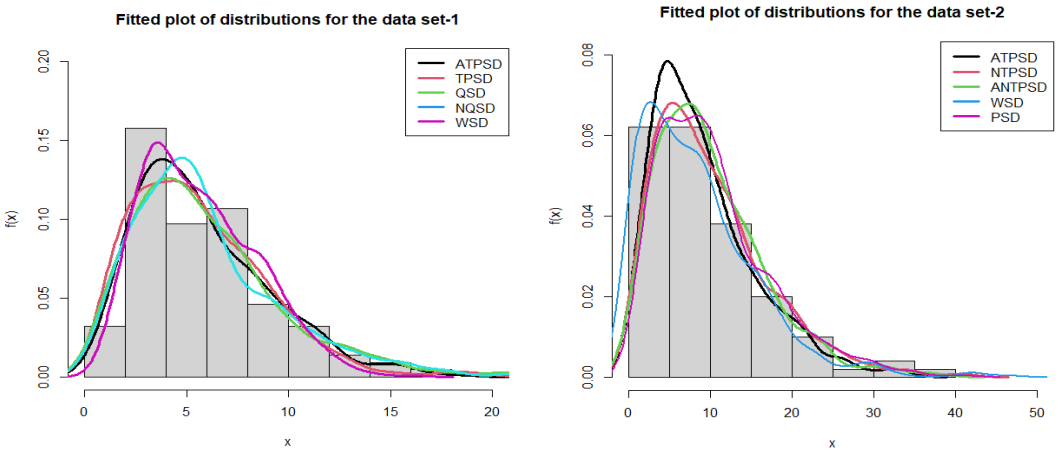


Figure 6: *Fitted plot of distributions of the data set-1 and data set-2.*

9. Conclusion

Some important and useful statistical properties of a three-parameter Sujatha distribution (ATPSD) including descriptive measures based on moments, reliability properties, mean deviations, stochastic ordering and Bonferroni and Lorenz curves have been derived and discussed. Maximum likelihood estimation has been discussed for estimating the parameters. Applications and goodness of fit of the ATPSD have been demonstrated with two real lifetime datasets and it shows better fit over several one parameter and two-parameter lifetime distributions.

10. Conflict of Interest

The Authors declare that there is no conflict of interest.

References

[1] Sharma, V., Shanker, R. and Shanker, R. (2019). On some one parameter Lifetime distributions and Their Applications, *Annals of Biostatistics & Biometrics Application*, 1(2):1 – 6.

[2] Nwikpe, B.J. and Iwok, I. A. (2020). A three-parameter Sujatha distribution with application, *European Journal of Statistics and Probability*, 8(3):22 – 34.

[3] Shanker, R. (2016). Sujatha Distribution and Its Applications**,** *Statistics in Transition-New Series*, 17(3):391- 410.

[4] Lindley, D.V. (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B*, 20(1):102 – 107.

[5] Shanker, R. (2015). Akash Distribution and Its Applications, *International Journal of Probability and Statistics*, 4 (3):65 – 75.

[6] Shanker, R. (2015). Shanker Distribution and Its Applications, *International Journal of Statistics and Applications*, 5 (6):338 – 348.

[7] Shanker, R. and Mishra, A. (2013). A quasi Lindley distribution, *African Journal of Mathematics and Computer Science Research (AJMCSR)*, 6 (4):64 – 71.

[8] Shaked, M. and Shanthikumar, J. (1994). Stochastic Orders and Their Applications. Academic Press, New York.

[9] Bonferroni, C.E. (1936). Teoria statistica delle classi e calcolo delle probability [Statistical class theory and calculation of probability], *Publicazioni del R Istituto Supriore di Scienze Economiche e Commerciali diFirenze, 8:*3-62.

[10] Almongy, H.M., Almetwally, E.M., Aljohani, H.M., Alghamdi, A.S. and Hafez, E.H. (2021). A new extended Rayleigh distribution with applications of COVID-19 data, *Results Physics*. 23:1-8.

[11] Ghitany, M.E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application, *Mathematics and Computers in Simulation*, 78:493–506.

Analysis of $MMAP/PH/1$ Classical Retrial Queue with Non-preemptive priority, Second optional service, Differentiate breakdowns, Phase type repair, Single vacation, Emergency vacation, Closedown, Setup and Discouragement

G. Ayyappan[1], G. Archana @ Gurulakshmi[2]

•

[1,2]Department of Mathematics,
Puducherry Technological University,
Puducherry, India.
[1]ayyappan@ptuniv.edu.in, [2]archanagurulakshmi@gmail.com

Abstract

A single server retrial queueing model with non-preemptive priority was examined in this research. The arrival of priority consumers follow a marked Markovian arrival pattern, and both high priority and low priority service times are according to phase type distribution. Matrix analytic method are used to examine the steady state analysis of this model. Various system performance measures, cost analysis and busy period analysis also examined in this model. In additionally, by using some system performance measures we provide the numerical illustration with numerically and graphically.

Keywords: Markovian arrival process, Priority queue, Phase type Repair, Second optional service, Differentiate breakdown, Closedown, Single vacation, Emergency vacation, Setup, Balking.

AMS Subject Classification (2010): **60K25, 68M30, 90B22** .

1. Introduction

Retrial queues in queueing theory have gained attention recently as a significant topic of study as a result of its numerous applications. System manufacturing, designing of local area communication networks and data communication networks are the most common examples. The customers are bound to be impatient in general. From the real-life experience, we can observe that the customers who require service must form a queue. However, some customers decide not to wait in queue due to time restrictions, and some customers who do wait in queue get impatient and leave the queue before receiving service.

The Markovian Arrival Process (MAP) is considered to be the most significant process tool in this theory. Neuts [26] pioneered the Versatile Markovian Point Process (VMPP). He has used the concept of point process which is Markovian arrival process. Chakravarthy [10] have analysed the MAP which is represented by n-dimensional parameter matrices (D_0, D_1) where D_0 governing the transition for no arrival and D_1 governing arrivals.

There are two types of priority services such as preemptive and non-preemptive. The arrival of priority customers have to wait until the regular customers service completed such as non-preemptive priority. The low priority consumers should be interrupted by the preemptive priority, also known as the high priority customers. Isotupa and Stanford [17] looked into a single server queue that takes connections that arrive from N classes of clients in a non-preemptive priority

manner. They found R matrix and waiting time distribution. Numerical results also provided in their model. Baek et al. [9] investigated a single server priority queueing system with two types of customers and consumable additional items. Additionally, they looked at buffer systems with zero buffers for type 1 customers, infinite buffers for type 2 customers, and buffers with K capacities for additional items. Krishnamoorthy and Divya [19] examined a single server queueing model with working vacation, non-preemptive priority, and two distinct N-policies. Their concept states that when the server is on vacation, responses to high priority (type I) clients continue, while responses to low priority (type II) clients must wait until the server resumes normal operations. Also they found busy period analysis, waiting time distribution and numerical illustrations.

A queueing model with two different kinds of clients in which arrival follows Markovian arrival process which was investigated by Chakravarthy and Dudin [13]. The steady state probability vector, waiting time distribution and several numerical illustrations are also found in their model. Sleptchenko et al. [28] has developed a single server queueing model along with arbitrary N client classes, class-dependent service rates, and priority classes. Krishnamoorthy et al. [20] looked into a multi-server queue with self-generated priorities and non-preemptive priority services. The arriving customer to a C-server counter follows MAP and service time follows PH for both priority customers. They found cost analysis and performance measures in their model. Nair et al. [24] analysed a $M/M/1$ queue with priority loss through feedback. They took into consideration the arrival of consumers with different priorities P_1 and P_2 in accordance with a marked Markovian arrival process and phase type distribution is used for service time. They discussed two types of model such as model 1 and 2. In model 1 was considered as non-preemptive service for P_2 customers and in model 2 was considered as preemptive policy of P_2 customers. Also they find waiting time analysis and system performance measures.

In real life situations, every working place and offices vacations are essential. Here we considered single and emergency vacation. The server can take the vacation after completion of service and also during the service time, the server can take emergency vacation. In many working places, the servers may take the vacation during the busy time and continue the remaining service of that customers. A queueing model with priority services was investigated by Ayyappan and Udayageetha [7]. They considered two types of vacation such as modified Bernoulli vacation and emergency vacation. After completing the requested service, the server goes on Bernoulli vacation if there are no high priority clients in the system; otherwise, the server is idle. During the service time of high priority customers, the server take emergency vacation. They presents some numerical examples and performance measures.

The arrival of negative consumers should detract the positive customers who gets the services from the server and they exit the system without the service being completed which is called as negative arrival. $M/M/1$ retrial queue with preemptive priority and a maximum of J vacations has been studied by Yuvarani and Saravanarajan [31]. They considered negative arrival of customer occurs in the busy period of positive customer. Due to the negative arrival, the positive customer spoiled their service and leave the system. Some performance measures and numerical illustrations are also given. A multi server queueing model with negative customer and partial protection of service has been done by Klimenok and Dudin [18]. A non-preemptive priority queue with server's walking process was done by Fukagawa et al. [15]. The stationary probability vector, queue waiting time and evaluation measures of the queue also done by them. A Single-server Discrete-time Retrial G-queue with server Breakdowns and Repairs was done by Wang and Zhang [30].

Ayyappan and Thilagavathy [6] accomplishes a single-server priority retrial queue with stand by server, breakdown, repair, vacation, negative arrival, balking and reneging. They used the concept of negative arrival while the main server is in busy. The negative arrival are affected to the positive customer those who gets the service to be removed completely from the system and server moved into repair process. Busy period analysis, cost analysis and graphical illustrations are all given. Retrial queueing model $MMAP/M_2/1$ with two orbits have been studied by Avrachenkov et al. [8]. They considered two orbits, one is an infinite capacity and another one is finite capacity. Some of the performance measures and numerical illustrations are also

provided. A multichannel queueing model with quasi-random input retrial times and phase type services has investigated by Artalejo and Corral [2]. By solving Quasi birth and death process, the stationary probability vector has been evaluated and some performance measures also evaluated.

The concept of optional secondary service is that after completion of primary customer service, the customer may need a secondary service with probability p or the customer leaves the system completely with probability q. A queue with single server subject to second optional service has been done by Madan [22]. They considered two types of service such as essential and optional service. The essential service are given to all the customers in the system while second optional service are provided only for some customers those who need the service once again. Waiting time distribution and particular cased are also derived.

Chakravarthy [11] looked into a $MAP/PH(1), PH(2)/1$ queueing structure whereby services were given on a first come, first served basis subject to vacations and optional secondary services. A single server queueing model with N-policy and second optional services have been evaluated by Das et al. [14]. They presented the cost analysis and various performance measures of their model. A single server queue with setup, closedown, multiple vacation, standby server, breakdown, repair and reneging was studied by Ayyappan and Thilagavathy [5]. Chakravarthy and Agarwal [12] explored a machine repair problem with Unreliable server. In their model, they considered phase type distribution for the service and repair time of server. They also determine the performance measures and some numerical illustrations. A single server retrial queueing model with Bernoulli vacation, feedback, breakdown and repair was analysed by Ayyappan and Gowthami [3]. Also they found the cost analysis, some performance measures and by using the performance measures they evaluate the numerical results.

There are two types of breakdown such as active and passive breakdown. The active breakdown occurs while the busy period of server and the passive breakdown occurs during the server idle period. Gao et al. [16] examined two kinds of breakdown and delayed repairs in an unreliable retry queue. They employed passive and active breakdowns in the periods of idle and busy, respectively. When a passive breakdown happens, the server cannot be repaired immediately and must wait for consumers to arrive from the outside or from orbit because the server lacks a monitoring system during idle times. They provided a few performance measures based on the likelihood that a server would be busy, idle, or undergoing maintenance, among other factors. By using performance measures they find the numerical values. A queueing model with single server subject to working vacation and two type of server breakdown have been analysed by Agarwal et al. [1]. They considered the server breakdown while server is in working vacation or normal busy period. Numerical illustrations are also examined by them.

Niu et al. [27] investigated a vacation queue with Setup and Closedown periods, as well as batch Markovian Arrival Processes. In their model after completion of service, the server closedown the system and setup the system when the server return from vacation. The arrival process follows Markovian arrival process and service time follows phase type distribution with the random variables Bernoulli vacation, setup, Bernoulli feedback, breakdown, repair and impatient customers was investigated by Ayyappan and Gowthami [4]. Now-a-days, in many places, most of the peoples does not prefer to wait a line at long time. Here we consider balking such as the customer does not enter into the system due to impatient. Swathi et al. [29] examined a queueing system with balking and reneging. In their model, they included the concept of customer balking and reneging as a result of the server's unavailability during vacation and breakdown times. The steady state analysis of the system and several performance measures were also derived by them.

The remainder of the article is organised as follows: The narration for our model is located in section 2. Section 3 discusses the matrix generating procedure and some notations. The system stableness, the invariant probability vector, and R matrix are all obtained in section 4. The busy period analysis is presented in section 5. Section 6 contains performance measures. Section 7 presents the cost analysis. Section 8 contains some numerical and graphical outcomes. Section 9 contains the conclusion part.

2. Narration of the Model

- In this article, we analyse a single server classical retrial policy with preemptive priority queue, differentiate breakdown, second optional service, phase type repair, two types of vacation, closedown, setup and balking.
- Arrival of both high priority and low priority (HP and LP) clients in accordance with $MMAP$, which is a generic version of MAP with parameter matrices of order (D_0, D_1, D_2) of order m_2. The D_0 matrix denotes the absence of positive customer arrivals, while the D_1 and D_2 matrix denotes customer arrivals.
- While low priority customers only have a "L" size finite buffer, high priority customers have an infinite capacity. The negative arrival of customers are also follows MAP with representation (C_0, C_1) of order m_1, where C_0 represents to no arrival and C_1 represents to arrival of customers.
- The service offered to the customers in the basis of first come first service. The customers receive the service immediately if server becomes idle. In idle time, the server may struck due to breakdown to starts the service and then moves to repair process.
- During the service period, the server experiences a breakdown owing to a negative arrival and immediately enters the repair process. At the same time, positive customer who receive service from the server will abandon the system totally.
- When a low priority client attempts to join an orbit that is already full, the action is deemed unsuccessful. If any low priority customers retrial from the orbit while the server is idle, the low priority customers will receive service from the server successfully.
- The duration of service time of both (HP/LP) customers which follows PH type distribution with the notations (α, T) of order n_1 where $T^0 + Te = 0$ such that $T^0 = -Te$ and the optional service of HP customers also follows PH type distribution with notation (α_1, T_1) of order n_2 where $T_1^0 + T_1 e = 0$ such that $T_1^0 = -T_1 e$.
- The server repair time follows a PH type distribution with representation (β, S) of order s where $S^0 + Se = 0$ and $S^0 = -Se$.
- During the service time of LP customers, the server takes emergency vacation and the customers those who are receives the service have to join the orbit and will get the service after the vacation completion by server. When the service is finished, the server shuts down the system and goes on vacation.
- The server will startup the system after completion of vacation. When on vacation, the customer may join the system with probability $(1 - b)$ or balk the system because of impatience with probability b.
- Inter-retrial times, emergency vacation, single vacation, breakdown times, closedown and setup times are all based on exponential distribution and its parameters as δ, η_1, η_2, τ, ϕ and ψ respectively. (see Figure 1).

3. The QBD process infinitesimal generation matrix

Let us narrate the few notation of this model which followed by generator matrix of the QBD process as follows:

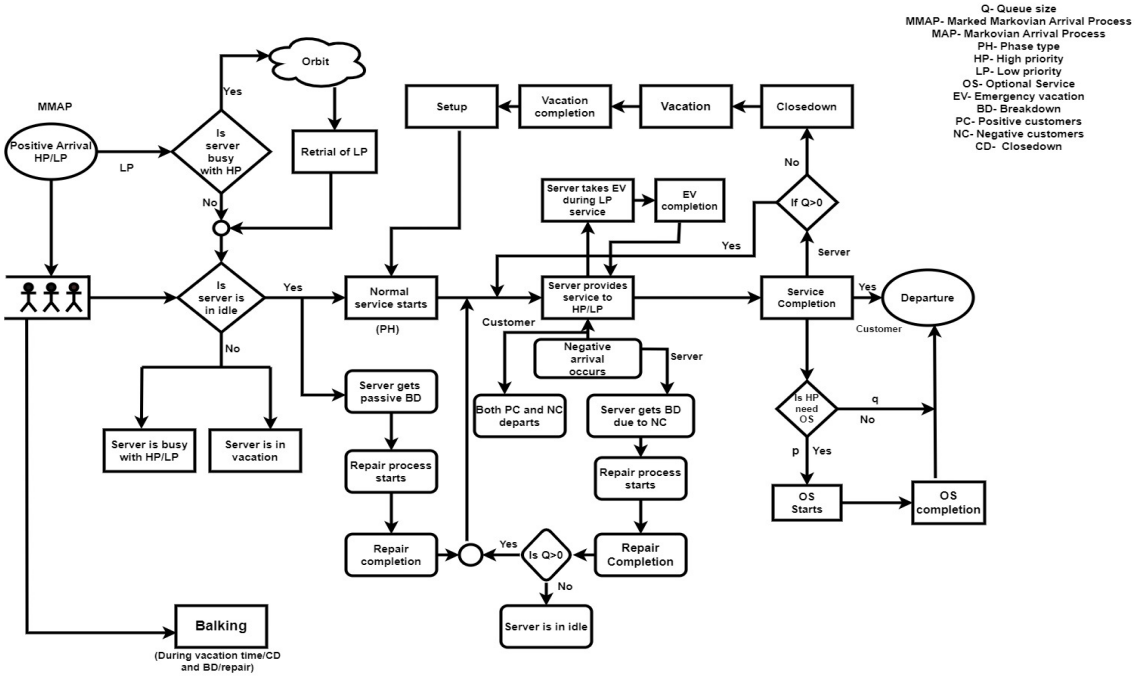


Figure 1: *Representations of our model in schematic form*

Notations:

- $\otimes$ - Two matrices with varying orders are combined in a Kronecker product.
- $\oplus$ - Two matrices with different orders are combined in a Kronecker sum.
- I_m - Represents the identity matrix of order $m \times m$.
- e - Each entry in a column vector has the required dimension, which is 1.
- 0 - It denotes an appropriate order of zero matrices.
- The fundamental arrival rate λ_i, where i=1,2 which is specified as $\lambda_i = \pi D_i e_{m_2}$, π represents the stationary probability vector of the generator matrix $D_0 + D_1 + D_2$ which determines MMAP transitions.
- The negative arrival rate be λ_3 which is specified as $\lambda_3 = \pi_1 C_1 e_{m_1}$, where π_1 is the steady state probability vector of generator matrix $C = C_0 + C_1$.
- The rate of normal service of server is indicated as $\mu = [\alpha(-T^{-1})e_{n_1}]^{-1}$.
- The rate of optional service of server is indicated as $\mu_1 = [\alpha_1(-T_1^{-1})e_{n_2}]^{-1}$.
- The repair rate for normal/optional service of server as represented by $\sigma = [\beta(-S^{-1})e_s]^{-1}$.
- $N_1(t)$ indicates the total number of customers with high priority in the system at time t.
- $N_2(t)$ indicates the total number of customers with low priority in the orbit.
- $C(t)$ stands for the server status at time t.

$$C(t) = \begin{cases} 0, & \text{server is in idle.} \\ 1, & \text{server is busy in HP normal Serivce.} \\ 2, & \text{server is busy in HP optional service.} \\ 3, & \text{server is busy for LP normal service.} \\ 4, & \text{server is in repair process.} \\ 5, & \text{sever is in emergency vacation.} \\ 6, & \text{server is in closedown.} \\ 7, & \text{server is in vacation} \\ 8, & \text{server is in setup.} \end{cases}$$

- $S(t)$ represents the service phase of server.
- $K(t)$ represents the repair phase of server.
- $A_i(t)$ stands for the arrival phase of negative and positive customers, where i=1,2.

Let $\{N_1(t), N_2(t), C(t), S(t), K(t), A_1(t), A_2(t), \quad t \geq 0\}$ is the $CTMC$ with the state space as follows,

$$\Omega = l(0) \bigcup_{i=1}^{\infty} l(i),$$

where

$$\begin{aligned} l(0)=&\{(0, i_2, 0, a_2) :\ 0 \leq i_2 \leq L,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(0, i_2, 3, k_1, a_1, a_2) :\ 0 \leq i_2 \leq L,\ 1 \leq k_1 \leq n_1,\ 1 \leq a_1 \leq m_1,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(0, i_2, 4, b, a_2) :\ 0 \leq i_2 \leq L,\ 1 \leq b \leq s,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(0, i_2, j, a_2) :\ 0 \leq i_2 \leq L,\ j = 5,6,7,8,\ 1 \leq a_2 \leq m_2\}, \end{aligned}$$

and for $i \geq 1$,

$$\begin{aligned} l(i)=&\{(i_1, i_2, 1, k_1, a_1, a_2) :\ i_1 \epsilon \mathbb{Z}^+,\ 0 \leq i_2 \leq L,\ 1 \leq k_1 \leq n_1,\ 1 \leq a_1 \leq m_1,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(i_1, i_2, 2, k_2, a_1, a_2) :\ i1 \epsilon \mathbb{Z}^+,\ 0 \leq i_2 \leq L,\ 1 \leq k_2 \leq n_2,\ 1 \leq a_1 \leq m_1,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(i_1, i_2, 3, k_1, a_1, a_2) :\ i \epsilon \mathbb{Z}^+,\ 0 \leq i_2 \leq L,\ 1 \leq k_1 \leq n_1,\ 1 \leq a_1 \leq m_1,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(i_1, i_2, 4, b, a_2) :\ i \epsilon \mathbb{Z}^+,\ 0 \leq i_2 \leq L,\ 1 \leq b \leq s,\ 1 \leq a_2 \leq m_2\} \\ &\cup\{(i_1, i_2, j, a_2) :\ i \epsilon \mathbb{Z}^+,\ j = 5,6,7,8,\ 0 \leq i_2 \leq L,\ 1 \leq a_2 \leq m_2\}. \end{aligned}$$

3.1. The Infinitesimal Matrix Generation

The quasi birth and death process has the generating matrix Q, is as follows:

$$Q = \begin{bmatrix} B_{00} & B_{01} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ B_{10} & A_1 & A_0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \dots & \dots \end{bmatrix},$$

where

$$B_{00} = \begin{bmatrix} B_{00}^1 & B_{00}^2 & B_{00}^3 & 0 & 0 & 0 & 0 \\ B_{00}^4 & B_{00}^5 & B_{00}^6 & B_{00}^7 & B_{00}^8 & 0 & 0 \\ B_{00}^9 & 0 & B_{00}^{10} & 0 & 0 & 0 & 0 \\ 0 & B_{00}^{11} & 0 & B_{00}^{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & B_{00}^{13} & B_{00}^{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & B_{00}^{15} & B_{00}^{16} \\ B_{00}^{17} & 0 & 0 & 0 & 0 & 0 & B_{00}^{18} \end{bmatrix},$$

where $B_{00}^1 = diag(D_0 - \tau I_{m_2}, D_0 - (\tau+\delta) I_{m_2}, ..., D_0 - (\tau + k\delta) I_{m_2})$

$$B_{00}^2 = \begin{bmatrix} e' \otimes \alpha \otimes D_2 & 0 & 0 & \dots & 0 \\ e' \otimes \delta\alpha \otimes I_{m_1} & e' \otimes \alpha \otimes D_2 & 0 & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & e' \otimes L\delta\alpha \otimes I_{m_1} & e' \otimes \alpha \otimes D_2 & \dots & 0 \end{bmatrix},$$

$$B_{00}^3 = I_{(L+1)} \otimes e' \otimes \tau I_{m_2},\ B_{00}^4 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & e_{m_1} \otimes T^0 \otimes I_{m_2} & 0 & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & e_{m_1} \otimes T^0 \otimes I_{m_2} \end{bmatrix},$$

$$B_{00}^5 = \begin{bmatrix} f_1 & f_2 & 0 & \dots & 0 \\ 0 & f_1 & f_2 & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & f_1 + f_2 \end{bmatrix}, \text{ where } f_1 = T \oplus D_0 \oplus C_0 - \eta_1 I_{nm_1m_2},\ f_2 = I_n \otimes I_m \otimes D_2,$$

$$B_{00}^6 = I_{(L+1)} \otimes I_{n_1} \otimes C_1 \otimes e_{m_2},\quad B_{00}^7 = I_{(L+1)} \otimes e_{nm_1} \otimes \eta_1 \otimes I_{m_2},$$

$$B_{00}^8 = \begin{bmatrix} e_n \otimes T^0 \otimes I_{m_2} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ & & \ddots & \ddots & \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix},\ B_{00}^9 = I_{(L+1)} \otimes R^0 \otimes I_{m_2},$$

$$B_{00}^{10} = \begin{bmatrix} f_3 & f_4 & 0 & \dots & 0 \\ 0 & f_3 & f_4 & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & f_3 + f_4 \end{bmatrix}, \text{ where } f_3 = D_0 \oplus R, f_4 = I_s \otimes D_2,$$

$$B_{00}^{11} = I_{(L+1)}) \otimes e' \otimes \alpha_1\eta_1 I_{m_2},$$

$$B_{00}^{12} = \begin{bmatrix} f_5 & f_6 & 0 & \dots & 0 \\ 0 & f_5 & f_6 & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & f_5 + f_6 \end{bmatrix}, \text{ where } f_5 = (D_0 + b(D_1 + D_2)) - \eta_1 I_{m_2}, f_6 = D_2(1-b)$$

$$B_{00}^{13} = \begin{bmatrix} f_7 & f_8 & 0 & \dots & 0 \\ 0 & f_7 & f_8 & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & f_7 + f_8 \end{bmatrix}, \text{ where } f_7 = (D_0 + b(D_1 + D_2)) - \phi_1 I_{m_2}, f_6 = f_8,$$

$$B_{00}^{15} = \begin{bmatrix} f_9 & f_{10} & 0 & \dots & 0 \\ 0 & f_9 & f_{10} & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & f_9 + f_{10} \end{bmatrix}, \text{ where } f_9 = (D_0 + b(D_1 + D_2)) - \eta_2 I_{m_2},\ f_{10} = f_6,$$

$$B_{00}^{14} = I_{(L+1)} \otimes \phi I_{m_2},\ B_{00}^{16} = I_{(L+1)} \otimes \eta_2 I_{m_2},\ B_{00}^{17} = I_{(L+1)} \otimes \psi I_{m_2},$$

$$B_{00}^{18} = \begin{bmatrix} f_{11} & f_{12} & 0 & \dots & 0 \\ 0 & f_{11} & f_{12} & \dots & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & f_{11}+f_{12} \end{bmatrix}, \text{ where } f_{11} = D_0 - \psi I_{m_2}, f_{12} = D_2,$$

$$B_{01} = \begin{bmatrix} B_{01}^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & B_{01}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & B_{01}^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & B_{01}^4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & B_{01}^5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & B_{01}^6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & B_{01}^7 \end{bmatrix}, \text{ where } B_{01}^1 = I_{(L+1)} \otimes e' \otimes \alpha \otimes D_1,$$

$B_{01}^2 = I_{(L+1)} \otimes I_{n_1} \otimes D_1 \otimes I_{m_1}$, $B_{01}^3 = I_{(L+1)} \otimes I_s \otimes D_1$, $B_{01}^4 = B_{01}^5 = B_{01}^6 = I_{(L+1)} \otimes D_1(1-b)$, $B_{01}^7 = I_{(L+1)} \otimes D_1$,

$$B_{10} = \begin{bmatrix} B_{10}^1 & 0 & B_{10}^2 & 0 & B_{10}^3 & 0 & 0 \\ B_{10}^4 & 0 & B_{10}^5 & 0 & B_{10}^6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where } B_{10}^1 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & e_{m_1} \otimes qT^0 \otimes I_{m_2} & 0 & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & e_{m_1} \otimes qT^0 \otimes I_{m_2} \end{bmatrix},$$

$B_{10}^2 = I_{(L+1)} \otimes e_s \otimes I_{m_2}$,

$$B_{10}^3 = \begin{bmatrix} e_{m_1} \otimes qT^0 \otimes I_{m_2} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}, B_{10}^4 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & e_{m_1} \otimes T_1^0 \otimes I_{m_2} & 0 & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & e_{m_1} \otimes T_1^0 \otimes I_{m_2} \end{bmatrix},$$

$$B_{10}^5 = I_{(L+1)} \otimes e_s \otimes C_1 \otimes I_{m_2}, B_{10}^6 = \begin{bmatrix} e_{m_1} \otimes T_1^0 \otimes I_{m_2} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

$$A_0 = \begin{bmatrix} A_0^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_0^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_0^3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A_0^4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A_0^5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_0^6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_0^7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & A_0^8 \end{bmatrix} \text{ where } A_0^1 = I_{(L+1)} \otimes I_{n_1} \otimes I_{m_1} \otimes D_1,$$

$A_0^2 = I_{(L+1)} \otimes I_{n_2} \otimes I_{m_1} \otimes D_1$, $A_0^3 = I_{(L+1)} \otimes I_{n_1} \otimes I_{m_1} \otimes D_1$, $A_0^4 = I_{(L+1)} \otimes I_s \otimes D_1$,

$A_0^5 = I_{(L+1)} \otimes D_1(1-b)$, $A_0^6 = A_0^7 = A_0^5$, $A_0^8 = I_{(L+1)} \otimes D_1$,

$$A_1 = \begin{bmatrix} A_1^1 & A_1^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & A_1^3 & 0 & 0 & 0 & 0 & 0 & 0 \\ A_1^4 & 0 & A_1^5 & A_1^6 & A_1^7 & 0 & 0 & 0 \\ A_1^8 & 0 & 0 & A_1^9 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_1^{10} & 0 & A_1^{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_1^{12} & A_1^{13} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_1^{14} & A_1^{15} \\ A_1^{16} & 0 & 0 & 0 & 0 & 0 & 0 & A_1^{17} \end{bmatrix}, \text{where } A_1^1 = \begin{bmatrix} f_{13} & f_{14} & 0 & \dots & 0 \\ 0 & f_{13} & f_{14} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{13}+f_{14} \end{bmatrix},$$

where $f_{13} = (T \oplus D_0) \oplus C0$, $f_{14} = D_2 \otimes I_{m_1 n_1}$.

$$A_1^3 = \begin{bmatrix} f_{15} & f_{16} & 0 & \dots & 0 \\ 0 & f_{15} & f_{16} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{15}+f_{16} \end{bmatrix}, \text{where } A_1^2 = I_{(L+1)} \otimes \alpha_1 \otimes pT^0 \otimes I_{m_1 m_2},$$

$f_{15} = (T_1 \oplus D_0) \oplus C0$, $f_{16} = D_2 \otimes I_{m_1 n_2}$, $A_1^4 = I_{(L+1)} \otimes \alpha \otimes T^0 \otimes I_{m_1 m_2}$,

$$A_1^5 = \begin{bmatrix} f_{17} & f_{18} & 0 & \dots & 0 \\ 0 & f_{17} & f_{18} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{17}+f_{18} \end{bmatrix} \text{where } f_{17} = ((T \oplus D_0) \oplus C_0) - \eta_1 I_{n_1 m_1 m_2},$$

$f_{18} = D_2 \otimes I_{m_1 n_1}$,

$A_1^6 = I_{(L+1)} \otimes e_{n_1} \otimes C_1 \otimes I_{m_2}$, $A_1^7 = I_{(L+1)} \otimes e_{n_1} \otimes e_{m_1} \otimes \eta_1 I_{m_2}$, $A_1^8 = I_{(L+1)} \otimes e'_{m_1} \otimes S^0 \alpha \otimes I_{m_2}$,

$$A_1^9 = \begin{bmatrix} f_{19} & f_{20} & 0 & \dots & 0 \\ 0 & f_{19} & f_{20} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots f_{19}+f_{20} & \end{bmatrix} \text{where } f_{19} = S \oplus D_0, f_{20} = I_s \otimes D_2,$$

$$A_1^{10} = I_{(L+1)} \otimes e'_{n_1} \otimes e'_{m_1} \otimes \eta_1 I_{m_2}, \quad A_1^{11} = \begin{bmatrix} f_{21} & f_{22} & 0 & \dots & 0 \\ 0 & f_{21} & f_{22} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{21}+f_{22} \end{bmatrix},$$

where $f_{21} = (D_0 + b(D_1 + D_2)) - \eta_1 I_{m_2}, f_{22} = D_2(1-b)$, $A_1^{12} = \begin{bmatrix} f_{23} & f_{24} & 0 & \dots & 0 \\ 0 & f_{23} & f_{24} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{23}+f_{24} \end{bmatrix},$

where $f_{23} = (D_0 + b(D_1 + D_2)) - \varphi I_{m_2}, f_{24} = D_2(1-b)$, $A_1^{13} = I_{(L+1)} \otimes \varphi I_{m_2}$,

$$A_1^{14} = \begin{bmatrix} f_{25} & f_{26} & 0 & \dots & 0 \\ 0 & f_{25} & f_{26} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{25}+f_{26} \end{bmatrix}, \text{where } f_{25} = (D_0 + b(D_1 + D_2)) - \eta_2 I_{m_2}, f_{26} = D_2(1-b),$$

$A_1^{15} = I_{(L+1)} \otimes \eta_2 I_{m_2}$, $A_1^{16} = I_{(L+1)} \otimes e' \otimes \alpha \otimes \psi I_{m_2}$,

$$A_1^{17} = \begin{bmatrix} f_{27} & f_{28} & 0 & \dots & 0 \\ 0 & f_{27} & f_{28} & \dots & 0 \\ & \ddots & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & f_{27}+f_{28} \end{bmatrix}, \text{where } f_{27} = D_0 - \psi I_{m_2}, \quad f_{28} = D_2,$$

$$A_2 = \begin{bmatrix} A_2^1 & 0 & 0 & A_2^2 & 0 & 0 & 0 & 0 \\ A_2^3 & 0 & 0 & A_2^4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \text{ where } A_2^1 = I_{(L+1)} \otimes \alpha \otimes qT^0 \otimes I_{m_1m_2},$$

$A_2^2 = I_{(L+1)} \otimes e_s \otimes C_1 \otimes I_{m_2}$, $A_2^3 = I_{(L+1)} \otimes \alpha \otimes T_1^0 \otimes I_{m_1m_2}$, $A_2^4 = I_{(L+1)} \otimes e_s \otimes C_1 \otimes I_{m_2}$.

4. System Analysis

We evaluate this model, beneath of the certain conditions to ensure that the system to be stable.

4.1. Stability condition for the system

Let A be the matrix , where $A = A_0 + A_1 + A_2$. The invariant probability vector ς, which is referred to as a generator matrix and its satisfying

$$\varsigma A = 0, \quad \varsigma e = 1.$$

The vector ς represents the states which is partitioned by $\varsigma = (\varsigma_0, \varsigma_1, \varsigma_2, \varsigma_3, \varsigma_4, \varsigma_5, \varsigma_6, \varsigma_7)$ and it is subdivided by $(\varsigma_{00}, \varsigma_{01}, ..., \varsigma_{0K}, \varsigma_{10}, \varsigma_{11}, ..., \varsigma_{1K}, \varsigma_{20}, \varsigma_{21}, ..., \varsigma_{2K}, \varsigma_{30}, \varsigma_{31}, ..., \varsigma_{3K}, \varsigma_{40}, \varsigma_{41}, ..., \varsigma_{4K}, \varsigma_{50}, \varsigma_{51}, ..., \varsigma_{5K}, \varsigma_{60}, \varsigma_{61}, ..., \varsigma_{6K}, \varsigma_{70}, \varsigma_{71}, ..., \varsigma_{7K})$ which is evaluated by the aid of subsequent equation:

$$\begin{aligned}
&\varsigma_{00}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) + (\alpha \otimes qT^0 \otimes I_{m_1m_2})] + \varsigma_{10}[\alpha \otimes T_1^0 \otimes I_{m_1m_2}] \\
&\qquad + \varsigma_{20}[\alpha \otimes T^0 \otimes I_{m_1m_2}] + \varsigma_{30}[e'_{m_1} \otimes S^0\alpha \otimes I_{m_2}] + \varsigma_{70}[e' \otimes \alpha \otimes \psi I_{m_2}] = 0, \\
&\varsigma_{0(i-1)}[D_2 \otimes I_{m_1n_1}] + \varsigma_{0i}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) + (\alpha \otimes qT^0 \otimes I_{m_1m_2}] \\
&\qquad + \varsigma_{1i}[\alpha \otimes T_1^0 \otimes I_{m_1m_2}] + \varsigma_{2i}[\alpha \otimes T^0 \otimes I_{m_1m_2}] + \varsigma_{3i}[e'_{m_1} \otimes S^0\alpha \otimes I_{m_2}] \\
&\qquad\qquad + \varsigma_{7i}[e' \otimes \alpha \otimes \psi I_{m_2}] = 0, 1 \le i \le K. \\
&\varsigma_{0K}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) + (\alpha \otimes qT^0 \otimes I_{m_1m_2}) + D_2 \otimes I_{m_1n_1}] + \varsigma_{1K}[\alpha \otimes T_1^0 \otimes I_{m_1m_2}] \\
&\qquad + \varsigma_{2K}[\alpha \otimes T^0 \otimes I_{m_1m_2}] + \varsigma_{3K}[e'_{m_1} \otimes S^0\alpha \otimes I_{m_2}] + \varsigma_{7K}[e' \otimes \alpha \otimes \psi I_{m_2}] = 0, \\
&\varsigma_{00}[\alpha_1 \otimes pT^0 \otimes I_{m_1m_2}] + \varsigma_{10}[(I_{n_2} \otimes I_{m_1} \otimes D_1) + ((T_1 \oplus D_0) \oplus C_0)] = 0, \\
&\varsigma_{0i}[\alpha_1 \otimes pT^0 \otimes I_{m_1m_2}] + \varsigma_{1(i-1)}[D_2 \otimes I_{m_1n_2}] + \varsigma_{1i}[(I_{n_2} \otimes I_{m_1} \otimes D_1) + ((T_1 \oplus D_0) \oplus C_0)] = 0, 1 \le i \le K-1. \\
&\varsigma_{0K}[\alpha_1 \otimes pT^0 \otimes I_{m_1m_2}] + \varsigma_{1K}[(I_{n_2} \otimes I_{m_1} \otimes D_1) + ((T_1 \oplus D_0) \oplus C_0) + (D_2 \otimes I_{m_1m_2})] = 0, \\
&\varsigma_{20}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) - \eta_1 I_{n_1m_1m_2}] + \varsigma_{40}[e'_{n_1} \otimes e'_{m_1} \otimes \eta_1 I_{m_2}] = 0, \\
&\varsigma_{2(i-1)}[D_2 \otimes I_{m_1n_1}] + \varsigma_{2i}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) - \eta_1 I_{n_1m_1m_2}] \\
&\qquad\qquad + \varsigma_{4i}[e'_{n_1} \otimes e'_{m_1} \otimes \eta_1 I_{m_2} = 0,\ 1 \le i \le K-1. \\
&\varsigma_{2K}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + (((T \oplus D_0) \oplus C_0) - \eta_1 I_{n_1m_1n_2}) + D_2 \otimes I_{m_1n_1}] + \varsigma_{4K}[e'_{n_1} \otimes e'_{m_1} \otimes \eta_1 I_{m_2}] = 0, \\
&\varsigma_{20}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) - \eta_1 I_{n_1m_1m_2}] + \varsigma_{40}[e'_{n_1} \otimes e'_{m_1} \otimes \eta_1 I_{m_2}] = 0, \\
&\varsigma_{2(i-1)}[D_2 \otimes I_{m_1n_1}] + \varsigma_{2i}[(I_{n_1} \otimes I_{m_1} \otimes D_1) + ((T \oplus D_0) \oplus C_0) - \eta_1 I_{n_1m_1m_2}] \\
&\qquad\qquad + \varsigma_{4i}[e'_{n_1} \otimes e'_{m_1} \otimes \eta_1 I_{m_2} = 0,\ 1 \le i \le K-1.
\end{aligned}$$

$$\varsigma_{20}[(I_{n_1}\otimes I_{m_1}\otimes D_1)+((T\oplus D_0)\oplus C_0)-\eta_1 I_{n_1m_1m_2}]+\varsigma_{40}[e'_{n_1}\otimes e'_{m_1}\otimes \eta_1 I_{m_2}]=0,$$
$$\varsigma_{2(i-1)}[D_2\otimes I_{m_1n_1}]+\varsigma_{2i}[(I_{n_1}\otimes I_{m_1}\otimes D_1)+((T\oplus D_0)\oplus C_0)-\eta_1 I_{n_1m_1m_2}]$$
$$+\varsigma_{4i}[e'_{n_1}\otimes e'_{m_1}\otimes \eta_1 I_{m_2}=0,\ 1\le i\le K-1.$$
$$\varsigma_{2K}[(I_{n_1}\otimes I_{m_1}\otimes D_1)+(((T\oplus D_0)\oplus C_0)-\eta_1 I_{n_1m_1n_2})+D_2\otimes I_{m_1n_1}]+\varsigma_{4K}[e'_{n_1}\otimes e'_{m_1}\otimes \eta_1 I_{m_2}]=0,$$
$$\varsigma_{00}[e_s\otimes C_1\otimes I_{m_2}]+\varsigma_{10}[e_s\otimes C_1\otimes I_{m_2}]+\varsigma_{20}[e_{n_1}\otimes C_1\otimes I_{m_2}]+\varsigma_{30}[(I_s\otimes D_1)+(S\oplus D_0)]=0,$$
$$\varsigma_{0i}[e_s\otimes C_1\otimes I_{m_2}]+\varsigma_{1i}[e_s\otimes C_1\otimes I_{m_2}]+\varsigma_{2i}[e_n\otimes C_1\otimes I_{m_2}]$$
$$+\varsigma_{3(i-1)}[I_s\otimes D_2]+\varsigma_{3i}[(I_s\otimes D_1)+(S\oplus D_0)]=0,1\le i\le K-1.$$
$$\varsigma_{0K}[e_s\otimes C_1\otimes I_{m_2}]+\varsigma_{1K}[e_s\otimes C_1\otimes I_{m_2}]+\varsigma_{2K}[e_n\otimes C_1\otimes I_{m_2}]$$
$$+\varsigma_{3K}[(I_s\otimes (D_1+D_2))+(S\oplus D_0)]=0,$$
$$\varsigma_{20}[e_{n_1}\otimes e_{m_1}\otimes \eta I_{m_2}]+\varsigma_{40}[(D_0+D_1)+bD_2-\eta_1 I_{m_2}]=0,$$
$$\varsigma_{2i}[e_{n_1}\otimes e_{m_1}\otimes \eta I_{m_2}]+\varsigma_{4(i-1)}[D_2(1-b)]+\varsigma_{4i}[(D_0+D_1)+bD_2-\eta_1 I_{m_2}]=0,1\le i\le K-1.$$
$$\varsigma_{2K}[e_{n_1}\otimes e_{m_1}\otimes \eta I_{m_2}]+\varsigma_{4K}[(D_0+D_1+D_2)-\eta_1 I_{m_2}]=0,$$
$$\varsigma_{50}[(D_0+D_1)+bD_2-\pi I_{m_2}]=0,$$
$$\varsigma_{5(i-1)}[D_2(1-b)]+\varsigma_{5i}[(D_0+D_1)+bD_2-\pi I_{m_2}]=0,1\le i\le K-1.$$
$$\varsigma_{5K}[(D_0+D_1+D_2)-\pi I_{m_2}]=0,$$
$$\varsigma_{50}[\pi I_{m_2}]+\varsigma_{60}[(D_0+D_1)+bD_2-\eta_2 I_{m_2}]=0,$$
$$\varsigma_{5i}[\pi I_{m_2}]+\varsigma_{6(i-1)}[D_2(1-b)]+\varsigma_{6i}[(D_0+D_1)+bD_2-\eta_2 I_{m_2}]=0,\ 1\le i\le K-1.$$
$$\varsigma_{5K}[\pi I_{m_2}]+\varsigma_{6K}[(D_0+D_1+D_2)-\eta_2 I_{m_2}]=0,$$
$$\varsigma_{60}[\eta_2 I_{m_2}]+\varsigma_{70}[(D_0+D_1)\psi I_{m_2}]=0,$$
$$\varsigma_{6i}[\eta_2 I_{m_2}]+\varsigma_{7(i-1)}[D_2]+\varsigma_{7i}[(D_0+D_1)-\psi I_{m_2}]=0,$$
$$\varsigma_{6K}[\eta_2 I_{m_2}]+\varsigma_{7K}[(D_0+D_1+D_2)-\psi I_{m_2}]=0,$$

subject to

$$[\sum_{i=0}^{K}\varsigma 0i+\sum_{i=0}^{K}\varsigma_{2i}]e_{n_2m_1m_2}+[\sum_{i=0}^{K}\varsigma_{1i}]e_{n_2m_1m_2}+[\sum_{i=0}^{K}\varsigma_{3i}]e_{sm_2}+[\sum_{r=4}^{7}\sum_{i=0}^{K}\varsigma_{ri}]e_{m_2}=1.$$

The necessary and sufficient condition of a QBD process which satisfy the condition $\varsigma A_0 e<\varsigma A_2 e$ that system to be stay in stable.
Therefore,

$$\begin{aligned}
&\varsigma_{00}[e_{n_1}\otimes e_{m_1}\otimes D_1e_{m_2}]+\varsigma_{01}[e_{n_1}\otimes e_{m_1}\otimes D_1e_{m_2}]+...+\varsigma_{0K}[e_{n_1}\otimes e_{m_1}\otimes D_1e_{m_2}]\\
&+\varsigma_{10}[e_{n_2}\otimes e_{m_1}\otimes D_1e_{m_2}]+\varsigma_{11}[e_{n_2}\otimes e_{m_1}\otimes D_1e_{m_2}]+...+\varsigma_{1K}[e_{n_2}\otimes e_{m_1}\otimes D_1e_{m_2}]\\
&+\varsigma_{20}[e_{n_1}\otimes e_{m_1}\otimes D_1e_{m_2}]+\varsigma_{21}[e_{n_1}\otimes e_{m_1}\otimes D_1e_{m_2}]+...+\varsigma_{2K}[e_{n_1}\otimes e_{m_1}\otimes D_1e_{m_2}\\
&+\varsigma_{30}[e_s\otimes D_1e_{m_2}]+\varsigma_{31}[e_s\otimes D_1e_{m_2}]+...+\varsigma_{3K}[e_s\otimes D_1e_{m_2}]\\
&+\varsigma_{40}[D_1(1-b)]+\varsigma_{41}[D_1(1-b)]+...+\varsigma_{4K}[D_1(1-b)]\\
&+\varsigma_{50}[D_1(1-b)]+\varsigma_{51}[D_1(1-b)]+...+\varsigma_{5K}[D_1(1-b)]\\
&+\varsigma_{60}[D_1(1-b)]+\varsigma_{61}[D_1(1-b)]+...+\varsigma_{6K}[D_1(1-b)]\\
&+\varsigma_{70}[D_1]+\varsigma_{71}[D_1]+...+\varsigma_{7K}[D_1]<\varsigma_{00}[qT^0\otimes e_{m_1}\otimes e_{m_2}]+\varsigma_{10}[T_1^0\otimes e_{m_1}\otimes e_{m_2}]\\
&+\varsigma_{01}[qT^0\otimes e_{m_1}\otimes e_{m_2}]+\varsigma_{11}[T_1^0\otimes e_{m_1}\otimes e_{m_2}]+...+\varsigma_{0K}[qT^0\otimes e_{m_1}\otimes e_{m_2}]\\
&+\varsigma_{1K}[T_1^0\otimes e_{m_1}\otimes e_{m_2}]+\varsigma_{00}[e_s\otimes C_1e_{m_1}\otimes e_{m_2}]\\
&+\varsigma_{10}[e_s\otimes C_1e_{m_1}\otimes e_{m_2}]+\varsigma_{01}[e_s\otimes C_1e_{m_1}\otimes e_{m_2}]\\
&+\varsigma_{11}[e_s\otimes C_1e_{m_1}\otimes e_{m_2}]+...+\varsigma_{0K}[e_s\otimes C_1e_{m_1}\otimes e_{m_2}]+\varsigma_{1K}[e_s\otimes C_1e_{m_1}\otimes e_{m_2}]
\end{aligned}$$

4.2. The Invariant Probability Vector

Let X represents the infinitesimal generator matrix Q and which is split by $X = (X_0, X_1, X_2, ...)$.

For X_0 is of dimension $((L+1)m_2 + (L+1)n_1m_1m_2 + (L+1)sm_2 + 4(L+1)m_2)$ and $X_i's$ are of dimension $(2(L+1)n_1m_1m_2 + (L+1)n_2m_1m_2 + (L+1)sm_2 + 4(L+1)m_2)$, $i \geq 1$.
As X is a vector of Q satisfies the relation

$$XQ = 0, \quad Xe = 1.$$

After satisfying the stability criterion, use the below equation to find the invariant probability vector X,

$$X_i = X_1R^{i-1}, \qquad i = 2, 3, ...$$

where R is the matrix created by solving the quadratic matrix equation, also known as the rate matrix.

$$R^2A_2 + RA_1 + A_0 = 0.$$

With the aid of succeeding equation we can find the vectors namely X_0, X_1 and X_2,

$$X_0B_{00} + X_1B_{10} = 0,$$
$$X_0B_{01} + X_1[A_1 + RA_2] = 0,$$

subject to normalizing condition

$$X_0e_{((L+1)m_2+(L+1)n_1m_1m_2+(L+1)sm_2+4(L+1)m_2)} + X_1[I-R]^{-1}e_{(2(L+1)n_1m_1m_2+(L+1)n_2m_1m_2+(L+1)sm_2+4(L+1)m_2)} = 1.$$

Therefore, the logarithmic reduction algorithm can be used to find the rate matrix R with the help of Latouche and Ramaswami [21].

5. Busy Period Analysis

- Under the busy period of $MMAP/PH/1$ queuing model, we will understand the epoch of the time interval starts from a new arrival which find the empty system and ends when the system becomes empty again at the completion of service.
- A busy cycle which is defined by the initial passage time of the level between 1 and 0 and the time return to level 0, requiring at least one visit to any other level.
- From level i to level $i-1$, where $i = 2, 3, 4, ...$ which is the initial passage time under the consideration of the QBD process. In the boundary states namely, $i = 0, 1$ which deals separately.
- For all the level i, where $i = 1, 2, 3, ...$, we seen that there are $(2(L+1)n_1m_1m_2 + (L+1)n_2m_1m_2 + (L+1)sm_2 + 4(L+1)m_2)$ states.

Notations:

- Let $G_{j,j'}(k, x)$ represent the conditional probability that the QBD process, starting at time $t = 0$ in the state (i, j) $t = 0$ and ends up in the state (i, j') by making meticulously k left jumps and obtaining both stages at the same period.
- Let the joint transform matrix

$$\tilde{G}_{j,j'}(z, s) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-sx} dG_{j,j'}(k, x); \quad |z| \leq 1, \quad Re(s) \geq 0$$

.

- The matrix $\tilde{G}(z,s) = \tilde{G}_{j,j'}(z,s)$. [Neuts [25]]
- Except for the boundary states, the matrix $G = G_{jj'} = \tilde{G}(1,0)$ be concerns the initial passage times.
- At time $t = 0$, when returning from stage 1 to stage 0, the conditional probability that described in the first return time and it is denoted as $G_{jj'}^{(1,0)}(k,x)$.
- At time $t = 0$, when returning to stage 0, the conditional probability that described and it is denoted as $G_{jj'}^{(0,0)}(k,x)$.
- At time $t = 0$, let S_{1j} represent the process's average initial passage time between stages i and $i-1$ and in the state (i,j).
- At time $t = 0$, in the initial passage procedure between levels i and $i-1$, which starts in the state (i,j), let S_{2j} be the average number of consumers that received service.
- $\tilde{S}_1$, $\tilde{S}_2$ be the column vectors along with S_{1j} and S_{2j} as their entries respectively.
- The expected first return time between stage 1 and stage 0 is represented by $\tilde{S}_1^{(1,0)}$.
- In the first return period from stage 1 to stage 0, the expected number of services completed and it is represented by $\tilde{S}_2^{(1,0)}$.
- The expected initial return time to stage 0 is represented by $\tilde{S}_1^{(0,0)}$.
- During the initial return time to stage 0, the expected number of services were rendered and it is represented by $\tilde{S}_2^{(0,0)}$.

We evaluate $\tilde{G}(z,s)$ matrix which satisfies the equation

$$\tilde{G}(z,s) = z(sI - A_1)^{-1}A_2 + (sI - A_1)^{-1}A_0\tilde{G}^2(z,s).$$

After found the rate matrix R, we can evaluate G matrix by using logarithmic reduction algorithm method which is given by Latouche and Ramaswami [21]

$$G = -(A_1 + RA_2)^{-1}A_2.$$

In the boundary states namely 1 and 0 and the equations represented by $\tilde{G}^{(1,0)}(z,s)$ and $\tilde{G}^{(0,0)}(z,s)$.

$$\tilde{G}^{(1,0)}(z,s) = z(sI - B_{11})^{-1}B_{10} + (sI - B_{11})^{-1}B_{12}\tilde{G}^{(2,1)}(z,s)\tilde{G}^{(1,0)}(z,s),$$
$$\tilde{G}^{(0,0)}(z,s) = (sI - B_{00})^{-1}B_{01}\tilde{G}^{(1,0)}(z,s).$$

The matrices $G, \tilde{G}^{(1,0)}(1,0)$ and $\tilde{G}^{(0,0)}(1,0)$ are stochastic.

The instants can be calculated as obeys:

$$\tilde{S}_1 = -\frac{\partial\tilde{G}(z,s)}{\partial s}\Big|_{s=0,z=1}e = -[A_0(G+I) + A_1]^{-1}e,$$
$$\tilde{S}_2 = \frac{\partial\tilde{G}(z,s)}{\partial z}\Big|_{s=0,z=1}e = -[A_0(G+I) + A_1]^{-1}]A_2e,$$
$$\tilde{S}_1^{(1,0)} = -\frac{\partial\tilde{G}^{(1,0)}(z,s)}{\partial s}\Big|_{s=0,z=1} = -[B_{11} + B_{12}\tilde{G}^{(2,1)}(1,0)]^{-1}[e + B_{12}\tilde{S}_1^{(2,1)}],$$
$$\tilde{S}_2^{(1,0)} = \frac{\partial\tilde{G}^{(1,0)}(z,s)}{\partial z}\Big|_{s=0,z=1}e = -[B_{12}\tilde{G}^{(2,1)}(1,0) + B_{11}]^{-1}[B_{12}\tilde{S}_2^{(2,1)} + B_{10}e],$$
$$\tilde{S}_1^{(0,0)} = -\frac{\partial\tilde{G}^{(0,0)}(z,s)}{\partial s}\Big|_{s=0,z=1}e = -B_{00}^{-1}[B_{01}\tilde{S}_1^{(1,0)} + e],$$
$$\tilde{S}_2^{(0,0)} = \frac{\partial\tilde{G}^{(0,0)}(z,s)}{\partial z}\Big|_{s=0,z=1}e = -B_{00}^{-1}[B_{01}\tilde{S}_2^{(1,0)}].$$

6. Performance Measure

- Probability of server being idle:

$$P_{SI} = \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{0i_21a_2}.$$

- Probability of server to be busy with HP customers:

$$P_{BH} = \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{k_1=1}^{n_1} \sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} X_{i_1i_21k_1a_1a_2}.$$

- Probability of server to be busy with LP customers:

$$P_{BL} = \sum_{i_2=0}^{L} \sum_{k_1=1}^{n_1} \sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} X_{0i_23k_1a_1a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{k_1=1}^{n_1} \sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} X_{i_1i_23k_1a_1a_2}.$$

- Probability of server to be busy with optional service of HP customers:

$$P_{BHOS} = \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{k_2=1}^{n_2} \sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} X_{i_1i_22k_2a_1a_2}.$$

- Probability of server being emergency vacation:

$$P_{EV} = \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{0i_25a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{i_1i_25a_2}.$$

- Probability of server being normal vacation:

$$P_{NV} = \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{0i_27a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{i_1i_27a_2}.$$

- Probability of server being closedown:

$$P_{CD} = \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{0i_26a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{i_1i_26a_2}.$$

- Probability of server being setup:

$$P_{SU} = \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{0i_28a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{a_2=1}^{m_2} X_{i_1i_28a_2}.$$

- Expected number of HP customers in the system:

$$\begin{aligned} E_{System} &= \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{L} \sum_{k_1=1}^{n_1} i_1 X_{i_1i_21k_1a_1a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{k_2=1}^{n_2} \sum_{a_1=1}^{m_1} i_1 X_{i_1i_22k_2a_1a_2} \\ &+ \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{k_2=1}^{n_2} \sum_{a_1=1}^{m_1} \sum_{a_2=1}^{m_2} i_1 X_{i_1i_23k_2a_1a_2} + \sum_{i_1=1}^{\infty} \sum_{i_2=0}^{L} \sum_{j=4}^{8} \sum_{a_2=1}^{m_2} i_2 X_{i_1i_2ja_2} \\ &= X_1(I-R)^{-2} e_{2(L+1)n_1m_1m_2+(L+1)n_2m_1m_2+(L+1)sm_2+4(L+1)m_2}. \end{aligned}$$

- Expected number of LP customers in the orbit:

$$\begin{aligned}E_{Orbit} = &\sum_{i_2=1}^{L}\sum_{a_2=1}^{m_2} i_2 X_{0i_20a_2} + \sum_{i_2=1}^{L}\sum_{k_1=1}^{n_1}\sum_{a_1=1}^{m_1}\sum_{a_2=1}^{m_2} i_2 X_{0i_23k_1a_1a_2} + \sum_{i_2=1}^{L}\sum_{b=1}^{s}\sum_{a_2=1}^{m_2} i_2 X_{0i_24ba_2}\\ &+ \sum_{i_2=1}^{L}\sum_{j=5}^{8}\sum_{a_2=1}^{m_2} i_2 X_{0i_2ja_2} + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{L}\sum_{k_1=1}^{n_1}\sum_{a_1=1}^{m_1}\sum_{a_2=1}^{m_2} i_2 X_{i_1i_21k_1a_1a_2}\\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{L}\sum_{k_2=1}^{n_2}\sum_{a_1=1}^{m_1}\sum_{a_2=1}^{m_2} i_2 X_{i_1i_2k_2a_1a_2} + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{L}\sum_{k_1=1}^{n_1}\sum_{a_1=1}^{m_1}\sum_{a_2=1}^{m_2} i_2 X_{i_1i_23k_1a_1a_2}\\ &+ \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{L}\sum_{b=1}^{s}\sum_{a_2=1}^{m_2} i_2 X_{i_1i_24ba2} + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{L}\sum_{j=5}^{8}\sum_{a_2=1}^{m_2} i_2 X_{i_1i_2ja_2}.\end{aligned}$$

7. Analysis of cost model

In this section, we introduce a cost function TC with the following assumption:

- TC - Total cost per unit time.
- C_{H_h} - Holding cost of each HP customer in the system at per unit time.
- C_{H_l} - Holding cost of each LP customer in the orbit at per unit time.
- C_{SI} - Per unit time cost during the server is in idle period.
- C_{BH} - Per unit time cost during the server is busy with HP customers.
- C_{BL} - Per unit time cost during the server is busy with LP customers.
- C_{BHOS} - Per unit time cost during the server is busy with optional service of HP customers.
- C_R - Per unit time cost during the server is in under repair process.
- C_{EV} - Per unit time cost during the server is in emergency vacation.
- C_{NV} - Per unit time cost during the server is in normal vacation.
- C_{CD} - Per unit time cost during the server is in closedown.
- C_{SU} - Per unit time cost during the server is in setup.
- C_1 - Cost obtained by the server in carrying out the normal service to HP/LP customers.
- C_2 - Cost obtained by the server in carrying out the optional service to HP customers.
- C_3 - Cost obtained by the server in carrying out the repair process.
- C_4 - Cost obtained by the server in carrying out the breakdown.
- C_5 - Cost obtained for the arrival of negative customers.
- C_6 - Cost obtained by the server in carrying out the emergency vacation.
- C_7 - Cost obtained by the server in carrying out the normal vacation.
- C_8 - Cost obtained in carrying out the closedown process.
- C_9 - Cost obtained by the server carrying out the setup process.

The total average cost per unit time is given by

$$\begin{aligned}TC = &C_{H_h}E_{System} + C_{H_l}E_{Orbit} + C_{SI}P_{SI} + C_{BH}P_{BH} + C_{BL}P_{BL} + C_{BHOS}P_{BHOS}\\ &+ C_RP_R + C_{EV}P_{EV} + C_{NV}P_{NV} + C_{CD}P_{CD} + C_{SU}P_{SU} + \mu C_1 + \mu_1 C_2\\ &+ \sigma C_3 + \tau C_4 + \lambda_3 C_5 + \eta_1 C_6 + \eta_2 C_7 + \phi C_8 + \psi C_9.\end{aligned}$$

8. Numerical Results

From this part, we determine the results of our model by representing numerically and graphically. The representations of MAP are distinct with the following variance and correlation structures, each of which has a mean value of 1. The arrival process as $ERL-A$, $EXP-A$ and $HYP-A$ correspond with renewal process, thus correlation is zero. This values are taken from Chakravarthy [10].

Positive Arrival in Erlang of order 2 (ERL-A):

$$D_0 = \begin{bmatrix} -4 & 4 \\ 0 & -4 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 \\ 2.8 & 0 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0 & 0 \\ 1.2 & 0 \end{bmatrix}$$

Positive Arrival in Exponential (EXP-A):

$$D_0 = [-1], \quad D_1 = [0.6], \quad D_2 = [0.4]$$

Positive Arrival in Hyper exponential (HYP-EXP-A):

$$D_0 = \begin{bmatrix} -1.90 & 0 \\ 0 & -0.19 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 1.026 & 0.114 \\ 0.1026 & 0.0114 \end{bmatrix}, \quad D_2 = \begin{bmatrix} 0.684 & 0.076 \\ 0.0684 & 0.0076 \end{bmatrix}$$

Negative Arrival in Erlang of order 2 (ERL-A):

$$C_0 = \begin{bmatrix} -0.5 & 0.5 \\ 0 & -0.5 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 0 & 0 \\ 0.5 & 0 \end{bmatrix}$$

Negative Arrival in Exponential (EXP-A):

$$C_0 = [-0.1], \qquad C_1 = [0.1]$$

Negative Arrival in Hyper exponential (HYP-EXP-A):

$$C_0 = \begin{bmatrix} -0.190 & 0 \\ 0 & -0.019 \end{bmatrix}, \qquad C_1 = \begin{bmatrix} 0.1710 & 0.0190 \\ 0.0171 & 0.0019 \end{bmatrix}$$

Let us consider the service and repair process as PH-distributions and these values are incurred from Chakravarthy [10] which are as follows:

ERL-S (Normal Service in Erlang of order 2):

$$\alpha = (1, 0), \qquad T = \begin{bmatrix} -25 & 5 \\ 8 & -25 \end{bmatrix}$$

ERL-S (Optional Service in Erlang of order 2):

$$\alpha = (1, 0), \qquad T = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

ERL-R (Repair in Erlang of order 2):

$$\beta = (1, 0), \qquad S = \begin{bmatrix} -2 & 2 \\ 0 & -2 \end{bmatrix}$$

EXP-S (Normal Service in Exponential):

$$\alpha = (1), \qquad T = [-1]$$

EXP-S (Optional Service in Exponential):

$$\alpha_1 = (1), \qquad T_1 = [-1]$$

EXP-R (Repair in Exponential):

$$\beta = (1), \qquad S = [-1]$$

HYP-EXP-S (Normal Service in Hyper exponential):

$$\alpha = (0.3, 0.7), \qquad T = \begin{bmatrix} -9 & 7 \\ 8 & -10 \end{bmatrix}$$

HYP-EXP-S (Optional Service in Hyper exponential):

$$\alpha_1 = (0.4, 0.6), \qquad T_1 = \begin{bmatrix} -12 & 6 \\ 5 & -10 \end{bmatrix}$$

HYP-EXP-R (Repair in Hyper exponential):

$$\beta = (0.4, 0.6), \qquad S = \begin{bmatrix} -6 & 4 \\ 3 & -4 \end{bmatrix}$$

8.1. Illustration 1

In tables 1,2 and 3, we determine the outcome of the repair rate of server (σ) on the expected system size (ES).
Fix $\lambda_1 = 0.8$, $\lambda_2 = 0.2$, λ_3, $\mu = 45$, $\mu_1 = 40$, $\eta_1 = 4$, $\eta_2 = 3$, $\varphi = 12$, $\psi = 12$, $\tau = 2$, $p = 0.5$, $q = 0.5$, $b = 0.05$, $\delta = 3$, $L = 3$.

Table 1: *Repair rate (σ) vs ES -* ***EXP-S***

σ	**EXP-A**	**ERL-A**	**HYP-A**
10	0.351226	0.119414	0.000351
10.5	0.351418	0.119473	0.000358
11	0.351632	0.119539	0.000365
11.5	0.351862	0.119609	0.000371
12	0.352100	0.119681	0.000376
12.5	0.352343	0.119756	0.000381
13	0.352588	0.119831	0.000385
13.5	0.352832	0.119906	0.000389
14	0.353074	0.119980	0.000392
14.5	0.353312	0.120054	0.000396

Table 2: *Repair rate (σ) vs ES -* ***ERL-S***

σ	**EXP-A**	**ERL-A**	**HYP-A**
10	0.211804	0.350087	0.042376
10.5	0.211885	0.350782	0.042475
11	0.211978	0.351445	0.042566
11.5	0.212078	0.352002	0.042650
12	0.212182	0.352487	0.042728
12.5	0.212289	0.352955	0.042800
13	0.212397	0.353404	0.042867
13.5	0.212504	0.353836	0.042930
14	0.212612	0.354250	0.042989
14.5	0.212717	0.354647	0.043044

Table 3: *Repair rate (σ) vs ES -* ***HYP-EXP-S***

σ	**EXP-A**	**ERL-A**	**HYP-A**
10	0.211852	0.112162	0.137225
10.5	0.211920	0.112284	0.137312
11	0.211995	0.112400	0.137393
11.5	0.212073	0.112509	0.137468
12	0.212154	0.112613	0.137538
12.5	0.212236	0.112712	0.137604
13	0.212319	0.112805	0.137666
13.5	0.212401	0.112893	0.137724
14	0.212481	0.112977	0.137779
14.5	0.212561	0.113057	0.137831

We observe that from the following tables 1,2, and 3.

- ES values rise for various combinations of arrival and service times as the server repair rate (σ) increases.
- When comparing the values of various service times, it can be seen that the expected system size increases more quickly for hyper exponential service times and slowly for Erlang service times.

8.2. Illustration 2

Using the two-dimensional graphs $2-10$, we investigate the impact of the normal service rate (μ) on the possibility that the server will be busy with high-priority customers (P_{BH}). Fix $\lambda_1 = 0.8$, $\lambda_2 = 0.2$, λ_3, $\mu_1 = 40$, $\sigma = 10$, $\eta_1 = 4$, $\eta_2 = 3$, $\tau = 2$, $\varphi = 12$, $\psi = 12$, $p = 0.5$, $q = 0.5$, $b = 0.05$, $\delta = 3$, $L = 3$.

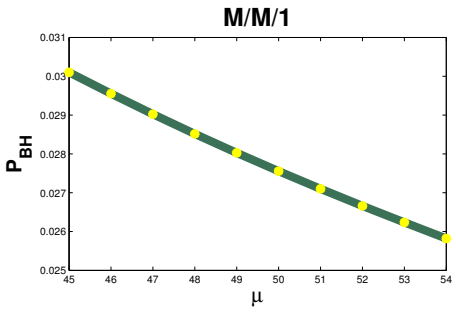


Figure 2: *Normal service rate (μ) vs. P_{BH}*

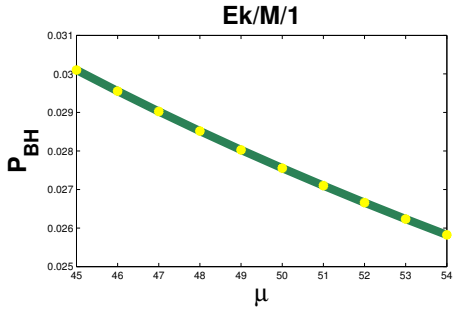


Figure 3: *Normal service rate (μ) vs. P_{BH}*

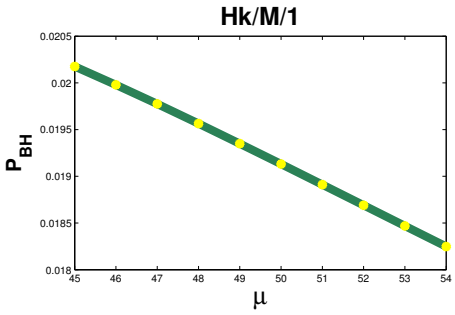


Figure 4: *Normal service rate (μ) vs. P_{BH}*

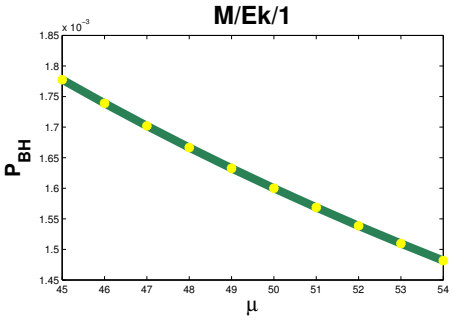


Figure 5: *Normal service rate (μ) vs. P_{BH}*

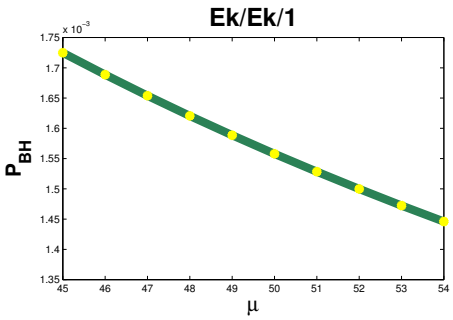


Figure 6: *Normal service rate (μ) vs. P_{BH}*

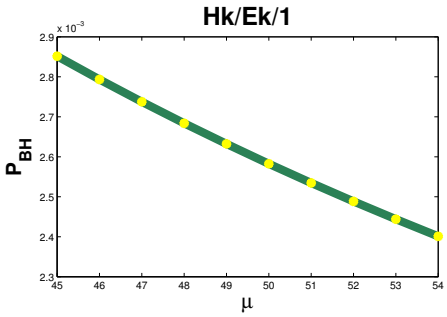


Figure 7: *Normal service rate (μ) vs. P_{BH}*

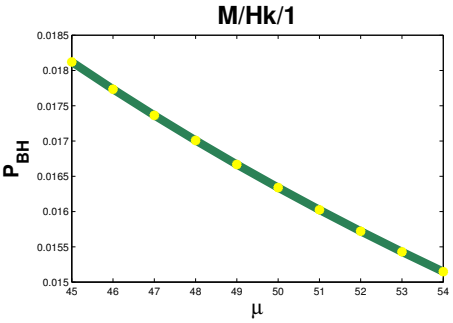


Figure 8: *Normal service rate (μ) vs. P_{BH}*

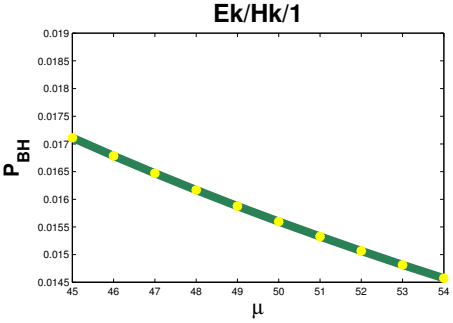


Figure 9: *Normal service rate (μ) vs. P_{BH}*

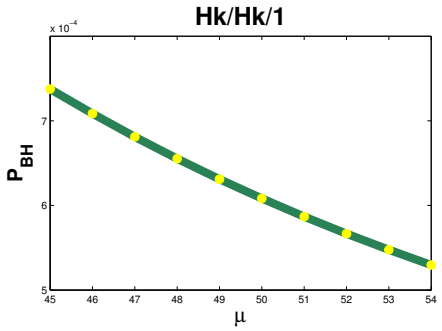


Figure 10: *Normal service rate (μ) vs. P_{BH}*

According to Figures $2-10$, as the normal service rate (μ) is raised, the probability that the server is busy with service (P_{BH}) increases for different arrival and service patterns. When increasing normal service rate (μ) on P_{BH} size increases much slower in Erlang arrival rather than hyper-exponential arrival.

8.3. Illustration 3

We investigate the impact of the normal service rate (μ) and breakdown rate (τ) on the probability that the server is busy with the normal service of high priority clients (P_{BH}) by using the three-dimensional graphs $11-19$. Fix $\lambda_1 = 0.8$, $\lambda_2 = 0.2$, λ_3, $\mu_1 = 40$, $\sigma = 10$, $\eta_1 = 4$, $\eta_2 = 3$, $\varphi = 12$, $\psi = 12$, $p = 0.5$, $q = 0.5$, $b = 0.05$, $\delta = 3$, $L = 3$.

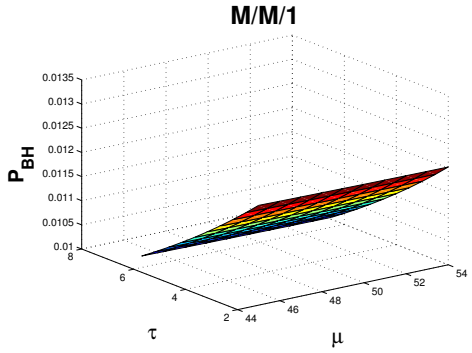


Figure 11: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

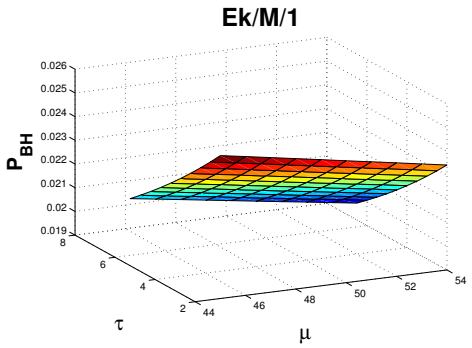


Figure 12: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

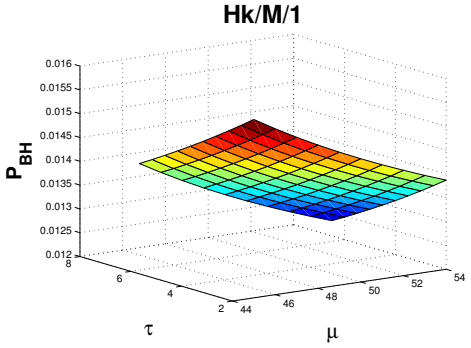


Figure 13: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

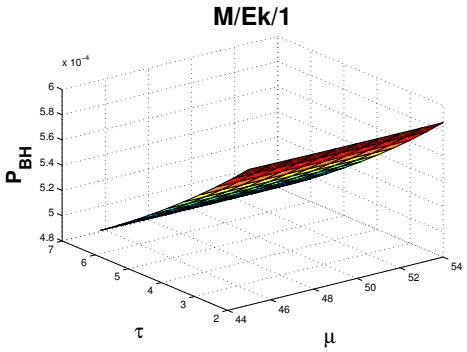


Figure 14: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

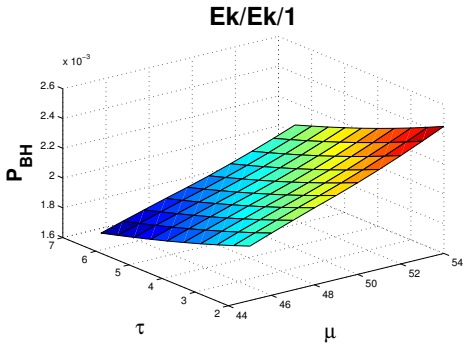


Figure 15: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

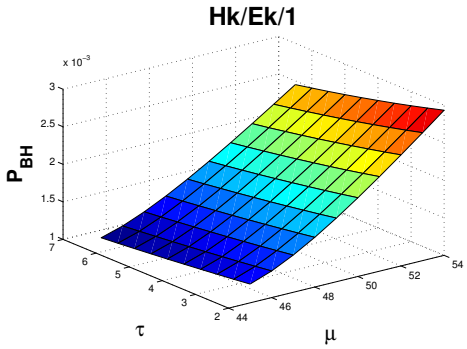


Figure 16: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

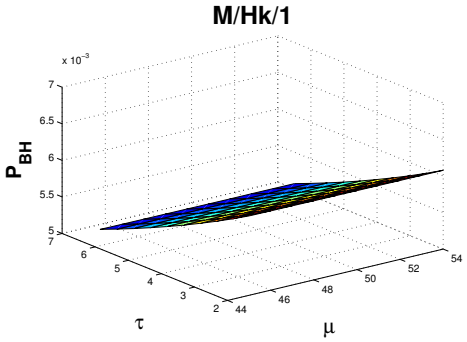


Figure 17: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

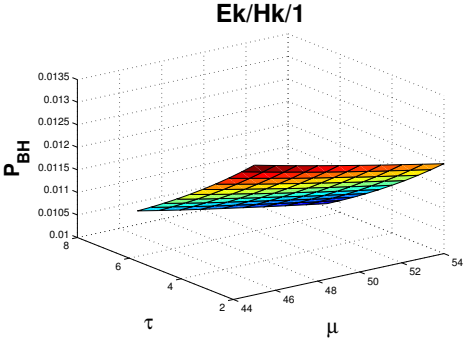


Figure 18: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

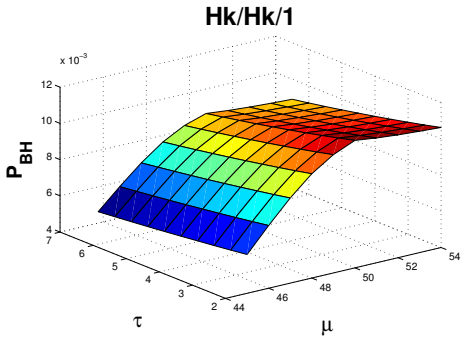


Figure 19: *(Service (normal) (μ) and Breakdown (τ) rates) vs. P_{BH}*

The probability that the server is busy with the normal service (P_{BH}) reduces for different arrival and service patterns when both the normal service rate (μ) and the breakdown rate (τ) are raised, as shown in Figures $11 - 19$. Rather than increasing at a hyper-exponential arrival, the Erlang arrival grows quickly. Similar to hyper-exponential services, the increment rate decreases for Erlang services.

9. Conclusion

In this paper, we have developed the queueing model with non preemptive priority queue, optional service, negative arrival, single vacation, emergency vacation, differentiate breakdown, repair, closedown , setup and balking. A queue with two categories of consumers with positive arrivals following $MMAP$, while negative arrival follows MAP and service times follows to be phase type distribution. By using matrix analytic method, we found the stationary probability since the queueing systems are Quasi Birth-Death process. The stability condition for the $MMAP/PH/1$ queuing system has analyzed and some performance measures for queueing system was selected and implemented in numerical illustrations by using three dimensional graphs. For further work, the model can be investigate with batch arrival and batch service which follows Markovian arrival process and various service rates with N-policy.

References

[1] Agarwal, P.K., Jain, A. and Jain, M. (2021). $M/M/1$ Queueing model with working vacation and two type of server breakdown, *Journal of Physics: Conference Series*, 1849(1): 1-15.

[2] Artalejo, J.R. and Corral, A.G. (2007). Modelling Communications Systems with Phase type service and Retrial times, *IEEE communications Letters*, 11(12): 955-957.

[3] Ayyappan, G. and Gowthami, R. (2019). Analysis of $MAP/PH/1$ Retrial Queue with Constant Retrial Rate, Bernoulli Schedule vacation, Bernoulli Feedback, Breakdown and Repair, *Reliability Theory and Applications*, 14(2(53)): 86-103.

[4] Ayyappan, G. and Gowthami, R. (2021). A $MAP/PH/1$ queue with Setup time, Bernoulli vacation, Reneging, Balking, Bernoulli feedback, Breakdown and Repair, *Reliability Theory and Applications*, 16(2(62)): 191-221.

[5] Ayyappan, G and Thilagavathy, K. (2020). Analysis of $MAP/PH/1$ Queueing model with Setup, Closedown, Multiple vacations, Standbyserver, Breakdown, Repair and Reneging, *Reliability Theory and Applications*, 15(2(57)): 104-143.

[6] Ayyappan, G. and Thilagavathy, K. (2021). Analysis of $MAP(1), MAP(2)/PH/1$ Non-preemptive priority Queueing model under classical Retrial policy with Breakdown, Repair, Discouragement, Single vacation, Standby server, Negative Arrival and Impatient customers, *Int. J. Appl. Comput. Math*, 7(184): 1-24.

[7] Ayyappan, G. and Udayageetha, J. (2018). Analysis of $M^{[X_1]}, M^{[X_2]}/G_1, G_2/1$ Retrial Queueing system with priority services, Working Breakdown, Non-Persistent customers, Modified Bernoulli vacation, Emergency vacation and Repair, *International Journal of Statistics and Systems*, 13(1): 23-39.

[8] Avrachenkov, K., Dudin, A. and Klimenok, V. (2010). Retrial Queueing Model $MMAP/M_2/1$ with two orbits, *Springer*, : 107-118.

[9] Baek, J., Dudina, O. and Kim, C. (2017). A Queueing system with Heterogeneous impatient customers and consumable additional items, *Int. J. Appl. Math. Comput. Sci*, 27(2): 367-384.

[10] Chakravarthy, S.R.(2011). Markovian Arrival Process, *Wiley Encyclopaedia of Operation Research and Management Science*.

[11] Chakravarthy, S.R.(2013). Analysis of $MAP/PH1, PH2/1$ Queue with Vacations and Optional secondary services, *Applied Mathematical Modelling*, 37(20-21): 8886-8902.

[12] Chakravarthy, S.R. and Agarwal, A.(2003). Analysis of a Machine Repair Problem with an Unreliable server and Phase type repairs and services, *Wiley Periodicals, Inc. Naval Research Logistics*, 50: 462-480.

[13] Chakravarthy, S.R. and Dudin, A. (2003). Analysis of a retrial queueing model with MAP arrivals and two types of customers, *Math. Comput. Modell*, 37: 343-363.

[14] Das, R.R., Rama Dev, V.N., Rathore, A. and Chandan, K. (2022). Analysis of Markovian queueing system with server failures, N-policy and second optional service, *Int. J. Nonlineaar Anal. Appl*, 13(1): 3073-3083.

[15] Fukagawa, Y., Yanaru, T. and Yoshida, S. (1987). Non-preemptive priority queue with server's walking process, *Electronics and Communications in Japan*, 70(11): pp. 49-57.

[16] Gao, S., Zhang, J. and Wang, X. (2020). Analysis of a retrial queue with two-type of Breakdowns and Delayed repairs, *IEEE Access*, 8(28-42): 172428-172442.

[17] Isotupa, K.P.S. and Stanford, D.A. (2002). An infinite- Phase Quasi-Birth and Death model for the non-preemptive priority $M/PH/1$ queue, *Stochastic Models*, 18(3): 387-424.

[18] Klimenok, V. and Dudin, A. (2012). A $BMAP/PH/N$ Queue with Negative customers and Partial Protection of service, *Communications in Statistics-Simulation*, 41: 1062-1082.

[19] Krishnamoorthy, A. and Divya, V. (2020). $(M, MAP)/(PH, PH)/1$ Queue with Non-preemptive priority and working vacation under N-policy, *Journal of the Indian Society for Probability and Statistis*, 21(3) 1-54.

[20] Krishnamoorthy, A., Babu, S. and Narayanan, V.C. (2008). $MAP/(PH, PH)/C$ Queue with Self- Generation of Priorities and Non-preemptive service, *Stochastic Analysis and Applications*, 26: 1250-1266.

[21] Latouche, G. and Ramaswami, V.(1999). Introduction of Matrix-Analytic Methods in stochastic modeling, *Society for Industrial and Applied Mathematics*, Philadelphia.

[22] Madan, K.C. (2000). An $M/G/1$ queue with second optional service, *Queueing systems*, 34: 37-46.

[23] Medhi,J.(1994). Stochastic processes, *J.Wiley*, New York.

[24] Nair, D.V., Krishnamoorthy, A., Melikov, A. and Aliyeva. S. (2021). $MMAP/(PH, PH)/1$ Queue with Priority Loss through Feedback, *Mathematics*, 9(15): 1-26.

[25] Neuts, M.F.(1981). Matrix-geometric solutions in stochastic models: An algorithmic approach, *The Johns Hopkins University Press*, Baltimore, London.

[26] Neuts, M.F.(1979). A Versatile Markovian point process, *Journal of Applied Probability*, 16(4): 764-779.

[27] Niu, Z., Shu, T. and Takahashi, Y.(2003). A Vacation queue with Setup and Closedown times and batch Markovian Arrival Processes, *Performance Evaulation*, 54: 225-248.

[28] Sleptchenko, A., Selen, J. and Adan, I. (2015). Joint queue length distribution of multi-class, single server queues with preemptive priorities, *Queueing System*, 81: 379-395.

[29] Swathi, Ch., Vasanta Kumar, V. and Hanumantha Rao, S. (2019). $M/M/1$ Queueing system with customer Balking and Renging, *International Journal of Innovative Technology and Exploring Engineering*, 8(9): 2636-2646.

[30] Wang, J.T. and Zhang, P.(2009). A Single-server Discrete-time Retrial G-queue with server Breakdowns and Repairs, *Acta Mathematicae Applicatae Sinica, English series*, 25(4): 675-684.

[31] Yuvarani, S. and Saravanarajan, M.c. (2017). Analysis of a preemptive priority retrial queue with negative customers, starting failure and at most J vacations, *Int. J. Knowledge Management in Tourism and Hospitality*, 1(1): 76-109 .

EXPONENTIAL - POISSON DISTRIBUTION IN RELIABILITY ACCEPTANCE SAMPLING PLAN FOR LIFE TESTING

Dr. V.Kaviyarasu[1] and A.Nagarajan[2]

•

(1).Associate Professor, Department of Statistics, Bharathiar University
(2).Research Scholar, Department of Statistics, Bharathiar University
E.mail: kaviyarasu@buc.edu.in , nagarajan.statistics@buc.edu.in

Abstract

Statistical Quality Control is an important field in production and maintenance of quality product in manufacturing environments. Reliability sampling plans (RSP) were widely employed in the sectors of manufacturing to monitor the quality of products in order to safe guard the producer as well as the consumer also the experimental costs and time can be saved. This article is developed on the reliability sampling plan when the evaluating life of the product is set to be truncated at pre-determined time follows Exponential-Poisson (EP) distribution. The probability of acceptance criteria for the single sampling is designed to achieve the lowest sample size for such proposed two parameter probability distribution with the corresponding decision rule. This study is conducted to design plan parameters on the basis of desired quality levels such as Acceptable Reliability Quality Level (ARQL), Indifference Reliability Quality Level (IRQL) and Rejectable Reliability Quality Level (RRQL). This study computes the median life for the specified producer's risk, its OC curve is provided along with the minimum ratio values. Furthermore, it determines the minimum size of the samples and the acceptance number. Table values have been obtained and provided for single sampling plan. Additionally, suitable examples are provided to conduct a study on a real time situations.

Keywords: Exponential – Poisson (EP) Distribution, Reliability, Median lifetime, Single Sampling Plan.

1. Introduction

Quality has become an inevitable term in the modern statistical society, especially in manufacturing sector. In such environment, every product must satisfy the required quality standards to achieve the goal. The act of employing statistical techniques to monitoring the quality and to maintaining the quality of the manufactured product in a systematic way is known as Statistical Quality Control (SQC). Due to technological advancements through mass production, it is an impossible to inspect every single product from a lot (i.e. 100% inspection is not feasible) and accepting a lot without inspection is also not acceptable hence both consumer and producer facing certain risks. So acceptance sampling is an important statistical technique to safeguard the consumer as well as the producer also. Here, the risks are termed as Producer risk (α) and Consumer risk (β) are the risk involved in the process of decision making.

Acceptance sampling is initially employed in the US military to test the quality of bullets from World War II and it acts as a vital tool in SQC, which focuses to make decisions about whether or not to accept a lot on the basis of the quality of randomly selected sample from a lot. This technique consists of the lot having size 'N' and 'n' is known to be the sample number of units and

'c' is the acceptance number. Reliability sampling plan is the one of the most important method in acceptance sampling which helps to assess the quality of product using time. Various techniques are employed to evaluate the quality of such a manufactured item to test the reliability of the item which is called as life test method here failure of an item follows a continuous probability distributions are adopted to model this methodology.

This paper is studied under the attribute sampling plan is studied to discover the life of an item to test the lower confidence limit on median life. According to Gupta and Groll (1961)[6], the median life constitutes a superior quality parameter than that of the average life for a skewed distribution. Our aim is to decrease the financial expenses and also the investment of time of the experimenter simultaneously truncated life test is studied to test the test termination time for the fixed time 't'. One can count the total number of failures occur during this process within the specified time, if no c failures occur prior to the scheduled time limit. If not, the experiment is terminated after the $(c+1)^{th}$ failure. Based on the values of the operating characteristics, the methodology for the smallest sample size is to be necessary for guaranteeing that the product's designated median life has been given along with the associated producer risk is presented here. An appropriate example have been discussed with suitable illustrations. The foremost objective of this article is to constitute a time truncated single sampling plan for median life under Exponential Poisson (EP) distribution.

2. Review of Literature

There has been extensive research about reliability sampling plans on the basis of truncated life tests done by various authors. Baklizi and El Masri (2004) [1] were studied acceptance sampling for Birnbaum–Saunders model, Barreto-Souva and Silva (2013) [2] pointed out that EP is better alternative to the gamma distribution, Cameron. J.M. (1952) [3] were studied about the construction of tables based on OC function of single sampling plans, Dodge. H.F and H.G. Romig (1959) [4] conducted a study on sampling inspection tables, Epstein (1954) [5] proposed a truncated test for the exponential case, Gupta and Groll (1961) [6] were conducted a study about acceptance sampling under Gamma distribution, Kaviyarasu and Fawaz (2017) [7] carried out a study on acceptance sampling on the modified weibull distribution, Kus (2007) [8] introduced a new life time distribution called as Exponential Poisson distribution(EP), Schilling and John (1980) [9] constructed a set of tables for various sampling plans, Sobel and Tischendrof (1959) [10] were studied about new life test objectives for acceptance sampling.

3. The Exponential-Poisson Distribution

The Exponential Poisson (EP) distribution is a two parameter continuous probability distribution that is used to model the time between events in a real time to test the life of an item. EP distribution is a compounded distribution under Exponential and zero truncated Poisson distribution. This distribution has several real time applications such as Network traffic modelling, manufacturing quality control, service queue management and stock price modelling etc.,.

According to Barreto-Souza and Silva (2013)[2], EP distribution is better alternative to the Gamma distribution. For a lifetime and reliability studies EP distribution performs a significant role in modelling the lifetime of the products.

The Cumulative distribution function of Exponential - Poisson distribution is

$$F(x;\lambda,\beta) = (e^{\lambda \exp(-\beta x)} - e^{\lambda})(1 - e^{\lambda})^{-1} \quad (1)$$

The Probability density function of Exponential - Poisson distribution is

$$f(x;\lambda,\beta) = \frac{\lambda\beta}{(1-e^{-\lambda})} e^{-\lambda-\beta x+\lambda \exp(-\beta x)} \quad (2)$$

Here $\lambda > 0$, $\beta > 0$ are the shape and scale parameter. Where λ is also known as Poisson parameter. When $\lambda \to 0$, the Exponential distribution is obtained by reducing the EP distribution.

The median function of EP distribution is

$$log\{\log [2^{-1}(e^{\lambda}+1)]^{-1}\lambda\}\beta^{-1} \quad (3)$$

i.e. $(T \leq \beta_0) = \beta$

$$\beta_0 = -\beta^{-1}\ log\{\log [2^{-1}(e^{\lambda}+1)]^{-1}\lambda\} \quad (4)$$

$$\eta = -\ log\{\log [2^{-1}(e^{\lambda}+1)]^{-1}\lambda\} \quad (5)$$

$$\Longrightarrow t_q = {}^{\eta}/_{\beta}$$

$$\Longrightarrow \beta = {}^{\eta}/_{\beta_0}$$

By substituting the scale parameter $\beta = {}^{\eta}/_{\beta_0}$, the CDF of EP distribution becomes

$$F(t) = (e^{\lambda \exp(-\frac{t}{\beta_0}*\eta)} - e^{\lambda})(1-e^{\lambda})^{-1} \quad t>0, \eta>0$$

Let $\delta = \frac{t}{\beta_0}$

$$F(t;\delta) = (e^{\lambda \exp(\delta)} - e^{\lambda})(1-e^{\lambda})^{-1} \quad , t>0,\ \delta>0 \quad (6)$$

4. Truncated Acceptance Sampling Plan

In acceptance sampling, a well-known simple plan is single sampling plan (SSP) and it has employed in many reliability studies. Here the product's lifetime (T) is assumed to follows the Exponential – Poisson (EP) distribution. The shape parameter λ is considered as a known parameter. The product's median life can be represented by m. In a truncated acceptance life test plans, the usual practice of testing the lifetime of the product is to terminating the experiment at a time (t) which has already determined. The decision of the acceptance criteria is purely based only the occurrence of defectives. If the total count of defectives is below the given acceptance number c, then it should be accepted. The main target of this experiment is to acquire a designated median life along with the help of probability P* (Consumer risk) and also to frame a lower confidence limit. For conducting a truncated life test experiment the following components should be considered.

- The total number of sample units on the experiment (n);
- The acceptance number (c); when the total count of defectives occurred is more than c at the final stage of pre-decided time then the inspected lot will be approved for the acceptance.
- The ratio t/β_0; Here 'β_0' is known to be a described median life and the maximum amount of time for the experiment is known as 't'.

5. Minimum Sample Size

The chance cause of accidently accepting a lot without knowing that the chosen lot is a poor is known to be consumer risk (α) and it is fixed to not greater than 1-P*. Since the chance of accepting a poor lot with a median is at the minimum of P*, then it is evident that P* represents the confidence level. The lot size must be taken into consideration as being infinite and must be assumed to be sufficiently large, so in this case binomial distribution is employed to evaluate the lot acceptance. To find smallest sample size (n) such that

$$\sum_{x=0}^{n}\binom{n}{x}p^{x}(1-p)^{n-x} \leq 1-p^{*} \quad (7)$$

In table 1, the minimum values (n) were presented that satisfies the above inequality, for $\frac{t}{\beta_0}$ = 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3 and P* = 0.75, 0.90, 0.95, 0.99 and c = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

TABLE 1: *Minimum sample sizes for the EP distribution.*

P*	N	t/β_0									
		0.3	**0.6**	**0.9**	**1.2**	**1.5**	**1.8**	**2.1**	**2.4**	**2.7**	**3**
0.75	0	7	4	3	2	2	2	2	1	1	1
0.75	1	13	7	5	4	4	3	3	3	3	3
0.75	2	19	11	8	6	6	5	5	4	4	4
0.75	3	25	14	10	8	7	7	6	6	5	5
0.75	4	31	17	13	10	9	8	8	7	7	7
0.75	5	36	20	15	12	11	10	9	9	8	8
0.75	6	42	23	17	14	12	11	11	10	9	9
0.75	7	47	27	20	16	14	13	12	11	11	10
0.75	8	53	30	22	18	16	14	13	13	12	12
0.75	9	58	33	24	20	18	16	15	14	13	13
0.75	10	64	36	27	22	19	18	16	15	15	14
0.9	0	11	6	4	3	3	2	2	2	2	2
0.9	1	18	10	8	6	5	4	4	4	3	3
0.9	2	25	14	10	8	7	6	6	5	5	5
0.9	3	32	17	13	10	9	8	7	7	6	6
0.9	4	38	21	15	12	11	10	9	8	8	7
0.9	5	44	24	18	14	13	11	10	10	9	9
0.9	6	50	28	20	17	14	13	12	11	11	10
0.9	7	56	31	23	19	16	15	13	13	12	11
0.9	8	62	34	25	21	18	16	15	14	13	13
0.9	9	68	38	28	23	20	18	16	15	15	14
0.9	10	74	41	30	25	22	19	18	17	16	15
0.95	0	13	7	5	4	3	3	3	2	2	2
0.95	1	22	12	8	7	6	5	5	4	4	4
0.95	2	29	14	11	9	8	7	6	6	5	5
0.95	3	36	19	14	11	10	9	8	7	7	7
0.95	4	43	23	17	14	12	10	10	9	8	8
0.95	5	49	27	19	16	14	12	11	10	10	9
0.95	6	55	30	22	18	16	14	13	12	11	11
0.95	7	61	34	25	20	17	16	14	13	13	12
0.95	8	68	37	27	22	19	17	16	15	14	13
0.95	9	74	41	30	24	21	19	17	16	15	15
0.95	10	80	44	32	26	23	21	19	18	17	16
0.99	0	20	10	7	5	4	4	3	3	3	3
0.99	1	30	15	10	8	7	6	5	5	5	4
0.99	2	35	19	14	11	9	8	7	7	6	6
0.99	3	45	24	17	13	11	10	9	8	8	7
0.99	4	54	28	19	16	13	12	11	10	9	9
0.99	5	61	32	22	18	16	14	13	12	11	10
0.99	6	68	36	25	21	18	16	14	13	12	12
0.99	7	75	40	29	23	20	17	16	16	14	13
0.99	8	82	42	31	25	21	19	17	16	15	15
0.99	9	88	48	34	27	23	21	19	18	17	16
0.99	10	95	49	37	29	25	23	21	19	18	17

6. Operating Characteristic Function

The Operating Characteristics (OC) function of the Acceptance sampling based on the Truncated Life Test (ASTLT) plan consists with the parameters of $(n, c, \frac{t}{\beta_0})$. For analysing the ASTLT, the probability is

$$L(p) = \text{Prob \{Accepting a good lot\}}$$

$$L(p) = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x} \quad (8)$$

Where p = F(t; θ) is a monotonically decreasing function of $\beta > \beta_0$. Based on the above inequality the operating characteristics (OC) values of $\frac{t}{\beta_0}$ were displayed in table 2.

TABLE 2: *OC values for (n, c =4, $t/\beta_0 = 0.60$) for a given P* under EP distribution.*

P*	N	t/β_0	β/β_0								
			2	4	6	8	10	12	14	16	18
0.75	31	0.3	0.1515	0.6882	0.8939	0.9587	0.9816	0.9909	0.9951	0.9972	0.9983
0.75	17	0.6	0.1567	0.681	0.8883	0.9556	0.98	0.99	0.9946	0.9969	0.9981
0.75	10	1.2	0.1726	0.6737	0.8801	0.9507	0.9772	0.9884	0.9937	0.9963	0.9977
0.75	8	1.8	0.1542	0.6271	0.8507	0.935	0.9688	0.9837	0.9909	0.9946	0.9966
0.75	7	2.4	0.1419	0.5882	0.8228	0.9188	0.9596	0.9783	0.9876	0.9925	0.9953
0.75	7	3	0.0655	0.4278	0.707	0.8494	0.9188	0.9538	0.9724	0.9828	0.9888
0.9	38	0.3	0.0182	0.3597	0.69	0.8515	0.9247	0.9593	0.9767	0.9859	0.9911
0.9	21	0.6	0.0173	0.3358	0.6639	0.8338	0.9138	0.9526	0.9724	0.9832	0.9893
0.9	12	1.2	0.0235	0.3399	0.6563	0.8251	0.9072	0.948	0.9694	0.9811	0.9879
0.9	10	1.8	0.0115	0.2359	0.5407	0.7405	0.8515	0.912	0.9459	0.9655	0.9772
0.9	8	2.4	0.0224	0.284	0.5814	0.7657	0.8662	0.9207	0.9511	0.9688	0.9794
0.9	7	3	0.0303	0.301	0.5882	0.7662	0.8646	0.9188	0.9494	0.9674	0.9783
0.95	43	0.3	0.0041	0.214	0.5458	0.7562	0.8667	0.924	0.9546	0.9718	0.9818
0.95	23	0.6	0.0052	0.2158	0.5412	0.7503	0.8619	0.9205	0.9522	0.9701	0.9806
0.95	14	1.2	0.0036	0.163	0.4595	0.6814	0.8129	0.8873	0.9299	0.9549	0.9701
0.95	10	1.8	0.0082	0.2021	0.4986	0.7073	0.8285	0.8966	0.9355	0.9584	0.9724
0.95	9	2.4	0.0042	0.1376	0.3977	0.6162	0.7595	0.8474	0.9009	0.934	0.9549
0.95	8	3	0.0044	0.1294	0.3745	0.5889	0.7354	0.8282	0.8864	0.9232	0.9469
0.99	54	0.3	0.0002	0.0659	0.3056	0.5499	0.719	0.8236	0.8869	0.9256	0.9498
0.99	28	0.6	0.0004	0.0756	0.3209	0.562	0.7268	0.8284	0.8899	0.9275	0.951
0.99	16	1.2	0.0026	0.0684	0.292	0.5251	0.6936	0.8021	0.8701	0.9129	0.9403
0.99	12	1.8	0.0006	0.0634	0.269	0.4934	0.6633	0.7772	0.8507	0.8982	0.9292
0.99	10	2.4	0.0007	0.0599	0.2507	0.4662	0.6359	0.7536	0.8318	0.8835	0.9178
0.99	9	3	0.0005	0.0481	0.2132	0.4155	0.5862	0.711	0.7975	0.8566	0.897

7. Producer Risk

The chance of rejecting a lot without knowing the chosen lot is satisfying the quality requirements is known as producer risk (β), when $\beta > \beta_0$ it will be computed as

$$\text{Prob(p) = Prob \{Rejecting a lot\} = 1 - Prob \{Accepting a lot\}}$$

For a single sampling plan and the definite values of producer's risk, one may very curious in finding the value estimates of $\frac{\beta}{\beta_0}$ is going to guarantee the producer's risk which is not greater than or equal to 0.05 on the basis of the employed sampling plan. $\frac{\beta}{\beta_0}$ is having the values which are known to be the smallest non-negative integer for $p = F\left(\frac{t}{\beta_0}\frac{\beta_0}{\beta}\right)$ which satisfies the below mentioned inequality.

$$\sum_{i=1}^{c} \binom{n}{x} p^i (1-p)^{n-i} \geq 0.95 \quad (9)$$

The minimum values of $\frac{t}{\beta_0}$ satisfying the above inequality to the proposed sampling plan $(n, c, \frac{t}{\beta_0})$ at a specific confidence level P* were presented in table - 3.

TABLE 3: *Minimum ratio of true mean life over β_0 at the producer's risk of 0.05*

		t/β_0									
P*	C	**0.3**	**0.6**	**0.9**	**1.2**	**1.5**	**1.8**	**2.1**	**2.4**	**2.7**	**3**
0.75	0	31.171	35.869	40.333	35.642	44.484	53.61	62.603	35.236	39.857	44.375
0.75	1	24.053	24.977	25.813	26.566	33.198	27.825	32.642	37.306	41.727	46.363
0.75	2	15.17	16.605	17.279	16.251	20.314	19.187	22.383	18.473	20.779	23.087
0.75	3	11.794	12.401	12.452	12.512	13.017	15.62	14.602	16.689	13.883	15.507
0.75	4	10.091	10.286	11.176	10.579	11.468	11.568	13.496	12.468	14.026	15.604
0.75	5	8.764	8.9857	9.4756	9.4254	10.447	10.881	10.737	12.281	11.265	12.519
0.75	6	8.1086	8.1481	8.4278	8.6487	8.6371	9.0264	10.531	10.292	9.5259	10.57
0.75	7	7.4575	7.9256	8.1975	8.0653	8.3097	8.8619	9.0839	8.8853	9.9959	9.1862
0.75	8	7.1177	7.4086	7.54	7.6484	8.0468	7.8036	7.9908	9.1324	8.8559	9.8399
0.75	9	6.7202	7.0283	7.028	7.2891	7.8226	7.7769	8.1294	8.1754	7.9518	8.8354
0.75	10	6.5121	6.7127	6.9987	7.0168	7.051	7.7368	7.3757	7.4428	8.3901	8.0692
0.9	0	63.773	70.043	70.271	70.099	87.897	70.592	81.507	93.52	105.22	117.13
0.9	1	14.854	16.199	19.378	19.065	19.77	18.346	21.481	24.549	19.835	22.132
0.9	2	8.9139	9.7523	10.292	10.765	11.623	11.729	13.679	12.646	14.227	15.806
0.9	3	6.8377	7.0802	7.9756	7.9752	8.8518	9.3044	9.3117	10.642	9.9191	11.022
0.9	4	5.6044	6.0612	6.3422	6.6176	7.4869	8.0562	8.3495	8.2644	9.2975	8.7511
0.9	5	4.9038	5.2175	5.7527	5.8153	6.675	6.6329	6.9162	7.9003	7.8167	8.6849
0.9	6	4.426	4.846	5.074	5.6602	5.676	6.268	6.6585	6.8541	7.7109	7.6222
0.9	7	4.0913	4.4232	4.8269	5.2174	5.3801	5.9826	5.9171	6.7624	6.9	6.8725
0.9	8	3.841	4.1194	4.444	4.8948	5.1468	5.3923	5.8277	6.1435	6.3087	7.0097
0.9	9	3.6471	3.9867	4.3196	4.6404	4.9662	5.2773	5.3647	5.6783	6.3881	6.5131
0.9	10	3.4921	3.787	4.0623	4.443	4.8175	4.8834	5.3427	5.713	5.965	6.128
0.95	0	75.159	81.468	87.582	93.706	87.907	105.88	123.15	93.52	105.47	117.38
0.95	1	18.232	19.636	19.368	22.382	23.766	23.512	27.359	24.51	27.58	30.645
0.95	2	10.412	9.7842	11.4	12.251	13.44	13.916	13.683	15.66	14.232	15.813
0.95	3	7.7168	7.9544	8.6374	8.8752	9.9753	10.623	10.875	10.642	11.972	13.239
0.95	4	6.3752	6.6457	7.2535	7.8503	8.277	8.0892	9.4373	9.5422	9.2972	10.359
0.95	5	5.4765	5.8962	6.0749	6.7414	7.2681	7.3145	7.7384	7.9003	8.8878	8.6849
0.95	6	4.8859	5.1978	5.6126	6.0094	6.6146	6.8245	7.2994	7.6096	7.7102	8.5677
0.95	7	4.4704	4.882	5.2738	5.5147	5.7538	6.4561	6.4517	6.7624	7.6077	7.6667
0.95	8	4.2249	4.4986	4.8191	5.1542	5.4735	5.7862	6.2791	6.6603	6.9115	7.0097
0.95	9	3.9791	4.3181	4.6546	4.8692	5.2427	5.6116	5.7678	6.1244	6.3853	7.0948
0.95	10	3.7839	4.0732	4.3627	4.6362	5.055	5.4839	5.6991	6.111	6.4303	6.6407
0.99	0	116.71	114.93	121.94	117.08	115.67	138.58	121.54	140.99	158.8	176.56
0.99	1	24.936	24.705	24.367	25.823	27.987	28.519	27.211	31.093	34.966	30.645
0.99	2	12.575	13.457	14.693	15.215	15.315	16.127	16.23	18.622	17.617	19.575
0.99	3	9.6924	10.114	10.59	10.629	11.091	11.973	12.369	12.43	13.983	13.239
0.99	4	8.0504	8.172	8.1848	9.0434	9.0172	9.9322	10.493	10.74	10.735	11.924
0.99	5	6.8508	7.0482	7.1377	7.6714	8.4186	8.6986	9.367	9.7673	9.9505	9.8752
0.99	6	5.4395	5.5758	5.6231	6.0237	6.6008	6.8107	7.2994	7.6096	7.7102	7.6222
0.99	7	5.5184	5.7897	6.1898	6.4359	6.8933	6.9043	7.5359	8.6124	8.2951	8.4531
0.99	8	5.1131	5.141	5.6015	5.9245	6.1088	6.5678	6.7506	7.1953	7.4928	8.3253
0.99	9	4.7508	5.0853	5.3166	5.5287	5.8006	6.2968	6.5557	7.0449	7.4067	7.6639
0.99	10	4.5104	4.5646	5.0915	5.2251	5.5537	6.0677	6.3979	6.5112	6.8822	7.1316

8. Numerical Illustration

In today's modern world, every region has its own food culture that is influenced by the environment, agriculture, whether, and so on. Nowadays, the beverage industry plays a significant role in the global food industry. Coca-Cola is the only beverage that is mass-produced and distributed globally. Coca-Cola was first sold in Atlanta in 1886. Despite the passage of many decades, the demand for Coca-Cola keeps on increasing on every single day. As a result, a beverage manufacturing company in the United States intends to increase Coca-Cola production. The exponential process is used to describe the increase in Coca-Cola production, and the Poisson process is used to describe the probability of manufacturing defects. Therefore, it is ensured that this production process is carried out using an exponential Poisson process.

Here it is considered that the Exponential-Poisson distribution is the appropriate distribution for evaluating life time of an item with the parameters λ=2. The quality inspector desires to investigate the median lifetime of an item has 1000 hours when the confidence level is P*= 0.75. The test was terminated after 600 hours. This leads to the ratio $t/\beta_0 = 600/1000 = 0.6$. The sampling plan which is used by the experimenter is (n=17, c=4, $t/\beta_0 = 0.60$).

Table-4: *OC curve for the plan (17, 4, 0.6) under EP for* $p^* = 0.75$

P	2	4	6	8	10	12	14	16	18
L(P)	0.15669	0.681	0.88834	0.95563	0.98003	0.99005	0.99462	0.9969	0.99811

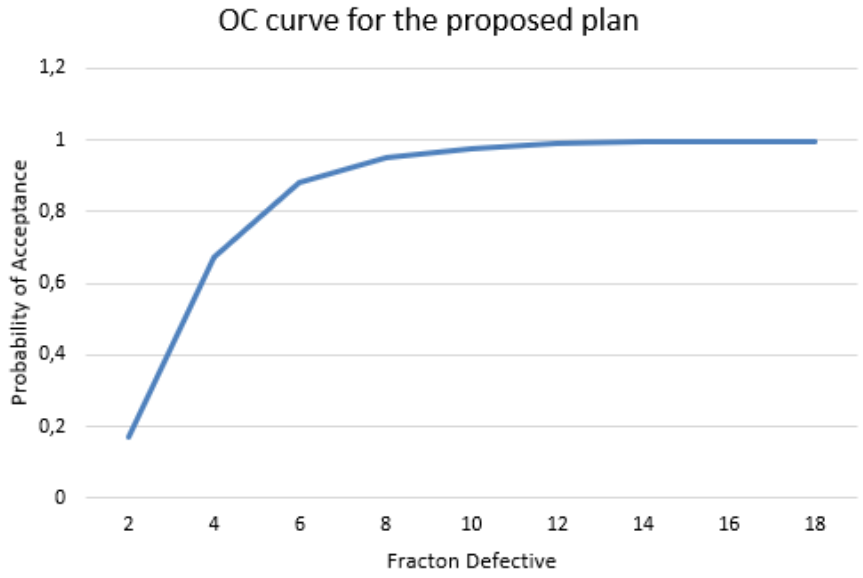


Figure-1: *OC curve for the plan (17, 4, 0.6) under EP for* $p^* = 0.75$.

9. Construction of tables

Step 1: Set the parameters λ=2 and the test termination ratio t/β_0 such as

0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3

Step 2: To find the value of η, substitute the parameters in the equation (5)

One can obtain η as 0.332832. Substitute $\beta = \eta/\beta_0$ in (1),

To find p, use the inequality p= $F(t, \delta)$

Step 3: By satisfying the inequality, determine the smallest sample size n.

$$\sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \leq 1 - p^*$$

Step 4: Utilize the inequality to determine the OC values.

$$L(p) = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i}$$

Step 5: By satisfying the given inequality, determine the minimum mean ratio at Producer risk $\alpha = 0.05$

$$\sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \geq 1 - \alpha$$

10. Conclusion

This article is developed for time truncated single sampling plan which follows a pre-fixed time when lifetime of the products follows an Exponential-Poisson distribution. The required minimum sample size and OC values of Producer risk were displayed in the given tables in order to guarantee the determined median life along with a confidence level that is given. This study reveals that the EP distribution proves that the sample size is much smaller than other statistical distributions which are used in acceptance sampling. Further, the table values are explained with suitable illustration.

References

[1] Baklizi, A. and El Masri, A. E. Q. (2004). Acceptance sampling based on truncated life tests in the Birnbaum Saunders model. *Risk Analysis: An International Journal*, Vol. *24*(6), pp. 1453-1457.

[2] Barreto-Souza, W. and Silva, R. B. (2015). A likelihood ratio test to discriminate exponential–Poisson and gamma distributions. *Journal of Statistical Computation and Simulation*, Vol. *85*(4), pp. 802-823.

[3] Cameron, J. M. (1952). Tables for constructing and for computing the operating characteristics of single sampling plans. *Industrial Quality Control*, Vol. *9*(1), pp. 37-39.

[4] Dodge, H. F. and Roming, H. G. (1959). *Sampling inspection tables* (No. 311.21 D63 1959).

[5] Epstein, B. (1954). Truncated life tests in the exponential case. *The Annals of Mathematical Statistics*, pp.555-564.

[6] Gupta, S. S. and Gupta, S. S. (1961). Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association*, Vol. *56*(296), pp. 942-970.

[7] Kaviayarasu, V. and Fawaz, P. (2017). A Reliability sampling plan to ensure percentiles through Weibull Poisson distribution. *International Journal of Pure and Applied Mathematics*, Vol. *117*(13), pp. 155-163.

[8] Kuş, C. (2007). A new lifetime distribution. *Computational Statistics & Data Analysis*, Vol. *51*(9), pp. 4497-4509.

[9] Schilling, E. G. and Johnson, L. I. (1980). Tables for the construction of matched single, double, and multiple sampling plans with application to MIL-STD-105D. *Journal of Quality Technology*, Vol. *12*(4), pp. 220-229.

[10] Sobel, M. and Tischendrof, J. A. (1959). Acceptance sampling with new life test objectives. In *Proceedings of fifth national symposium on reliability and quality control* Vol. 108, p. 118.

CHARACTERIZATION OF POISSON TYPE LENGTH BIASED EXPONENTIAL CLASS SOFTWARE RELIABILITY GROWTH MODEL AND PARAMETER ESTIMATION

Rajesh Singh[1], Kailash R. Kale[2] , Pritee Singh[3]

1. R. T. M. Nagpur University, Nagpur-440033,
rsinghamt@hotmail.com
2. G. N. A. ACS College, Barshitakli, Dist-Akola.
kailashkale10@gmail.com
3. Institute of Science, Nagpur;
priteesingh25@gmail.com

Abstract

The authors of this study set out to build a software reliability growth model (SRGM). Software reliability is a crucial attribute that has to be quantified and evaluated. In most cases, software errors happen at unpredictable times. In this article, the failure intensity of the single parameter length-biased exponential class SRGM has been characterized taking into account the Poisson process of the incidence of software faults. The parameters of the proposed SRGM under investigation are the scale parameter (θ_1) and the total number of failures (θ_0). It is considered that the experimenter may have previous knowledge of the parameters from past or earlier experiences in the form of gamma priors. The posterior probability may be obtained by combining the prior probability with the likelihood of the data, and Bayes estimators can then be suggested.

Keywords: Binomial process, gamma prior, maximum likelihood estimator (MLE), Rayleigh class, software reliability growth model (SRGM), incomplete gamma function, confluent hyper-geometric function.

I. Introduction

Beginning in the early 1970s of the previous century, research on software reliability has advanced until the present. Various kinds of software have dominated many fields of the humanities, sciences, and technology as well as daily life for all people. The film industry, education sector, E-commerce sector, medical and healthcare sectors, space agencies, banking sector and various government agencies, all employ different types of software for the convenience and development the fields.

The software is the end product of several intricate code sequences developed by the humans according to the needs of above sectors within stipulated time period. Due to huge magnitude of complicated code sequences, there is a greater likelihood of failures or ineffective performance. These software failures may result from a variety of issues, including memory faults, language-specific issues, calling third-party libraries, standard library issues, etc. Such flaws may have operational repercussions that cause system failure and unanticipated dangerous outcomes.

As a result, the aforementioned areas require software that operates reliably. Hence, the evaluation and quantification of the software's performance are therefore crucial. The reliability of the software is one of the performance indicators. To put it another way, it becomes crucial to create reliable software that serves the needs of users or systems.

Software reliability is now thought to be a crucial factor in determining customer satisfaction, along with software functionality and performance. Software reliability growth models (SRGM) outline the broad link between software failure occurrences and the key process influences (such as fault introduction, fault removal, operational profile, etc.). The statistical relation between data on defects and the known characteristics of probabilistic behavior is known as the SRGM. The basic goal of software reliability modeling is to represent a relationship in which, when defects are found and removed, there is a reduction in the number of failures per time interval or an increase in the time interval between failures. The SRGM is often characterized by the mean failure function or failure intensity function. The pattern of occurrence of software failure is its type, and the mathematical functional form of failure intensity is its class. The software reliability growth models are categorized according to the system described by [8].

The length-biased distributions have been presented by [2] and formalized by [9]. These distributions are sometimes referred to as size-biased probability distributions. Reliability theory may also use these distributions (see [4], [5], and [6]). Modeling software reliability may be done using length-biased distributions. In this study, using the Poisson pattern of occurrence of software failure and the length-biased exponential form of failure intensity, the Poisson type length-biased exponential class model is introduced as per the classification system provided by [8]. As this SRGM is being described, it is assumed that the failure occurring at time t has a Poisson occurrence (i.e., Type) and that the mean failure function's functional form is characterized by a length-biased exponential distribution (i.e., Class). The software failures in this model are presumed to be independent of one another and dependent on the duration of the time interval that comprises the same software failure. For the estimation part of parameters the gamma priors taking into account. The Bayes estimators of the parameters are obtained in this study by the methods of [7], [12], [10], and [11], and they are compared with MLEs in subsequent parts.

II. Model Section

Suppose time to failure follows length biased exponential distribution denoted by $f(t)$ with scale parameter θ_1 and software failures occur in Poisson pattern then

$$f(t) = \begin{cases} t\theta_1^2 e^{-\theta_1 t} ; & t > 0, \theta_1 > 0, E[t] \neq 0 \\ 0 ; & otherwise \end{cases} \quad (1)$$

Also let the total number of faults remaining in the software at time $t = 0$ is a Poisson random variate with mean θ_0 then the failure intensity $\lambda(t) = \theta_0 f(t)$ (cf. [7]) can be obtained as

$$\lambda(t) = \theta_0 t\theta_1^2 e^{-\theta_1 t} ; \quad t > 0, \theta_1 > 0, \theta_0 > 0 \quad \text{and } E[t] \neq 0 \quad (2)$$

The mean failures function at time t comes out to be

$$\mu(t) = \theta_0 \left[1 - (1 + \theta_1 t)e^{-\theta_1 t}\right] \quad (3)$$

The details about number of failures experienced by time t, performance of failure intensity λ(t) and $\mu(t)$ have been discussed in [13].

III. Maximum Likelihood Estimation

The most important and extensively used technique of point estimation is maximum likelihood estimation when underlying distribution of data is known. The maximum likelihood estimation is considered for failure times. The base of maximum likelihood estimation is likelihood function

which can be obtained by assuming that m_e failures are experienced at times $t_i, i = 1,2, \dots, m_e$ up to execution time is t_e $(\geq t_{m_e})$. Also using the failure intensity at each $t_i, i = 1,2, \dots, m_e$ obtained in (2) and mean failure function at time t_e obtained by replacing $t = t_e$, the likelihood function of θ_0 and θ_1 can be obtained as $L(\theta_0, \theta_1) = \{\prod_{i=1}^{m_e} \lambda(t_i)\} exp\,(-\mu(t_e))$ (cf. Musa et al. (1987)).

The $L(\theta_0, \theta_1)$ can take following form for this model

$$L(\theta_0, \theta_1) = \theta_0^{m_e} \theta_1^{2m_e} \left[\prod_{i=1}^{m_e} t_i\right] e^{-T\theta_1} e^{-\theta_0 \left[1-(1+\theta_1 t_e)e^{-\theta_1 t_e}\right]} \quad (4)$$

where

$\sum_{i=1}^{m_e} t_i = T$

The Maximum Likelihood Estimators for the parameters θ_0 and θ_1 are

$$\hat{\theta}_{m0} = m_e\left(1 - (1 + \hat{\theta}_{m1} t_e)e^{-\hat{\theta}_{m1} t_e}\right)^{-1} \quad (5)$$

and

$$\hat{\theta}_{m1} = \left[\hat{\theta}_{m0}^{-1} t_e^{-2} (2m_e - T\theta_{m1}) e^{\hat{\theta}_{m1} t_e}\right]^{1/2} \quad (6)$$

respectively. The values of $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ can be obtained after simultaneous solution of equations (5) and (6).

IV. Bayesian parameter estimation

The bayesian technique is used to put the subjective and objective data sources together into the analysis. In this technique the parameters are considered as a random variables having known probability pattern. This known probability pattern is termed as prior in Bayesian technique. Whole the analysis is based on this prior and using Bayes theorem combines this prior and likelihood of data. In present case, it is considered that the experimenter have prior information about both the parameters θ_0 and θ_1 in the form of gamma probability function. Then the following prior distributions $g(\theta_0)$ and $g(\theta_1)$ can be considered for parameters θ_0 and θ_1 respectively.

$$g(\theta_0) \propto \begin{cases} \theta_0^{\nu-1} e^{-\eta\theta_0} & , \theta_0 \epsilon\, [0, \infty) \\ 0 & , otherwise \end{cases} \quad (7)$$

and

$$g(\theta_1) \propto \begin{cases} \theta_1^{\alpha-1} e^{-\beta\theta_1} & , \theta_1 \epsilon\, [0, \infty) \\ 0 & , otherwise \end{cases} \quad (8)$$

Now, Consider the total execution time is t_e and during this time m_e failures are experienced at times t_i, $i = 1,2, \dots, m_e$ then, the joint posterior of θ_0 and θ_1 given $\underline{t}(= t_i, i = 1,2, \dots, m_e)$ is

$$\pi\left(\theta_0, \theta_1 \middle| \underline{t}\right) \propto \theta_0^{m_e+\nu-1} \theta_1^{2m_e+\alpha-1} e^{-(T+\beta)\theta_1} e^{-(\eta+1)\theta_0}\, e^{\left[\theta_0(1+\theta_1 t_e)e^{-\theta_1 t_e}\right]} \;\; \theta_0 > m_e\,, \theta_1 > 0 \quad (9)$$

In this section, the point estimates (posterior mean) of both the parameters θ_0 and θ_1 under study are obtained by Bayesian technique considering the squared error loss as

$$\hat{\theta}_{B0} = D^{-1} \sum_{j=0}^{\infty} \frac{\Gamma(m_e+\nu+j+1, (\eta+1)m_e)}{j!(\eta+1)^{(j+1)}} \Psi(2m_e + \alpha, 2m_e + \alpha + j + 1, T^* t_e^{-1}) \quad (10)$$

and

$$\hat{\theta}_{B1} = \frac{(2m_e+\alpha)}{D t_e} \sum_{j=0}^{\infty} \frac{\Gamma(m_e+\nu+j, (\eta+1)m_e)}{j!(\eta+1)^j} \Psi(2m_e + \alpha + 1, 2m_e + \alpha + j + 2, T^* t_e^{-1}) \quad (11)$$

where $\Psi(\alpha, \beta; x)$ is Confluent Hypergeometric Function (cf. [1] and [3]), normalizing constant is

$D = \sum_{j=0}^{\infty} \frac{\Gamma(m_e+\nu+j, (\eta+1)m_e)}{j!(\eta+1)^j} \Psi(2m_e + \alpha, 2m_e + \alpha + j + 1, T^* t_e^{-1})$

and

$T^* = T + \beta + j t_e.$

V. Discussion

I. Subsection One

The proposed Bayes estimators i.e. $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ are compared with corresponding maximum likelihood estimators i.e. $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ respectively on the basis of risk efficiencies $RE_j = R_j' R_j^{-1}$ where $R_j = E\left[\hat{\theta}_{Bj} - \theta_j\right]^2$ and $R_j' = E\left[\hat{\theta}_{mj} - \theta_j\right]^2$; $j = 0,1$. Here, the performance of proposed Bayes estimators $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ over MLEs $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ have been compared on the basis of risks efficiencies using Monte Carlo simulation technique. The risks efficiencies are obtained by generating sample of size, say m_e failures upto total execution time t_e and it was repeated 10^3 times from the length biased exponential distribution. Then, using Monte Carlo simulation technique risks efficiencies has been evaluated and is presented in the graphs Figure 1 to 9.

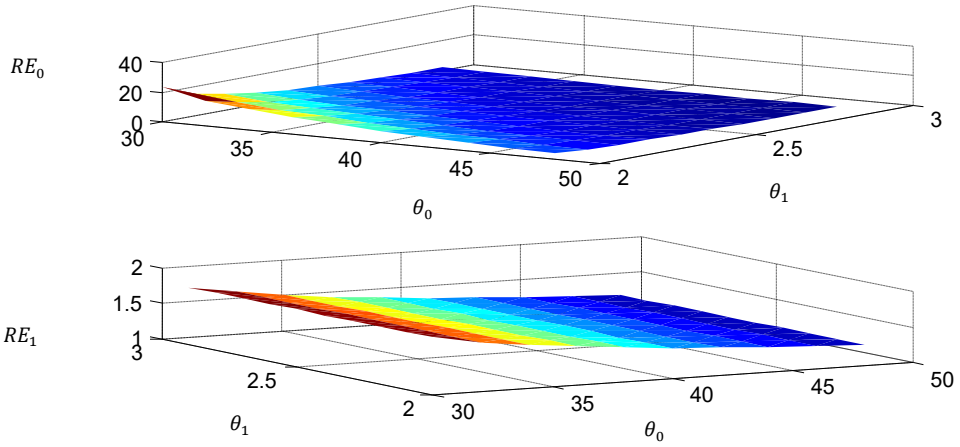


Figure 1: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ for t_e = 100; ϑ = 1, η = 1; α = 1, β = 1,*

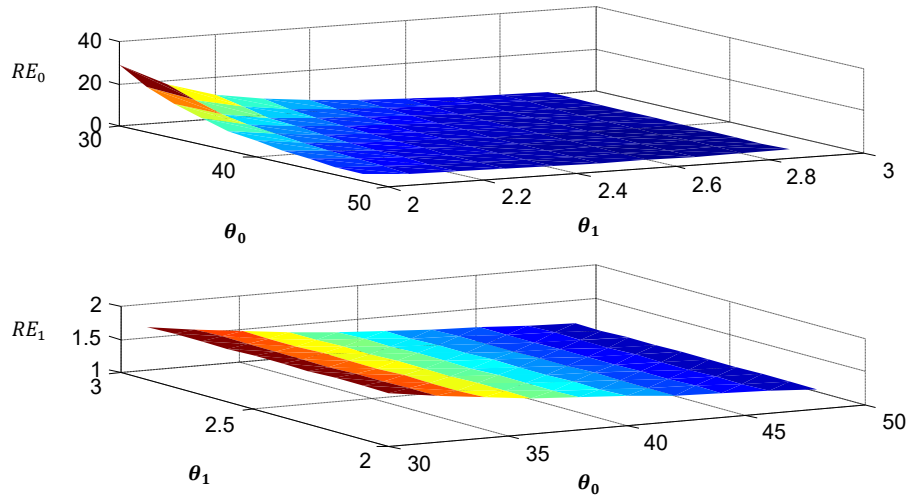


Figure 2: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 125; ϑ = 1, η = 1; α = 1, β = 1*

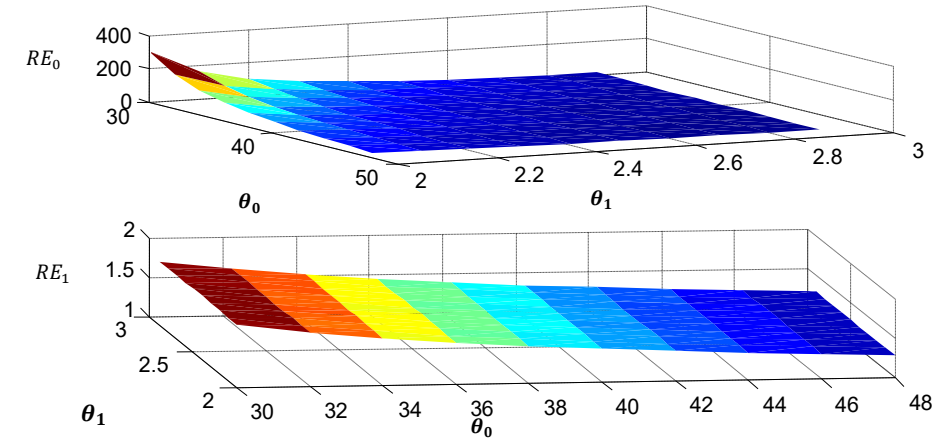


Figure 3: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_{B1}$ $t_e = 150$; $\vartheta = 1$, $\eta = 1$; $\alpha = 1$, $\beta = 1$,

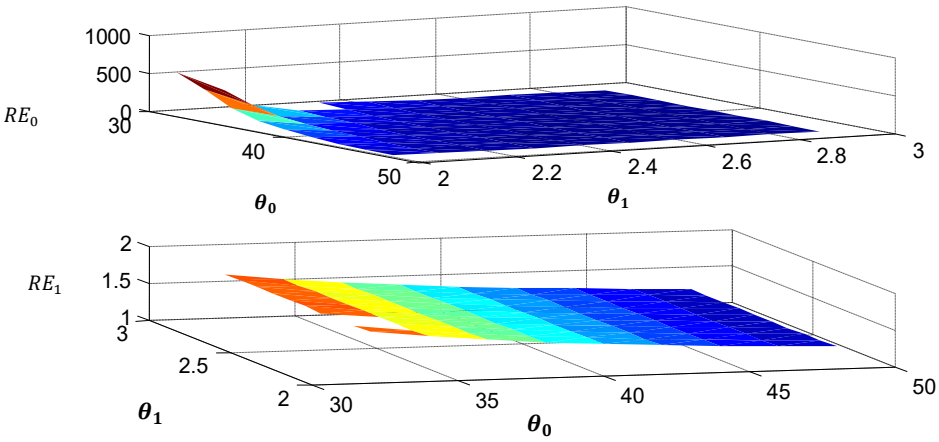


Figure 4: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_{B1}$ $t_e = 200$; $\vartheta = 1$, $\eta = 1$; $\alpha = 1$, $\beta = 1$

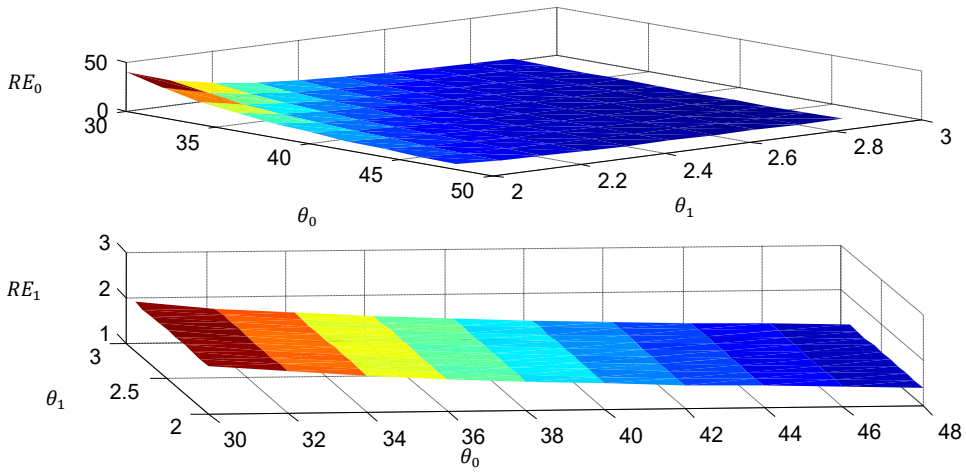


Figure 5: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_{B1}$ $t_e = 100$; $\vartheta = 10$, $\eta = 1$; $\alpha = 10$, $\beta = 1$,

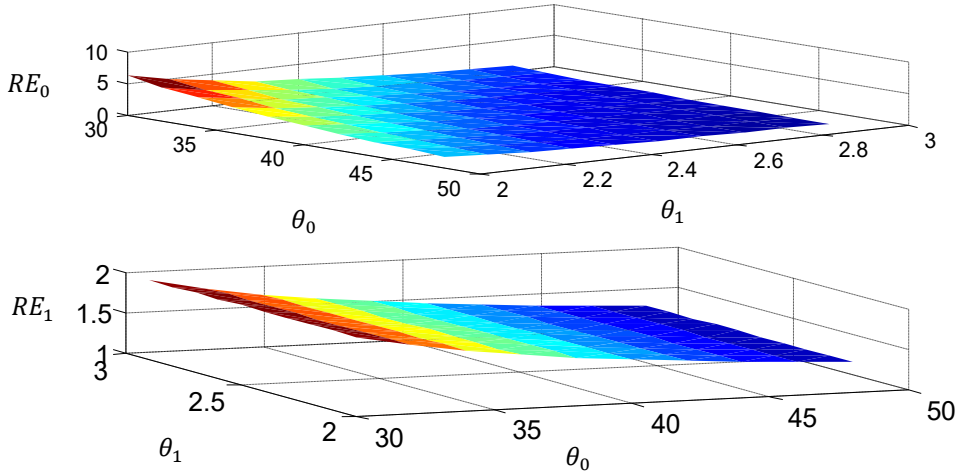


Figure 6: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_{B1}$ $t_e = 100$; $\vartheta = 10$, $\eta = 5$; $\alpha = 10$, $\beta = 5$,

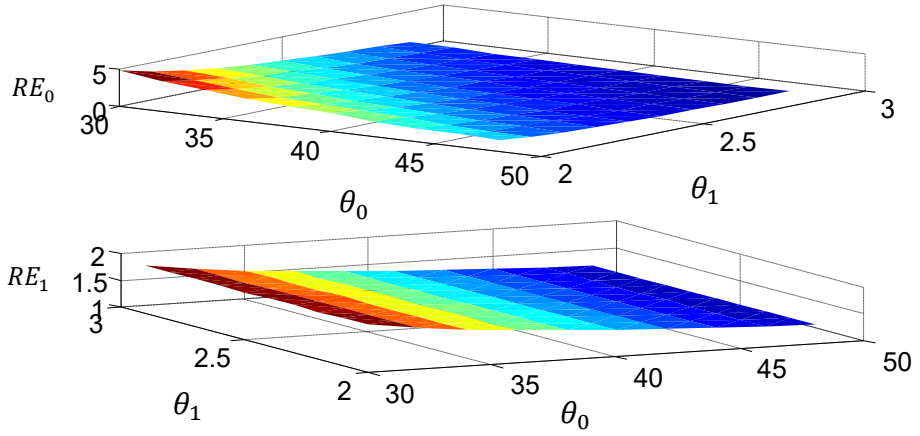


Figure 7: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_{B1}$ $t_e = 100$; $\vartheta = 10$, $\eta = 10$; $\alpha = 10$, $\beta = 10$,

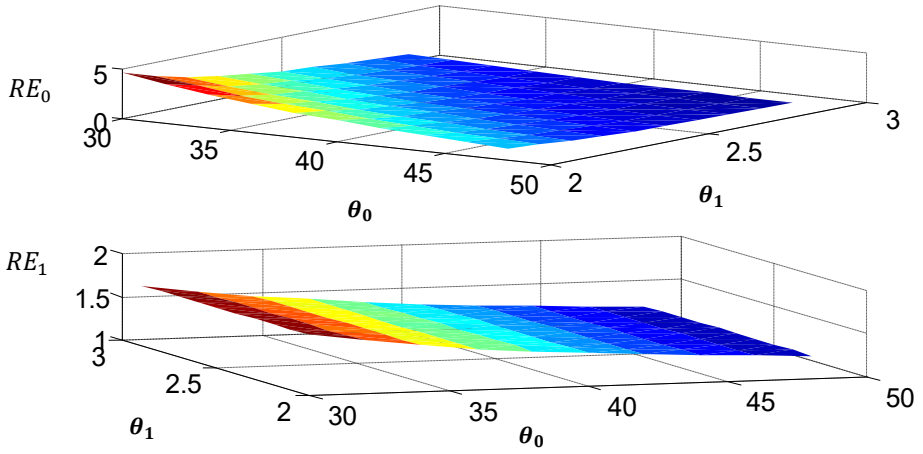


Figure 8: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_{B1}$ $t_e = 100$; $\vartheta = 1$, $\eta = 10$; $\alpha = 1$, $\beta = 10$,

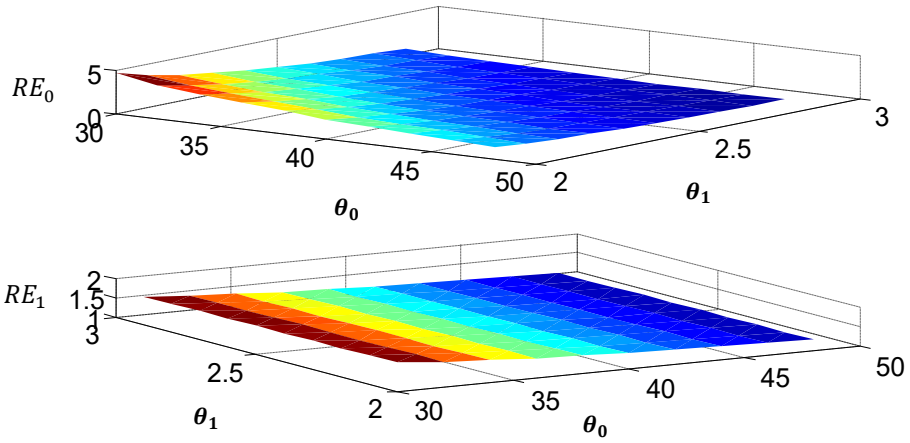


Figure 9: *Risk Efficiencies* $\hat{\theta}_{B0}$ *and* $\hat{\theta}_B$ t_e *= 100;* ϑ *= 3,* η *= 10;* α *= 3,* β *= 10,*

II. SUBSECTION TWO

Based on the above graphical representation from Figure 1 to Figure 9 of performance of proposed Bayes estimators against their corresponding MLE, it can be seen that the risk efficiencies RE_0 of $\hat{\theta}_{B0}$ decrease as θ_0 and θ_1 increase. It can also be seen that for large values of θ_1 and θ_0 the proposed Bayes estimator of θ_0 is not better than MLE otherwise proposed Bayes estimator $\hat{\theta}_{B0}$ is better than MLE. Moreover, when the value of t_e is small, the values of RE_0 first increase, attain a maxima and then decrease as the value of t_e increase. Similarly, it can be observed that the risk efficiencies of $\hat{\theta}_{B1}$ i.e. RE_1 decrease as the value of θ_1 and $\theta_0$increase but the values of risk efficiencies RE_1 are almost constant for the increase in values of θ_0. Further, The values of RE_1 are uniform over the variation in value of t_e. It is important to note that the proposed Bayes estimator $\hat{\theta}_{B1}$ is always better than MLE. Due to increase in values of shape and scale parameter of both the priors the values of RE_0 decrease for constant values of scale parameters.

On the basis of better performance of risk efficiencies of $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ over $\hat{\theta}_{m0}$ and $\hat{\theta}_{m0}$ following conclusions are drawn.

VI. Conclusions

After having experience or prior knowledge about the software failure process to researchers. These proposed Bayes estimators can perform better than their corresponding MLEs for the proper choices of prior parameters. The proposed Bayes estimator of θ_0 can be preferred over MLE if it is felt that total number of failures may not be very large and failure rate may be small. The proposed Bayes estimator of θ_1 can be preferred over MLE. Under this prior belief these estimators can be preferred for calendar time modeling.

References

[1] Abramowitz, M. and Stegun, I. A. "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", New York: Dover publications, 1965.

[2] Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies," *Ann. Eugenics*, 6:13-25.

[3] Gradshteyn, I. S. and Ryzhik, I. M. Table of Integrals, Series, and Products, Alan Jeffrey (editor) 5th Ed., New York: Academic Press, 1994.

[4] Gupta, R. C. and Keating, J. P. (1986). Relations for Reliability Measures under Length Biased Sampling, *Scand Journal of Statistics*, 13: 49-56.

[5] Gupta, R. C. and Tripathi, R. C. (1990). Effect of length-biased sampling on the modeling error, *Communication in statistics –Theory and Methods*, 19(4):1483-1491.

[6] Khatree, R. (1989). Characterization of Inverse-Gaussian and Gamma distributions through their length-biased distributions, *IEEE Transactions on Reliability*, 38(5):610-611.

[7] Musa, J. D. Iannino, A. and Okumoto, K. Software Reliability: Measurement, Prediction, Application, New York:McGraw-Hill, 1987.

[8] Musa, J. D. and Okumoto, K. (1984). A logarithmic Poisson execution time model for software reliability measurement, *Proceedings of Seventh International conference on software engineering Orlando,* 230-238.

[9] Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment." In Classical and Contagious Discrete Distributions, Eds. G.P. Patil, *Pergamon Press and Statistical Publishing Society. Calcutta*, 320-332.

[10] Singh, R. and Andure, N. W. (2008). Bayes estimators for the parameters of the Poisson type exponential distribution", *IAPQR transactions,* 33(2):121-128.

[11] Singh, R. and Singh, P. (2012). Characterization of some Software Reliability Growth Models and Parameter Estimation, *MRP-UGC, New Delhi*, 2012 (Submitted to UGC New Delhi).

[12] Singh, R. Vidhale, A. A. and Carpenter, M. (2009). Bayes estimators of parameters of Poisson Type Exponential Class Software Model considering generalized Poisson and Gamma priors, *Journal of Model Assist. Statis. Appl.,* 4(2)83-89.

[13] Singh, R. Singh, P. and Kale, K. R. (2016). Bayes estimators for the parameters of Poisson Type Length Biased Exponential Class Model using Non-Informative Priors, *Journal of Reliability and Statistical Studies*, 9(1):21-28.

[14] Singh, R. Singh, P. and Kale, K. R. (2022). Estimation of Parameters of PTRC SRGM using Non-informative Priors. *International Journal of Advanced Research in Science, Communication and Technology*, 2(1)172-178.

POWER WEIGHTED SUJATHA DISTRIBUTION WITH PROPERTIES AND APPLICATION TO SURVIVAL TIMES OF PATIENTS OF HEAD AND NECK CANCER

Rama Shanker

•

Department of Statistics, Assam University, Silchar, Assam, India
Shankerrama2009@gmail.com

Kamlesh Kumar Shukla

•

Department of Community Medicine, Noida International Institute of Medical Sciences, Noida International University, Gautam Budh Nagar, India
Kkshukla22@gmail.com

Abstract

In this paper a power weighted Sujatha distribution, which includes power Sujatha distribution, weighted Sujatha distribution and Sujatha distribution as particular cases, has been proposed. Its statistical properties including behavior of density function, moments, hazard rate function, and mean residual life function have been discussed. Estimation of parameters has been discussed using the method of maximum likelihood. A simulation study has been presented to know the performance of maximum likelihood estimates of parameters. Application of the proposed distribution have been explained with a real lifetime data relating to patients suffering from head and neck cancer and goodness of fit shows quite satisfactory fit.

Keywords: *Sujatha distribution, Weighted Sujatha distribution, Power Sujatha distribution, Hazard rate function, Mean residual life function, Maximum Likelihood estimation*

1. Introduction

It has been observed that the survival times of patients suffering from head and neck cancer needs special consideration to find a suitable distribution which can be used to model the data. During recent decades several one parameter, two-parameter and three-parameter lifetime distributions have been proposed in statistics literature to model survival times of patients suffering from head and neck cancer and observed that all proposed distributions are not very much suitable due to theoretical or applied point of view. It has been observed that, in general, the survival times of patients suffering from head and neck cancer are stochastic in nature and while discussing the goodness of fit of several one parameter , two-parameter and three-parameter well-known distributions which were earlier proposed by different researchers that these distribution does not give good fit.

Shanker [1] proposed a one parameter Sujatha distribution having its probability density function (pdf) and cumulative distribution function (cdf) as

$$f_1(y;\theta)=\frac{\theta^3}{\theta^2+\theta+2}(1+y+y^2)e^{-\theta y};y>0,\theta>0 \tag{1.1}$$

$$F_1(y;\theta)=1-\left[1+\frac{\theta y(\theta y+\theta+2)}{\theta^2+\theta+2}\right]e^{-\theta y};y>0,\theta>0(1.2) \tag{1.2}$$

Shanker and Shukla [2], taking a weight function $x^{\alpha-1}$in (1.1), proposed a two-parameter weighted Sujatha distribution (WSD) defined by its pdf and cdf

$$f_2(y;\theta,\alpha)=\frac{\theta^{\alpha+2}}{\theta^2+\alpha\theta+\alpha(\alpha+1)}\frac{y^{\alpha-1}}{\Gamma(\alpha)}(1+y+y^2)e^{-\theta y};y>0,\theta>0,\alpha>0 \tag{1.3}$$

$$F_2(y;\theta,\alpha)=1-\frac{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha,\theta y)+(\theta y)^{\alpha}(\theta y+\theta+\alpha+1)e^{-\theta y}}{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}, \tag{1.4}$$

where $\Gamma(\alpha,z)=\int_z^{\infty}e^{-y}y^{\alpha-1}dy;y\geq 0,\alpha>0$ is the upper incomplete gamma function.

Shanker and Shukla [3], taking a power transformation $x=y^{\frac{1}{\beta}}$ in (1.1), proposed a two-parameter power Sujatha distribution(PSD) defined by its pdf and cdf

$$f_3(x;\theta,\alpha)=\frac{\alpha\theta^3}{\theta^2+\theta+2}x^{\alpha-1}(1+x^{\alpha}+x^{2\alpha})e^{-\theta x^{\alpha}};x>0,\theta>0,\alpha>0 \tag{1.5}$$

$$F_3(y;\theta,\alpha)=1-\left[1+\frac{\theta y^{\alpha}(\theta y^{\alpha}+\theta+2)}{\theta^2+\theta+2}\right]e^{-\theta y^{\alpha}};y>0,\theta>0,\alpha>0 \tag{1.6}$$

Ghitany *et al*. [4] proposed a two-parameter weighted Lindley distribution (WLD) having parameters θ and α defined by its probability density function (pdf) and cumulative distribution function (cdf)

$$f_4(y;\theta,\alpha)=\frac{\theta^{\alpha+1}}{\theta+\alpha}\frac{y^{\alpha-1}}{\Gamma(\alpha+1)}(1+y)e^{-\theta y};y>0,\theta>0,\alpha>0 \tag{1.7}$$

$$F_4(y;\theta,\alpha)=1-\frac{(\theta+\alpha)\Gamma(\alpha,\theta y)+(\theta y)^{\alpha}e^{-\theta y}}{(\theta+\alpha)\Gamma(\alpha)};y>0,\theta>0,\alpha>0, \tag{1.8}$$

where $\Gamma(\alpha)$ and $\Gamma(\alpha,z)$ are the complete gamma function and the upper incomplete gamma function. Its structural properties including moments, hazard rate function, mean residual life function, estimation of parameters and applications for modeling survival time data has been discussed by Ghitany *et al*. [4]. Shanker *et al.* [5] discussed various moments based properties including coefficient of variation, coefficient of skewness, coefficient of kurtosis and index of dispersion of weighted Lindley distribution and its applications to model lifetime data from biomedical sciences and engineering.

Ghitany *et al.* [6] proposed a power Lindley distribution (PLD) having parameters θ and α defined by its pdf and cdf

$$f_5(y;\theta,\alpha)=\frac{\alpha\theta^2}{(\theta+1)}y^{\alpha-1}(1+y^{\alpha})e^{-\theta y^{\alpha}};y>0,\theta>0,\alpha>0 \tag{1.9}$$

$$F_5(x;\theta,\alpha)=1-\left[1+\frac{\theta x^{\alpha}}{\theta+1}\right]e^{-\theta x^{\alpha}};x>0,\theta>0,\alpha>0 \tag{1.10}$$

Note that the PLD is a convex combination of Weibull (α,θ) and a generalized gamma $(2,\alpha,\theta)$distribution with mixing proportion$\frac{\theta}{\theta+1}$. Ghitany *et al*. [6] has discussed the properties of PLD including the shapes of the density, hazard rate functions, moments, skewness and kurtosis measures, estimation of parameters using maximum likelihood estimation and application to model a real lifetime data from engineering. Recall that at $\alpha=1$ both WLD in (1.7) and PLD in (1.9) reduce to Lindley distribution introduced by Lindley [7] having pdf and cdf

$$f_6(y;\theta)=\frac{\theta^2}{\theta+1}(1+y)e^{-\theta y};y>0,\theta>0 \tag{1.10}$$

$$F_6(y;\theta)=1-\left[1+\frac{\theta y}{\theta+1}\right]e^{-\theta y};y>0,\theta>0 \tag{1.11}$$

Ghitany *et al*. [8] have discussed its various statistical and mathematical properties and application. Shanker *et al*. [9] have detailed study on modeling of lifetime data using both exponential and

Lindley distributions and observed that there are many lifetime data where exponential distribution gives better fit than Lindley distribution.

Zakerzadeh and Dolati [10] have introduced a three parameter generalized Lindley distribution (GLD) having pdf and cdf given by

$$f_7(x;\theta,\alpha,\beta)=\frac{\theta^{\alpha+1}}{\theta+\beta}\frac{x^{\alpha-1}}{\Gamma(\alpha+1)}(\alpha+\beta x)e^{-\theta x};x>0,\theta>0,\alpha>0,\beta>0 \quad (1.12)$$

$$F_7(x;\theta,\alpha,\beta)=1-\frac{\alpha(\beta+\theta)\Gamma(\alpha,\theta x)+\beta(\theta x)^{\alpha}e^{-\theta x}}{(\beta+\theta)\Gamma(\alpha+1)};x>0,\theta>0,\alpha>0,\beta>0 \quad (1.13)$$

Lindley distribution, gamma distribution and weighted Lindley distribution (WLD) are particular cases of (1.9) at $(\alpha=\beta=1)$, $(\beta=0)$ and $(\beta=\alpha)$, respectively. Shanker [11] obtained various raw moments and central moments of GLD and discussed properties based on moments including coefficient of variation, skewness, kurtosis and index of dispersion of GLD and its comparative study with generalized gamma distribution (GGD) introduced by Stacy [12] to model various lifetime data from engineering and biomedical sciences and concluded that in many cases GGD gives much better fit than GLD. Shanker and Shukla [13] have detailed comparative study on modeling of real lifetime data from engineering and biomedical sciences using GLD and generalized gamma distribution (GGD) introduced by Stacy (1962) [12] and concluded that there are several lifetime data where GGD gives much better fit than GLD.

In the present paper, a three - parameter power weighted Sujatha distribution (PWSD) which includes PSD, WSD and Sujatha distribution as particular cases, has been proposed and discussed. Its raw moments been obtained. The survival function and the hazard rate function of the distribution have been derived and their shapes have been discussed for varying values of the parameters. The estimation of its parameters has been discussed using maximum likelihood method. Finally, the goodness of fit and the application of the distribution have been explained through a real lifetime data relating to patients suffering from head and cancer and the fit has been compared with other one parameter, two-parameter and three-parameter lifetime distributions.

2. Power weighted Sujatha distribution

Assuming the power transformation $X=Y^{\frac{1}{\beta}}$ in the pdf of WSD (1.3), the pdf of the random variable Xcan be obtained as

$$f_8(x;\theta,\alpha,\beta)=\frac{\beta\theta^{\alpha+2}}{\theta+\alpha\theta+\alpha(\alpha+1)}\frac{x^{\beta\alpha-1}}{\Gamma(\alpha)}\left(1+x^{\beta}+x^{2\beta}\right)e^{-\theta x^{\beta}};x>0,\theta>0,\alpha>0,\beta>0 \quad (2.1)$$

We would call the distribution in (2.1) as the power weighted Sujatha distribution (PWSD) and we denote it as PWSD(θ,α,β). It can be easily verified that the PWSD(θ,α,β) in (2.1) reduces to Sujatha distribution, WSD and PSD for$(\alpha=\beta=1)$,$(\beta=1)$ and$(\alpha=1)$, respectively. It can be easily verified that PWSD is a convex combination of generalized gamma distributions with different parameters, namely GGD(θ,α,β), GGD$(\theta,\alpha+1,\beta)$ and GGD$(\theta,\alpha+2,\beta)$. That is

$$f_5(x;\theta,\alpha,\beta)=p_1g_1(\theta,\alpha,\beta)+p_2g_2(\theta,\alpha+1,\beta)+(1-p_1-p_2)g_3(\theta,\alpha+2,\beta),$$

Where

$$p_1=\frac{\theta^2}{\theta^2+\alpha\theta+\alpha(\alpha+1)},\ p_2=\frac{\alpha\theta}{\theta^2+\alpha\theta+\alpha(\alpha+1)}$$

$$g_1(\theta,\alpha,\beta)=\frac{\beta\theta^{\alpha}}{\Gamma(\alpha)}x^{\beta\alpha-1}e^{-\theta x^{\beta}};x>0,\theta>0,\alpha>0,\beta>0$$

$$g_2(\theta,\alpha+1,\beta)=\frac{\beta\theta^{\alpha+1}}{\Gamma(\alpha+1)}x^{\beta(\alpha+1)-1}e^{-\theta x^{\beta}};x>0,\theta>0,\alpha>0,\beta>0$$

$$g_3(\theta,\alpha+2,\beta)=\frac{\beta\theta^{\alpha+2}}{\Gamma(\alpha+2)}x^{\beta(\alpha+2)-1}e^{-\theta x^{\beta}};x>0,\theta>0,\alpha>0,\beta>0$$

Graphs of density function of PWSD for varying values of parameters θ,αandβ have been drawn and presented in figure 1. From the figure 1, It was observed that pdf of PWSD is increasing with

increased value of theta at fixed value of alpha and beta, whereas its value is decreasing with increased value of alpha at fixed valued of theta and beta.

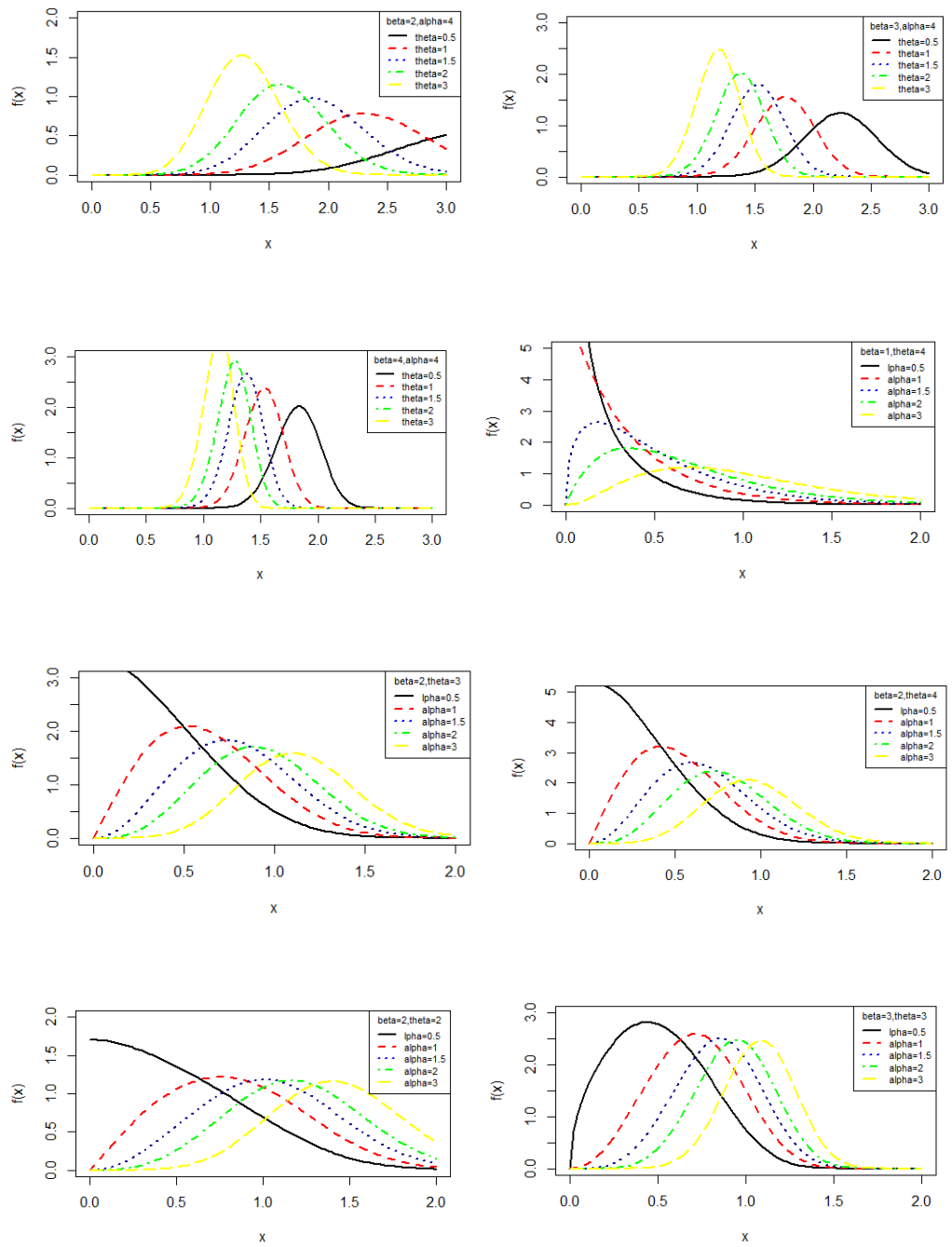


Figure 1: Graphs of the pdf of PWSD for varying values of parametersθ, αandβ

3. Moments

The rth moment about origin of PWSD (2.1) can be obtained as

$$\mu_r' = \frac{\beta\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\int_0^\infty x^{\beta\alpha+r-1}\left(1+x^\beta+x^{2\beta}\right)e^{-\theta x^\beta}dx$$

$$= \frac{\beta\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\int_0^\infty e^{-\theta x^\beta}x^{\beta\alpha+r-1}\,dx+\int_0^\infty e^{-\theta x^\beta}x^{\beta\alpha+\beta+r-1}\,dx+\int_0^\infty e^{-\theta x^\beta}x^{\beta\alpha+2\beta+r-1}\,dx\right]$$

Assuming $u = \theta x^{\beta}$, we have $du = \theta\beta x^{\beta-1}dx$, $x = \left(\frac{u}{\theta}\right)^{\frac{1}{\beta}}$ and $dx = \frac{du}{\theta\beta\left(\frac{u}{\theta}\right)^{\frac{\beta-1}{\beta}}}$

Thus, we have

$$\mu_r{}' = \frac{\theta^{\alpha+1}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\begin{matrix}\frac{1}{\theta^{\alpha+\frac{r}{\beta}-1}}\int_0^{\infty} e^{-u}u^{\alpha+\frac{r}{\beta}-1}\,du + \frac{1}{\theta^{\alpha+\frac{r}{\beta}}}\int_0^{\infty} e^{-u}u^{\alpha+\frac{r}{\beta}+1-1}\,du \\ + \frac{1}{\theta^{\alpha+\frac{r}{\beta}+1}}\int_0^{\infty} e^{-u}u^{\alpha+\frac{r}{\beta}+2-1}\,du\end{matrix}\right]$$

$$= \frac{\theta^{\alpha+1}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\frac{\Gamma\left(\alpha+\frac{r}{\beta}\right)}{\theta^{\alpha+\frac{r}{\beta}-1}} + \frac{\Gamma\left(\alpha+\frac{r}{\beta}+1\right)}{\theta^{\alpha+\frac{r}{\beta}}} + \frac{\Gamma\left(\alpha+\frac{r}{\beta}+2\right)}{\theta^{\alpha+\frac{r}{\beta}+1}}\right]$$

$$= \frac{\theta^{\alpha+1}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\frac{\theta^2\Gamma\left(\alpha+\frac{r}{\beta}\right)+\theta\Gamma\left(\alpha+\frac{r}{\beta}+1\right)+\Gamma\left(\alpha+\frac{r}{\beta}+2\right)}{\theta^{\alpha+\frac{r}{\beta}+1}}\right]$$

$$= \frac{\theta^2+\theta\left(\alpha+\frac{r}{\beta}\right)+\left(\alpha+\frac{r}{\beta}\right)\left(\alpha+\frac{r}{\beta}+1\right)}{\theta^{\frac{r}{\beta}}\{\theta+\alpha\theta+\alpha(\alpha+1)\}}\frac{\Gamma\left(\alpha+\frac{r}{\beta}\right)}{\Gamma(\alpha)}$$

$$= \frac{\theta^2\beta^2+\theta\beta(\alpha\beta+r)+(\alpha\beta+r)(\alpha\beta+\beta+r)}{\beta\theta^{\frac{r}{\beta}}\{\theta+\alpha\theta+\alpha(\alpha+1)\}}\frac{\Gamma\left(\alpha+\frac{r}{\beta}\right)}{\Gamma(\alpha)}; r = 1,2,3,\ldots \qquad (3.1)$$

Substituting r = 1,2,3 and 4 in (3.1), the first four moments about the origin of PWSD can be obtained. Again using the relationship between moments about origin and central moments, central moments can be obtained. Since the expressions for central moments are complicated, central moments are not being given.

4. Hazard Rate Function

The survival function $S(x;\theta,\alpha,\beta)$of PWSD can be obtained as

$$S(x;\theta,\alpha,\beta) = P(X > x) = \int_x^{\infty} f_5(t;\theta,\alpha,\beta)\,dt$$

$$= \frac{\beta\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\int_x^{\infty} t^{\beta\alpha-1}\left(1+t^{\beta}+t^{2\beta}\right)e^{-\theta t^{\beta}}dt$$

$$= \frac{\beta\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\int_x^{\infty} t^{\beta\alpha-1}\,e^{-\theta t^{\beta}}dt + \int_x^{\infty} t^{\beta\alpha+\beta-1}\,e^{-\theta t^{\beta}}dt + \int_x^{\infty} t^{\beta\alpha+2\beta-1}\,e^{-\theta t^{\beta}}dt\right]$$

Taking $u = t^{\beta}$and $t = (u)^{\frac{1}{\beta}}$gives $dt = \frac{du}{\beta u^{\frac{\beta-1}{\beta}}}$ and thus we have

$$S(x;\theta,\alpha,\beta) = \frac{\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\begin{matrix}\int_{x^{\beta}}^{\infty} e^{-\theta u}u^{\frac{\beta\alpha-1}{\beta}}\frac{du}{u^{\frac{\beta-1}{\beta}}} + \int_{x^{\beta}}^{\infty} e^{-\theta u}u^{\frac{\beta\alpha+\beta-1}{\beta}}\frac{du}{u^{\frac{\beta-1}{\beta}}} \\ + \int_{x^{\beta}}^{\infty} e^{-\theta u}u^{\frac{\beta\alpha+2\beta-1}{\beta}}\frac{du}{u^{\frac{\beta-1}{\beta}}}\end{matrix}\right]$$

$$= \frac{\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\int_{x^{\beta}}^{\infty} e^{-\theta u}u^{\alpha-1}du + \int_{x^{\beta}}^{\infty} e^{-\theta u}u^{\alpha}du + \int_{x^{\beta}}^{\infty} e^{-\theta u}u^{\alpha+1}du\right]$$

$$= \frac{\theta^{\alpha+2}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}\left[\begin{matrix}\frac{\Gamma\left(\alpha,\theta x^{\beta}\right)}{\theta^{\alpha}} + \frac{e^{-\theta x^{\beta}}\left(\theta x^{\beta}\right)^{\alpha} + \alpha\Gamma\left(\alpha,\theta x^{\beta}\right)}{\theta^{\alpha+1}} \\ + \frac{e^{-\theta x^{\beta}}\left(\theta x^{\beta}\right)^{\alpha}\left(\theta x^{\beta}+\alpha+1\right) + \alpha(\alpha+1)\Gamma\left(\alpha,\theta x^{\beta}\right)}{\theta^{\alpha+2}}\end{matrix}\right]$$

$$= \frac{\theta^2\Gamma\left(\alpha,\theta x^{\beta}\right)+\theta\left\{e^{-\theta x^{\beta}}\left(\theta x^{\beta}\right)^{\alpha}+\alpha\Gamma\left(\alpha,\theta x^{\beta}\right)\right\}+e^{-\theta x^{\beta}}\left(\theta x^{\beta}\right)^{\alpha}\left(\theta x^{\beta}+\alpha+1\right)+\alpha(\alpha+1)\Gamma\left(\alpha,\theta x^{\beta}\right)}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}$$

$$= \frac{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma\left(\alpha,\theta x^{\beta}\right)+\theta e^{-\theta x^{\beta}}\left(\theta x^{\beta}\right)^{\alpha}+e^{-\theta x^{\beta}}\left(\theta x^{\beta}\right)^{\alpha}\left(\theta x^{\beta}+\alpha+1\right)}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)}$$

$$= \frac{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma\left(\alpha,\theta x^{\beta}\right)+e^{-\theta x^{\beta}}\left(\theta x^{\beta}+\theta+\alpha+1\right)\left(\theta x^{\beta}\right)^{\alpha}}{\{\theta+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha)},$$

where $\Gamma\left(\alpha,\theta x^{\beta}\right)$ is the upper incomplete gamma function defined as

$$\Gamma\left(\alpha,\theta x^{\beta}\right) = \int_{\theta x^{\beta}}^{\infty} y^{\alpha-1}e^{-y}dy; \alpha > 0, \theta x^{\beta} > 0.$$

It can be easily verified that at $(\beta = 1)$,$(\alpha = 1)$and $(\alpha = \beta = 1)$, the survival function of PWSD reduce to the survival function of WSD, PSD and Sujatha distribution, respectively.

Thus the cdf of PWSD can be obtained using $F_5(x;\theta,\alpha,\beta) = 1 - S(x;\theta,\alpha,\beta)$.

The hazard rate function, $h(x;\theta,\alpha,\beta)$, of PWSD can be given by

$$h(x;\theta,\alpha,\beta) = \frac{f_5(x;\theta,\alpha,\beta)}{S(x;\theta,\alpha,\beta)}$$

$$= \frac{\beta\theta^{\alpha+2}x^{\beta\alpha-1}(1+x^{\beta}+x^{2\beta})e^{-\theta x^{\beta}}}{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha,\theta x^{\beta})+e^{-\theta x^{\beta}}(\theta x^{\beta}+\theta+\alpha+1)(\theta x^{\beta})^{\alpha}}.$$

Graphs of $h(x;\theta,\alpha,\beta)$for varying values of parameters θ,αandβ are shown in figure 2.

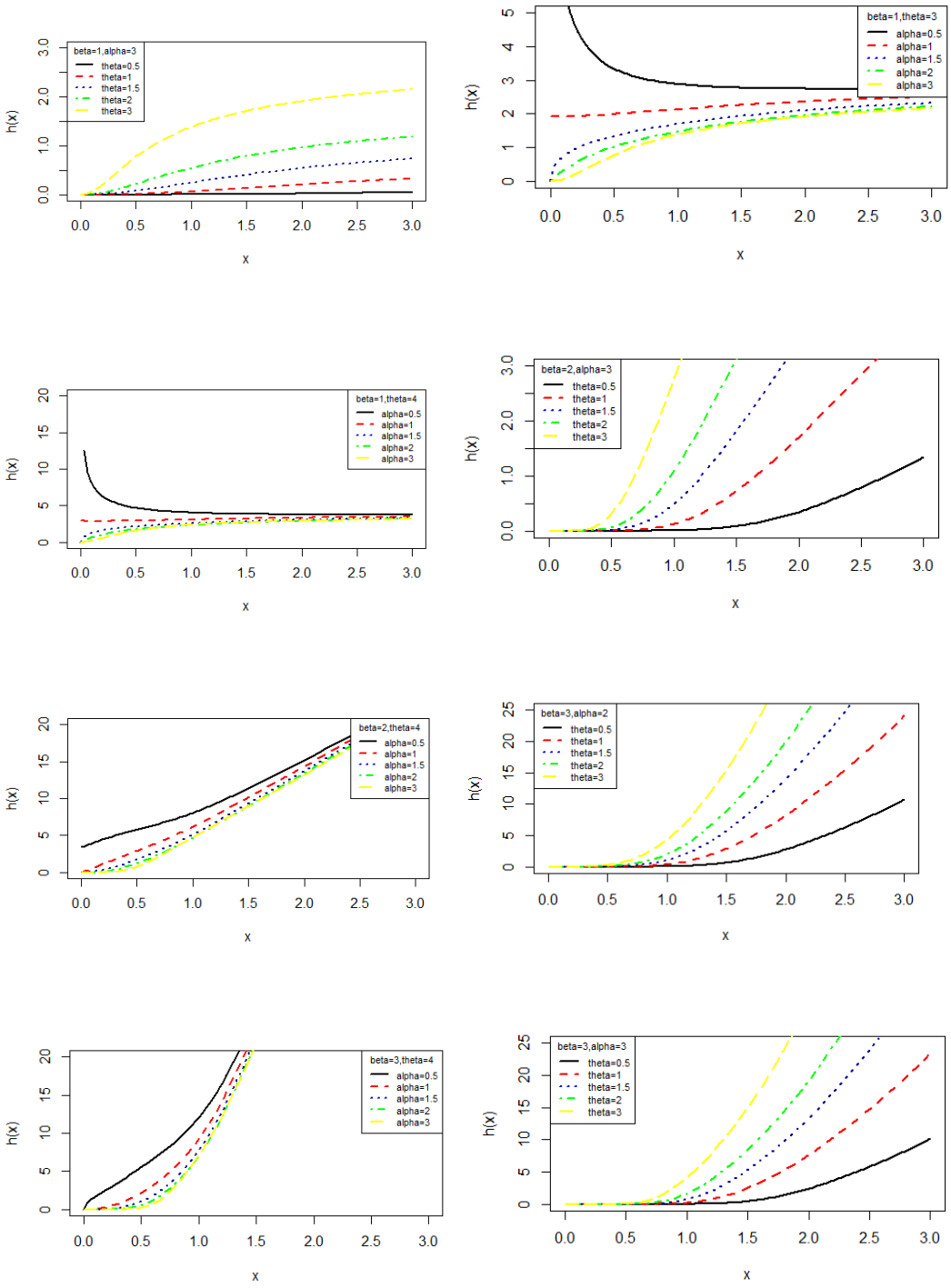


Figure 2: Graphs of hazard rate function of PWSD for varying values of parametersθ,αandβ.

5. Mean Residual life function

The mean residual life function of PWSD can be obtained as

$$m(x;\theta,\alpha,\beta)=\frac{1}{S(x;\theta,\alpha,\beta)}\int_x^\infty tf(t;\theta,\alpha,\beta)\,dt-x$$

$$=\frac{\beta\theta^{\alpha+2}}{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha,\theta x^\beta)+e^{-\theta x^\beta}(\theta x^\beta+\theta+\alpha+1)(\theta x^\beta)^\alpha}$$
$$\times\int_x^\infty t^{\beta\alpha}\left(1+t^\beta+t^{2\beta}\right)e^{-\theta t^\beta}dt-x$$

$$=\frac{\beta\theta^{\alpha+2}}{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha,\theta x^\beta)+e^{-\theta x^\beta}(\theta x^\beta+\theta+\alpha+1)(\theta x^\beta)^\alpha}$$
$$\times\left[\int_x^\infty e^{-\theta t^\beta}t^{\beta\alpha}dt+\int_x^\infty e^{-\theta t^\beta}t^{\beta\alpha+\beta}dt+\int_x^\infty e^{-\theta t^\beta}t^{\beta\alpha+2\beta}dt\right]-x$$

Taking $u=t^\beta$ and $t=(u)^{\frac{1}{\beta}}$ gives $dt=\frac{du}{\beta u^{\frac{\beta-1}{\beta}}}$ and thus we have

$$m(x;\theta,\alpha,\beta)=\frac{\theta^{\alpha+2}}{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha,\theta x^\beta)+e^{-\theta x^\beta}(\theta x^\beta+\theta+\alpha+1)(\theta x^\beta)^\alpha}$$
$$\times\left[\int_{x^\beta}^\infty e^{-\theta u}u^{\alpha+\frac{1}{\beta}-1}du+\int_{x^\beta}^\infty e^{-\theta u}u^{\alpha+\frac{1}{\beta}+1-1}du+\int_{x^\beta}^\infty e^{-\theta u}u^{\alpha+\frac{1}{\beta}+2-1}du\right]-x$$

$$=\frac{\theta^{\alpha+2}}{\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma(\alpha,\theta x^\beta)+e^{-\theta x^\beta}(\theta x^\beta+\theta+\alpha+1)(\theta x^\beta)^\alpha}$$
$$\times\left[\frac{\Gamma\left(\alpha+\frac{1}{\beta},\theta x^\beta\right)}{\theta^{\alpha+\frac{1}{\beta}}}+\frac{\Gamma\left(\alpha+\frac{1}{\beta}+1,\theta x^\beta\right)}{\theta^{\alpha+\frac{1}{\beta}+1}}+\frac{\Gamma\left(\alpha+\frac{1}{\beta}+2,\theta x^\beta\right)}{\theta^{\alpha+\frac{1}{\beta}+2}}\right]-x$$

$$=\frac{\theta^2\Gamma\left(\alpha+\frac{1}{\beta},\theta x^\beta\right)+\theta\Gamma\left(\alpha+\frac{1}{\beta}+1,\theta x^\beta\right)+\Gamma\left(\alpha+\frac{1}{\beta}+2,\theta x^\beta\right)}{\theta^{\frac{1}{\beta}}\{\theta^2+\alpha\theta+\alpha(\alpha+1)\}\Gamma\left(\alpha,\theta x^\beta\right)+e^{-\theta x^\beta}\left(\theta x^\beta+\theta+\alpha+1\right)\left(\theta x^\beta\right)^\alpha}-x.$$

Graphs of mean residual function of PWSD for varying values of parameters θ,α andβ have been drawn and presented in figure 3. From the figure 3, It was observed that mean residual of PWSD is decreasing with increased value of theta at fixed value of alpha and beta.

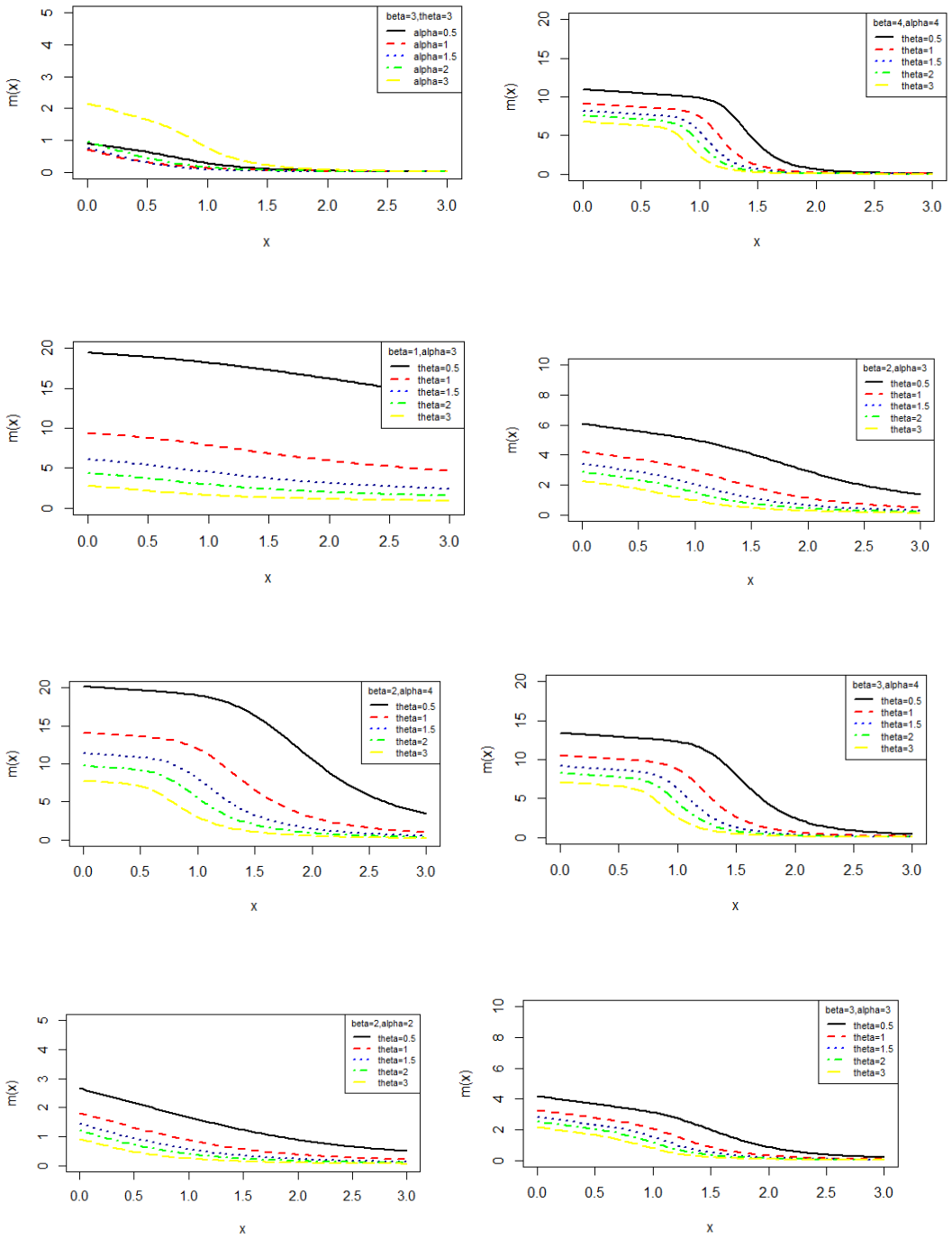


Figure 3: *Graphs of mean residual life function of PWSD for varying values of parameters* θ, α and β.

6. Maximum Likelihood Estimation of parameters

Suppose $(x_1, x_2, x_3, \ldots, x_n)$ be a random sample of size n from PWSD (2.1). The natural log likelihood function is thus obtained as

$$ln\, L = \sum_{i=1}^{n} ln\, f_8\,(x_i; \theta, \alpha, \beta)$$

$$= n[ln\, \beta + (\alpha + 2)\, ln\, \theta - ln(\theta^2 + \alpha\theta + \alpha^2 + \alpha) - ln\, \Gamma\,(\alpha)] + (\beta\alpha - 1) \sum_{i=1}^{n} ln(x_i)$$
$$+ \sum_{i=1}^{n} ln\left(1 + {x_i}^{\beta} + {x_i}^{2\beta}\right) - \theta \sum_{i=1}^{n} {x_i}^{\beta}.$$

The maximum likelihood estimates (MLEs) $(\hat{\theta}, \hat{\alpha}, \hat{\beta})$ of parameters (θ, α, β) of PWSD are the solution of the following nonlinear log likelihood equations

$$\frac{\partial \ln L}{\partial \theta} = \frac{n(\alpha+2)}{\theta} - \frac{n(2\theta+\alpha)}{\theta^2+\alpha\theta+\alpha^2+\alpha} - \sum_{i=1}^{n} {x_i}^{\beta} = 0$$

$$\frac{\partial \ln L}{\partial \alpha} = n \ln \theta - \frac{n(\theta+2\alpha+1)}{\theta^2+\alpha\theta+\alpha^2+\alpha} - n\psi(\alpha) + \beta \sum_{i=1}^{n} ln(x_i) = 0$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{n}{\beta} + \alpha \sum_{i=1}^{n} ln\, x_i + \sum_{i=1}^{n} \frac{{x_i}^{\beta}\left(1+2{x_i}^{\beta}\right) ln(x_i)}{1+{x_i}^{\beta}+{x_i}^{2\beta}} - \theta \sum_{i=1}^{n} {x_i}^{\beta}\, ln(x_i) = 0$$

where $\psi(\alpha) = \frac{d}{d\alpha} ln\, \Gamma(\alpha)$ is the digamma function.

These three natural log likelihood equations do not seem to be solved directly, because they cannot be expressed in closed forms. The (MLE's) $(\hat{\theta}, \hat{\alpha}, \hat{\beta})$ of parameters (θ, α, β) can be computed directly by solving the natural log likelihood equation using Newton-Raphson iteration method available in R-software till sufficiently close values of $\hat{\theta}, \hat{\alpha}$ and $\hat{\beta}$ are obtained. The initial values of parameters θ, α and β are taken as $\theta = 0.5, \alpha = 0.5$ and $\hat{\beta} = 1.5$.

7. A Simulation Study

In this section, a simulation study has been carried to check the performance of maximum likelihood estimates by taking sample sizes (n = 40,80,120,160) for values of $\theta = 0.5,1.0,1.5,2.0$ and $\alpha = 0.5$ $\beta = 1.0$. Similarly $\alpha = 0.5,2.0,4.0,5.0 \& \theta = 1.5, \beta = 1.0$, and $\beta = 0.1,0.5,1.5,2.0$ & $\theta = 1.5, \alpha = 2.5$. Acceptance and rejection method is used to generate random number for data simulation using R-software. The process were repeated 1,000 times for the calculation of Average Bias error (ABE) and MSE (Mean square error) of parameters θ, α and β are presented in tables1, 2 &3 respectively.

Table1. *ABE and MSE of MLE parameters θ, α and β for fixed value of $\alpha = 0.5$, $\beta = 1.0$*

n	θ	$ABE(\hat{\theta})$	$MSE(\hat{\theta})$	$ABE(\hat{\alpha})$	$MSE(\hat{\alpha})$	$ABE(\hat{\beta})$	$MSE(\hat{\beta})$
40	0.5	0.0183	0.0134	-0.0909	0.0005	0.0249	0.0495
	1.0	0.0058	0.0014	-0.0409	0.0669	0.0198	0.0316
	1.5	-0.0066	0.0017	-0.0909	0.3305	0.0074	0.0044
	2.0	-0.0192	0.0147	-0.1159	0.5374	0.0248	0.0495
80	0.5	0.0131	0.0138	-0.0465	0.0006	0.0248	0.0495
	1.0	0.0068	0.0038	-0.0215	0.0369	0.0198	0.0316
	1.5	0.0006	0.0003	-0.0464	0.1729	0.0073	0.0043
	2.0	-0.0056	0.0025	-0.0589	0.2784	0.0011	0.0001
120	0.5	0.0132	0.0210	-0.0295	0.0002	0.0112	0.0151
	1.0	0.0091	0.0098	-0.0129	0.0199	0.0078	0.0074
	1.5	0.0049	0.0028	-0.0296	0.10491	-0.0004	0.0002
	2.0	0.0007	0.0000	-0.0379	0.1723	-0.0046	0.0025
160	0.5	0.0101	0.0164	-0.0219	0.0000	0.0080	0.0105
	1.0	0.0070	0.0078	-0.0095	0.0144	0.0056	0.0049
	1.5	0.0038	0.0024	-0.0219	0.0773	-0.0006	0.0007
	2.0	0.0007	0.0093	-0.0282	0.0127	-0.0037	0.0023

Table2.*ABE and MSE of MLE parameters $\theta, \alpha$and β for fixed value of$\theta = 1.5, \beta = 1.0$*

N	α	$ABE(\hat{\theta})$	$MSE(\hat{\theta})$	$ABE(\hat{\alpha})$	$MSE(\hat{\alpha})$	$ABE(\hat{\beta})$	$MSE(\hat{\beta})$
40	0.5	0.0157	0.0099	-0.0903	0.0003	0.0594	0.1412
	2.0	0.0033	0.0004	-0.04033	0.0650	0.0494	0.0977
	4.0	-0.0092	0.0034	-0.0903	0.3263	0.0244	0.0238
	5.0	-0.0217	0.0188	-0.1153	0.5320	0.0119	0.0056
80	0.5	0.0078	0.0049	-0.0451	0.0002	0.0297	0.0706
	2.0	0.0016	0.0002	-0.0201	0.0325	0.0247	0.0488
	4.0	-0.0046	0.0016	-0.0451	0.1632	0 .0122	0.0119
	5.0	-0.0108	0.0094	-0.0576	0.2660	0.0059	0.0028
120	0.5	0.0071	0.0060	0.0293	0.0449	0.0185	0.0413
	2.5	0.0029	0.0010	0.0168	0.0341	0.0152	0.0277
	4.0	0.0293	0.1034	0.0012	0.0002	0.0068	0.0056
	5.0	0.0377	0.1704	0.0054	0.0035	0.0027	0.0008
160	0.5	0.0056	0.0052	-0.0217	0.0000	0.0138	0.0305
	2.5	0.0025	0.0010	-0.0124	0.0245	0.0113	0.0204
	4.0	-0.0005	0.0005	-0.0217	0.0757	0.0051	0.0041
	5.0	-0.0037	0.0022	-0.0280	0.1256	0.0019	0.0005

Table 3. *ABE and MSE of MLE parameters $\theta, \alpha$and β for fixed value of $\theta = 1.5, \alpha = 2.5$*

N	β	$ABE(\hat{\theta})$	$MSE(\hat{\theta})$	$ABE(\hat{\alpha})$	$MSE(\hat{\alpha})$	$ABE(\hat{\beta})$	$MSE(\hat{\beta})$
40	0.1	0.79609	25.3479	0.7246	26.3839	0.0121	0.0059
40	0.5	0.7835519	24.5581	0.7621	23.2352	0.0021	0.0001
40	1.5	0.7710519	23.7808	0.7246	21.0050	-0.0228	0.0208
40	2.0	0.7585519	23.0160	0.6996	19.5807	-0.0353	0.0499
80	0.1	1.24654	93.2320	1.2157	97.3935	0.0051	0.0015
80	0.5	1.2382	91.9897	1.2407	92.3639	-0.0015	0.0001
80	1.5	1.2298	90.7557	1.2157	88.6793	-0.0182	0.0198
80	2.0	1.2215	89.5300	1.1990	86.2645	-0.0265	0.0422
120	0.1	0.3259	12.7464	0.3011	13.0932	0.0025	7.7334
120	0.5	0.3217	12.4226	0.3136	11.8053	-0.0007	7.5789
120	1.5	0.3175	12.1029	0.3011	10.8831	-0.0091	9.9985
120	2.0	0.3134	11.7874	0.2928	10.2891	-0.0132	2.1209
160	0.1	0.1996	6.3794	0.1856	6.8906	0.0018	5.7328
160	0.5	0.1965	6.1813	0.1950	6.0855	-0.0006	5.8974
160	1.5	0.1934	5.9863	0.1856	5.5145	-0.0068	7.5232
160	2.0	0.1903	5.7944	0.1794	05.1495	-0.0099	1.5942

8. Applications

The applications and goodness of fit of the power weighted Sujatha distribution (PWSD) has been discussed for one real dataset reported by Efron [14] relating to the survival times of a group of patients suffering from Head and Neck cancer disease and treated using a combination of radiotherapy and chemotherapy (RT+CT).The goodness of fit is compared with exponential distribution, Lindley distribution, Weibull distribution introduced by Weibull [15], Gamma distribution, Generalized exponential distribution (GED) introduced by Gupta and Kundu [16] , Power Lindley distribution (PLD), Weighted Lindley distribution (WLD), Power Sujatha distribution (PSD), weighted Sujatha distribution (WSD), and Generalized Lindley distribution (GLD). The dataset is as follow:

12.20	23.56	23.74	25.87	31.98	37	41.35	47.38	55.46	58.36
63.47	68.46	78.26	74.47	81.43	84	92	94	110	112
119	127	130	133	140	146	155	159	173	179
194	195	209	249	281	319	339	432	469	519
633	725	817	1776						

In order to compare the goodness of fit of the considered distributions for the dataset, values of $-2\,ln\,L$, AIC (Akaike information criterion), K-S Statistic (Kolmogorov-Smirnov Statistic) and p-value for two datasets have been computed The formulae for AIC and K-S Statistics are as follows: $AIC = -2\,ln\,L + 2k$, and $K - S = \underset{x}{\text{Sup}}|F_n(x) - F_0(x)|$, where$k$being the number of parameters involved in the respective distributions, n is the sample size and $F_n(x)$is the empirical distribution function. The best distribution corresponds to the lower values of$-2\,ln\,L$, AIC and K-S statistic. Note that the estimates of parameters of the considered distributions are based on maximum likelihood estimates. The initial values of the parameters for ML estimates of PWSD have been selected as$\theta = 1.5$,$\alpha = 0.5$ and$\beta = 1.5$, as the log-likelihood function is non-linear.

The ML estimate of parameters of the considered distributions for dataset is given in table 4. The goodness of fit by K-S statistics for dataset with considered distributions are presented in tables 5. From the goodness of fit given in table 5, it is crystal clear that PWSD gives much closure fit to dataset relating to survival times of patients suffering from head and neck cancer and hence it can be considered as an important distribution for modeling data relating to survival times of patients suffering from head and neck cancer. The variance-covariance matrix of the parameters (θ, α, β) of PWSD for dataset is given in table 6. The survival plots for the dataset, fitted plot of cdf for given dataset, and the fitted plot of given dataset are presented in figures 4, 5 and 6 respectively.

Table 4: *Summary of the ML estimates of parameters for dataset*

Model	ML Estimates		
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$
PWSD	26.9742	50.3682	0.1325
GLD	0.0047	0.0524	5.0750
PSD	0.1539	……	0.5690
WSD	0.0089	0.01637	…….
WLD	0.00531	0.21191	……
PLD	0.05301	0.68893	…….
GED	0.00482	1.09367	……
Gamma	0.00489	1.08501	……
Weibull	0.00710	0.92327	…….
Lindley	0.00892	…….	……..
Exponential	0.00447	……	…….

Table 5: *Summary of Goodness of fit by K-S Statistic for dataset*

Model	$-2\ln L$	AIC	K-S	p-value
PWSD	555.41	561.41	0.104	0.572
GLD	564.09	570.09	0.150	0.248
PSD	559.45	563.45	0.135	0.512
WSD	579.96	583.96	0.350	0.000
WLD	565.91	569.91	0.161	0.181
PLD	560.78	564.78	0.118	0.529
GED	563.93	567.93	0.145	0.280
Gamma	564.10	568.10	0.149	0.249
Weibull	563.71	567.71	0.298	0.005
Lindley	579.16	581.16	0.219	0.025
Exponential	564.01	566.01	0.145	0.282

Table 6: *Variance-covariance matrix of the parameters θ, α and β of PWSD for dataset*

$$\begin{array}{c} \\ \hat{\theta} \\ \hat{\alpha} \\ \hat{\beta} \end{array} \begin{array}{c} \begin{array}{ccc} \hat{\theta} & \hat{\alpha} & \hat{\beta} \end{array} \\ \begin{bmatrix} 983.70956 & 1475.4914 & -2.33334 \\ 1475.4914 & 2219.87726 & -3.47289 \\ -2.33334 & -3.47289 & 0.005661 \end{bmatrix} \end{array}$$

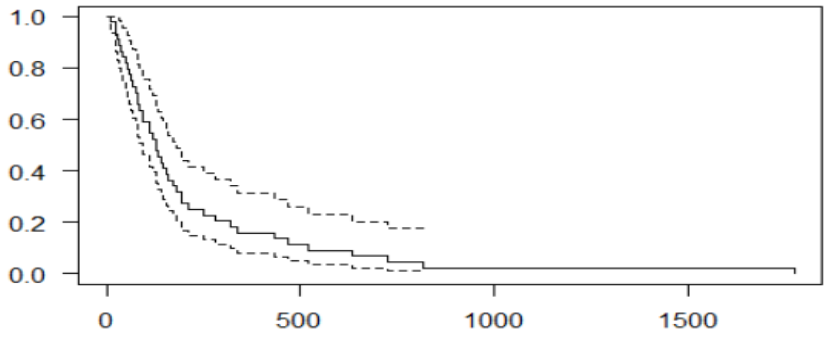


Figure 4. *Survival plots for the given dataset.*

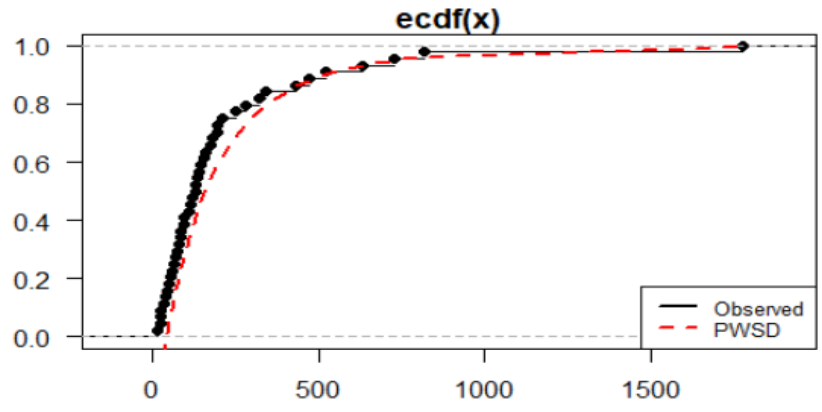


Figure 5. *Fitted plot of CDF on given dataset*

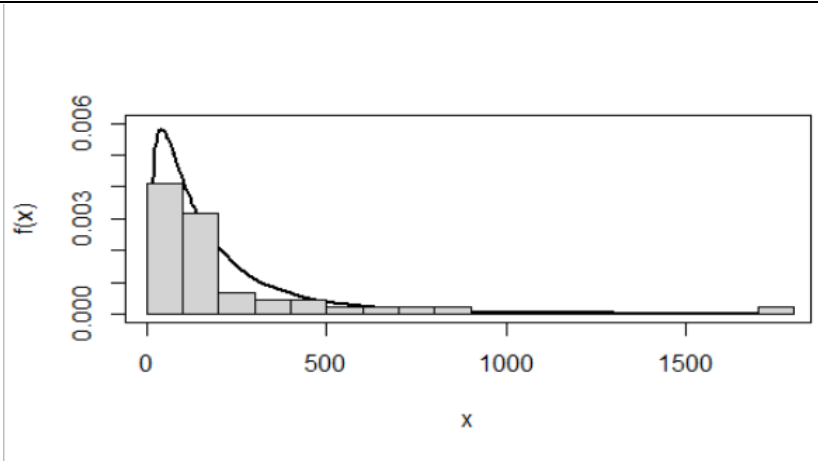


Figure 6. *Fitted plot of pdf of PWSD on dataset*

9. Concluding Remarks

In this paper a three-parameter power weighted Sujatha distribution (PWSD) which includes weighted Sujatha distribution (WSD), power Sujatha distribution (PSD) and Sujatha distribution as particular cases, has been proposed. Its statistical properties including behavior it density function, moments, hazard rate function, mean residual life function have been discussed. Estimation of parameters has been discussed using the method of maximum likelihood. A simulation study has been presented to know the performance of maximum likelihood estimates of parameters. Application of the proposed distribution have been explained with a real lifetime data which is related to patients suffering from head and neck cancer and its goodness of fit has been compared with other lifetime distributions. The proposed distribution shows better fit over the other one parameter, two-parameter and three-parameter distributions.

References

[1] Shanker, R (2016a): Sujatha distribution and its applications, *Statistics in Transition New series*, 17(3), 1-20

[2] Shanker, R. and Shukla, K.K. (2018): A two-parameter Weighted Sujatha distribution and Its Application to Model Lifetime data, International *Journal of Applied Mathematics and statistics*, 57 (3), 106 – 121

[3] Shanker, R. and Shukla, K. K. (2019): A two-parameter Power Sujatha distribution with properties and Application, *International Journal of Mathematics and Statistics*, 20(3), 11 – 22

[4] Ghitany, M.E., Alqallaf ,F., Al-Mutairi, D.K., Husain, H.A. (2011): A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in simulation*, 81, 1190-1201.

[5] Shanker, R., Shukla, K.K. and Hagos, F. (2016): On Weighted Lindley distribution and Its Applications to Model Lifetime Data, Jacobs Journal of Biostatistics,1(1), 1- 9

[6] Ghitany, M.E., Al-Mutairi, D.K., Balakrishnan, N., and Al-Enezi, L.J. (2013): Power Lindley distribution and Associated Inference, Computational Statistics and Data Analysis 64, 20 –33.

[7] Lindley, D.V. (1958): Fiducial distributions and Bayes' Theorem, *Journal of the Royal Statistical Society, Series* B, 20, 102 – 107.

[8] Ghitany, M.E. , Atieh, B. and Nadarajah, S. (2008): Lindley distribution and its Application, *Mathematics Computing and Simulation*, 78, 493 – 506.

[9] Shanker, R., Hagos, F, and Sujatha, S. (2015): On modeling of Lifetimes data using exponential and Lindley distributions, *Biometrics & Biostatistics International Journal*, 2 (5), 1-9.

[10] Zakerzadeh, H. and Dolati, A. (2009): Generalized Lindley distribution, *Journal of Mathematical extension*, 3 (2), 13 – 25.

[11] Shanker, R. (2016b): On Generalized Lindley Distribution and its Applications to model lifetime data from B*iomedical science and engineering, Insights in Biomedicine*, 1(2), 1 – 6

[12] Stacy, E.W. (1962): A generalization of the gamma distribution, *Annals of Mathematical Statistical*, 33, 1187 – 1192

[13] Shanker, R. and Shukla, K.K. (2016): On modeling of Lifetime data using three-parameter generalized Lindley and generalized gamma distributions, *Biometrics & Biostatistics International journal*, 4(7), 283 -288

[14] Weibull, W. (1951): A statistical distribution of wide applicability, *Journal of Applied Mathematics*, 18, 293 – 297.

[15] Efron, B. (1988): Logistic regression, survival analysis and the Kaplan-Meier curve, Journal of the American Statistical Association, 83, 414-425

[16] Gupta, R.D. and Kundu, D. (1999): Generalized Exponential Distribution, *Austalian & New Zealand Journal of Statistics*, 41(2), 173 – 188.

A STUDY ON STATISTICAL PROPERTIES OF A NEW CLASS OF Q-EXPONENTIAL-WEIBULL DISTRIBUTION WITH APPLICATION TO REAL-LIFE FAILURE TIME DATA

N. Sundaram[1] and G. Jayakodi[2]

•

[1,2] Department of Statistics,
Presidency College,
Chennai, Tamil Nadu, India
ncsundar77@gmail.com[1], jayakodimcc@gmail.com[2]

Abstract

This article introduces a new four-parameter probability distribution called the q-Exponential-Weibull distribution based on the q-Exponential-G family of distribution. The proposed new distribution has to decrease and increase failure rates which are more common in reliability scenarios and can be used instead of Weibull and the exponential distribution. It also includes some sub-models like q-Exponential-Exponential, q-Exponential-Rayleigh, Exponential-Weibull, Exponential-Exponential and Exponential-Rayleigh lifetime distributions. Various Mathematical and statistical Properties are investigated, which include Limiting behavior, Moments and Moment Generating functions, Quantile function and Order Statistics. The Maximum Likelihood estimator is used for estimating the model parameters. This new distribution is compared with other lifetime distributions using different kinds of real-life failure time data.

Keywords: q-Exponential-Weibull, Quantile, Reliability Measures, Maximum Likelihood Estimation, failure time data.

1. Introduction

Lifetime distributions are very useful statistical tool for analyzing the various characteristics of lifetime data. The developments and applications of lifetime distribution are essential in numerous fields. Hence, the major aspects of generating new families of probability distributions are they offer greater flexibility and a better fit at the expense of one or more extra parameters.

The non-extensive statistical mechanism plays a vital growth in the past few years. This new formulation is not based on the usual statistical mechanism, provided that will give a better description of the complex system developed by [26]. In the recent decade's probability distribution, which emerge from the non-extensive statistical mechanism called q-type distribution attracted several statisticians to develop new distribution [11], [20] and [23]. Studying this type of distribution is quite interesting because of its complex system and power-law behavior. The application of this type of distribution has been found in many research areas like Physics, Biology, Mathematics, Chemistry, Economics, Medicine etc.

The q-Exponential distribution emerged from maximizing the non-extensive statistical mechanism under appropriate constraints [26]. This theory is a generalization of the classical Boltzmann-Gibbs (BG) statistical mechanism. So, the q-Exponential distribution has found varieties of applications in the research field including in the field of complex systems. This article introduces a new four-parameter probability distribution called the q-Exponential Weibull distribution.

The well-known q-distributions are q-Exponential distributions discussed by Malacarne et al. [15], q-Gamma distribution due to Duarte et al. [7], q-Weibull distribution due to Picoli et al. [21], q-Gaussian distribution due to Adrian et al. [1]. Picoli J.R. et al. [22] discussed q-distribution in complex systems. Ana Claudia souza [3] studied the reliability data analysis of systems in the wear-out phase using q-Exponential likelihood. Fode Zhang et al. [9] discovered the information geometry on the curved q-Exponential family with application to survival analysis. Shalizi [25] express the Maximum Likelihood Estimation for q-Exponential distribution. The geometry of q-exponential distribution with dependent competing risk and accelerated life testing is given by Fode Zhang et al. [10]. Keith Briggs [12] demonstrates the modelling train delay with the q-Exponential distribution. The reliability of stress strength and its estimation of exponentiated q-Exponential distribution is given by Mohammed et al [18]. Modelling censored survival data with q-Exponential distribution discussed by Sundaram [19].

In reliability and survival analysis most commonly, used distributions are Exponential and Weibull distributions [16], q-Exponential is an alternative one. The q-Exponential distribution is a higher version of an Exponential distribution. It has two parameters: q and α, where q is the shape parameter (entropy index/control parameter) and α is the scale parameter. As compared to the Exponential distribution that has just one parameter (α), the q-Exponential distribution has more flexibility regarding the decay of the pdf [3]. Indeed, the Exponential probability distribution is a special case of the q-Exponential when $q \to 1$. Another feature of this distribution is that it does not have the limitation of a constant hazard rate as the Exponential one, thus allowing the modelling of either system improvement ($1 < q < 2$) or degradation ($q < 1$). The pdf of q-Exponential distribution [26], is given by

$$f_q(x) = (2-q)\,\alpha\,[1-(1-q)\,\alpha x]^{1/(1-q)} \qquad \text{for } x, \alpha > 0, q<2 \tag{1}$$

This can also be rewritten as

$$f_q(x) = (2-q)\,\alpha\, e_q(-\alpha x)$$

$$\text{Where} \quad e_q(x) = [1 + (1-q)\,x]^{\frac{1}{1-q}}$$

Which is the q-exponential if $q \neq 1$. When $q = 1$, $e_q(x)$ is just $\exp(x)$.

The cumulative distribution function (cdf) of the q-Exponential-generated family is given by.

$$F(x) = \int_0^{G(x)/(1-G(x))} (2-q)\alpha[1-(1-q)\alpha x]^{\frac{1}{1-q}} \tag{2}$$

The simplified form of (2) is.

$$F(X) = 1 - [1 - (1-q)\,\alpha\,\frac{G(X)}{1-G(X)}]^{\frac{2-q}{1-q}} \qquad x, \alpha > 0, q<2 \tag{3}$$

where q is the shape parameter (entropy index) and α is the scale parameter. The corresponding probability density function is given by

$$f(x) = (2-q)\,\alpha\,\frac{g(x)}{[1-G(x)]^2}\,[1 - (1-q)\,\alpha\,\frac{G(X)}{1-G(X)}]^{\frac{1}{1-q}} \tag{4}$$

where $X>0$, $\alpha>0$, $q<2$.

The rest of the paper are as follows. In Section 2, The new class q-Exponential-Weibull distribution is introduced and presented its particular cases. The mathematical and statistical properties are discussed in section 3 and in section 4, the maximum likelihood estimation method and their asymptotic behaviors have been discussed. Simulation techniques has been explained in section 5. Real life failure time data has been applied and the results are presented in section 6. In section 7, we have discussed the conclusion of the new class of q-Exponential Weibull distribution.

2. The q-Exponential-Weibull Distribution

The q-Exponential distribution combined with Weibull distribution gives the q-Exponential Weibull distribution. Here the q-exponential is the generator distribution, and the two-parameter Weibull distribution (Waloddi Weibull, 1951) is a parent distribution whose pdf and cdf are given by

$$g(x, \lambda, \gamma) = \lambda\gamma x^{\gamma-1} e^{-\lambda x^{\gamma}} \qquad x, \gamma, \lambda > 0 \tag{5}$$

$$G(x, \lambda, \gamma) = 1 - e^{-\lambda x^{\gamma}} \tag{6}$$

using (6) in (3), we get the new cdf of q-Exponential-Weibull distribution. The simplified form of q-Exponential-Weibull distribution is

$$F(x, \Omega) = 1-[1 - (1 - q)\, \alpha\, (e^{\lambda x^{\gamma}} - 1)]^{\frac{2-q}{1-q}} \tag{7}$$

where $\Omega = \{q, \alpha, \lambda, \gamma\}$ be the set of parameters, here q and γ are the shape parameters and α, λ are the scale parameters. The equation (7) is called the cdf of q-Exponential-Weibull distribution.
Substituting (5) and (6) in (4), we get the new pdf. The new pdf is,

$$f(x, \Omega) = (2-q)\, \alpha\lambda\gamma\, e^{\lambda x^{\gamma}} x^{\gamma-1} [1 - (1 - q)\, \alpha\, (e^{\lambda x^{\gamma}} - 1)]^{\frac{1}{1-q}} \tag{8}$$

Rewriting the above equation (8), we get

$$f(x, \Omega) = (2-q)\, \alpha\lambda\gamma\, e^{\lambda x^{\gamma}} x^{\gamma-1}\, e_q [-\alpha\, (e^{\lambda x^{\gamma}} - 1)] \tag{9}$$

The equation (8) and (9) are called the pdf of q-Exponential-Weibull distribution (q-EW). The particular case of our new q-Exponential-Weibull distribution is presented in Table 1.

Table 1: *The particular case of q-Exponential-Weibull distribution*

Model	q	α	λ	γ	Cdf	References
q-Exponential-Exponential	q	α	λ	1	$1-[1-(1-q)\,\alpha\,(e^{\lambda x}-1)]^{\frac{2-q}{1-q}}$	New
q- Exponential-Rayleigh	q	α	$\frac{\lambda}{2}$	2	$1-[1-(1-q)\,\alpha\,(e^{\frac{\lambda x^2}{2}}-1)]^{\frac{2-q}{1-q}}$	New
Exponential – Weibull	1	α	λ	γ	$1-e^{-\alpha(e^{\lambda x^{\gamma}}-1)}$	New
Exponential – Exponential	1	α	λ	1	$1-e^{-\alpha(e^{\lambda x}-1)}$	Elgarhy et al. (2017)
Exponential- Rayleigh	1	α	$\frac{\lambda}{2}$	2	$1-e^{-\alpha(e^{\frac{\lambda}{2}x^2}-1)}$	Kawsar Fatima and, S.P Ahmad (2017)

2.1 Reliability Measures: Survival function: (survivor function)

The survivor function for the new distribution S(x) is defined to be the probability that the survival time is greater than or equal to t, and it is given by

$$S(x) = P\ (X \geq t) = 1- F(x)$$

$$S\ (x, \Omega) =[1 -\ (1 - q)\ \alpha\ (e^{\lambda x^{\gamma}} -\ 1)]^{\frac{2-q}{1-q}} \tag{10}$$

2.2 Hazard function:

The hazard function is used to express the risk or hazard of an event such as death occurring at some time t, and it is given by

$$h\ (x) = \frac{f(x)}{s(x)}$$

Substituting (8) and (10) we get the hazard function of q-Exponential-Weibull distribution. which is defined below.

$$h\ (x, \Omega) = (2-q)\ \alpha\lambda\gamma\ e^{\lambda x^{\gamma}} x^{\gamma-1}[1 -\ (1 - q)\ \alpha\ (e^{\lambda x^{\gamma}} -\ 1)]^{\frac{-(1+q)}{(1-q)}} \tag{11}$$

2.3 Reverse hazard rate function

The reverse hazard function of q-Exponential-Weibull distribution is defined by

$$r\ (x) = \frac{f(x)}{F(x)}$$

$$r\ (x, \Omega) = \frac{(2-q)\ \alpha\lambda\gamma\ e^{\lambda x^{\gamma}} x^{\gamma-1}[1-\ (1-q)\ \alpha\ (e^{\lambda x^{\gamma}} - 1)]^{\frac{1}{1-q}}}{1-[1-\ (1-q)\ \alpha\ (e^{\lambda x^{\gamma}} - 1)]^{\frac{2-q}{1-q}}} \tag{12}$$

2.4 Cumulative hazard function

Cumulative hazard function is presented below,

$$H\ (x, \Omega) = -\log\ (1-F(x))$$

$$H\ (x, \Omega) = -\ln\ (s\ (x, \Omega)) = -\ln\ [[1 -\ (1 - q)\ \alpha\ (e^{\lambda x^{\gamma}} -\ 1)]^{\frac{2-q}{1-q}}] \tag{13}$$

The above equation is known as cumulative hazard function of q-Exponential-Weibull distribution.

2.5 Graphical Study of q-Exponential Weibull distribution under various functions:

In this section, we studied the structure of the cdf, pdf, S(x) and h(x) of q-Exponential-Weibull distribution using different values of the parameters. The illustrative figures are presented below.

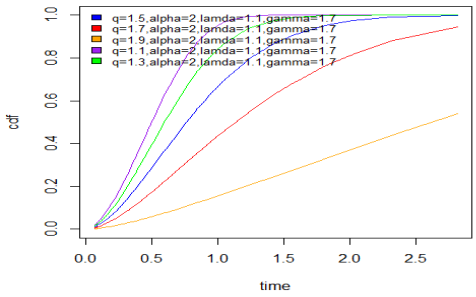


Figure 2.a: *The graph of the cdf of the q-EW distribution with different values of the parameter*

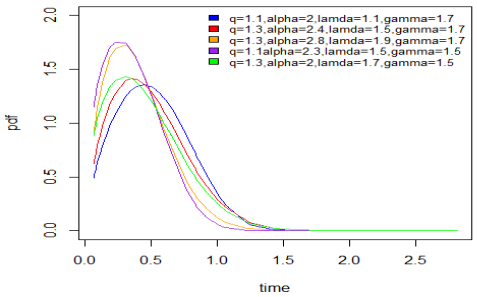


Figure 2.b1: *Graph of the pdf of the q-EW distribution when all the parameters are changed*

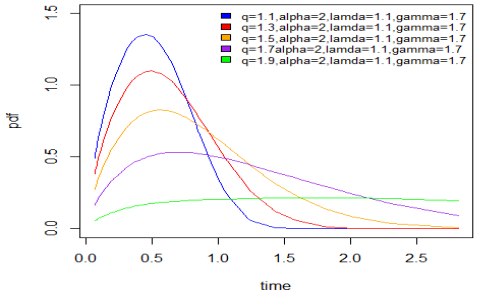


Figure 2.b2:*The graph of the pdf of the q-EW distribution when changing first shape parameter (q) values and other parameters are fixed*

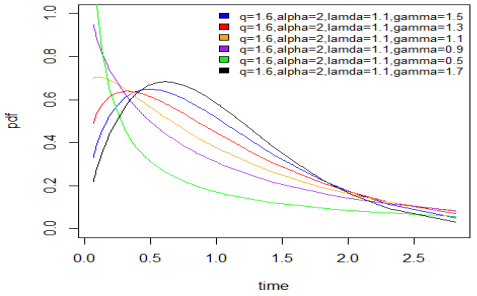


Figure 2.b3: *The graph of the pdf the of q-EW distribution when changing second shape parameter (γ)values and other parameters are fixed*

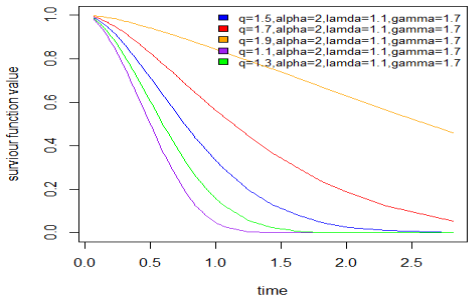


Figure 2.c: *The graph of the survival function of the q-EW distribution with different parameter values*

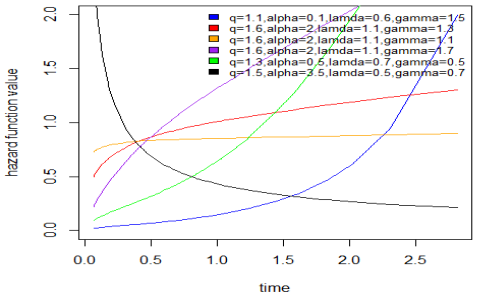


Figure 2.d: *The graph of the Hazard rate of the q-EW distribution for with different parameter values*

Figure 2.a cumulative density plot demonstrates the validity of the distribution as a probability distribution. The probability density function graphs (2.b1,2. b2 and 2.b3) shows that it is skewed and more adaptable for various parameter values. The graph of the hazard function (2.d) demonstrates that it can take on various shapes, including constant, increasing, and decreasing. As a result, fitting data sets of different forms may be done and which was quite well using the q-Exponential Weibull distribution.

3.Properties

In this section we study some mathematical and statistical properties of q-Exponential-Weibull distribution.

3.1 Mixture Representation:

Several properties of the new distribution can be derived using the concept of exponentiated distribution. The mixture representation of q-exponential-Weibull distribution is derived in the following sections.

Using the generalized binomial theorem, where β>0 is real non integer and |z|<1,

$$(1-z)^{\beta-1} = \sum_{k=0}^{\infty}(-1)^k \binom{\beta-1}{k}(z)^k$$

$$f(x) = (2\text{-}q)\,\alpha \frac{g(x)}{[1-G(x)]^2}\left[1-(1-q)\,\alpha\,\frac{G(X)}{1-G(X)}\right]^{\frac{2-q}{1-q}-1} \qquad \text{since } \beta = \frac{2-q}{1-q}$$

$$= (2\text{-}q)\,\alpha \frac{g(x)}{[1-G(x)]^2}\sum_{k=0}^{\infty}(-1)^k \binom{\beta-1}{k}(1\text{-}q)^k\,\alpha^k \left[\frac{G(X)}{1-G(X)}\right]^k$$

$$= (2\text{-}q)\,\alpha \frac{g(x)}{[1-G(x)]^2}\sum_{k=0}^{\infty}(-1)^k \binom{\beta-1}{k}(1\text{-}q)^k\,\alpha^k \frac{[G(x)]^k}{[1-G(x)]^k}$$

$$=\sum_{k=0}^{\infty}(-1)^k \binom{\beta-1}{k}(1\text{-}q)^k\,\alpha^{k+1}\,(2\text{-}q)\,\frac{g(x)\,[G(x)]^k}{[1-G(x)]^{k+2}}$$

Generalized binomial theorem

$$[1-G(x)]^{-(k+2)} = \sum_{j=0}^{\infty}\frac{\Gamma(k+j+2)}{j!\,\Gamma(k+2)}[G(x)]^j$$

$$= \sum_{k=0}^{\infty}(-1)^k \binom{\beta-1}{k}(1\text{-}q)^k\,\alpha^{k+1}\,(2\text{-}q)\sum_{j=0}^{\infty}\frac{\Gamma(k+j+2)}{j!\,\Gamma(k+2)} g(x)\,[G(x)]^{k+j+1-1}$$

$$=\sum_{k=0}^{\infty}(-1)^k \binom{\beta-1}{k}(1\text{-}q)^k\,\alpha^{k+1}\,(2\text{-}q)\sum_{j=0}^{\infty}\frac{\Gamma(k+j+2)}{j!(k+j+1)\,\Gamma(k+2)}(k+j+1)\,g(x)\,[G(x)]^{k+j+1-1}$$

$$f(x,\Omega) = \sum_{j,k=0}^{\infty} W_{j,k}\, h_{(k+j+1)}(x,\Omega) \qquad (14)$$

where $W_{j,k} = (-1)^k \binom{\beta-1}{k}(1\text{-}q)^k\,\alpha^{k+1}\,(2\text{-}q)\frac{\Gamma(k+j+2)}{j!(k+j+1)\,\Gamma(k+2)}$

$$h_a(x,\Omega) = a\,g(x,\Omega)\,[G(x,\Omega)]^{a-1}$$

The q-Exponential Weibull density can be expressed as an infinite linear combination of exponentiated – G density function.

Then, $$[F(x)]^R = [1-[1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{2-q}{1-q}}]^R$$

Using the generalized binomial theorem, where β>0 is real non integer and |z|<1,

$$(1-z)^{\beta-1} = \sum_{l=0}^{\infty}(-1)^l \binom{\beta-1}{l}(z)^l$$

$$[F(x)]^R = \sum_{l=0}^{\infty}(-1)^l \binom{R}{l}[1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{(2-q)l}{(1-q)}}$$

Which can also be written as,

$$[F(x)]^R = \sum_{l,m=0}^{\infty}(-1)^{l+m}\binom{R}{l}\binom{\vartheta}{m}(1\text{-q})^{m}\alpha^{m}(e^{\lambda x^{\gamma}}-1)^{m} \text{ where } \vartheta = \frac{(2-q)l}{(1-q)}$$

$$[F(x)]^R = \sum_{l,m=0}^{\infty}(-1)^{l+m}(1\text{-q})^{m}\alpha^{m}\binom{R}{l}\binom{\vartheta}{m}\left[\frac{1-e^{-\lambda x^{\gamma}}}{1-(1-e^{-\lambda x^{\gamma}})}\right]^{m}$$

Using Generalized binomial theorem, the above equation can be written as,

$$[F(x)]^R = \sum_{l,m,n=0}^{\infty}(-1)^{l+m}(1\text{-q})^{m}\alpha^{m}\binom{R}{l}\binom{\vartheta}{m}\binom{m+n-1}{n}\left[1-e^{-\lambda x^{\gamma}}\right]^{m+n}$$

Simply further we get,

$$[F(x)]^R = \sum_{l,m,n,r=0}^{\infty}(-1)^{l+m+r}(1\text{-q})^{m}\alpha^{m}\binom{R}{l}\binom{\vartheta}{m}\binom{m+n-1}{n}\binom{m+n}{r}(e^{-\lambda x^{\gamma}})^{r}$$

$$[F(x)]^R = \sum_{l,m,n,r=0}^{\infty} W_{l,m,n,r}(e^{-\lambda x^{\gamma}})^{r} \quad (15)$$

Where $W_{l,m,n,r} = (-1)^{l+m+r}(1\text{-q})^{m}\alpha^{m}\binom{R}{l}\binom{\vartheta}{m}\binom{m+n-1}{n}\binom{m+n}{r}$

3.2 Limiting Behavior:

Lemma 1: The limit of the cdf of the q-Exponential-Weibull, F(x) as X approaches infinity, $x\to\infty$ is equal to one and limit of the cdf of the q-Exponential-Weibull, F(x) as X tends to zero, $x \to 0$ is equal to zero.

$$\lim_{x\to\infty} F(x) = 1$$

Proof: The cdf of the q-Exponential Weibull F(x) as X approaches infinity ($x \to \infty$), from 7 we get Using equation (9)

$$\lim_{x\to\infty} F(x, \Omega) = \lim_{x\to\infty} 1\text{-}[1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{2-q}{1-q}}$$

$$= 1\text{-}[1-(1-q)\,\alpha\,(e^{\lambda(\infty)^{\gamma}}-1)]^{\frac{2-q}{1-q}}$$

$$= [1-0] = 1$$

Hence, the lemma is proved under limiting property.

$$\lim_{x\to 0} F(x) = 0$$

$$\lim_{x\to 0} F(x, \Omega) = \lim_{x\to 0} 1\text{-}[1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{2-q}{1-q}}$$

$$= 1\text{-}[1-(1-q)\,\alpha\,(e^{\lambda(0)^{\gamma}}-1)]\ \text{^}2\text{-q/ (1-q)}$$

$$= 1\text{-}[1-0] = 0$$

Lemma 2: In probability theory, of a continuous random variable has the following property

(i) $f(x) \geq 0$; where $-\infty<x<\infty$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

Using above definition, the validity of the model f(x) is checked. In our survival model the range of x is $0<x<\infty$.

$$\int_{0}^{\infty} f(x)dx = 1$$

$$= \int_0^{\infty} (2-q)\,\alpha\lambda\gamma\, e^{\lambda x^{\gamma}} x^{\gamma-1} [1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{1}{1-q}} dx$$

$$= (2-q)\,\alpha\lambda\gamma \int_0^{\infty} e^{\lambda x^{\gamma}} x^{\gamma-1} [1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{1}{1-q}} dx$$

Now y = $[1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]$

$$\frac{dy}{dx} = 0 - (1\text{-}q)\alpha\lambda\gamma e^{\lambda x^{\gamma}} x^{\gamma-1}$$

$$\frac{dy}{(1-q)\alpha\lambda\gamma e^{\lambda x^{\gamma}} x^{\gamma-1}} = dx$$

$$= (2-q)\,\alpha\lambda\gamma \int_0^{\infty} \frac{e^{\lambda x^{\gamma}} x^{\gamma-1}}{[1-(1-q)\,\alpha\,(e^{\lambda x^{\gamma}}-1)]^{\frac{1}{1-q}}} * \frac{dy}{(1-q)\alpha\lambda\gamma e^{\lambda x^{\gamma}} x^{\gamma-1}}$$

$$= -\frac{2-q}{1-q} \int_0^{\infty} y^{1/(1-q)}\, dy$$

$$\int_0^{\infty} f(x)dx = -[y^{(2-q)/(1-q)}]_0^{\infty}$$

$$= -\{\, [[1-(1-q)\alpha\,(e^{\lambda x^{\gamma}}-1)]^{(2-q)/(1-q)}]_{x=\infty} - [[1-(1-q)\alpha\,(e^{\lambda x^{\gamma}}-1)]^{(2-q)/(1-q)}]_{x=0}\,\}$$

$$\int_0^{\infty} f(x)dx = -[0-1] = 1$$

Hence q-Exponential-Weibull distribution is a valid pdf.

3.3 Quantile Function:

The quantile function of X= Q(u) = $F^{-1}(u)$ can be obtained by inverting equation (7) as follows,

$$Q(u) = \left[\frac{1}{\lambda} \ln\left[1 + \frac{1}{(1-q)\alpha}\left[1-(1-u)^{\frac{1-q}{2-q}}\right]\right]\right]^{\frac{1}{\gamma}} \quad (16)$$

Simulation of q-Exponential-Weibull random variable is straightforward. Let u be the uniform variable on the interval [0,1], then the random variable X = $F^{-1}(u)$ follows q-Exponential-Weibull distribution given in equation (8) with the parameters (q, α, λ, γ). By using equation (20), we can obtain the first, second and third quantiles by replacing u as 0.25, 0.5 and 0.75, respectively.

3.4 Moments:

This section provides the moment and moment generating function of q-Exponential-Weibull distribution. The moments of the functions are quantitative measures related to the shape of the function. The first four moments, skewness and kurtosis of q-Exponential-Weibull distribution can be obtained as

$$\mu_r' = E[x^r] = \int_{-\infty}^{\infty} x^r f(x,\Omega)dx$$

Using equation (13) we have,

$$\mu_r' = \int_{-\infty}^{\infty} x^r \sum_{j,k=0}^{\infty} W_{j,k}\, h_{(k+j+1)}(x,\Omega)\, dx$$

$$= \sum_{j,k=0}^{\infty} W_{j,k} \int_{-\infty}^{\infty} x^r h_{(k+j+1)}(x,\Omega)dx$$

Where $I_{j,k}(x,\Omega) = \int_{-\infty}^{\infty} x^r h_{(k+j+1)}(x,\Omega)dx$

$$\mu'_r = E(x^r) = \sum_{j,k=0}^{\infty} W_{j,k}\ I_{j,k}(x, \Omega) \quad (17)$$

The mean, variance, skewness and kurtosis can be obtained from equation (14).

when r =1 gives mean = E(x)

variance = $E(x^2) - [E(x)]^2$

skewness = $\frac{\mu_3(\theta) - 3\mu_3(\theta)\mu_2(\theta) + 2\mu_1^3(\theta)}{[\mu_2(\theta) - \mu_1^2(\theta)]^{3/2}}$

kurtosis = $\frac{\mu_4(\theta) - 4\mu_1(\theta)\mu_3(\theta) + 6\mu_1^2(\theta)\mu_2(\theta) - 3\mu_1^4(\theta)}{[\mu_2(\theta) - \mu_1^2(\theta)]^2}$

Generally, the moment generating function of q-Exponential-Weibull distribution is obtained through the following relation

$$M_x(t, \Omega) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^{\infty} x^r f(x)dx = \sum_{r=0}^{\infty} \frac{t^r}{r!} E(x^r) = \sum_{r,j,k=0}^{\infty} \frac{t^r}{r!} W_{j,k}\ I_{j,k}(x, \Omega) \quad (18)$$

The Characteristic function of q-Exponential-Weibull distribution is obtained through the following relation

$$\phi_x(t, \Omega) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \int_0^{\infty} x^r f(x)dx = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} E(x^r) = \sum_{r,j,k=0}^{\infty} \frac{(it)^r}{r!} W_{j,k}\ I_{j,k}(x, \Omega) \quad (19)$$

The cumulant generating function of q-Exponential-Weibull distribution is given by

$$k_x(t, \Omega) = log[[\sum_{r=0}^{\infty} \frac{(t)^r}{r!} \int_0^{\infty} x^r f(x)dx] = \log [\sum_{r=0}^{\infty} \frac{(t)^r}{r!} E(x^r)]$$

$$= log[\sum_{r,j,k=0}^{\infty} \frac{(t)^r}{r!} W_{j,k}\ I_{j,k}(x, \Omega)] \quad (20)$$

3.5 Order Statistics:

Let $X_{1:n} < X_{2:n} < X_{3:n} < \ldots < X_{n:n}$ be the order statistics of a random sample of size n following q-Exponential-Weibull distribution with the parameter α, q, λ, γ then the probability density function of p^{th} order statistic can be written as,

$$f_{(x_p)}[x_{(p)}] = \frac{f(x_p)}{B(p, n-p+1)} \sum_{v=0}^{n-p} (-1)^v \binom{n-p}{v} [F(x_p)]^{v+p-1} \quad (21)$$

Substituting (13) and (14) in (18) and replacing R= $(v + p - 1)$ we get

$$f_{(x_p)}[x_{(p)}] = \frac{\sum_{j,k=0}^{\infty} W_{j,k} h_{(k+j+1)}(X, \Omega)}{B(p, n-p+1)} \sum_{v=0}^{n-p} (-1)^v \binom{n-p}{v} \sum_{l,m,n,r=0}^{\infty} W_{l,m,n,r} (e^{-\lambda x^{\gamma}})^r$$

$$f_{(x_p)}[x_{(p)}] = \frac{1}{B(p, n-p+1)} \sum_{j,k=0}^{\infty} \sum_{v=0}^{n-p} \sum_{l,m,n,r=0}^{\infty} \omega^* h_{(k+j+1)}(X, \Omega)(e^{-\lambda x^{\gamma}})^r \quad (22)$$

Where $\omega^* = (-1)^v \binom{n-p}{v} W_{j,k} W_{l,m,n,r}$

4. Method of Estimation

In this section, the maximum likelihood estimates (MLE) of the unknown parameters for the q-Exponential-Weibull distribution are determined based on complete samples. Let $x_1, x_2 \ldots x_n$ be a random sample from q-Exponential-Weibull distribution with unknown parameter vector $\Omega = \{q, \alpha, \lambda, \gamma\}$. The likelihood function for the proposed distribution $\mathcal{L}$ is given by

$$\mathcal{L}(x, \Omega) = (2-q)^n \alpha^n \lambda^n \gamma^n \prod_{i=1}^{n} e^{\lambda x_i{}^{\gamma}} x_i{}^{\gamma-1} [1 - (1-q)\,\alpha\,(e^{\lambda x_i{}^{\gamma}} - 1)]^{\frac{1}{1-q}}$$

Then the log likelihood of the equation is

$$\ell(\Omega) = \log \mathcal{L}(t, \Omega) = n \log (2\text{-}q) + n \log \alpha + n \log \lambda + n \log \gamma + \lambda \sum_{i=1}^{n} x_i{}^{\gamma} + (\gamma - 1) \sum_{i=1}^{n} \log x_i + \left(\frac{1}{1-q}\right) \sum_{i=1}^{n} \log [1 - (1-q)\,\alpha\,(e^{\lambda x_i{}^{\gamma}} - 1)] \quad (23)$$

The maximum likelihood estimates of the parameters (q, α, λ, γ) are found by taking a partial derivative of $\ell(\Omega)$ with respect to q, α, λ, γ, equating the derivatives to zero, and evaluating them at $\hat{q}$, $\hat{\alpha}$, $\hat{\lambda}$, $\hat{\gamma}$.

$$\frac{\partial \ell(\Omega)}{\partial q} = \frac{n}{2-q} + \sum_{i=1}^{n} \frac{1}{1-q} \left[\frac{\alpha (e^{\lambda x_i{}^{\gamma}} - 1)}{1 - (1-q)\,\alpha\,(e^{\lambda x_i{}^{\gamma}} - 1)}\right] - \sum_{i=1}^{n} \log [1 - (1-q)\alpha \left(e^{\lambda x_i{}^{\gamma}} - 1\right) \quad (24)$$

$$\frac{\partial \ell(\Omega)}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \frac{1}{1-q} \left[\frac{(1-q)(e^{\lambda x_i{}^{\gamma}} - 1)}{1 - (1-q)\,\alpha\,(e^{\lambda x_i{}^{\gamma}} - 1)}\right] \quad (25)$$

$$\frac{\partial \ell(\Omega)}{\partial \lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n} x_i{}^{\gamma} - \sum_{i=1}^{n} \frac{1}{1-q} \left[\frac{\alpha (1-q)\, x_i{}^{\gamma} e^{\lambda x_i{}^{\gamma}}}{1 - (1-q)\,\alpha\,(e^{\lambda x_i{}^{\gamma}} - 1)}\right] \quad (26)$$

$$\frac{\partial \ell(\Omega)}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n} \log x_i + \sum_{i=1}^{n} \lambda\, x^{\gamma} \log x_i - \sum_{i=1}^{n} \frac{1}{1-q} \left[\frac{\alpha\lambda (1-q)\, x_i{}^{\gamma} e^{\lambda x_i{}^{\gamma}} * \log x_i}{1 - (1-q)\,\alpha\,(e^{\lambda x_i{}^{\gamma}} - 1)}\right] \quad (27)$$

For solving these non-linear equation's, we can use any iteration method such as Newton-Raphson technique.

5. Generating random samples from q-Exponential Weibull distribution

The Inverse CDF method is used for generating random numbers from a particular distribution. In this method, random numbers from a particular distribution are generated by solving the equation obtained on equating CDF of a distribution to a number u. The number u is itself being generated from u~$U(0,1)$. In this section we made an attempt to q-Exponential-Weibull distribution to generate the random number using equation 16 at a fixed values of parameters $(q, \alpha, \lambda, \gamma)$.

$$X = F^{-1}(u)$$

$$X = \left[\frac{1}{\lambda} \ln \left[1 + \frac{1}{(1-q)\alpha} \left[1 - (1-u)^{\frac{1-q}{2-q}}\right]\right]\right]^{\frac{1}{\gamma}} \quad (28)$$

For uniform over (0,1) then x~$q - EW(1.2,2,1.1,1.7)$ can be generated random sample of size 50 are presented below.

0.2471, 0.8991, 0.9387, 1.5307, 0.6133, 0.8110, 0.8077, 0.7611, 0.8320, 1.5590,
1.0539, 1.7640, 1.4155, 0.8662, 1.2408, 1.9081, 1.0625, 0.6137, 0.5943, 0.7125,
0.9593, 0.3809, 0.1623, 0.2987, 0.9664, 1.31036, 0.6269, 1.3524, 0.6302, 1.0810,
2.1260, 1.4057, 1.1020, 0.6074, 1.7022, 1.1539, 1.1613, 0.5775, 0.1133, 0.9533,
1.1283, 1.2516, 1.6930, 0.9185, 1.3880, 0.8035, 0.9471, 0.1955, 2.4077, 0.7141

Here we have used one of the goodness of fit tests "Kolmogorov-Smirnov (KS)" test for the above-generated data for testing the q-Exponential-Weibull distribution. The null hypothesis is that

the samples are drawn from the q-Exponential-Weibull distribution against the alternative hypothesis is that the samples are not drawn from the q-Exponential-Weibull distribution. The test statistic value of the KS test for the generated samples is (D value) 0.097 at 5% level of significance with the p-value of 0.76. Since the p-value is greater than 0.05, the null hypothesis is accepted. Hence, the samples are drawn from the proposed distribution. Therefore, the q-Exponential-Weibull distribution has satisfied the goodness of fit test.

6. Application to real life data

In this section, we have used different kinds of real-life failure time data to show the suitability of the q-Exponential Weibull distribution, also we have compared to some other related distributions namely Exponentiated Weibull-Exponential (EWE) and Generalized Exponential-Weibull (GEW) distributions. The pdf of the respective distributions is represented below:

- The Exponentiated Weibull-Exponential (EWE) distribution introduced by Elgarhy et.al [8], with pdf

$$f(x) = q\alpha\gamma\lambda[e^{\lambda x}-1]^{\gamma-1}\exp\left[-\left\{\alpha\left[e^{\lambda x}-1\right]^{\gamma}-\lambda x\right\}\right]\left[1-exp\left(-\alpha\left[e^{\lambda x}-1\right]^{\gamma}\right)\right]^{q-1} \quad x,q,\alpha,\lambda,\gamma>0 \qquad (29)$$

- The Generalized Exponential-Weibull (GEW) distribution introduced by Dikko and Faisal [6], with the pdf

$$f(x) = q(\alpha+\gamma\lambda x^{\lambda-1})e^{-(\alpha x+\gamma x^{\lambda})}[1-e^{-(\alpha x+\gamma x^{\lambda})}]^{q-1} \qquad x,\alpha,\gamma,\lambda,q>0 \qquad (30)$$

In order to assess the flexibility of the proposed distribution, we have considered some model selection criteria like, -2loglikelihood and AIC (Akaike Information Criterion) are used and analyses performed with the aid of R software.

Dataset1: The first data set is the failure times of 84 aircraft windshields. This failure time data set is available in Murthy et al's book "Weibull Models" (2004, page 297). A large aircraft's windscreen is a sophisticated piece of equipment made up of multiple layers of material, including a very touchy outer skin with a heated layer just behind it, all laminated under high temperature and pressure. These failures do not cause aircraft damage, but they do require the repair of the windscreen. The failure times of 84 aircraft windshields are given below:

0.040, 1.866, 2.385, 3.443, 0.301, 1.876, 2.481, 3.467, 0.309, 1.899, 2.610, 3.478, 0.557, 1.911, 2.625, 3.578, 0.943, 1.912, 2.632, 3.595, 1.070, 1.914, 2.646, 3.699, 1.124, 1.981, 2.661, 3.779,1.248, 2.010, 2.688, 3.924, 1.281, 2.038, 2.823, 4.035, 1.281, 2.085, 2.890, 4.121, 1.303, 2.089, 2.902, 4.167, 1.432, 2.097, 2.934, 4.240, 1.480, 2.135, 2.962, 4.255, 1.505, 2.154, 2.964, 4.278, 1.506, 2.190, 3.000, 4.305, 1.568, 2.194, 3.103, 4.376, 1.615, 2.223, 3.114, 4.449, 1.619, 2.224, 3.117, 4.485, 1.652, 2.229, 3.166, 4.570, 1.652, 2.300, 3.344, 4.602, 1.757, 2.324, 3.376, 4.663

Table 2: *Estimates of fitted distribution for aircraft windshield failure data*

Model	Estimated Parameters				Model Selection	
	$\hat{q}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\gamma}$	-2LL	AIC
q-EW	1.729287	4.629657	0.006168	1.539852	**251**	**259**
EWE	15.46262	1.38606	4.08592	0.07846	253	261
GEW	0.04796	0.31873	0.43050	0.68102	419	427

Dataset 2: The second data set represents the survival times (in days) of 72 guinea pigs infected with virulent tubercle bacilli, observed and reported by Bjerkedal. The data is presented below:

0.1, 0.33, 0.44, 0.56, 0.59, 0.72, 0.74, 0.77, 0.92, 0.93, 0.96, 1, 1, 1.02, 1.05, 1.07, 07, .08, 1.08, 1.08, 1.09, 1.12, 1.13, 1.15, 1.16, 1.2, 1.21, 1.22, 1.22, 1.24, 1.3, 1.34, 1.36, 1.39, 1.44, 1.46, 1.53, 1.59, 1.6, 1.63, 1.63, 1.68, 1.71, 1.72, 1.76, 1.83, 1.95, 1.96, 1.97, 2.02, 2.13, 2.15, 2.16, 2.22, 2.3, 2.31, 2.4, 2.45, 2.51, 2.53, 2.54, 2.54, 2.78, 2.93, 3.27, 3.42, 3.47, 3.61, 4.02, 4.32, 4.58, 5.55.

Table 3: *Estimates of fitted distribution for guinea pig failure data*

Model	Estimated Parameters				Model Selection	
	$\hat{q}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\gamma}$	-2LL	AIC
q-EW	1.54493	33.49094	0.01988	2.98777	**184**	**192**
EWE	1.08287	83.46078	0.01628	2.99604	187	195
GEW	0.66675	0.07983	0.27917	0.45848	242	250

We observed from the above tables 2 and 3, the -2LL and AIC values of the q-Exponential-Weibull distribution have the smallest among the other distributions. Therefore, the q-Exponential-Weibull distribution has performed well than the other distributions. So, we conclude from this section, the q-Exponential-Weibull distribution has achieved the goal of the suitability of the different kinds of real-life failure time data.

8. Conclusion

In this research article, we have introduced a new class of four-parameter distribution referred to as "q-Exponential-Weibull distribution" by taking the Weibull distribution as the base distribution and the q-Exponential distribution as the generator distribution by using the generator technique. The q-Exponential-Weibull density can be expressed as a linear combination of exponentiated - G densities. For checking the model properties, we have derived survival, hazard, cumulative hazard and reverse hazard functions from q-Exponential-Weibull distribution, and also studied graphically. In the graphical study of the q-Exponential-Weibull distribution under various functions with different parameter values, the proposed distribution has achieved the properties of the density function. The mathematical and statistical properties are applied to q-Exponential-Weibull distribution. The q-Exponential-Weibull distribution has satisfied the above said properties. The parameters of the q-Exponential-Weibull distributions are estimated using the maximum likelihood estimation method. The random samples have been generated from the q-Exponential-Weibull distribution and the goodness of fit test has been verified using Kolmogorov-Smirnov (KS) test, also we have studied the application of real-time failure time data to q-Exponential-Weibull distribution. The proposed distribution performed well than the other distribution based on the model selection criteria. Based on the above-said results, the q-Exponential-Weibull distribution is more adaptable and more flexible to fit the real-life failure time data. We hope that the proposed distribution would draw more widespread applications in different areas of research such as reliability analysis, medicine engineering and economics etc.

References

[1] Adrian, A. Budini. (2015). Extended q-Gaussian and q-exponential distributions from Gamma random variables. *Physical Review E*,91,052113.

[2] Alzaatreh, A., Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71 (1):63-79.

[3] Ana Claudia Souza Vidal de Negreirosa, Isis Didier Lins, Marcio Jose das Chagas Moura and Enrique Lopez Droguettc. (2020). Reliability data analysis of systems in the wear-out phase using a (corrected) q-Exponential likelihood. *Reliability Engineering and System Safety*,197,106787.

[4] Bourguignon, M., Silva, R.B. and Cordeiro, G.M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12(1):53-68.

[5] Collett D. modelling survival data in medical Research, chapman and hall, London,2003.

[6] Dikko, H.G. and Faisal, A.M. (2017). A New generalized Exponential-Weibull distribution: Its properties and application. *Bayero Journal of Pure and Applied Sciences*, 10(2):29-37.

[7] Duarte Queiros, S.M., L. G. Moyano, J. de Souza, and Tsallis. C. (2007). A non-extensive approach to the dynamics of financial observables. *The European Physical Journal B*, 55, 161

[8] Elgarhy.M., Shakil.M., and Golam Kibria B.M. (2017). Exponentiated Weibull-Exponential distribution with applications. *An International journal Applications and applied mathematics,* 12(2):710-725.

[9] Fode Zhang, Hon Keung Tony Ng and Yimin Sh. (2018). Information geometry on the curved q-exponential family with application to survival data analysis. *Physica A,* 512:788–802.

[10] Fode Zhang, Yimin Shi and Ruibing Wang. (2017). Geometry of the q-exponential distribution with dependent competing risks and accelerated life testing. *Physica A,* 468:552–565.

[11] Islam, B and AI-Talib, M. (2019). Exponentiated Q-Exponential distribution proceedings the 6th International Arab Conference on mathematics and computations (IACMC 2019).

[12] Keith Briggs and Christian Beck. (2007). Modelling train delays with q-exponential functions. *Physica A: Statistical mechanics and its applications*, 378(2):498–504.

[13] Lawless, J.F. Statistical Models and Methods for Lifetime Data, John Wiley & sons, New York, 1982.

[14] Lee, E.T. and Wang, J.W., Statistical methods for survival data analysis, 3rd Edition, John Wiley and Sons, New York, ISBN: 9780471458555; pages:534.

[15] Malacarne, l.c., mende, r.s. and lenzi, e.k. (2001). q-exponential distribution in urban agglomeration. *Physical Review E*,65:017106.

[16] Marshall, A., Olkin, I. (1997). A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families. *Biometrika*, 84:641– 652.

[17] Moeschberger. (2006). Survival Analysis: Techniques for Censored and Truncated Data. *Second edition, springer,* ISBN 978-0-387-21645-4.

[18] Mohammed. S. Jalal and Ferash. M. Batol. (2023). Reliability of stress-strength and its estimation of Exponentiated exponential distribution. *Iraqi journal of science*, 64(3):1299-1306.

[19] Narayanaswamy Sundaram (2019). Modelling Censored Survival Data With q-Exponential Distributions. *Global Journal for Research Analysis*, 8 (7).

[20] Nicy Sebastian, Jeena Joseph and Princy T. (2022). Type 1 Topp-Leone q−Exponential Distribution and its Applications. *Reliability Theory & Applications*, 3 (69), 17:361-375.

[21] Picoli Jr., R.S. Mendes and L.C. Malacarne. (2003). q-exponential, Weibull, and q-Weibull distributions: an empirical analysis. *Physica A,* 324:678–688.

[22] Picoli Jr., R.S. Mendes, L. C. Malacarne, R. P. B. Santos. (2009). q-distributions in complex systems: a brief review. *Brazilian Journal of Physics*, 39:468-474.

[23] Sebastian, N. Rasin, R. S. and Silviya, P. O. (2019). Topp-Leone Generator Distributions and its Applications. Proceedings of National Conference on Advances in Statistical Methods, 127-139.

[24] Sales Filho R, Lopes Droguett E, Lins I, Moura M. C, Azevedo R. (2016). Stress-strength reliability estimation based on the q-Exponential distribution. *Quality and Reliability Engineering International*, 4:51.

[25] Shalizi c.r. (2007), Maximum likelihood estimation for q-exponential (Tsallis) distributions, http://arxiv.org/abs/math/0701854.

[26] Tsallis. C, Introduction to Non-extensive Statistical Mechanics- Approaching a Complex World, Springer, New York (2009).

RT&A, No.3 (74)
Volume 18, September 2023

Reliability Acceptance Sampling Plan for One Parameter Polynomial Exponential Distribution

Anumita Mondal and Sudhansu S. Maiti

•

Department of Statistics, Visva-Bharati University
Santiniketan-731 235, West Bengal, India
anumitamndl35@gmail.com, dssm1@rediffmail.com

Abstract

In this study, we construct a reliability acceptance sampling inspection plan to decide whether to accept or reject a lot of products where the One Parameter Polynomial Exponential (OPPE) family of distributions governs the lifetimes. The OPPE distribution has infinite support. To utilise finite support, it has transformed into its unit form, i.e. having the support (0, 1). The design of the plan, Operating characteristic curve, and Sampling procedure are discussed. Determination of the plan parameters using an algorithm is stated. The optimal sample size is determined to protect the consumer's confidence level. Two simplest particular choices of the OPPE family - the exponential and the Lindley are chosen as examples, and optimal plan parameters are tabulated and compared. The plan is executed with three real-life data sets.

Keywords: Consumer's risk, Operating characteristic function, Scale-invariant family of distributions, Truncated life test, unit-Lindley distribution.

1. Introduction

Quality control takes centre stage if one wants multiple copies of a product. The first question that arises throughout the repetitive process is what should the quality features of the product be for it to be satisfactory? The same prototype of products can be seen with the naked eye in the early days. As a result, making items identical is one of the most crucial aspects. Variation in the products is inevitable. In 1924, Walter A. Shewhart introduced many statistical approaches to assess product quality variation.

The sampling inspection plan aims to sentence a lot of products and make decisions to reject or accept that lot. In practice, two types of sampling plans are attribute sampling and variable sampling. In attribute sampling, products of a lot are judged as defective or non-defective, whereas the variable sampling plan measures the product's actual quantitative information. The main advantage of a variable sampling plan is that it requires fewer samples than an attribute plan with the same protection.

Acceptance sampling is traditionally used to decide on acceptance or rejection of the lot, not to determine the quality of the product by using estimation methods. As a result, most acceptance sampling plans are not adequately designed and worrying truth because buying a lot without knowing its quality seems risky. Therefore, to make an appropriate decision, it is better to develop a procedure for evaluating the value of the fraction defective of the lot.

In a sampling inspection plan, if sample observations represent the lifetime of products in the test, one may be interested in the hypothesis that the population average surpasses a specified

average. Suppose the population average represents the average lifetime, defined as μ. If μ_0 is the specified minimum value, then one would like to test the hypothesis $\mu \geq \mu_0$, which means the population average surpasses the specified average. In the test, n samples are placed for testing over a period of time t. The lot cannot be accepted if the observed failures exceed acceptance number (c). This sampling inspection plan is called Reliability Acceptance Sampling Plan (RASP) and it is characterised by the triplet $(n,\ c,\ t)$.

Many authors chalked out the RASP for quality characteristics following different parametric distributions, like exponential [22] , Weibull [9] , Gamma [11], Normal and Log-Normal [10], half Logistics [12], Log-Logistic [17], Birnbaum-Saunders [14] , exponential Frechet [1], three-parameter kappa [2] , generalised inverse exponential [21], generalised Weibull[8], Ishita [3] , transmuted generalised inverse Weibull [4] , generalised Pareto [20] , Quasi Shanker [5], etc.

The popularity of exponential distribution is well known in the context of life testing because of its simplicity in analysis. The constancy properties of failure rate and residual mean life limit the distribution in the present industrial scenario. The Lindley distribution is an excellent alternative to an exponential distribution, but it is overlooked for life-testing purposes. It is a mixture of two parametric distributions, exponential and gamma. The Lindley distribution is more flexible than the exponential distribution because of its increasing, decreasing and upside-down bathtub failure rate for the parameter at different values. [7] has proposed a new and unified approach in generalizing the Lindley's distribution. They investigated some structural properties like moments, skewness, kurtosis, median, mean deviations, Lorenz curve, entropies and limiting distribution of extreme order statistics; reliability properties like reliability function, hazard rate, stress-strength reliability, stochastic ordering; and estimation methods like the method of moment and maximum likelihood. We call the distribution as the one-parameter polynomial exponential (OPPE) family of distributions. The proposed RASP will be constructed assuming only OPPE distribution considering the rejectable quality level. The RASP for the Lindley distribution will be discussed in detail as particular example and compare it with that of the exponential distribution.

Few works have been done on acceptance sampling inspection plans assuming the Lindley distribution (see [15]; [**?**]); [6] and [23] worked on an acceptance sampling plan under a truncated life test assuming two-parameter Lindley distribution. Plan parameters are estimated based on two-point approaches on Operating characteristic (OC) curve-acceptable and rejectable quality levels.

Almost all works on RASP are done for scale-invariant distributions. Minimum sample size n and acceptance number (c) are determined for different times per mean ($\frac{t}{\mu_0}$). However, the OPPE distribution does not belong to a scale-invariant family of distributions. Therefore, the utilisation of time per mean is beyond our scope. We may directly chalk the plan with plan parameters $(n,\ c,\ t)$. Since the OPPE distribution has support (0,∞), we may utilise it with the finite support by transforming into its unit form, i.e. having the support (0,1) with the transformation $V = e^{-T}$. We make tables using the unit-Lindley form by choosing V and the mean μ_v in the interval (0,1). Utilising this benefit, we choose optimal $(n,\ c)$ and then revert to plan parameter t from the relation of the transformation. So, in a nutshell, our objective is to develop a RASP for the OPPE distributed quality characteristic. Based on the time-truncated life test, the plan has the advantage of saving the organisation's time and cost while also being very helpful in determining whether to accept or reject a lot. The OC is derived for choosing the optimal plan based on the consumer's confidence level. Tables of minimum sample sizes are examples for easy understanding and execution of the proposed plan. It is put into practice for real-life experimental data, and the OC surface is depicted to provide a clear picture of the plan.

The following is the arrangement of the rest of the paper. The OPPE and unit-OPPE distributions are described in section 2. In section 3, we describe the sampling design, operating

characteristics function, and operating procedure. In section 4, an algorithm for calculating the minimal sample size of the proposed RASP is stated for the OPPE distribution, and examples for the Lindley and exponential distributions, in particular, are presented in tabular form. In section 5, we use the said sampling plan to work on real-world data. Section 6 concludes.

2. The One Parameter Polynomial Exponential Distribution and its unit version

The probability density function (PDF) of a random variable T of the OPPE distribution can be written as

$$f_T(t,\theta) = h(\theta)p(t)e^{-\theta t},\ t,\ \theta > 0, \tag{1}$$

where, $h(\theta) = \frac{1}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}}$, $p(t) = \sum_{k=0}^{r} a_k t^k$, a_k's are known non-negative constants and r is known non-negative integer.

The distribution can also be written as

$$\begin{aligned} f_T(t,\theta) &= h(\theta)\sum_{k=0}^{r} a_k t^k e^{-\theta t} \\ &= \frac{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}} f_{GA}(t;k+1,\theta)}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}}, \end{aligned} \tag{2}$$

where $f_{GA}(t;k+1,\theta)$ is the PDF of a gamma distribution with shape parameter $(k+1)$ and scale parameter θ. The distribution is a finite mixture of $(r+1)$ gamma distributions.

The cumulative density function (CDF) is given by

$$F_T(t,\theta) = 1 - \left(\frac{\sum_{k=0}^{r} \frac{a_k \Gamma(k+1)\Gamma(k+1,\theta t)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_k \frac{k!}{\theta^{k+1}}} \right),\ t,\ \theta > 0, \tag{3}$$

where $\Gamma(m,t) = \frac{1}{\Gamma(m)} \int_t^{\infty} e^{-u} u^{m-1} du$.
The s-th order raw moment of OPPE is given by

$$\begin{aligned} \mu_s' &= E(T^s) \\ &= \frac{\sum_{k=0}^{r} a_k \frac{\Gamma(k+s+1)}{\theta^{k+s+1}}}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}}. \end{aligned} \tag{4}$$

Now, if we take a transformation $V = e^{-T}$, then the OPPE turns into unit-OPPE in range of (0,1). The PDF and CDF of unit-OPPE is given by ,

$$\begin{aligned} f_V(v,\theta) &= h(\theta)\sum_{k=0}^{r} a_k (-\ln v)^k v^{\theta-1} \\ &= \frac{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}} f_{UGA}(v;k+1,\theta)}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}},\ 0 < v < 1, \end{aligned} \tag{5}$$

where $f_{UGA}(v;k+1,\theta) = \frac{\theta^{k+1}}{\Gamma(k+1)}(-\ln v)^{k+1} v^{\theta-1}$ is the PDF of a unit-gamma distribution with shape parameter $(k+1)$ and scale parameter θ, and

$$F_V(v,\theta) = 1 - \left(\frac{\sum_{k=0}^{r} \frac{a_k \Gamma(k+1)\Gamma(k+1,-\theta \ln v)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_k \frac{k!}{\theta^{k+1}}} \right),\ 0 < v < 1,\ \theta > 0, \tag{6}$$

respectively.
The s-th order raw moment of unit-OPPE is given by

$$\begin{aligned}\mu'_s &= E(V^s) \\ &= \frac{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{(s+\theta)^{k+1}}}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}}.\end{aligned} \tag{7}$$

The Lindley distribution (for $r = 1,\ a_0 = a_1 = 1$), introduced by Lindley (1958) to analyse failure time data has the PDF, CDF and hazard rate function (HRF) as

$$f_T(t;\theta) = \frac{\theta^2}{\theta+1}(1+t)e^{-\theta t} \qquad t > 0, \theta > 0, \tag{8}$$

$$F_T(t;\theta) = 1 - \frac{1+\theta+\theta t}{\theta+1}e^{-\theta t} \tag{9}$$

and

$$h_T(t;\theta) = \frac{\theta^2(1+t)}{1+\theta+\theta t} \qquad t > 0, \theta > 0, \tag{10}$$

respectively.

The mean of the random variable T is

$$\mu = \frac{\theta+2}{\theta(1+\theta)}. \tag{11}$$

The unit-Lindley distribution with parameter θ has the PDF , CDF and HRF respectively , as follows:

$$f(v;\theta) = \frac{\theta^2}{1+\theta}(1 - log(v))\left(v^{\theta-1}\right) \qquad 0 < v < 1, \theta > 0, \tag{12}$$

$$F(v;\theta) = \frac{v^\theta(1+\theta(1-log(v)))}{1+\theta} \qquad 0 < v \leq 1, \theta > 0, \tag{13}$$

$$h(v;\theta) = \frac{\theta^2(1-log(v))}{v(\theta log(v) - (1+\theta)(1-v^{-\theta}))} \qquad 0 < v < 1, \theta > 0. \tag{14}$$

The shapes of the PDF, CDF and HRF of unit-Lindley distribution for different θ are shown in Figure 1. Notably, the HRF is an increasing function. So, the distribution is capable of modeling life time data. Furthermore, the first moment about the origin of unit-Lindley distribution can be obtained as $m'_1 = \frac{\theta^2(2+\theta)}{(1+\theta)^3} = \mu_v$.
For comparison purpose, we will choose the exponential distribution ($r = 0,\ a_0 = 1$) and its corresponding unit version. The unit version of the exponential distribution with parameter θ has the PDF, CDF and HRF as

$$f(v,\theta) = \theta v^{\theta-1}, \qquad 0 < v < 1, \theta > 0, \tag{15}$$

$$F(v,\theta) = v^\theta, \qquad 0 < v \leq 1, \theta > 0, \tag{16}$$

$$h(v,\theta) = \frac{\theta v^{\theta-1}}{1-v^\theta}, \qquad 0 < v < 1, \theta > 0, \tag{17}$$

respectively. In this case, $\mu_v = \frac{\theta}{1+\theta}$ implies $\theta = \frac{\mu_v}{1-\mu_v}$.

3. Reliability Acceptance Sampling Plan for OPPE distribution

According to the product's mean life, a product lot is labelled as good or bad in this sampling plan. The RASP has the plan parameters n, c, and t. For proper implementation of the plan, engineers and practitioners use the tabulated value or algorithm. Tables are presented for fixed t and c, the optimal value of n. Since the value of $t \in (0,\ \infty)$, fixing t is tedious, whereas choosing $v \in (0,\ 1)$ is easy and comprehensive.

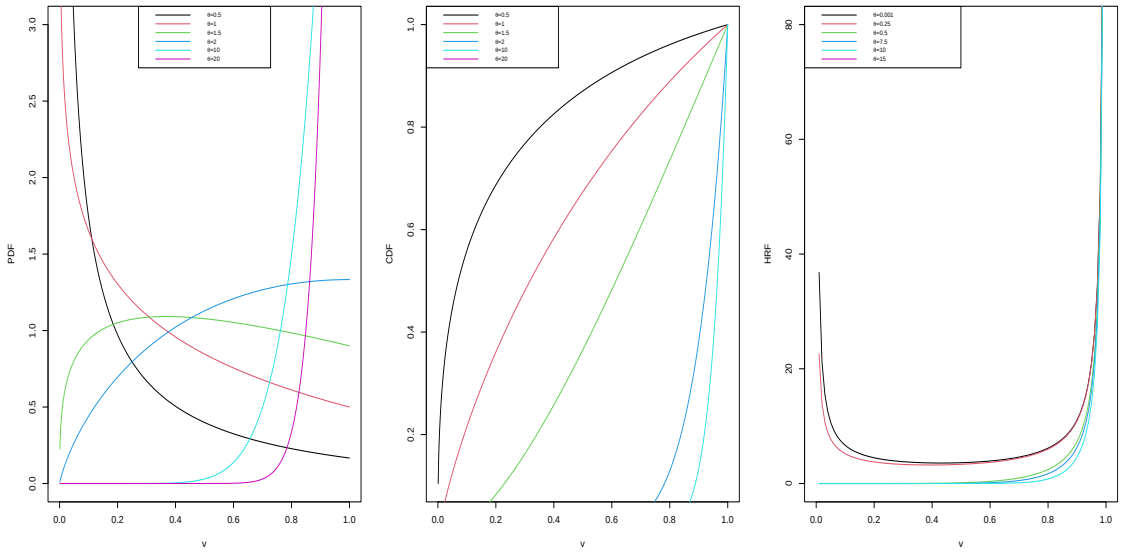


Figure 1: *PDF,CDF and HRF of unit-Lindley distribution for different θ*

3.1. Design of the sampling plan

A product is defective if it fails before truncation time v. The fraction defective i.e. the probability that a product is defective is

$$p(v) = F(v;\theta) = 1 - \left(\frac{\sum_{k=0}^{r} \frac{a_k \Gamma(k+1)\Gamma(k+1,v^\theta)}{\theta^{k+1}}}{\sum_{k=0}^{r} a_k \frac{k!}{\theta^{k+1}}} \right),\ 0 < v < 1,\ \theta > 0. \tag{18}$$

In particular, for the Lindley distribution,

$$p(v) = F(v;\theta) = \frac{v^\theta(1+\theta(1-log(v)))}{1+\theta} \qquad 0 < v < 1, \theta > 0. \tag{19}$$

In this equation, we replace the shape parameter θ by product's mean life (μ). We can say that θ = g(μ), and we get the value of θ by solving the equation by numerical method and hence we have $p(v) = F(v, \mu_v)$.
The Operating Characteristic (OC) function plays a vital role in product control techniques. It gives the probability of acceptance of an individual lot from finite production. For some fixed p, our sampling plan characterized by $(n,\ c,\ t)$ or equivalently by $(n,\ c,\ v)$, where $v = e^{-t}$. For sufficient large lots, the binomial distribution can be applied. The OC function can be formulated as

$$\pi(p) = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} = 1 - B_p(c+1,\ n-c)$$

with $p = F(v;\mu_v)$, and $B_p(c+1,\ n-c)$ is the incomplete beta function.
For determining small positive integer n for given c, v and μ_v^0 , we use

$$\pi(p_0) = \sum_{i=0}^{c} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq 1 - P^*, \tag{20}$$

where, P^* is the level of confidence. If p is very small and n is very large, and $\beta = np$ is finite, then the binomial distribution can be approximated by the Poisson distribution. Then, the OC function becomes

$$\pi(p) = \sum_{i=0}^{c} e^{-\beta} \frac{\beta^i}{i!} = 1 - \Gamma(c+1,\ \beta),$$

with $\Gamma(k,\ w) = \frac{1}{\Gamma(k)} \int_0^w x^{k-1} e^{-x} dx$, the incomplete gamma function.

3.2. Sampling Procedure

The RASP is conducted as follows.

1. Put n items on test.
2. Choose the acceptance number c and specify the maximum test time t.
3. Perform the experiment and count the number of failures.
4. Accept the lot if the number of failures is at most c by the experiment time t.
5. Terminate the experiment as soon as (c+1)th failure occurs and reject the lot.

4. Estimation of the plan parameters

A sampling scheme for unit-OPPE distribution with parameter $(n,\ c,\ v)$ satisfies the consumer's risk given in the previous section. To determine the smallest integer of sample size n for given $(c,\ v)$ is

$$Min_{(n|c,\mu_v^0,P^*)} n$$
subject to

$$\sum_{i=0}^{c} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq 1 - P^*, \tag{21}$$

where $p_0 = 1 - \left(\frac{\sum_{k=0}^{r} \frac{a_k \Gamma(k+1)\Gamma(k+1, v g(\mu_v^0)}{g(\mu_v^0)^{k+1}}}{\sum_{k=0}^{r} a_k \frac{k!}{g(\mu_v^0)^{k+1}}} \right),\ 0 < v < 1,\ g(\mu_v^0) > 0.$

Algorithm for determination of the plan parameters is as follows.

1. Specify the confidence level P^*.
2. Choose t and μ_0.
3. Calculate $v = e^{-t}$ and $\mu_v^0 = \frac{\sum_{k=0}^{r} a_k \frac{\Gamma(k+s+1)}{\theta^{k+s+1}}}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}}$, θ_0 is obtained by solving $\mu_0 = \frac{\sum_{k=0}^{r} a_k \frac{\Gamma(k+s+1)}{\theta^{k+s+1}}}{\sum_{k=0}^{r} a_k \frac{\Gamma(k+1)}{\theta^{k+1}}}$ numerically.
4. Choose a value of c.
5. For given (c,t,μ_v^0), choose minimize n such that (21) satisfies.

The minimum values of n and c satisfying the inequality are obtained and shown in Table 1- 2 for the Lindley and that are in Tables 3-4 for the exponential distribution for $P^* = 0.95, 0.99$. The representative tables are shown for an easy and comprehensive understanding of the proposed plan. The contents of each table are described as follows. The first row is reserved for specifying confidence level (P^*). The first column represents the pre-specified mean value, μ_0^v of unit-Lindley/unit-exponential distribution and the corresponding mean value, μ_0 of Lindley/exponential distribution in the parenthesis. The second row represents the truncation time, v, of unit-Lindley/unit-exponential distribution and the corresponding time, t, of Lindley/exponential distribution in the parenthesis. For a combination of $(\mu_0^v,\ v)$ or $(\mu_0,\ t)$, in the cell, the optimal choices of n, the sample size and c, the acceptance number are presented.

Table 2: *Determination of optimal sample size for Lindley set up*

	$P^* = 0.99$					
$\mu_v^0(\mu_0)$v(t)	$6.74 \times 10^{-3}(5)$		$4.54 \times 10^{-5}(10)$		$3.05 \times 10^{-8}(15)$	
	c	n	c	n	c	n
0.1162(5)	0	3	0	5	0	8
	1	5	1	8	1	12
	2	6	2	10	2	16
0.0447(10)	0	2	0	2	0	3
	1	3	1	4	1	5
	2	4	2	5	2	6
0.0236(15)	0	1	0	2	0	2
	1	3	1	3	1	3
	2	4	2	4	2	5

Table 1: *Determination of optimal sample size for Lindley set up*

	$P^* = 0.95$					
$\mu_v^0(\mu_0)$v(t)	$6.74 \times 10^{-3}(5)$		$4.54 \times 10^{-5}(10)$		$3.05 \times 10^{-8}(15)$	
	c	n	c	n	c	n
0.1162(5)	0	2	0	3	0	6
	1	4	1	6	1	9
	2	5	2	8	2	12
0.0447(10)	0	1	0	2	0	2
	1	3	1	3	1	4
	2	4	2	4	2	5
0.0236(15)	0	1	0	1	0	2
	1	2	1	3	1	3
	2	3	2	4	2	4

Table 3: *Determination of optimal sample size for Exponential set up*

	$P^* = 0.95$					
$\mu_v^0(\mu_0)$v(t)	$6.74 \times 10^{-3}(5)$		$4.54 \times 10^{-5}(10)$		$3.05 \times 10^{-8}(15)$	
	c	n	c	n	c	n
0.1662(5)	0	7	0	21	0	94
	1	11	1	34	1	150
	2	15	2	45	2	197
0.0909(10)	0	4	0	7	0	16
	1	6	1	11	1	25
	2	8	2	15	2	34
0.0625(15)	0	3	0	5	0	8
	1	5	1	8	1	14
	2	7	2	10	2	18

A few observations from the Tables are noted below.
(i) With the confidence level (P^*) increase, the minimum sample size increases.
(ii) The optimal sample size for the Lindley distribution is smaller than that for the exponential distribution.
(iii) The sample size increases with the increase of truncation time (t) or mean (μ) or both.

Table 4: *Determination of optimal sample size for Exponential set up*

	$P^* = 0.99$					
$\mu_v^0(\mu_0)$v(t)	$6.74 \times 10^{-3}(5)$		$4.54 \times 10^{-5}(10)$		$3.05 \times 10^{-8}(15)$	
	c	n	c	n	c	n
0.1662(5)	0	10	0	32	0	72
	1	15	1	46	1	122
	2	20	2	60	2	166
0.0909(10)	0	5	0	11	0	12
	1	8	1	16	1	21
	2	11	2	20	2	27
0.0625(15)	0	4	0	7	0	7
	1	6	1	10	1	11
	2	8	2	13	2	16

5. Real life examples

Data Set 1 : The data is from [16] and arose in test on the cycle at which the yarn failed. The data are the number of cycles until failure of the yarn(100 units):
15, 20, 20, 38, 38, 40, 40, 42, 55, 55, 61, 61, 65, 71, 76, 81, 86, 88, 90, 93, 98, 105, 121, 124, 124, 131, 135, 135, 137, 143, 146, 149, 151, 157, 166, 169, 175, 176, 180, 180, 180, 182, 185, 185, 186, 188, 188, 193, 194, 195, 196, 198, 198, 203, 203, 211, 220, 224, 229, 229, 236, 239, 244, 246, 246, 249, 250, 251, 262, 264, 264, 264, 277, 279, 282, 284, 286, 290, 292, 315, 321, 325, 337, 338, 341, 350, 353, 364, 393, 396, 398, 400, 400, 423, 497, 568, 571, 597, 653, 829.
First, we have checked whether the considered data set is well fitted with the exponential or Lindley distribution by goodness-of-fit test. For this purpose, we have used the Akaike Information Criterion [AIC=-2log(likelihood)+2k, k is the parameter number] to verify which data fit better. The model that best fits the data could be the one with the lowest AIC value.

Table 5: *Comparison of Exponential and Lindley Distribution for Data Set 1*

Distribution	Estimate of θ	Negative Log-likelihood	AIC
Exponential	0.00450	640.2587	1282.517
Lindley	0.0089687	625.6708	1253.3410

Table 5 shows that the Lindley distribution gives a better fit as the AIC value is less than the exponential distribution. The histogram with fitted distributions and P-P plots is shown in Figure 2.
Suppose the truncation time of the testing number of cycles until failure of the yarn is 150, 200, 250 and 300. We construct the decision table for specified mean life as 150, 200, 150, and 300. The estimated average life of the number of cycles is 221.98. The sample size n(=100) is fixed, so Tables 6 and 7 are constructed for (c,v) or (c,t) values, and the decision regarding acceptance or rejection of the lot is made accordingly for the Lindley and the exponential distributions.
Based on the observations, we must make a sentence for the lot, whether it will be accepted or rejected. In this example, let us assume that truncation time, t is 200, that after transformation of t ($v = e^{-t}$), v is 0.0497, and that we take the specified mean of the Lindley distribution to be 150, with the corresponding unit-Lindley mean to be 0.3001. We accept the lot only if the number of failures before the specified mean,150, is less than or equal to acceptance number 60 (see Table 6). In this case, the decision is to accept the lot. We arrive at the same decision for exponential assumption, but the acceptance number, c(=94), is larger than that for the Lindley distribution. So, there may be a chance to come to a wrong decision if Lindley is a better fit, which will be

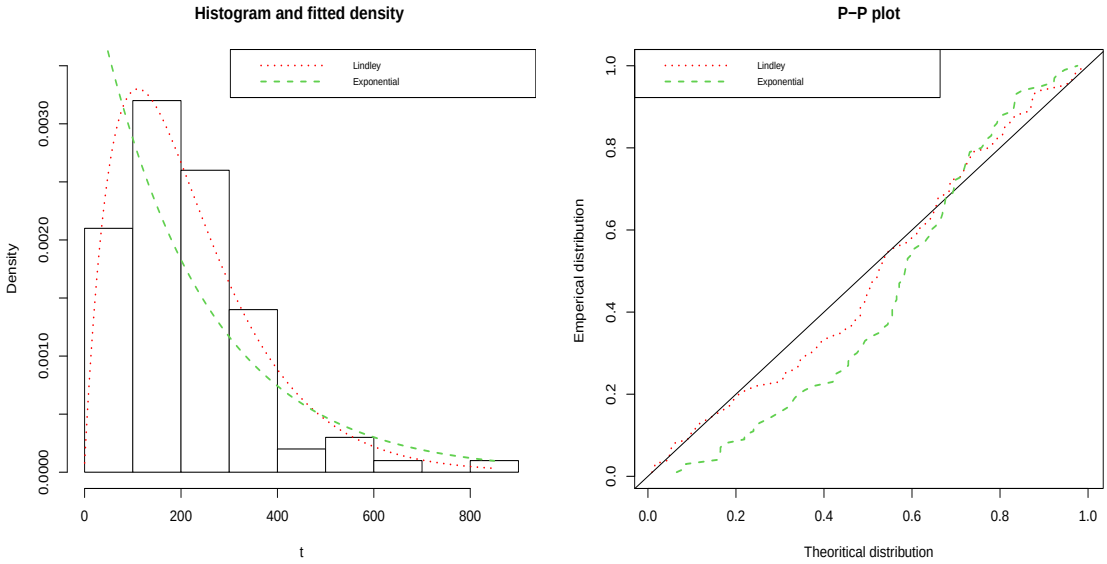


Figure 2: *Histogram with fitted distributions and P-P plot for Data Set 1*

noticed in the case of Data Set 2. The OC surface of the plan for $n = 100$ and $c = 72$ has been shown in Figure 3.

Table 6: *RASP for Lindley distribution for Data Set 1*

	$P^* = 0.95$											
$\mu_v^0(\mu_0)$v(t)	0.1353 (150)			0.0497 (200)			0.0183 (250)			0.0067 (300)		
	c	n	decision	c	n	decision	c	n	decision	c	n	decision
0.3001(150)	72	100	Accept	60	100	Accept	49	100	Accept	39	100	Accept
0.2065(200)	82	100	Accept	74	100	Accept	65	100	Accept	57	100	Accept
0.1515(250)	88	100	Accept	82	100	Accept	76	100	Accept	70	100	Accept
0.1162(300)	91	100	Accept	87	100	Accept	83	100	Accept	78	100	Accept

Table 7: *RASP for Exponential distribution for Data set 1*

	$P^* = 0.95$											
$\mu_v^0(\mu_0)$v(t)	0.1353 (150)			0.0497 (200)			0.0183 (250)			0.0067 (300)		
	c	n	decision	c	n	decision	c	n	decision	c	n	decision
0.0067(150)	96	100	Accept	94	100	Accept	93	100	Accept	93	100	Accept
0.0049(200)	96	100	Accept	95	100	Accept	95	100	Accept	94	100	Accept
0.0039(250)	97	100	Accept	96	100	Accept	95	100	Accept	95	100	Accept
0.0033(300)	97	100	Accept	96	100	Accept	96	100	Accept	95	100	Accept

Data Set 2:
This data represents the failure times in minutes for a sample of 15 electronic component in accelerated life test (see, [13]), also used by [19] and [18] :
1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2, 23, 30.6, 37.3, 46.3, 53.9, 59.8, 66.2.

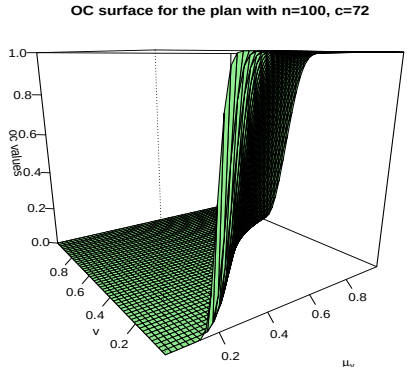


Figure 3: *OC Surface of the proposed plan for Data Set 1*

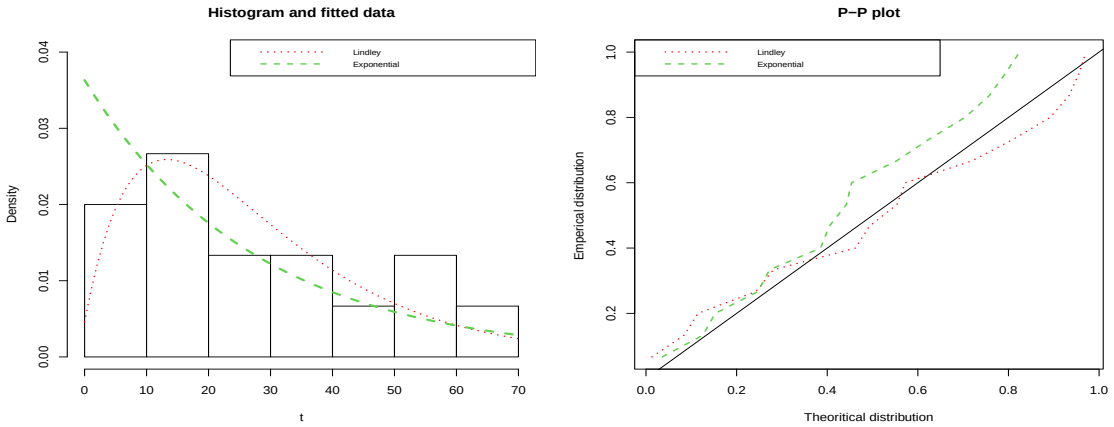


Figure 4: *Histogram with fitted distributions and P-P plot for Data Set 2*

Table 8: *Comparison of Exponential and Lindley Distributions for Data Set 2:*

Distribution	Estimate of θ	Negative Log-likelihood	AIC
Exponential	0.03630	64.7386	131.4764
Lindley	0.07025	64.40554	130.8110

Table 8 shows that the Lindley distribution gives a better fit as the AIC value is less than that for the exponential distribution. The histogram with fitted distributions and the P-P plot shown in Figure 4 substantiates the claim.
The estimated mean life is 27.54 for the fitted Lindley distribution. We have selected the testing time for failure times as 20, 30, 35 and 40. The corresponding hypothesis for testing is $\mu \geq \mu_0$ against $\mu < \mu_0$, where μ_0 is the specified mean, and μ is the average lifetime. We construct Tables 9 and 10 with sample size n =15 for the Lindley and exponential distributions, respectively.
For example, let the truncation time is 30, and the specified mean is 35. The decision regarding the lot is to reject under the Lindley assumption and accept under the exponential assumption. Hence the exponential assumption leads to a wrong conclusion as the data fits better with the Lindley distribution. Figure 5 shows the OC surface of the plan for $n = 15$ and $c = 9$.

Table 9: *RASP for the Lindley distribution for Data set 2*

	$P^* = 0.95$											
$\mu_v^0(\mu_0)$v(t)	0.3679 (20)			0.1353 (30)			0.0820 (35)			0.0498 (40)		
	c	n	decision	c	n	decision	c	n	decision	c	n	decision
0.4853(20)	9	15	Accept	5	15	Accept	4	15	Reject	3	15	Reject
0.3002(30)	11	15	Accept	8	15	Reject	7	15	Reject	6	15	Reject
0.2465(35)	11	15	Accept	9	15	Reject	8	15	Reject	8	15	Reject
0.2065(40)	12	15	Accept	10	15	Reject	9	15	Reject	9	15	Reject

Table 10: *RASP for Exponential distribution for Data set 2*

	$P^* = 0.95$											
$\mu_v^0(\mu_0)$v(t)	0.3679 (20)			0.1353 (30)			0.0820 (35)			0.0498 (40)		
	c	n	decision	c	n	decision	c	n	decision	c	n	decision
0.0476(20)	12	15	Accept	11	15	Accept	10	15	Accept	10	15	Accept
0.0323(30)	12	15	Accept	11	15	Accept	11	15	Accept	11	15	Accept
0.0278(35)	12	15	Accept	12	15	Accept	11	15	Accept	11	15	Accept
0.0244(40)	12	15	Accept	12	15	Accept	11	15	Accept	11	15	Accept

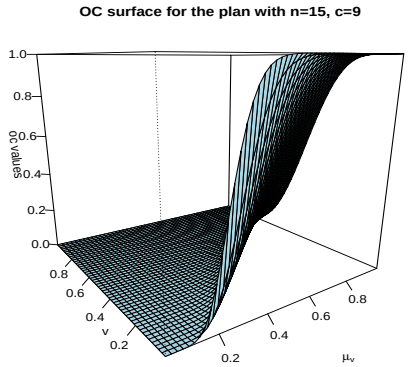


Figure 5: *OC Surface of the proposed plan for Data Set 2*

Data set 3: This data set shows that the cycle-to-failure numbers for 25 (100-cm) specimens of yarn tested at a particular strain level (see, [13]) are:
15, 20, 38, 42, 61, 76, 86, 98, 121, 146, 149, 157, 175, 176, 180, 180, 198, 220, 224, 251, 264, 282, 321, 325, 653.
Table 11 shows that the OPPE with $a_0 = 9, a_1 = 4, a_2 = 0.005$ distribution is a better fit as the AIC value is less than that for the exponential distribution. The histogram with fitted distributions and the P-P plot in Figure 6 justifies the claim. Table 12 shows different plans and their decisions for the fitted OPPE model. Table 13 shows that for the exponential model. The OPPE assumption shows its superiority. The OC surface of the plan for $n = 25$ and $c = 18$ with the fitted OPPE is at Figure 7.

Table 11: *Comparison of Exponential and OPPE(9,4,0.005) for Data Set 3*

Distribution	Estimate of θ	Negative Log-likelihood	AIC
Exponential	0.0056	154.5078	311.179
OPPE(9,4,0.005)	0.01115	152.5078	307.0156

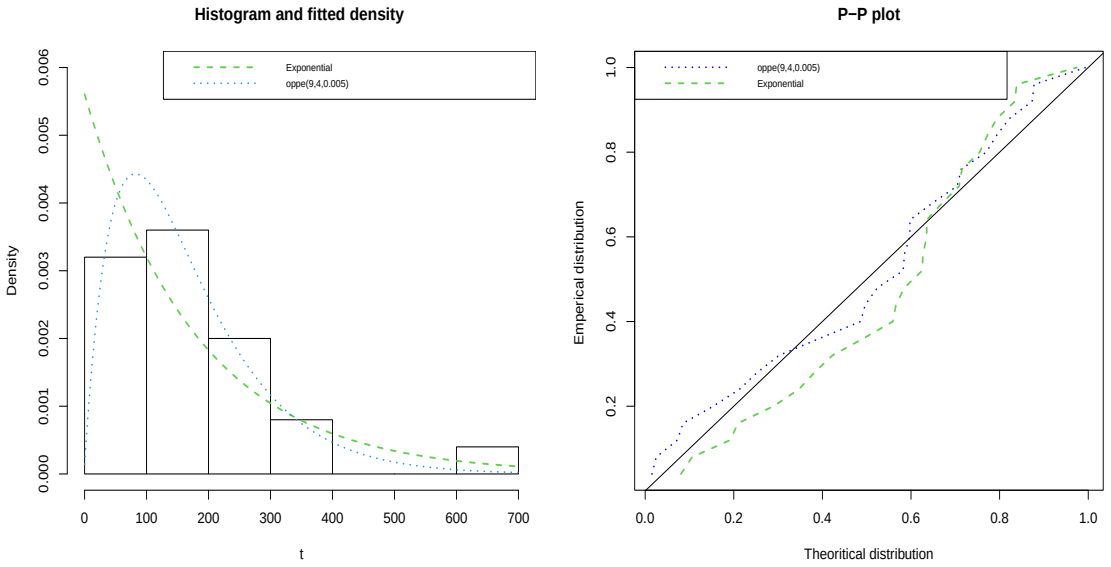


Figure 6: *Histogram with fitted distributions and P-P plot for Data Set 3*

Table 12: *RASP for the OPPE(9,4,0.005) distribution for Data set 3*

	$P^* = 0.95$											
$\mu_v^0(\mu_0)$v(t)	0.4065 (100)			0.2466 (150)			0.1469 (200)			0.0907 (250)		
	c	n	decision	c	n	decision	c	n	decision	c	n	decision
0.5208(100)	18	25	Accept	11	25	Accept	8	25	Accept	6	25	Reject
0.4054(150)	16	25	Accept	14	25	Accept	11	25	Accept	9	25	Reject
0.3282(200)	18	25	Accept	16	25	Reject	13	25	Reject	12	25	Reject
0.2730(250)	19	25	Accept	17	25	Reject	15	25	Reject	14	25	Reject

Table 13: *RASP for the Exponential distribution for Data set 3*

	$P^* = 0.95$											
$\mu_v^0(\mu_0)$v(t)	0.4065 (100)			0.2466 (150)			0.1469 (200)			0.0907 (250)		
	c	n	decision	c	n	decision	c	n	decision	c	n	decision
0.0099(100)	23	25	Accept	23	25	Accept	22	25	Accept	22	25	Accept
0.0066(150)	23	25	Accept	23	25	Accept	23	25	Accept	22	25	Accept
0.0049(200)	23	25	Accept	23	25	Reject	23	25	Accept	23	25	Accept
0.0039(250)	23	25	Accept	23	25	Accept	23	25	Accept	23	25	Accept

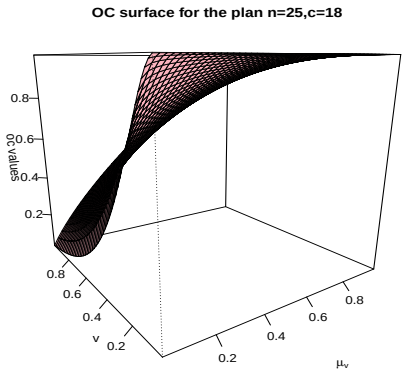


Figure 7: *OC Surface of the proposed plan for Data Set 3*

6. Concluding remarks

A reliability acceptance sampling plan is formulated for the OPPE distributed quality characteristic. The OPPE family of distributions does not belong to the scale-invariant family, whereas most of the RASP chalked out for the scale-invariant family in the literature. The optimal plan parameters are estimated by transforming the OPPE distribution into its unit form to utilize the advantage of finite range (in this case, (0,1)). A few examples are presented for finding optimal sample sizes for the proposed plan for the Lindley distribution, a particular choice of the OPPE family, which will be helpful to scientists and quality practitioners for implementation. Three data sets are analyzed for implementing the proposed plan. The approach may be adopted to construct RASP for other lifetime quality characteristic distributions that do not belong to the scale-invariant family.

Acknowledgement
The authors thank the anonymous referee for insightful comments that help improve the paper.

Declarations

Disclosure of Conflicts of interest/competing interests: The authors declare that they have no conflict of interest.

Authors contributions: Each author has equal contribution. All authors jointly write, review and edit the manuscript.

Funding: The authors received no specific funding for this study.

Data Availability Statements: All cited data analysed in the article are included in References. Data sets are also provided in the article.

Ethical Approval: This article does not contain any studies with human participants performed by any of the authors.

Code availability: Codes are available on request.

References

[1] Al-Nasser, A. D. and Al-Omari, A. I. (2013): Acceptance sampling plan based on truncated life tests for exponentiated Frechet distribution, *Journal of Statistics and Management*, 16(1), 13-24.

[2] Al-Omari, A. I. (2014): Acceptance Sampling Plan Based on Truncated Life Tests for Three Parameter Kappa Distribution, *Economic Quality Control*, 29, 53-62.

[3] Al-Nasser, A. D., Al-Omari, A. I., Bani-Mustafa, A. and Jaber, K. (2018): Developing Single-Acceptance Sampling Plans Based On A Truncated Lifetime Test For An Ishita Distribution, *Statistics in Transition New Series*, 19, 393-406.

[4] Al-Omari, A. I. (2018): The transmuted generalized inverse Weibull distribution in acceptance sampling plans based on life tests, *Transactions of the Institute of Measurement and Control*, 40(16), 4432-4443.

[5] Al-Omari, A. I., Almanjahie, I. M., and Dar, J. G. (2021): Acceptance sampling plans under two-parameter Quasi Shanker distribution assuming mean life with an application to manufacturing data, *Science Progress* 104(2), 1-17.

[6] Al-Omari, A. I. and Al-Nasser, A. D. (2019): A Two-Parameter Quasi Lindley Distribution in Acceptance Sampling Plans from Truncated Life Tests, *Pakistan Journal of Statistics and Operation Research*, 15(1), 39-47

[7] Bouchahed, L. and Zeghdoudi, H. (2018): A new and unified approach in generalizing the Lindley's distribution with applications, *Statistics in Transition*, 19(1), 61-74.

[8] Chowdhury, S. (2016): Acceptance sampling plans based on truncated life test for the generalized Weibull model, *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 886-889.

[9] Goode, H. P. and Kao, J. H. K. (1961): Sampling plans based on the Weibull distribution, *Proceedings of Seventh National Symposium on Reliability and Quality Control*, Philadelphia, Pennsylvania, 24-40.

[10] Gupta, S. (1962): Life Test sampling plans for Normal and Log-Normal Distribution, *Technometrics* , 4, 151-175.

[11] Gupta, S. S. and Groll, P. A. (1961): Gamma distribution in acceptance sampling based on life tests, *Journal of the American Statistical Association*, 56, 942-970.

[12] Kantam, R.R. L. and Rosaiah, K. (1998): Half logistic distribution in acceptance sampling based on life tests, *IAPQR Transactions*, 23, 117-125.

[13] Lawless,J. F., Statistical Models and Methods for Lifetime Data, John Wiley Sons, New York, 2003.

[14] Lio, Y. L., Tsai, T. R., and Wu, S. J. (2010): Acceptance sampling plans from truncated life tests based on the Birnbaum-Saunders distribution for percentiles, *Communications in Statistics-Simulation and Computation*, 39, 119-136.

[15] Mukherjee, S. and Maiti, S. S. (2014): Sampling Inspection Plan by variable for Lindley distributed quality characteristic, *Proceedings of IMBIC - MSAST* , 3, 213-223.

[16] Picciotto R. Tensile fatigue characteristics of a sized polyester viscose yarn and their effect on weaving performance. North Carolina State, University at Raleigh, USA, 1970.

[17] Rosaiah, K., Kantam, R.R. L. and Kumar,S. Ch. (2006): Reliability Test plans for Exponentiated Log-Logistic Distribution, *Economic Quality Control*, 21(2), 279-289.

[18] Saha, M., Tripathi, H., Dey, S., and Maiti, S. S. (2021): Acceptance sampling inspection plan for the Lindley and power Lindley distributed quality characteristics, *International Journal of System Assurance Engineering and Management*, 12, 1410-1419.

[19] Shanker, R. and Shukla, K. K (2016): On modelling of lifetime data using three-parameter generalized Lindley and generalized gamma distributions, *Biometrics* & *Biostatistics International Journal*, 4, 283-288.

[20] Singh, N., Sood, A., and Buttar, G. S. (2020): Acceptance Sampling Plan for Truncated Life Tests Based on Generalized Pareto Distribution using Mean Life, textitIndustrial Engineering and Management Systems, 19, 694-703.

[21] Singh, S.,Tripathi,Y. M, and Jun, C. H. (2015): Sampling Plans Based on Truncated Life Test for a Generalized Inverted Exponential Distribution, *Industrial Engineering* & *Management Systems*, 14(2), 183-195.

[22] Sobel, M and Tischendorf,J. A. (1959): Acceptance sampling with new life test objectives, *Proceedings of fifth National Symposium on Reliability and Quality Control*, Philadelphia Pennsylvania, 108-118.

[23] Wu, C. W., Shu, M. H., and Wu, N. Y. (2020): Acceptance sampling schemes for two-parameter Lindley lifetime products under a truncated life test, *Quality Technology and Quantitative Management*, 18, 382-395.

RT&A, No.3 (74)
Volume 18, September 2023

TWO-CLASSES FOR REGRESSION TYPE OF ESTIMATORS FOR THE RATIO OF TWO POPULATION MEANS IN TWO-PHASE SAMPLING IN THE PRESENCE OF NON-RESPONSE FOR STRATIFIED POPULATION

Manish Mishra, B. B. Khare and Sachin Singh

•

Department of Statistics, Banaras Hindu University, Varanasi-221005, India.
manish.mishra10@bhu.ac.in, bbkhare56@yahoo.com and singhat619@gmail.com

Abstract

Utilizing the auxiliary information in stratified population, in the current study, we have discussed two classes for the regression type of estimators to estimate the ratio of two population means in the presence of non-response with the unknown population mean of the auxiliary variable. To estimate the unknown value of the population mean of auxiliary variable, we have used two-phase sampling method. For the suggested classes of estimators, we have considered two situations for the use of auxiliary information along with the non-response in the study variable such as incomplete information on the study variable and incomplete information on the corresponding units of the auxiliary variable and in another situation we have considered incomplete information on the study variable and complete information on the auxiliary variable. To estimate the non-response in study variable and auxiliary variables, we have used the Hansen and Hurwitz method of sub-sampling from the non-respondents. For the suggested classes of estimators, some members have been recognized. Using large sample approximation, the expressions for bias and mean square error have been derived for the suggested classes. The optimum values of the constants involving in the expression of mean square error have also been calculated. Mean square errors of the Suggested classes are found to be equal in theoretical study and real data study. An empirical study has been conducted with the help of a real data set (The Primary Census Abstract-2011 published by the Office of the Registrar General & Census Commissioner, India.) in order to compare the proposed classes of estimators with the conventional estimator for the different rates of non-response and different choices of sub-sampling fraction. The Suggested classes are found to be most efficient with respect to the conventional estimator for the different rates of non-response and different choices of sub-sampling fraction in empirical study.

Keywords: Ratio of two population means, Regression type estimator, Two-phase sampling, Auxiliary variable, Non-response, Mean square error

1. Introduction

In the literature of sample surveys, the ratio of two population means plays a crucial role. In this context to have a better understanding, we have several examples in the field of scientific and Socio-economic studies such as:

- **Agricultural surveys**: The crop production per acre in a crop survey, agriculture labor per cultivator and the ratio of production of corn acres to wheat acres, etc.
- **Industrial surveys**: the outlay of total expenses per employee, Proportion of liquid to total asset, etc.

- **Medical surveys**: In the estimation of growth index, the estimate of ratio using the measurements on weight to height. The skull or chest circumference may be used as an auxiliary variable.

In the case of finite population, the estimate for the ratio of two population means by using known and unknown population mean of the auxiliary variable have been studied by several authors such as Singh [13], Tripathi [17], Khare [3], Upadhyaya et al. [18], Singh and Naqvi [14]and Kumar and Srivastava [10]. In a recent study, Ahuja et al.[16] have suggested a generalized two-phase sampling estimator for ratio of two population means.

The occurrence of non-response is very common in the field of sample surveys. Hansen and Hurwitz [2] have suggested a technique of sub-sampling from the non-respondents to treat the problem of non-response. Further, El-Badry [1] has made some improvements to reduce the effect of non-response. In case of finite population in the presence of non-response, the estimation of ratio of two population means using known and unknown population mean of auxiliary variable have been studied by Khare and Pandey [4], Khare and Sinha [[5],[6]], Khare and Sinha [7] and Khare et al. [8].

For the stratified population in the presence of non-response, Khare and Jha [9] and Singh et al. [12] have suggested the classes of estimators for estimating the population mean utilizing auxiliary information with known and unknown population mean of auxiliary variable.

Following the research work of Singh et al. [12], we have made an effort by suggesting two classes of regression type of estimators for the ratio of two population means utilizing two-phase sampling method for the estimation of unknown population mean of auxiliary variable in the presence of non-response for the stratified population. Some members of the suggested classes of estimators have been recognized. The properties of the suggested classes have been obtained. To support the effectiveness of the proposed classes of estimators with respect to the relevant estimator, an empirical study is conducted with the help of a real data set.

2. Notations and Sampling Procedure

We have a heterogeneous population of size $\eta : [\eta_1, \eta_2, \eta_3 ... \eta_N]$ study variables (y_1, y_2) and auxiliary variable x with respective population means $\overline{Y}_1, \overline{Y}_2$ and population mean $(\overline{X})$ of auxiliary variable is unknown, which is divided into L homogeneous strata. The population parameters used in this study are denoted as follows:

we have,

$$\overline{Y}_1 = \frac{1}{N}\sum_{i=1}^{N} y_{1i}, \overline{Y}_2 = \frac{1}{N}\sum_{i=1}^{N} y_{2i}, \overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ and } R = \frac{\overline{Y}_1}{\overline{Y}_2} \tag{1}$$

In this present study, we are dealing with the stratified population with unknown $\overline{X}$ in the presence of non-response. In such a situation, to estimate the unknown population mean of an auxiliary variable, we use the technique of two-phase sampling which is described as follows:

In the very first step, we have the population of size N divided in L homogeneous strata of sizes $N_1, N_2, N_3 ... N_L$. In the first-phase, using simple random sampling without replacement (SRSWOR), we draw a larger preliminary sample of size $n_i^{'}$ from i^{th} stratum of size N_i.

Further, in the second-phase, we draw a relatively small sample of size $n_i^{'}(n_i < n_i^{'})$ using SRSWOR from $n_i^{'}$ units of i^{th} stratum.

For the study (y_1, y_2)variables and auxiliary variable (x), a sample of size n_i, we draw from i^{th} stratum of size N_i. Due to the non-response in the population, we observe that in the sample of size n_i there are n_{i1} responding and n_{i2} non-responding units such that $[n_i = n_{i1} + n_{i2}]$.

In the next step, to get the estimate of these n_{i2}non-responding units, we draw a sub-sample of size $r_i [= \frac{n_{i2}}{l_i}, l_i > 1]$.

Hence, we have responding n_{i1} and r_i sub-sampled units for the i^{th} stratum which we use in defining the Hansen and Hurwitz [2] estimators. The population means of the i^{th} stratum for

(y_1, y_2) and x are given by $\overline{Y}_{1i}$,$\overline{Y}_{2i}$ and $\overline{X}_i$ and their estimators in the presence of non-response using Hansen and Hurwitz [2] sub-sampling method from non-respondents are given as follows:

$$\overline{y}^*_{1i} = \frac{n_{i1}}{n_i}\overline{y}_{1i(1)} + \frac{n_{i2}}{n_i}\overline{y}'_{1i(2)}, \overline{y}^*_{2i} = \frac{n_{i1}}{n_i}\overline{y}_{2i(1)} + \frac{n_{i2}}{n_i}\overline{y}'_{2i(2)} \text{ and } \overline{x}^*_i = \frac{n_{i1}}{n_i}\overline{x}_{i(1)} + \frac{n_{i2}}{n_i}\overline{x}'_{i(2)}. \quad (2)$$

where, $\overline{y}_{1i(1)}$, $\overline{y}_{2i(1)}$ and $\overline{x}_{i(1)}$ are the sample means for the variable (y_1, y_2) and x for the n_{i1} units in i^{th} stratum. The sample means based on r_i units sub-sampled from n_{i2} units in the i^{th} stratum are denoted by $\overline{y}'_{1i(2)}$, $\overline{y}'_{2i(2)}$ and $\overline{x}'_{i(2)}$ for study variables (y_1, y_2) and the auxiliary variable x.

In the presence of non-response, the stratified sample means for $\overline{Y}_1$,$\overline{Y}_2$ and $\overline{X}$ for the i^{th} stratum are given as follows:

$$\overline{y}^*_{1st} = \sum_{i=1}^{L} W_i\overline{y}^*_{1i}, \overline{y}^*_{2st} = \sum_{i=1}^{L} W_i\overline{y}^*_{2i} \text{ and } \overline{x}^*_{st} = \sum_{i=1}^{L} W_i\overline{x}^*_i. \quad (3)$$

For the auxiliary variable x,the stratified sample mean for estimating $\overline{X}$ based on first-phase sample of size n'_i is given as follows:

$$\overline{x}'_{st} = \sum_{i=1}^{L} W_i\overline{x}'_i. \quad (4)$$

where,$\overline{x}'_i$ is the sample mean based on first-phase sample of size n'_i,$(n'_i > n_i)$ drawn from N_i units of the i^{th} stratum is given as follows:

$$\overline{x}'_i = \frac{1}{n'_i}\sum_{i=1}^{n'_i} x_{ij} \quad (5)$$

.

The stratified sample mean based on n_i units in i^{th} stratum for auxiliary variable x is given as follows:

$$\overline{x}_{st} = \sum_{i=1}^{L} W_i\overline{x}_i. \quad (6)$$

.

The population variance and co-variance for i^{th} stratum and non-responding part of the i^{th} stratum used in this study are given as follows:

$$S^2_{y1i} = \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{1i,j} - \overline{Y}_{1i})^2,$$

$$S^2_{y2i} = \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{2i,j} - \overline{Y}_{2i})^2,$$

$$S^2_{xi} = \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(X_{ij} - \overline{X}_{1i})^2,$$

$$S^2_{y1i(2)} = \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{1ij(2)} - \overline{Y}_{1i(2)})^2,$$

$$S^2_{y2i(2)} = \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{2ij(2)} - \overline{Y}_{2i(2)})^2,$$

$$S^2_{xi(2)} = \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(X_{ij(2)} - \overline{X}_{i(2)})^2,$$

$$S^2_{y1iy2i} = \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{1i,j}-\overline{Y}_{1i})(Y_{2i,j}-\overline{Y}_{2i}),$$

$$S^2_{y1ixi} = \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{1i,j}-\overline{Y}_{1i})(X_{i,j}-\overline{X}_{i}),$$

$$S^2_{y2ixi} = \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{2i,j}-\overline{Y}_{2i})(X_{i,j}-\overline{X}_{i}),$$

$$S^2_{y1iy2i(2)} = \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{1i,j(2)}-\overline{Y}_{1i(2)})(Y_{2i,j(2)}-\overline{Y}_{2i(2)}),$$

$$S^2_{y1ixi(2)} = \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{1i,j(2)}-\overline{Y}_{1i(2)})(X_{i,j(2)}-\overline{X}_{i(2)}),$$

$$S^2_{y2ixi(2)} = \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{2i,j(2)}-\overline{Y}_{2i(2)})(X_{i,j(2)}-\overline{X}_{i(2)}). \tag{7}$$

where, $y_{1i,j}$: j^{th} value of y_1 in the i^{th} stratum, $y_{2i,j}$: j^{th} value of y_2 in the i^{th} stratum, $x_{i,j}$: j^{th} value of x in the i^{th} stratum, $y_{1i,j(2)}$: j^{th} value of y_1 for non-responding units in the i^{th} stratum, $y_{2i,j(2)}$: j^{th} value of y_2 for non-responding units in the i^{th} stratum and $x_{i,j(2)}$: j^{th} value of x for non-responding units in the i^{th} stratum.

We denote,$W_{i1} = \frac{N_{i1}}{N_i}$,$W_{i2} = \frac{N_{i2}}{N_i}$ such that $[N_i = N_{i1} + N_{i2}]\ \forall,\ i = 1,2,3...L$. Where, N_{i1} and N_{i2} are the size of the responding and non-responding units in i^{th} stratum.

3. Proposed Classes of Regression Type of Estimators for R in Two-phase Sampling

In the case of unknown population mean of the auxiliary variable, we are considering two different situations of non-response i.e. incomplete information on y_1, y_2 and corresponding information on the auxiliary variable x and also we use complete information on auxiliary variable x. Here, we propose two classes of two-phase sampling regression type of estimators which are given as follows:

$$\hat{R}_{1st} = [\hat{R}_{st} + \alpha_1(\overline{x}'_{st} - \overline{x}'_{st})]\phi_{(1)}(u_1) \tag{8}$$

$$\hat{R}_{2st} = [\hat{R}_{st} + \alpha_2(\overline{x}'_{st} - \overline{x}'_{st})]\phi_{(1)}(u_2) \tag{9}$$

where, $\hat{R}_{st} = \frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}$, $u_1 = \frac{\overline{x}^*_{st}}{\overline{x}'_{st}}$ and $u_2 = \frac{\overline{x}_{st}}{\overline{x}'_{st}}$.

Such that,

$$\phi_1(1) = 1, \phi_1(2) = 1, \phi_{1(1)}(1) = \left(\frac{\delta}{\delta u_1}\phi_1(1)\right)_1 \text{ and } \phi_{1(2)}(1) = \left(\frac{\delta}{\delta u_2}\phi_2(1)\right)_1 \tag{10}$$

$\phi_{(1)}(u_1)$ and $\phi_{(2)}(u_2)$ are the function of u_1 and u_2 satisfy the regularity conditions given as follows:

- The functions of u_1 and u_1 assume values in a bounded closed convex subset U^* of the two-dimensional real line containing the point (1).
- In U^*, the function of u_1 and u_2 are continuous and bounded.
- The first-order and second-order partial derivatives of the given function of u_1 and u_1 exist and are continuous and bounded U^*.

The first and second-order partial derivatives of the function $\phi_{(1)}(u_1)$ w.r.to u_1 and $\phi_{(2)}(u_2)$ w.r.to u_2 are denoted by $[\phi_{1(1)}(u_1), \phi_{11(1)}(u_1)]$ and $[\phi_{1(2)}(u_2), \phi_{11(2)}(u_2)]$ respectively.

Now, using Taylor's series expansion we expand the function $\phi_{(1)}(u_1)$ and $\phi_{(2)}(u_2)$ upto the second-order partial derivative about the point (1), we have

$$\hat{R}_{1st} = [\hat{R}_{st} + \alpha_1 \overline{x}'_{st}(u_2 - 1)][\phi_{(1)}(1) + (u_1 - 1)\phi_{1(1)}(1) + \frac{1}{2}(u_1 - 1)^2\phi_{11(1)}(1)] \tag{11}$$

$$\hat{R}_{2st} = [\hat{R}_{st} + \alpha_2 \overline{x}'_{st}(u_1 - 1)][\phi_{(2)}(1) + (u_2 - 1)\phi_{1(2)}(1) + \frac{1}{2}(u_2 - 1)^2\phi_{11(2)}(1)] \tag{12}$$

Using the condition given in equation (10) and regularity conditions, the expression (11) and (12) can be written as:

$$\begin{aligned}\hat{R}_{1st} &= [\hat{R}_{st} + \alpha_1 \overline{x}'_{st}(u_2 - 1)][1 + (u_1 - 1)\phi_{1(1)}(1) + \frac{1}{2}(u_1 - 1)^2\phi_{11(1)}(1)] \\ &= \hat{R}_{st} + \hat{R}_{st}(u_1 - 1)\phi_{1(1)}(1) + \hat{R}_{st}\frac{1}{2}(u_1 - 1)^2\phi_{11(1)}(1) + \alpha_1 \overline{x}'_{st}(u_2 - 1) \\ &+ \alpha_1 \overline{x}'_{st}(u_1 - 1)(u_2 - 1)\phi_{1(1)}(1) + \frac{1}{2}\alpha_1 \overline{x}'_{st}(u_1 - 1)^2(u_2 - 1)\phi_{11(1)}(1)\end{aligned} \tag{13}$$

$$\begin{aligned}\hat{R}_{2st} &= [\hat{R}_{st} + \alpha_2 \overline{x}'_{st}(u_1 - 1)][1 + (u_2 - 1)\phi_{1(2)}(1) + \frac{1}{2}(u_2 - 1)^2\phi_{11(1)}(2)] \\ &= \hat{R}_{st} + \hat{R}_{st}(u_2 - 1)\phi_{1(2)}(1) + \hat{R}_{st}\frac{1}{2}(u_2 - 1)^2\phi_{11(2)}(1) + \alpha_2 \overline{x}'_{st}(u_1 - 1) \\ &+ \alpha_2 \overline{x}'_{st}(u_1 - 1)(u_2 - 1)\phi_{1(2)}(1) + \frac{1}{2}\alpha_2 \overline{x}'_{st}(u_2 - 1)^2(u_1 - 1)\phi_{11(2)}(1)\end{aligned} \tag{14}$$

4. Properties of the proposed classes of regression type of estimators

To obtain the expression for bias and MSE of the suggested classes of estimators, we define:

$$\begin{aligned}\overline{y}^*_{1st} &= \overline{Y}_1(1 + \epsilon_0) \quad , \overline{y}^*_{2st} = \overline{Y}_2(1 + \epsilon_1),\ \overline{x}_{st} = \overline{X}(1 + \epsilon_2),\ \overline{x}'_{st} = \overline{X}(1 + \epsilon_3) \\ \overline{x}^*_{st} &= \overline{X}_1(1 + \epsilon_4)\end{aligned} \tag{15}$$

such that, $|\epsilon_i| < 1$ and $E(\epsilon_i) = 0\ \forall\ i = 1, 2, 3, 4$.

Here, we ignore the finite population correction term because the population size is large enough under consideration. i.e. by using large sample approximation, we have

$$\begin{aligned}E(\epsilon_0^2) &= \frac{V(\overline{y}^*_{1st})}{\overline{Y}_1^2} = \frac{1}{\overline{Y}_1^2}\sum_{i=1}^{L}[W_i^2\lambda_i S^2_{y_{1i}} + \frac{(l_i - 1)}{n_i}W_{i2}S^2_{y_{1i}(2)}], \\ E(\epsilon_1^2) &= \frac{V(\overline{y}^*_{2st})}{\overline{Y}_2^2} = \frac{1}{\overline{Y}_2^2}\sum_{i=1}^{L}[W_i^2\lambda_i S^2_{y_{2i}} + \frac{(l_i - 1)}{n_i}W_{i2}S^2_{y_{2i}(2)}], \\ E(\epsilon_2^2) = E(\epsilon_2\epsilon_4) &= \frac{V(\overline{x}_{st})}{\overline{X}^2} = \frac{1}{\overline{X}^2}\sum_{i=1}^{L}[W_i^2\lambda_i S^2_{x_i}], \\ E(\epsilon_3^2) = E(\epsilon_2\epsilon_3) = E(\epsilon_3\epsilon_4) &= \frac{V(\overline{x}'_{st})}{\overline{X}^2} = \frac{1}{\overline{X}^2}\sum_{i=1}^{L}[W_i^2\lambda'_i S^2_{x_i}], \\ E(\epsilon_4^2) &= \frac{V(\overline{x}^*_{st})}{\overline{X}^2} = \frac{1}{\overline{X}^2}\sum_{i=1}^{L}\left\{W_i^2\lambda_i S^2_{x_i} + \frac{(l_i - 1)}{n_i}W_{i2}S^2_{x_i(2)}\right\},\end{aligned} \tag{16}$$

$$E(\epsilon_0\epsilon_1) = \frac{Cov(\overline{y}^*_{1st}\overline{y}^*_{1st})}{\overline{Y}_1\overline{Y}_2} = \frac{1}{\overline{Y}_1\overline{Y}_2}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{1i}y_{2i}} + \frac{(l_i-1)}{n_i}W_{i2}S_{y_{1i}y_{2i}(2)}\right\},$$

$$E(\epsilon_0\epsilon_4) = \frac{Cov(\overline{y}^*_{1st}\overline{x}^*_{st})}{\overline{Y}_1\overline{X}} = \frac{1}{\overline{Y}_1\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{1i}x_i} + \frac{(l_i-1)}{n_i}W_{i2}S_{y_{1i}x_i(2)}\right\},$$

$$E(\epsilon_1\epsilon_4) = \frac{Cov(\overline{y}^*_{2st}\overline{x}^*_{st})}{\overline{Y}_2\overline{X}} = \frac{1}{\overline{Y}_2\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{2i}x_i} + \frac{(l_i-1)}{n_i}W_{i2}S_{y_{2i}x_i(2)}\right\},$$

$$E(\epsilon_0\epsilon_2) = \frac{Cov(\overline{y}^*_{1st}\overline{x}_{st})}{\overline{Y}_1\overline{X}} = \frac{1}{\overline{Y}_1\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_i x_i}\right\},$$

$$E(\epsilon_0\epsilon_3) = \frac{Cov(\overline{y}^*_{1st}\overline{x}^{'}_{st})}{\overline{Y}_1\overline{X}} = \frac{1}{\overline{Y}_1\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i^{'} S_{y_{1i}x_i}\right\},$$

$$E(\epsilon_1\epsilon_3) = \frac{Cov(\overline{y}^*_{2st}\overline{x}^{'}_{st})}{\overline{Y}_2\overline{X}} = \frac{1}{\overline{Y}_2\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i^{'} S_{y_{2i}x_i}\right\}$$

$$E(\epsilon_1\epsilon_2) = \frac{Cov(\overline{y}^*_{2st}\overline{x}_{st})}{\overline{Y}_2\overline{X}} = \frac{1}{\overline{Y}_2\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{2i}x_i}\right\} \tag{17}$$

Where, $\lambda_i = \left(\frac{1}{n_i} - \frac{1}{N_i}\right)$ and $\lambda_i^{'} = \left(\frac{1}{n_i^{'}} - \frac{1}{N_i}\right)$

Now, using the condition given in equations (15) and (16) on equations (13) and (14), The expressions for bias and MSE of suggested classes of estimators are given as follows:

$$\begin{aligned} Bias(\hat{R}_{1st}) = Bias(\hat{R}_{st}) &+ R[E(\epsilon_0\epsilon_4) - E(\epsilon_0\epsilon_3) - E(\epsilon_3\epsilon_4) + E(\epsilon_1\epsilon_4) + E(\epsilon_1\epsilon_3) \\ &+ E(\epsilon_3^2)]\phi_{1(1)}(1) + \overline{X}[E(\epsilon_2\epsilon_4) - E(\epsilon_2\epsilon_3) - E(\epsilon_3\epsilon_4) - E(\epsilon_3^2)]\alpha_1\phi_{1(1)}(1) \end{aligned} \tag{18}$$

$$MSE(\hat{R}_{1st}) = MSE(\hat{R}_{st}) + A\phi^2_{1(1)}(1) + B\alpha_1^2 + 2C\phi_{1(1)}(1) + 2D\alpha_1 + 2E\alpha_1\phi_{1(1)}(1) \tag{19}$$

$$\begin{aligned} Bias(\hat{R}_{2st}) = Bias(\hat{R}_{st}) &+ R[E(\epsilon_0\epsilon_2) - E(\epsilon_0\epsilon_3) - E(\epsilon_3\epsilon_2) + E(\epsilon_1\epsilon_2) + E(\epsilon_1\epsilon_3) \\ &+ E(\epsilon_3^2)]\phi_{1(2)}(1) + \overline{X}[E(\epsilon_2\epsilon_4) - E(\epsilon_4\epsilon_3) - E(\epsilon_2\epsilon_3) - E(\epsilon_3^2)]\alpha_2\phi_{1(2)}(1) \end{aligned} \tag{20}$$

$$MSE(\hat{R}_{2st}) = MSE(\hat{R}_{st}) + A_1\phi^2_{1(2)}(1) + B_1\alpha_2^2 + 2C_1\phi_{1(2)}(1) + 2D_1\alpha_2 + 2E_1\alpha_2\phi_{1(2)}(1) \tag{21}$$

The optimum values of $\phi_{1(1)}(1)$, $\phi_{1(2)}(1)$, α_1 and α_2 are obtained to get the minimum value of mean square error. These are given as follows:

$$\phi_{1(1)}(1) = \frac{(CE - AD)}{(AB - E^2)},\ \alpha_1 = \frac{(DE - BC)}{(AB - E^2)} \tag{22}$$

$$\phi_{1(2)}(1) = \frac{(C_1E_1 - A_1D_1)}{(A_1B_1 - E_1^2)},\ \alpha_2 = \frac{(D_1E_1 - B_1C_1)}{(A_1B_1 - E_1^2)} \tag{23}$$

The expressions for the minimum MSEs after substituting the optimum values of $\phi_{1(1)}(1)$, α_1, $\phi_{1(2)}(1)$ and α_2 from equations (21) and (22) in equations (18) and (20). we get,

$$MSE(\hat{R_{1st}}) = MSE(\hat{R_{st}}) - \left[\frac{BC^2 + AD^2 - 2CDE}{AB - E^2}\right] \tag{24}$$

$$MSE(\hat{R_{2st}}) = MSE(\hat{R_{st}}) - \left[\frac{B_1C_1^2 + A_1D_1^2 - 2C_1D_1E_1}{A_1B_1 - E_1^2}\right] \tag{25}$$

where,

$$A = R^2\left[E(\epsilon_3^2) + E(\epsilon_4^2) - 2E(\epsilon_3\epsilon_4)\right],$$

$$B = \overline{X}^2\left[E(\epsilon_2^2) + E(\epsilon_3^2) - 2E(\epsilon_2\epsilon_3)\right],$$

$$C = R^2\left[E(\epsilon_0\epsilon_4) - E(\epsilon_1\epsilon_4) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$D = R\overline{X}\left[E(\epsilon_0\epsilon_2) - E(\epsilon_1\epsilon_2) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$E = R\overline{X}\left[E(\epsilon_2\epsilon_4) - E(\epsilon_2\epsilon_3) - E(\epsilon_3\epsilon_4) + E(\epsilon_3^2)\right],$$

$$A_1 = R^2\left[E(\epsilon_2^2) + E(\epsilon_3^2) - 2E(\epsilon_2\epsilon_3)\right],$$

$$B_1 = \overline{X}^2\left[E(\epsilon_3^2) + E(\epsilon_4^2) - 2E(\epsilon_3\epsilon_4)\right],$$

$$C_1 = R^2\left[E(\epsilon_0\epsilon_2) - E(\epsilon_1\epsilon_2) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$D_1 = R\overline{X}\left[E(\epsilon_0\epsilon_4) - E(\epsilon_1\epsilon_4) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$E_1 = R\overline{X}\left[E(\epsilon_2\epsilon_4) - E(\epsilon_2\epsilon_3) - E(\epsilon_3\epsilon_4) + E(\epsilon_3^2)\right] \text{ and}$$

$$MSE(\hat{R}_{st}) = R^2[E(\epsilon_0^2) + E(\epsilon_1^2) - E(\epsilon_0\epsilon_1)] \tag{26}$$

Here, we observe that $A_1 = R^2\frac{B}{\overline{X}^2}$, $B_1 = \overline{X}^2\frac{A}{R^2}$, $C_1 = R\frac{D}{\overline{X}}$, $D_1 = \overline{X}\frac{C}{R}$ and $E_1 = E$.

After substituting these values in equation (24), we find that $MSE(\hat{R}_{1st})$ and $MSE(\hat{R}_{2st})$ are equal. i.e. $MSE(\hat{R}_{1st})$ =$MSE(\hat{R}_{2st})$ The optimum MSE of $\hat{R}_{1st}$ and $\hat{R}_{2st}$ are the same because they are utilizing the same information on y_1, y_2 and x in both cases.

5. Members of the Suggested classes

All the members of the suggested classes satisfy the conditions given in equation (10). Hence if the optimum values of the constants presented in suggested members are calculated by the expression given in equations (21) and (22) then all the members shown in table 1 will attain the minimum mean square error equal to the expression of MSE given in equation (23) and (24). The optimum values of constants are sometimes in the form of some unknown parameters and sometimes in the form of value of unknown constants. The optimum values of constants, in this situation, may be obtained from past data on the value (Reddy [11]), or by estimating the parameters included in the optimum value of the constant based on sample values. The minimum values of mean square error of the estimator up to the term of order $\frac{1}{n}$ are unchanged if we estimate the optimum values of the constants by using the sample values [Srivastava and Jhajj [15]]. If the condition given in equation (10) is satisfied by any parametric function $\phi_{(1)}(u_1)$ and $\phi_{(2)}(u_2)$ then they can generate a class of asymptotic estimators. Such classes have a large number of members. Some of them are given as follows:

Table 1: *Members of the classes*

Member of class $\hat{R}_{1st}$	Member of class $\hat{R}_{2st}$
$\hat{R}_{11st} = \left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}} + \alpha_1 \overline{x}'_{st}(u_2 - 1)\right](u_1)^\gamma$	$\hat{R}_{21st} = \left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}} + \alpha_2 \overline{x}'_{st}(u_1 - 1)\right](u_2)^\gamma$
$\hat{R}_{12st} = \left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}} + \alpha_1 \overline{x}'_{st}(u_2 - 1)\right]\left(\frac{\exp(u_1-1)}{\exp(u_1+1)}\right)$	$\hat{R}_{22st} = \left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}} + \alpha_2 \overline{x}'_{st}(u_1 - 1)\right]\left(\frac{\exp(u_2-1)}{\exp(u_2+1)}\right)$
$\hat{R}_{13st} = \left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}} + \alpha_1 \overline{x}'_{st}(u_2 - 1)\right](2 - u_1{}^\beta)$	$\hat{R}_{23st} = \left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}} + \alpha_2 \overline{x}'_{st}(u_1 - 1)\right](2 - u_2{}^\beta)$

Here, γ and β are the constants.

6. An Empirical study

Table 2: *Population parameters for each stratum*

Parameters		(i)	(ii)	(iii)	(iv)	(v)	(vi)
N_i		152	112	85	82	109	96
$n_i^{'}$		108	79	60	58	77	68
n_i		42	31	24	23	30	27
$\rho_{x_i y_{1i}}$		-0.3075	-0.0559	-0.0514	-0.2213	-0.6455	-0.5745
$\rho_{x_i y_{2i}}$		0.0739	0.6171	0.6025	0.1958	-0.245	0.5606
$\rho_{y_{1i} y_{2i}}$		0.2427	0.1695	0.4927	-0.0041	0.3483	-0.4074
$\overline{Y}_{1i}$		107.96	68.95	74.19	161.05	68.98	120.38
$\overline{Y}_{2i}$		203.1	190.41	282.74	222.39	251.98	250.28
$\overline{X}_i$		581.75	575.71	651.97	631.26	692.6	643.72
$S_{y_{1i}}$		52.07	37.79	61.75	80.31	50.22	57.27
$S_{y_{2i}}$		34.01	40.19	71.51	29.3	27.97	39.47
S_{x_i}		83.32	107.88	97.21	99.47	100.12	96.2
10% Non-resposne	$\rho_{x_i y_{1i}(2)}$	-0.6188	0.3031	0.535	-0.6709	-0.8781	-0.5204
	$\rho_{x_i y_{2i}(2)}$	0.3532	-0.2579	0.8087	-0.228	-0.3799	0.1083
	$\rho_{y_{1i} y_{2i}(2)}$	-0.3266	-0.0073	0.6923	-0.3547	0.4499	-0.2732
	$S_{y_{1i}(2)}$	61.7	22.09	79.44	65.24	68.36	29.71
	$S_{y_{2i}(2)}$	18.6	9.05	26.63	65.41	38.7	25.43
	$S_{x_i(2)}$	13.06	12.36	90.48	62.8	12.52	21.67
20 % Non-response	$\rho_{x_i y_{1i}(2)}$	-0.4829	-0.2723	0.0952	-0.7095	-0.8498	-0.263
	$\rho_{x_i y_{2i}(2)}$	0.1189	-0.653	0.7614	0.0035	0.0355	0.1717
	$\rho_{y_{1i} y_{2i}(2)}$	-0.0311	-0.3092	0.3305	-0.0823	-0.1392	-0.118
	$S_{y_{1i}(2)}$	67.53	52.33	69.71	77.85	59.14	60.05
	$S_{y_{2i}(2)}$	30.65	9.76	29.04	63.53	33.07	34.87
	$S_{x_i(2)}$	15.18	22.89	64.53	53.64	15.05	18.7
30% Non-response	$\rho_{x_i y_{1i}(2)}$	-0.4373	-0.1605	0.2785	-0.2861	-0.5	-0.3592
	$\rho_{x_i y_{2i}(2)}$	0.0884	-0.428	0.605	0.0022	0.2985	0.2831
	$\rho_{y_{1i} y_{2i}(2)}$	0.1563	0.2582	0.7877	-0.0932	-0.092	0.2379
	$S_{y_{1i}(2)}$	64.88	61.92	97.94	105.66	64.33	58.05
	$S_{y_{2i}(2)}$	31.78	10.76	72.99	76.89	32.25	46.2
	$S_{x_i(2)}$	20.39	20.51	88.09	43.81	15.46	26.26

The data used in the study has been taken from the Primary Census Abstract-2011 published by the Office of the Registrar General & Census Commissioner, India.

The number of cultivators=y_1 (Study variable)

The number of main workers=y_2 (Study variable)

The number of literate persons= x (auxiliary variable)

Here, we are considering the number of cultivators, main workers and literate persons per thousand population.

In the given population, we have six strata which are given as: Strata (i) = Central states, Strata (ii)= Eastern states, Strata (iii) = Northern states, Strata (iv)=North- East states, Strata (v) = Southern states, and Strata (vi) = Western states. And their parameters are given in table 2.

For the different choices of sub-sampling fraction i.e. $\frac{1}{l} = \frac{1}{2}$, $\frac{1}{l} = \frac{1}{3}$ and $\frac{1}{l} = \frac{1}{4}$ and for the different non-response rates 10%, 20% and 30% the optimum values of $\phi_{1(1)}(1)$,$\phi_{1(2)}(1)$, α_1 and α_2 are given in table 3.

Table 3: *Optimum values of* $\phi_{1(1)}(1)$,$\phi_{1(2)}(1)$, α_1 *and* α_2

Non-response rate(%)	Constants	$1/l = \frac{1}{2}$	$1/l = \frac{1}{3}$	$1/l = \frac{1}{4}$
10	$\phi_{1(1)}(1)$	1.1023	1.1985	1.2786
	α_1	0.00063	0.00059	0.00056
	$\phi_{1(2)}(1)$	0.92	0.8672	0.8233
	α_2	0.00075	0.00082	0.00087
20	$\phi_{1(1)}(1)$	1.1784	1.3019	1.3875
	α_1	0.0006	0.00055	0.00052
	$\phi_{1(2)}(1)$	0.8783	0.8105	0.7636
	α_2	0.00081	0.00089	0.00095
30	$\phi_{1(1)}(1)$	1.0054	1.0144	1.0204
	α_1	0.00067	0.00066	0.00066
	$\phi_{1(2)}(1)$	0.9732	0.9679	0.9649
	α_2	0.00069	0.00069	0.0007

Table 4: *The percentage relative efficiency (PRE) of* $\hat{R}_{1st}$ *and* $\hat{R}_{1st}$ *with* $\hat{R}_{st}$

Non-response rates(%)	**Estimators**	(N=636)		
		$1/l$		
		1/2	1/3	1/4
10	$\hat{R}_{st}$	**100(124.85)**	**100(134.11)**	**100(143.37)**
	$\hat{R}_{1st} = \hat{R}_{2st}$	**122.20(102.16)**	**122.53(109.45)**	**122.91(116.64)**
20	$\hat{R}_{st}$	**100(134.73)**	**100(153.88)**	**100(173.02)**
	$\hat{R}_{1st} = \hat{R}_{2st}$	**122.54(109.94)**	**123.26(124.83)**	**123.97(139.56)**
30	$\hat{R}_{st}$	**100(155.29)**	**100(194.99)**	**100(234.69)**
	$\hat{R}_{1st} = \hat{R}_{2st}$	**118.21(131.36)**	**116.08(167.97)**	**114.72(204.57)**

Note:Figures in parenthesis show MSE of the estimators in 10^{-6}.

7. Discussion

Table 4 shows the PREs and MSEs (Figures in parenthesis) of the estimators $\hat{R}_{st}$, $\hat{R}_{1st}$ and $\hat{R}_{2st}$ for the different rates of non-response and choices of the sub-sampling fraction. We can see that in table 4, as we increase the non-response rates the MSE increases for each sub-sampling fractions.

The PRE for the 10% and 20% non-response rates shows almost same value. However, for 30% non-response rate the PRE decreases for each sub-sampling fraction.

By increasing the value of l the MSE increases and PRE for 10% and 20% non-response rates are almost same and PRE decreases slightly for 30% non-response rate.

The MSE and PRE of the suggested classes of estimators are found to be equal for different rates of non-response and different choices of sub-sampling fraction.

8. Conclusion

Hence, the findings on the basis of empirical study are justified that the proposed classes of estimators for estimating the ratio of two population means with unknown population mean of auxiliary variable in the presence of non-response for stratified population performed better than the usual estimator $\hat{R}_{st}$. So, we recommend the suggested classes of estimators for the use in practice.

References

[1] El-Badry, M. A. (1956). A sampling procedure for mailed questionnaires. *Journal of the American Statistical Association* , 51(274), 209-227.

[2] Hansen, M. H., Hurwitz, W. N. (1946). The problem of non-response in sample surveys. of the American Statistical Association, 41(236), 517-529.

[3] Khare, B. B. (1991). Determination of sample sizes for a class of two phase sampling estimators for ratio and product of two population means using auxiliary character. *Metron*, 49(1-4), 185-197.

[4] Khare, B.B. Pandey, S.K. (2000). A class of estimators for ratio of two population means using auxiliary character in presence of non-response. *J. Sc. Res. B.H.U.*, 50, 115-124.

[5] Khare, B.B. Sinha, R.R. (2002). Estimators of the ratio of two population means using auxiliary character with unknown population mean in presence of non-response. *Prog. Maths. B.H.U.*, Vol.36, No.(1,2) 337-348.

[6] Khare, B. B., Sinha, R. R. (2004). Estimation of finite population ratio using two-phase sampling scheme in the presence of non-response. *Aligarh Journal of Statistics*, 24, 43-56.

[7] Khare, B. B., Sinha, R. R. (2007). Estimation of the ratio of the two population means using multi auxiliary characters in the presence of non-response. Statistical Techniques in Life Testing, *Reliability, Sampling Theory and Quality Control*, 1, 63-171.

[8] Khare, B. B., Jha, P. S., Kumar, K. (2014). Improved generalized chain estimators for ratio and product of two population means using two auxiliary characters in the presence of non-response. *International J. Stats Economics*, 13(1), 108-121.

[9] Khare, B. B., Jha, P. S. (2017). Classes of estimators for population mean using auxiliary variable in stratified population in the presence of non-response. *Communications in Statistics-Theory and Methods*, 46(13), 6579-6589.

[10] Kumar, K. Srivastava, U. (2018). Estimation of ratio and product of two population means using exponential type estimators in sample surveys. *International Journal of Mathematics and Statistics*, 19“3, pp. 102“109.

[11] Reddy, V. N. (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya C*, 40, 29-37.

[12] Singh, R. , Khare, S. , Khare, B.B. Jha , P.S. (2020). The general classes of estimators for population mean under stratified two phase random sampling in the presence of non-response. *Int. J. Agricult. Stat. Sci.*, Vol. 16, No. 2, pp. 557-565, 2020.

[13] Singh, M. P. (1965). On the estimation of ratio and product of the population parameters. *Sankhy´: The Indian Journal of Statistics*, Series B, 321-328.

[14] Singh, R. K. Naqvi, N. (2015). A generalized class of estimator of the ratio of two population means using auxiliary information. *Sri Lankan Journal of Applied Statistics*, 16“3, pp. 179“193.

[15] Srivastava, S. K., Jhajj, H. S. (1983). A class of estimators of the population mean using multiauxiliary information. *Calcutta Statistical Association Bulletin*, 32(1-2), 47-56.

[16] Ahuja, T. K. ., Misra, P. ., Belwal, O. K. . (2021). A Generalized Two Phase Sampling Estimator of Ratio of Population Means Using Auxiliary Information. *Journal of Reliability and Statistical Studies*, 14(01), 1“16.

[17] Tripathi, T.P. (1980). A general class of estimators for population ratio. *Sankhya*, Ser. C, 42, 63-75.

[18] Upadhyaya, L. N., Singh, G. N., Singh, H. P. (2000). Use of transformed auxiliary variable in the estimation of population ratio in sample survey. *Statistics in Transition*, 4(6),1019-102.

DISTRIBUTIONAL PROPERTIES OF ORDER STATISTICS AND RECORD STATISTICS FROM ERLANG-TRUNCATED EXPONENTIAL FAMILY OF DISTRIBUTION AND ITS CHARACTERIZATIONS

Imtiyaz A. Shah
•
Department of Community Medicine (Bio Statistics), Sher-I-Kashmir Institute of Medical Sciences, Srinagar (J&K) India
driashah03@gmail.com

Abstract

Erlang Truncated Exponential Distributions are characterized by distributional properties of order statistics. These characterizations include known results for ordinary order statistics based on two non-adjacent order statistics coming from two independent Erlang truncated exponential distributions. Using this method and compared with an efficient recent method given by [20], three examples of real lifetime data-sets are analyzed by that deals with non-random samples. Such type of examples predicts the accumulative new cases per million foe infection of the new COVID-19. Corollaries for Pareto and power function distributions are also derived.

Keywords: Order statistics; characterization of distributions; reliability characteristics; Erlang truncated exponential; random translation

1. Introduction

Various characterizations of Erlang truncated exponential distributions based on distributional properties of order statistics are found in the literature. Let $X_{1,n}, X_{2,n}, \cdots X_{n,n}$ denote the order statistics of a identically independent distributed (i.i.d) random variables $X_1, X_2, \cdots, X_n$, $n \geq 2$, each with distribution function $F_X(x)$. Furthermore, a variety of other models of ordered random variables are contained in this concept. For a detailed discussion of several of these models, such as sequential order statistics, k^{th} record values and Pfeifer's record model.

In this paper we present characterizations of Erlang truncated exponential distributions DF exp($\beta\alpha_\lambda$), with mean $\frac{1}{(\beta\alpha_\lambda)}$, $\beta > 0, \propto > 0, \lambda > 0$. via distributional properties of generalized order statistics including the known results for ordinary order statistics.

Consider a sequence of real numbers $X_1, X_2, \cdots, X_n$ which are independently and identically, distributed with common cumulative distribution (DF) $F_X(x)$ and the probability density function PDF $f_X(x)$ and the distribution function Then the PDF and DF of $X_{U(r)}$, r^{th} upper record is [5] and [9].

$$f_{X_{U(r)}}(x) = \frac{1}{(r-1)!}\,[R(x)]^{r-1} f(x) \tag{1}$$

and

$$\bar{F}_{X_{U(r)}}(x) = 1 - F_{X_{U(r)}}(x) = e^{-R(x)} \sum_{j=0}^{r-1} \frac{[R(x)]^j}{j!} \tag{2}$$

where

$R(x) = -\ln\bar{F}(x), \bar{F}(x) = 1 - F(x)$ (3)

The PDF and DF of $X_{r:n}$, the r^{th} order statistic from a sample of size n is given as [8] and [13].

$$f_{X_{r:n}}(x) = \frac{n!}{(r-1)!(n-r)!}\,[F(x)]^{r-1}[1-F(x)]^{n-r}f(x) \quad (4)$$

and

$$F_{X_{r:n}}(x) = \sum_{j=r}^{n}\binom{n}{j}[F(x)]^{j}[1-F(x)]^{n-j} \quad (5)$$

2. Model

The cumulative distribution function CDF $F_X(x)$and probability density function PDF $f_X(x)$ of the Extended Erlang-Truncated Exponential (EETE) distribution are given by

$F_X(x) = [1 - e^{-\beta(\alpha_\lambda)x}]^{\alpha}$, $0 \le x < \infty,\ \alpha,\ \beta,\ \lambda > 0,$ (6)

and

$$f_X(x) = \alpha\,\beta\,(\alpha_\lambda)\,e^{-\beta(\alpha_\lambda)x}[1 - e^{-\beta(\alpha_\lambda)x}]^{\alpha-1}\,,\ 0 \le x < \infty,\ \alpha,\beta,\ \lambda > 0 \quad (7)$$

where α and β are the shape parameters and λ is the scale parameter.

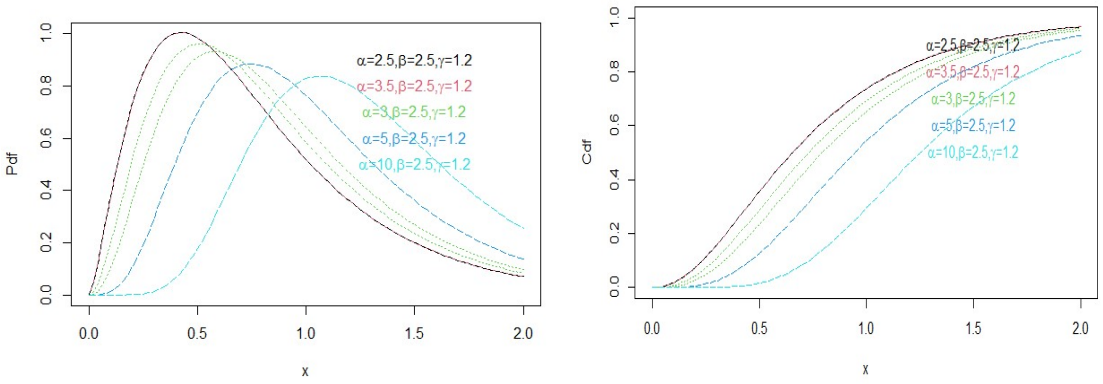


Figure 1. *Possible shapes of the probability density function $f(x)$ (left) and cumulative distribution function $F(x)$ (right) of the Extended Erlang-Truncated Exponential (EETE) distribution for fixed parameter values of β and λ.*

The Extended Erlang-Truncated Exponential (EETE) distribution reduces to Erlang-Truncated Exponential (ETE) when $\alpha = 1$.

Erlang-Truncated Exponential (ETE) distribution was originally introduced by [15] as an extension of the standard one parameter exponential distribution. The Erlang-Truncated Exponential (ETE) distribution results from the mixture of Erlang distribution and the left truncated one-parameter exponential distribution. The cumulative distribution function CDF $F_X(x)$, and probability density function PDF$f_X(x)$ of the Erlang-Truncated Exponential (ETE) distribution are given by

$F_X(x) = [1 - e^{-\beta(\alpha_\lambda)x}]\,,\quad 0 \le x < \infty,\ \beta,\ \lambda > 0,$ (8)

where $\alpha_\lambda = \left(1 - e^{-\lambda}\right)$

and

$$f_X(x) = \beta\,(\alpha_\lambda)\,e^{-\beta(\alpha_\lambda)x}\,,\qquad 0 \le x < \infty,\ \beta,\ \lambda > 0 \quad (9)$$

respectively, where β is the shape parameter and λ is the scale parameter. The Erlang-Truncated Exponential (ETE) distribution collapses to the classical one-parameter exponential distribution with parameter β and $\lambda \to \infty$.

$X \sim Par(\beta(\alpha_\lambda))$

if X has a Pareto distribution with the DF

$$F(x) = [1 - x^{-\beta(\alpha_\lambda)}], \quad 1 < x < \infty, \ \beta > 0, \alpha_\lambda > 0 \tag{10}$$

$X \sim pow\ (\beta(\alpha_\lambda))$

if X has a power function distribution with the DF

$$F(x) = x^{\beta(\alpha_\lambda)}, \quad 0 < x < 1, \ \beta > 0, \alpha_\lambda > 0 \tag{11}$$

It may further be noted that

if $\log X \sim$ Erlang-truncated exp $(\beta(\alpha_\lambda))$ then $X \sim$ Par $(\beta(\alpha_\lambda))$ (12)

if $-\log X \sim$ Erlang-truncated exp $(\beta(\alpha_\lambda))$ then $X \sim$ pow $(\beta(\alpha_\lambda))$ (13)

.

3. RELIABILITY CHARACTERISTICS

The reliability function R(x) is an important tool for characterizing life phenomenon. R(x) is analytically expressed as $R(x) = 1 - F(x)$. Under certain predefined conditions, the reliability function R(x) gives the probability that a system will operate without failure until a specified time x. The reliability function of the Extended Erlang-Truncated Exponential (EETE) distribution is given by

$$R(x) = 1 - \left(1 - e^{-\beta(\alpha_\lambda)x}\right)^{\alpha}, 0 \le x < \infty, \ \alpha, \ \beta, \ \lambda > 0 \tag{14}$$

Another important reliability characteristics is the failure rate function. The failure rate function gives the probability of failure for a system that has survived up to time x. The failure rate function h(x) is mathematically expressed $h(x) = f(x)/R(x)$. The failure rate function the Extended Erlang-Truncated Exponential (EETE) distribution is given by:

$$h(x) = \frac{\alpha\, \beta\, (\alpha_\lambda)\, e^{-\beta(\alpha_\lambda)x}[1-e^{-\beta(\alpha_\lambda)x}]^{\alpha-1}}{1-[1-e^{-\beta(\alpha_\lambda)x}]^{\alpha}}, \ 0 \le x < \infty, \ \alpha, \beta, \ \lambda > 0$$

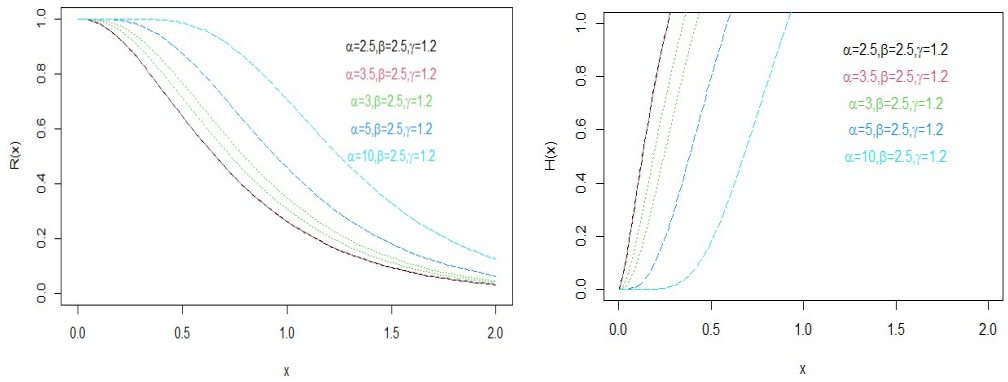


Figure 2. *Possible shapes of the reliability function $R(x)$ (left) and failure rate function $h(x)$ (right) of the Extended Erlang-Truncated Exponential (EETE) distribution for fixed parameter values of β and λ*

4. CHARACTERISTION RESULTS BASED ON UPPER RECORDS

In this section we consider a relation characterizing the Erlang-Truncated Exponential distribution based on order statistics and record statistics. This generalizes some previous characterization results and uses upper as well as lower order statistics. It has been assumed here throughout that the *df* is differentiable *w.r.t.* its argument.

THEOREM 4.1 :-

A random variable $X_{U(r)}$ be a sequence of i.i.d. non-negative random variables with an absolutely continuous distribution having the r^{th} upper statistic from a sample of size n drawn from a continuous DF $F_X(x)$ with PDF $f_X(x)$. Furthermore, let $Y_{U(j)}$ be the r^{th} upper statistic based on a sample of size n, which is drawn from a continuous DF $F_z(z) = P(Z \leq z)$, where Y is independent of X. Finally, let the relation

$$X_{U(N_o)} \overset{d}{=} X_{U(R)} + \tilde{Z} \tag{15}$$

be satisfied for all $1 \leq R < N_o \leq n$, Then, $\tilde{z} \overset{d}{=} X_{U(N_o-R)}$ and Z ~ Erlang truncated exponential $(\beta\alpha_\lambda)$ if and if $Y \sim$ Erlang truncated exponential$(\beta\alpha_\lambda)$, $\beta > 0,\ \alpha > 0, \lambda > 0$.

Proof. We first prove the necessary part. Let the moment generating function (MGF) of $X_{U(N_o)}$ be $M_{X_{U(N_o)}}(t)$. Then, (15) implies that

$$M_{X_{U(N_o)}}(t) = M_{X_{U(R)}}(t) \cdot M_{\tilde{Z}}(t) \tag{16}$$

Let us now derive the MGF of the $X_{U(R)}$based on Erlang truncated exp$(\beta\alpha_\lambda)$. Clearly, in view of (15), we get

$$M_{X_{U(R)}}(t) = \frac{\beta(\alpha_\lambda)}{(r-1)!}\int_0^\infty e^{-x((\beta\alpha_\lambda - t)}x^{R-1}dx = \left(\frac{\propto}{\propto - t}\right)^R \tag{17}$$

Where $\Gamma(.)$ is the usual gamma function. On the other hand, in view of (16)

$$M_{\tilde{Z}}(t) = \frac{M_{X_{U(N_o)}}(t)}{M_{X_{U(R)}}(t)} = \left(\frac{\propto}{\propto - t}\right)^{N_o - R} \tag{18}$$

On comparing (18) with (17), we deduce that $M_{\tilde{Z}}(t)$ is the MGF of $Y(N_o - R)$, i.e., the $(N_o - \mathrm{R})^{th}$ upper record statistics from a sample of size R and is independent of $X_{U(R)}$drawn from the DF Erlang truncated exp$(\beta(\alpha_\lambda)$. Hence the proved Necessity part.

W To prove the sufficiency part. In view of (15) be satisfied with $\tilde{Z} \overset{d}{=} Y_{U(N_o-R)}$ and Y ~ exp$(\beta(\alpha_\lambda))$. Furthermore, let $X_{U(N_o)}$and $X_{U(R)}$ in (15) be upper statistic, which are based on an unknown DF $F_X(x)$ and they are independent of $Y_{U(N_o)}$. Therefore, the convolution relation (3.1) implies that

$$\begin{aligned} f_{X_{U(N_o)}}(x) &= \int_0^\infty f_{X_{U(R)}}(y) f_{Y_{U(N_o)}}(x-y)dy \\ &= \frac{(\beta(\alpha_\lambda))^{N_o-R}}{(N_o-R-1)!}\int_0^\infty e^{-\beta(\alpha_\lambda)(x-y)} \times [x-y]^{N_o-R-1} f_{X_{U(R)}}(y)dy \end{aligned} \tag{19}$$

Differentiating both the sides of (19) $w.r.t.$ x, we get

$$\begin{aligned} \frac{d}{dx}f_{X_{U(N_o)}}(x) &= \frac{(\beta(\alpha_\lambda))^{N_o-R}}{(N_o-R-2)!}\int_0^\infty e^{-\beta(\alpha_\lambda)(x-y)} \times [x-y]^{N_o-R-2} f_{X_{U(R)}}(y)dy \\ &\quad - \frac{(\beta(\alpha_\lambda))^{N_o-R+1}}{(N_o-R-1)!}\int_0^\infty e^{-\beta(\alpha_\lambda)(x-y)} \times [x-y]^{N_o-R-1} f_{X_{U(R)}}(y)dy \end{aligned} \tag{20}$$

and by using the representation (19), we get

$$f_{X_{U(N_o-1)}}(x) = \frac{(\beta(\alpha_\lambda))^{N_o-R-1}}{(N_o-R-2)!}\int_0^\infty e^{-\beta(\alpha_\lambda)(x-y)} \times [x-y]^{N_o-R-1} f_{X_{U(r)}}(y)dy \tag{21}$$

and by combing (20) and (21), weget

$$\frac{d}{dx}f_{X_{U(N_o)}}(x) = \beta(\alpha_\lambda)[f_{X_{U(N_o-1)}}(x) - f_{X_{U(N_o)}}(x)]$$

or equivalently, by integrating from 0 to x

$$f_{X_{U(N_o)}}(x) = \beta(\alpha_\lambda)[F_{X_{U(N_o-1)}}(x) - F_{X_{U(N_o)}}(x)] \tag{22}$$

Now, by using the relation (II) of [5] and [9] on page 75, we get

$F_{X_{U(N_o-1)}}(x) - F_{X_{U(N_o)}}(x) = \frac{[R(x)]^{N_o-1}}{(N_o-1)!}\ [\bar{F}_X(x)]$ (23)

Therefore, by combing (1), (22) and (323), we have

$$\frac{f_X(x)}{\bar{F}_X(x)} = \beta(\alpha_\lambda)$$

Hence, the complete sufficient part, $F_X(x) = [1 - e^{-\beta(\alpha_\lambda)x}]\ , x > 0, \beta > 0, \propto > 0, \lambda > 0.$

Remark 4.1. ([7], Remark 1) have shown that for two adjacent upper records

$X_{U(2)} \overset{d}{=} X_{U(1)} + \tilde{Y}$

Then, $\tilde{Y} \overset{d}{=} X_{U(1)}$ and Y ~ exp(βα$_\lambda$) if and if $X \sim \exp(1)$, $\beta > 0, \propto > 0, \lambda > 0$.

Remark 4.2. [14] and [6] have shown that

$X_{U(R+1)} \overset{d}{=} X_{U(R)} + \tilde{Z}$

Then, $\tilde{Z} \overset{d}{=} X_{U(1)}$ and Y ~ exp(βα$_\lambda$) if and if $X \sim \exp(1)$, $\beta > 0, \propto > 0, \lambda > 0.$

Remark 4.3. [11] have shown

$X_{U(N_o)} \overset{d}{=} X_{U(R)} + \tilde{Z}$

Then, $\tilde{Z} \overset{d}{=} X_{U(R)}$ and Y ~ Ga$(N_o - R, 1)$ if and if $X \sim \exp(1)$, $\beta > 0, \propto > 0, \lambda > 0.$

Corollary 4.1. Assume that the RVs X and Y are independent, as we assumed in Theorem 4.1. By replacing the additive relation (15) by the multiplication relation

$X_{U(N_o)} \overset{d}{=} X_{U(R)} + \tilde{Z}$ (24)

be satisfied for all $1 \le R < N_o \le n$, Then, $\tilde{Z} \overset{d}{=} X_{U(N_o-R)}$ and Y ~ exp(βα$_\lambda$) if and if $X \sim Par(\beta\alpha_\lambda)$, $\beta > 0, \propto > 0, \lambda > 0.$

Proof. Here the proof immediately follows, by noting that if X ~ $Pareto(\beta(\alpha_\lambda))$, then log $X \sim exp\ (\beta(\alpha_\lambda))$ and

$\log X_{U(N_o)} \overset{d}{=} \log X_{U(R)} + \log \tilde{Z}$

which implies

$X_{U(N_o)} \overset{d}{=} X_{U(R)} + \tilde{Z}$

Corollary 4.2. Assume that the RVs X and Y are independent, as we assumed in Theorem 4.1. By replacing the additive relation (15) by the multiplication relation

$X_{L(N_o)} \overset{d}{=} X_{L(R)} + \tilde{Z}$ (25)

be satisfied for all $1 \le R < N_o \le n$, Then, $\tilde{Z} \overset{d}{=} X_{L(N_o-R)}$ and Y ~ exp(βα$_\lambda$) if and if $X \sim Pow(\beta\alpha_\lambda)$, $\beta > 0, \propto > 0, \lambda > 0.$

Proof. The Corollary can be proved by considering if X ~ $Power(\beta(\alpha_\lambda))$, then −logX ~ $exp(\beta(\alpha_\lambda))$ and

$$-log X^*_{L(N_o)} \overset{d}{=} -log\ X^* X^*_{L(R)} - \log Y^*$$

which implies

$$X^*_{L(N_o)} \overset{d}{=} X^*_{L(R)} \cdot Y^*$$

5. CHARACTERISTION RESULTS BASED ON ORDER STATISTICS

THEOREM 5.1 :-

A random variable $X_{R:n}$ be a sequence of *i.i.d.* non-negative random variables with an absolutely continuous distribution having the R^{th} order statistics from a sample of size n drawn from a continuous DF F_X (x) with PDF $f_X(x)$. Furthermore, let $Y_{r:n}$ be the r^{th} order statistics based on a sample of size n , which is drawn from a continuous DF $F_y(y)$, where Y is independent of X. Finally, let the relation

$X_{N_o:n} \overset{d}{=} X_{R:n} + \tilde{Z},$ (26)

be satisfied for all $1 \le R < N_o$, Then, $\tilde{Z} \overset{d}{=} X_{N_o-R:n-R}$ and Y ~ exp(βα$_\lambda$) if and if $X \sim \exp(\beta\alpha_\lambda)$, $\beta > 0, \propto > 0, \lambda > 0.$

Proof. The necessary part can be proved easily using mgf. Namely, let in view of (26) be satisfied with $X_{R:n}$ be $M_{X_{R:n}}(t)$. Then, (26) implies that

$$M_{X_{N_o:n}}(t) = M_{X_{N_o:n}}(t) \cdot M_{\tilde{Z}}(t) \tag{27}$$

Let us now derive the MGF of the $X_{X_{R:n}}$ based on Erlang truncated exp(βα$_\lambda$). Clearly, in view of (26), we get

$$M_{X_{R:n}}(t) = \frac{\beta(\alpha_\lambda)\,\Gamma(n+1)}{(R-1)!\,\Gamma(n-R+1)} \int_0^\infty [e^{-x(\beta\alpha_\lambda)}]^{n-R+\frac{t}{\alpha}}\,[1-e^{-x(\beta\alpha_\lambda}]^{R-1}\; e^{-x(\beta\alpha_\lambda)}dx \tag{28}$$

Which by using the transformation y $= e^{-x(\beta\alpha_\lambda)}$ takes the form

$$M_{X_{R:n}}(t) = \frac{\Gamma(n+1)\,\Gamma(n-R-\frac{t}{\propto}+1)}{\Gamma(n-R+1)\,\Gamma(n-\frac{t}{\propto}+1)} \tag{29}$$

Where Γ(.) is the usual gamma function. On the other hand, in view of (28)

$$M_{\tilde{Z}}(t) = \frac{M_{X_{N_O:n}}(t)}{M_{X_{R:n}}(t)} = \frac{\Gamma(n-R+1)\,\Gamma(n-N_o-\frac{t}{\propto}+1)}{\Gamma(n-N_o+1)\,\Gamma(n-R-\frac{t}{\propto}+1)} \tag{30}$$

On comparing (30) with (29), we deduce that $M_{\tilde{Z}}(t)$ is the MGF of $Y_{N_o-R:n-R}$, i.e., the $(N_o - \mathrm{R})^{th}$ order statistics from a sample of size (n − R)drawn from the DF Erlang truncated exp($\beta(\alpha_\lambda)$) and is independent of $X_{R:n}$ drawn from . This completes the proof of the necessity part.

while the proof of the sufficiency part follows closely as the sufficiency part of Theorem 5.1. Namely, let the representation (26) be satisfied with $\tilde{Y} \overset{d}{=} X_{N_o-R:n-R}$ and Y ~ exp(βα$_\lambda$). Furthermore, let $X_{N_o:n}$ and $X_{N_o-R:n-R}$ in (26) be order statistics, which are based on an unknown DF $F_X(x)$ and they are independent of $X_{R:n}$. Therefore, the convolution relation (26) implies that

$$f_{X_{N_o:n}}(x) = \int_0^x f_{X_{R:n}}(y) f_{X_{N_o-R:n-R}}(x-y)dy$$

$$= \frac{\beta(\alpha_\lambda)\,(n-R)!}{(N_o-R-1)!\,(n-N_o)!} \int_0^x e^{-\beta(\alpha_\lambda)(x-y)} \times [1-(e^{-\beta(\alpha_\lambda)(x-y)})]^{n-N_o+1} f_{X_{R:n}}(y)dy \tag{31}$$

By differentiating both the sides of (31) with respect to x, we get

$$\frac{df_{X_{N_o:n}}(x)}{dx} = \frac{(\beta(\alpha_\lambda))^2\,(N_o-R-1)\,(n-R)!}{(N_o-R-1)!\,(n-N_o)!} \int_0^x [e^{-\beta(\alpha_\lambda)\,(x-y)}]^{(n-N_o+2)} \times [1-e^{-\beta(\alpha_\lambda)(x-y)}]^{N_o-R-2} f_{X_{R:n}}(y)d$$

$$-\frac{(\beta(\alpha_\lambda))^2\,(n-N_o+1)\,(n-r)!}{(N_o-R-1)!\,(n-N_o)!} \int_0^x [e^{-\beta(\alpha_\lambda)(x-y)}]^{n-N_o+1} \times [1-(e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{N_o-R-1} f_{X_{R:n}}(y)dy$$

$$= \beta(\alpha_\lambda)\,(n-N_o+1)\,[f_{X_{N_o-1:n}}(x) - f_{X_{N_o:n}}(x)]$$

Or equivalently, by integrating from 0 to x,

$$f_{X(N_o,n)}(x) = \beta(\alpha_\lambda)(n-N_o+1)[F_{X(N_o-1,n-1)}(x) - F_{X(N_o,n)}(x)] \tag{32}$$

Now, by using the relation of [13],

$$\frac{f_X(x)}{\overline{F}_X(x)} = \beta(\alpha_\lambda)$$

which implies that

$$\mathrm{F_X}(x) = [1-e^{-\beta(\alpha_\lambda)y}],\ \beta > 0, \propto > 0, \lambda > 0, \mathrm{x} > 0$$

This complete the proof of the sufficiency part, as well as the proof of Theorem 4.1.

Corollary 5.1. A random variables (RVs) X and Y are independent, as we assumed in Theorem 5.1. By replacing the additive relation (26) by the multiplicative relation

$$X_{N_o:n} \overset{d}{=} X_{R:n}\ .\tilde{Z} \tag{33}$$

Then, $\tilde{Z} \overset{d}{=} Y_{N_o-R:n-rR}$ and Y ~ $Pareto(\beta(\alpha_\lambda))$ if and only if X ~ $Pareto(\beta(\alpha_\lambda))$

Proof. The proof follows exactly as the proof of Corollary 4.1.

Remark 5.1. [7] have proved that

$$X_{R:n} \overset{d}{=} X_{R-1:n} + U$$

Where U ~ exp (n − R + 1) if and only if X_1~exp (1).

Remark 5.2. [11] have shown that

$$X_{N_o:n} \overset{d}{=} X_{R:n} + U$$

Where U $\overset{d}{=} -log\ M$ W ~ Be (n − R + 1, N_o − R if and only if X_1~exp (1).

Corollary 5.2. A random variables (RVs) X and Y are independent, as we assumed in Theorem 5.1. By replacing the additive relation (26) by the multiplicative relation

$$X^*_{R:n} \overset{d}{=} X^*_{N_o:n} \cdot Z^*, \tag{34}$$

Then, $Z^* \overset{d}{=} Y^*_{r:s-1}$ and $Y^* \sim Power(\beta(\alpha_\lambda))$ if and if $X^* \sim Power(\beta(\alpha_\lambda), \beta > 0, \propto > 0, \lambda > 0$.

Proof. To prove the corollary, we note that

$-\log X_{N_o:n} \overset{d}{=} -log\ X_{R:n} - logX_{N_o-R:n-R}$

implies

$X_{n-N_o+1:n} \overset{d}{=} X_{n-R+1:n} + X_{n-N_o+1:n-R}$

Or,

$X_{n-N_o+1:n} \overset{d}{=} X_{n-R+1:n} + X_{n-N_o+1:n-R}$

THEOREM 5.2 :-

A random variable $X_{R:n}$ be a sequence of i.i.d. non-negative random variables with an absolutely continuous distribution having the R^{th} order statistics from a sample of size n drawn from a continuous DF F_X (x) with PDF $f_X(x)$. Furthermore, let $Y_{R:n}$ be the R^{th} order statistics based on a sample of size n , which is drawn from a continuous DF $F_z(z)$, where Z is independent of X. Finally, let the relation

$$X_{N_o:n} \overset{d}{=} X_{N_o-R:n-R} + \tilde{Z}, \tag{35}$$

be satisfied for all $1 \leq R < N_o$, Then, $\tilde{Z} \overset{d}{=} X_{R:n}$ and Y ~ exp(βα$_\lambda$) if and if $X \sim \exp(\beta\alpha_\lambda)$, $\beta > 0$, $\propto > 0, \lambda > 0$.

Proof. We first prove the necessary part. Let the moment generating function (MGF) of $X_{R:n}$ be $M_{X_{R:n}}(t)$. Then, (38) implies that

$$M_{X_{N_o:n}}(t) = M_{X_{N_o:n}}(t) \cdot M_{\tilde{Z}}(t) \tag{36}$$

Let us now derive the MGF of the $X_{X_{R:n}}$ based on Erlang truncated exp(βα$_\lambda$). Clearly, in view of (26), we get

$$M_{X_{R:n}}(t) = \frac{\beta(\alpha_\lambda)\,\Gamma(n+1)}{(R-1)!\,\Gamma(n-R+1)} \int_0^\infty [e^{-x(\beta\alpha_\lambda)}]^{n-R+\frac{t}{\alpha}}\ [1 - e^{-x(\beta\alpha_\lambda}]^{R-1}\ e^{-x(\beta\alpha_\lambda)} dx \tag{37}$$

Which by using the transformation $y = e^{-x(\beta\alpha_\lambda)}$ takes the form

$$M_{X_{R:n}}(t) = \frac{\Gamma(n+1)\,\Gamma(n-R-\frac{t}{\propto}+1)}{\Gamma(n-R+1)\,\Gamma(n-\frac{t}{\propto}+1)} \tag{38}$$

Where Γ(.) is the usual gamma function. On the other hand, in view of (3.14)

$$M_{\tilde{Z}}(t) = \frac{M_{X_{N_O:n}}(t)}{M_{X_{R:n}}(t)} = \frac{\Gamma(n-R+1)\,\Gamma(n-s-\frac{t}{\propto}+1)}{\Gamma(n-N_o+1)\,\Gamma(n-R-\frac{t}{\propto}+1)} \tag{39}$$

On comparing (39) with (38), we deduce that $M_{\tilde{Z}}(t)$ is the MGF of $Y_{N_o-R:n-R}$, i.e., the $(N_o - R)^{th}$ order statistics from a sample of size (n − R)drawn from the DF Erlang truncated exp($\beta(\alpha_\lambda)$) and is independent of $X_{R:n}$ drawn from . This completes the proof of the necessity part.

while the proof of the sufficiency part follows closely as the sufficiency part of Theorem 4.1. Namely, let the representation (26) be satisfied with $\tilde{Z} \overset{d}{=} X_{N_o-R:n-R}$ and Y ~ exp(βα$_\lambda$). Furthermore, let $X_{N_o:n}$ and $X_{N_o-R:n-R}$ in (26) be order statistics, which are based on an unknown DF $F_X(x)$ and they are independent of $X_{R:n}$. Therefore, the convolution relation (26) implies that

$$f_{X_{N_o:n}}(x) = \int_0^x f_{X_{N_o-R:n-R}}(y) f_{X_{r:n}}(x-y)dy$$

$$= \frac{\beta(\alpha_\lambda)}{B(n-R+1,R)} \int_0^x e^{-\beta(\alpha_\lambda)(x-y)^{n-R+1}} \times [1-(e^{-\beta(\alpha_\lambda)(x-y)})]^{R-1} f_{X_{N_o-R:n-R}}(y)dy \quad (40)$$

By differentiating both the sides of (40) with respect to x, we get

$$\frac{df_{X_{N_o:n}}(x)}{dx} = \frac{(\beta(\alpha_\lambda))^2\ (R-1)}{B(n-R+1,R)} \int_0^x [e^{-\beta(\alpha_\lambda)\ (x-y)}]^{(n-R+2)} \times [1-e^{-\beta(\alpha_\lambda)(x-y)}]^{R-2} f_{X_{N_o-R:n-R}}(y)dy$$

$$-\frac{(\beta(\alpha_\lambda))^2\ (n-R+1)}{B(n-R+1,R)} \int_0^x [e^{-\beta(\alpha_\lambda)(x-y)}]^{n-R+1} \times [1-(e^{-\beta(\alpha_\lambda)(x-y)})]^{R-1} f_{X_{N_o-R:n-R}}(y)dy$$

$$= \beta(\alpha_\lambda)\ (n)\ [f_{X_{N_o-1:n-1}}(x) - f_{X_{N_o:n}}(x)]$$

Or equivalently, by integrating from 0 to x,

$$f_{X_{N_o:n}}(x) = \beta(\alpha_\lambda)(n)[F_{X(N_o-1,n-1)}(x) - F_{X(N_o,n)}(x)] \quad (41)$$

Now, by using the relation of [13], we get

$$F_{X(N_o-1,n-1)}(x) - F_{X(N_o,n)}(x) = \binom{n-1}{N_o-1} [F_X(x)]^{N_o-1} [1-F_X(x)]^{n-N_o+1} \quad (42)$$

Therefore, by combing (1), (41) and (42), we get

$$\frac{f_X(x)}{\overline{F}_X(x)} = \beta(\alpha_\lambda)$$

which implies that

$F_X(x) = [1-e^{-\beta(\alpha_\lambda)y}]$, $\beta > 0, \propto > 0, \lambda > 0, x > 0$

This complete the proof of the sufficiency part, as well as the proof of Theorem 4.1.

Corollary 5.1. A random variables (RVs) X and Y are independent, as we assumed in Theorem 5.2. By replacing the additive relation (35) by the multiplicative relation

$$X_{N_o:n} \overset{d}{=} X_{R:n}\ .\tilde{Z} \quad (43)$$

Then, $\tilde{Z} \overset{d}{=} Y_{N_o-R:n-R}$ and Y $\sim$ $Pareto(\beta(\alpha_\lambda))$ if and only if X $\sim$ $Pareto(\beta(\alpha_\lambda))$.

Proof. The proof follows exactly as the proof of Corollary 4.1.

Corollary 5.2. A random variables (RVs) X and Y are independent, as we assumed in Theorem 3.2. By replacing the additive relation (26) by the multiplicative relation

$$X^*_{R:n} \overset{d}{=} X^*_{N_o:n} \cdot Z^*, \quad (44)$$

Then, $Z^* \overset{d}{=} Y^*_{r:s-1}$ and $Y^* \sim Power(\beta(\alpha_\lambda))$ if and if $X^* \sim Power(\beta(\alpha_\lambda), \beta > 0, \propto > 0, \lambda > 0$.

Proof. To prove the corollary, in view of (11) and (20).

6. APPLICATIONS

Many authors have considered prediction problems based on samples of random sizes, The importance of the order statistics in the reliability theory is attributed to the fact that the r^{th} order statistics $(n-r+1)$ out-of-n system made up of n identical components with independent life lengths. On the other hand, in dealing with censored samples, where the life-test is terminated after observing the r^{th} failure (Type II censoring), or the termination of the test occurs after a given time lapse (Type I censoring), the complete survival times can not usually be observed (due to time or cost). In many biological and agriculture problems, we often come across a situation where the sample size is not deterministic because either some observations get lost

for various reasons, or the size of the target population and its representative sample cannot be determined well. For example, assume that the inhabitants of a populous town are exposed to a dose of radiation resulting from an atomic accident, or exposed to an infection of an unknown epidemic. Furthermore, assume that our interest focuses on the time at which r persons would die among a big random sample of size n that is drawn from the residents of this town. Since the number of infected people in this town is unknown and changes randomly with time, the drawn sample contains a random number of infected and non-infected people. Accordingly, the sample size of the sub-sample of the infected people will be a non-negative integer valued RV, e.g.N, and it will be described by a sequence of independent and identically distributed RVs $X_1, X_2, \cdots, X_N$. Therefore, the r^{th} smallest order statistic will be denoted by $X_{r:N}$, which represents the time at which r persons will die.

7. CONCLUSIONS

In this paper we consider the equality by distribution of the form $Y \stackrel{d}{=} XV$, where X and V are two independent random variables. It should be noted that the random contraction–dilation schemes have important applications in many areas such as economic modeling and reliability. The characterization results given in Section 4 can be used in developing goodness-of-fit tests for the corresponding probability distributions. This paper deals with the generalized order statistics and dual generalized order statistics within a class of Erlang-Truncated Exponential distribution. Two theorems for characterizing the general form of distribution based on generalized order statistics dual generalized order statistics are given. Special cases are also deduced.

References

[1] Okorie, I. E., Akpanta, A.C and Ohakwe J. (2016). Transmuted Erlang-truncated exponential distribution, Econ. Qual. Control 31 (2) 71–84.

[2] Okorie, I.E, Akpanta, A.C. and Ohakwe, J. (2017). Marshall–Olkin generalized Erlang-truncated exponential distribution: properties and applications, Cogent Math. 4 (1) 1285093.

[3] Nasiru, S., Luguterah, A. and Iddrisu, M.M. (2016). Generalized Erlang-truncated exponential distribution, Adv. Appl. Stat. 48 (4) 273–301.

[4] Alosey-EL, R. A. (2007). Random sum of new type of mixture of distribution. Int. J. Stat. Syst. 2: 49-57.

[5] Ahsanullah, M. (1995). Record Statistics. New York: Nova Science Publishers, Inc. Commack.

[6] Ahsanullah, M. (2006). The generalized order statistics from exponential distribution. Pak. J. Statist. **22(2)**:121-128.

[7] Alzaid, A. A. and Ahsanullah, M. (2003). A characterization of the Gumbel distribution based on Record values. Commun. Statist. Theor. Meth. **32**: 2101-2108.

[8] Arnold, B.C., Balakrishan, N. and Nagaraja, H. N. (1992). A First Course in Order Statistics. New York: Wiley, USA.

[9] Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). Records. New York: Wiley, USA.

[10] Beutner, E. and Kamps, U. (2008). Random contraction and random dilation of generalized order statistics. Commun. Statist. Theor. Meth. **37**:2185-2201.

[11] Castaño-Martínez, A., López-Blázquez, F. and Salamanca-Miño, B. (2010). Random translations, contractions and dilations of order statistics and records. Statist.:

[12] David, H. A. and Nagaraja, H. N. (2003). Order Statistics, New York: John Wiley.

[13] Oncel S.Y., Ahsanullah, M., Aliev, F.A. and Aygun, F. (2005). Switching record and order statistics via random contraction. Statist. Probab. Lett. **73**: 207-217.

[14] Wesolowski, J. and Ahsanullah, M. (2004). Switching order statistics through random power contractions. Aust. N.Z.J. Statist. **46(2)**: 297-303.

[15] Alosey-EL, R. A. (2007). Random sum of new type of mixture of distribution. Int. J. Stat. Syst. 2: 49-57.

[16] Rao, C. R. & Chanabhng, D. N. (1998), Recent Approaches to Characterizations Based on Order Statistics and Record Values (Handbook of Statistics, 16, 231- 256), Elsevier.

[17] Tavangar, M. & Hashemi, M. (2013), 'On Characterizations of the Generalized Pareto Distributions Based on Progressively Censored Order Statistics', *Statistical Papers* 54, 381–390.

[18] Fagiuoli, E., Pellerey, F. and Shaked, M. (1999). A characterization of the dilation order and its applications. Statist. Papers 40: 393-406.

[19] Rao, C. R. & Chanabhng, D. N. (1998), Recent Approaches to Characterizations Based on Order Statistics and Record Values (Handbook of Statistics, 16, 231- 256), Elsevier.

[20] Barakat et. Al. (2021), Predicting future order statistics with random sample size. AIMS Mathematics, 6(5): 5133–5147.

ESTIMATION OF FRECHET PARAMETERS WITH TIME-CENSORED DATA IN ACCELERATED LIFE TESTING UTILISING THE GEOMETRIC PROCESS

Abdul Kalam [1*], Cheng Weihu[2], Ahmadur Rahman[3],Mohammad Ahmad[4]

•

[1,2,4]Faculty of Science, Beijing University of Technology, China,
[3] Department of Statistics and O.R., Aligarh Muslim University, Aligarh-202002, India
[1]faisal.stats@gmail.com, [2]chengweihu@bjut.edu.cn, [3]ahmadur.st@gmail.com,
[4]mahmador@gmail.com
[*]Corresponding Author

Abstract

The geometric process (GP) has been applied to estimate constant stress accelerated life testing for the Frechet failure item with time-censored data. A geometric process (GP) is developed by the failure time of tested items when stress levels are constantly rising. The estimates of the various parameters are calculated using the maximum likelihood estimation procedure. The asymptotic variance of estimates is obtained using a Fisher information matrix. The asymptotic variance is then used to calculate the distribution parameter asymptotic interval values. The statistical properties and confidence intervals of the required parameters are then illustrated using a simulation technique.

Keywords: Frechet Distribution, Geometric Process, Maximum Liklihood Estimate, Asymptotic Confidence Interval Estimate, Simulation Study.

I. Introduction

The most common approach to product evaluation is accelerated life testing (ALT), which gives the necessary details on the product's life under normal usage. It's widely employed in the manufacturing sector for the purpose of enhancing product quality. In order to gather data swiftly than under normal conditions, it enables the researcher to enhance the stresses on the life distribution parameters. Because working under normal conditions would be time consuming, it is impractical to test items under greater stress than is normal in order to induce early failure. The life distribution of a product, as well as any related characteristics under normal stress, must be extrapolated from test data using accelerated life analysis. Comparing such a test to tests conducted under normal conditions, time and money are saved.

Making the decision as to what stress should be imposed and how is the most challenging job in Accelerated Life Testing. The ALT contains a variety of stress loading types, such as constant stress , progressive stress, step stress, random stress and cyclic stress. ALT has mainly two types of data that is complete (every failure time is available) and censored (some failure time is unavailable).
Numerous authors have given their perspective on accelerated life testing (for constant stress), references includes [1, 2, 3, 4]. Yang[5], introduced optimal design using a four-level ALT and compared it with three-level ALT for different censoring schemes. ALT utilised Lam's[6] GP concept to investigate the problem of repair replacement. ALT plans for Generalized Exponential

Distribution under GP was analyse by Lone[7]. A great number of literature on ALT under GP model are (see,[8, 11, 9, 10]). Zhou[12] showed ALT with progressive hybrid censoring under geometric process for Rayleigh distribution. And the same time Huang[13] preseneted GP for exponential failure model in respect of complete as well censored observation. Fan [14] explored the constant ALT design for the generalised gamma model. For life distributions like exponential and lognormal distributions, Chen[15] discovered Bayesian approximations of the parameters in a generalised linear model (GP). So many works has done on GP in ALT, see Lone SA[16], Kamal M[17], Lone SA[18], Kamal M[19], Zarrin S[20], Lone SA[22], Lone SA[23], Ismail[24], Lone SA[25], Aly H[26], Alam I[27]. Using Informative and Noninformative Priors, Sindhu[28] performed a Bayesian Study for Censored Shifted Gompertz Mixture Distributions. Nassr, SG[29] extended Weibull distribution under adaptive type II progressive hybrid censoring. Hemmati, F[30] provided the log-normal distribution under type-II progressive hybrid censoring . For the Modified Kies exponential distribution, Hussam E[31] investigated simple and multiple ramp-stres ALT design for type-II censored data and Binomial Removal.

II. Model Description and Testing Procedure

I. Geometric Process (GP)

A sequence of stochastic variable $\{X_n\}$, $n = 1, 2, 3...$ is referred to as a Geometric Process (GP) if $\{\lambda^{n-1}X_n, n = 1, 2, 3...\}$ established a renewal process. Where, ratio of GP $\lambda(> 0)$ is a real valued. It may be demonstrated that if $\{X_n\}$, $n = 1, 2, 3...$ develops a GP and there exist a random variable having pdf $f(x)$ with mean γ and variance σ^2 then the subsequent pdf of X_n will be given as $\lambda^{n-1}f(\lambda^{n-1}x)$ with mean $E(X_n) = \gamma/\lambda^{n-1}$ and variance $V(X_n) = \sigma^2/\lambda^{2(n-1)}$.

II. Frechet Failure Model

Probability density function of the Frechet variable is :

$$f(x) = \alpha\beta^{\alpha}x^{-\alpha-1}e^{-(\frac{x}{\beta})^{-\alpha}}, \qquad x > 0, \alpha > 0, \beta > 0 \tag{1}$$

where, α(shape) and β (scale) are parameters of the life distribution:

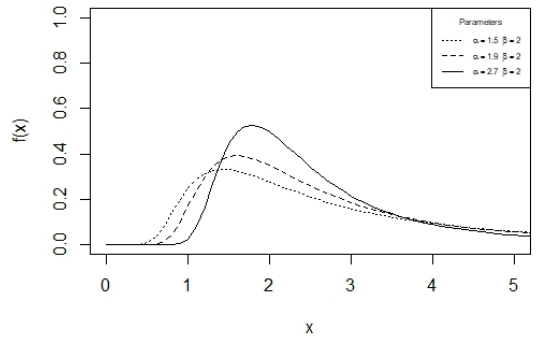


Figure 1: *PDF at different shapes and fixed scale ($\beta = 2$)*

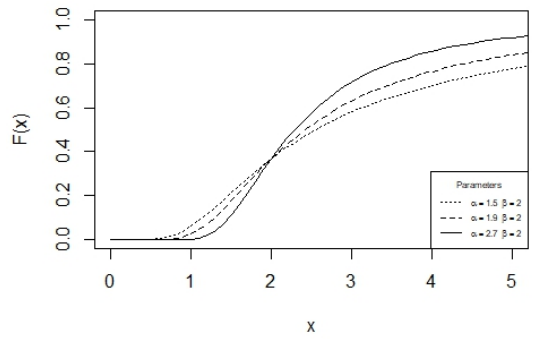


Figure 2: *CDF at different shapes and fixed scale ($\beta = 2$)*

CDF of Frechet variable takes the following expression:

$$F(x) = e^{-(\frac{x}{\beta})^{-\alpha}}, \qquad x > 0, \alpha > 0, \beta > 0 \tag{2}$$

Survival function for the Frechet variable is given as:

$$S(x) = 1 - e^{-(\frac{x}{\beta})^{-\alpha}}, \qquad x > 0 \tag{3}$$

The hazard function (HF) is

$$h(x) = \frac{\alpha\beta^{\alpha}x^{-\alpha-1}e^{-(\frac{x}{\beta})^{-\alpha}}}{1 - e^{-(\frac{x}{\beta})^{-\alpha}}} \tag{4}$$

III. Assumptions

1. The lifetime of failed items follows Frechet distribution at each stress.
2. Supoose a life test is organized with s number of stresses(increasing order). A random sample of n items are place on each stress and begin to operate simultaneously. Let the failure time of i^{th}(ranges from 1 to n) item in k^{th}(ranges from 1 to s) stress is denoted by x_{ki}. Now, failed items are removed and the test will run till the complete sample is exhausted at a predetermined t(censoring time)at each stress.
3. Stress is a log-linear function of the scale parameter β i.e., $log(\beta_k) = a + bS_k$, where a and b are unknown parameters, values dependind on the nature of products and test method.
4. Let say the lifespan of items on each stress is represented by random variables $X_0, X_1, X_2, ..., X_s$, where X_0 is the lifespan of the items under normal stress and sequence $X_k, k = 1, 2, ...s$ formulates a GP with ratio parameter $\lambda > 0$.

All of the preceding assumptions, except for the last one assumption (4^{th} assumption), are generally accepted in the ALT. Last one is based on the notion of a geometric procedure that is better than the traditional one without making computation more difficult. The following theorem, which presupposes a log linear function between life and stress is demonstrates as:

Theorem 1. If stresses in an ALT increases constantly, then the lifespan of products at each stress develops a GP. i.e., if the difference $(S_{k+1} - S_k) = (\Delta S)constant$, for $k = 1, 2, ..., s - 1$, then $\{X_k\}$, $k = 1, 2, ..., s$ develops a GP.

Proof. Last assumption (4^{th} assumption) states that,

$$\log\left(\frac{\beta_{k+1}}{\beta_k}\right) = b(S_{k+1} - S_k) = b\ (\Delta S\ ) \tag{5}$$

This demonstrates that the excessive stresses comprise an arithmetic series with a difference ΔS. (constant) Now, the previous expression can be rewritten as:

$$\frac{\beta_{k+1}}{\beta_k} = e^{b\Delta S} = \frac{1}{\lambda}, \tag{6}$$

from (6), we have

$$\beta_k = \frac{1}{\lambda}\beta_{k-1} = \frac{1}{\lambda^2}\beta_{k-2} = ... = \frac{1}{\lambda^k}\beta$$

The lifetime pdf of an object has the following structure at the kth stress level

$$f(x) = \alpha\beta^{\alpha}x^{-\alpha-1}e^{-(\frac{x}{\beta})^{-\alpha}}$$

$$f_{X_k}(x) = \lambda^{-\alpha k r_k}\ \alpha^{r_k}\beta^{\alpha r_k}x_{ki}\ e^{\left(-\frac{\lambda^k x_{ki}}{\beta}\right)} \tag{7}$$

And the cdf is written as

$$F_{X_k}(x) = e^{\left(-\frac{\lambda^k x_{ki}}{\beta}\right)} \tag{8}$$

This shows that

$$f_{X_k}(x) = \lambda^k f_{X_0}(\lambda^k x) \tag{9}$$

Hence, from the theory of GP and using the equation(9) it is obvious that if the pdf of lifespan of the X_0 (normal stress) is $f_{X_0}(x)$, then the pdf of the lifespan X_k (kth stress) is given by $\lambda^k f_{X_0}(\lambda^k x)$. Finally , it is evident that lifespan of a arithmetically increasing stresses results a GP. ■

IV. Maximum Likelihood Estimation

The most significant and frequently applied estimation technique is the maximum likelihood (ML) approach. The applicability of other methods is restricted, whereas it can be used with any probability distribution. ML estimation achievement in ALT is more complicated, and closed-form estimates of parameters are typically unavailable. Consequently, to calculate them, an arithmetic technique Newton Raphson (NR) method has been used.
Let's say the test is stopped at time t for each stress level, and only $x_{ki}(\leq t)$ failures are recorded. Suppose that $r_k \leq n$ failures at the k^{th} stress levels are obtained prior to suspending the test, and remaing $(n - r_k)$ items are survived till the entire test without any failure. The likelihood function of a particular stress is provided for time-censored Frechet failure data under GP with s number of stress:

$$L_k = \frac{n!}{(n-r_k)!}\left[\lambda^{-\alpha k r_k}\ \alpha^{r_k}\beta^{\alpha r_k}\prod_{i=1}^{r_k}x_{ki}\ e^{\left(-\frac{\lambda^k x_{ki}}{\beta}\right)}\right]\left[1 - e^{\left(-\frac{\lambda^k t}{\beta}\right)}\right]^{n-r_k} \tag{10}$$

Consequently, the likelihood function for overall stresses is

$$\begin{aligned} L_s(\alpha,\beta,\lambda) &= \prod_{s=1}^{k} L_k \\ &= \prod_{k=1}^{s} \frac{n!}{(n-r_k)!}\left[\lambda^{-\alpha k r_k}\ \alpha^{r_k}\beta^{\alpha r_k}\prod_{i=1}^{r_k} x_{ki}\ e^{\left(-\frac{\lambda^k x_{ki}}{\beta}\right)}\right]\left[1-e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\right]^{n-r_k} \end{aligned} \tag{11}$$

Taking log both sides and the likelihood function for the above equation is

$$\begin{aligned} l &= \ln L_s(\alpha,\beta,\lambda) \\ &= \sum_{k=1}^{s}\left[\ln\left(\frac{n!}{(n-r_k)!}\right) + r_k(\ln\alpha+\ln\beta-\alpha k\ln\lambda) - (\alpha+1)\sum_{i=1}^{r_k}\ln x_{ki} - \left(\frac{\lambda^k}{\beta}\right)^{-\alpha}\sum_{i=1}^{r_k} x_{ki}^{-\alpha}\right. \\ &\left. + (n-r_k)\ln\left\{1-e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\right\}\right] \end{aligned} \tag{12}$$

MLE's of different parameters α, β and λ has found after solving the these normal equations $\frac{\partial l}{\partial \alpha}$, $\frac{\partial l}{\partial \beta}$ and $\frac{\partial l}{\partial \lambda} = 0$.

$$\begin{aligned} \frac{\partial l}{\partial \alpha} = \sum_{k=1}^{s}\left[-k\ r_k\ln\lambda + \alpha^{-1}r_k - \sum_{i=1}^{r_k}\ln x_{ki} + \left(\frac{\lambda^k}{\beta}\right)^{-\alpha}\ln\left(\frac{\lambda^k}{\beta}\right)\sum_{i=1}^{r_k} x_{ki} - (n-r_k)A^{-1}\right. \\ \left.\times\left\{e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}\ln\left(-\frac{\lambda^k t}{\beta}\right)\right\}\right] \end{aligned} \tag{13}$$

$$\frac{\partial l}{\partial \beta} = \sum_{k=1}^{s}\left[\beta^{-1}r_k - \left(\frac{\lambda^k}{\beta}\right)^{-\alpha}\ln\beta\sum_{i=1}^{r_k} x_{ki} + (n-r_k)A^{-1}\left\{e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}\ln\beta\right\}\right] \tag{14}$$

$$\frac{\partial l}{\partial \lambda} = \sum_{k=1}^{s}\left[-\alpha\ k\ r_k\lambda^{-1} - \beta^{\alpha}\alpha\ k\ \lambda^{-(\alpha k+1)}\sum_{i=1}^{r_k} x_{ki} - (n-r_k)A^{-1}\left\{\alpha\ k\ \lambda^{-(\alpha k+1)}\left(\frac{t}{\beta}\right)^{-\alpha}e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\right\}\right] \tag{15}$$

where $A = 1 - e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}$

As we can see,expressions(13), (14) and (15) are not linear. Consequently, it is challenging to find a closed-form answer. Therefore, the estimate of α, β and λ is obtained by concurrently solving these expressions using the NR method.

V. Asymptotic Confidence Interval

Under some specific regularity restrictions, large sample theory assures the consistency and normality of ML estimators. Because the estimate of parameters are not forming closed form, exact confidence intervals of the parameters cannot be determined . As a consequence, asymptotic confidence intervals rather than exact confidence intervals develop using the asymptotic property of MLE's.

Providing the The Fisher Information Matrix (FIM) as:

$$F = \begin{bmatrix} -\frac{\partial^2 l}{\partial \alpha^2} & -\frac{\partial^2 l}{\partial \alpha \partial \beta} & -\frac{\partial^2 l}{\partial \alpha \partial \lambda} \\ -\frac{\partial^2 l}{\partial \beta \partial \alpha} & -\frac{\partial^2 l}{\partial \beta^2} & -\frac{\partial^2 l}{\partial \beta \partial \lambda} \\ -\frac{\partial^2 l}{\partial \lambda \partial \alpha} & -\frac{\partial^2 l}{\partial \lambda \partial \beta} & -\frac{\partial^2 l}{\partial \lambda^2} \end{bmatrix}$$

The components of FIM can be obtained as follows:

$$\frac{\partial^2 l}{\partial \alpha^2} = \sum_{k=1}^{s} \left[-\alpha^{-1} r_k - \left(\frac{\lambda^k}{\beta}\right)^{-\alpha} \left\{ \ln\left(\frac{\lambda^k}{\beta}\right) \right\}^2 \sum_{i=1}^{r_k} x_{ki} - (n - r_k) A^{-2} \right.$$
$$\left. \times \left\{ e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} \ln\left(-\frac{\lambda^k t}{\beta}\right) \right\} \left\{ A \left(\left(\frac{\lambda^k t}{\beta}\right)^{-\alpha} - 1 \right) - e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} \right\} \right] \tag{16}$$

$$\frac{\partial^2 l}{\partial \beta^2} = \sum_{k=1}^{s} \left[\beta^{-2}\, r_k - \left(\frac{\lambda^k}{\beta}\right)^{-\alpha} (\ln \beta)^2 \sum_{i=1}^{r_k} x_{ki} + (n - r_k) A^{-2} \right.$$
$$\left. \times \left\{ e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} \ln \beta \left(A \left(\beta^{\alpha} \ln \beta - e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} \ln \beta \right) - 1 \right) \right\} \right] \tag{17}$$

$$\frac{\partial^2 l}{\partial \lambda^2} = \sum_{k=1}^{s} \left[-\alpha\ k\ r_k \lambda^{-2} + \beta^{\alpha}\ \alpha(\alpha+1)\ k\ \lambda^{-(\alpha k+2)} \sum_{i=1}^{r_k} x_{ki} + (n - r_k) A^{-2} \right.$$
$$\left. \times \left\{ \alpha\ k\ \lambda^{-(\alpha k+2)} \left(\frac{t}{\beta}\right)^{-\alpha} e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(A \left(\alpha\ k \lambda^{-(\alpha k+1)} \left(\frac{t}{\beta}\right)^{-\alpha} \right) + \left(\frac{t}{\beta}\right)^{-\alpha} e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \right) \right\} \right] \tag{18}$$

$$\frac{\partial^2 l}{\partial \alpha \partial \beta} = \sum_{k=1}^{s} \left[-\left(\frac{\lambda^k}{\beta}\right)^{-\alpha} \left(\frac{1}{\beta}\right) \left(1 + \alpha \ln\left(\frac{\lambda^k}{\beta}\right)\right) \sum_{i=1}^{r_k} x_{ki} + (n - r_k) A^{-2} \right.$$
$$\left. \times \left\{ e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} \left(\frac{1}{\beta} + \alpha \ln\left(\frac{\lambda^k}{\beta}\right)\right) + \left(\frac{1}{\beta}\right) \ln \beta - e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} \ln \beta \right\} \right] \tag{19}$$

$$\frac{\partial^2 l}{\partial \alpha \partial \lambda} = \sum_{k=1}^{s} \left[-\lambda^{-1} k\ r_k + \left(\frac{\lambda^k}{\beta}\right)^{-\alpha} \frac{k}{\lambda} \left(\frac{1}{\beta} + \alpha \ln\left(\frac{\lambda^k}{\beta}\right)^{-\alpha} \right) + (n - r_k) A^{-2} \right.$$
$$\left. \times \left\{ \alpha\ k\ \lambda^{-(\alpha k-1)} \left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha} e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \left(\alpha\ k \lambda^{-(\alpha k+1)} \left(\frac{t}{\beta}\right)^{-\alpha} \right) + \left(\frac{t}{\beta}\right)^{-\alpha} e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}} \right\} \right] \tag{20}$$

$$\frac{\partial^2 l}{\partial\beta\partial\lambda} = \sum_{k=1}^{s}\left[-\lambda^{-1}k\alpha\left(\frac{\lambda^k}{\beta}\right)^{-\alpha}\ln\beta\sum_{i=1}^{r_k}x_{ki} - (n-r_k)A^{-2} \times\left\{e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}\ln\beta\left(\frac{1}{\beta}+\alpha\ln\left(\frac{\lambda^k}{\beta}\right)\right) - e^{\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}}\left(-\frac{\lambda^k t}{\beta}\right)^{-\alpha}\ln\beta\right\}\right] \tag{21}$$

The var-covariance matrix is:

$$F^{-1} = \begin{bmatrix} -\frac{\partial^2 l}{\partial\alpha^2} & -\frac{\partial^2 l}{\partial\alpha\partial\beta} & -\frac{\partial^2 l}{\partial\alpha\partial\lambda} \\ -\frac{\partial^2 l}{\partial\beta\partial\alpha} & -\frac{\partial^2 l}{\partial\beta^2} & -\frac{\partial^2 l}{\partial\beta\partial\lambda} \\ -\frac{\partial^2 l}{\partial\lambda\partial\alpha} & -\frac{\partial^2 l}{\partial\lambda\partial\beta} & -\frac{\partial^2 l}{\partial\lambda^2} \end{bmatrix}^{-1}$$

The diagonal elements in the above matrix are representing variances terms, and the non-diagonal are indicating covariances .

Estimates of asymptotic confidence interval for the proposed parameters α, β and λ are written as:

$\left[\hat{\alpha} \pm Z_{1-\frac{\varphi}{2}}\left(SE(\hat{\alpha})\right)\right]$, $\left[\hat{\beta} \pm Z_{1-\frac{\varphi}{2}}\left(SE(\hat{\beta})\right)\right]$ and $\left[\hat{\lambda} \pm Z_{1-\frac{\varphi}{2}}\left(SE(\hat{\lambda})\right)\right]$ respectively.

III. Simulation study

Simulation study is a computational approach to analyze the behavior of the function. The uniform distribution has used to better understand the characteristics of the parameters. The proposed simulation method going through these steps.

1. Generating a pseudo random sample using the distribution $u[0,1]$.
2. Inverse-cdf method applied to transform the equation (8) in terms of u. Expression of $x_{ki}(\leq t)$ is:
$$x_{ki} = \frac{\beta}{\lambda^k\left(-\log u\right)^{1/\alpha}}$$
3. 5000 random samples of size 25, 50, 80, 120 and 150 have been produced from the Frechet Distribution.
4. Opted a fixed censoring time $t = 3.5$ at normal condition.
5. Taken values of number of failed items $r_k = (0.8 \times n)$, where n is sample size.
6. Chosen stress levels are $s = 4, 6$ and 8 along with parameter values $\alpha = 1.4$, $\lambda = 0.5$, $\beta = 1.1$.
7. Finally, optim () function in R-Programming Software has been used to calculate the ML estimates of mean along with various statistical measurements such as root mean squared error (RMSE), relative absolute bias (RAB), and lower and upper limit of 95% confidence intervals for different samples of different sizes.

Table 1: *Simulation findings for* $\alpha = 1.4$, $\lambda = 0.5$, $\beta = 1.1$ *and* $s = 4$

Sample size(n)	Failed items(r_k)	Estimates	Mean	RMSE	RAB	Lower limit	Upper limit
		α	1.3478	0.0721	0.0373	1.2065	1.4891
25	20	λ	0.4010	0.0523	0.1980	0.2985	0.5035
		β	1.0501	0.1230	0.0454	0.8090	1.2912
		α	1.3731	0.0515	0.0192	1.2722	1.4740
50	40	λ	0.3044	0.0415	0.3912	0.2231	0.3857
		β	1.0623	0.0871	0.0343	0.8916	1.2330
		α	1.3836	0.0409	0.0117	1.3034	1.4638
80	64	λ	0.7133	0.0328	0.4266	0.6490	0.7776
		β	1.0603	0.0688	0.0361	0.9255	1.1951
		α	1.3926	0.0335	0.0053	1.3269	1.4583
120	96	λ	0.5040	0.0456	0.0080	0.4146	0.5934
		β	1.0847	0.0561	0.0139	0.9747	1.1946
		α	1.3963	0.0300	0.0026	1.3375	1.4551
150	120	λ	0.4659	0.0431	0.0688	0.3814	0.5504
		β	1.0847	0.0501	0.0139	0.9865	1.1829

Table 2: *Simulation findings for* $\alpha = 1.4$, $\lambda = 0.5$, $\beta = 1.1$ *and* $s = 6$

Sample size(n)	Failed items(r_k)	Estimates	Mean	RMSE	RAB	Lower limit	Upper limit
		α	1.3218	0.0508	0.0558	1.2222	1.4214
25	20	λ	0.8050	0.0714	0.6100	0.6651	0.9449
		β	1.1030	0.1230	0.0027	0.8619	1.3441
		α	1.3429	0.0363	0.0407	1.2718	1.4140
50	50	λ	0.8024	0.0513	0.6048	0.7019	0.9029
		β	1.1260	0.0710	0.0236	0.9868	1.2652
		α	1.3571	0.0288	0.0306	1.3006	1.4135
80	64	λ	0.3097	0.0363	0.3806	0.2386	0.3808
		β	1.0979	0.0561	0.0019	0.9879	1.2079
		α	1.3648	0.0236	0.0251	1.3185	1.4110
120	96	λ	0.1058	0.0213	0.7884	0.0641	0.1475
		β	1.0921	0.0459	0.0072	1.0021	1.1821
		α	1.3683	0.0212	0.0226	1.3267	1.4098
150	120	λ	0.0504	0.0113	0.8992	0.0283	0.0725
		β	1.0900	0.0411	0.0091	1.0094	1.1706

Table 3: *Simulation findings for* $\alpha = 1.4$, $\lambda = 0.5$, $\beta = 1.1$ *and* $s = 8$

Sample size(n)	Failed items(r_k)	Estimates	Mean	RMSE	RAB	Lower limit	Upper limit
		α	1.3812	0.0454	0.0134	1.2922	1.4702
25	20	λ	0.2087	0.0583	0.5826	0.0944	0.3229
		β	1.1030	0.0977	0.0027	0.9115	1.2945
		α	1.3690	0.0285	0.0221	1.3131	1.4249
50	40	λ	0.8024	0.0367	0.6048	0.7305	0.8743
		β	1.1505	0.0614	0.0459	1.0301	1.2708
		α	1.3692	0.0226	0.0220	1.3249	1.4135
80	64	λ	0.7041	0.0513	0.4082	0.6035	0.8046
		β	1.1447	0.0484	0.0406	1.0498	1.2396
		α	1.3742	0.0184	0.0184	1.3381	1.4103
120	96	λ	0.7084	0.0273	0.4168	0.6549	0.7619
		β	1.1943	0.0395	0.0857	1.1169	1.2717
		α	1.3767	0.0165	0.0166	1.3443	1.4090
150	120	λ	0.6099	0.0127	0.2198	0.5850	0.6348
		β	1.1681	0.0355	0.0619	1.0985	1.2376

In this study, various measures such as average mean values, RMSE and RAB are calculated using 5000 replications of different samples to avoid randomness. The results presented in Table 1-3 are based on different sample sizes with parameter values $\alpha = 1.4$, $\lambda = 0.5$, $\beta = 1.1$ and stresses $s = 4, 6$ and 8 to analyse the performance of the MLEs of the Frechet parameters.
Table 1-3 shows that, nearly all of the parameter estimates in Table 3 result in lesser RMSEs and RABs compared with the estimates in Table 1-2. In every situation, the RMSEs of the MLEs of the parameters in Table 1-3 decrease as sample size increases.

IV. Conclusion

In the current study, a Frechet failure item accelerated life testing (ALT) design with time-censored data has been taken into consideration. The likelihood equation for the Frechet parameter is built using a geometric process, which is produced by the failure time of tested objects under constantly increasing stress leveles. Since the likelihood equation does not have the closed-form, the Newton-Raphson technique is used to calculate the mean, root mean square error (RMSE), and relative absolute bias (RAB) for the parameters. The results provided in Table 1-3 show that the estimates are reasonably near to their true values with low RMSEs. A larger sample number results in lower RMSE values and a narrower confidence interval. This work can be extended for various censoring schemes such progressive censoring.

References

[1] Kundu, D. and Gupta, R.D., (1999). Generalized exponential distribution. Aust NZJ Stat, 41, pp.173-188.
[2] Ahmad, K.E., Fakhry, M.E. and Jaheen, Z.F., (1997). Empirical Bayes estimation of P (Y< X) and characterizations of Burr-type X model. Journal of Statistical Planning and Inference, 64(2), pp.297-308.

[3] Rahman, A. and Lone, S.A., (2019). Mathematical Model of Accelerated Life Testing Plan Using Geometric Process. Bayesian Analysis and Reliability Estimation of Generalized Probability Distributions, p.30.
[4] Ahmad, N., (2010). Designing accelerated life tests for generalised exponential distribution with log-linear model. International Journal of Reliability and Safety, 4(2-3), pp.238-264.
[5] Yang, G.B., (1994). Optimum constant-stress accelerated life-test plans. IEEE transactions on reliability, 43(4), pp.575-581.
[6] Lin, Y.L.Y., (1988). Geometric processes and replacement problem. Acta Mathematicae Applicatae Sinica, 4, pp.366-377.
[7] Lone, S.A., Alam, I. and Rahman, A., (2022). Statistical analysis under geometric process in accelerated life testing plans for generalized exponential distribution. Annals of Data Science, pp.1-13.
[8] Lone, S.A. and Rahman, A., (2016). Arif-Ul-Islam, Estimation in Step-Stress Partially Accelerated Life Tests for the Mukherjee-Islam Distribution Using Time Constrain. International Journal of Modern Mathematical Sciences, 14(3), pp.227-238.
[9] Rahman, A., Sindhu, T.N., Lone, S.A. and Kamal, M., (2020). Statistical inference for Burr Type X distribution using geometric process in accelerated life testing design for time censored data. Pakistan journal of statistics and operation research, pp.577-586.
[10] Mohamed, A.E.R., Abu-Youssef, S.E., Ali, N.S. and Abd El-Raheem, A.M., (2018). Inference on constant-stress accelerated life testing based on geometric process for extension of the exponential distribution under type-II progressive censoring. Pakistan Journal of Statistics and Operation Research, pp.233-251.
[11] Watkins, A.J. and John, A.M., (2008). On constant stress accelerated life tests terminated by Type II censoring at one of the stress levels. Journal of statistical Planning and Inference, 138(3), pp.768-786.
[12] Zhou, K., Shi, Y.M. and Sun, T.Y., (2012). Reliability analysis for accelerated life-test with progressive hybrid censored data using geometric process.
[13] Huang, S., (2011). Statistical inference in accelerated life testing with geometric process model (Doctoral dissertation, Sciences).
[14] Fan, T.H. and Yu, C.H., (2013). Statistical inference on constant stress accelerated life tests under generalized gamma lifetime distributions. Quality and Reliability Engineering International, 29(5), pp.631-638.
[15] Chen, J., Li, K.H. and Lam, Y., (2010). Bayesian computation for geometric process in maintenance problems. Mathematics and computers in simulation, 81(4), pp.771-781.
[16] Lone, S.A. and Ahmed, A., (2021). Design and Analysis of Accelerated Life Testing and its Application Under Rebate Warranty: Accelerated Life Testing. Sankhya A, 83, pp.393-407.
[17] Kamal, M., Khan, S., Rahman, A., Aldallal, R.A., Abd El-Raouf, M.M., Muse, A.H. and Rabie, A., (2022). Reliability Analysis of Hybrid System Using Geometric Process in Multiple Level of Constant Stress Accelerated Life Test through Simulation Study for Type-II Progressive Censored Masked Data. Mathematical Problems in Engineering, 2022.
[18] Lone, S.A. and Rahman, A., (2017). Step Stress Partially Accelerated Life Testing Plan For Competing Risk Using Adaptive Type-I Progressive Hybrid Censoring. Pakistan Journal of Statistics, 33(4).
[19] Kamal, M., (2013). Design Of Accelerated Life Testing Using Geometric Process For Pareto Distribution With Type-I Censoring. Journal of Global Research in Mathematical Archives (JGRMA), 1(8), pp.59-66.
[20] Zarrin, S., (2013). Designs Of Accelerated Life Tests (Doctoral Dissertation, Aligarh Muslim University Aligarh).
[21] Kamal, M., Rahman, A., Zarrin, S. and Kausar, H., (2021). Statistical inference under step stress partially accelerated life testing for adaptive type-II progressive hybrid censored data. Journal of Reliability and Statistical Studies, pp.585-614.

[22] Lone, S.A., Panahi, H. and Shah, I., (2021). Bayesian prediction interval for a constant-stress partially accelerated life test model under censored data. Journal of Taibah University for Science, 15(1), pp.1178-1187.

[23] Lone, S.A., Rahman, A. and Tarray, T.A., (2021). Inference for step-stress partially accelerated life test model with an adaptive type-I progressively hybrid censored data. Journal of Modern Applied Statistical Methods, 19(1), p.13.

[24] Ismail, A.A., (2014). Inference for a step-stress partially accelerated life test model with an adaptive Type-II progressively hybrid censored data from Weibull distribution. Journal of Computational and Applied Mathematics, 260, pp.533-542.

[25] Lone, S.A. and Rahman, A., (2019). Designing Accelerated Life Testing for Product Reliability Under Warranty Prospective. Bayesian Analysis and Reliability Estimation of Generalized Probability Distributions, p.68.

[26] Aly, H.M., Bleed, S.O. and Muhammed, H.Z., (2020). Inference and optimal design of accelerated life test using geometric process for generalized half-logistic distribution under progressive type-II censoring. Journal of Data Science, pp.361-379.

[27] Alam, I., Ahmad, H.H., Ahmed, A. and Ali, I., (2022). Inference on adaptive progressively hybrid censoring schemes under partially accelerated life test for OLiHL distribution. Quality and Reliability Engineering International.

[28] Sindhu, T.N., Aslam, M. and Hussain, Z., (2016). A simulation study of parameters for the censored shifted Gompertz mixture distribution: A Bayesian approach. Journal of statistics and management systems, 19(3), pp.423-450.

[29] Nassr, S.G., Almetwally, E.M. and El Azm, W.S.A., (2021). Statistical inference for the extended weibull distribution based on adaptive type-II progressive hybrid censored competing risks data. Thailand Statistician, 19(3), pp.547-564.

[30] Hemmati, F. and Khorram, E., (2013). Statistical analysis of the log-normal distribution under type-II progressive hybrid censoring schemes. Communications in Statistics-simulation and Computation, 42(1), pp.52-75.

[31] Hussam, E., Alharbi, R., Almetwally, E.M., Alruwaili, B., Gemeay, A.M. and Riad, F.H., (2022). Single and multiple ramp progressive stress with binomial removal: practical application for industry. Mathematical Problems in Engineering, 2022.

RELIABILITY ANALYSES OF AN INDUSTRIAL SYSTEM BASED ON HERMITE POLYNOMIAL AND 2-PARAMETER WEIBULL

Anas Sani Maihulla
•
Department of Mathematics, Sokoto State University, Sokoto- Nigeria
anas.maihulla@ssu.edu.ng

Michael Khoo Boon Chong
•
School of Mathematical Sciences, University Science Malaysia
mkbc@usm.my

Ibrahim Yusuf
•
Department of Mathematical sciences, Bayero University, Kano- Nigeria
iyusuf.mth@buk.edu.ng

Abstract

The current study aims to assess the structure's reliability using stochastic Hermite surface methodology. This approach uses series expansion of standard normal random variables to model uncertainty. (i.e., polynomial chaos expansion). The coefficients of the polynomial chaos expansion are found through stochastic collocation, which only requires a few performance function evaluations. After determining the order of the polynomial and its coefficients, first order methods calculates the reliability index. To demonstrate the applicability of the suggested based reliability analysis, failure rates were used and the duration for the evaluated were set to be 360 days. Numerical result are provided on monthly basics. On the other hand, to achieve our goal, we proposed the 2-parameter modified Weibull distribution. The simulation was performed using Maple software. The evaluation for each subsystem was displayed in the result and analyses section. The conclusion, however, draws a broad conclusion about the study.

Keywords: Reliability, Variables, Polynomials, Stochastic, Simulation, Distributions

I. Introduction

Derivatives are challenging to assess. For reliability analysis in this situation, Bucher and Bourgand [1] developed the response surface method (RSM). Near the failure region, a multidimensional quadratic polynomial in RSM is used to approximate the unknown limit surface. This method has recently been widely used by engineers and researchers for a variety of applications, including performance evaluation, crash simulation, and reliability-based design optimization. However, since the failure region is close to the polynomial approximation of the original limit state, it is frequently challenging to determine the ideal separation for limit states with multiple design points.Additionally, because this is a deterministic representation, it is unable to accurately depict

the stochastic features of the initial limit state. The Askey-Wilson polynomials, which share the characteristics of more conventional orthogonal polynomials like the Legendre polynomials or the spherical harmonics, are the most general orthogonal polynomials [2].

A new dynamic reliability assessment method based on the decomposition method and the Hermite polynomials approximation is presented for the dynamic reliability analysis of stochastic structures under stationary random excitation. In this method, a multi-dimensional dynamic reliability response function is additively broken down into a one-dimensional function, and the one-dimensional dynamic reliability response function is then approximated using Hermite polynomials. Last but not least, the Maple software simulation yields the unconditional reliability of the explicit response function, and two expressions show the logic of the suggested approach. A systematic account of analytical calculus for distributions can be found in Potthoff-Yan [3]. Ordinary differentiation and integration can be thought of as being generalized to arbitrary order in the calculus of fractional order. The uncertainty of the rational order of the derivation gave rise to G.W [4]. Leibniz's question, which gave rise to the fractional calculus, in 1695 [5]. Recent years have seen a significant increase in interest in both theory and application for Jacobi polynomials. For the purpose of solving the multi-term fractional differential equations, the shifted Jacobi operational matrix of fractional derivatives and the spectral tau approach are combined. [6]. To calculate multi-term FDEs, a new explicit solution is created that is targeted for shifted Chebyshev polynomials with flexible degree and fractional order. The work in [7] can be used to solve the same linear problem. Laguerre polynomials have been used in a number of attempts to solve FDEs using various spectral methods within the realm of numerical methods [8]. The research done in [9] led to the creation of a novel time-dependent problem-targeting algorithm built on spectral Laguerre approximations. A novel tau method was proposed in a recent paper that examined the modified Laguerre functions [10]. Comparative analyses of Weibull distribution with Reliability, Availability, Maintainability, and Dependability analyses of an industrial system was conducted [14]. The reliability analyses of filtration system using Copula approach was conducted [15]. Complex reverse osmosis system was studied by [16] and [17] using RAMD analyses. And that of Photovoltaic using the same RAMD analyses was conducted by [18].

There is no comparison of the industrial system's performance using the Hermite polynomial and the two-parameter Weibull distribution in the literature currently available, this prompt our research. Additionally, notice how an industrial process innovation is developing.

II. Methods

Initial value conditions of multi-term fractional differential equations serve as a driving force behind many practical issues. The Hermite tau method is modified in this section to include the operational matrix in order to solve fractional differential equations. The following lists every step in the entire process.

2.2. The Properties of Hermite Polynomials

Let $\alpha = (-\infty, \infty)$ and $\gamma(l) = e^{-l^2}$be the reliability function on α. Hermite polynomials of degree m are defined in their analytical form [11]

$$Rm(l) = \sum_{d=0}^{\lfloor \frac{m}{2} \rfloor} \frac{(-1)^k m!(2l)^{m-2d}}{d!(m-2d)!} \quad (1)$$

where $R_0(l) = 1$ and $R_1(l) = 2l$

This recurrence relation is satisfied by hermite polynomials.

$$R_{m+1}(l) = 2lRm(l) + 2mR_{m-1}(l) \quad (2)$$

An orthogonal system is one in which the set of Hermite polynomials is an orthogonal polynomial

$$\int_{-\infty}^{\infty} Rm(l)R_j(l)\gamma(l)dx = h_j\delta_{mj} \quad (3)$$

where $\delta_{mj} = 0 \; \forall \; m \neq j$ and $\delta_{mj} = m \; \forall \; m = j$ describes the role of Kronecker and $h_j = 2^j j!\sqrt{\pi}$

2.3. Fractional integration hermite operational reliability

We aim to derive a modified reliability equation of an industrial system for Hermite polynomials in this section.

Let $\lambda(l) \in L^2(\alpha)$, subsequently, $\lambda(l)$ can be defined as follows using Hermite polynomials

$$\lambda(l) = \sum_{j=0}^{\infty} a_j R_j(l) \quad (4)$$

Then, coefficient a_j can be written as

$$a_j = \frac{1}{2^j j!\sqrt{\pi}} \int_{-\infty}^{\infty} \lambda(l)R_j(l)\gamma(l)dl\,, \quad j = 0,1,\dots\,. \quad (5)$$

The initial $(N+1)$ the only consideration is given to terms of Hermite polynomials, such that the modified failure rate of the industrial system is in the equation below:

$$\lambda_N(l) = \sum_{j=0}^{N} a_j R_j(l) = A^T \boldsymbol{\rho}(l) \quad (6)$$

where

$$A^T = [a_0 \quad a_1 \quad \dots \quad a_N] \text{ and } \phi(l) = [R_0(l) \quad R_1(l) \quad \dots \quad R_N(l)]^T \quad (7)$$

Integration of Hermite vector $\boldsymbol{\rho}(x)$ by $J^q\boldsymbol{\rho}(x)$ the system reliability in table 1 is obtained from the following equation.

$$J^q\phi(l) = \boldsymbol{P}^{(q)}\boldsymbol{\rho}(l) \quad (8)$$

where q indicates a fined integer value and $\boldsymbol{P}^{(q)}$ represents the actual operational reliability of integrated $\boldsymbol{\rho}(l)$.

2.4. For the reliability analyses using 2-parameter Weibull distribution

It is clearly stated by Quek and Ang [12]. If the lifetimes follow the Weibull distribution, the p.d.f.

$$f(t) = \beta\alpha^{-\beta}t^{\beta-1}e^{-(\frac{t}{\alpha})^{\beta}} \quad (9)$$

Where α and β are the scale and shape parameter of the Weibull distribution respectively.
The size of the units in which the random variable, t, is measured is reflected by the scale parameter, α. The distribution's form changes depending on the shape parameter,
β. We can create a diverse set of curves that reflect real lifetime failure distributions by modifying the value of β.
From (1), the Cumulative Distribution Function is given by:

$$F(t) = 1 - e^{-(\frac{t}{\alpha})^{\beta}} \quad (10)$$

From the relation,

$$R(t) = 1 - F(t) \quad (11)$$

We can substitute (2) into (3), and we have:

$$R(t) = e^{-(\frac{t}{\alpha})^{\beta}} \quad (12)$$

Where $R(t)$ is the reliability or survival function.

The failure rate function or the hazard function can therefore be derived from the following relation:

$$h(t) = \frac{f(t)}{R(t)} \tag{13}$$

Substituting (9) and (12) into (13) we have:

$$h(t) = \frac{\beta\alpha^{-\beta}t^{\beta-1}e^{-(\frac{t}{\alpha})^{\beta}}}{e^{-(\frac{t}{\alpha})^{\beta}}} \tag{14}$$

$$h(t) = \beta\alpha^{-\beta}t^{\beta-1} \tag{15}$$

If we consider M_0 to be the number of hours in a year (the size of the population). Out of which M_s units (the number of hours that the system is upstate) survive the test. While M_f fail, then reliability function $R(t)$ is given by:

$$R(t) = \frac{M_s}{M_0} = \frac{M_0 - M_f}{M_0} \tag{16}$$

Differentiating both sides of (16) and taking M_0 fixed, the following equation will result

$$\frac{dR(t)}{dt} = \frac{1}{M_0}\frac{dM_f}{dt} \tag{17}$$

The rate at which component fails can therefore be defined as:

$$\frac{dM_f}{dt} = -M_0\frac{dR(t)}{dt} \tag{18}$$

Dividing both sides of the above equation byM_s, we obtain the instantaneous probability $g(t)$ of failure, this is:

$$g(t) = \frac{1}{M_S}\frac{dM_f}{dt} = -\frac{M_0}{M_S}\frac{dR(t)}{dt} \tag{19}$$

Using equation (16) into (19) we get:

$$g(t) = -\frac{1}{R(t)}\frac{dR(t)}{dt} \tag{20}$$

Integrating both sides of the equation (20) we have:

$$\int g(t)dt = -\log R(t) \tag{21}$$

From the above equation, the $R(t)$ will be:

$$R(t) = e^{-\int_0^x g(t)dt} \tag{22}$$

Where x is variable.

The function $g(t)$ is called the hazard function or failure rate. Equation (22) can be considered as a generic expression of failure as it is applicable to both exponential and non-exponential failure distribution.

For our modified Weibull distribution, we compared (22) with (12)

$$(\frac{t}{\alpha})^{\beta} = \int_0^x g(t)dt \tag{23}$$

For,

$g(t) = -\frac{1}{R(t)}\frac{dR(t)}{dt}$, and the below facts:

The derivative of a definite integral of a function is the function itself only when the lower limit of the integral is a constant and the upper limit is the variable with respect to which we are differentiating. To summarize Ulrich et al. [13]:

- The derivative of an indefinite integral of a function is the function itself. i.e., $\frac{d}{dx}\int f(x)dx = f(x)$
- The derivative of a definite integral with constant limits is 0. i.e., $\frac{d}{dx}\int_a^b f(x)dx = 0$
- The derivative of a definite integral where the lower limit is a constant and the upper limit is a variable is a function itself in terms of the given variable (upper bound).
  i.e., $\frac{d}{dx}\int_a^x f(x)dx = f(x)$ where 'a' is a constant and 'x' is a variable.
  $$(\frac{t}{\alpha})^{\beta} = -\frac{M_0}{M_S}(R(t)) \tag{24}$$
  The NMW proposed for the present research follows;

$$R(t) = -\frac{M_S}{M_0}(\frac{t}{\alpha})^{\beta} \tag{25}$$

III. Results

The numerical values for the 2-parameter Weibull distribution and Hermite polynomial derived equations were presented in this section. For the two approaches, comparative analyses were conducted. 360 days were allotted for the analyses using the failure rate range of 0.75 to 0.95.

Table 1: *Hermite Polynomials reliability with time*

Time↓ X→	0.75	0.85	0.95	Exp(-m)
30	0.9483	0.9385	0.8383	0.7653
60	0.9245	0.9134	0.8134	0.7446
90	0.9188	0.9076	0.8075	0.7146
120	0.9176	0.8962	0.8056	0.7087
150	0.9171	0.8759	0.7946	0.6801
180	0.9148	0.8451	0.7676	0.6479
210	0.9117	0.8153	0.7373	0.6178
240	0.9092	0.8009	0.7187	0.5929
270	0.8977	0.7852	0.7089	0.5708
300	0.8832	0.7657	0.6806	0.5479
330	0.8633	0.7352	0.6735	0.5188
360	0.8558	0.7143	0.6566	0.5070

Table 2: *2-Parameter Weibull distribution's reliability with time*

Time↓ α→	0.75	0.85	0.95	Exp(-m)
30	0.9954	0.7854	0.5855	0.3667
60	0.9857	0.7789	0.5734	0.3584
90	0.9787	0.7607	0.5677	0.3499
120	0.9644	0.7525	0.5537	0.3312
150	0.9567	0.7457	0.5344	0.3290
180	0.9439	0.7346	0.5270	0.3043
210	0.9387	0.7277	0.5145	0.2932
240	0.9298	0.7190	0.5044	0.2823
270	0.9124	0.6932	0.4944	0.2732
300	0.9097	0.6854	0.4797	0.2617
330	0.8934	0.6724	0.4653	0.2476
360	0.8862	0.6656	0.4522	0.2332

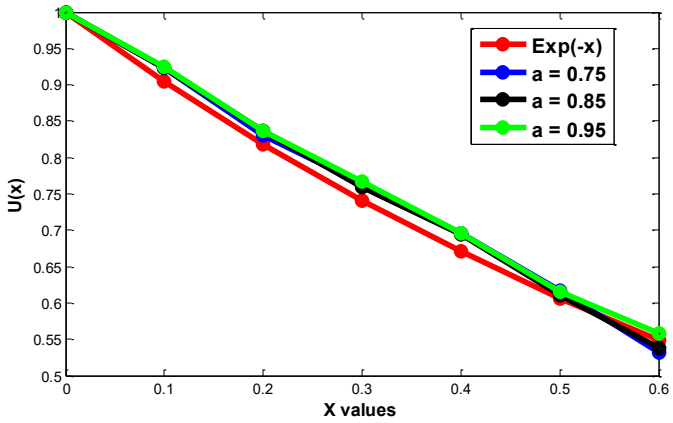


Figure 1: *Reliability analyses For failure against repair rate of a system using Hermite Polynomial*

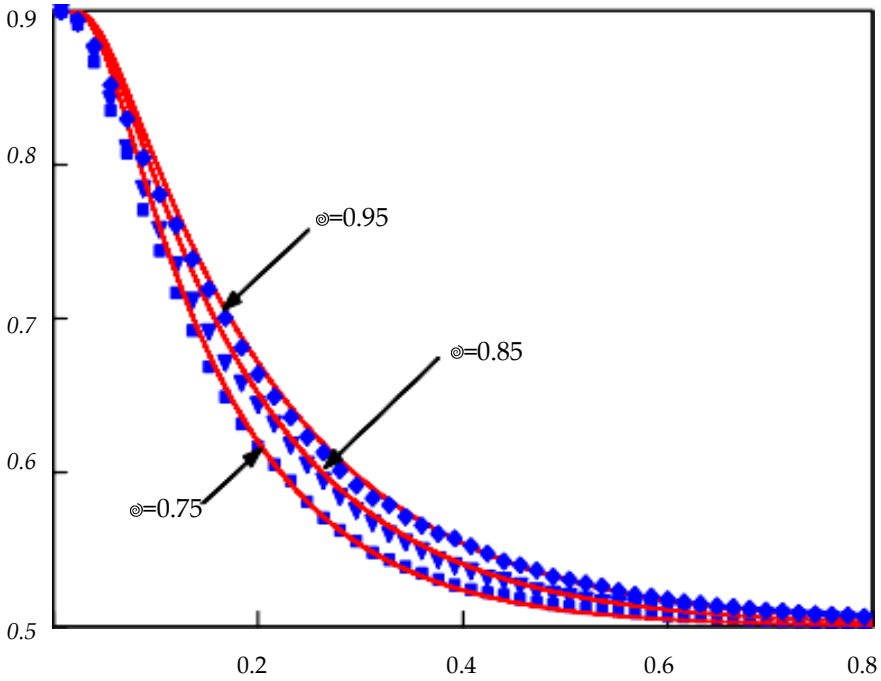


Figure 2: *Reliability analyses For failure against repair rate of a system using 2-parameter Weibull distribution*

IV. Discussion

Using the Hermite polynomial method, the result shown in table 1 that corresponds to Figure 1 demonstrated how the reliability of an industrial system is impacted by the rise in failure rate. The reliability of an industrial system was then shown to be affected as the failure rate was changed from 0.75 to 0.95 in table 2 that corresponds to Figure 2. When comparing the results in figures 1 and 2, it is obvious that the 2-parameter Weibull yields a more accurate result than the Hermite polynomials in terms of reliability estimation. Therefore, it is advised that future studies modify the Hermite polynomial to increase the dependability of industrial systems.

Acknowledgement

I recommended the effort by Tertiary Education Trust Fund (TETFund) for their kindness to sponsorship for Ph.D research as well as the benchwork here in Malaysia.

References

[1] Bucher, C.G. Bourgund, U. (1990). A fast and efficient response surface approach for structural reliability problems. *Struct Saf* 7:57–66

[2] Szego, G. (1975). Orthogonal polynomials. *4th ed., Amer. Math. Soc, Providence, RI*

[3] Yan, A. (1994). Linear stochastic differential equations and Wick products, Probability.*Theory Rei. Fields*, Vol. 99, 501-526

[4] Abramowitz, M. and I. S. Stegun, I. S. (1965). Handbook of Mathematical Functions." *Dover Publications, Inc.*, New York, NY.

[5] Carpinteri, A. and Mainardi, F. (1997). Fractals and Fractional Calculus in Continium Mechanics. *Springer-Verlag Wien New-York.*

[6] Doha, E.H. Bhrawy, A. H. and Ezz-Eldien, S.S. (2013). A New Jacobi Operational Matrix: An application for Solving Fractional Differential Equations. *Appl. Math. Modell*, 36:4931-4943. https://doi.org/10.1016/j.apm.2011.12.031.

[7] Bhrawy A. H. and Alofi, A. S. (2013). The Operational Matrix of Fractional Integration for Shifted Chebyshev Polynomials. *Applied Mathematics Letters*, 26:25-31. https://doi.org/10.1016/j.aml.2012.01.027.

[8] Guo, B. Y. and Zhang, X. Y. (2005). A New Generalized Laugerre Approximation and Its Applications. *J. Of Comput. App. Math*, 181:342-363. https://doi.org/10.1016/j.cam.2004.12.008.

[9] Mikhailenko, B. G. (1999). Spectral Laugerre Method for the Approximate Solution of time dependent problems. *Applied Mathematics Letters*, 12:105-110. https://doi.org/10.1016/S0893-9659(99)00043-9.

[10] Bhrawy, A. H. Alghamdi M. M. and Taha, M. T. (2012). A New Modified Generalized Laugerre Operational Matrix of Fractional Integration for Solving Fractional Differential Equations on the Half Line Vol. 179. *https://doi.org/10.1186/1687-1847- 2012-179.*

[11] A.D Poularikas, , (1999) "The Handbook of Formulas and Tables for Signal Processing, Hermite Polynomials."

[12] Quek, S. T. and Ang A. H. S. (1986). Structural System Reliability by The Method of Stable Configuration. *National Science Foundation, University of Illinois* , ISSN: 0069-4274, NO. 529, Washington.

[13] Ulrich, L. Rohde, G. C. Jain, A. K. Poddar, G. (2011). Introduction to Differential Calculus. *Systematic Studies with Engineering Applications for Beginners* ISBN:9781118117750 |DOI:10.1002/9781118130155

[14] Anas, S. M. Ibrahim Y. Saminu I. B. (2023). WEIBULL COMPARISON BASED ON RELIABILITY, AVAILABILITY, MAINTAINABILITY, AND DEPENDABILITY (RAMD) ANALYSIS. *Reliability: Theory & Applications. March 1(72):*120-132. https://doi.org/10.24412/1932-2321- 2023-172-120-132

[15] Maihulla, A. S. and Yusuf, I. (2022). RELIABILITY ANALYSIS OF REVERSE OSMOSIS FILTRATION SYSTEM USING COPULA. *Reliability: Theory & Applications*, *17*(2): 163-177. Retrieved from http://www.gnedenko.net/RTA/index.php/rta/article/view/890.

[16] Maihulla, A. S. Yusuf, I. and Bala, S. I. (2021). Performance Evaluation of a Complex Reverse Osmosis Machine System in Water Purification using Reliability, Availability, Maintainability and Dependability Analysis. *Reliability: Theory & Applications*, *16*(3):115-131. Retrieved from http://www.gnedenko.net/RTA/index.php/rta/article/view/764

[17] Maihulla, A. Yusuf, I. (2022). Reliability modeling and performance analysis of reverse osmosis machine in water purification using Gumbel–Hougaard family copula," *Life Cycle Reliab Saf Eng* . https://doi.org/10.1007/s41872-022-00214-2

[18] Maihulla, A. S. and Yusuf, I. (2021). Performance Analysis of Photovoltaic Systems Using (RAMD) Analysis. *Journal of the Nigerian Society of Physical Sciences*, *3*(3): 172–180. https://doi.org/10.46481/jnsps.2021.194

COMPLETELY RANDOMIZED DESIGN IN FUZZY OBSERVATIONS

Kirthik VairaMariappan A[1] and Manigandan P[2]

•

[1]Department of Statistics, Government Arts College, Dharmapuri - 5, Tamil Nadu, India.
maryaaindia@gmail.com
[2]Department of Statistics, Periyar University, Salem - 11, Tamil Nadu, India.
srimanigandan95@gmail.com

Abstract

The real world is vague, unclear and full of ambiguity, and are inevitable. The classical statistics disregards the extreme, aberrant, uncertain values, and hence a new appropriate tool had to surface. The Analysis of Variance (ANOVA) method is used to compare the response variable's means between several groups that are specified by the factor variable. Another method of data analysis offered by ANOVA is one that is based on statistics and is experimental design-driven, or Design of Experiment (DOE). In DOE, there are single and two-factor experimental designs depending on, observing the effect of number of factor(s) on output variable as a primary interest. Among all the single factor experimental designs, Completely Randomized Design (CRD) is the simplest and flexible design. In this design, treatments are randomly allocated to the experimental units over the entire experimental material. Each treatment is repeated to increase the efficiency of the design. CRD is more appropriate to use when the data is homogenous. The objective that deals with the preparation and analysis of experiments is experimental design. The treatments are apportioned to the exploratory units at random in the fully randomized experimental design. When the observed data are fuzzy observations rather than precise numerical values, the CRD is expanded in this study. In this paper, an innovative Triangular Fuzzy Number (TFN) in the fuzzy Completely Randomized Design (FCRD) analysis statistical method for evaluating CRD model hypotheses on fuzzy data is presented. To convert the fuzzy totally randomized design model into two crisps CRD models using the suggested way, and then convert to lower and upper models are used in fuzzy hypothesis. Determine the fuzzy hypothesis for the fuzzy CRD model based on the hypotheses of the two crisp CRD models using the decision rules. The fuzzy test appears to be a competitive tool in circumstances with ambiguous data, particularly linguistic ambiguity because it is more adaptable than the conventional test of significance. This paper presents and illustrates a novel fuzzy triangular number-based approach to fuzzy CRD analysis. This paper also explores how flexible a CRD may be when handling uncertain elements. This study provides an example of a new method for fuzzy CRD analysis employing TFN.

Keywords: Fuzzy Set, CRD, Fuzzy CRD, Triangular Fuzzy Number, Decision Rules

1. Introduction

The Analysis of Variance (ANOVA) was introduced in the 1920s by Prof. R.A. Fisher. This method can be used to solve the problems of variations, especially in the agricultural sector. The ANOVA has many independent demographic variables and it is a most powerful tool of the test of significance. The significance test in terms of t-distribution is the only adequate procedure to test the significance of the difference between the two-sample means. In such a situation, when three or

more sample means are considered simultaneously, an alternative procedure is needed to test the hypothesis that all samples are taken from the same population. This is called ANOVA. The foundation for experimental designs was laid in 1935 by Prof. R.A. Fisher. The term design of tests is said to be the logical construction of tests, in which the degree of uncertainty can be well defined. The basic principles of experimental designs are randomization, replication, and local control. The local control is the method of increasing efficiency in test designs. A Completely Randomized Design (CRD) means that treatments are assigned to a completely randomized group so that each test unit has the same chance of receiving any one treatment. Since the principle of local control is not used, the CRD is considered simple and the experimental material is observed, but it is seen that the experimental material is not completely homogeneous. It is specifically designed to address mathematical uncertainty and inadequate specification and provide a systematic tool for dealing with the inherent fuzzy of many problems. The word fuzzy means ambiguity (vagueness). Fuzziness occurs when the boundary of information is not clearly cut. In 1965s Lotfi .A. Zadeh introduced fuzzy sets as an extension of the set with classical notations. The classical set theory allows membership of elements in a set of binary terms to be inside. Fuzzy sets theory allows the estimated membership function in intervals [0,1]. Sometimes, agricultural data is not recorded for natural calamities.

Therefore, fuzzy synthesis is the most inevitable. In both cases, the observed variable of the fuzziness often occurs. In the first case, due to technical problems, the response variable cannot be measured properly. So, in this case the data cannot be clearly recorded with the exact numbers and the measurement errors are computed linguistically to justify the required tolerance. The second phenomenon is that the response variable is presented in terms of linguistic forms such as a special linguistic report or variance report. As for his products, they are not counted. In both of the above cases the data can be represented by the concept of fuzzy sets for analyzing the test (Zadeh [23]). An example is cited by H.C. Wu [21], to illustrate in this situation. There are many real-life populations in which imprecise values can be assigned to their experimental outcomes. Some practical reasons may not be accurate for the agricultural observations so that fuzzy sets used and the fuzzy was introduced by Zadeh [22], to represent manipulate data and in order to non-statistical uncertainties. D. Dubois et al. Brett [6], defined any fuzzy numbers as a fuzzy subset of the real line. The symmetric triangular fuzzy approximation was presented by M. Ma et al. [15]. S. Chanas [2], presented a formula for determining approximations of intervals under humming distance. S. Chandrasekaran et al. Tamilmani [3], proposed the arithmetic operations of fuzzy numbers using the alpha-cut interval method. The Triangular Fuzzy Numbers (TFNs) result from addition or subtraction between TFNs results. Therefore, addition and subtraction between fuzzy numbers become a TFNs. Such areas include approximate reasoning, decision making, optimization, control, and so on. R.R. Hocking [11], has been the traditional statistical testing, the sample observations are crisp and a statistical experiment leads to a binary conclusion.

Applying fuzzy set theory to Statistics. K.G. Manton et al. [16], proposed a fuzzy test for testing hypotheses with fuzzy data and fuzzy testing created the acceptance of null and alternative hypotheses. Statistical hypothesis testing for ambiguous data by presenting the notions of pessimism and pessimism by H.C. Wu [20]. We provide decision rules that can be used to accept or reject ambiguous null and alternative hypotheses. The observed values of the classical random variable can be considered an ambiguous number, while the model for the observed values in the linear model. Note that Filsmoser and Viertl [7] and Viertl [17], use a similar idea. The proposed technique, ambiguous data as well, given the vague assumptions of the tests were imprecise data, along with two hypothesis test, replacing CRD models crisp data, that is the lower-level model and the upper-level model, then each CRD hypothesis, testing the crisp data, models and results, getting after using the results obtained in terms of the provisions of the proposed decision of the population receive the decision. In the decision rules of the proposed test technique, we did not

use the confidence, distrust, and h-level set used in the H.C. Wu [21]. In this way, fuzzy numbers are appropriate models for formalizing and manipulating these populations (Gill et al. [9]). According to Kumari et al. [13] the methodology was expanded by introducing a fuzzy regression approach to randomized block designs that takes into account qualitative predictor factors in multiple linear regression. The concept from this study was a helpful thread for developing thorough connectedness between regression and randomized block designs. According to the researchers, fuzzy MLR can predict far more accurately than MLR alone. By comparing the RMSEs from various forecasting techniques, Koul et al. [12] suggested a study to identify the variance analysis experimental model approach between stock exchange trends. In this paper, we introduce a new technique using triangular fuzzy numbers in the fuzzy CRD analysis with an example.

2. Preliminaries

2.1 Completely Randomized Design (CRD)

CRD is the basic single factor design. In this design the treatments are assigned completely at random so that each experimental unit has the same chance of receiving any one treatment. But CRD is appropriate only when the experimental material is homogeneous. As there is generally large variation among experimental plots due to many factors CRD is not preferred in field experiments. In laboratory experiments and greenhouse studies it is easy to achieve homogeneity of experimental materials and therefore CRD is most useful in such experiments.

2.2 Triangular Fuzzy Number

The triangular fuzzy number membership function is defined by

$$\mu_{\tilde{A}}(x)=\begin{cases}\dfrac{x-a}{b-a} & ;\ a\le x\le b\\[2ex] \dfrac{x-c}{b-c} & ;\ b\le x\le c\end{cases}$$

Where a is indicate lower point, b is indicate centre point and c is indicate upper point.

A Triangular fuzzy number can be represented as an interval number form as follows.

$$\tilde{A}=\left[\{(b-a)r+a\}^{L};\{-(d-c)h+c\}^{U}\right];\ 0\le r,h\le 1.$$

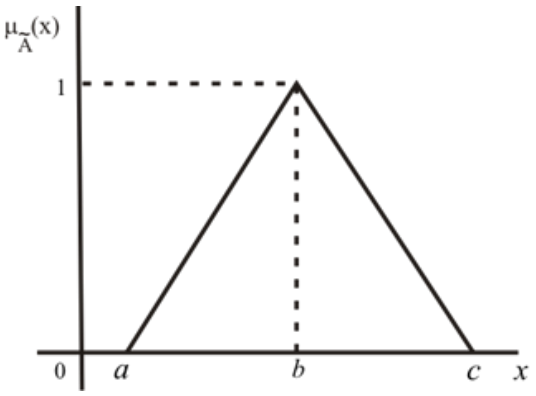


Figure 1: *Figure Triangular Fuzzy Numbers*

Note that r is the level of pessimistic value and h is the level of optimistic value of the fuzzy number $\tilde{A}=(a,b,c)$.

3. Statistical Analysis of CRD

The *CRD* is the one in which all the experimental units are taken in a single group which are homogeneous as far as possible. Suppose there are t treatments in an experiment. Let i^{th} treatment be replicates n_i times then, the total number of experimental units in the design is $\sum_{i=1}^{t} n_i = N$. Then, the treatment is allocated at random to entire experimental area. In this design provides a one-way classified data with different levels of a single factor is called treatments. For instance, y_{ij} can be the productivity of the j^{th} week in the i^{th} varieties, or the paddy seedling of the j^{th} week grown of the j^{th} type of shelf display. Since the number of cases or trials for the i^{th} factor level is denoted by N, so, $j = 1,2,...,n_i$. Now, the Statistical analysis of *CRD* is analogue to *ANOVA* one-way classified data, linear model becomes

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \; ; i = 1,2,...,t \; ; j = 1,2,...,n_i \tag{1}$$

In which, y_{ij}'s is the j^{th} observations of the i^{th} treatment; μ is the general mean effect which is fixed; α_i is the fixed effect due to the i^{th} treatment and ε_{ij} is the random error effect which distributed as normal $N \square (0, \sigma^2) ; i = 1,2,...,t$ and $j = 1,2,...,n_i$.

The grand total of *n* observations of *CRD* is $\sum_{i=1}^{t}\sum_{j=1}^{n_i} y_{ij} = y.. = G$; the correction factor is $cf = \frac{y_{..}^2}{N}$ and the i^{th} treatment total taken is $\sum_{i=1}^{t}\sum_{j=1}^{n_i} y_{ij} = y_{i.} = T_i$.

Apply the *ANOVA* for one way classify and compute the total sum of squares (sst), the treatment sum of squares $(sstr)$ and the error sum of squares (sse) are given below;

$$Q_{sst} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}..)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N} \tag{2}$$

$$Q_{sstr} = \sum_{i=1}^{r} n_i(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{r} \frac{y_{i.}^{\;2}}{n_i} - \frac{y_{..}^{\;2}}{N} \tag{3}$$

and

$$Q_{sse} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} y_{ij}^{\;2} - \sum_{i=1}^{r} \frac{y_{i.}^{\;2}}{n_i} \tag{4}$$

Where, sst, $sstr$ and sse which has $(N-1)$, $(t-1)$ and $(N-t)$ degrees of freedom (df), respectively. The mean sum squares are obtained as follows:

$$msstr = \frac{sstr}{t-1} \text{ and } msse = \frac{sse}{N-t} \tag{5}$$

Where, $msstr$ and $msse$ stands for treatment mean square and error mean square.

In order to test whether or not the factor level means μ are equal, the following classical testing hypotheses are considered.

$$H_0 : \mu_1 = \mu_2 = ... = \mu_t \text{ Vs } H_1 : \mu_1 \neq \mu_2 \neq ... \neq \mu_t$$

The test statistic to be used is

$$F = \frac{msstr}{msse} \square F_{(t-1),(N-1)} \tag{6}$$

When the null hypothesis H_0 holds true, it is known that F is distributed as with degrees of freedom $(t-1)$ and $(N-t)$ that is $F_{(t-1),(N-t)}$.

All these values are referring in the *ANOVA* table and inference is drawn.

Table 1: *ANOVA table for CRD*

sv	*df*	*ss*	*mss*	$\tilde{F}-ratio$
Between Treatmen	$(t-1)$	Q_{sstr}	$msstr$	$F=\frac{msstr}{msse}$
Within Treatmen	$(N-t)$	Q_{sse}	$msse$	
Total	$(N-1)$	Q_{sst}		

Decision Rule:

The decision rules in the F test to accept or reject the null hypothesis and alternative hypothesis are the level of significance α is given by

(i) If $msstr > msse$ and $\tilde{F}_c=\frac{msstr}{msse}<\tilde{F}_t$ where $\tilde{F}_t$ and $\tilde{F}_c$ is the tabulated and calculated values of $\tilde{F}$ with $(t-1)\,(N-t)$, degrees of freedom at α level of significance, then we accept the null hypothesis $\tilde{H}_0$, otherwise the alternative hypothesis $\tilde{H}_1$ is accepted.

(ii) If $msstr < msse$ and $\tilde{F}_C=\frac{msse}{msstr}<\tilde{F}_T$ where $\tilde{F}_t$ and $\tilde{F}_c$ is the tabulated and calculated values of $\tilde{F}$ with $(N-t)\,(t-1)$, degrees of α level of significance, then we accept then null hypothesis $\tilde{H}_0$, otherwise the alternative hypothesis $\tilde{H}_1$ is accepted.

3.1 Statistical Analysis of Fuzzy CRD

In this real-world, sometimes agricultural data cannot be accurately recorded. For example, the growth of seeds grown in a field due to fluctuation cannot be exactly measured. Therefore, the fuzzy set theory provides an appropriate tool for processing naturally imprecise data. Under this consideration, the more appropriate way to describe the paddy seedlings level is to say that the initial stage paddy seedlings are around 10 centimeters. The phrase about 10 centimeters should be considered an ambiguous number, which is realized by the fuzzy set theory. Therefore, our aim is the statistical analysis of fuzzy *CRD* using the *TFNs* method. In this case, observations and recorded data are treated as *TFNs*. Statistical hypotheses and populations parameter are crisp and hence the linear model is considered as $\tilde{y}_{ij}=\mu+\alpha_i+\varepsilon_{ij}$; in which $\tilde{y}_{ij}$'s is the j^{th} observations of the i^{th} treatment; μ is the general mean effect which is fixed; α_i is the fixed effect due to the i^{th} treatment and ε_{ij} is the random error effect which distributed as normal $N\square(0,\sigma^2);i=1,2,...,t$ and $j=1,2,...,n_i$.

Statistical hypotheses are considered as classical ones:

$$\tilde{H}_0=\tilde{\mu}_1=\tilde{\mu}_2=...=\tilde{\mu}_t \text{ Vs } \tilde{H}_1:\tilde{\mu}_1\neq\tilde{\mu}_2\neq...\neq\tilde{\mu}_t.$$

But one point that deviates from the classical *ANOVA* assumptions in the linear model is that the sample observations did not change anything else in the *CRD* model before collecting *TFNs* and data rather than actual numbers. Regarding the fuzzy arithmetic of *TFNs* described in the observed values of statistics for simplicity of calculations, Zadeh's [20] fuzzy extension principle can be explained lower level and upper-level model as follows:

$$\tilde{Q}_{ssro}^{L}=\sum_{i=1}^{t}\sum_{j=1}^{n_i}\tilde{y}_{ij}^{2};\tilde{Q}_{G}^{L}=\tilde{y}..=\sum_{i=1}^{t}\tilde{y}_{i.}\;;\;cf=\frac{\tilde{y}_{..}^{2}}{N};\tilde{y}_{i.}=\sum_{j=1}^{n_i}\tilde{y}_{ij} \tag{7}$$

$$\tilde{Q}_{sst}^{L}=\sum_{i=1}^{t}\sum_{j=1}^{n_i}\tilde{y}_{ij}^{2}-\frac{\tilde{y}_{..}^{2}}{N} \tag{8}$$

$$\tilde{Q}_{sstr}^{L}=\sum_{i=1}^{t}\frac{\tilde{y}_{i.}^{2}}{n_i}-\frac{\tilde{y}_{..}^{2}}{N} \tag{9}$$

and

$$\tilde{Q}_{sse}^{L}=\tilde{Q}_{sst}^{L}-\tilde{Q}_{sstr}^{L} \tag{10}$$

Table 2: *ANOVA table for lower level Fuzzy CRD*

sv	*df*	*ss*	*mss*	$\tilde{F}-ratio$
Between Treatmen	$(t-1)$	$\tilde{Q}_{sstr}^{L}$	$msstr^{L}$	$\tilde{F}^{L}=\frac{msstr}{msse}$
Within Treatmen	$(N-t)$	$\tilde{Q}_{sse}^{L}$	$msse^{L}$	
Total	$(N-1)$	$\tilde{Q}_{sst}^{L}$		

$$\tilde{Q}_{ssro}^{U}=\sum_{i=1}^{t}\sum_{j=1}^{n_i}\tilde{y}_{ij}^{2};\tilde{Q}_{G}^{U}=\tilde{y}..=\sum_{i=1}^{t}\tilde{y}_{i.}\;;\;cf=\frac{\tilde{y}_{..}^{2}}{N};\tilde{y}_{i.}=\sum_{j=1}^{n_i}\tilde{y}_{ij} \tag{11}$$

$$\tilde{Q}_{sst}^{U}=\sum_{i=1}^{t}\sum_{j=1}^{n_i}\tilde{y}_{ij}^{2}-\frac{\tilde{y}_{..}^{2}}{N} \tag{12}$$

$$\tilde{Q}_{sstr}^{U}=\sum_{i=1}^{t}\frac{\tilde{y}_{i.}^{2}}{n_i}-\frac{\tilde{y}_{..}^{2}}{N} \tag{13}$$

and

$$\tilde{Q}_{sse}^{U}=\tilde{Q}_{tss}^{U}-\tilde{Q}_{sstr}^{U} \tag{14}$$

Table 3: *ANOVA table for upper level Fuzzy CRD*

sv	*df*	*ss*	*mss*	$\tilde{F}-ratio$
Between Treatmen	$(t-1)$	$\tilde{Q}_{sstr}^{U}$	$msstr^{U}$	$\tilde{F}^{U}=\frac{msstr}{msse}$
Within Treatmen	$(N-t)$	$\tilde{Q}_{sse}^{U}$	$msse^{U}$	
Total	$(N-1)$	$\tilde{Q}_{sst}^{U}$		

3.2 Fuzzy Decision Rules of $\tilde{F}$ -Test

Suppose that if at α level of significance, the null hypothesis of the lower level model is accepted for $0\le h\le\tilde{F}_t$ where $0\le\tilde{F}_t\le 1$ and the null hypothesis of the upper level model is accepted for $0\le r\le\tilde{F}_t$ where $0\le\tilde{F}_t\le 1$ then, the fuzzy null hypothesis of the fuzzy ANOVA model is accepted for and at α level of significance. Otherwise, the fuzzy alternative hypothesis of the fuzzy ANOVA model is accepted at α level of significance.

4. Applications

Following application to each of the three types of paddy in a CRD, the yield in kilograms (kgs.) per four plots. Due to some unforeseen circumstances, it is impossible to record the precise amount of yields in kgs. in a sample; nonetheless, there is some fuzzy information available. Below are the triangular fuzzy data:

Table 4: *Fuzzy CRD table forTFNs*

Yields in kgs. (*i*)	Varieties of paddy (*j*)		
	V1	V2	V3
Y1	4,6,8	6,8,10	-
Y2	5,7,10	4,6,8	6,8,10
Y3	7,9,11	8,10,12	9,12,14
Y4	5,9,12	7,9,11	-

Test that there is a significant difference in the varieties of paddy performance of the yields in kgs. per plots.

Let $\tilde{\mu}_i$ be the mean number of varieties of paddy for the i^{th} yields in kgs. per plots.

Now, the null hypothesis, $\tilde{H}_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3 = \tilde{\mu}_4$ and the alternative hypothesis, $\tilde{H}_A$: not all $\tilde{\mu}_i$'s are equal.

Now, the ANOVA model for " *r* is the lower level of pessimistic value" and " *h* is the upper level of optimistic value" the interval model for the triangular fuzzy number is given below:

Table 5: *Fuzzy CRD table for lower and upper level models*

Yields in kgs. (*i*)	Varieties of paddy (*j*)		
	V1	V2	V3
Y1	$2r+4,8-2h$	$2r+6,10-2h$	-
Y2	$2r+5,10-3h$	$2r+4,8-2h$	$2r+6,10-2h$
Y3	$2r+7,11-2h$	$2r+8,12-2h$	$9r+3,14-2h$
Y4	$4r+5,12-3h$	$2r+7,11-2h$	-

Table 6: *Fuzzy CRD table for lower level model*

Yields in kgs. (*i*)	Varieties of paddy (*j*)		
	V1	V2	V3
Y1	$2r+4$	$2r+6$	-
Y2	$2r+5$	$2r+4$	$2r+6$
Y3	$2r+7$	$2r+8$	$3r+9$
Y4	$4r+5$	$2r+7$	-

The null hypothesis $H_0^{LL} : \mu_1^{LL} = \mu_2^{LL} = \mu_3^{LL} = \mu_4^{LL}$ against the alternative hypothesis H_A^{LL}: not all μ_i^{LL}' s are equal.

Here, $N = 10$ and $n_i = 2,3,3,2$ the yields in kgs. per plot and the varieties of paddy for the 1,2,3,4 respectively.

Total sum of squares for lower level model is

$$sst_r^L = 4.1r^2 + 1.4r + 24.9$$

Treatment sum of squares of lower level model is

$$sstr_r^L = 1.4r^2 + 3.4r + 16.9$$

Error sum of squares of lower level model is

$$sse_r^L = 2.7r^2 - 2r + 8$$

MSTR and *MSE* lower level model is

$$msstr_r^L = 0.47r^2 + 1.13r + 5.63 \text{ and } msse_r^L = 0.45r^2 - 0.33r + 1.33$$

$\tilde{F}$ – Ratio of lower level model is

$$\tilde{F}_r^L = \frac{0.47r^2 + 1.13r + 5.63}{0.45r^2 - 0.33r + 1.33}$$

All these values are referring in *ANOVA* table and inference is drawn.

Table 7: *ANOVA table for Lower Level of Fuzzy CRD*

sv	*df*	*ss*	*mss*	$\tilde{F} - ratio$
Between Treatments	3	$1.4r^2 + 3.4r + 16.9$	$0.47r^2 + 1.13r + 5.63$	$\frac{0.47r^2 + 1.13r + 5.63}{0.45r^2 - 0.33r + 1.33}$
Within Treatments	6	$2.7r^2 - 2r + 8$	$0.45r^2 - 0.33r + 1.33$	
Total	9	$4.1r^2 + 1.4r + 24.9$	-	-

Now, $F_r^L > F_t^L$, for all $r; 0 \le r \le 0.33$ where $F_t^L = 4.76$ is the F table value of α at 5% level of significance with $(3,6)$ degrees of freedom. Therefore, the null hypothesis H_0^L of the lower level model is accepted for the $r; 0 \le r \le 0.33$.

Table 8: *Fuzzy CRD table for upper level model*

Yields in kgs. (*i*)	Varieties of paddy (*j*)		
	V1	V2	V3
Y1	$8-2h$	$10-2h$	-
Y2	$10-3h$	$8-2h$	$10-2h$
Y3	$11-2h$	$12-2h$	$14-2h$
Y4	$12-3h$	$11-2h$	-

The null hypothesis $H_0^{UL}: \mu_1^{UL} = \mu_2^{UL} = \mu_3^{UL} = \mu_4^{UL}$ against the alternative hypothesis H_A^{UL}: not all μ_i^{UL}'s are equal.

Here, $N = 10$ and $n_i = 2,3,3,2$ the varieties of yields in kg. 1,2,3,4 respectively.

Total sum of squares for upper level model is

$$sst_h^U = 1.6h^2 - 1.6h + 30.4$$

Treatment sum of squares of upper level model is

$$sstr_h^U = 0.43h^2 + 0.73h + 20.57$$

Error sum of squares of upper level model is

$$sse_h^U = 1.17h^2 - 2.33h + 9.83$$

MSTR and *MSE* upper level model is

$$msstr_h^U = 0.14h^2 + 0.24h + 6.85 \text{ and } msse_h^U = 0.19h^2 - 0.38h + 1.64$$

$\tilde{F}$ – Ratio of upper level model is

$$\tilde{F}_h^U = \frac{0.14h^2 + 0.24h + 6.85}{0.19h^2 - 0.38h + 1.64}$$

All these values are referring in *ANOVA* table and inference is drawn.

Table 9: *ANOVA table for Upper Level of Fuzzy CRD*

sv	*df*	*ss*	*mss*	$\tilde{F} - ratio$
Between Treatments	3	$0.43h^2 + 0.73h + 20.57$	$0.14h^2 + 0.24h + 6.85$	$\frac{0.14h^2 + 0.24h + 6.85}{0.19h^2 - 0.38h + 1.64}$
Within Treatments	6	$1.17h^2 - 2.33h + 9.83$	$0.19h^2 - 0.38h + 1.64$	
Total	9	$1.6h^2 - 1.6h + 30.4$	-	-

Now, $F_h^U < F_t^U$, for all $h; 0 \le h \le 0.61$ where $F_t^U = 4.76$ is the F table value of 5% at α level of significance with $(3,6)$ degrees of freedom. Therefore, the null hypothesis H_0^U of the upper level model is accepted for the $h; 0 \le h \le 0.61$.

Thus, since the null hypothesis H_0^L and H_0^U of the lower level data and upper level data are accepted for all $r; 0 \le r \le 0.33$ and $h; 0 \le h \le 0.61$ (note that null hypotheses are not rejected at $r = 1$ and $h = 1$, that is the centre level), the fuzzy null hypothesis $\tilde{H}_0$ of the fuzzy ANOVA model is accepted for all $r; 0 \le r \le 0.33$ and $h; 0 \le h \le 0.61$. Thus, we conclude that four yields of kgs. per plots are equal only if $r; 0 \le r \le 0.33$ and $h; 0 \le h \le 0.61$. That is, the maximum level of pessimistic value is 0.33 and the maximum level of optimistic value is 0.61. From the applications, thus observe that the acceptance of the fuzzy null hypothesis for not all *r* and *h* always, but for some specific levels of *r* and *h*, that is $r; 0 \le r \le 0.42$ and $h; 0 \le h \le 0.61$.

5. Conclusion

In this paper, the propose a new statistical fuzzy hypothesis testing of completely randomized design model with the fuzzy data. In the proposed technique, do transfer the fuzzy completely randomized design model into two crisp CRD models. Based on the decisions of hypotheses of two crisp CRD models, to take a decision on the fuzzy hypothesis of the fuzzy CRD model. Since our fuzzy test is more flexible than the traditional test of significance, it seems to be a competitive tool in situations with imprecise data, especially of the linguistic type. Since the proposed technique in this paper is mainly based only on the crisp models, the proposed technique can be extended to the experimental design analysis having fuzzy data and RBD, LSD etc.

References

[1] Buckley J.J. Fuzzy statistics, Springer-Verlag, New York, 2005.

[2] Chanas S. (1999). On the interval approximation of a fuzzy numbers, *Fuzzy sets and*

Systems, 102, 221-226.

[3] Chandrasekaran S. and Tamilmani E. (2015). Arithemetic Operation of Fuzzy Numbers using - cut Method, *International Journal of Innovative Science, Engineering and Technology*, vol. 2, Issue 10.

[4] Cochran W.G. and Cox G.M., Experimental Designs (2nd ed.), New York: Wiley, 1957.

[5] Dubois D. and Prade H., Fuzzy Sets and Systems: Theory and Application, Academic Press, New York, 1980.

[6] Dubois D. and Prade H. (1978). Operations on fuzzy numbers, *Int. J. Syst. Sci.*, 9, 613-626.

[7] Filzmoser P. and Viertl R. (2009). Testing Hypotheses with Fuzzy Data: The Fuzzy P value, Metrika, 59, 21-29.

[8] George.J. Klir and Bo Yuan, Fuzzy sets and fuzzy logic, Theory and Applications, Prentice-Hall, New Jersey, 2008.

[9] Gil et al. (2006). Bootstrap Approach to the Multi-sample Test of Means with Imprecise Data *Computer Statistics and Data Analysis* 51, 148-162.

[10] Gupta S.C. and Kapoor V.K., Fundamentals of applied statistics, Sultan Chand & Sons, New Delhi, India, 2007.

[11] Hocking R.R., Methods and applications of linear models: regression and the analysis of variance, New York: John Wiley & Sons, 1996.

[12] Koul, S., Awasthi, A. K., & Garov, A. K. (2019). Experimental model approach for decision making in Stock Index. *Think India Journal*, *22*(37), 1272-1276.

[13] Kumari, S. (2022). A Study on Neutrosophic Completely Randomised Design. *Mathematical Statistician and Engineering Applications*, *71*(4), 3738-3747.

[14] Ma M. et al. (2000). A new approach for defuzzification, *Fuzzy Sets and Systems*, 111, 351-356.

[15] Manton K.G. et al., Statistical applications using fuzzy sets, New York: John Wiley & Sons, 1994.

[16] Montgomery D.C., Design and Analysis of Experiments (8th ed.), New York: John Wiley & Sons, 1991.

[17] Nguyen H.T. and Walker E.A., A First Course in Fuzzy Logic (3rd ed.), Paris: Chapman Hall/CRC, 2005.

[18] Viertl R. (2006). Univariate statistical analysis with fuzzy data, *Computational Statistics and Data Analysis*, 51, 33-147.

[19] Viertl R., Statistical methods for fuzzy data, John Wiley and Sons, 2011.

[20] Wu H.C., (2005). Statistical hypotheses testing for fuzzy data, *Information Sciences*, 175, 30 -56.

[21] Wu H.C., (2007). Analysis of variance for fuzzy data, *International Journal of Systems Science*, 38, 235-246.

[22] Zadeh L.A., (1965). Fuzzy sets, Information and Control, 8, 338-353.

[23] Zadeh L.A., (1975). The Concept of a Linguistic Variable and its application to approximate reasoning, *Information Sciences*, 8, 199-249.

[24] Zimmermann H.J., Fuzzy set theory and its applications, Academic Publishers, Kluwer, 1991.

A COLD STANDBY SYSTEM WITH IMPERFECT SWITCH AND PREVENTIVE MAINTENANCE: A STOCHASTIC STUDY

R.K. Bhardwaj[1]*, Purnima Sonkar[1]

•

[1]Department of Statistics, Punjabi University Patiala-147002, India
rkb_mstates@rediffmail.com, pani.sonkar@gmail.com,

Abstract

The aim of this paper is to develop a probabilistic model for a cold standby system that consists of an imperfect switching device and a servicing facility. The model aims to address the issue of unexpected random failures of the switch by implementing preventive maintenance measures. The system has two identical units. It starts with one unit in active operation and another unit in cold standby mode. In standby mode the unit remains in perfect state. No failure is allowed in standby mode. As the operating unit fails, the standby unit needs to be switched into operation, to keep the system working. A servicing facility is present in the system to perform necessary servicing related tasks. The servicing facility referred to as the server, also takes care of all necessary remedial activities like preventive maintenance and repairs. The switch used as switching mechanism to place the standby unit into operation may found imperfect when needed. Similarly, the server too can fail while doing job. A preventive maintenance scheme is used for the switch whereas treatment is given to server. The method of semi-Markov process and regenerative point technique is used for model developing and solving, respectively. The expressions are derived to determine different system performance measures such as mean time to system failure, availability, busy period, expected number of preventive maintenances and the profit. The distributions of random time elapsed in repairs, replacements, preventive maintenances and treatments are general. This study highlights the usefulness of switch's preventive maintenance in long run. To study the asymptotic behavior of the system model, all the expressions for system performance measures are obtained in steady state. A simulation study is conducted using a presumed data set and assuming a Weibull probability distribution. The numerical results are shown in tabular form. The simulation results serve to highlight the significance of preventive maintenance for the switch. The findings of the paper can provide guidelines to the people engaged in designing, framing and implementing standby switching systems in real applications.

Keywords: Cold standby, Transition probabilities, Imperfect switch, Server, Preventive maintenance, Semi-Markov process, Steady state, Regenerative point techniques, Weibull distribution.

1. Introduction

The standby redundancy is always at the core of a backup system. A cold standby system is primarily characterized by the standby unit and the switch mechanism, needed to switch the standby into operation at the failure of operating unit. Implementing an appropriate standby scheme results in improved system performance. Though the operating time of a system can be enhanced

by using standby redundancy but impossible to make it full failure proof. The working component of a standby system may fail due to aging, deterioration and some other factors such as shocks [1]. Such factors that are responsible for failures need appropriate repairs to improve system's functioning [2]. Different types of failures demand for different repair strategies like in-house repairs or specialized repairs, carried out by internal or external repairmen [3]. The availability of repairmen and their skills can have a significant impact on system performance and reliability.

Therefore another crucial element of a functioning system is the server; as an efficient and robust server can keep a system functioning for long. The server may get exhausted after working for a long and then needs some short of refreshment [4]. Employing additional repairmen can address the issue of system downtime caused by refreshing the server [5]. In addition to cold standby systems, the server is capable of handling failures in various other types of standby systems as well [6]. The primary task of repairmen in cold standby systems is to activate the cold standby unit which is accomplished through the switching mechanism.

The switch plays a vital role in standby system as it is responsible for switching a failed unit with a standby one. The decision to switch can be made either at a predetermined time or when the operating unit fails [7]. The functioning of the system can be impacted if the switch mechanism itself fails [8]. When a switch failure occurs the process of rebooting and repairing a standby system can have a notable impact on its overall reliability [9]. The switch in the standby system can exhibit either perfect or imperfect behavior depending on its level of functionality [10]. Due to the prevalence of switch failures the switching mechanism in standby systems is some time regarded as imperfect. [11]. The imperfections in the switching mechanism of both cold standby and warm systems have a negative impact on their reliability [12]. The combination of switch and server in a system plays a vital role in maintaining system reliability. The switch activates cold standby unit when the main unit fails while the server handles all repair activities. Such configuration effectively reduces system downtime. However, if either the server or the switch or both fail during task performance the system performance declines [13]. In case of failure the switch may be found imperfect and the server may be unreliable both adversely affect the profit of the system [14]. In such cases the repair process may be disrupted that leads to reduced system performance. So to examine impact of such failures on the profit and availability of the system a probabilistic model may be helpful [15].

A standby system requires regular maintenance and repair to ensure higher availability and reliability as it deteriorates with time [16]. Neglecting the maintenance and repair can lead to equipment failure and financial losses. So a robust maintenance or repair plan is the necessity for optimal system performance [17]. Condition-based maintenance is an effective approach among several others to optimize maintenance costs and ensure efficient operation of standby systems [18]. Though, maintenance presents challenges stemming from factors such as complexity, cost and competition but implementing suitable strategies can significantly reduce maintenance costs [19]. In most of cases the cost of breakdown maintenance is typically higher than preventive maintenance. Therefore implementing a preventive maintenance scheme can help minimize the risk of equipment failure and reduce overall maintenance costs while increasing availability and reliability [20]. Periodic system's inspections are usually conducted to decide about maintenance strategy [21]. In some cases preference is given to preventive maintenance over others [22]. Optimizing the inspection intervals is necessary to achieve the goal of effective preventive maintenance [23]. An effective maintenance plan may involves interval based inspections or multiple inspections in stages [24].

In this paper, we have developed a stochastic model for a cold standby system consisting of two identical units, a switch and a server. The switch is responsible for activating the standby unit and can experience random failures. It undergoes preventive maintenance after a specific time threshold. Similarly the server handles all remedial activities but can only perform one operation at a time and is also prone to failure during tasks. The model gives priority to switch repairs over other

remedial activities. It assumes perfect repairs and server treatments. The study focuses on analyzing the impact of switch preventive maintenance on system performance. Steady state expressions for system performance measures are derived considering all random variables as statistically independent. The system model is developed using the theory of semi-Markov processes. The regenerative point technique and Laplace transforms are used to solve the model. A simulation study is conducted using a dataset and Weibull probability distribution with the results presented graphically.

2. Notations

O:	The unit is in operative mode.
Cs:	The unit is kept as cold standby.
Sh:	The switch is ok
Sv:	The server is ok.
Cs_w:	The cold standby unit is in waiting.
p/q:	Probability that switch is operational/failed.
F_{ur}/ F_{UR}:	The unit is under repair/ continuously under repair from the previous state.
F_{wr}/F_{WR}:	The failed unit is waiting for repair/ waiting for repair continuously from the previous state.
ShF_{ur} /ShF_{UR}:	The switch is under repair / under repair continuously from the previous state.
Sh_{pm}/ShF_{PM}:	The switch is under preventive maintenance / under continuously preventive maintenance from the previous state.
ShF_{wr}/ShF_{WR}:	The switch is waiting for repair / continuously from the previous state.
SvF_{ut}/SvF_{UT}:	The server is under treatment/ continuously from the previous state.
z(t)/ Z(t):	pdf/ cdf of the failure time of the unit.
r (t)/ R(t):	pdf / cdf of the failure time of the server.
f(t)/ F(t):	pdf / cdf of repair time of the failed unit.
h(t)/ H(t):	pdf / cdf of repair time of the failed switch.
s(t)/ S(t):	pdf / cdf of the treatment time of the server.
$q_{ij}(t)/Q_{ij}(t)$:	pdf / cdf of direct transition time from a regenerative state S_i to a regenerative state S_j without visiting any other regenerative state.
$q_{ij.k}(t)/Q_{ij.k}(t)$:	pdf / cdf of first passage time from a regenerative state S_i to a regenerative state S_j or to a failed state S_j visiting state S_k once in (0,t].
(s)/(c):	Symbol for Stieltjes convolution / Laplace convolution.
˜ / * :	Symbol for Laplace Stieltjes Transform(LST) / Laplace Transform(LT).

3. Model Development

3.1. States of the System

The following are the possible states of the system model Figure 1.

The regenerative up states:

$S_0 = (O, Cs, Sh, Sv)$, $S_1 = (F_{ur}, O)$, $S_3 = (F_{wr}, O, Sv_{F_{ut}})$, $S_4 = (O, Cs_w, Sh_{pm})$

The failed regenerative down state:

$$S_2 = (F_{wr}, Cs_w, Sh_{Fur})$$

Non-regenerative states:

$S_5 = (F_{wr}, Cs_w, Sh_{PM})$, $S_6 = (F_{UR}, F_{wr})$, $S_7 = (F_{wr}, F_{WR}, Sv_{F_{ut}})$, $S_8 = (F_{WR}, Cs_w, Sh_{wr}, Sv_{F_{ut}})$, $S_9 = (F_{wr}, F_{WR}, Sv_{F_{UT}})$, $S_{10} = (F_{ur}, F_{WR})$, $S_{11} = (F_{WR}, Cs_w, Sh_{Fur})$,

3.2. State Transition Diagram

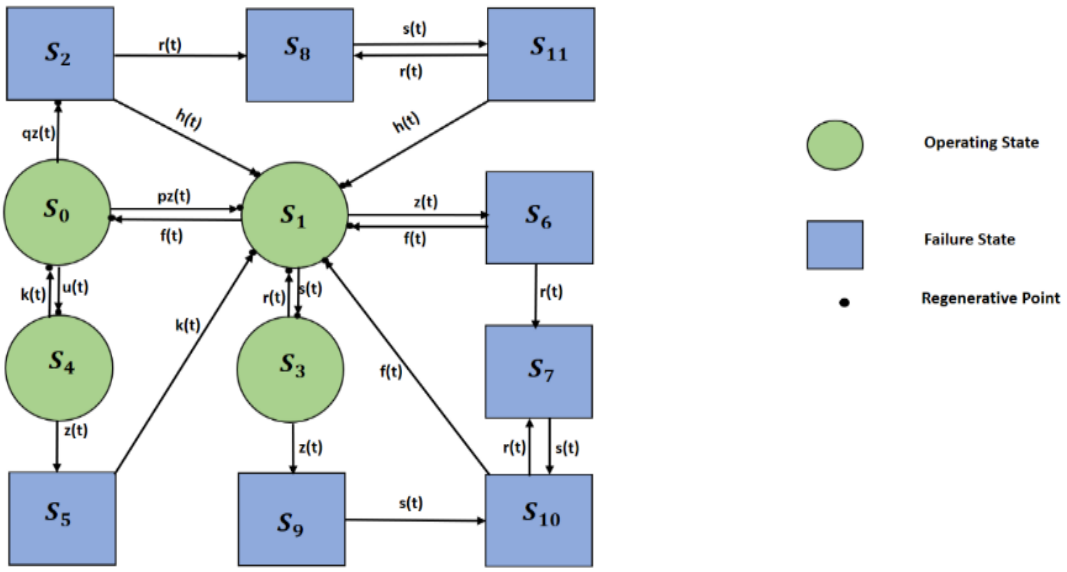


Figure 1: *State transition diagram*

3.3. Transition Probabilities and Mean Sojourn Times

Simple probabilistic considerations yield the following expressions for the non-zero elements

$$p_{ij} = Q_{ij}(\infty) = \int_0^\infty q_{ij}(t)dt \tag{1}$$

Also, the Mean Sojourn time μ_i in state S_i are given by:

$$\mu_i = E(t) = \int_0^\infty P(T > t)dt \tag{2}$$

We get

$$p_{01} = \int_0^\infty pz(t)\bar{U}(t)dt,\ p_{02} = \int_0^\infty qz(t)\bar{U}(t)dt,\quad p_{10} = \int_0^\infty f(t)\bar{R}(t)\bar{Z}(t)dt,$$
$$p_{13} = \int_0^\infty r(t)\bar{F}(t)\bar{Z}(t)dt,\quad p_{16} = \int_0^\infty z(t)\bar{R}(t)\bar{F}(t)dt,\ p_{21} = \int_0^\infty h(t)\bar{R}(t)dt,$$
$$p_{27} = \int_0^\infty r(t)\bar{H}(t)dt,\quad p_{31} = \int_0^\infty s(t)\bar{Z}(t)dt,\ p_{39} = \int_0^\infty z(t)\bar{S}(t)dt,$$
$$p_{40} = \int_0^\infty k(t)\bar{Z}(t)dt,\ p_{45} = \int_0^\infty z(t)\bar{F}(t)dt,\ p_{51} = \int_0^\infty k(t)dt,$$
$$p_{61} = \int_0^\infty f(t)\bar{Z}(t)dt,\ p_{67} = \int_0^\infty z(t)\bar{F}(t)dt,\ p_{7,10} = \int_0^\infty s(t)dt,$$
$$p_{8,11} = \int_0^\infty s(t)dt,\ p_{9,11} = \int_0^\infty s(t)dt,\ p_{10,1} = \int_0^\infty f(t)\bar{R}(t)dt,$$
$$p_{10,7} = \int_0^\infty r(t)\bar{F}(t)dt,\ p_{11,1} = \int_0^\infty h(t)\bar{R}(t)dt,\ p_{11,8} = \int_0^\infty r(t)\bar{H}(t)dt,$$

Here, it can be checked that sum of simple probabilities originating from a single state is unity. The expressions for mean sojourn times are as follows:

$$\mu_0 = \int_0^\infty \bar{Z}(t)dt,\quad \mu_1 = \int_0^\infty \bar{F}(t)\bar{R}(t)\bar{Z}(t)dt,\qquad \mu_2 = \int_0^\infty \bar{R}(t)\bar{H}(t)dt,$$
$$\mu_3 = \int_0^\infty \bar{S}(t)\bar{Z}(t)dt,\quad \mu_4 = \int_0^\infty \bar{K}(t)\bar{Z}(t)dt$$

4. System's Performance Measures

4.1. MTSF

Let $\emptyset_i(t)$ be the c.d.f of the first passage time from the regenerative state S_i to a failed state. Regarding the failed state as an absorbing state, we have the following recursive relations for $\emptyset_i(t)$:

$$\begin{aligned}
\emptyset_0(t) &= Q_{01}(t)\ (s)\ \emptyset_1(t) + Q_{04}(t)\ (s)\ \emptyset_4(t) + Q_{02}(t)\\
\emptyset_1(t) &= Q_{10}(t)\ (s)\ \emptyset_0(t) + Q_{13}(t)\ (s)\ \emptyset_3(t) + Q_{16}(t)\\
\emptyset_3(t) &= Q_{31}(t)\ (s)\ \emptyset_1(t) + Q_{39}(t)\\
\emptyset_4(t) &= Q_{40}(t)\ (s)\ \emptyset_1(t) + Q_{45}(t)
\end{aligned} \tag{4}$$

Taking LST of equation (4) and solving for $\widetilde{\emptyset_0}(s)$, we have

$$R^*(s) = \frac{1-\widetilde{\emptyset_0}(s)}{s} \tag{5}$$

The reliability R(t) can be obtained by taking the inverse Laplace transition of (5) and MTSF is given by

$$MTSF = \lim_{s\to 0} R^*(s) = \lim_{s\to 0} \frac{1-\widetilde{\emptyset_0}(s)}{s} \tag{6}$$

$$MTSF = \frac{(\mu_0 + p_{04}\mu_4)(1-p_{13}p_{31}) + p_{01}(\mu_1 + p_{13}\mu_3)}{(1-p_{13}p_{31})(1-p_{04}p_{40}) - p_{01}p_{10}} \tag{7}$$

4.2. Steady State Availability

$M_i(t)$ is the probability that the system is up initially in state $S_i \in E$ is up at time t without visiting to any other regenerative state, we have

$$M_0 = \int_0^\infty \bar{Z}(t)dt,\quad M_1 = \int_0^\infty \bar{F}(t)\bar{R}(t)\bar{Z}(t)dt,\quad M_3 = \int_0^\infty \bar{S}(t)\bar{Z}(t)dt$$

Let $A_i(t)$ be the probability that the system is in up-state at an instant 't' given that the system entered regenerative state S_i at t=0. The recursive relations for $A_i(t)$ are as follows:

$$\begin{aligned}
A_0(t) &= M_0(t) + q_{01}(t)\ (c)\ A_1(t) + q_{02}(t)(c)A_2(t)\\
A_1(t) &= M_1(t) + q_{10}(t)\ (c)\ A_0(t) + \left(q_{1,1.6}(t)(t) + q_{1,1.6,(7,10)^n}\right)(c)A_1(t) + q_{13}(t)\ (c)\ A_3(t)\\
A_2(t) &= \left(q_{21}(t) + q_{2,1.8,11}(t)\right)(c)A_1(t)\\
A_3(t) &= M_3(t) + \left(q_{31}(t) + q_{3,1.9,10}(t) + q_{3,1.9,(10,7)^n}(t)\right)(c)A_1(t)\\
A_4(t) &= M_4(t) + q_{4,0}(t)\ (c)\ A_0(t) + q_{4,1.5}(t)(c)A_1(t)
\end{aligned} \tag{8}$$

Where S_j is any successive regenerative state to which the regenerative state S_i can transit through n transitions. Taking LT of equation (8) and solving for $A_0^*(s)$, the steady state availability is given

$$A_0 = \lim_{s\to 0} sA_0^*(s) = \frac{(\mu_0 + p_{04}\mu_4)p_{1,0} + (1-p_{13}p_{31})(\mu_1 + p_{13}\mu_3)}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1-p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \tag{9}$$

4.3. Busy period Analysis for the Server

Let $B_i(t)$ be the probability that the server is busy in repair of the unit or switch at an instant t given that the system entered regenerative state S_i at t = 0. The recursive relations for $B_i(t)$ are as follows:

$$\begin{aligned}
B_0(t) &= q_{01}(t)\ (c)\ B_1(t) + q_{02}(t)(c)B_2(t)\\
B_1(t) &= W_1(t) + q_{10}(t)\ (c)\ B_0(t) + q_{13}(t)\ (c)\ B_3(t) + q_{11.6}(t)(c)B_1(t) + q_{11.6,(7,10)^n}(c)B_1(t)\\
B_2(t) &= W_2(t) + q_{21}(t)\ (c)\ B_1(t) + q_{21.8,11}(t)(c)B_1(t)\\
B_3(t) &= q_{31}(t)\ (c)\ B_1(t) + q_{31.9,10}(t)(c)B_1(t) + q_{31.9,(10,7)^n}(t)(c)B_1(t)\\
B_4(t) &= q_{4,0}(t)\ (c)\ B_0(t) + q_{4,1.5}(t)(c)B_1(t)
\end{aligned} \tag{10}$$

$W_i(t)$ be the probability that the server is busy in state S_i due to repair of the unit or switch up to time 't' without making any transition to any other regenerative state or returning to the same via

one or more non-regenerative state

$W_1(t) = \bar{Z}(t)\bar{F}(t)\bar{R}(t) + (z((t)\bar{F}(t)\bar{R}(t)©1)\bar{F}(t) + (z(t)\bar{F}(t)\bar{R}(t)©r(t)\bar{F}(t)©1)\bar{S}(t)$
$+ (z(t)\bar{F}(t)\bar{R}(t)©r(t)\bar{F}(t)©s(t)©1)\bar{F}(t)$

$W_2(t) = \bar{R}(t)\bar{H}(t) + (r(t)\bar{R}(t)©1)\bar{S}(t) + (r(t)\bar{R}(t)©s(t)©1)\bar{H}(t)$

Using LT, of equation (10) and solving for $B_0^*(s)$, the time for which server is busy due to repair of unit or switch is given by

$$B_0 = \lim_{s\to 0} sB_0^*(s) = \frac{W_1^*(0)(1-p_{04}p_{40}) + p_{02}p_{10}W_2^*(0)}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1-p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \quad (11)$$

4.4. Busy Period Analysis for the Server due to Switch Preventive Maintenance

Let $B^P{}_i(t)$ be the probability that the server is busy in preventive maintenance of switch after a fixed time period at an instant t given that the system entered regenerative state S_i at t = 0. The recursive relations for $B_i(t)$ are as follows:

$B^P{}_0(t) = q_{01}(t)\,(c)\,B^P{}_1(t) + q_{02}(t)(c)B^P{}_2(t) + q_{04}(t)(c)B^P{}_4(t)$
$B^P{}_1(t) = q_{10}(t)\,(c)\,B^P{}_0(t) + q_{13}(t)\,(c)\,B^P{}_3(t) + q_{11.6}(t) + q_{11.6,(7,10)^n}(c)B^P{}_1(t)$
$B^P{}_2(t) = q_{21}(t)\,(c)\,B^P{}_1(t) + q_{21.8,11}(t)(c)B^P{}_1(t)$
$B^P{}_3(t) = q_{31}(t)\,(c)\,B^P{}_1(t) + q_{31.9,10}(t)(c)B^P{}_1(t) + q_{31.9,(10,7)^n}(t)(c)B^P{}_1(t)$
$B^P{}_4(t) = W_4(t) + q_{4,0}(t)\,(c)\,B^P{}_0(t) + q_{4,1.5}(t)(c)B^P{}_1(t)$ (12)

$W_i(t)$ be the probability that the server is busy in state S_i due to switch preventive maintenance up to time 't' without making any transition to any other regenerative state or returning to the same via one or more non-regenerative state

$W_4(t) = \bar{Z}(t)\bar{K}(t) + (z(t)©1)\bar{K}(t)$

Using LT, of equation (12) and solving for $B^{P*}{}_0(s)$, the time for which server is busy due to switch preventive maintenance is given by

$$B^P{}_0 = \lim_{s\to 0} sB^{P*}{}_0(s) = \frac{p_{04}p_{10}W_4^*(0)}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1-p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \quad (13)$$

4.5. Expected Number of Server Treatments

Let $T_i(t)$ be the expected number of treatments given to the server in (0,t] given that the system entered regenerative state S_i at t=0. The recursive relations for $T_i(t)$ are as follow:

$T_0(t) = Q_{01}(t)\,(s)\,T_1(t) + Q_{02}(t)(s)T_2(t) + Q_{04}(t)(s)T_4(t)$
$T_1(t) = Q_{10}(t)\,(s)\,T_0(t) + Q_{13}(t)\,(s)\,T_3(t) + Q_{11.6}(t)(s)T_1(t) + Q_{11.6,(7,10)^n}(s)T_1(t)$
$T_2(t) = Q_{21}(t)\,(s)\,T_1(t) + Q_{21.8,11}(t)(s)T_1(t)$
$T_3(t) = Q_{31}(t)\,(s)\,(1 + T_1(t)) + Q_{31.9,10}(t)(s)T_1(t) + Q_{31.9,(10,7)^n}(t)(s)T_1(t)$
$T_4(t) = Q_{4,0}(t)\,(s)\,T_0(t) + Q_{4,1.5}(t)(s)T_1(t)$ (14)

Using LT, of equation (14) and solving for $\widetilde{T_0}(s)$, the expected number of the treatments given to the server are given by

$$T_0 = \lim_{s\to 0} s\widetilde{T_0}(s) = \frac{p_{31}p_{13}(1-p_{04}p_{40})}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1-p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \quad (15)$$

4.6. Expected Number of Switch Repairs

Let $U_i(t)$ be the expected number of repairs given to the switch in (0,t] given that the system entered regenerative state S_i at t=0. The recursive relations for $U_i(t)$ are as follow:

$U_0(t) = Q_{01}(t)\,(s)\,U_1(t) + Q_{02}(t)(s)U_2(t) + Q_{04}(t)(s)U_4(t)$
$U_1(t) = Q_{10}(t)\,(s)\,U_0(t) + Q_{13}(t)\,(s)\,U_3(t) + Q_{11.6}(t)(s)U_1(t) + Q_{11.6,(7,10)^n}(s)U_1(t)$
$U_2(t) = Q_{21}(t)\,(s)\,(1 + U_1(t)) + Q_{21.8,11}(t)(s)(1 + U_1(t))$
$U_3(t) = Q_{31}(t)\,(s)\,U_1(t) + Q_{31.9,10}(t)(s)U_1(t) + Q_{31.9,(10,6)^n}(t)(s)U_1(t)$
$U_4(t) = Q_{4,0}(t)\,(s)\,U_0(t) + Q_{4,1.5}(t)(s)U_1(t)$ (16)

Using LT, of equation (16) and solving for $\widetilde{U_0}(s)$, the expected number of the repairs given to the switch are given by

$$U_0 = \lim_{s\to 0} s\widetilde{U_0}(s) = \frac{p_{02}p_{10}}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1 - p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \quad (17)$$

4.7. Expected Number of Repairs of the Unit

Let $O_i(t)$ be the expected number of repairs given to the server in (0,t] given that the system entered regenerative state S_i at t=0. The recursive relations for $O_i(t)$ are as follow:

$O_0(t) = Q_{01}(t)\ (s)\ O_1(t) + Q_{02}(t)(s)O_2(t) + Q_{04}(t)(s)O_4(t)$
$O_1(t) = Q_{10}(t)\ (s)\ (1 + O_0(t)) + Q_{13}(t)\ (s)\ O_3(t) + Q_{11.6}(t)(s)(1 + O_1(t)) + Q_{11.6,(7,10)^n}(s)(1 + O_1(t))$
$O_2(t) = Q_{21}(t)\ (s)\ O_1(t) + Q_{21.8,11}(t)(s)O_1(t)$
$O_3(t) = Q_{31}(t)\ (s)\ O_1(t) + Q_{31.9,10}(t)(s)(1 + O_1(t)) + Q_{31.9,(10,6)^n}(t)(s)(1 + O_1(t))$
$O_4(t) = Q_{4,0}(t)\ (s)\ O_0(t) + Q_{4,1.5}(t)(s)O_1(t)$ (18)

Where S_j is any regenerative state to which the given regenerative state S_i transits. Using LT, of equation (18) and solving for $\widetilde{O_0}(s)$, the expected number of the repairs given to the unit are given by

$$O_0 = \lim_{s\to 0} s\widetilde{O_0}(s) = \frac{(p_{10} + p_{11,6} + p_{11,6,(7,10)^n})(1 - p_{04}p_{40})}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1 - p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \quad (19)$$

4.8. Expected Number of Switch Preventive Maintenances

Let $N_i(t)$ be the expected number of preventive maintenance given to switch after a fixed time in (0,t] given that the system entered regenerative state S_i at t=0. The recursive relations for are as follows:

$N_0(t) = Q_{01}(t)\ (s)\ N_1(t) + Q_{02}(t)(s)N_2(t) + Q_{04}(t)(s)N_4(t)$
$N_1(t) = Q_{10}(t)\ (s)\ N_0(t) + Q_{13}(t)\ (s)\ N_3(t) + Q_{11.6}(t)(s)N_1(t) + Q_{11.6,(7,10)^n}(s)N_1(t)$
$N_2(t) = Q_{21}(t)\ (s)\ N_1(t) + Q_{21.8,11}(t)(s)N_1(t)$
$N_3(t) = Q_{31}(t)\ (s)(1 + N_1(t)) + Q_{31.9,10}(t)(s)N_1(t) + Q_{31.9,(10,7)^n}(t)(s)N_1(t)$
$N_4(t) = Q_{4,0}(t)\ (s)\ (1 + N_0(t)) + Q_{4,1.5}(t)(s)(1 + N_1(t))$ (20)

Where S_j is any regenerative state to which the given regenerative state S_i transits. Using LT, of equation (20) and solving for $\widetilde{N_0}(s)$, the expected number of PM of switch are given by

$$N_0 = \lim_{s\to 0} s\widetilde{N}_0(s) = \frac{p_{04}p_{10}}{(\mu_0 + p_{04}\mu_4')p_{1,0} + (1 - p_{04}p_{40})(\mu_1' + p_{13}\mu_3') + p_{10}p_{02}\mu_2'} \quad (21)$$

4.9. Cost Analysis

The Profit incurred to the system is given by

$P = C_0A_0 - C_1B_0 - C_2B^P - C_3T_0 - C_4U_0 - C_5O_0 - C_6N_0$ (22)

C_0 = Revenue per unit up time of the system.
C_1 = Cost per unit time for which server is busy in repairing
C_2 = Cost per unit time for which server is busy in preventive maintenance.
C_3 = Cost per treatment of the server.
C_4 = Cost per repair of the switch.
C_5 = Cost per repair of the unit.
C_6 = Cost per preventive maintenance of the switch.

5. Simulation Study (Weibull Distribution)

Let us suppose that all the random variables follow Weibull distribution as given below:

$$z(t) = \lambda\eta t^{\eta-1}e^{-\lambda t^{\eta}}, \qquad f(t) = \alpha\eta t^{\eta-1}e^{-\alpha t^{\eta}}$$

$$r(t) = \mu\eta t^{\eta-1}e^{-\mu\eta t^{\eta}}, \quad s(t) = \beta\eta t^{\eta-1}e^{-\beta t^{\eta}}$$

$$h(t) = \Upsilon\eta t^{\eta-1}e^{-\Upsilon t^{\eta}}, \quad k(t) = \Theta\eta t^{\eta-1}e^{-\Theta t^{\eta}}, \quad u(t) = \xi\eta t^{\eta-1}e^{-\xi t^{\eta}}$$

For simulation, we assumed values of different parameters and costs. The impact of different parameters on system performance is shown in the tables.

Table 1: *Effect of various parameters on MTSF*

MTSF η=0.5							
Failure rate (λ)	$\alpha = 0.2, \beta = 0.7, \Upsilon = 0.3, \mu = 0.03, \Theta = 0.77, \xi = 0.4$ p = 0.4	γ=0.4	α=0.5	β=0.8	p=0.7	θ=0.9	ξ=0.8
0.01	896.56	896.56	918.05	917.72	917.51	997.48	1195.09
0.02	409.57	409.57	420.27	419.97	419.77	492.65	509.27
0.03	251.26	251.26	258.36	258.07	257.89	324.51	294.96
0.04	174.47	174.47	179.76	179.48	179.32	240.53	194.91
0.05	129.91	129.91	134.11	133.85	133.70	190.21	138.91
η=1							
0.01	108.16	108.16	110.72	110.74	110.74	117.63	283.87
0.02	54.43	54.43	55.82	55.84	55.83	58.56	143.77
0.03	36.51	36.51	37.52	37.53	37.53	38.89	96.62
0.04	27.55	27.55	28.36	28.38	28.37	29.07	72.79
0.05	22.18	22.18	22.86	22.88	22.88	23.18	58.32
η=2							
0.01	43.09	43.09	44.15	44.14	44.14	47.54	74.30
0.02	22.77	22.77	23.40	23.38	23.39	22.95	41.47
0.03	15.98	15.98	16.46	16.45	16.45	14.83	30.27
0.04	12.57	12.57	12.98	12.97	12.97	10.81	24.50
0.05	10.51	10.51	10.88	10.87	10.87	8.44	20.92

Table1 summarizes significant findings on Mean Time to System Failure (MTSF). It reveals the impact of different parameter values on MTSF. Higher values of failure rates (λ) lead to decreased MTSF implying shorter system lifespans. Increasing the value of shape parameter (η) also lowers MTSF and system reliability. Notably, increasing values of switch repair rate (γ) from 0.3 to 0.4, rate of switch goes under preventive maintenance (ξ) from 0.4 to 0.8, server treatment rate (β) from 0.8 to 0.9, unit repair rate (α) from 0.2 to 0.5 and switch preventive maintenance rate (θ) from 0.77 to 0.9 all enhance the observed trends of MTSF. It implies that the switch preventive maintenance impacts the system reliability significantly.

Likewise, Tables 2 shows similar trends of system availability. The measure declines with higher values of shape parameter (η). It demonstrate that higher failure rate of the unit reduces system availability. For a fixed value of failure rates (λ) the system's availability improves with

increasing values of rate with which switch goes under preventive maintenance (ξ), server treatment rate (β), repair rate of unit (α) and switch preventive maintenance rate (θ). The values of system availability also rises significantly along with improved switch preventive maintenance.

Table 2: *Effect of various parameters on Availability*

Availability η=0.5							
Failure rate (λ)	$\alpha = 0.2, \beta = 0.7, \Upsilon = 0.3, \mu = 0.03, \theta = 0.77, \xi = 0.4$ p = 0.4	γ=0.4	α=0.5	β=0.8	p=0.7	θ=0.9	ξ=0.8
0.01	0.9787	0.9912	0.9900	0.9902	0.9908	0.9888	0.9996
0.02	0.9574	0.9813	0.9797	0.9799	0.9806	0.9768	0.9986
0.03	0.9362	0.9704	0.9689	0.9689	0.9695	0.9640	0.9976
0.04	0.9151	0.9586	0.9578	0.9575	0.9575	0.9505	0.9961
0.05	0.8943	0.9461	0.9464	0.9456	0.9449	0.9364	0.9941
η=1							
0.01	0.9102	0.9912	0.9921	0.9921	0.9855	0.9926	0.9940
0.02	0.8388	0.9813	0.9839	0.9837	0.9707	0.9844	0.9887
0.03	0.7807	0.9704	0.9754	0.9748	0.9557	0.9755	0.9816
0.04	0.7325	0.9586	0.9667	0.9656	0.9405	0.9659	0.9737
0.05	0.6918	0.9461	0.9578	0.9560	0.9253	0.9557	0.9651
η=2							
0.01	0.8422	0.9901	0.9899	0.9929	0.9916	0.9884	0.9880
0.02	0.8331	0.9822	0.9816	0.9876	0.9850	0.9789	0.9785
0.03	0.8237	0.9737	0.9726	0.9819	0.9780	0.9691	0.9687
0.04	0.8141	0.9649	0.9632	0.9757	0.9705	0.9589	0.9586
0.05	0.8045	0.9557	0.9533	0.9692	0.9627	0.9485	0.9484

The table 3 highlights the effect of various parameters on system profit. It reveals that system profit declines with an increasing failure rate of unit (λ) as well as higher values of shape parameter (η). The profit expands with repairs, treatments and preventive maintenance.

The numerical simulation results have indicated that effect of preventive maintenance on overall system's performance is encouraging. All the performance measures MTSF, availability and the profit improves with switch preventive maintenance rate (θ). The higher frequency of preventive maintenance (ξ) too ensures improved system performance. The MTSF almost doubles with change in value of shape parameter (η) from 0.5 to 1. These results underline the importance of managing failure rates and implementing effective preventive maintenance strategies to keep a system reliable and profitable.

Table 3: *Effect of various parameters on system profit*

Profit η=0.5							
Failure rate (λ)	$\alpha = 0.2, \beta = 0.7, \Upsilon = 0.3, \mu = 0.03, \theta = 0.77, \xi = 0.4, p = 0.4$	γ=0.4	α=0.5	β=0.8	p=0.7	θ=0.9	ξ=0.8
0.01	7324.6	7931.7	8351.3	8796.9	9412.2	8178.6	9195.4
0.02	7048.7	7610.7	8095.4	8519.8	9097.0	7917.9	8536.6
0.03	6809.4	7338.3	7876.9	8238.1	8760.1	7663.9	7967.8
0.04	6590.2	7093.0	7678.5	7984.0	8460.9	7416.7	7470.4
0.05	6384.4	6865.4	7492.8	7747.6	8185.9	7176.4	7030.8
η=1							
0.01	8028.0	8784.7	9666.5	9662.9	9590.1	9253.3	9759.9
0.02	7907.3	8723.4	9361.3	9379.0	9252.0	8905.2	9434.4
0.03	7780.8	8654.9	9089.5	9130.8	8958.1	8604.7	9134.0
0.04	7649.1	8580.5	8840.1	8904.5	8691.3	8332.6	8850.1
0.05	7513.2	8501.3	8607.8	8693.8	8444.0	8080.8	8601.8
η=2							
0.01	8085.3	8534.8	8369.4	9045.3	9723.2	9664.2	8571.4
0.02	7797.9	8263.8	8108.4	8753.1	9393.2	8694.1	8547.5
0.03	7530.9	8019.6	7872.4	8488.7	9096.0	7899.1	7799.3
0.04	7280.6	7795.7	7655.6	8245.9	8823.7	7234.4	7178.6
0.05	7047.3	7588.6	7454.7	8021.0	8572.2	6670.1	6655.2

6. Applications

In practice the switch is an integral component that holds significant importance in various applications. It plays a critical role in diverse sectors few including Rail Tracking, Wind Power Plants, and DSLAM Networks. In Rail Tracking systems switches are responsible for ensuring the safe and efficient movement of trains by facilitating the switching of tracks. In Wind Power Plants switches are utilized to control the flow of electricity from the turbines to the grid. Similarly in DSLAM Networks the switches manage the traffic and direct data packets to their intended destinations. As the switch preventive maintenance refers to the proactive measures taken to anticipate and mitigate potential failures before they actually occur. Hence by conducting regular inspections and performing necessary preventive maintenance helps minimize the risk of sudden failures within systems using switching phenomena.

7. Conclusion

The occurrence of unexpected system breakdowns significantly reduces system outputs. Therefore implementing appropriate maintenance schemas are always required. These schemas aimed to enhance system performance by minimizing sudden breakdowns. In this paper the application of preventive maintenance specifically to the switch is focused on. The idea of implementing preventive maintenance effectively reduces the failure likelihood of switch. The simulation results have illustrated the impact of switch preventive maintenance on mean time to system failure, system availability and profitability. These findings indicated that switch preventive maintenance influences overall system performance and profitability.

References

[1] Wu, Q. (2012). Reliability analysis of a cold standby system attacked by shocks. *Applied Mathematics and Computation*, *218*(23):11654-11673. doi: 10.1016/j.amc.2012.05.051.

[2] Wang, J. & Ye, J. (2022). A new repair model and its optimization for cold standby system. *Operational Research*, 1-18. doi: 10.1007/s12351-020-00545-x.

[3] Andalib, V. & Sarkar, J. (2019). A repairable system supported by two spare units and serviced by two types of repairers. doi: https://doi.org/10.48550/arXiv.1908.02547

[4] Kumar, A. Garg, R. and Barak, M. S. (2023). Reliability measures of a cold standby system subject to refreshment. *International Journal of System Assurance Engineering and Management*, *14*(1):147-155. doi: 10.1007/s13198-021-01317-2.

[5] Andalib, V. and Sarkar, J. (2022). A system with two spare units, two repair facilities, and two types of repairers. *Mathematics*, *10*(6):852. doi: 10.3390/math10060852.

[6] Wang, J. Xie, N. and Yang, N. (2021). Reliability analysis of a two-dissimilar-unit warm standby repairable system with priority in use. *Communications in Statistics-Theory and Methods*, *50*(4):792-814. doi: 10.1080/03610926.2019.1642488.

[7] Jia, X. Chen, H. Cheng, Z. and Guo, B. (2016). A comparison between two switching policies for two-unit standby system. *Reliability engineering & system safety*, *148*:109-118. doi: 10.1016/j.ress.2015.12.006.

[8] Ke, J. C. Liu, T. H. and Yang, D. Y. (2016). Machine repairing systems with standby switching failure. *Computers & Industrial Engineering*, *99*:223-228. doi: 10.1016/j.cie.2016.07.016.

[9] Hsu, Y. L. Ke, J. C. Liu, T. H. and Wu, C. H. (2014). Modeling of multi-server repair problem with switching failure and reboot delay and related profit analysis. *Computers & Industrial Engineering*, *69*:21-28. doi: 10.1016/j.cie.2013.12.003.

[10] Hu, L. Cao, X. and Li, Z. (2019). Reliability analysis of discrete time redundant system with imperfect switch and random uncertain lifetime. *Journal of Intelligent & Fuzzy Systems*, *37*(1):723-735. doi: 10.3233/JIFS-181260.

[11] Bai, S. Jia, X. Cheng, Z. Guo, B. Zhao, Q. and Zhang, X. (2022). Operation optimization model for warm standby system based on non-periodic and imperfect multiple active switching policy. *Computers & Industrial Engineering*, *167*:108001. doi: 10.1016/J.CIE.2022.108001.

[12] Sadeghi, M. and Roghanian, E. (2017). Reliability analysis of a warm standby repairable system with two cases of imperfect switching mechanism. *Scientia Iranica*, *24*(2):808-822. doi: 10.24200/sci.2017.4063.

[13] Kuo, C. C. and Ke, J. C. (2016). Comparative analysis of standby systems with unreliable server and switching failure. *Reliability Engineering & System Safety*, *145*:74-82. doi: 10.1016/j.ress.2015.09.001.

[14] Ke, J. C. Liu, T. H. and Yang, D. Y. (2018). Modeling of machine interference problem with unreliable repairman and standbys imperfect switchover. *Reliability Engineering & System*

Safety, *174*:12-18. doi: 10.1016/j.ress.2018.01.013.

[15] Bhardwaj, R. K. Sonker, P. and Singh, R. (2022). A semi-Markov model of a system working under uncertainty. In *System Assurances*. Academic Press, 175-187. doi: 10.1016/B978-0-323-90240-3.00011-4.

[16] Samuelson, A. Haigh, A. O'Reilly, M. M. and Bean, N. G. (2017). Stochastic model for maintenance in continuously deteriorating systems. *European Journal of Operational Research*, *259*(3):1169-1179. doi: 10.1016/j.ejor.2016.11.032.

[17] Sultan, K. S. and Moshref, M. E. (2021). Stochastic Analysis of a Priority Standby System under Preventive Maintenance. *Applied Sciences*, *11*(9):3861. doi: 10.3390/app11093861.

[18] Wang, Y. Li, X. Chen, J. and Liu, Y. (2022). A condition-based maintenance policy for multi-component systems subject to stochastic and economic dependencies. *Reliability Engineering & System Safety*, *219*:108174. doi: 10.1016/j.ress.2021.108174.

[19] Levitin, G. Xing, L. and Dai, Y. (2021). Optimal operation and maintenance scheduling in m-out-of-n standby systems with reusable elements. *Reliability Engineering & System Safety*, *211*:107582. doi: 10.1016/j.ress.2021.107582.

[20] Hamdan, K. Tavangar, M. and Asadi, M. (2021). Optimal preventive maintenance for repairable weighted k-out-of-n systems. *Reliability Engineering & System Safety*, *205*:107267. doi: 10.1016/j.ress.2020.107267.

[21] Wang, W. Wu, Z. Xiong, J. and Xu, Y. (2018). Redundancy optimization of cold-standby systems under periodic inspection and maintenance. *Reliability engineering & system safety*, *180*:394-402. doi: 10.1016/j.ress.2018.08.004.

[22] Purnima, A. K. and Kumar, N. (2022). Mathematical Modeling and Profit Analysis of a Standby System with Priority for Preventive Maintenance. *Mathematical Statistician and Engineering Applications*, *71*(4):9129-9139.

[23] Lu, X. Wang, W. Yang, H. Zuo, M. J. and Zhou, D. (2012). Optimizing the Periodic Inspection Interval for a 1-out-of-2 Cold Standby System Using the Delay-Time Concept. *Quality and Reliability Engineering International*, *28*(6):648-662. doi: 10.1002/qre.1395.

[24] Zhang, F. Shen, J. Liao, H. and Ma, Y. (2021). Optimal preventive maintenance policy for a system subject to two-phase imperfect inspections. *Reliability Engineering & System Safety*, *205*:107254. doi: 10.1016/J.RESS.2020.107254.

MOVING BLOCK BOOTSTRAP METHOD WITH BETTER ELEMENTS REPRESENTATION FOR UNIVARIATE TIME SERIES DATA

[1]Kayode AYINDE, *[2]James DANIEL, [1]Akinola ADEPETUN and
[1]Olusegun S. EWEMOOJE

•

[1]Federal University of Technology, Akure, Nigeria
[2]National Bureau of Statistics, Abuja, Nigeria
*futathesis@gmail.com

Abstract

Bootstrap method was initially used to determine accuracy measures for sample estimates of independent and identical distributions (i.i.d.). In order to apply bootstrap method to time-dependent data, blocking technique is introduced to preserve serial correlation of the original time series data. In the past, resampling techniques for time-dependent data were implemented using Non-overlapping Block Bootstrap (NBB) method but its dichotomous block arrangement restricts the number of blocks. As a result, improvement becomes necessary. Although the Moving Block Bootstrap (MBB) method improves upon NBB with regard to many more blocks, it introduces an uneven representation of the time series elements which eventually influences its accuracy. In this paper, an innovative method called Moving Block Bootstrap Method with better element Representation (MBBR) is developed to ensure that the time series elements within the block are better represented with minimum number of elements. To compare MBB and MBBR, simulated studies were carried out on some set time series data following each classes of Autoregressive Moving Average (ARMA) model with different parameters, sample sizes and standard deviation using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Results show that by improving the representation of time series data in the blocking arrangement, the accuracy of the proposed method (MBBR) consistently outperforms the existing one (MBB) and thus, provides more efficient estimates of the dependent variable.

Keywords: Bootstrap method, Autoregressive Moving Average (ARMA), measures of accuracy, Moving block bootstrap (MBB), Moving Block Bootstrap Method with better element Representation.

1. Introduction

Statistics, the bedrock of rational and scientific decision-making, often relies on random samples of observations to draw conclusions. As a result, one can safely say that proper sampling is the backbone of Statistics, and bootstrapping is one of its dynamics. Bootstrap method relies on using original sample or some parts of it as an artificial population from which random samples are selected. [5] introduced the concept of bootstrap which keeps spreading like bushfire in the field of statistical sciences in a couple of decades. The resampling scheme as provided by [5] is discussed as follows. Suppose that $X\iota = x_1, x_2, \cdots, x_n$ is a random sample of observation from independently identical distribution (i.i.d.). If X_ι^* is selected at random from X_ι (that is $x_1^*, x_2^*, \cdots, x_k^*$ from $x_1, x_2, \cdots, x_n$) with probability $P(X_\iota) = \frac{1}{n}$, $\iota = 1, 2, \cdots, n$; To generate a bootstrap random sample by resampling X_ι, generate n random integers $\iota = 1, 2, \cdots, n$ and select the bootstrap sample $X_\iota^* = x_1^*, \cdots, x_k^*$ with replacement. Suppose θ is the parameter of interest (θ

could be a vector), and $\hat{\theta}$ is an estimator of θ.

The moving block bootstrap (MBB) was developed in [8] as an improvement of the Non-overlapping Block Bootstrap NBB. The MBB was developed to give room for more blocks as against the NBB in [3]. However, while making provision for more blocks, some elements appear less frequently than others, especially the extreme values. Thus, this eventually became a problem as such uneven representation lower its accuracy values. Illustration of MBB method as demonstrated by [8] shows the problem as follows: think of $X = x_1, x_2, \cdots, x_{10}$ as the original time series data. If a block length of $l = 3$ is considered such that $B_1 = x_1, x_2, x_3$, $B_2 = x_2, x_3, x_4$, $B_3 = x_3, x_4, x_5$, $B_4 = x_4, x_5, x_6$, $B_5 = x_5, x_6, x_7$, $B_6 = x_6, x_7, x_8$, $B_7 = x_7, x_8, x_9$, $B_8 = x_8, x_9, x_{10}$, it is therefore observed that x_1, x_10 each appears once and x_2, x_9 each appears just twice while every other observation appears three times each in the MBB blocking scheme. This portrays the problem of uneven representation of the time series elements within the block arrangement at the two edges of the original the time series data. The presence of such uneven representation is capable of influencing the accuracy measure of MBB method. To profile solution to the problem of this uneven representation without creating additional one, effort is made in this paper to design a synthetic blocking scheme within its block pot as an improvement on the MBB

In [3], data-based Markov chain to sample blocks is prioritized in order to increase the possibility that subsequent blocks match at their conclusion. The description of the circumstances in which the bias of a bootstrap estimator of a variance is reduced by this matched-block bootstrap is made. However, the moving block bootstrap only accelerates the pace of bias convergence when a Markov process generates the data. The estimator's variance is not decreased by the moving block bootstrap. Extension of the bootstrap method proposed by [5] through [8]. Series was split into $nl + 1$ overlapping blocks of length l. Observations 1 to l is considered to be block 1, observations 2 to $l + 1$ is considered to be block 2, etc until all the elements of the the time series data are exhausted. Then, $sfracnl$ blocks were generated at random with replacement from these $nl + 1$ blocks. Following that, aligning the $sfracnl$ blocks in the selection order yields the bootstrap observations. Although the bootstrapped observations are no longer steady with this construction, this type of bootstrap approach still works with dependent data. However, [16] stressed that the issue can be solved by arbitrarily changing the block length. The stationary bootstrap is the name of this technique. Under the condition that ρ^{-1} is roughly equal to l, where l is the block length and the parameter of the geometric distribution, the stationary bootstrap estimate of variance and the moving block estimate of variance are relatively close ([17]). The Markovian bootstrap and a stationary bootstrap approach that matches succeeding blocks based on standard deviation matching are other related variants of the moving block bootstrap ([6]).

Circular Block Bootstrap (CBB), Moving Block Bootstrap (MBB), and Stationary Block Bootstrap (SBB) are asymptotically comparable in the sense of mean squared error, according to [10], who compared the asymptotic minimal values of the mean square error of each of the four block bootstrap methods (MSE). The study confirmed that, even with moderately sized samples, there are benefits to employing MBB and CBB rather than the stationary block bootstrap approach. Investigated on how the optimal block bootstrap method might be particularly sensitive to the choice of block size was made by [1], While [12] pointed out the stationary difficulty of the resampled series by the moving block bootstrapping. Tapered Block Bootstrap was proposed by [13], a new variation of the block bootstrap covering approximately linear statistics, and represented an improvement over the original block bootstrap ([5]). Tapered block bootstraps are shown to have asymptotic validity and favorable bias properties for smooth functions of means and M-estimators. Instead of using the block bootstrap, tapering is typically applied to the random weights in the bootstrapped empirical distribution ([15]; [14]). A detailed discussion of optimally selecting window shapes and block sizes were also presented along with some finite-sample simulations by [14].

A new block bootstrap procedure for time series data called the extended tapered block bootstrap

for estimating the variance and estimating the sampling distribution of a large category of approximately linear statistics was proposed by [20]. The paper established the consistency of the distribution approximation under the smooth function model by obtaining asymptotic bias and variance expansions. The extended tapered block bootstrap has a wider applicability than the tapered block bootstrap, while preserving the favorable bias and mean squared error properties of the tapered block bootstrap. A small simulation study was done to compares the performance of block-based bootstrapping methods on a finite-sample basis. ARIMA methodology was introduced by [2]. Before then, statisticians analyzed time series data without taking into account how non-stationarity might affect their analyses. It was shown that non-stationary data could be made stationary by "differencing" the time series data. In this way, one could pull apart a juicy trend at a specific time period from a growth/decline that would be expected anyway. For an instance, given the non-stationarity of the time series data. They also stressed that the Partial Autocorrelation Function (PAF) ϕ_k is non-zero for k less than or equal to ρ and zero for k greater than ρ for an autoregressive process of order ρ. In other words, a cut-off after lag k is present in the partial autocorrelation function (PAF) of the ρ^{th} order autoregressive process. The autoregressive model is typically expressed as:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \ldots + \phi_p X_{t-p} + \varepsilon_t \tag{1}$$

The complexity of this model was given a relief by [2]. When an autocorrelation coefficient deviates from their confidence range, they established a cut off. The Box and Jenkins $ARIMA$ approaches are predicated on the notion that a time series with highly correlated successive values can be viewed as having been produced by a string of independent shocks.

Using dependent data, [9] developed a novel fast bootstrap theory. In our scheme, smoothed moment indicators were resampled. Effectiveness of this method is demonstrated for parametric as well as semi-parametric estimation problems. The novel method's asymptotic improvements demonstrate that it is higher-order correct under reasonable time series, estimating function, and smoothing kernel assumptions. The use and benefits of the generalized technique of moments estimation, generalized empirical likelihood estimation, and the M-estimation method were shown by [9] using the method described in this article. The autoregressive conditional duration approach was put up against other current, frequently used first- and higher-order correct methods in a Monte Carlo research. The innovative bootstrap generates higher-order accurate confidence intervals while being computationally lighter than higher-order correct rivals. A real-data example on the dynamics of trading volumes of US stocks serves as an excellent example of the empirical applicability of our methodology. It was pointed by [4] that a brand-new bootstrap method based on generative adversarial networks for time series data (GANs). They show that GANs can understand the dynamics of typical stationary time series processes and that they can be used to produce additional samples from processes using GANs trained on a single sample path. A vector chosen from a normal distribution with a zero mean and an identity variance-covariance matrix can be utilized in the study to create credible samples, and temporal convolutional neural networks have a design that works well as both a generator and a discriminator. The simulations used in the article to evaluate the performance of the recommended bootstrap to circular block bootstrapping when resampling an AR(1) time series process also highlight the finite sample features of GAN sampling. According to the study, resampling with the GAN can provide better empirical coverage than circular block bootstrapping. The Sharpe ratio was given an empirical application at the end.

2. Methods

2.1. The Existing and Proposed Methods

The existing method considered in this research is the Moving Block Bootstrap (MBB) upon which the proposed method called the Moving Block Bootstrap with better element Representation

(MBBR) is built. Detail of the existing and the proposed methods is discussed below.

The Existing Method

The MBB is an extension of Non-overlapping Block Bootstrap (NBB) presented in [3] in which rooms are provided for observations in the original observation(s) that are cut off from the last block because, n which is the sample size may not be divisible by l, the block length. It invariably provides for more number of blocks than the NBB method. Given a time series data, $\{X_t\} = \{x_1, x_2, \cdots, x_n\}$ which follows an AR, MA or ARMA process presented in Equation 7 and the statistic of interest $\hat{\theta}_{MBB} = f(X_t) = f(x_1, x_2, \cdots, x_n)$, B_ι is defined as the block of size l consecutive observations starting from x_ι, that is:

$$B_\iota = \{X_\iota, \ldots, X_{\iota+l-1}\}; \quad \iota = 1, 2, \cdots, b \text{ for } \begin{cases} \lim\limits_{l,n\to\infty} \frac{l}{n} \to \infty \\ b = n - l + 1 \end{cases} \tag{2}$$

Let $b = n - l + 1$ and $n^* = k \times l$, while k and n^* are positive integers, such that n^* is the smallest multiple of $l \geqslant n$. A random sample of k blocks, $(n < k < \infty)$, $\{B_1^*, B_2^*, \cdots, B_k^*\}$ is resampled independently with replacement from $\{B_1, \cdots, B_b\} \sim unif(1, b)$ with probability of b^{-1}; where each $B_\iota, \iota = 1, 2, \cdots, k$ has a block of size l; $(n > l > 1)$. If $l = 1$, the block bootstrap returns to independent and identically distributed bootstrap originally proposed by [5]. Figure 1 provides a schematic representation of Equation 2 as a blocking scheme for moving block bootstrap.

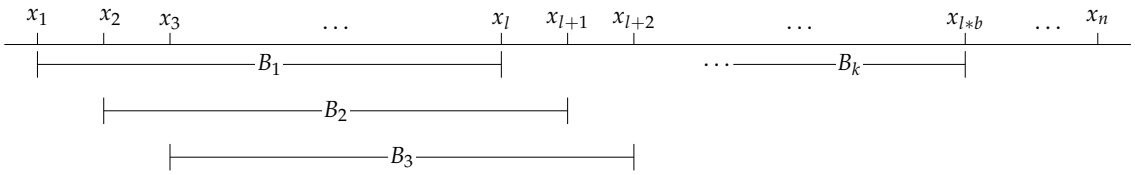


Figure 1: *Moving Block Bootstrap*

The MBB is then formed by collapsing elements of B_l^*; $\iota = 1, 2, \cdots, k$ into a single time series data to form

$$\begin{aligned} \{X_\iota^*\} &= \{x^*_{(\iota-1)l+1}, \cdots, x^*_{n^*}\} \\ &= \{x_1^*, x_2^*, \cdots, x_{n^*}^*\} \end{aligned} \tag{3}$$

where $n^* = k \times l$ length of the resampled series. The resampled statistics is then calculated according to $\hat{\theta}_{MBB} = f_{n^*}(\{x_1^*, x_2^*, \cdots, x_{n^*}^*\})$. One keep varying the block length to test for the minimum $\sqrt{(MSE)}$ For instance, given a time series data as $X_t = x_1, x_2, \cdots, x_{10}$, is chunked into equal block length of 3 as can be seen in array of Equation 4.

$$X_t = \begin{pmatrix} x_1 & x_2 & x_3 & & & & & & & \\ & x_2 & x_3 & x_4 & & & & & & \\ & & x_3 & x_4 & x_5 & & & & & \\ & & & x_4 & x_5 & x_6 & & & & \\ & & & & x_5 & x_6 & x_7 & & & \\ & & & & & x_6 & x_7 & x_8 & & \\ & & & & & & x_7 & x_8 & x_9 & \\ & & & & & & & x_8 & x_9 & x_{10} \end{pmatrix} \begin{matrix} \mathbf{B_1} \\ \mathbf{B_2} \\ \mathbf{B_3} \\ \mathbf{B_4} \\ \mathbf{B_5} \\ \mathbf{B_6} \\ \mathbf{B_7} \\ \mathbf{B_8} \end{matrix} \tag{4}$$

The Proposed Method

Consider modifications to the following already existing block resampling techniques for dependent data: the MBB in [8]. The reason for modification is to introduce equal number of presence of every element of the parent time series in the blocking procedure. Take for instance,

given a time series data as $X_t = x_1, x_2, \cdots, x_{10}$, is chunked into blocks as follows according to each method (MBB and MBBR). The Moving Block Bootstrap Method with better element Representation (MBBR) is formed to reduce the less representative presence of extreme member of the time series data from $2l$ to just 2. as can be seen by comparing Equation 4 and 6. Reduction of less-represented elements of the time series data helps to increase the performance of model evaluation metrics (RMSE and MAE). The MBBR method is an extension of MBB method from [8]. Given a time series data, $\{X_t\} = \{x_1, x_2, \cdots, x_n\}$ which follows an AR MA or ARMA process of Equation 7 and the statistic of interest $\hat{\theta}_{MBB} = f(X_t) = f(x_1, x_2, \cdots, x_n)$, B_ι is defined as the block of size l consecutive observations starting from x_ι, that is:

$$B_\iota = \{X_\iota, \ldots, X_{\iota+(\iota-1)}\}; \quad \iota = 1, 2, \cdots, b \text{ for } \begin{cases} \lim\limits_{l,n\to\infty} \frac{l}{n} \to \infty \\ b = n - 2(l-1) \end{cases} \tag{5}$$

Let $b = n - 2(l-1)$ and $n^* = k \times l$, while k and n^* are positive integers, such that n^* is the smallest multiple of $l \geqslant n$.

A random sample of k blocks, $(n < k < \infty)$, $\{B_1^*, B_2^*, \cdots, B_k^*\}$ is resampled independently with replacement from $\{B_1, \cdots, B_b\} \sim unif(1, b)$ with probability of b^{-1}; where each $B_\iota, \iota = 1, 2, \cdots, k$ has a block of size l; $(n > l > 1)$. If $l = 1$, the block bootstrap returns to independent and identically distributed bootstrap originally proposed by [5]. Figure 1 provides a schematic representation of Equation 5 as a blocking scheme for moving block bootstrap. The array in Equation 6 demonstrates the churning procedure of a typical time series data of $n = 10$ sample size with a block length of $l = 3$. This proposed method can be executed with the help of an R package in [7] using R programming Language in [19] as demonstrated in Listing 1.

$$X_t = \begin{pmatrix} x_1 & x_2 & x_3 & & & & & & & \\ & x_2 & x_3 & x_4 & & & & & & \\ & & & x_4 & x_5 & x_6 & & & & \\ & & & & x_5 & x_6 & x_7 & & & \\ & & & & & & x_7 & x_8 & x_9 & \\ & & & & & & & x_8 & x_9 & x_{10} \end{pmatrix} \begin{matrix} \mathbf{B_1} \\ \mathbf{B_2} \\ \mathbf{B_3} \\ \mathbf{B_4} \\ \mathbf{B_5} \\ \mathbf{B_6} \end{matrix} \tag{6}$$

Listing 1: *OBL Package for Minimum RMSE Values*

```
> install.packages(OBL)
> df <- OBL::blockboot(ts, 10, 123, 4)
> df$RMSE[2]
# [1] 0.3398036
> df$RMSE[4]
# [1] 0.3303526
```

2.2. Data Generation for Simulation Study

2.2.1 Time Series Data Using ARIMA Model Methodology

Recall that:

$$X_t = x_1, x_2, \cdots, x_n \tag{7}$$

obtained from an AR(p), MA(q) or ARMA(p, q) process is given as follows:

$$X_t = \mu + \sum_{t=1}^{p} \phi_j X_{t-j} + \sum_{t=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon)$$
$$\text{for} \begin{cases} j = 1, 2, \cdots, p; \\ k = 1, 2, \cdots, q; \\ t = 1, 2, \cdots, n; \\ 0 \leqslant |\phi_j| \leqslant 1; \\ 0 \leqslant |\theta_k| \leqslant 1; \\ \phi_j \neq \theta_k; \\ \sum_{t=1}^{n} \varepsilon_t = 0; \\ \varepsilon_t \sim N(0, \sigma^2); \\ \sigma_\varepsilon^2 = \sigma^2; \\ |\mu| \geq 0. \end{cases} \tag{8}$$

The use of "R" package demonstrated in R code Listing 2 below shows commands written for "R" to grid-search for a "seed" that produces exactly or approximately an ARIMA(1, 0, 0) with $\varphi = 0.8$ and sample size ($n = 10$) starting from seed 280000 to 290000 with an increase of one unit. From the grid-search result printed, seed 289805 is deemed appropriate for the example. Similar effort is put in for every time series data simulated for this study to ensure that the output of each simulated time series data depict its specified parameters.

Listing 2: *Seed Searching Using "ARIMASS" Package in R*

```
devtools::install_github("sta189332/searchar")
# @example
# searchar::arsearch(a = 280000,  z = 290000, n = 10, p = 1, d = 0, q = 0, ar11 = 0.8, sd = 1,
    j1 = 4, arr1 = "0.80")
#          ar1    seed
# 7 0.8079816 282327
# 5 0.8062789 283176
# 6 0.8074425 284165
# 8 0.8081475 284461
# 4 0.8026127 287720
# 9 0.8084755 288160
# 3 0.8023778 289053
# 1 0.8000000 289805
# 2 0.8000368 289989
```

Haven got the set of program seeds that simulates the specified parameters of the ARIMA time series data that is described in each simulated data, one can then use such program seed to simulate the desired ARIMA time series data. In Listing 3 bellow, "seed(289805)" is used as a demonstration to simulate for the $ARIMA(1,0,0)$ with sample size ($n = 10$) and $\varphi = 0.8$ in line 1 while on line 2 one checked for the empirical characteristics of the $ARIMA(1,0,0)$ to make sure that the simulated data does not just come from the population of $ARIMA(1,0,0)$ and $\varphi = 0.8$ alone but that the simulated data itself has the characteristics one calls for. The below paragraphs show how time series data for AR, MA and ARMA are simulated with there respective parameters.

Listing 3: *Illustration of ARIMA Time Series Data Simulation*

```
set.seed(289805)
ar_1 <- stats::arima.sim(n = 10, model = list(ar = c(0.8), order = c(1, 0, 0)), sd = 1)
forecast::auto.arima(ar_1, ic = "aicc")
# Series: ar_1
# ARIMA(1,0,0) with zero mean
# Coefficients:
# ar1
# 0.8000
```

Data Simulation for Autoregressive (AR) Model

Forty eight (48) time series data were simulated to follow (AR(1) models with different parameters coefficients $\varphi = (0.8, 0.9, 0.95)$, sample size $n = (10, 15, 20, 25)$ and standard deviation of varying levels values $sd = (1, 3, 5, 10)$. Stationary conditions are confirmed to be true for AR(1) time series data with different levels of φ values in Equation 9. $AR(1)$ models of consideration are spelled out in the below Equation 9 with their different levels of autocorrelation (φ), standard deviation (σ) and sample sizes (n).

$$\begin{aligned} \hat{X}_t =& \varphi \hat{X}_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon) \\ \text{for } & \begin{cases} \varphi = 0.8, 0.9, 0.95; \\ \sigma_\varepsilon = 1, 3, 5, 10; \\ n = 10, 15, 20, 25; \\ |\varphi| \leqslant 1. \end{cases} \end{aligned} \tag{9}$$

Forty eight (48) time series data were simulated to follow AR(2) models with different parameters coefficients $\varphi = ((0.4, 0.4), (0.45, 0.45), (0.35, 0.6))$, sample size $n = (10, 15, 20, 25)$ and standard deviation of varying levels values $sd = (1, 3, 5, 10)$. Stationary conditions are confirmed to be true for $AR(2)$ time series data at different levels of φ values; $\varphi = (0.4, 0.4), (0.45, 0.45), (0.35, 0.6)$ in Equation 10.

$$\begin{aligned} \hat{X}_t =& \varphi_1 \hat{X}_{t-1} + \varphi_2 \hat{X}_{t-2} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon) \\ \text{for } & \begin{cases} \varphi_1 = 0.4, 0.45, 0.35; \\ \varphi_2 = 0.4, 0.45, 0.6; \\ \sigma_\varepsilon = 1, 3, 5, 10; \\ |\varphi_1| \leqslant 1; \\ |\varphi_2| \leqslant 1; \\ n = 10, 15, 20, 25; \\ |\varphi_1| + |\varphi_2| \leqslant 1; \\ |\varphi_1| - |\varphi_2| \leqslant 1. \end{cases} \end{aligned} \tag{10}$$

Data Simulation for Moving Average (MA) Model

Forty eight (48) time series data were simulated to follow MA(1) models with different parameters coefficients $\vartheta = (0.8, 0.9, 0.95)$, sample size $n = (10, 15, 20, 25)$ and standard deviation of varying levels values $sd = (1, 3, 5, 10)$. Stationary conditions are confirmed to be true for MA(1) time series data with different levels of ϑ values in Equation 11. $MA(1)$ models of consideration are spelled out in the below Equation 11 with their different levels of autocorrelation and standard deviation.

$$\begin{aligned} \hat{X}_t =& \vartheta \hat{X}_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon) \\ \text{for } & \begin{cases} \vartheta = 0.8, 0.9, 0.95; \\ \sigma_\varepsilon = 1, 3, 5, 10; \\ n = 10, 15, 20, 25; \\ |\varphi| \leqslant 1. \end{cases} \end{aligned} \tag{11}$$

Forty eight (48) time series data were simulated to follow MA(2) models with different parameters coefficients $\vartheta = ((0.4, 0.4), (0.45, 0.45), (0.35, 0.6))$, sample size $n = (10, 15, 20, 25)$ and standard deviation of varying levels values $sd = (1, 3, 5, 10)$. Stationary conditions are confirmed to be true

for $AR(2)$ time series data at different levels of ϑ values; $\vartheta = (0.4, 0.4), (0.45, 0.45), (0.35, 0.6)$ in Equation 12.

$$\begin{aligned} \hat{X}_t =& \vartheta_1 \hat{X}_{t-1} + \vartheta_2 \hat{X}_{t-2} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon) \\ \text{for } & \begin{cases} \vartheta_1 = 0.4, 0.45, 0.35; \\ \vartheta_2 = 0.4, 0.45, 0.6; \\ \sigma_\varepsilon = 1, 3, 5, 10; \\ n = 10, 15, 20, 25; \\ |\vartheta_1| \leqslant 1; \\ |\vartheta_2| \leqslant 1; \\ |\vartheta_1| + |\vartheta_2| \leqslant 1; \\ |\vartheta_1| - |\vartheta_2| \leqslant 1. \end{cases} \end{aligned} \tag{12}$$

Data Simulation for Autoregressive Moving Average (ARMA) Model

Forty eight (48) time series data were simulated to follow ARMA(1, 1)) models with different parameters coefficients $\psi = ((0.5, 0.3), (0.5, 0.4), (0.35, 0.6))$, sample size $n = (10, 15, 20, 25)$ and standard deviation of varying levels values $sd = (1, 3, 5, 10)$. Stationary conditions are confirmed to be true for $ARMA(1, 1)$ time series data at different levels of ψ values; $\psi = (0.4, 0.4), (0.45, 0.45), (0.35, 0.6)$ in Equation 13.

$$\begin{aligned} \hat{X}_t =& \phi X_{t-1} + \theta \varepsilon_{t-1}; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon) \\ \text{for } & \begin{cases} \phi = 0.5, 0.5, 0.35; \\ \theta = 0.3, 0.4, 0.6; \\ \sigma_\varepsilon = 1, 3, 5, 10; \\ n = 10, 15, 20, 25; \\ |\phi| < 1; \\ |\theta| < 1; \\ \phi \neq \theta. \end{cases} \end{aligned} \tag{13}$$

2.3. Criteria for Model and Method Selection

In order to choose the better-performing method between the two methods (MBB and MBBR) discussed above, Root Mean Squared Error (RMSE) is used to choose the best-performing model that results in the better method. Note that [18] deployed RMSE to choose the best-performing method for forecasting the carbon dioxide (CO_2) emission of Bahrain. Mean Absolute Error (MAE) is another metric used in this paper for method evaluation. To evaluate the robustness of data-model comparisons, [11] concluded that RMSE is not enough, rather MAE or other relevant measures are needed. The most common accuracy-fit-performance metric is RMSE. Prediction quality is often evaluated by the root mean square error or root mean square deviation. Based on Euclidean distance, it shows how far predictions differ from measured true values. To calculate RMSE, calculate the residual (difference between prediction and truth) for each data point, compute its norm, calculate its mean, and then take its square root. As RMSE relies on and requires true measurements at every predicted data point.

$$RMSE = \sqrt{\frac{1}{n} \sum_{\iota=1}^{n} (X_\iota - \hat{X}_\iota)^2} \tag{14}$$

$$MAE = \frac{1}{n} \sum_{\iota=1}^{n} |X_\iota - \hat{X}_\iota| \tag{15}$$

From the above Equations 14 and 15, n is the number of data points, X_t is the i-th measurement, and $\hat{X}_t$ is its corresponding prediction. MAE also has the same units as RMSE. Usually, MAE is smaller than RMSE, although it can be the opposite if the predicted values are very close to the observed ones.

3. Results

Methods Comparison for AR(1) Models

It can be seen in Table 1 and Figure 2 that MBBR method has smaller RMSE and MAE values indicating that the newly proposed method (MBBR) presents a better quality of prediction than the existing method (MBB). It can also be seen in Table 1 and Figure 2 that RMSE and MAE values increase as the value of standard deviation increases. Comparing the values of RMSE in Table 1 with the values of MAE in Figure 2 pairwise, MAE values are smaller than that of RMSE. The proposed method (MBBR) has minimum elements yet has better representation and more efficient than the existing method MBB as seen in 1 and Figure 2.

Methods Comparison for AR(2) Models

It can be seen in Table 2 and Figure 3 that MBBR method has smaller RMSE and MAE values indicating that the newly proposed method (MBBR) presents a better quality of prediction than the existing method (MBB). It can also be seen in Table 2 and Figure 3 that RMSE and MAE values increase as the value of standard deviation increases. Comparing the values of RMSE in Table 2 with the values of MAE in Figure 3 pairwise, MAE values are smaller than that of RMSE. The proposed method (MBBR) has minimum elements yet has better representation and more efficient than the existing method MBB as seen in 2 and Figure 3.

Methods Comparison for MA(1) Models

It can be seen in Table 3 and Figure 4 that MBBR method has smaller RMSE and MAE values indicating that the newly proposed method (MBBR) presents a better quality of prediction than the existing method (MBB). It can also be seen in Table 3 and Figure 4 that RMSE and MAE values increase as the value of standard deviation increases. Comparing the values of RMSE in Table 3 with the values of MAE in Figure 4 pairwise, MAE values are smaller than that of RMSE. The proposed method (MBBR) has minimum elements yet has better representation and more efficient than the existing method MBB as seen in 3 and Figure 4.

Methods Comparison for MA(2) Models

It can be seen in Table 4 and Figure 5 that MBBR method has smaller RMSE and MAE values indicating that the newly proposed method (MBBR) presents a better quality of prediction than the existing method (MBB). It can also be seen in Table 4 and Figure 5 that RMSE and MAE values increase as the value of standard deviation increases. Comparing the values of RMSE in Table 4 with the values of MAE in Figure 5 pairwise, MAE values are smaller than that of RMSE. The proposed method (MBBR) has minimum elements yet has better representation and more efficient than the existing method MBB as seen in 4 and Figure 5.

Methods Comparison for ARMA(1, 1) Models

It can be seen in Table 5 and Figure 6 that MBBR method has smaller RMSE and MAE values indicating that the newly proposed method (MBBR) presents a better quality of prediction than the existing method (MBB). It can also be seen in Table 5 and Figure 6 that RMSE and MAE values increase as the value of standard deviation increases. Comparing the values of RMSE in Table 5 with the values of MAE in Figure 6 pairwise, MAE values are smaller than that of RMSE.

The proposed method (MBBR) has minimum elements yet has better representation and more efficient than the existing method MBB as seen in 5 and Figure 6.

Table 1: *Minimum RMSE Criterion of MBB and MBBR Methods for AR(1)*

	$\varphi = 0.8$				$\varphi = 0.9$				$\varphi = 0.95$				
n	10	15	20	25	10	15	20	25	10	15	20	25	
MBB	1.58	0.89	1.50	1.85	0.89	1.20	1.45	1.39	1.26	0.88	1.29	1.01	$sd = 1$
MBBR	1.58	0.80	1.43	1.85	0.83	1.01	1.40	1.34	1.12	0.17	1.21	0.87	
MBB	4.75	2.67	4.65	5.55	2.67	3.60	4.72	4.18	3.79	3.21	4.01	3.03	$sd = 3$
MBBR	4.73	2.41	4.29	5.55	2.50	3.04	4.19	4.02	3.37	2.63	3.64	2.60	
MBB	7.92	4.45	7.76	9.26	4.45	6.01	7.87	6.97	6.32	5.34	6.69	5.0	$sd = 5$
MBBR	7.89	4.02	7.15	9.26	4.17	5.07	6.98	6.70	5.62	4.39	6.07	4.34	
MBB	15.83	8.90	15.52	18.52	8.90	12.01	15.74	13.93	12.65	10.68	13.38	10.11	$sd = 10$
MBBR	15.77	8.04	14.30	18.51	8.34	10.13	13.96	13.41	11.23	8.78	12.14	8.68	

Table 2: *Minimum RMSE Criterion of MBB and MBBR Methods for AR(2)*

	$\varphi_1 = 0.4,\ \varphi_2 = 0.4$				$\varphi_1 = 0.45,\ \varphi_2 = 0.45$				$\varphi_1 = 0.35,\ \varphi_2 = 0.6$				
n	10	15	20	25	10	15	20	25	10	15	20	25	
MBB	0.81	1.37	1.11	1.28	1.15	1.08	1.47	1.26	1.22	1.09	1.39	1.29	$sd = 1$
MBBR	2.42	4.1	3.33	3.83	3.45	3.25	4.4	3.79	3.67	3.27	4.18	3.88	
MBB	2.42	4.1	3.33	3.83	3.45	3.25	4.4	3.79	3.67	3.27	4.18	3.88	$sd = 3$
MBBR	2.28	4.05	3.28	3.32	3.3	3.16	4.3	3.54	3.66	3.15	3.70	3.39	
MBB	4.03	6.83	5.55	6.38	5.74	5.42	7.34	6.32	6.12	5.45	6.97	6.46	$sd = 5$
MBBR	3.8	6.76	5.46	5.53	5.55	5.27	7.17	5.9	6.11	5.25	6.17	5.66	
MBB	12.1	13.66	11.11	12.76	11.48	10.85	14.68	12.65	12.24	10.89	13.94	12.92	$sd = 10$
MBBR	8.06	13.51	10.93	11.05	11	10.55	14.34	11.81	12.21	10.5	12.33	11.31	

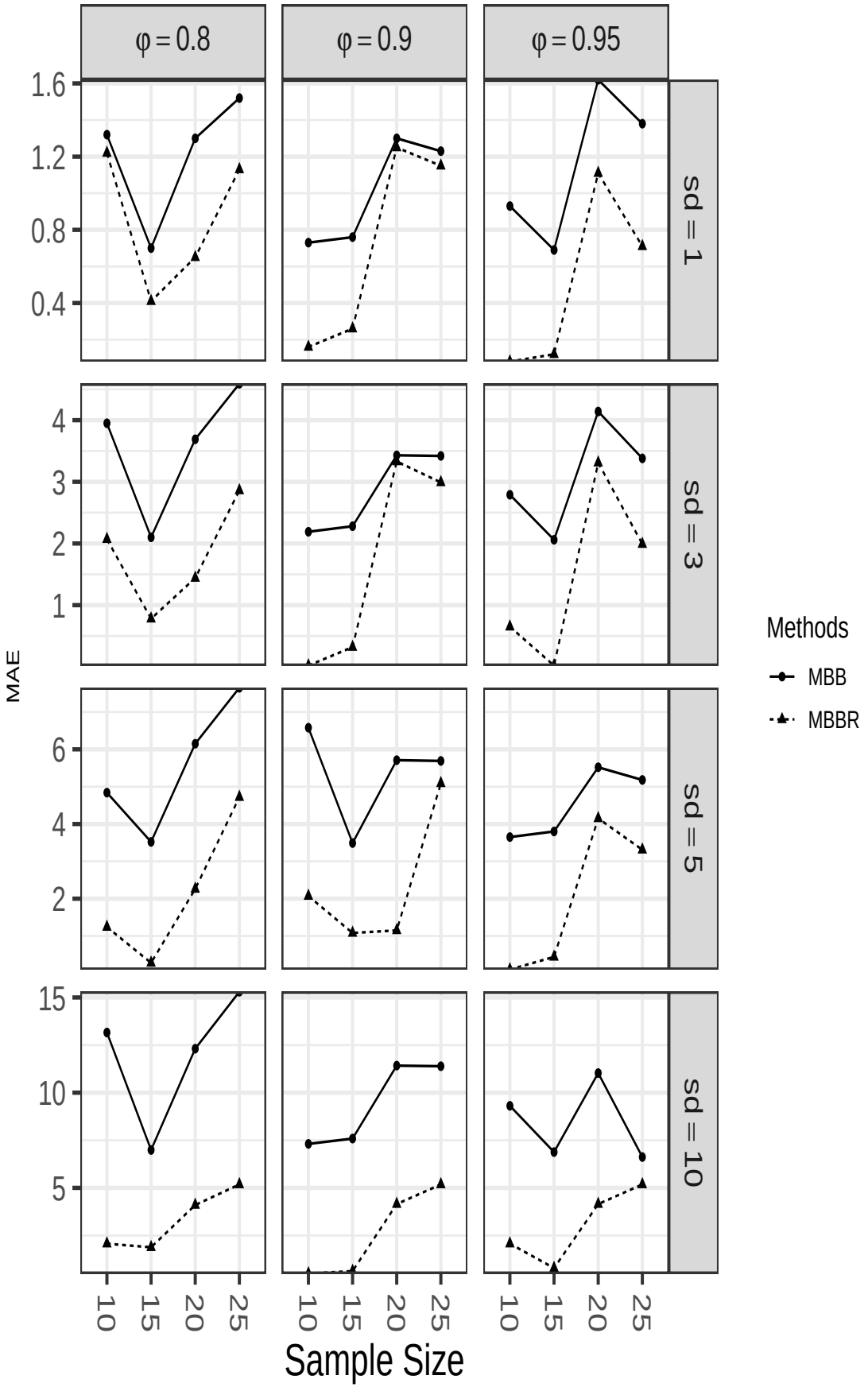


Figure 2: *Facet-Line Plot Showing Minimum MAE of MBB and MBBR Methods for AR(1)*

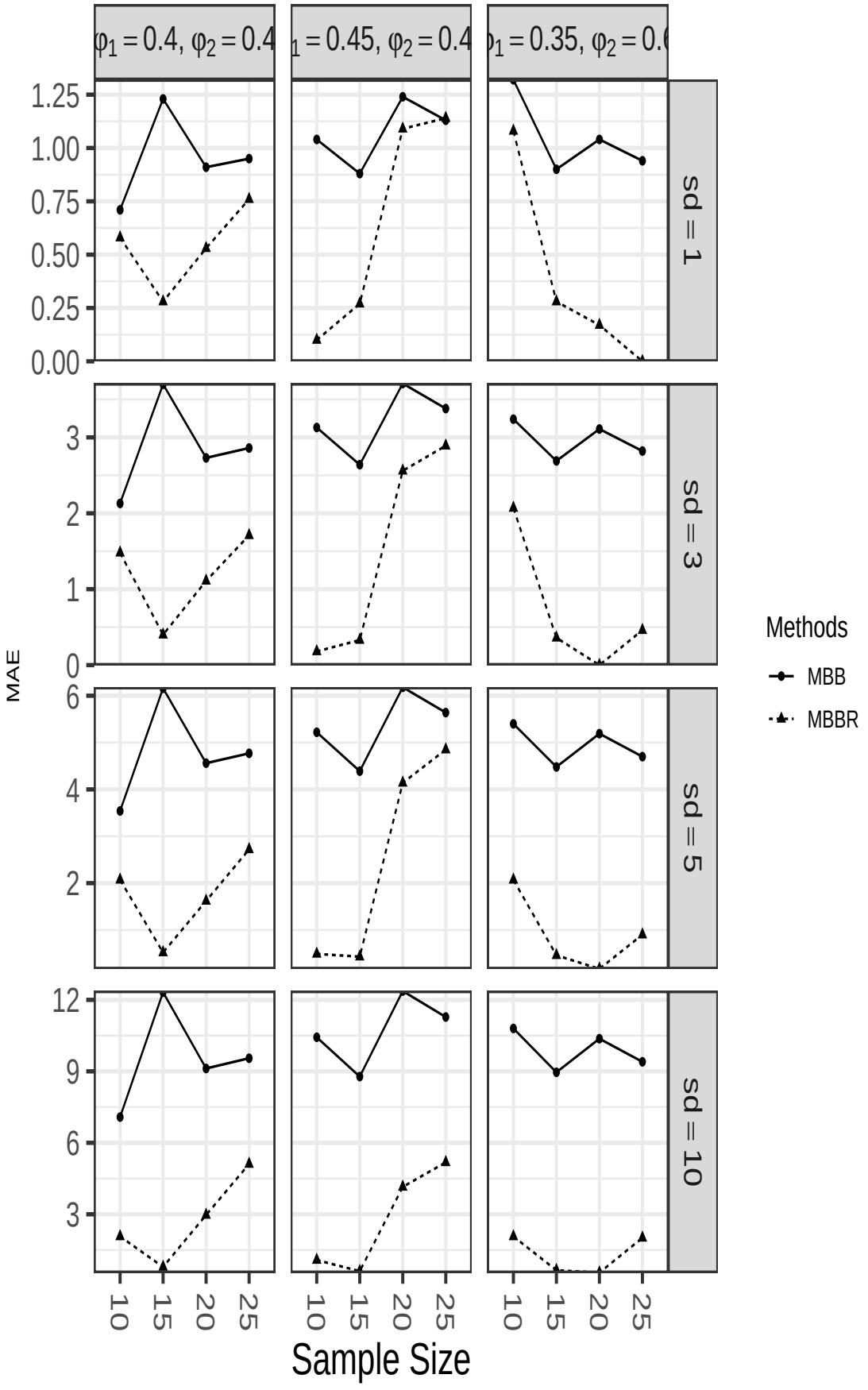


Figure 3: *Facet-Line Plot Showing Minimum MAE of MBB and MBBR Methods for AR(2)*

Table 3: *Minimum RMSE Criterion of MBB and MBBR Methods for MA(1)*

	$\vartheta = 0.8$				$\vartheta = 0.9$				$\vartheta = 0.95$				
n	10	15	20	25	10	15	20	25	10	15	20	25	
MBB	0.77	0.96	1.29	1.63	1.21	0.98	1.25	1.27	1.46	1.26	1.12	1.31	$sd = 1$
MBBR	0.72	0.94	1.29	1.57	1.13	0.92	1.24	1.16	1.42	1.23	1.09	1.31	
MBB	2.30	2.88	3.88	4.89	3.63	2.93	3.76	3.8	4.38	3.77	3.37	3.93	$sd = 3$
MBBR	2.15	2.81	3.84	4.71	3.39	2.76	3.25	3.49	4.25	3.68	3.17	3.93	
MBB	3.83	4.80	6.46	8.15	6.05	4.88	6.27	6.33	7.3	6.29	5.62	6.55	$sd = 5$
MBBR	3.59	4.68	6.40	7.85	5.64	4.59	5.42	5.81	7.17	6.13	5.28	6.54	
MBB	7.66	9.61	12.92	16.31	12.1	9.76	12.53	12.66	14.6	12.58	11.23	13.1	$sd = 10$
MBBR	7.17	9.36	12.81	15.7	11.28	9.19	10.83	11.63	14.17	12.26	10.55	13.08	

Table 4: *Minimum RMSE Criterion of MBB and MBBR Methods for MA(2)*

	$\vartheta_1 = 0.4,\ \vartheta_2 = 0.4$				$\vartheta_1 = 0.45,\ \vartheta_2 = 0.45$				$\vartheta_1 = 0.35,\ \vartheta_2 = 0.6$				
n	10	15	20	25	10	15	20	25	10	15	20	25	
MBB	0.86	1.35	1.22	1.27	1.74	1.29	1.49	1.3	1.23	1.28	1.17	1.23	$sd = 1$
MBBR	0.90	1.35	1.21	1.27	1.73	1.29	1.49	1.3	1.23	1.23	1.16	1.22	
MBB	2.69	4.04	3.67	3.8	5.22	3.87	4.48	3.91	3.70	3.83	3.52	3.69	$sd = 3$
MBBR	2.59	4.04	3.62	3.8	5.19	3.87	4.48	3.9	3.70	3.69	3.49	3.65	
MBB	4.49	6.74	6.12	6.34	8.69	6.45	7.47	6.52	6.17	6.38	5.87	6.15	$sd = 5$
MBBR	4.32	6.74	6.03	6.33	8.66	6.45	7.47	6.5	6.17	6.15	5.82	6.08	
MBB	11.77	13.48	12.24	12.67	17.38	12.91	14.94	12.99	12.34	12.75	11.74	12.30	$sd = 10$
MBBR	8.97	13.48	12.07	12.65	17.31	12.91	14.94	13.01	12.33	12.31	11.65	12.16	

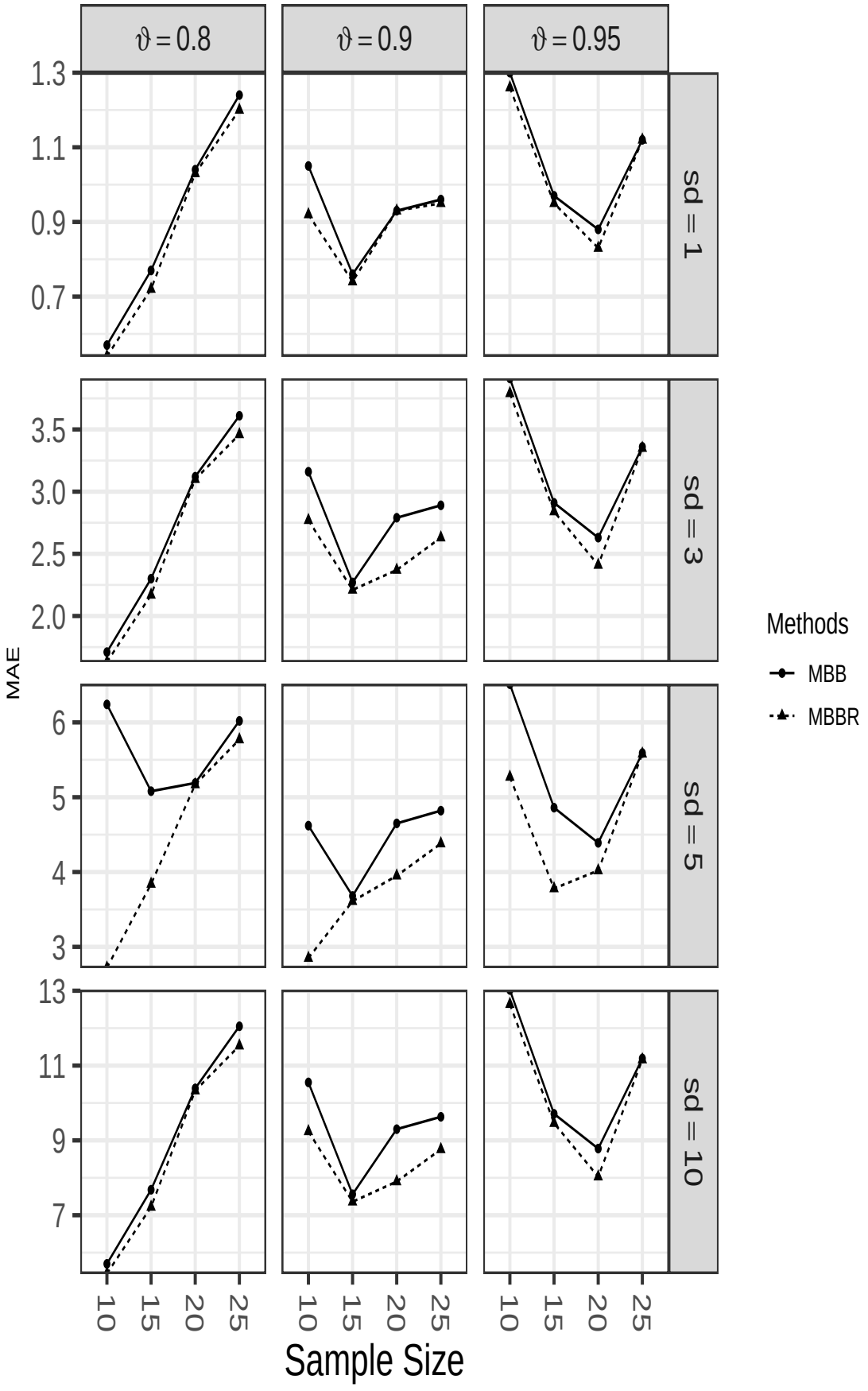


Figure 4: *Facet-Line Plot Showing Minimum RMSE of MBB and MBBR Methods for MA(1)*

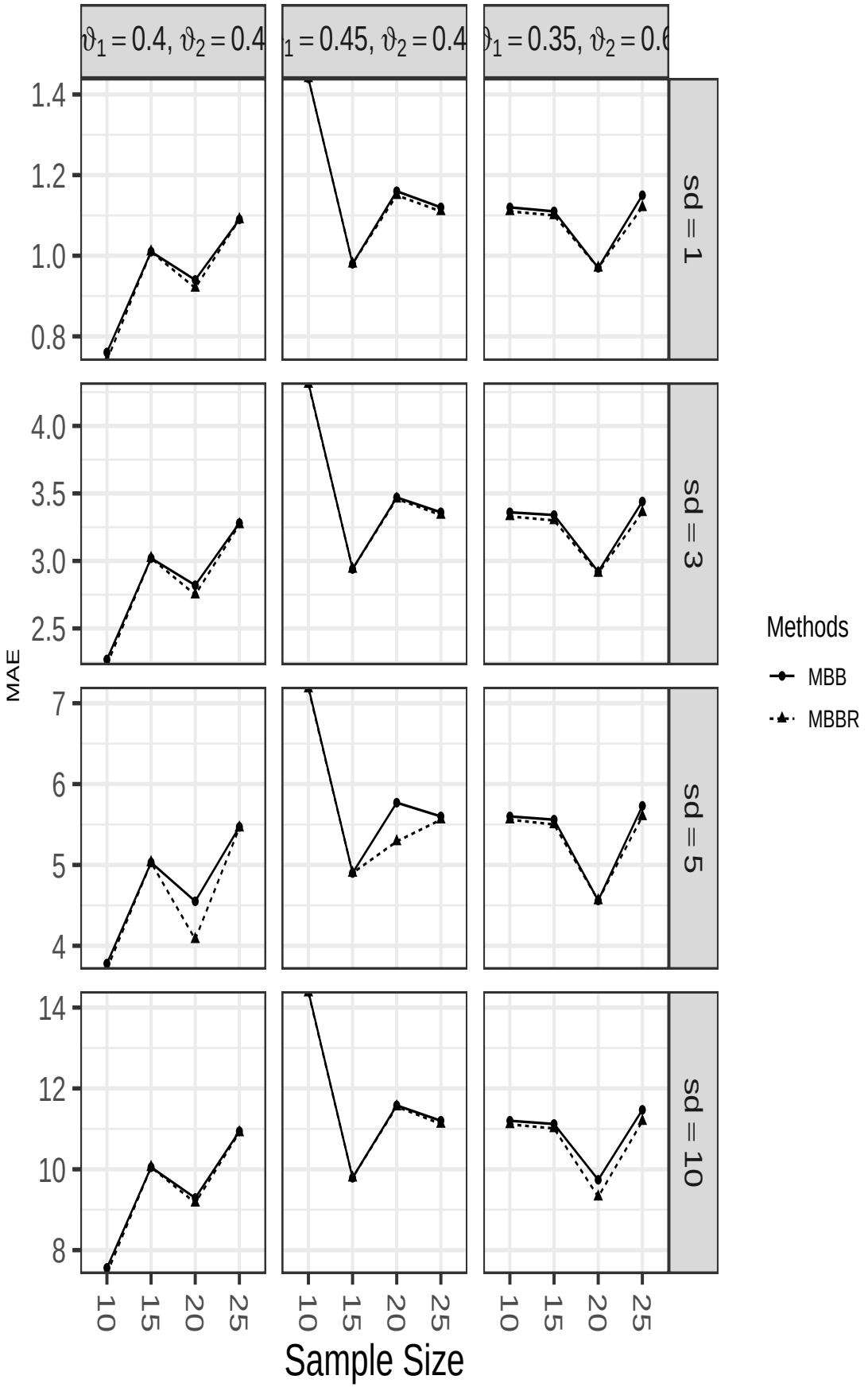


Figure 5: *Facet-Line Plot Showing Minimum MAE of MBB and MBBR Methods for MA(2)*

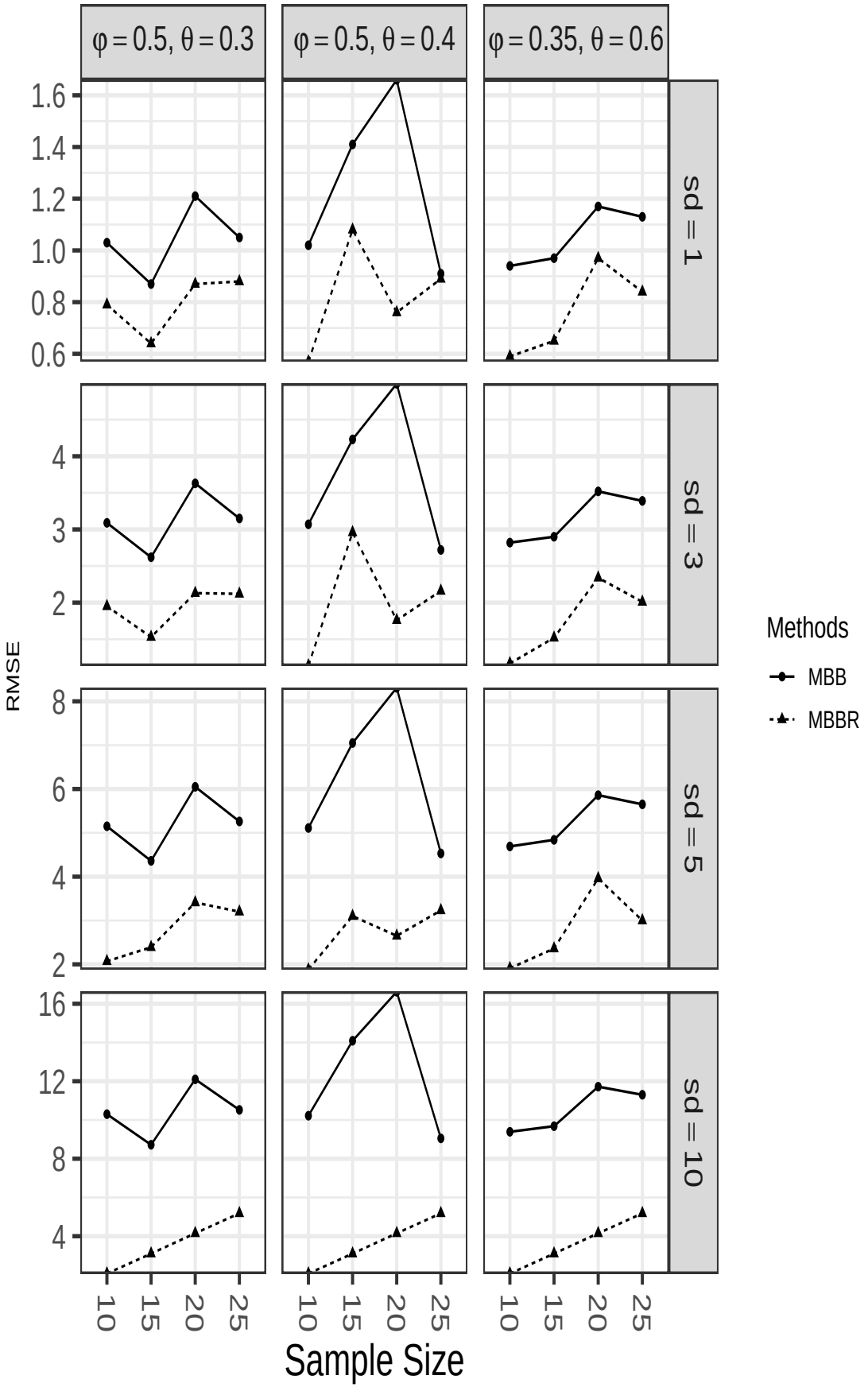


Figure 6: *Facet-Line Plot Showing Minimum MAE of MBB and MBBR Methods for ARMA(1, 1)*

Table 5: *Minimum RMSE Criterion of MBB and MBBR Methods for ARMA(1, 1)*

	$\varphi_1 = 0.5,\ \vartheta_2 = 0.3$				$\varphi_1 = 0.5,\ \vartheta_2 = 0.4$				$\varphi_1 = 0.35,\ \vartheta_2 = 0.6$				
n	10	15	20	25	10	15	20	25	10	15	20	25	
MBB	1.23	1.09	1.5	1.35	1.29	1.62	1.95	1.07	1.27	1.19	1.43	1.34	$sd = 1$
MBBR	1.17	1.01	1.49	1.21	1.2	1.62	1.93	1.04	1.11	1.12	1.41	1.31	
MBB	3.68	3.27	4.5	4.05	3.87	4.87	5.84	3.2	3.81	3.57	4.29	4.03	$sd = 3$
MBBR	3.5	3.03	4.48	3.62	3.61	4.85	5.78	3.12	3.34	3.37	4.24	3.93	
MBB	6.14	5.45	7.5	6.75	6.45	8.11	9.73	5.34	6.34	5.96	7.15	6.72	$sd = 5$
MBBR	5.83	5.06	7.47	6.03	6.01	8.08	9.63	5.2	5.57	5.62	7.07	6.56	
MBB	12.28	10.91	15	13.49	12.9	16.22	19.45	10.67	12.69	11.91	14.3	13.44	$sd = 10$
MBBR	12.08	10.11	14.95	12.06	12.02	16.17	19.27	10.4	11.14	11.23	14.15	13.11	

4. Discussion

It is important to mention interesting points that emerge from this study. The study shows that the newly proposed method (MBBR) represents a better quality of prediction than the existing method (MBB). It is also shown that neither the sample size nor the model parameter(s) has significant impact on the accuracy measures. It is also worthy of note that the varying level of standard deviation has a direct and positive impact the values of accuracy measure. In general, MAE values are smaller than their corresponding values of RMSE. The proposed method has minimum elements yet has better representation than the existing method. Although this research focus mainly on comparison of Moving Block Bootstrap method with the newly proposed (Moving Block Bootstrap with better element representation), Future research will focus on comparing the proposed method with other existing methods of block bootstrap.

References

[1] Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews*, 19(1):1–48.

[2] Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

[3] Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The annals of statistics*, pages 1171–1179.

[4] Dahl, C. M. and Sørensen, E. N. (2022). Time series (re) sampling using generative adversarial networks. *Neural Networks*.

[5] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7(1):569–593.

[6] Horowitz, J. L. (2003). Bootstrap methods for markov processes. *Econometrica*, 71(4):1049–1082.

[7] James, D. and Kayode, A. (2022). *OBL: Optimum Block Length*. R package version 0.2.1.

[8] Kunsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, pages 1217–1241.

[9] La Vecchia, D., Moor, A., and Scaillet, O. (2022). A higher-order correct fast moving-average bootstrap for dependent data. *Journal of Econometrics*.

[10] Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404.

[11] Liemohn, M. W., Shane, A. D., Azari, A. R., Petersen, A. K., Swiger, B. M., and Mukhopadhyay, A. (2021). Rmse is not enough: Guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, 218:105624.

[12] Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, 225.

[13] Paparoditis, E. and Politis, D. (2002a). The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal*, 5(1):131–148.

[14] Paparoditis, E. and Politis, D. (2002b). The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal*, 5(1):131–148.

[15] Paparoditis, E. and Politis, D. N. (2001). Tapered block bootstrap. *Biometrika*, 88(4):1105–1119.

[16] Politis, D. and Romano, J. (1992). Circular block'resampling procedure for stationary data, rin lepage, r. and billard, l.(eds) exploring the limits of bootstrap. *Wiley: New York*, 263:270.

[17] Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, pages 2031–2050.

[18] Qader, M. R., Khan, S., Kamal, M., Usman, M., and Haseeb, M. (2022). Forecasting carbon emissions due to electricity power generation in bahrain. *Environmental Science and Pollution Research*, 29(12):17346–17357.

[19] R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[20] Shao, X. (2010). Extended tapered block bootstrap. *Statistica Sinica*, pages 807–821.

RT&A, No.3 (74)
Volume 18, September 2023

RELIABILITY AND SENSITIVITY ANALYSIS OF A SYSTEM WITH CONDITIONAL AND EXTENDED WARRANTY

Kajal Sachdeva

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
kajal.rs.maths@mdurohtak.ac.in

Gulshan Taneja

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
drgtaneja@gmail.com

Amit Manocha*

•

Department of Applied Sciences, TITS Bhiwani, Haryana, India
amitmanocha80@yahoo.com

*Corresponding Author

Abstract

System reliability and maintenance cost are the most crucial and decisive factors influencing consumers' buying behaviour. The manufacturer attempts to address the consumer concern by offering a warranty in accordance with the reliability of the system and maintenance costs. This study aims to examine the stochastic behaviour of a single unit system operating in three different time frames, namely normal, extended and expired warranty time duration. The system user can prolong the normal warranty period at an extra cost. This prolonged warranty is termed an 'Extended Warranty'. However, the manufacturer provides a warranty on a system with certain conditions. If the failures are covered under the warranty conditions, the repair/replacement is done free of cost; otherwise, all charges are borne by the system user. Markov and regenerative processes are used to derive the system's reliability and other performability measures. Time distributions used in the study are taken as arbitrary. The profit function for the manufacturer and the user is formulated and analysed. Sensitivity analysis for system availabilities in different time zones and profit functions is also done. Numerical examples for exponential, Weibull and Erlang time distributions are discussed to illustrate the derived measures.

Keywords: Reliability; Extended warranty; Regenerative process; Profit function; Sensitivity analysis.

1. Introduction

The recent advancement in various technological aspects has paved the way for endless technologies and innovations to hit the mainstream, forcing manufacturers to offer a gamut of consumer options. Despite the numerous advantages of technological development, the flip side is the system's complexity. Consequently, the consumers are apprehensive of system reliability which may adversely affect the sales of the developed product. Hence, in the pursuit of ensuring system

reliability and addressing the consumer's concerns and dilemmas, the manufacturer offers a warranty and extended warranty. System reliability and warranty may be considered as interrelated concepts. Warranty is the written agreement provided to system users for cumulative product acceptance without any strain of product manufacturing or faults. Ives and Vitale [1], Ritchken et al. [2], Singpurwalla and Wilson [3] proposed that the warranty on various products caters to risk reduction, quality bench-marking and enhanced market competitiveness. Different types of warranty policies were introduced to optimize profit. Free Replacement Warranty, Full-Service Warranty and Renewing Pro-Rata Warranty were described by Blischke, and Murthy [4], Jain and Maheshwari [5], Bai and Pham [6] respectively. Huang et al. [7] discussed the future problems and challenges in reliability and warranty. Kadyan and Ramniwas [8] proposed a probabilistic model for a single-unit system protected by warranty conditions. Rahman and Chattopadhyay [9] developed cost models on long-term/service contract policies. Pham and Bai [10] discussed warranty costs, compared different warranty policies and evaluated the warranty benefits. The warranties offered to consumers can safeguard them from exorbitant maintenance costs considering the ever-increasing maintenance cost resulting from advancements. A conditional warranty covers the cost associated with defects stipulated in the agreement at the time of purchase.
Taneja [11] developed a stochastic model in which repair/replacement is done by the manufacturer on predefined warranty conditions. Lei et al. [12] discussed the characterization of warranty price policies and optimized the product price, which is profitable to the user. Further, Solkhe and Taneja [13] considered the system with conditional warranty and compared the performability measures before and after the expiry of warranty periods. Niwas [14] analysed a warranted system with waiting time for repair and cost paid by the user if failures happen due to unauthorised modifications. Hooti et al. [15] optimised the warranty duration and the repair. Another type of warranty, known as an extended warranty, is worth considering these days. There are systems used in daily life for which manufacturers, dealers or third parties provide extended warranties, such as home appliances and electronic equipment, particularly in the automotive industry. Though optional, it offers users a sense of security regarding system reliability, maintenance costs, etc. Suiter and Lorson [16] described the pros and cons of an extended warranty for the system user. Huang [17] considered a system with minor, degraded and catastrophic failures where warranty was provided for either one year at a fixed lump sum price or monthly warranty plan which may be extended for another month. Salmasnia and Hatami [18] considered a model with an extended warranty where the failures are controlled using technology and non-periodic maintenance activities.
However, stochastic modelling of a system with prolonged conditional warranty and identification of key parameters influencing the most on the manufacturer/user profit is yet to be reported in the literature. So, the objective of our study is to stochastically analyze a system functioning in normal, extended and beyond warranty periods with a focus on its reliability characteristics and economic viability. The study also identifies the parameter that significantly impacts system profitability from the manufacturer and user perspectives. This article is structured as follows. System description and assumptions made are described in Section 2. Various notations used in the study are cited in Section 3. A probabilistic model for the described system is developed in Section 4. Transition probabilities and mean sojourn times are also evaluated in this Section. Expressions for reliability indices and metrics impacting the system's profitability are derived in Sections 5, 6 and 7, respectively. In Section 8, the profit functions for the manufacturer and the user are established. Section 9 focused on the sensitivity analysis of availabilities and profit functions. To illustrate the developed model, numerical examples for different density functions are discussed in Section 10. Conclusions regarding reliability, system availabilities, the profitability of the manufacturer/user and their sensitivity are also drawn in this Section. Finally, Section 11 provides several concluding insightful interpretations.

2. System Descriptions and Assumptions

Descriptions for the system under consideration and assumptions made for the analysis are as follows:

1. System has a single unit that operates in three different warranty periods named normal, extended and expiry warranty periods.
2. The failed system is inspected by the manufacturer or external source to ensure
   (a) Whether faults occur in a system comes under warranty claims or not.
   (b) Whether the system is repairable or needs replacement.
3. If an inspection reveals that a fault falls inside the purview of a normal or extended warranty conditions, the manufacturer is liable for paying the cost of repair/replacement. Otherwise, all the charges are borne by the user itself.
4. In the expiry (beyond warranty) period, the user manages the expenditure for repair/replacement.
5. Transition time distributions have been taken general.

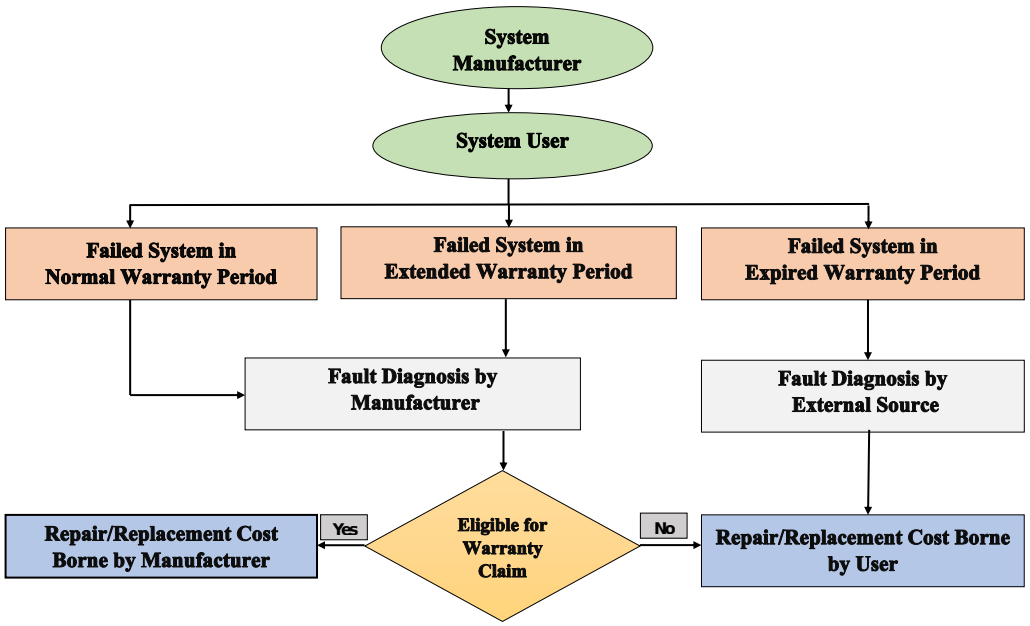


Figure 1: *System description.*

Figure 1 shows the description of the system. Markov and regenerative processes are applied to develop a stochastic model for the system defined. Mathematical expressions for reliability, Mean time to failure (MTTF), availabilities, expected busy period of repairman, and the number of replacements are derived. Profit functions are formulated. Sensitivity analysis is also done for availabilities in three different time zones and for the profit function of the manufacturer and the user. Exponential, Weibull and Erlang distributions are used for numerical calculations. Various conclusions on reliability indices, profitability and sensitivity, are drawn.

3. Nomenclature

The notations for various probabilities/transition densities are:

S_0	state of system at t=0
$\odot$	symbol for Laplace transform
p/q	probability that fault is covered/not covered by warranty terms.
$p_1/p_2/p_3$	probability that system gets failed during normal/extended/ expired warranty period.
r_1/r_2	probability that fault is repairable/ irreparable and only to be replaced.
$f(t)$	p.d.f. of failure time.
$i_m(t)/i_u(t)$	p.d.f. of inspection time by the repairman engaged by manufacturer/ user itself.
$g_k(t)/h_k(t)$	p.d.f. of repair/replacement time within warranty period 'k=n/et/ex'
$I_i^k(t)$	P{repairman of manufacturer is engaged in inspection at instant t $\mid S_0 = i$ within warranty period 'k'}.
$BM_i^k(t)(BU_i^k(t))$	P{manufacturer repairman is occupied with repair/replacement when the associated cost are to be paid by the manufacturer (user) at instant t $\mid S_0 = i$ within warranty period 'k'}.
$RM_i^k(t)(RU_i^k(t))$	expected number of replacement in (0,t], when expenses are met by manufacturer(user) $\mid S_0 = i$ during warranty period 'k'.

where k denotes normal(n), extended(et), expired(ex).

4. Stochastic Model

The probable states of described system are:

State 0: (O_{nw});	State 1: (F_{in});	State 2: $(F_{rn}^{(m)})$;
State 3: $(F_{rpn}^{(m)})$;	State 4: $(F_{rn}^{(u)})$;	State 5: $(F_{rpn}^{(u)})$;
State 6: (F_{iet});	State 7: $(F_{ret}^{(u)})$;	State 8: $(F_{rpet}^{(u)})$;
State 9: $(F_{ret}^{(m)})$;	State 10: $(F_{rpet}^{(m)})$;	State 11: (O_{etw});
State 12: (F_{iex});	State 13: $(F_{rex}^{(u)})$;	State 14: $(F_{rpex}^{(u)})$;
State 15: (O_{exw});		

where,

O_{kw}	operative system in warranty period 'k'.
F_{ik}	failed system under inspection in warranty period 'k'.
$F_{rn}^{(m)}(F_{rpn}^{(m)})/F_{ret}^{(m)}(F_{rpet}^{(m)})$	failed system under repair(replacement) in normal/extended warranty period, for which expenses are to be borne by manufacturer.
$F_{rk}^{(u)}/F_{rpk}^{(u)}$	failed system under repair/ replacement in warranty period 'k', for which charges are to be borne by user itself.

Here, k denotes normal(n), extended(et), expired(ex).
By employing markov and regenerative process, the transition between various states is represented by Figure 2. The state space constitutes the set of regenerative states i.e., S={0, 1, 2, ..., 15}, where O={0, 11, 15} is operative, and F={1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14} is failed state space respectively.
The transition densities from state i to state j ($Q_{ij}(t)$) are:

$$Q_{01}(t) = \int_0^t dF(u), \quad Q_{12}(t) = pr_1 \int_0^t dI_m(u), \quad Q_{13}(t) = pr_2 \int_0^t dI_m(u),$$
$$Q_{14}(t) = qr_1 \int_0^t dI_m(u), \quad Q_{15}(t) = qr_2 \int_0^t dI_m(u), \quad Q_{20}(t) = \int_0^t dG_n(u),$$
$$Q_{30}(t) = \int_0^t dH_n(u), \quad Q_{40}(t) = \int_0^t dG_n(u), \quad Q_{50}(t) = \int_0^t dH_n(u),$$
$$Q_{06}(t) = \int_0^t dF(u), \quad Q_{67}(t) = qr_1 \int_0^t dI_m(u), \quad Q_{68}(t) = qr_2 \int_0^t dI_m(u),$$

$Q_{69}(t) = pr_1 \int_0^t dI_m(u)$, $Q_{6,10}(t) = pr_2 \int_0^t dI_m(u)$, $Q_{7,11}(t) = \int_0^t dG_{et}(u)$,
$Q_{8,11}(t) = \int_0^t dH_{et}(u)$, $Q_{9,11}(t) = \int_0^t dG_{et}(u)$, $Q_{10,11}(t) = \int_0^t dH_{et}(u)$,
$Q_{0,12}(t) = \int_0^t dF(u)$, $Q_{11,12}(t) = \int_0^t dF(u)$, $Q_{11,6}(t) = \int_0^t dF(u)$,
$Q_{12,13}(t) = r_1 \int_0^t dI_u(u)$, $Q_{12,14}(t) = r_2 \int_0^t dI_u(u)$, $Q_{13,15}(t) = \int_0^t dG_{ex}(u)$,
$Q_{14,15}(t) = \int_0^t dH_{ex}(u)$, $Q_{15,12}(t) = \int_0^t dF(u)$.

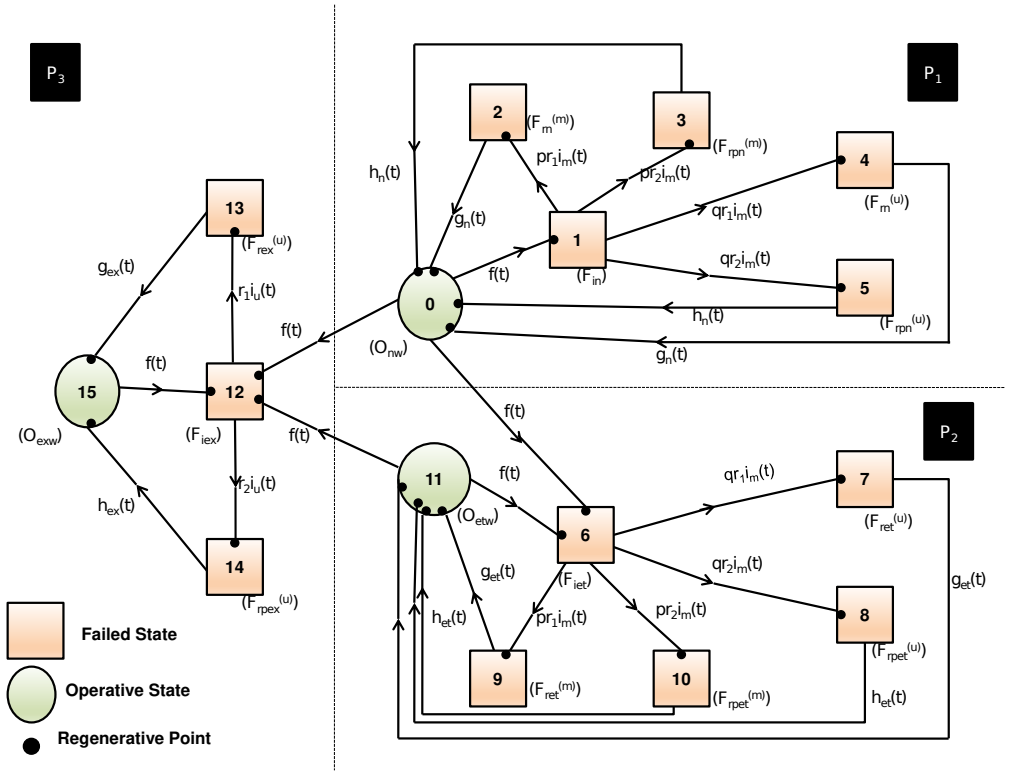


Figure 2: *State transition diagram*

Thus, transition probabilities from state i to j are:

$$p_{ij} = \lim_{s \to 0} q_{ij}^*(s), \quad where \quad q_{ij}(t) = \frac{dQ_{ij}(t)}{dt}.$$

These probabilities follow the property of transition probability matrix for each measure of system effectiveness.
Mean sojourn time (μ_i) in state i are:

$$\mu_0 = \int_0^\infty t f(t) dt = \int_0^\infty \overline{F}(t) dt.$$

Similarly,

$$\mu_1 = \int_0^\infty \overline{I}_m(t) dt, \quad \mu_2 = \int_0^\infty \overline{G}_n(t) dt, \quad \mu_3 = \int_0^\infty \overline{H}_n(t) dt, \quad \mu_4 = \int_0^\infty \overline{G}_n(t) dt,$$
$$\mu_5 = \int_0^\infty \overline{H}_n(t) dt, \quad \mu_6 = \int_0^\infty \overline{I}_m(t) dt, \quad \mu_7 = \int_0^\infty \overline{G}_{et}(t) dt, \quad \mu_8 = \int_0^\infty \overline{H}_{et}(t) dt,$$
$$\mu_9 = \int_0^\infty \overline{G}_{et}(t) dt, \quad \mu_{10} = \int_0^\infty \overline{H}_{et}(t) dt, \quad \mu_{11} = \int_0^\infty \overline{F}(t) dt, \quad \mu_{12} = \int_0^\infty \overline{I}_u(t) dt,$$
$$\mu_{13} = \int_0^\infty \overline{G}_{ex}(t) dt, \quad \mu_{14} = \int_0^\infty \overline{H}_{ex}(t) dt, \quad \mu_{15} = \int_0^\infty \overline{F}(t) dt.$$

Defining m_{ij}= $E(q_{ij}(t))$= $\int_0^\infty tq_{ij}(t)dt$, we have

$$m_{01} = \int_0^\infty tq_{01}(t)dt = \int_0^\infty tf(t)dt = \mu_0.$$

Similarly, we get the following relations

$m_{06} = m_{0,12} = \mu_0$, $m_{20} = \mu_2$, $m_{30} = \mu_3$, $m_{40} = \mu_4$,
$m_{50} = \mu_5$, $m_{7,11} = \mu_7$, $m_{8,11} = \mu_8$, $m_{9,11} = \mu_9$,
$m_{10,11} = \mu_{10}$, $m_{11,6} = m_{11,12} = \mu_{11}$, $m_{12,13} + m_{12,14} = \mu_{12}$, $m_{13,15} = \mu_{13}$,
$m_{14,15} = \mu_{14}$, $m_{15,12} = \mu_{15}$, $m_{12} + m_{13} + m_{14} + m_{15} = \mu_1$,
$m_{67} + m_{68} + m_{69} + m_{6,10} = \mu_6$.

The reliability and performance-indicating characteristics of the system are determined in the following section.

5. Reliability (R(t)) and Mean Time to Failure (MTTF)

Theorem 1. R(t) and MTTF are given as

$$R(t) = \overline{F(t)}, \qquad MTTF = \mu_0.$$

Proof. Let $\psi_0(t)$=P[system is operative until time t $\mid$ $S_0 = 0$], then using probabilistic arguments, it can be seen from transition diagram, that

$$\psi_0(t) = p_1Q_{01}(t) + p_2Q_{06}(t) + p_3Q_{0,12}(t). \tag{1}$$

The expression on R.H.S. of equation (1) shows the system transit from state 0 to failed state 1 or 6 or 12, with probability $p_1Q_{01}(t)$, $p_2Q_{06}(t)$ and $p_3Q_{0,12}(t)$ respectively in time t. Taking Laplace-Stieltjes transformation of the above equation, we get

$$\psi_0^{**}(s) = p_1Q_{01}^{**}(s) + p_2Q_{06}^{**}(s) + p_3Q_{0,12}^{**}(s). \tag{2}$$

Thus,

$$\begin{aligned}
R(t) &= L^{-1}[\frac{1-\psi_0^{**}(s)}{s}] \\
&= L^{-1}[\frac{1-(p_1Q_{01}^{**}(s) + p_2Q_{06}^{**}(s) + p_3Q_{0,12}^{**}(s))}{s}] \qquad [Substituting(2)] \\
&= L^{-1}[\frac{1}{s}] - p_1L^{-1}[\frac{Q_{01}^{**}(s)}{s}] - p_2L^{-1}[\frac{Q_{06}^{**}(s)}{s}] - p_3L^{-1}[\frac{Q_{0,12}^{**}(s)}{s}] \\
&= 1 - p_1\int_0^t dF(u) - p_2\int_0^t dF(u) - p_3\int_0^t dF(u) \quad [Using\ transition\ densities] \\
&= 1 - (p_1+p_2+p_3)\int_0^t dF(u) \\
&= 1 - \int_0^t dF(u) \quad (\because p_1+p_2+p_3 = 1) \\
&= 1 - \int_0^t f(u)du = 1 - F(t) = \overline{F(t)}.
\end{aligned} \tag{3}$$

and

$$\begin{aligned}
MTTF &= \lim_{s\to 0}[\frac{1-\psi_0^{**}(s)}{s}] \quad [\frac{0}{0}] \\
&= -\psi_0^{**'}(0).
\end{aligned} \tag{4}$$

Differentiating eq^n (2) both sides w.r.t. s and then taking $\lim s \to 0$, we get

$$\begin{aligned} \psi_0^{**'}(0) &= p_1 Q_{01}^{**'}(0) + p_2 Q_{06}^{**'}(0) + p_3 Q_{0,12}^{**'}(0) \\ &= -\mu_0. \end{aligned} \tag{5}$$

From eq^n (4) and (5),

$$\text{MTTF}=\mu_0.$$

■

6. System Availability

Theorem 2. The Laplace transformation of point-wise availability during **extended warranty period** is given by

$$A_0^{et*}(s) = \frac{N_1^{et*}(s)}{D_1^{et*}(s)},$$

where
$N_1^{et*}(s)$=

$$\begin{vmatrix} M_0^*(s) & -q_{06}^*(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -q_{67}^*(s) & -q_{68}^*(s) & -q_{69}^*(s) & -q_{6,10}^*(s) & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) \\ 0 & 0 & 0 & 1 & 0 & 0 & -q_{8,11}^*(s) \\ 0 & 0 & 0 & 0 & 1 & 0 & -q_{9,11}^*(s) \\ 0 & 0 & 0 & 0 & 0 & 1 & -q_{10,11}^*(s) \\ M_{11}^*(s) & -q_{11,6}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

$M_0^*(s) = M_{11}^*(s) = \frac{1-f^*(s)}{s}$,

$D_1^{et*}(s)$=

$$\begin{vmatrix} 1 & -q_{06}^*(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -q_{67}^*(s) & -q_{68}^*(s) & -q_{69}^*(s) & -q_{6,10}^*(s) & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) \\ 0 & 0 & 0 & 1 & 0 & 0 & -q_{8,11}^*(s) \\ 0 & 0 & 0 & 0 & 1 & 0 & -q_{9,11}^*(s) \\ 0 & 0 & 0 & 0 & 0 & 1 & -q_{10,11}^*(s) \\ 0 & -q_{11,6}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

The system steady-state availability is given as

$$A_0^{et} = \frac{N_1^{et}}{D_1^{et}},$$

where

$$N_1^{et} = \mu_{11},$$

and

$$D_1^{et} = \mu_{11} + \mu_6 + \mu_7 q r_1 + \mu_8 q r_2 + \mu_9 p r_1 + \mu_{10} p r_2.$$

Proof. Considering $A_0^{et}(t)$ = P(system is operative at time t in extended warranty period | $S_0 = 0$), then from transition diagram, we have

$$\begin{aligned} A_0^{et}(t) &= \overline{F(t)} + \int_0^t q_{06}(u) A_6^{et}(t-u)du \\ &= M_0(t) + q_{06}(t) \odot A_6^{et}(t) \end{aligned} \tag{6}$$

where $M_0(t)$ represents that the system remains operative in state 0 instead of moving to any other state. The term $q_{06}(t)$ denotes the system transition probability from state 0 to state 6 in time u<t and thereafter remains operative from state 6 onwards for t-u time.
Similarly,

$$\begin{cases} A_6^{et}(t) = q_{67}(t) \odot A_7^{et}(t) + q_{68}(t) \odot A_8^{et}(t) + q_{69}(t) \odot A_9^{et}(t) + q_{6,10}(t) \odot A_{10}^{et}(t), \\ A_7^{et}(t) = q_{7,11}(t) \odot A_{11}^{et}(t), \\ A_8^{et}(t) = q_{8,11}(t) \odot A_{11}^{et}(t), \\ A_9^{et}(t) = q_{9,11}(t) \odot A_{11}^{et}(t), \\ A_{10}^{et}(t) = q_{10,11}(t) \odot A_{11}^{et}(t), \\ A_{11}^{et}(t) = M_{11}(t) + q_{11,6}(t) \odot A_6^{et}(t). \end{cases} \tag{7}$$

Taking Laplace Transformation of eq^n(6)-(7) and solving them for $A_0^{et*}(s)$ by method of determinants, we get

$$A_0^{et*}(s) = \frac{L_1(s)}{M_1(s)},$$

$$\begin{aligned} L_1(s) &= M_o^*(s) + M_{11}^*(s)q_{06}^*(s)q_{67}^*(s)q_{7,11}^*(s) + M_{11}^*(s)q_{06}^*(s)q_{68}^*(s)q_{8,11}^*(s) \\ &\quad + M_{11}^*(s)q_{06}^*(s)q_{69}^*(s)q_{9,11}^*(s) + M_{11}^*(s)q_{06}^*(s)q_{6,10}^*(s)q_{10,11}^*(s) \\ &\quad - M_o^*(s)q_{67}^*(s)q_{11,6}^*(s)q_{7,11}^*(s) - M_o^*(s)q_{68}^*(s)q_{11,6}^*(s)q_{8,11}^*(s) \\ &\quad - M_o^*(s)q_{69}^*(s)q_{11,6}^*(s)q_{9,11}^*(s) - M_o^*(s)q_{11,6}^*(s)q_{6,10}^*(s)q_{10,11}^*(s) \\ &= N_1^{et*}(s) \end{aligned} \tag{8}$$

$$\begin{aligned} M_1(s) &= 1 - q_{68}^*(s)q_{11,6}^*(s)q_{8,11}^*(s) - q_{69}^*(s)q_{11,6}^*(s)q_{9,11}^*(s) - q_{11,6}^*(s)q_{6,10}^*(s)q_{10,11}^*(s) \\ &\quad - q_{67}^*(s)q_{11,6}^*(s)q_{7,11}^*(s) \\ &= D_1^{et*}(s) \end{aligned} \tag{9}$$

Using Abel's lemma, the system's steady state availability is

$$A_0^{et} = \lim_{s \to 0} sA_0^{et*}(s) = \frac{N_1^{et*}(0)}{D_1^{et*\prime}(0)} = \frac{N_1^{et}}{D_1^{et}}, \tag{10}$$

Differentiating eq^n (9) w.r.t. s,

$$\begin{aligned} D_1^{et*\prime}(s) &= q_{11,6}^{*\prime}(s)(-q_{67}^*(s)q_{7,11}^*(s) - q_{68}^*(s)q_{8,11}^*(s) - q_{69}^*(s)q_{9,11}^*(s) - q_{6,10}^*(s)q_{10,11}^*(s)) \\ &\quad - q_{67}^{*\prime}(s)q_{7,11}^*(s)q_{11,6}^*(s) - q_{68}^{*\prime}(s)q_{8,11}^*(s)q_{11,6}^*(s) - q_{69}^{*\prime}(s)q_{9,11}^*(s)q_{11,6}^*(s) \\ &\quad - q_{6,10}^{*\prime}(s)q_{10,11}^*(s)q_{11,6}^*(s) - q_{7,11}^{*\prime}(s)q_{11,6}^*(s)q_{6,7}^*(s) \\ &\quad - q_{8,11}^{*\prime}(s)q_{11,6}^*(s)q_{6,8}^*(s) - q_{9,11}^{*\prime}(s)q_{11,6}^*(s)q_{6,9}^*(s) \\ &\quad - q_{10,11}^{*\prime}(s)q_{11,6}^*(s)q_{6,11}^*(s). \end{aligned} \tag{11}$$

Setting lim s $\to$ 0 in eq^n (8) and (11), we obtain

$$N_1^{et} = \mu_{11}, \tag{12}$$

$$\begin{aligned} D_1^{et} &= \mu_{11}(pr_1 + pr_2 + qr_1 + qr_2) + pr_1\mu_9 + m_{69} + pr_2\mu_{10} + m_{6,10} + qr_1\mu_7 \\ &\quad + m_{67} + m_{68} + \mu_8 qr_2 \\ &= \mu_{11} + \mu_6 + \mu_9 pr_1 + \mu_{10} pr_2 + \mu_7 qr_1 + \mu_8 qr_2. \end{aligned} \tag{13}$$

■

Similarly, availabilities during normal and expired warranty period are given as

$$A_0^n = \frac{N_1^n}{D_1^n}, \qquad A_0^{ex} = \frac{N_1^{ex}}{D_1^{ex}}. \tag{14}$$

where

$$\begin{aligned} N_1^n &= \mu_0, D_1^n = \mu_0 + \mu_1 + \mu_2 pr_1 + \mu_3 pr_2 + \mu_4 qr_1 + \mu_5 qr_2, \\ N_1^{ex} &= \mu_{15}, D_1^{ex} = \mu_{15} + \mu_{12} + \mu_{13} r_1 + \mu_{14} r_2. \end{aligned} \tag{15}$$

7. Expected Busy Period and Number of Replacements

Employing the definitions of BU_i^k, BM_i^k and I_i^k, i ∈ S (defined in Section 3) and follow the same probabilistic arguments as discussed in preceding Section 6, the expected time for which repairman remain involved in repair/replacement/inspection of a failed system in different warranty zones, in steady-state given as:

$BU_0^k = \frac{N_2^k}{D_1^k}$; $BM_0^k = \frac{N_3^k}{D_1^k}$; $I_0^k = \frac{N_4^k}{D_1^k}$; k=n,et,ex.

$N_2^n = q(\mu_4 r_1 + \mu_5 r_2)$; $N_2^{et} = q(\mu_7 r_1 + \mu_8 r_2)$; $N_2^{ex} = \mu_{13} r_1 + \mu_{14} r_2$;
$N_3^n = p(\mu_2 r_1 + \mu_3 r_2)$; $N_3^{et} = p(\mu_9 r_1 + \mu_{10} r_2)$; $N_4^n = \mu_1$;
$N_4^{et} = \mu_6$; $N_4^{ex} = \mu_{12}$;

Further, by definitions of RU_i^k and RM_i^k,i ∈ S (defined in Section 3), in steady state, the expected number of replacements during different warranty time zones are:

$RU_0^k = \frac{N_5^k}{D_1^k}$; $RM_0^k = \frac{N_6^k}{D_1^k}$; k=n,et,ex
$N_5^n = qr_2$; $N_5^{et} = qr_2$; $N_5^{ex} = r_2$;
$N_6^n = pr_2$; $N_6^{et} = pr_2$;

D_1^n, D_1^{et} and D_1^{ex} are mentioned in eq^n (13) and (15).
Focusing on economic viability of defined system, cost-benefit analysis is performed in the succeeding section. Profitability analysis assists both the manufacturer and the user in categorizing the parameters which may cause long-term loss.

8. Cost-Benefit Analysis

To carry out a cost-benefit analysis, the profit function is defined for the manufacturer and user. Mathematically, the profit function for a system is the difference between total revenue generated and total expenditure incurred in a given period. So, in steady-state, the profit function for the manufacturer and user is formulated as follows:

Manufacturer Profit

$$\begin{aligned} P^m &= CP + EC - MC - C_1^m(p_1 I_0^n + p_2 I_0^{et}) - C_2^m(p_1 BM_0^n + p_2 BM_0^{et}) \\ &\quad - C_3^m(p_1 RM_0^n + p_2 RM_0^{et}). \end{aligned} \tag{16}$$

User Profit

$$\begin{aligned} P^u &= C_0(p_1A_0^n + p_2A_0^{et} + p_3A_0^{ex}) - C_1^u(p_3I_0^{ex}) - C_2^u(p_1BU_0^n + p_2BU_0^{et} + p_3BU_0^{ex}) \\ &\quad -C_3^u(p_1RU_0^n + p_2RU_0^{et} + p_3RU_0^{ex}) - CP - EC. \end{aligned} \tag{17}$$

where,
CP=Cost price of the system
MC=Manufacturing cost of the system
EC=Cost for extended the warranty period.
C_0=Revenue generated by system.
$C_1^m(C_1^u)$= Cost of engaging the repairman by the manufacturer(user) for inspection.
$C_2^m(C_2^u)$=Cost of engaging the repairman by the manufacturer(user) for repair/replacement.
$C_3^m(C_3^u)$=Cost per replacement of system borne by manufacturer(user).
The above-mentioned costs are considered per unit of time.

9. Sensitivity Analysis

Sensitivity analysis is a concept that determines as to which parameter (independent variable) the obtained measures (dependent variable) are highly or least affected. Relative sensitivity analysis is used to assess the impact of different parameters because the numerical values for various parameters differ significantly. A normalised form of a sensitivity function is known as relative sensitivity function. Using the $eq^n s$ (10), (14), (16) and (17), the sensitivity (D_y^k, Z_y^s) and relative sensitivity functions (d_y^k, z_y^s) for availabilities (A_0^n, A_0^{et}, A_0^{ex}) and profit functions (P^m,P^u) respectively are defined as

$$D_y^k = \frac{\partial(A_0^k)}{\partial y}; \quad d_y^k = \frac{D_y^k y}{A_0^k}; \quad k = n, et, ex. \tag{18}$$

and

$$Z_y^s = \frac{\partial(P^s)}{\partial y}; \quad z_y^s = \frac{Z_y^s y}{P^s}; \quad s = u, m. \tag{19}$$

10. Results and Discussions

In this section, measures obtained in Sections 5-9 respectively are illustrated via numerical examples.
Expressions for various measures have been derived using general probability time distributions. The system user may choose specific values for the parameters involved to illustrate the model based on records of failures, repairs, costs, and available probabilities. The particular distribution can be identified by applying the appropriate test to such data. Because real data on failures, repairs, costs, etc., could not be collected in our study, we used exponential, Weibull and Erlang distributions and assumed values for parameters involved to illustrate the model numerically.

10.1. Example 1:All the Time Distributions follows Exponential Distribution

Assuming failure/inspection/replacement/repair times are exponentially distributed with their p.d.f. given as:

$$\begin{cases} f(t) = \lambda_0 e^{-\lambda_0 t}, & i_m(t) = \gamma_m e^{-\gamma_m t}, & i_u(t) = \gamma_u e^{-\gamma_u t}, & h_n(t) = \beta_n e^{-\beta_n t}, \\ h_{et}(t) = \beta_{et} e^{-\beta_{et} t}, & g_n(t) = \alpha_n e^{-\alpha_n t}, & g_{et}(t) = \alpha_{et} e^{-\alpha_{et} t}, & g_{ex}(t) = \alpha_{ex} e^{-\alpha_{ex} t}, \\ h_{ex}(t) = \beta_{ex} e^{-\beta_{ex} t}. \end{cases} \tag{20}$$

Consider the value of parameters as

$$p = 0.7, q = 0.3, p_1 = 0.2, p_2 = 0.3, p_3 = 0.5, r_1 = 0.7, r_2 = 0.3, \lambda_0 = 0.0005,$$
$$\gamma_m = 1.5, \gamma_u = 1.2, \alpha_n = 0.5, \alpha_{et} = 0.4, \beta_n = 0.02, \beta_{et} = 0.02, \alpha_{ex} = 0.25,$$
$$\beta_{ex} = 0.01, CP = 150, EC = 15, MC = 120, C_0 = 500, C_1^m = 80, C_2^m = 100,$$
$$C_3^m = 15,000, C_1^u = 100, C_2^u = 120, C_3^u = 15,000. \quad (21)$$

10.1.1 Effect of Time (t) and Failure Rate (λ_0) on Reliability Measures

Taking the values of other parameters constant as mentioned in eq^n (21), the effect of parameters (t, λ_0) on reliability function (R(t)) is shown in Figure 3.
As both the parameters t and λ_0 increase, R(t) decreases. Further, Table 1 represents the values of availabilities (A_0^n, A_0^{et},A_0^{ex}) in three different time zones for varied λ_0. All three availabilities decrease with an increase in λ_0. Also for any particular λ_0, the availabilities satisfy the relation $A_0^n > A_0^{et} > A_0^{ex}$. In other words, during the expired warranty period, the system's availability is much influenced by λ_0 as compared to the other time zones.

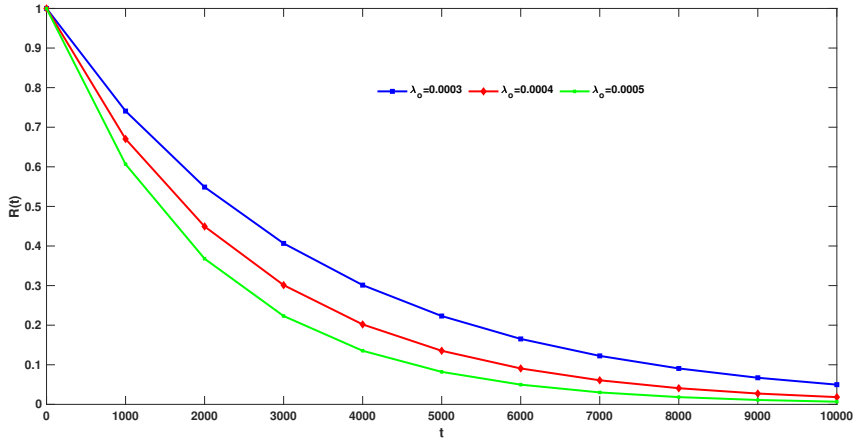


Figure 3: *Reliability w.r.t varied t and* λ_0

Table 1: *Value of Availability for varied* λ_0

	Availability		
λ_0	A_0^n	A_0^{et}	A_0^{ex}
0.001	0.9832	0.9829	0.9675
0.002	0.9670	0.9663	0.9370
0.003	0.9513	0.9503	0.9083
0.004	0.9361	0.9349	0.8814
0.005	0.9214	0.9199	0.8560
0.006	0.9071	0.9054	0.8321
0.007	0.8933	0.8913	0.8094
0.008	0.8799	0.8777	0.7880
0.009	0.8669	0.8645	0.7676
0.010	0.8542	0.8517	0.7483

10.1.2 Effect of Various Rates / Costs on Profit Functions

The outcomes of profit functions, P^m for varied (EC, MC), (CP, p) and P^u for varied (C_0, C_3^u), (λ_0, r_1) are studied. The other parameters are kept fixed, and their values are taken as in $eq^n(21)$. The results obtained are represented by Figure 4, 5, 6, 7 and Table 2 respectively and summarised as follows:

1. P^m goes down as MC and p increases but hike in its value is observed when EC and CP increases.
2. As the parameter C_0 increases, P^u increases. Moreover, the rise in the values of λ_0, r_1, C_3^u respectively, results in the decreasing P^u.

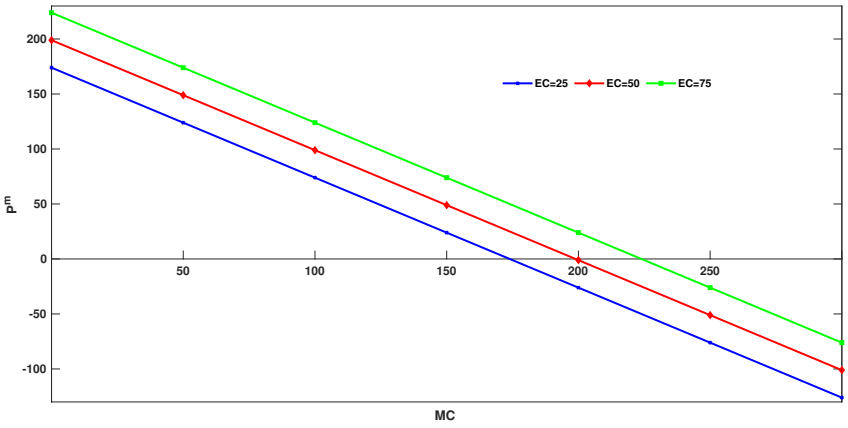


Figure 4: *Variation in P^m for varied MC and EC*

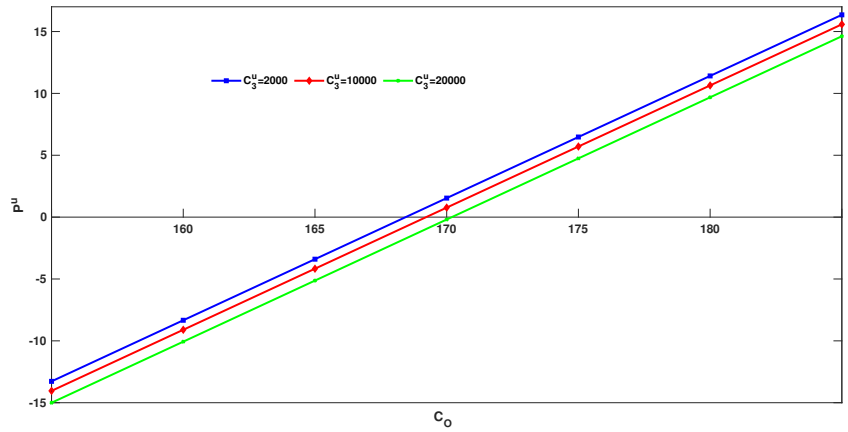


Figure 5: *Variation in P^u for varied C_0 and C_3^u*

3. The system should remain profitable for user as well as manufacturer.

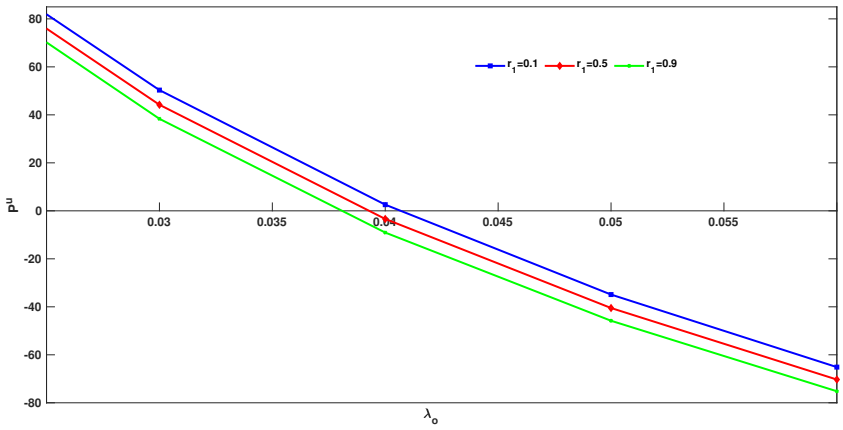


Figure 6: *Variation in P^u for varied λ_0 and r_1*

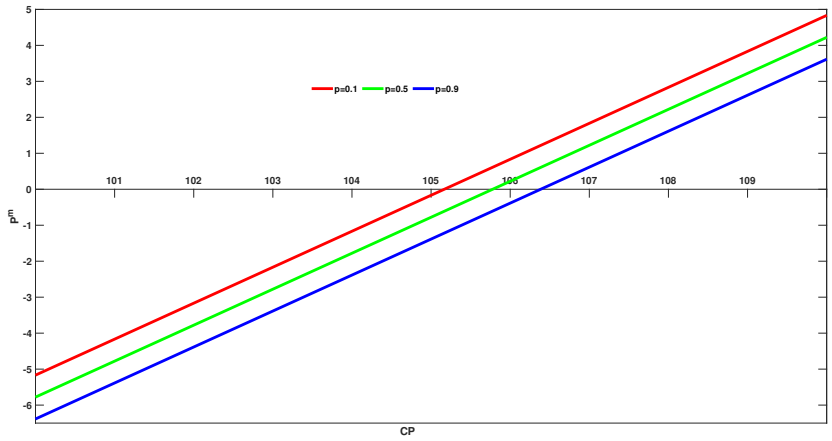


Figure 7: *Variation in P^m for varied CP and p*

Keeping that in mind, the bounds for some of the parameters are evaluated as:

(a) For EC=25, $P^m > 0$ iff MC < 175

(b) For C_3^u = 2000, $P^u > 0$ iff $C_0 > 168$

(c) For r_1=0.1, $P^u > 0$ iff $\lambda_0 < 0.038$

(d) For p=0.1, $P^m > 0$ iff $CP > 105.112$

Bounds for other values of EC, C_3^u and r_1 are also mentioned in Table 2.

Table 2: *Bounds for Revenue/Cost/Rate*

Cost/Revenue/Rate	Varied Parameter	Bounds For Profitability(P^u / P^m>0)
MC	EC=25	MC<175
	EC=50	MC<200
	EC=75	MC< 220
C_0	C_3^u=2000	C_0>168
	C_3^u=10000	C_0>169
	C_3^u=20000	C_0>170
λ_0	$r_1 = 0.1$	λ_0< 0.038
	$r_1 = 0.5$	λ_0< 0.039
	$r_1 = 0.9$	λ_0< 0.040
CP	$p = 0.1$	CP>105.112
	$p = 0.5$	CP>105.784
	$p = 0.9$	CP>106.396

10.2. Example 2: Failure Time follows Weibull Distribution

Assuming the failure time follows Weibull distribution with p.d.f.

$$f(t) = \frac{\delta}{\eta}.(\frac{t}{n})^{\delta-1}.e^{-(\frac{t}{n})^{\delta}} \tag{22}$$

where δ and η are shape and scale parameters respectively.
All the other time distributions follows exponential distribution with same p.d.f. and numerical value as taken in Section 10.1. Figure 8 reveals that the reliability decreases with time and for t<200, it is on the higher side for higher values of η and δ. However, the reverse trend of its values for η and δ is noticed for t>200.
Taking δ=0.3 and η=150, the behaviour of A_0^n, A_0^{et} w.r.t γ_m and A_0^{ex} w.r.t γ_u is shown in Table 3. The increasing trend of availabilities with an increase in γ_u and γ_m respectively are observed.However,the system availability (A_0^{ex}) during expired warranty period is lesser as compared to A_0^n, A_0^{et} for any particular value of γ_u(=γ_m).

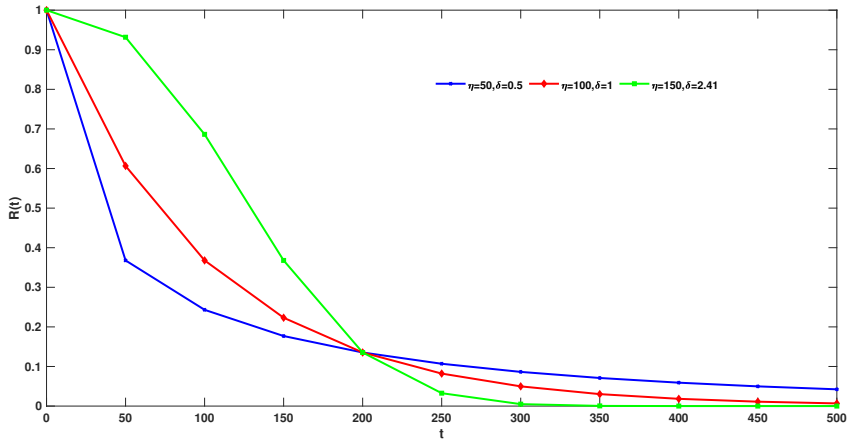


Figure 8: *Reliability w.r.t. varied t,η and δ*

Table 3: *Value of Availabilities for varied γ_m/γ_u*

	Availability		
γ_m/γ_u	A_0^n	A_0^{et}	A_0^{ex}
0.1	0.9813	0.9811	0.9701
0.2	0.9848	0.9846	0.9735
0.3	0.9860	0.9857	0.9746
0.4	0.9866	0.9863	0.9752
0.5	0.9869	0.9867	0.9756
0.6	0.9872	0.9869	0.9758
0.7	0.9873	0.9871	0.9760
0.8	0.9875	0.9872	0.9761
0.9	0.9876	0.9873	0.9762
1.0	0.9876	0.9874	0.9762

10.3. Example 3: Repair Time follows Erlang Distribution

Considering the repair time in different warranty time follows Erlang distribution with p.d.f.

$$\begin{cases} g_n(x) = \xi_n.\frac{x^{k_n-1}.e^{-\xi_n x}}{(k_n-1)!} \\ g_{et}(x) = \xi_{et}.\frac{x^{k_{et}-1}.e^{-\xi_{et} x}}{(k_{et}-1)!} \\ g_{ex}(x) = \xi_{ex}.\frac{x^{k_{ex}-1}.e^{-\xi_{ex} x}}{(k_{ex}-1)!} \end{cases} \tag{23}$$

where
$k_n/k_{et}/k_{ex}$ and $\xi_n/\xi_{et}/\xi_{ex}$ are shape and scale parameter during normal/extended/expired warranty period,
All of the other time distributions have the same p.d.f. ,i.e., exponential distribution and numerical values as in Section 10.1.

Table 4: *Value of Availabilities for varied λ_o*

	Availability		
λ_o	A_o^n	A_o^{et}	A_o^{ex}
0.0001	0.9907	0.9396	0.9383
0.0002	0.9817	0.8861	0.8837
0.0003	0.9727	0.8383	0.8351
0.0004	0.9640	0.7955	0.7916
0.0005	0.9554	0.7568	0.7524
0.0006	0.9469	0.7216	0.7169
0.0007	0.9386	0.6897	0.6846
0.0008	0.9304	0.6604	0.6551
0.0009	0.9224	0.6335	0.6280
0.001	0.9145	0.6087	0.6031

Taking ξ_n=ξ_{et}=ξ_{ex}=0.5 and k_n=k_{et}=k_{ex}=7, the behaviour of A_0^n, A_0^{et} and A_0^{ex} w.r.t λ_o is shown in Table 4. The decreasing trend of availabilities with an increase in λ_o is observed.

10.4. Numerical Calculations for Sensitivity Analysis

Considering all the p.d.f. involved as exponential as taken in eq^n (20) and assuming the values of parameters as mentioned in eq^n (21), the sensitivity analysis is performed. Using eq^n (18) and (19), the outcomes for the sensitivity and relative sensitivity functions of availability and profit

functions are summarized in Table 5, 6 and 7 respectively. The magnitude of these functions is taken into account while drawing inferences about parameters and the order in which they influence the different measures.

Table 5: *Sensitivity and relative sensitivity analysis of Availabilities*

Parameter	Sensitivity Function	Relative Sensitivity Function
(y)	$D_y = \frac{\partial(A_0)}{\partial y}$	$d_y = \frac{D_y y}{A_0}$
	Normal Warranty Period	
λ_0	-16.7791	-0.0085
γ_m	2.1848×10^{-4}	3.3052×10^{-4}
α_n	0.0014	7.0597×10^{-4}
β_n	0.3687	0.0074
p	-0.0081	-0.0057
q	-0.0081	-0.0025
r_1	-9.8315×10^{-4}	-6.9408×10^{-4}
r_2	-0.0246	-0.0073
	Extended Warranty Period	
λ_0	-16.1172	-0.0086
γ_m	2.1840×10^{-4}	3.3045×10^{-4}
α_{et}	0.0021	8.4732×10^{-4}
β_{et}	0.3686	0.0074
p	-0.0082	-0.0058
q	-0.0082	-0.0025
r_1	-0.0012	-8.4732×10^{-4}
r_2	-0.0246	-0.0074
	Expired Warranty Period	
λ_0	-32.54	-0.0166
α_{ex}	0.0053	0.0015
β_{ex}	1.4507	0.0147
γ_u	3.3582×10^{-4}	4.0978×10^{-4}
r_1	-0.0019	-0.0014
r_2	-0.0484	-0.0148

It has been determined that

1. Availabilities (A_0^n, A_0^{et}, A_0^{ex}) in three different periods are highly influenced by λ_0. Though these are least affected by variation in γ_u and γ_m respectively.
2. P^m and P^u both the profit functions are extremely sensitive to λ_0.
3. Variation in C_3^u and C_3^m results in a nominal change in P^m and P^u respectively.

Table 6: *Sensitivity and relative sensitivity analysis of User Profit*

Parameter	Sensitivity Function	Relative Sensitivity Function
(y)	$Z_y^u = \frac{\partial(P^u)}{\partial y}$	$z_y^u = \frac{Z_y^u y}{P^u}$
λ_0	-1.7455×10^4	-0.0268
α_{ex}	1.6730	0.0013
β_{ex}	448.1202	0.0137
γ_u	0.1003	3.6900×10^{-4}
γ_m	0.0545	2.5063×10^{-4}
α_n	0.1474	2.2597×10^{-4}
β_n	39.4737	0.0024
α_{et}	0.3453	4.2349×10^{-4}
β_{et}	59.1900	0.0036
C_0	0.9874	1.5137
C_1^u	$-2.0489 \times 10-4$	-6.2816×10^{-5}
C_2^u	-0.0093	-0.0034
C_3^u	-9.6069×10^{-5}	-0.0044
CP	-1	-0.4599
EC	-1	-0.0460
p	-2.0369	-0.0044
q	-3.6463	-0.0034
r_1	-0.9001	-0.0019
r_2	-26.3183	-0.0242
p_1	494.8074	0.3034
p_2	494.7154	0.4551
p_3	487.5415	0.7474

Moreover, the order or sequence in which different parameters influence the availabilities (A_0^n, A_0^{et}, A_0^{ex}) and profit functions (P^m,P^u) are

- Availability(A_0^n): $\lambda_0 > \beta_n > r_2 > p > q > \alpha_n > r_1 > \gamma_m$.
- Availability(A_0^{et}): $\lambda_0 > \beta_{et} > r_2 > \alpha_{et} > p > q > r_1 > \gamma_m$.
- Availability(A_0^{ex}): $\lambda_0 > \beta_{ex} > r_2 > \alpha_{ex} > r_1 > \gamma_u$.
- Profit Function(P^u):$C_0 > p_3 > CPp_2 > p_1 >> EC > \lambda_0 > r_2 > \beta_{ex} > C_3^u p >> \beta_{et} > C_2^u > q > \beta_n > r_1 > \alpha_{ex} > \alpha_{et} > \gamma_u > \gamma_m > \alpha_n > C_1^u$.
- Profit Function(P^m):$CP > MC > EC > \lambda_0 > p > r_2 > C_3^m > p_2 > p_1 > C_2^m > \beta_{et} > C_1^m > \beta_n > r_1 > \alpha_{et} > \alpha_n > \gamma_m > q$.

Table 7: *Sensitivity and relative sensitivity analysis of Manufacturer Profit*

Parameter	Sensitivity Function	Relative Sensitivity Function
(y)	$Z_y^m = \frac{\partial(P^m)}{\partial y}$	$z_y^m = \frac{Z_y^m y}{P^m}$
λ_0	-2.1458×10^3	-0.0244
γ_m	0.0086	2.9373×10^{-4}
α_n	0.0188	2.1404×10^{-4}
β_n	5.0452	0.0023
α_{et}	0.0441	4.0166×10^{-4}
β_{et}	7.5651	0.0034
C_1^m	$-1.6524 \times 10-4$	-3.0100×10^{-4}
C_2^m	-0.0029	-0.0066
C_3^m	-5.2050×10^{-5}	-0.0178
CP	1	3.4155
EC	1	0.3415
MC	-1	-2.7324
p	-1.5181	-0.0242
q	-0.0089	6.0795×10^{-5}
r_1	-0.0387	-6.1683×10^{-4}
r_2	-3.4432	-0.0235
p_1	-2.1573	-0.0098
p_2	-2.1690	-0.0148

11. Conclusion

A stochastic model of system functioning in normal, extended and expiry warranty conditions is developed in this paper. Markov and regenerative processes are employed to derive various reliability characteristics and profit functions for the manufacturer as well as the user of the system. The derived measures are further illustrated by discussing numerical examples for exponential, Weibull and Erlang cases. System is found available for a longer period in normal as compared to extended or expiry warranty periods. Upper/ lower bounds are obtained for involved rates/ costs, which can affect the system's profitability. Availabilities and profit functions are observed to be most sensitive to the failure rate. Further, for cost consideration, manufacturer and user profit functions are influenced most by the cost price of the system (CP) and revenue generated (C_o), respectively. Since the results for a described system are obtained using general probability time distribution, the finding of the study is lucrative from the standpoints of both the manufacturer and the user if they have real data on failures, repairs, costs, and so on.

Funding

The first author delightedly acknowledges the University Grants Commission (UGC), New Delhi, India for providing financial support.

Disclosure statement

The authors declare that they have no conflict of interest.

References

[1] Ives, B., and Vitale, M. (1988). After the sale: Leveraging maintenance with information technology. *MIS Quarterly*, 12(1): 7–21.

[2] Ritchken, P. H.,Chandramohan, J., and Tapiero, C. S. (1989). Servicing, quality design and control. *IIE Transactions*, 21(3): 213–220.

[3] Singpurwalla, N. D., and Wilson, S. (1993). The warranty problem: its statistical and game-theoretic aspects. *SIAM review*, 35(1): 17–42.

[4] Blischke, W. R., and Murthy, D. N. P. (1992). Product warranty management - I: A taxonomy for warranty policies. *European journal of operational research*, 62(2): 127–148.

[5] Jain, M., and Maheshwari, S. (2006). Discounted costs for repairable units under hybrid warranty. *Applied Mathematics and Computation*, 173(2): 887–901.

[6] Bai, J., and Pham, H. (2006). Cost analysis on renewable full-service warranties for multi-component systems. *European Journal of Operational Research*, 168(2): 492–508.

[7] Huang, H. Z., Liu, Z. J., and Murthy, D. (2007). Optimal reliability, warranty and price for new products. *IIE Transactions*, 39(8): 819–827.

[8] Kadyan, M., and Ramnivas (2013). Cost benefit analysis of a single-unit system with warranty for repair. *Applied Mathematics and Computation*, 223: 346–353.

[9] Rahman, A., and Chattopadhyay, G. (2015). Long term warranty and after sales service. *In Long Term Warranty and After Sales Service*, 1–15.

[10] Pham, H., and Bai, J. (2023). Warranty Policies: Analysis and Perspectives. *In Springer Handbook of Engineering Statistics*, 185–195.

[11] Taneja, G. (2010). Reliability Model for a System with Conditional Warranty and various types of Repair/Replacement. *Advances in Information Theory and Operations Research*, 214–220.

[12] Lei, Y., Liu, Q., and Shum, S. (2017). Warranty pricing with consumer learning. *European Journal of Operational Research*, 263(2): 596–610.

[13] Solkhe, M.,and Taneja, G. (2019). Reliability and Availability Analysis of a system with Periods Before and After Expiry of Conditional Warranty. *International Journal of Agricultural and Statistical Sciences* 15: 281–288.

[14] Niwas, R. (2023). Reliability and profit analysis of a markov model having cost-free warranty with waiting repair facility. *In Engineering Reliability and Risk Assessment*, 145–160.

[15] Hooti, F., Ahmadi, J., and Longobardi, M. (2020). Optimal extended warranty length with limited number of repairs in the warranty period. *Reliability Engineering & System Safety*, 203: 107111.

[16] Lorson, E., and Suiter, M. (2021). Do you want an extended warranty with that? *Page One Economics Newsletter.*

[17] Huang, Y. S., Fang, c. C., Lu, C. M., and Tseng, T. L. B. (2022). Optimal warranty policy for consumer electronics with dependent competing failure processes. *Reliability Engineering & System Safety*, 222: 108418.

[18] Salmasnia, A.,and Hatami, A. (2019). A three-level servicing contract with an integrated maintenance plan, warranty policy, technology level, and pricing model that considers the time value of money. *Scientia Iranica* 30(2): 674–690.

ANALYZING LOAD SHARING SYSTEM RELIABILITY: A MODIFIED WEIBULL DISTRIBUTION APPROACH

Santosh S. Sutar

Yashwantrao Chavan School of Rural Development, Shivaji University, Kolhapur - 416 004, India,
sutarssantosh@gmail.com

•

Chandrakant G. Gardi*

Department of Statistics, Punyashlok Ahilyadevi Holkar Solapur University,
Solapur, 413255, India,
chandrakant.gardi@gmail.com

•

Somnath D. Pawar

Department of Statistics, Shivaji University, Kolhapur - 416 004, India.
sdpawar23@gmail.com

* Corresponding author

Abstract

Load sharing systems have the ability to distribute the workload among its components. For analyzing a two component parallel load sharing system, the accelerated failure time (AFT) based model with component lifetimes as linear failure rate distribution have been recently proposed in the literature. In the present study, the component lifetimes are assumed to follow a modified Weibull distribution, which is the generalization of many standard lifetime distributions such as exponential, Weibull, Rayleigh, and linear failure rate. The use of modified Weibull distribution leads to a new family of bivariate distributions for ordered random variables. We have also looked into the associated inference techniques for the proposed model. In order to evaluate the effectiveness of the suggested estimating approaches, we conducted a simulation study. In order to provide a practical application and better understanding, we carefully examine a dataset related to motors.

Keywords: accelerated failure time model, conditional distribution, load sharing, modified Weibull distribution, order statistics

1. Introduction

Load sharing systems are characterized by their ability to distribute the workload among multiple components, such that if one component fails, the remaining components bear the additional workload. This can either increase or decrease the load on each surviving component. Load sharing systems have been extensively investigated in various engineering domains, such as software and hardware reliability, power plants, computing workload analysis, and fiber composites (Wang *et al.* [1]).

Liu [2] presents various instances that illustrate the concept of load sharing systems. These include scenarios like electric generators distributing an electrical load within a power plant, CPUs operating in a multiprocessor computer system, cables supporting a suspension bridge, bolts

fastening a wheel assembly onto a truck, and valves or pumps functioning in a hydraulic system. When any of these components fail, the remaining components must bear an additional load, which can elevate their failure rates. In an intriguing study by Drummond et al. [3] conducted on vertebrate species, it was observed that when a litter mate dies due to food shortage, the surviving offsprings receive a larger portion of the available food supply, leading to improved growth. This finding highlights how the failure of one individual can inadvertently benefit the surviving members. Furthermore, in the realm of software testing, the detection of a fault can uncover previously undetected critical faults. This demonstrates that a component's failure can facilitate the discovery of other hidden issues, thereby enhancing the overall reliability of the system. These examples collectively illustrate that when a component fails, it can actually enhance the remaining components' remaining lifespan, resulting in a higher growth rate for the surviving components.

Daniels [4] conducted the first study on the phenomenon of load sharing and load sharing systems. A thorough analysis of load sharing systems up till 2009 is present in Dewan and Naik-Nimbalkar [5]. Deshpande et al. [6], Park [7], Singh and Gupta [8], Park [9], Gurov and Utkin [10], Sutar and Naik-Nimbalkar [11]-[12], Krivtsov et al. [13], Wang et al. [1] and Sutar and Naik-Nimbalkar [14] have all published studies and modeled the load sharing phenomenon since then.

The study of load-sharing systems with a k-out-of-m configuration was suggested by Sutar and Naik-Nimbalkar [11] with modelling strategy based on the accelerated failure time (AFT) model. They concentrated on a particular configuration, a parallel load sharing system consisting of two components with baseline as the linear failure rate distribution. The associated inference techniques were also covered by the researchers. The distributions used there in for the ordered random variables are a subset of a larger family of distributions known as sequential order statistics. Kamps [15] first described this family of distributions and Cramer and Kamps [16] further developed them.

This study utilizes the load sharing model based on accelerated failure time (AFT), proposed by Sutar and Naik-Nimbalkar [11], to examine the load sharing phenomenon within a parallel system consisting of two components. We adopt a modified Weibull distribution (MWD) as the baseline distribution for the components in the system, which is characterized by three parameters: λ_1, λ_2, and λ_3. The introduction of this three-parameter MWD was done by Sarhan and Zaindin [17].

It is important to highlight that the three-parameter MWD provides a comprehensive representation of various distributions, including exponential, Weibull, Rayleigh and linear failure rate. Thus, the MWD serves as a versatile baseline distribution for the component lifetime in any load sharing system. The subsequent sections of this paper are organized as follows.

In Section 2, we address the AFT-based load-sharing model for a parallel load sharing system consisting of two components and with a modified Weibull distribution for component lifetimes. In Section 3, the inference procedures are thoroughly examined, while Section 4 focuses on the simulation study. Section 5 demonstrates an application using real data, and the last section presents the concluding remarks.

2. Proposed AFT based Load sharing model

We investigate a parallel system consisting of two components. The cumulative distribution function (c.d.f.) of the components follows a MWD characterized by three parameters: λ_1, λ_2, and λ_3. The probability density function (p.d.f.), survival function (s.f.), and hazard rate function of the MWD with parameters λ_1, λ_2, and λ_3 are provided as follows.

$$f(u) = \left(\lambda_1 + \lambda_2\lambda_3 u^{\lambda_3-1}\right) \exp\left\{-\left(\lambda_1 u + \lambda_2 u^{\lambda_3}\right)\right\},\ u > 0, \lambda_1, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_2 > 0 \quad (1)$$

$$\bar{F}(u) = \exp\left\{-\left(\lambda_1 u + \lambda_2 u^{\lambda_3}\right)\right\}, u > 0, \lambda_1, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_2 > 0$$
$$h(u) = \left(\lambda_1 + \lambda_2 \lambda_3 u^{\lambda_3 - 1}\right), u > 0, \lambda_1, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_2 > 0$$

The three-parameter modified Weibull distribution, denoted as MWD$(\lambda_1, \lambda_2, \lambda_3)$, serves as a generalization of the following distributions.

(a) It represents the Exponential distribution, ED(λ_1), when λ_2 is set to 0 and λ_3 is finite.

(b) It encompasses the Weibull distribution, WD(λ_2, λ_3), when λ_1 is set to 0.

(c) It corresponds to the Rayleigh distribution, RD(λ_2), when λ_3 is set to 2 and λ_1 is set to 0.

(d) It encompasses the Linear failure rate, LFR(λ_1, λ_2), when λ_3 is set to 2.

For more details on MWD$(\lambda_1, \lambda_2, \lambda_3)$, one can refer Sarhan and Zaindin [17] and references cited therein.

The load sharing behavior observed in a system comprising two components is captured by the AFT model, which was introduced by Sutar and Naik-Nimbalkar [11]. We denote the lifetimes of the two components in the system as V_1 and V_2. These lifetimes are considered independent and identically distributed random variables. The baseline densities of V_1 and V_2 are denoted as $f_1(\cdot)$ and $f_2(\cdot)$, respectively, while their corresponding baseline survival functions are denoted as $\bar{F}_1(\cdot)$ and $\bar{F}_2(\cdot)$. Let $X = \min(V_1, V_2)$ denote time of the first failure and $Y = \max(V_1, V_2)$ denote the time of the second failure or the system failure time. Consequently, the marginal density of the first failure can be expressed as follows.

$$g(x) = \left(2\lambda_1 + 2\lambda_2 \lambda_3 x^{\lambda_3 - 1}\right) \exp\left\{-\left(2\lambda_1 x + 2\lambda_2 x^{\lambda_3}\right)\right\}, \; x > 0, \lambda_1 > 0, \lambda_2, \lambda_3 \geq 0, \lambda_1 + \lambda_2 > 0. \tag{2}$$

It is worth mentioning that the distribution of the first failure is identical to the baseline distribution, which is a modified Weibull distribution with parameters $(2\lambda_1, 2\lambda_2, \lambda_3)$. Following the AFT load sharing model, the conditional density of variable Y given that $X = x$, as well as the joint density of the variables X and Y, can be expressed in the following manner.

$$g(y|x) = \left\{\frac{\lambda_1}{\beta} + \frac{\lambda_2 \lambda_3 y^{\lambda_3 - 1}}{\beta^{\lambda_3}}\right\} \exp\left\{-\frac{\lambda_1}{\beta}(y - x) - \frac{\lambda_2}{\beta^{\lambda_3}}(y^{\lambda_3} - x^{\lambda_3})\right\}, \tag{3}$$
$$0 < x < y < \infty, \; \beta > 0, \; \lambda_1 > 0, \lambda_2 \geq 0, \lambda_3 \geq 0, \lambda_1 + \lambda_2 > 0,$$

$$g(x, y) = 2\left(\lambda_1 + \lambda_2 \lambda_3 x^{\lambda_3 - 1}\right)\left\{\frac{\lambda_1}{\beta} + \frac{\lambda_2 \lambda_3 y^{\lambda_3 - 1}}{\beta^{\lambda_3}}\right\}$$
$$\times \exp\left\{-\frac{\lambda_1}{\beta}(y - x) - \frac{\lambda_2}{\beta^{\lambda_3}}(y^{\lambda_3} - x^{\lambda_3}) - \left(2\lambda_1 x + 2\lambda_2 x^{\lambda_3}\right)\right\}, \tag{4}$$
$$0 < x < y < \infty, \; \beta > 0, \; \lambda_1 > 0, \lambda_2 \geq 0, \lambda_3 \geq 0, \lambda_1 + \lambda_2 > 0,$$

It is important to note that when the parameter β is equal to 1, the joint density described in equation (4) simplifies to the joint density of independent random variables X and Y. Essentially, when $\beta = 1$, it indicates no load sharing effect, and hence the occurrences of the two failures (first and second) are independent of each other. The parameter β is referred to as the load sharing parameter in this context. In the following section, we will delve into the inference procedures associated with this concept.

3. Inference procedures

In this section, we examine different methods for estimating the unknown parameters and introduce a testing procedure to assess the presence of the load sharing effect.

3.1. Direct estimation procedure

The complete data, denoted as $(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = \{(x_i, y_i) : x_i \leq y_i;\ i = 1, 2, ..., n\}$, represents the set of observations. The log-likelihood function based on this complete data can be expressed as follows.

$$\log L = n \log 2 + \sum_{i=1}^{n} \log\left(\lambda_1 + \lambda_2\lambda_3 x_i^{\lambda_3 - 1}\right) + \sum_{i=1}^{n} \log\left(\lambda_1\beta^{\lambda_3} + \lambda_2\lambda_3\beta y_i^{\lambda_3 - 1}\right) - n(\lambda_3 + 1)\log\beta$$
$$- \frac{\lambda_1}{\beta}\sum_{i=1}^{n}(y_i - x_i) - \frac{\lambda_2}{\beta^{\lambda_3}}\sum_{i=1}^{n}(y_i^{\lambda_3} - x_i^{\lambda_3}) - 2\lambda_1\sum_{i=1}^{n} x_i - 2\lambda_2\sum_{i=1}^{n} x_i^{\lambda_3}.$$

We observe that, the log-likelihood equations $\frac{\partial \log L}{\partial \beta} = 0$, $\frac{\partial \log L}{\partial \lambda_1} = 0$, $\frac{\partial \log L}{\partial \lambda_2} = 0$ and $\frac{\partial \log L}{\partial \lambda_3} = 0$ do not have explicit solutions for the parameters λ_1, λ_2, λ_3, β. The mathematical expressions for the score functions, specifically $\frac{\partial \log L}{\partial \beta} = 0$, $\frac{\partial \log L}{\partial \lambda_1} = 0$, $\frac{\partial \log L}{\partial \lambda_2} = 0$, and $\frac{\partial \log L}{\partial \lambda_3} = 0$, can be found in Appendix (A).

In the following subsection, we outline a two-step approach for determining the values of the unknown parameters λ_1, λ_2, λ_3, and β.

3.2. Two-step parameter estimation procedure

The process of estimating the values of λ_1, λ_2, λ_3, and β has been conducted using a two-step methodology.

Step 1. We observe the first failure, X, and estimate baseline parameters, namely, λ_1, λ_2 and λ_3 by using the MCEM procedure proposed by Sutar [18].

Step 2. In order to estimate the load sharing parameter β, we utilize the conditional distribution of Y given $X = x$, as expressed in equation (3). The estimates of λ_1, λ_2, and λ_3 obtained in *Step 1* are then substituted into that equation to perform the estimation. We refer to this estimation process as a two-step estimation procedure, and the subsequent subsections outline these two steps in detail.

3.2.1 Estimation of λ_1, λ_2 and λ_3 (*Step 1*)

It is worth noting that the distribution of the first failure, X, is also a modified Weibull distribution (MWD) with parameters $2\lambda_1$, $2\lambda_2$, and λ_3. Let us denote $2\lambda_1 = \gamma_1$, $2\lambda_2 = \gamma_2$, thus distribution of X is MWD with parameters γ_1, γ_2 and λ_3. We use the MCEM algorithm proposed by Sutar [18] for finding the estimates of γ_1, γ_2 and λ_3. To implement the proposed MCEM algorithm, we take two independent random variables U_1 and U_2, which has exponential (γ_1) and Weibull (γ_2, λ_3) distributions, with their respective survival functions as $\exp(-\gamma_1 u_1)$ and $\exp(-\gamma_2 u_2^{\lambda_3})$. Let $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\hat{\lambda}_3$ be the MLEs of γ_1, γ_2 and λ_3 obtained through MCEM algorithm, then the MLEs of λ_1, λ_2 and λ_3 can be obtained as $\hat{\lambda}_1 = \frac{\hat{\gamma}_1}{2}$, $\hat{\lambda}_2 = \frac{\hat{\gamma}_2}{2}$ and $\hat{\lambda}_3$.

3.2.2 Estimation of β (*Step 2*)

To estimate the load sharing parameter β, we utilize the conditional distribution of Y given $X = x$ as described in equation (3). In this study, two methods are proposed for estimating β, which are discussed as follows.

Method I : It can be noted, the conditional distribution of Y given $X = x$ as truncated MWD with parameters $\frac{\lambda_1}{\beta} = \theta_1$(say), $\frac{\lambda_2}{\beta^{\lambda_3}} = \theta_2$ (say), $\lambda_3 = \theta_3$ (say) truncated at $X = x$. Furthermore, the conditional distribution mentioned is equivalent to the distribution of the minimum value between

two independent random variables, denoted as W_1 and W_2. Specifically, W_1 follows a truncated exponential distribution, which is truncated below x and has a parameter θ_1. On the other hand, W_2 follows a truncated Weibull distribution, also truncated below x, with parameters θ_2 and θ_3. The survival functions of W_1 and W_2 are $\exp\{-\theta_1(w_1 - x)\}$ and $\exp\{-\theta_2(w_2^{\theta_3} - x^{\theta_3})\}$, respectively. Let us consider the complete data for $i = 1, 2, ..., n$ as a set of $2n$ independent random variables denoted as (W_{1i}, W_{2i}). The random variables W_{1i} represent truncated exponential distributions, truncated below x_i, with a parameter θ_1, while W_{2i} represent truncated Weibull distributions, truncated at x_i, with parameters (θ_2, θ_3). Additionally, we define Z_{2i} as the minimum value between W_{1i} and W_{2i}. Consequently, Z_{2i} follows a truncated MWD (Minimum of Weibull and Exponential Distribution) distribution, characterized by a probability density function (p.d.f)

$$g(z_{2i}) = \left(\theta_1 + \theta_2\theta_3 z_{2i}^{\theta_3 - 1}\right) \exp\left\{-\theta_1(z_{2i} - x_i) - \theta_2(z_{2i}^{\theta_3} - x_i^{\theta_3})\right\},$$

$$0 < x_i < z_{2i} < \infty,\ \beta > 0, \theta_1 > 0, \theta_2 \geq 0, \theta_3 \geq 0, \theta_1 + \theta_2 > 0.$$

We can regard the observed values $\underline{y} \equiv (y_1, y_2, ..., y_n)$ as corresponding to the values of $\underline{Z}_2 \equiv (Z_{21}, Z_{22}, ..., Z_{2n})$.

The joint density of W_1 and W_2 given $\underline{X} = \underline{x}(\equiv (x_1, x_2, ..., x_n))$ can be written as

$$g(w_1, w_2|\underline{\mathbf{x}}) = \{\theta_1, \theta_2, \theta_3\}^n \prod_{i=1}^{n} w_{2i}^{\theta_3 - 1} \exp\left\{-\theta_1(w_{1i} - x_i) - \theta_2(w_{2i}^{\theta_3} - x_i^{\theta_3})\right\}, \tag{5}$$

The log-likelihood can be expressed in the following manner.

$$\log L = n \log(\theta_1\theta_2\theta_3) + \theta_3 \sum_{i=1}^{n} \log(u_{2i}) - \theta_1 \sum_{i=1}^{n} u_{1i} - \theta_2 \sum_{i=1}^{n} u_{2i}^{\theta_3}.$$

In order to perform the E step, it is necessary to calculate the conditional expectation of $E_c[\log L|\underline{Z}_2]$. This can be represented as follows.

$$E_c[\log L|\underline{Z}_2] = n \log(\theta_1^*\theta_2^*\theta_3^*) + \theta_3^* E_c\left[\sum_{i=1}^{n} \log(U_{2i})|\underline{Z}_2\right] - \theta_1^* E_c\left[\sum_{i=1}^{n} U_{1i}|\underline{Z}_2\right] - \theta_2^* E_c\left[\sum_{i=1}^{n} U_{2i}^{\theta_3^*}|\underline{Z}_2\right]. \tag{6}$$

Remark 1. For i equal to 1, 2, and 3, the variables θ_i and θ_i^* represent the values of θ_i at the current iteration and the next iteration of the MCEM (Monte Carlo Expectation-Maximization) algorithm. Specifically, if $\theta_i = \theta_i^{(p)}$ represents the estimated value of θ_i at the p-th iteration, and $\theta_i^* = \theta_i^{(p+1)}$ represents the estimated value of θ_i at the $(p+1)$-th iteration, then θ_i and θ_i^* respectively denote the values of θ_i at the p-th and $(p+1)$-th iterations of the MCEM algorithm.

The conditional density of W_{11} given $X = x$ and $Z_{21} = z_{21}$ is a mixed probability density function (p.d.f.) and can be expressed as follows.

$$g(w_1|x, z_{21}) = \frac{\theta_1}{\left(\theta_1 + \theta_2\theta_3 z_{21}^{\theta_3 - 1}\right)} \left\{I_{(w_1 = z_{21})} + \theta_2\theta_3 z_{21}^{\theta_3 - 1} \exp\{-\theta_1(w_1 - z_{21})\} I_{(w_1 > z_{21})}\right\}.$$

where, $I_A(\cdot)$ is indicator function defined on set A. The details regarding the same are Appendix (B). Thus, the conditional expectation of W_{11} given X and Z_{21} can be obtained as

$$E(W_1|X, Z_{21}) = \int w_1 g(w_1|x, z_{21}) dw_1$$

$$= \frac{\theta_1}{\left(\theta_1 + \theta_2\theta_3 z_{21}^{\theta_3 - 1}\right)} \left\{Z_{21} + \theta_2\theta_3 Z_{21}^{\theta_3 - 1} \exp\{\theta_1 Z_{21}\} \int_{Z_{21}}^{\infty} w_1 \exp\{-\theta_1 w_1\} dw_1\right\}.$$

By using the result,

$$\int_{Z_{21}}^{\infty} w_1 \exp\{-\theta_1 w_1\} dw_1 = \frac{(\theta_1 Z_{21} + 1)\exp\{-\theta_1 Z_{21}\}}{\theta_1^2},$$

we get

$$E(W_1|X, Z_{21}) = \frac{\theta_1}{\left(\theta_1 + \theta_2\theta_3 z_{21}^{\theta_3-1}\right)} \left\{ Z_{21} + \theta_2\theta_3 Z_{21}^{\theta_3-1} \frac{\exp\{\theta_1 Z_{21}\}(\theta_1 Z_{21}+1)\exp\{-\theta_1 Z_{21}\}}{\theta_1^2} \right\}$$
$$= \frac{1}{\theta_1} + \left\{ Z_{21} - \left(\theta_1 + \theta_2\theta_3 Z_{21}^{\theta_3-1}\right)^{-1} \right\} = K(Z_{21}) \text{(say)},$$

and hence

$$E\left(\sum_{i=1}^{n} W_{1i}|\underline{X}, \underline{Z}_2\right) = \frac{n}{\theta_1} + \sum_{i=1}^{n} Z_{2i} - \sum_{i=1}^{n} \frac{1}{\left(\theta_1 + \theta_2\theta_3 Z_{21}^{\theta_3-1}\right)}.$$

Thus, given $\{Z_{2i} = y_i, i = 1, 2, ..., n\}$ and $\{X_i = x_i, i = 1, 2, ..., n\}$, we get

$$E\left(\sum_{i=1}^{n} W_{1i}|\underline{x}, \underline{y}\right) = \frac{n}{\theta_1} + \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{1}{\left(\theta_1 + \theta_2\theta_3 y_i^{\theta_3-1}\right)}.$$

Likewise, the conditional density function of W_{21} given $X = x$, $Z_{21} = z_{21}$, and the corresponding conditional expectation can be expressed as follows.

$$g(w_2|x, z_{21}) = \frac{\theta_2\theta_3}{\left(\theta_1 + \theta_2\theta_3 z_{21}^{\theta_3-1}\right)} \left\{ z_{21}^{\theta_3-1} I_{(w_2=z_{21})} + \theta_1 w_2^{\theta_3-1} e^{\{-\theta_2(W_2^{\theta_3} - z_{21}^{\theta_3})\}} I_{(w_2>z_{21})} \right\}$$

and

$$E_c\left\{W_{21}^{\theta_3^*}|\underline{X}, Z_{21}\right\} = \frac{\theta_2\theta_3 Z_{21}^{\theta_3^*+\theta_3+1}}{\left(\theta_1 + \theta_2\theta_3 Z_{21}^{\theta_3-1}\right)} + \frac{\theta_1\theta_2\theta_3 e^{\theta_2 Z_{21}^{\theta_3}}}{\left(\theta_1 + \theta_2\theta_3 Z_{21}^{\theta_3-1}\right)} \int_{Z_1}^{\infty} W_2^{\theta_3^*+\theta_3-1} e^{-\theta_2 W_2^{\theta_3}} dW_2.$$

After simplification, we get,

$$E_c\left\{W_{21}^{\theta_3^*}|\underline{X}, Z_{21}\right\} = \frac{\theta_2\theta_3 Z_{21}^{\theta_3^*+\theta_3+1}}{\left(\theta_1 + \theta_2\theta_3 Z_{21}^{\theta_3-1}\right)} + \frac{\theta_1\theta_2^{-\frac{\theta_3^*}{\theta_3}}}{\left(\theta_1 + \theta_2\theta_3 Z_{21}^{\theta_3-1}\right)} \int_0^{\infty} \left(u + \theta_2 Z_{21}^{\theta_3}\right)^{\frac{\theta_3^*}{\theta_3}} e^{-u} du$$

Let

$$T = \int_0^{\infty} \left(u + \theta_2 Z_{21}^{\theta_3}\right)^{\frac{\theta_3^*}{\theta_3}} e^{-u} du = E\left[V + \theta_2 Z_{21}^{\theta_3}\right]^{\frac{\theta_3^*}{\theta_3}},$$

where V has an exponential distribution with mean 1. By employing the Monte Carlo technique, we can replace T with a Monte Carlo sum, which is given as follows.

$$T = E\left[V + \theta_2 Z_{21}^{\theta_3}\right]^{\frac{\theta_3^*}{\theta_3}} = \frac{1}{m}\sum_{j=1}^{m} \left(v_j + \theta_2 Z_{21}^{\theta_3}\right)^{\frac{\theta_3^*}{\theta_3}},$$

where, $\{v_1, v_2, ..., v_m\}$ is random sample of size m (sufficiently large) from exponential distribution with mean 1. Thus, we get

$$E_c\left\{\sum_{i=1}^{n} W_{2i}^{\theta_3^*}|\underline{X}, \underline{Z}_2\right\} = \sum_{i=1}^{n} \left\{ \frac{\theta_2\theta_3 Z_{2i}^{\theta_3^*+\theta_3-1}}{\left(\theta_1 + \theta_2\theta_3 Z_{2i}^{\theta_3-1}\right)} + \frac{\theta_1\theta_2^{-\frac{\theta_3^*}{\theta_3}}}{m\left(\theta_1 + \theta_2\theta_3 Z_{2i}^{\theta_3-1}\right)} \sum_{j=1}^{m} \left(v_j + \theta_2 Z_{2i}^{\theta_3}\right)^{\frac{\theta_3^*}{\theta_3}} \right\}$$

and hence $\{Z_{2i} = y_i, i = 1, 2, ..., n\}$, that is $\underline{Z}_2 = \underline{y}$, we get

$$E_c\left\{\sum_{i=1}^{n} W_{2i}^{\theta_3^*}|\underline{x},\underline{y}\right\} = \sum_{i=1}^{n}\left\{\frac{\theta_2\theta_3 y_i^{\theta_3^*+\theta_3-1}}{\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)} + \frac{\theta_1\theta_2^{-\frac{\theta_3^*}{\theta_3}}}{m\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}\sum_{j=1}^{m}\left(v_j+\theta_2 y_i^{\theta_3}\right)^{\frac{\theta_3^*}{\theta_3}}\right\}.$$

By applying similar arguments to calculate $E_c\left\{\sum_{i=1}^{n}\log\left(W_{2i}\right)|\underline{x},\underline{y}\right\}$, we obtain

$$E_c\left\{\sum_{i=1}^{n}\log\left(W_{2i}\right)|\underline{x},\underline{y}\right\} = \sum_{i=1}^{n}\left\{\frac{\theta_2\theta_3\log y_i}{\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}\right\} + \sum_{i=1}^{n}\left\{\frac{\theta_1}{m\theta_2\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}\left(\sum_{j=1}^{m}\left(v_j+\theta_2 y_i^{\theta_3}\right)-\log\theta_2\right)\right\}.$$

To carry out the M-step, we obtain the following expression from equation (6).

$$\frac{\partial E_c\left[\log L|\underline{X},\underline{Z}_2\right]}{\partial\theta_1^*} = \frac{n}{\theta_1^*} - E_c\left[\sum_{i=1}^{n} W_{1i}|\underline{X},\underline{Z}_2\right], \tag{7}$$

$$\frac{\partial E_c\left[\log L|\underline{X},\underline{Z}_2\right]}{\partial\theta_2^*} = \frac{n}{\theta_2^*} - E_c\left[\sum_{i=1}^{n} W_{2i}^{\theta_3^*}|\underline{X},\underline{Z}_2\right], \tag{8}$$

$$\frac{\partial E_c\left[\log L|\underline{X},\underline{Z}_2\right]}{\partial\theta_3^*} = \frac{n}{\theta_3^*} - E_c\left[\sum_{i=1}^{n}\log(W_{2i})|\underline{X},\underline{Z}_2\right] - \theta_2^*\frac{\partial E_c\left[\sum_{i=1}^{n} W_{2i}^{\theta_3^*}|\underline{X},\underline{Z}_2\right]}{\partial\theta_3^*}. \tag{9}$$

From (7) and (8), we get

$$\frac{\partial E_c\left[\log L|\underline{X},\underline{Z}_2\right]}{\partial\theta_1^*} = 0 \Rightarrow \theta_1^* = \frac{n}{E_c\left[\sum_{i=1}^{n} W_{1i}|\underline{X},\underline{Z}_2\right]}, \tag{10}$$

$$\frac{\partial E_c\left[\log L|\underline{X},\underline{Z}_2\right]}{\partial\theta_2^*} = 0 \Rightarrow \theta_2^* = \frac{n}{E_c\left[\sum_{i=1}^{n} W_{2i}^{\theta_3^*}|\underline{X},\underline{Z}_2\right]}. \tag{11}$$

Based on the expression (5), it can be inferred that $t(W_1) = \sum_{i=1}^{n} W_{1i}$ serves as the sufficient statistic for θ_1. In order to carry out the M-step, we equate the sufficient statistic to its expectation, which takes the following form in our scenario.

$$E[W_1] = \frac{1}{\theta_1}. \tag{12}$$

The EM iterations alternate between expressions (12) and (10). Let $\theta_1^{(p)}$ represent the estimate of θ_1 at the p-th iteration step of the MCEM algorithm. The updated estimate $\theta_1^{(p+1)}$ is determined by the following equation.

$$\theta_1^{(p+1)} = \left\{\frac{1}{\theta_1^{(p)}} + \frac{1}{n}\left[\sum_{i=1}^{n}(y_i - x_i) - \sum_{i=1}^{n}\frac{1}{\left(\theta_1^{(p)}+\theta_2^{(p)}\theta_3^{(p)} y_i^{\theta_3^{(p)}-1}\right)}\right]\right\}^{-1}. \tag{13}$$

To estimate θ_3, we substitute equation (11) into equation (9), resulting in a nonlinear equation in terms of θ_3. This equation can be expressed as follows.

$$\frac{\partial E_c\left[\log L|\underline{x},\underline{y}\right]}{\partial\theta_3^*} = \frac{n}{\theta_3^*} - E_c\left[\sum_{i=1}^{n}\left(\log(W_{2i})\right)|\underline{x},\underline{y}\right] - n\frac{\left\{\frac{\partial E_c\left[\log W_{2i}^{\theta_3^*}|\underline{x},\underline{y}\right]}{\partial\theta_3^*}\right\}}{E_c\left[\sum_{i=1}^{n} W_{2i}^{\theta_3^*}|\underline{x},\underline{y}\right]}.$$

We employ the Newton-Raphson iterative method to estimate θ_3^*. Let $\theta_3^{(r)}$ denote the estimate of θ_3 at the r-th iteration of the Newton-Raphson method. The updated estimate $\theta_3^{(r+1)}$ is determined by the following equation.

$$\theta_3^{(r+1)} = \theta_3^{(r)} - \frac{\left\{\frac{\partial E_c\left[\log L|\underline{x},\underline{y}\right]}{\partial \theta_3^{(r)}}\right\}}{\left\{\frac{\partial^2 E_c\left[\log L|\underline{x},\underline{y}\right]}{\partial (\theta_3^{(r)})^2}\right\}}. \tag{14}$$

The detailed expressions used in equation (14) can be found in Appendix (C). The process should be iterated until the convergence criterion is satisfied. The value of θ_3 at the $(p+1)$-th iteration of MCEM, denoted as $\theta_3^* = \theta_3^{(p+1)}$, represents the stabilized value of θ_3 obtained through the Newton-Raphson method. Once we get the estimate of θ_3, we can obtain θ_2 at $(p+1)^{th}$ iteration by substituting (14) in (11). That is $\theta_2^{(p+1)}$ is given by

$$\theta_2^{(p+1)} = \frac{n}{E_c\left[\sum_{i=1}^n W_{2i}^{\theta_3^{(p+1)}}|\underline{x},\underline{y}\right]}. \tag{15}$$

Thus, we get the estimates $\widehat{\theta}_1 = \widehat{(\lambda_1/\beta)}$, $\widehat{\theta}_2 = \widehat{(\lambda_2/\beta^{\lambda_3})}$ of (λ_1/β) and $\left(\lambda_2/\beta^{\lambda_3}\right)$ respectively. Now, we have the two estimates of β as $\widehat{\beta}_1 = (\widehat{\lambda}_1/\widehat{\theta}_1)$ and $\widehat{\beta}_2 = (\widehat{\lambda}_2/\widehat{\theta}_2)^{1/\widehat{\lambda}_3}$.Here, $\widehat{\lambda}_1$, $\widehat{\lambda}_2$, and $\widehat{\lambda}_3$ represent the estimates of λ_1, λ_2, and λ_3, respectively, obtained in Step 1. The estimate of β is obtained by taking the average of $\widehat{\beta}_1$ and $\widehat{\beta}_2$. This estimation method is referred to as the 'average method' for estimating β.

Method II : The likelihood, based on the conditional density of Y given $X = x$, can be expressed as follows.

$$L = \prod_{i=1}^n \left\{\frac{\lambda_1}{\beta} + \frac{\lambda_2\lambda_3 y_i^{\lambda_3-1}}{\beta^{\lambda_3}}\right\} \exp\left\{-\frac{\lambda_1}{\beta}\sum_{i=1}^n (y_i - x_i) - \frac{\lambda_2}{\beta^{\lambda_3}}\sum_{i=1}^n (y_i^{\lambda_3} - x_i^{\lambda_3})\right\}.$$

Then the log-likelihood can be written as

$$\log L = \sum_{i=1}^n \log\left\{\frac{\lambda_1}{\beta} + \frac{\lambda_2\lambda_3 y_i^{\lambda_3-1}}{\beta^{\lambda_3}}\right\} - \frac{\lambda_1}{\beta}\sum_{i=1}^n (y_i - x_i) - \frac{\lambda_2}{\beta^{\lambda_3}}\sum_{i=1}^n (y_i^{\lambda_3} - x_i^{\lambda_3}).$$

Then we have,

$$\frac{\partial}{\partial\beta}\log L = -\sum_{i=1}^n \frac{\left\{\frac{\lambda_1}{\beta^2} + \frac{\lambda_2\lambda_3^2 y_i^{\lambda_3-1}}{\beta^{\lambda_3+1}}\right\}}{\left\{\frac{\lambda_1}{\beta} + \frac{\lambda_2\lambda_3 y_i^{\lambda_3-1}}{\beta^{\lambda_3}}\right\}} + \frac{\lambda_1}{\beta^2}\sum_{i=1}^n (y_i - x_i) - \frac{\lambda_2\lambda_3}{\beta^{\lambda_3+1}}\sum_{i=1}^n (y_i^{\lambda_3} - x_i^{\lambda_3}),$$

$$\frac{\partial^2}{\partial\beta^2}\log L = -\sum_{i=1}^n \frac{\left\{\frac{\lambda_1}{\beta^2} + \frac{\lambda_2\lambda_3^2 y_i^{\lambda_3-1}}{\beta^{\lambda_3+1}}\right\}\left\{\frac{2\lambda_1}{\beta^3} + \frac{\lambda_2\lambda_3^2(\lambda_3+1) y_i^{\lambda_3-1}}{\beta^{\lambda_3+2}}\right\} + \left\{\frac{\lambda_1}{\beta^2} + \frac{\lambda_2\lambda_3^2 y_i^{\lambda_3-1}}{\beta^{\lambda_3+1}}\right\}^2}{\left\{\frac{\lambda_1}{\beta} + \frac{\lambda_2\lambda_3 y_i^{\lambda_3-1}}{\beta^{\lambda_3}}\right\}^2}$$

$$-\frac{2\lambda_1}{\beta^3}\sum_{i=1}^n (y_i - x_i) - \frac{\lambda_2\lambda_3(\lambda_3+1)}{\beta^{\lambda_3+2}}\sum_{i=1}^n (y_i^{\lambda_3} - x_i^{\lambda_3}).$$

To estimate the load sharing parameter β, we substitute the maximum likelihood estimates (MLEs) $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\lambda}_3$ of λ_1, λ_2, and λ_3, respectively, into the aforementioned expressions. The Newton-Raphson iteration method is then employed to obtain the estimate of β. Let $\beta^{(m)}$ be the estimate of β at m^{th} iteration. The estimate of β at $(m+1)^{th}$ iteration is given by

$$\beta^{(m+1)} = \beta^{(m)} - \frac{\left(\frac{\partial}{\partial\beta}\log L\right)|_{(\hat{\lambda}_1,\hat{\lambda}_2,\hat{\lambda}_3)}}{\left(\frac{\partial^2}{\partial\beta^2}\log L\right)|_{(\hat{\lambda}_1,\hat{\lambda}_2,\hat{\lambda}_3)}}.$$

We termed this procedure to estimate β as the 'iteration method'. The subsequent section presents the test procedure used to assess the presence of the load sharing effect.

3.3. Test Procedure

To test the load sharing effect, we set up the null hypothesis H_0 stating that the failure of a component does not affect the survival components, and the alternative hypothesis H_1 stating that there exists a load sharing phenomenon. Specifically, we express the null hypothesis as $H_0 : \beta = 1$, indicating no load sharing effect, and the alternative hypothesis as $H_1 : \beta \neq 1$, indicating the presence of a load sharing effect. To test these hypotheses, we employ a score type test, as used by Sutar and Naik-Nimbalkar [11]. The test statistic follows an asymptotic χ^2 distribution with 1 degree of freedom.

In the subsequent section, we present a simulation study to evaluate and compare the performance of two estimation methods: the average method and the iterative method. The simulation study aims to assess the accuracy and efficiency of these methods under various scenarios and conditions. By conducting simulations and analyzing the results, we can gain insights into the strengths and limitations of each method and make informed decisions about their suitability for practical applications.

4. Simulation study

In this section, we performed a simulation study to assess the performance of the proposed estimation procedure in estimating unknown parameters. We generated a total of 10,000 samples from the joint density described in equation (4) for different combinations of sample sizes (n) and parameter values. This allowed us to examine the behavior and accuracy of the estimation procedure under various scenarios and conditions.

For sample sizes of $n = 20$, 30, 50, and 100, we considered different parameter combinations, namely $(\lambda_1, \lambda_2, \lambda_3, \beta)$ as (1,2,0.5,0.5), (1,2,0.5,1), (1,2,0.5,1.5), (1,2,1,0.5), (1,2,1,1), (1,2,1,1.5), (1,2,2,0.5), (1,2,2,1), (1,2,2,1.5), (2,2,0.5,0.5), (2,2,0.5,1), (2,2,0.5,1.5), (2,2,1,0.5), (2,2,1,1), (2,2,1,1.5), (2,2,2,0.5), (2,2,2,1), and (2,2,2,1.5).

The average estimates of the unknown parameters $(\lambda_1, \lambda_2, \lambda_3, \beta)$ obtained through Method-I (Two-step Procedure), denoted as $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\beta})$, along with their corresponding standard errors (SE), i.e., SE$(\hat{\lambda}_1)$, SE$(\hat{\lambda}_2)$, SE$(\hat{\lambda}_3)$, and SE$(\hat{\beta})$, are presented in Table 1 and Table 2. The simulation results reveal a clear pattern: as the sample size grows larger, the standard errors exhibit a decreasing trend. This indicates that larger sample sizes lead to enhanced precision in estimating the parameters, implying that the estimates become more accurate and reliable.

We also conducted a simulation study for the iterative method. We generated 10,000 samples with sizes $n = 30, 50$, and 100 from the joint density given in equation (4). We considered different parameter combinations as (1,2,0.5,0.5), (1,2,0.5,1), (1,2,0.5,1.5), (1,2,1,0.5), (1,2,1,1), (1,2,1,1.5), (1,2,2,0.5), (1,2,2,1), (1,2,2,1.5).

The estimates of the unknown parameters $(\lambda_1, \lambda_2, \lambda_3, \beta)$, where the estimate of β obtained through the Method-II (iterative method), along with their corresponding standard errors (SE$(\hat{\lambda}_1)$, SE$(\hat{\lambda}_2)$, SE$(\hat{\lambda}_3)$, and SE$(\hat{\beta})$), are reported in Table 3. We observed that compared to the estimates obtained by the average method, the estimates obtained by the iterative method had higher standard errors and tended to be overestimated.

Same phenomenon is observed for the simulation for study corresponding to the parameter combination $(\lambda_1, \lambda_2, \lambda_3, \beta)$ as (2,2,0.5,0.5), (2,2,0.5,1), (2,2,0.5,1.5), (2,2,1,0.5), (2,2,1,1), (2,2,1,1.5), (2,2,2,0.5), (2,2,2,1) and (2,2,2,1.5). Hence, we decided not to report the simulation results corresponding to these parameter combination.

Table 1: *The average estimates of $(\lambda_1, \lambda_2, \lambda_3, \beta)$ obtained through the two-step procedure.*

n	λ_1	λ_2	λ_3	β	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\beta}$	SE($\widehat{\lambda}_1$)	SE($\widehat{\lambda}_2$)	SE($\widehat{\lambda}_1$3)	SE($\widehat{\beta}$)
30	1	2	0.5	0.5	1.6347	3.2042	1.4232	0.7157	0.0798	0.0854	0.0693	0.0516
50	1	2	0.5	0.5	1.5741	2.5862	1.0873	0.6217	0.0759	0.0871	0.0669	0.0463
100	1	2	0.5	0.5	1.2432	2.3124	0.7554	0.5482	0.0748	0.0854	0.0675	0.0454
30	1	2	0.5	1	1.6865	3.256	1.8661	1.4245	0.0981	0.0847	0.0775	0.0611
50	1	2	0.5	1	1.4885	2.5789	1.0832	1.3029	0.0868	0.1001	0.0798	0.0579
100	1	2	0.5	1	1.1941	2.2879	0.7286	1.1229	0.0782	0.0861	0.0692	0.0461
30	1	2	0.5	1.5	1.8624	3.2677	1.9431	1.8289	0.1031	0.0962	0.0885	0.0721
50	1	2	0.5	1.5	1.5765	2.4989	0.9867	1.7110	0.0887	0.1042	0.0875	0.0598
100	1	2	0.5	1.5	1.2299	2.3093	0.7589	1.5603	0.0798	0.0954	0.0781	0.0476
30	1	2	1	0.5	1.8921	3.2776	2.4102	0.7372	0.0986	0.0865	0.0703	0.0627
50	1	2	1	0.5	1.5132	2.5867	1.5305	0.6134	0.0764	0.0967	0.0686	0.0511
100	1	2	1	0.5	1.2389	2.3123	1.2397	0.5623	0.0831	0.0876	0.0779	0.0467
30	1	2	1	1	1.7868	3.2682	2.3405	1.4321	0.1031	0.0881	0.0832	0.0684
50	1	2	1	1	1.5105	2.6105	1.5193	1.3139	0.0872	0.0989	0.0794	0.0673
100	1	2	1	1	1.1962	2.2961	1.2204	1.1283	0.0864	0.0872	0.0689	0.0551
30	1	2	1	1.5	1.8611	3.2692	2.3952	1.8382	0.1084	0.1051	0.0967	0.0685
50	1	2	1	1.5	1.5902	2.5156	1.4734	1.7102	0.0972	0.1098	0.0935	0.0637
100	1	2	1	1.5	1.2346	2.3203	1.2087	1.5589	0.0876	0.0971	0.0798	0.0472
30	1	2	2	0.5	1.9614	3.2658	3.4231	0.7267	0.0889	0.0847	0.0704	0.0542
50	1	2	2	0.5	1.5837	2.5991	2.5237	0.6079	0.0769	0.0885	0.0679	0.0476
100	1	2	2	0.5	1.2437	2.2984	2.2389	0.5472	0.0768	0.0853	0.0672	0.0462
30	1	2	2	1	1.8773	3.2674	3.4212	1.4298	0.0974	0.0869	0.0773	0.0616
50	1	2	2	1	1.4932	2.5813	2.5326	1.2998	0.0867	0.0983	0.0795	0.0568
100	1	2	2	1	1.2167	2.2916	2.2193	1.1183	0.0792	0.0851	0.0693	0.0463
30	1	2	2	1.5	1.8823	3.2593	3.4261	1.8672	0.1006	0.0975	0.0869	0.0578
50	1	2	2	1.5	1.5824	2.5139	2.4934	1.6672	0.0891	0.1092	0.0879	0.0573
100	1	2	2	1.5	1.2305	2.3027	2.1979	1.5723	0.07967	0.0945	0.0797	0.0493

Table 2: *The average estimates of $(\lambda_1, \lambda_2, \lambda_3, \beta)$ obtained through the two-step procedure.*

n	λ_1	λ_2	λ_3	β	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\beta}$	SE($\widehat{\lambda}_1$)	SE($\widehat{\lambda}_2$)	SE($\widehat{\lambda}_1$3)	SE($\widehat{\beta}$)
30	2	2	0.5	0.5	2.9587	3.2681	1.9672	0.7361	0.0893	0.0842	0.0669	0.0578
50	2	2	0.5	0.5	2.5831	2.5951	1.0851	0.6138	0.0765	0.0879	0.0677	0.0456
100	2	2	0.5	0.5	2.2360	2.3013	0.7542	0.5479	0.0754	0.0851	0.0679	0.0442
30	2	2	0.5	1	2.8476	3.2821	1.8567	1.4310	0.0981	0.0843	0.0774	0.0621
50	2	2	0.5	1	2.4881	2.5813	1.0829	1.3021	0.0870	0.0995	0.0792	0.0582
100	2	2	0.5	1	2.1933	2.2929	0.7300	1.1232	0.0798	0.0856	0.0686	0.0450
30	2	2	0.5	1.5	2.8621	3.2713	1.9413	1.8348	0.1116	0.0961	0.0873	0.0635
50	2	2	0.5	1.5	2.5855	2.5018	0.9964	1.7027	0.0877	0.1030	0.0862	0.0604
100	2	2	0.5	1.5	2.2320	2.3058	0.7542	1.5590	0.0802	0.0934	0.0770	0.0482
30	2	2	1	0.5	2.9745	3.2799	2.3941	0.7416	0.0971	0.0845	0.0689	0.0618
50	2	2	1	0.5	2.5941	2.5979	1.5238	0.6156	0.0771	0.0962	0.0690	0.0504
100	2	2	1	0.5	2.2468	2.3015	1.2416	0.5601	0.0815	0.0860	0.0758	0.0485
30	2	2	1	1	2.8891	3.2801	2.3407	1.4392	0.1028	0.0871	0.0817	0.0671
50	2	2	1	1	2.4932	2.5918	1.5262	1.3124	0.0881	0.0998	0.0811	0.0656
100	2	2	1	1	2.2003	2.3056	1.2185	1.1273	0.0841	0.0887	0.0694	0.0529
30	2	2	1	1.5	2.8601	3.2713	2.4006	1.8425	0.1061	0.1027	0.0946	0.0684
50	2	2	1	1.5	2.5888	2.5139	1.4866	1.7073	0.0926	0.1115	0.0915	0.0630
100	2	2	1	1.5	2.2387	2.3117	1.2032	1.5622	0.0857	0.0965	0.0831	0.0490
30	2	2	2	0.5	2.9601	3.2589	3.4098	0.7244	0.0862	0.0818	0.0671	0.0512
50	2	2	2	0.5	2.5785	2.5853	2.5164	0.6021	0.0741	0.0861	0.0664	0.0447
100	2	2	2	0.5	2.2351	2.2937	2.2336	0.5423	0.0727	0.0815	0.0639	0.0413
30	2	2	2	1	2.8744	3.2701	3.4189	1.4251	0.0948	0.0841	0.0751	0.0591
50	2	2	2	1	2.4821	2.5784	2.5261	1.2982	0.0836	0.0971	0.0783	0.0555
100	2	2	2	1	2.1927	2.2891	2.2164	1.1153	0.0774	0.0816	0.0683	0.0436
30	2	2	2	1.5	2.8513	3.2587	3.4228	1.8294	0.0987	0.0939	0.0849	0.0611
50	2	2	2	1.5	2.5751	2.4992	2.4839	1.6982	0.0838	0.1018	0.0839	0.0572
100	2	2	2	1.5	2.2194	2.2943	2.1926	1.5468	0.0782	0.0896	0.0744	0.0451

Table 3: *The average estimates of $(\lambda_1, \lambda_2, \lambda_3, \beta)$ obtained through the iterative method.*

n	λ_1	λ_2	λ_3	β	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\beta}$	SE($\widehat{\lambda}_1$)	SE($\widehat{\lambda}_2$)	SE($\widehat{\lambda}_1$3)	SE($\widehat{\beta}$)
30	1	2	0.5	0.5	1.9887	3.3081	1.9873	0.8154	0.0893	0.0902	0.0678	0.0603
50	1	2	0.5	0.5	1.6912	2.6332	1.0923	0.7385	0.0782	0.0893	0.0691	0.0472
100	1	2	0.5	0.5	1.3497	2.3213	0.7743	0.6293	0.0772	0.0869	0.0695	0.0449
30	1	2	0.5	1	1.9534	3.3109	1.9576	1.4734	0.1023	0.0953	0.0867	0.0703
50	1	2	0.5	1	1.5934	2.6156	1.0921	1.3921	0.0899	0.1012	0.0823	0.0612
100	1	2	0.5	1	1.3173	2.4679	0.7591	1.1581	0.0823	0.0897	0.0728	0.0499
30	1	2	0.5	1.5	1.9727	3.3278	1.9825	1.9568	0.1792	0.1034	0.0957	0.0684
50	1	2	0.5	1.5	1.7455	2.5357	0.9764	1.8627	0.0893	0.1125	0.0897	0.0692
100	1	2	0.5	1.5	1.5671	2.3513	0.7934	1.6193	0.0934	0.1045	0.0842	0.0502
30	1	2	1	0.5	1.9834	3.2895	2.4231	0.8245	0.1034	0.0957	0.0725	0.0769
50	1	2	1	0.5	1.6761	2.6281	1.5482	0.7372	0.0821	0.0993	0.0723	0.0584
100	1	2	1	0.5	1.3182	2.3756	1.2949	0.6429	0.0902	0.0931	0.0784	0.0521
30	1	2	1	1	1.9233	3.2956	2.3756	1.5334	0.1342	0.0913	0.0882	0.0705
50	1	2	1	1	1.5421	2.6492	1.5849	1.4294	0.0917	0.1034	0.0942	0.0736
100	1	2	1	1	1.2951	2.3682	1.2735	1.2661	0.0879	0.0921	0.0704	0.0569
30	1	2	1	1.5	1.9349	3.2937	2.4623	1.9173	0.1236	0.1412	0.1034	0.0756
50	1	2	1	1.5	1.5923	2.5634	1.4954	1.8728	0.1054	0.1532	0.1031	0.0713
100	1	2	1	1.5	1.3681	2.3542	1.2348	1.6212	0.0886	0.0993	0.0902	0.0534
30	1	2	2	0.5	1.9789	3.2782	3.4267	0.8178	0.0921	0.0941	0.0714	0.0589
50	1	2	2	0.5	1.5845	2.5936	2.5372	0.6912	0.0797	0.0901	0.0686	0.0484
100	1	2	2	0.5	1.3383	2.3417	2.2756	0.6389	0.0810	0.0889	0.0725	0.0467
30	1	2	2	1	1.8972	3.2852	3.4462	1.5343	0.1034	0.0872	0.0792	0.0602
50	1	2	2	1	1.4939	2.5789	2.5319	1.4014	0.0913	0.0989	0.0810	0.0579
100	1	2	2	1	1.3214	2.2973	2.2682	1.2314	0.0824	0.0901	0.0713	0.0498
30	1	2	2	1.5	1.9344	3.2604	3.4610	1.9282	0.1083	0.0991	0.0892	0.0659
50	1	2	2	1.5	1.6952	2.5021	2.4934	1.8317	0.0884	0.1153	0.0897	0.0604
100	1	2	2	1.5	1.3156	2.3023	2.2103	1.7116	0.0816	0.0927	0.0829	0.0523

5. Illustration

In this section, we have applied the AFT based load sharing model and estimation procedures to motor data obtained from Reliability Edge Home [19]. The dataset consists of 18 systems, each consisting of two motors operating continuously in parallel. The failure times of both motors, along with their identification labels A and B, were recorded.

Our objectives were twofold. Firstly, we aimed to determine whether the modified Weibull distribution (MWD) is an appropriate baseline distribution for modeling the lifetimes of both components. Secondly, we aimed to test whether there exists a load sharing phenomenon, where the failure of one motor affects the working of the other.

To assess the appropriateness of the MWD as the baseline distribution, we conducted a

Kolmogorov-Smirnov type test, which confirmed its suitability. However, it should be noted that the test was conservative due to the estimation of unknown parameters. We utilized a two-step estimation procedure, with the estimation of β being conducted using the 'average method'. The estimated values of λ_1, λ_2, λ_3, and β were found to be 0.0028, 2.08168×10^{16}, 31.6118, and 1.9847, respectively.

To investigate the presence of load sharing among the motor failure times, we employed a score-type test proposed by Sutar and Naik-Nimbalkar [11]. The computed test statistic value was 19.564, which surpassed the critical values at both the 1% and 5% significance levels. Consequently, we can infer that the failure of one motor has a significant impact on the lifetime of the other. This finding supports the existence of a load sharing phenomenon, where the failures of individual components influence the performance of the remaining components in the system. This conclusion is further supported by the estimated value of $\hat{\beta}$ being 1.9847 (significatly different than 1), suggesting that these 18 parallel systems exhibit load sharing or a load sharing phenomenon among the component failures.

6. Conclusions

In our study, we focused on a two-component parallel load sharing system and utilized the accelerated failure time based load sharing model to capture the load sharing behavior observed in this system. We chose the modified Weibull distribution as the baseline distribution for the component lifetime. We proposed two procedures for estimating the model parameters within this framework and also discussed a test procedure for assessing the presence of load sharing in such systems. Furthermore, we conducted a simulation study to evaluate the performance of the proposed estimation procedures, which demonstrated satisfactory results. To illustrate the practical applicability of the load sharing system, we analyzed a specific dataset. It is worth mentioning that the modeling and analysis of load sharing phenomena can be extended to more complex systems, such as a *k*-out-of-*m* system.

Acknowledgment

The Santosh S. Sutar is grateful for the financial support provided by Department of Science and Technology (DST), Science and Engineering Research Board (SERB), Government of India under a Core Research Grant (File No. CRG/2021/005672).

References

[1] Wang, D., Jiang, C. and Park, C. (2019). Reliability analysis of load-sharing systems with memory. *Lifetime Data Analysis*, 25:341–360.

[2] Liu, H. (1998). Reliability of a Load-Sharing k-out-of-n:G System: Non-iid Components with Arbitrary Distributions. *I.E.E.E. Transactions on Reliability*, 47:279-284.

[3] Drummond, H., Vazquez, E., Sanchez-Colon, S., Martinez-Gomez, M., Hudson, R. (2000). Competition for milk in the domestic rabbit: survivors benefit from littermate deaths. *Ethology*, 106:511-526.

[4] Daniels, H. E. (1945). The statistical theory of the strength of bundles of threads. I. *Proceedings of the Royal Society London A*, 183:404-435.

[5] Dewan I. and Naik-Nimbalkar U. V. (2010). Load-Sharing Systems. *Wiley Encyclopedia of Operations, Research and Management Science*, https://doi.org/10.1002/9780470400531.eorms0476.

[6] Deshpande, J. V., Dewan I. and Naik-Nimbalkar, U. V. (2010). A family of distributions to model load sharing systems. *J. Statist. Plann. Infer.*, 140:1441-1451.

[7] Park, C. (2010). Parameter estimation for the reliability of load-sharing systems. *IIE Trans*, 42:753-765.

[8] Singh B. and Gupta, P. K. (2012). Load-sharing system model and its application to the real data set. *Mathematics and Computers in Simulation* 82:1615-1629.

[9] Park, C. (2013). Parameter estimation from load-sharing system data using the expectation maximization algorithm. *IIE Trans*, 45:147-163.
[10] Gurov S. V. and Utkin, L. V. (2014). A load-share reliability model under the changeable piecewise smooth load. *Journal of Quality and Reliability Engineering*, 1-11.
[11] Sutar S. S. and Naik-Nimbalkar, U. V. (2014). Accelerated Failure Time Models For Load Sharing Systems. *I.E.E.E. Transactions on Reliability*, 63:706-714.
[12] Sutar S. S. and Naik-Nimbalkar, U. V. (2016). A model for *k*-out-of-*m* load sharing systems. *Communications in Statistics - Theory and Methods*, 45(20):5946-5960.
[13] Krivtsov, V., Amari S. and Gurevich V. (2017). Load sharing in series configuration. *Quality and Reliability Engineering International*, 1-12.
[14] Sutar S. S. and Naik-Nimbalkar, U. V. (2019). A load share model for non-identical components of a k-out-of-m system. *Applied Mathematical Modelling*, 72:486-498.
[15] Kamps, U. (1995). A concept of generalized order statistics. *J. Statist. Plann. Infer.* 48:1-23.
[16] Cramer, E. and Kamps, U. (2003). Marginal distributions of sequential and generalized order statistics. *Metrika*, 58:293-310.
[17] Sarhan A. M. and Zaindin M. (2009). Modified Weibull distribution. *Applied Science*, 11:123-136.
[18] Sutar S. S. (2014). Parameter estimation of the modified Weibull distribution using Monte Carlo Expectation Maximization algorithm. *Model Assisted Statistics and Applications*, 11:171–178.
[19] Reliability Edge Home, (2003), Quarter 2, 3, (http://www.reliasoftindia.com/newsletter/4q2002/qalt.htm).

Appendix (A): Expressions involved in Score functions

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{\left(\lambda_1\lambda_3\beta^{\lambda_3-1} + \lambda_2\lambda_3 y_i^{\lambda_3-1}\right)}{\left(\lambda_1\beta^{\lambda_3} + \lambda_2\lambda_3\beta y_i^{\lambda_3-1}\right)} - \frac{n(\lambda_3+1)}{\beta} + \frac{\lambda_1}{\beta^2}\sum_{i=1}^{n}(y_i - x_i) - \frac{\lambda_2\lambda_3}{\beta^{\lambda_3+1}}\sum_{i=1}^{n}(y_i^{\lambda_3} - x_i^{\lambda_3}) = 0, \tag{16}$$

$$\frac{\partial \log L}{\partial \lambda_1} = \sum_{i=1}^{n} \frac{\beta^{\lambda_3-1}}{\left(\lambda_1\beta^{\lambda_3} + \lambda_2\lambda_3\beta y_i^{\lambda_3-1}\right)} + \sum_{i=1}^{n} \frac{1}{\left(\lambda_1 + \lambda_2\lambda_3 x_i^{\lambda_3-1}\right)} - \frac{1}{\beta}\sum_{i=1}^{n}(y_i - x_i) - 2\sum_{i=1}^{n} x_i = 0,$$

$$\frac{\partial \log L}{\partial \lambda_2} = \sum_{i=1}^{n} \frac{\beta\lambda_3 y_i^{\lambda_3-1}}{\left(\lambda_1\beta^{\lambda_3} + \lambda_2\lambda_3\beta y_i^{\lambda_3-1}\right)} + \sum_{i=1}^{n} \frac{\lambda_3 x_i^{\lambda_3-1}}{\left(\lambda_1 + \lambda_2\lambda_3 x_i^{\lambda_3-1}\right)} - \frac{\lambda_2}{\beta^{\lambda_3}}\sum_{i=1}^{n}(y_i^{\lambda_3} - x_i^{\lambda_3}) - 2\lambda_2\sum_{i=1}^{n} x_i^{\lambda_3} = 0,$$

and

$$\frac{\partial \log L}{\partial \lambda_3} = \sum_{i=1}^{n} \frac{\left(\lambda_1\beta^{\lambda_3}\log\beta + \lambda_2\beta y_i^{\lambda_3-1} + \lambda_2\lambda_3\beta y_i^{\lambda_3-1}\log y_i\right)}{\left(\lambda_1\beta^{\lambda_3} + \lambda_2\lambda_3\beta y_i^{\lambda_3-1}\right)} + \sum_{i=1}^{n} \frac{\lambda_2 x_i^{\lambda_3-1} + \lambda_2\lambda_3 x_i^{\lambda_3-1}\log(x_i)}{\left(\lambda_1 + \lambda_2\lambda_3 x_i^{\lambda_3-1}\right)} - n\log\beta - \lambda_2\sum_{i=1}^{n}\left(\frac{y_i}{\beta}\right)^{\lambda_3}\log\left(\frac{x_i}{\beta}\right) - \lambda_2\sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\lambda_3}\log\left(\frac{y_i}{\beta}\right) - 2\lambda_2\sum_{i=1}^{n} x_i^{\lambda_3}\log(x_i) = 0.$$

Appendix (B): The information pertaining to the conditional density of W_{11} given $X = x$ and $Z_{21} = z_{21}$

Consider

$$\begin{aligned}\bar{G}(w_1, z_{21}|X = x) &= P(Z_{21} > z_{21}, W_{11} > w_1|X = x)\\ &= \bar{F}_{W_1|X=x}(\max(z_{21}, w_1))\bar{F}_{W_2|X=x}(z_{21})\\ &= \exp\left\{-\theta_1[\max(z_{21}, w_1) - x] - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\right\}.\end{aligned}$$

Due to ordering in z_{21} and w_1, we have following three cases-

1. $z_{21} > w_1$ i.e. $\max(z_{21}, w_1) = z_{21}$.
2. $z_{21} < w_1$ i.e. $\max(z_{21}, w_1) = w_1$.
3. $z_{21} = w_1$ i.e. $\max(z_{21}, w_1) = z_{21}$ *or* w_1.

When, $z_{21} > w_1$ we have,

$$\bar{G}(w_1, z_{21}|X = x) = \exp\left\{-\theta_1(z_{21} - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\right\}.$$

Thus, we have

$$g(w_1, z_{21}|X = x) = \frac{\partial^2}{\partial z_{21}\partial w_1}\bar{G}(z_{21}, w_1|x) = 0,\ z_{21} > w_1 > 0.$$

When, $z_{21} < w_1$ we have,

$$\bar{G}(w_1, z_{21}|X = x) = \exp\left\{-\theta_1(w_1 - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\right\}\ , w_1 > z_{21} > 0,$$

and hence

$$g(w_1, z_{21}|X = x) = \frac{\partial^2}{\partial z_{21}\partial w_1}\bar{G}(z_{21}, w_1|x)\ , w_1 > z_{21} > 0.$$

That is

$$g(w_1, z_{21}|X = x) = \theta_1\theta_2\theta_3 z_{21}^{\theta_3-1}\exp\left\{-\theta_1(w_1 - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\right\}\ , w_1 > z_{21} > 0.$$

When, $z_{21} = w_1$ we have,

$$\bar{G}(w_1, z_{21}|X = x) = \exp\{-\theta_1(z_{21} - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\}\ , w_1 = z_{21} > 0.$$

Thus, we have

$$\begin{aligned}g(w_1, z_{21}|X = x) &= \frac{\partial^2}{\partial z_{21}\partial w_1}\bar{G}(z_{21}, w_1|x)\ , w_1 = z_{21} > 0\\ &= \theta_1\exp\left\{-\theta_1(z_{21} - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\}\right\}, z_{21} = w_1\ , w_1 = z_{21} > 0.\end{aligned}$$

Thus, by combining all the above cases, we can write the joint density as

$$\begin{aligned}g(w_1, z_{21}|X = x) = \theta_1\theta_2\theta_3 z_{21}^{\theta_3-1}&\exp\{-\theta_1(w_1 - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\}I(z_{21} = w_1)\\ &+\theta_1\exp\{-\theta_1(z_{21} - x) - \theta_2(z_{21}^{\theta_3} - x^{\theta_3})\}I(z_{21} < w_1).\end{aligned}$$

Hence, the conditional density of W_{11} given $X = x, Z_{21} = z_{21}$ can be obtained as

$$\begin{aligned}g(w_1|x, z_{21}) &= \frac{g(w_1, z_{21}|X = x)}{g(z_{21}|X = x)}\\ &= \frac{\theta_1}{\left(\theta_1 + \theta_2\theta_3 z_{21}^{\theta_3-1}\right)}\left\{I_{(w_1=z_{21})} + \theta_2\theta_3 z_{21}^{\theta_3-1}\exp\left\{-\theta_1(w_1 - z_{21})\right\}I_{(w_1>z_{21})}\right\}.\end{aligned}$$

Appendix (C): Expressions involved in (14)

$$\frac{\partial^2 E_c\left[\log L|\underline{x},\underline{y}\right]}{\partial(\theta_3^{(r)})^2} = -\frac{n}{(\theta_3^{(r)})^2} - n\frac{\left\{E_c\left[W_{2i}^{\theta_3^{(r)}}|\underline{x},\underline{y}\right]\frac{\partial^2 E_c\left[W_{2i}^{\theta_3^{(r)}}|\underline{x},\underline{y}\right]}{\partial(\theta_3^{(r)})^2}\right\} - \left\{\frac{\partial E_c\left[W_{2i}^{\theta_3^{(r)}}|\underline{x},\underline{y}\right]}{\partial(\theta_3^{(r)})^2}\right\}}{\left\{E_c\left[W_{2i}^{\theta_3^{(r)}}|\underline{x},\underline{y}\right]\right\}^2},$$

with

$$\frac{\partial E_c\left\{\sum_{i=1}^n W_{2i}^{\theta_3^{(r+1)}}|\underline{x},\underline{y}\right\}}{\partial\theta_3^{(r)}} = \sum_{i=1}^n \frac{\theta_2\theta_3 y_i^{\theta_3^{(r+1)}+\theta_3-1}\log(y_i)}{\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

$$-\sum_{i=1}^n\sum_{j=1}^m \frac{\theta_1^{(r)}\theta_2^{(r)^{-\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}}\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)^{\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}}{m\theta_3^{(r)}\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

$$+\sum_{i=1}^n\sum_{j=1}^m \frac{\theta_1^{(r)}\theta_2^{(r)^{-\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}}\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)^{\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}\log\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)}{m\theta_3^{(r)}\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

and

$$\frac{\partial^2 E_c\left\{\sum_{i=1}^n W_{2i}^{\theta_3^{(r+1)}}|\underline{x},\underline{y}\right\}}{\partial(\theta_3^{(r)})^2} = \sum_{i=1}^n \frac{\theta_2\theta_3 y_i^{\theta_3^{(r+1)}+\theta_3-1}(\log(y_i))^2}{\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

$$+\sum_{i=1}^n\sum_{j=1}^m \frac{\theta_1^{(r)}\theta_2^{(r)^{-\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}}\left(\log\theta_2^{(r)}\right)^2\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)^{\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}}{m\left(\theta_3^{(r)}\right)^2\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

$$-\sum_{i=1}^n\sum_{j=1}^m \frac{\theta_1^{(r)}\theta_2^{(r)^{-\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}}\left(\log(\theta_2^{(r)})\right)\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)^{\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}\log\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)}{m\left(\theta_3^{(r)}\right)^2\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

$$-\sum_{i=1}^n\sum_{j=1}^m \frac{\theta_1^{(r)}\theta_2^{(r)^{\frac{-\theta_3^{(r+1)}}{\theta_3^{(r)}}}}\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)^{\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}\log\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)}{m\left(\theta_3^{(r)}\right)^2\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}$$

$$+\sum_{i=1}^n\sum_{j=1}^m \frac{\theta_1^{(r)}\theta_2^{(r)^{\frac{-\theta_3^{(r+1)}}{\theta_3^{(r)}}}}\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)^{\frac{\theta_3^{(r+1)}}{\theta_3^{(r)}}}\left(\log\left(v_j+\theta_2^{(r)}y_i^{\theta_3}\right)\right)^2}{m\left(\theta_3^{(r)}\right)^2\left(\theta_1+\theta_2\theta_3 y_i^{\theta_3-1}\right)}.$$

RANDOMIZED BLOCK DESIGN IN FUZZY ENVIRONMENTS

A. Kirthik VairaMariappan[1] and Manigandan Palanisamy[2]

•

[1]Department of Statistics, Government Arts College, Dharmapuri, Tamilnadu, India
maryaaindia@gmail.com
[2]Department of Statistics, Periyar University, Salem, Tamilnadu, India
srmanigandan95@gmail.com

Abstract

On the basis of the statistics, ANOVA also provides a method of data analysis that is motivated by consideration of the experimental design or Design of Experiment (DOE). Experimental design plays an essential part in statistical analysis and data interpretation. One factor of criteria forms the basis of a one-way classification. Two factors or two criteria form the basis of two-way classification. Innovations and creations require experimentation as their foundation. Replication, randomization, and local control are the three fundamental tenets of experimental designs, which are used to determine the cause and effect of interactions. The error of any treatment can be isolated and any number of treatments may be omitted from the analysis without complicating it. The data provided in this study are vague and need an extended version of the RBD to investigate these vague observations. The simplest of all designs based on the principles of randomization and replication are Completely Randomized Designs (CRD). When the experimental materials aren't uniform in some circumstances. Divide the experimental region into smaller, homogeneous blocks in RBD. The treatment is applied at random to each block, and each block is reproduced. Since uncertainty is a common feature of all real-world issues and denotes fuzziness and unpredictability, Randomized Block Design has long been widely used in the agricultural and industrial sectors. It is therefore impossible to avoid using statistical RBD analysis with fuzzy observations. The objective of this study was to develop the problem of a Randomized Block Design (RBD) test for Triangular Fuzzy Numbers (TFN) is discussed in this paper. However, in a scenario that is actual, the underlying relationship is not a clear-cut function of a particular form; it has some ambiguity or imprecision. The estimated numbers are very similar to the actual ones. This approach may generally be used for any real-time triangle fuzzy number calculation. In this proposed methodology, it is obvious that if the value of the observed fuzzy test statistics is similar to real numbers in the testing crisp hypotheses, then fuzzy RBD is very sensitive for making the determinations as to whether to accept or reject the fuzzy null hypotheses and also debates the application of the method for example.

Keywords: RBD, Fuzzy RBD, TFN, Decision Rule

1. Introduction

The Completely Randomized Design (CRD) was simple because the principle of local control was not used, and experimental material was assumed to be homogeneous, but it is noted that the experimental material is not absolutely homogeneous. A fertility gradient in one direction is often present in agricultural field experiments. The simple method of regulating the variability of the

experimental material in such a situation consists of stratifying or grouping the entire experimental area into relatively homogeneous strata or subgroups (called blocks) perpendicular to the fertility gradient direction. These blocks are so designed that plots are homogeneous within a block, and heterogeneous between blocks. In other words, inside a block there might be less variation, and the main difference or variation between blocks. It should be held in mind that for an effective blocking of the content, familiarity with the design of experimental units is important. The method of dividing experimental material into a number of blocks gives rise to a design known as RBD that can be described as an arrangement of t treatments in r blocks such that each treatment takes place exactly once in each block. Fuzzy set theory [23] was extended to several areas that need to handle ambiguous and unclear data. These areas include estimated logic, decision-making, optimization, power, etc.

The sample findings are crisp in conventional statistical research, and a statistical test leads to the binary decision. Many authors have studied the statistical theories that are evaluated in fuzzy environments using the fuzzy set theory principles introduced by Zadeh [24]. Chachi et al. [4] are proposing a new approach to the issue of evaluating statistical hypotheses. As a fuzzy subset of the real line, Dubois and Prade [6] identified some of the fuzzy numbers. Mikihiko Konishi et al. [16] suggested an Analysis of Variance (ANOVA) for the fuzzy interval data using the definition of the fuzzy set. Wu [21, 22] introduced hypothesis testing of a single factor ANOVA model for fuzzy data by solving optimization problems using the h-level and the notions of pessimistic degree and optimistic degree. The two-factor ANOVA test were analysed by Gajivaradhan and Parthiban [8] using an alpha cut interval method for trapezoidal fuzzy numbers. A bootstrap approach to the multi-sample test of means with imprecise data was suggested by Gil et al. [10]. When both the theories and the available data are fuzzy, Arefi and Taheri [2] formed the testing hypothesis. Filzmoser and Viertl [7], proposed to test hypotheses on the fuzzy p value with fuzzy data. Nakama et al. [15] Discuss derive statistical tests that are ideal for testing the null hypotheses, and develop a bootstrap scheme to estimate the p values of the test statistics observed. Ahmed et, al. [1], proposed, a new dimension of the methodology involving a fuzzy regression approach to RBD introduced, which is involving qualitative predictor variables under consideration on multiple linear regression. The idea from this research will be a useful thread for establishing comprehensive connectivity between RBD and regression. The researchers concluded that fuzzy MLR can predict much better compared to MLR itself. Mariappan and Pachamuthu [14] suggested the statistical testing of hypotheses for fuzzy CRD using TFN. Magno do N et, al. [13], developed a fuzzy model that could predict weight loss as a function of the rapid cooling of table grapes in different plastic film bags. Modelling was performed using three types of plastic film bags (micro-perforated, macro-perforated, and non-perforated) at three levels of palletization (lower, intermediate, and upper), arranged in an experimental design in randomized blocks, in a 3 × 3 factorial scheme, with three blocks. The influence of the relative humidity and amplitude of humidity on the variable weight loss percentage of the Arra 15 grape variety was measured. The average percentage error of the fuzzy model was 9.78%. The intermediate level alone showed an error of 4.02%. Thus, the developed fuzzy model provided a good prediction of the weight loss of table grapes. The classical RBD model for TFN is analyzed in this paper using a numerical example.

2. Preliminaries

2.1 Triangular Fuzzy Number

A triangular fuzzy number $\tilde{A}$ is a fuzzy number fully specified by tribles (a,b,c) such that $a \le b \le c$ with membership function defined as

$$\mu_{\tilde{A}}(x)=\begin{cases}0 & if\ x\le a\\ \dfrac{x-a}{b-a} & if\ a\le x\le b\\ \dfrac{x-c}{b-c} & if\ b\le x\le c\\ 0 & if\ x\ge c\end{cases}$$

where a is the indicates of lower point, b is the indicates of centre point and c is the indicates of upper point.

The triangular fuzzy number is represented diagramatically as

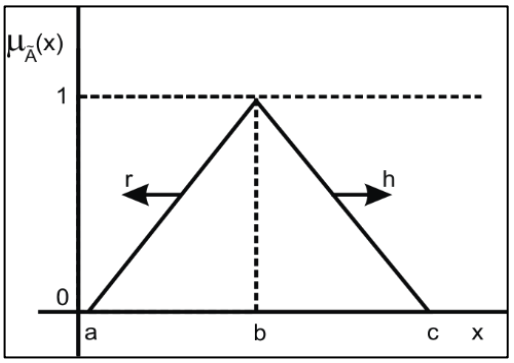


Figure 1: *Triangular Fuzzy Numbers*

The form of a fuzzy interval number can be expressed as a triangular fuzzy number follows:

$$\tilde{A}=\left[\{(b-a)r+a\}^{L};\{(b-c)h+c\}^{U}\right];0\le h,r\le 1$$

where r is the level of pessimistic and h is the level of optimistic of the fuzzy numbers $\tilde{A}=(a,b,c)$.

3. Statistical Analysis of RBD

Through proving the local control (blocking) measure in the design, an increase in *CRD* can be obtained. One such design is entirely *RBD*. *ANOVA* technique for two-way data classification is applicable to the *RBD* layout experiment. The data obtained from the experiment is graded by two factors namely treatments and blocks according to different levels. For *RBD*, the linear model is described by

$$y_{ij}=\mu+a_i+b_j+e_{ij};i=1,2,...,t;j=1,2,...,b$$

where, y_{ij} is the observations corresponding to the i^{th} treatment and j^{th} block, μ is the general mean effect which is fixed, a_i is the fixed effect due to the i^{th} treatment, b_j is the fixed effect due to the j^{th} block and e_{ij} is the random error effect. To determine whether the factor level means μ_i equal or not. The following testing hypotheses are known as

$$H_0:\mu_1=\mu_2=\mu_3=\mu_4 \text{ against } H_1:\text{not all } \mu_i \text{ are equal}$$

Let $\sum_{ij}^{tb} y_{ij}=y_{..}=G$ be the grand total of tb observations, $\sum_{ij}^{tb} y_{ij}=y_{i.}=T_i$ be the i^{th} treatment total, $\sum_{ij}^{tb} y_{ij}=y_{.j}=B_j$ is the j^{th} block total and also $cf=\dfrac{G^2}{kr}$. Then, the various sum of squares, mean sum of squares and *F* Ratio listed given below:

$SST = Q_{SST} = \sum_{ij}^{tb} y_{ij}^2 - \frac{G^2}{kr}$ which has $(tb-1)\,df$, $SSBTR = Q_{SSBTR} = \sum_{ij}^{tb} \frac{T_i^2}{r} - \frac{G^2}{kr}$ which has $(t-1)$ df, $SSB = Q_{SSB} = \sum_{ij}^{tb} \frac{B_j^2}{k} - \frac{G^2}{kr}$ which has $(b-1)$ df, $SSE = Q_{SSE} = Q_{SST} - Q_{SSBTR} - Q_{SSBB}$ which has $(t-1)(b-1)$ df, $MSBTR = \frac{SSBTR}{(t-1)}$, $MSBB = \frac{SSBB}{(b-1)}$, $MSE = \frac{SSE}{(t-1)(b-1)}$, $F_{BTR} = \frac{MSBTR}{MSE}$ and $F_{BB} = \frac{MSBB}{MSE}$.

In the ANOVA table, all these values are referred to and inferences are drawn.

Table 1: *ANOVA Table for Crisp RBD*

SV	*df*	*SS*	*MSS*	*F Ratio*
Between Treatments	$(t-1)$	Q_{SSBTR}	$MSBTR$	F_{BTR}
Between Blocks	$(b-1)$	Q_{SSBB}	$MSBB$	F_{BB}
Experimental Error	$(t-1)(b-1)$	Q_{SSE}	MSE	-
Total	$(tb-1)$	Q_{SST}	-	-

3.1 Decision Rule of Between Treatments and Between Blocks

The decision rules of *F* test to accept or reject between treatments and between blocks at $\alpha\%$ significance level the null hypothesis and alternative hypothesis. Suppose that if $F_T < F_C$, [where F_T is the tabulated value for $(t-1),(t-1)(b-1)$ and $(b-1),(t-1)(b-1)$ degrees of freedom, and F_C is the calculated value], then the null hypothesis H_0 is rejected. Otherwise, alternative hypothesis H_0 is rejected.

3.2 Fuzzy Analysis of RBD

The triangular fuzzy approach to the fuzzy statistical analysis of *RBD*. Throughout this case, the data recorded as well as the observations are regarded as *TFN*. Below is the mathematical general linear model:

$$\tilde{y}_{ij} = \tilde{\mu} + \tilde{a}_i + \tilde{b}_j + e_{ij};\ i = 1,2,...,t;\ j = 1,2,...,b$$

In the fuzzy interval *RBD* models, the general linear model of classical *RBD* is classified; the fuzzy lower and upper level models are regarded as: $\tilde{y}_{ij}^L = (\tilde{\mu})_r^L + (\tilde{a}_i)_r^L + (\tilde{b}_j)_r^L + (\varepsilon_{ij})_r^L$ and $\tilde{y}_{ij}^U = (\tilde{\mu})_h^U + (\tilde{a}_i)_h^U + (\tilde{b}_j)_h^U + (\varepsilon_{ij})_h^U$ in which $(\tilde{y}_{ij})_r^L$ and $(\tilde{y}_{ij})_h^U$ is the observation corresponding to the i^{th} level of factor A and j^{th} level of factor B. $(\tilde{\mu})_r^L$ and $(\tilde{\mu})_h^U$ is the general mean effect which is fixed. $(\tilde{a}_i)_r^L$ and $(\tilde{a}_i)_h^U$ is the fixed effect due to the i^{th} level of factor A. $(\tilde{b}_j)_r^L$ and $(\tilde{b}_j)_h^U$ is the fixed effect due to the j^{th} level of factor B. $(\tilde{\varepsilon}_{ij})_r^L$ and $(\tilde{\varepsilon}_{ij})_h^U$ is the random error effect which is independent identically distributed (iid) with mean is 0 and constant variance is σ^2; $i = 1,2,...,t$ and $j = 1,2,...,b$. After this, to test the lower and upper level model and the fuzzy null hypotheses and fuzzy alternative hypotheses respectively, utilizing classical RBD methods. The simplest the fuzzy null hypotheses $\tilde{H}_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = ... = \tilde{\mu}_r$ against the fuzzy alternative hypotheses $\tilde{H}_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2 \neq ... \neq \tilde{\mu}_r$. This implies the following two sets (Lower and Upper levels) of hypotheses are given below.

3.3 Fuzzy Hypotheses of Lower and Upper Level Models

The fuzzy null hypotheses of between treatment and between block is $\tilde{H}_0^L : \tilde{\mu}_1^L = \tilde{\mu}_2^L = ... = \tilde{\mu}_r^L$ against the fuzzy alternative hypotheses between treatment and between block is $\tilde{H}_1^L : \tilde{\mu}_1^L \neq \tilde{\mu}_2^L \neq ... \neq \tilde{\mu}_r^L$.

The fuzzy null hypotheses of between treatment and between block is $\tilde{H}_0^U : \tilde{\mu}_1^U = \tilde{\mu}_2^U = ... = \tilde{\mu}_r^U$ against the fuzzy alternative hypotheses of between treatment and between block is $\tilde{H}_1^U : \tilde{\mu}_1^U \neq \tilde{\mu}_2^U \neq ... \neq \tilde{\mu}_r^U$.

In *TFN* pessimistic and optimistic for the fuzzy lower and upper level models from the null hypothesis of acceptance or rejection direction levels. Through the use of triangular fuzzy lower and upper levels formulas are $(b_{ij} - a_{ij})r + a_{ij}$ where $0 \le i \le t; 0 \le j \le b$ and $(b_{ij} - c_{ij})h + c_{ij}$ where $0 \le i \le t; 0 \le j \le b$. (Note that $r^L = 1$ and $h^U = 1$, centre level). Then the required formula for lower level of fuzzy *RBD* is given below:

$$SST_r^L = \sum_{i=1}^{t}\sum_{j=1}^{b}[(\tilde{y}_{ij})_r^L] - \frac{[(\tilde{y}_{..}^2)_r^L]}{tb}, \quad SSBTR_r^L = \sum_{i=1}^{t}\frac{[(\tilde{y}_{i.}^2)_r^L]}{b} - \frac{[(\tilde{y}_{..}^2)_r^L]}{tb}$$

$$SSBB_r^L = \sum_{j=1}^{b}\frac{[(\tilde{y}_{i.}^2)_h^U]}{t} - \frac{[(\tilde{y}_{..}^2)_h^U]}{tb}, \quad SSE_r^L = SST_r^L - SSBTR_r^L - SSBB_r^L$$

$$MSBTR_r^L = \frac{SSBTR_r^L}{(t-1)}, \quad MSBB_r^L = \frac{SSBB_r^L}{(b-1)} \text{ and } MSE_r^L = \frac{SSE_r^L}{(t-1)(b-1)}$$

$$(\tilde{F}_{BTR})_r^L = \frac{MSBTR}{MSE} \text{ and } (\tilde{F}_{BB})_r^L = \frac{MSBB}{MSE}$$

In the lower level of *ANOVA* table for fuzzy *RBD* table, all these values are represented and fuzzy decision rule is drawn.

Table 2: *ANOVA Table for Lower Level of Fuzzy RBD*

SV	*df*	*SS*	*MSS*	$\tilde{F} - Ratio$
Between Treatments	$(t-1)$	$SSBTR_r^L$	$MSBTR_r^L$	$(\tilde{F}_{BTR})_r^L$
Between Blocks	$(b-1)$	$SSBB_r^L$	$MSBB_r^L$	$(\tilde{F}_{BB})_r^L$
Experimental Error	$(t-1)(p-1)$	SSE_r^L	MSE_r^L	-
Total	$(tb-1)$	SST_r^L	-	-

3.2 Fuzzy Decision Rule of (Lower and Upper levels) Between Treatments and Between Blocks

Suppose that if $F_T < F_C$, [where F_T is the tabulated value for $(t-1),(t-1)(b-1)$ and $(b-1),(t-1)(b-1)$ degrees of freedom, and F_C is the calculated value (using 3.1)], then the fuzzy null hypotheses of lower level for $\tilde{H}_0^L$ and fuzzy null hypotheses of upper level for $\tilde{H}_0^U$ is rejected for $0 \le r^L \le r_T$ where $0 \le r_T \le 1$ and $0 \le h^U \le h_T$ where $0 \le h_T \le 1$. Otherwise, fuzzy alternative hypotheses of lower level for $\tilde{H}_0^L$ and fuzzy alternative hypotheses of upper level for $\tilde{H}_0^U$ is rejected for $0 \le h^U \le h_T$ where $0 \le h_T \le 1$.

The proposed classical technique for evaluating *RBD* model fuzzy hypotheses with fuzzy

data is illustrated with an example below.

$$e = mc^2 \tag{1}$$

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

4. Applications

In our study, to collect the yields of primary data groundnut varieties at Omalur, Salem District of Tamil Nādu. Three replicates of various groundnut varieties (TMV 2, TMV 7, VRI 2) in kilograms and four yields of (Y1, Y2, Y3, Y4). Via an *RBD*, with four replications of groundnut in kilograms for yields per plot, three varieties of crops are tested, the layout being *TFN* due to certain work friction is given as below.

Table 3: *Table for Classical RBD using TFN*

Varieties of Groundnut	Yields in kilograms			
	Y1	Y2	Y3	Y4
TMV 2	56,58,60	54,58,62	53,56,59	54,58,62
TMV 7	58,60,62	53,58,63	56,59,62	57,60,63
VRI 2	59,62,65	58,60,62	57,60,63	57,59,61

To test if there is any substantial difference between the production of the groundnut varieties in the yields in kilograms per plot. Let $\tilde{\mu}_i$ be the mean number of yields in kilograms per plots for the i^{th} varieties of groundnut. Now, the null hypothesis, $\tilde{H}_0 : \tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}_3 = \tilde{\mu}_4$ and the alternative hypothesis, $\tilde{H}_1$: not all $\tilde{\mu}_i$'s are equal.

$\tilde{H}_0$: To test whether groundnut varieties do not vary significantly with respect to yields.

$\tilde{H}_1$: To test if groundnut varieties vary significantly with respect to yields.

Let us consider the lower-level model is given below

4.1. Lower Level Model

Table 4: *Table for Upper Level Model*

Varieties of Groundnut	Yields in kilograms			
	Y1	Y2	Y3	Y4
TMV 2	$2r+56$	$4r+54$	$3r+53$	$4r+54$
TMV 7	$2r+58$	$5r+53$	$3r+56$	$3r+57$
VRI 2	$3r+59$	$2r+58$	$3r+57$	$2r+57$

$$SST_r^L = 10r^2 - 30r + 46,\ SSBTR_r^L = 1.5r^2 - 10.5r + 24.5$$
$$SSBB_r^L = 2.67r^2 - 10.67r + 12.67,\ SSE_r^L = 5.83r^2 - 8.83r + 8.83$$

$$MSBTR_r^L = 0.75r^2 - 5.25r + 12.25\,,\ MSBB_r^L = 0.89r^2 - 3.56r + 4.22\,,$$

$$MSE_r^L = 0.97r^2 - 1.47r + 1.47$$

$$(\tilde{F}_{BTR})_r^L = \frac{0.75r^2 - 5.25r + 12.25}{0.97r^2 - 1.47r + 1.47},\ (\tilde{F}_{BB})_r^L = \frac{0.89r^2 - 3.56r + 4.22}{0.97r^2 - 1.47r + 1.47}$$

4.2. Fuzzy Decision Rule of Between Treatments

If $\tilde{F}_r^L > F_T$, for all $r; 0 \le r \le 1$ where $F_T = 5.14$ is the F table value of α at 5% level of significance with $(2,6)\,df$ then, the fuzzy null hypotheses $\tilde{H}_0^L$ is rejected for the $r; 0 \le r \le 1$. Thus, the disparity between the treatments is substantial. Therefore, groundnut varieties vary greatly in yields.

4.3. Fuzzy Decision Rule of Between Blocks

If $\tilde{F}_r^L < F_T$, for all $r; 0 \le r \le 1$ where $F_T = 4.76$ is the F table value of α at 5% level of significance with $(3,6)\,df$ then, the fuzzy null hypotheses $\tilde{H}_0^L$ is accepted for the $r; 0 \le r \le 1$. Furthermore, the difference between treatments is not significant. Therefore, the groundnut varieties are not substantially different in terms of yields.

Let us consider the upper level model is given below

4.4. Upper Level Model

Table 4: *Table for Upper Level Model*

Varieties of Groundnut	Yields in kilograms			
	Y1	Y2	Y3	Y4
TMV 2	$-2h+60$	$-4h+62$	$-3h+59$	$-4h+62$
TMV 7	$-2h+62$	$-5h+63$	$-3h+62$	$-3h+63$
VRI 2	$-3h+65$	$-2h+62$	$-3h+63$	$-2h+61$

Likewise, upper level models of ambiguous *RBD* use formula and table, thus avoiding calculation.

4.5. Fuzzy Decision Rule of Between Treatments

If $F_T > \tilde{F}_h^U$, for all $h; 0 \le h \le 1$ where $F_T = 5.14$ is the F table value of α at 5% level of significance with $(2,6)\,df$ then, the fuzzy null hypotheses $\tilde{H}_0^U$ is accepted for the $h; 0 \le h \le 1$. Consequently, the disparity between treatments is not significant. Therefore, with regard to yields in kilograms, groundnut varieties do not vary significantly.

4.6. Fuzzy Decision Rule of Between Blocks

If $F_T > \tilde{F}_h^U$, for all $h; 0 \le h \le 1$ where $F_T = 8.14$ is the F table value of α at 5% level of significance with $(6,3)\,df$ then, the null hypothesis of the $\tilde{H}_0^U$ is accepted for the $h; 0 \le h \le 1$. Therefore, the discrepancy between treatments is not significant. Consequently, with regard to yields in kilograms, groundnut varieties do not vary significantly.

Therefore, so because fuzzy null hypotheses between treatments $\tilde{H}_0^L$ and $\tilde{H}_0^U$ of the lower level data is rejected and upper level data is accepted for all $r;0\le r\le 1$ and $h;0\le h\le 1$ (note that null hypotheses of accepted or rejected at $r=1$ and $h=1$, that is the centre level), the between blocks of fuzzy null hypothesis $\tilde{H}_0^L$ and $\tilde{H}_0^U$ of the lower and upper level data is accepted for all $r;0\le r\le 1$ and $h;0\le h\le 1$ the fuzzy null hypotheses $\tilde{H}_0$ of the fuzzy *RBD* model is accepted and rejected for all $r;0\le r\le 1$ and $h;0\le h\le 1$. Thus, we conclude that four yields of groundnut kilograms are equal if $r;0\le r\le 1$ and $h;0\le h\le 1$.

5. Conclusion

A statistical test of the hypothesis for RBD model using TFN for fuzzy data is suggested in this study. Can make a decision on the fuzzy RBD model hypothesis based on the hypothesis determinations of two crisp RBD models. Since our fuzzy test is rather than standard significance tests, it appears to be a useful method in circumstances with imprecise data and also extend the crisp RBD to LSD, BIBD, PBIBD for fuzzy environments.

References

[1] Ahmed, W. M. A. W., Khan, S. Q., Rohim, R. A. A., Aleng, N. A., & Ghazali, F. M. M. (2020). Approximation of Randomized block design towards fuzzy multiple linear regression: A case study in health sciences. *International Journal of Scientific and Technology Research,* 9(1), 1303-8.

[2] Arefi M. and Taheri S.M. (2011). Testing Fuzzy Hypotheses Using Fuzzy Data Based on Fuzzy Test Statistic, *Journal of Uncertain Systems,* 5(1), 45-61.

[3] Chanas S. (1999). On the interval approximation of a fuzzy numbers. *Fuzzy sets and Systems*, 102:221-226.

[4] Chachi J., Taheri S. M. and Viertl R. (2012). Testing Statistical Hypotheses based on Fuzzy Confidence Intervals. *Forschungsbericht SM-2012-2, Technische Universitat Wien, Austria*,

[5] Dubois D. and Prade H. Fuzzy Sets and Systems: Theory and Application, Academic Press, New York, 1980.

[6] Dubois D. and Prade H. (1978). Operations on fuzzy numbers, *International Journal of System Science*, 9:613-626.

[7] Filzmoser P. and Viertl R. (2009). Testing Hypotheses with Fuzzy Data: The Fuzzy P value, *Metrika*, 59:21-29.

[8] Gajivaradhan P. and Parthiban S. (2016). A Comparative Study of Two Factor ANOVA Model Under Fuzzy Environments using Trapezoidal Fuzzy Numbers. *International Journal of Fuzzy Mathematical Archive*, 10(1):1-25.

[9] George.J. Klir and Bo Yuan, Fuzzy sets and fuzzy logic, Theory and Applications, Prentice-Hall, New Jersey, 2008.

[10] Gil et al. (2006). Bootstrap Approach to the Multi-sample Test of Means with Imprecise Data. *Computer Statistics and Data Analysis* 51:148-162.

[11] Gupta S.C. and Kapoor V.K., Fundamentals of applied statistics, Sultan Chand & Sons, New Delhi, India, 2007.

[12] Hocking R.R., Methods and applications of linear models: regression and the analysis of variance, New York: John Wiley & Sons, 1996.

[13] Magno do N et, al. (2022). Fuzzy Modeling for Rapid Cooling of Table Grapes in Different Plastic Film Bags. *Scientific Paper Eng. Agríc*, 42 (1).

[14] K.G. Manton et al., Statistical applications using fuzzy sets, New York: John Wiley & Sons, 1994.

[15] Mariappan A. and Pachamuthu M. (2020). Statistical Testing of Hypotheses for Fuzzy Completely Randomized Design. *Journal of Xidian University*, Vol. 14, Issue 9.

[16] Mikihiko Konishi, Tetsuji Okuda and Kiyoji Asai, (2006). Analysis of Variance based on Fuzzy Interval Data using Moment Correction Method. *International Journal of Innovative Computing, Information and Control,* 2:83-99.

[17] Nakama T., Colubi A., and Lubiano M.A. (2010). Two-Way Analysis of Variance for Interval-Valued Data. *Combining Soft Computing & Stats. Methods*, AISC 77:475–482.

[18] Nguyen H.T. and Walker E.A. A First Course in Fuzzy Logic (3rd ed.), Paris: Chapman Hall/CRC, 2005.

[19] Viertl R. (2006). Univariate statistical analysis with fuzzy data. *Computational Statistics and Data Analysis,* 51:33-147.

[20] Viertl R. Statistical methods for fuzzy data, John Wiley and Sons, 2011.

[21] Wu H.C. (2005). Statistical hypotheses testing for fuzzy data. *Information Sciences*, 175: 30-56.

[22] Wu H.C. (2007). Analysis of variance for fuzzy data. *International Journal of Systems Science*, 38:235-246.

[23] Zadeh L.A. (1965). Fuzzy sets. *Information and Control*, 8:338-353.

[24] Zadeh L.A. (1975). The Concept of a Linguistic Variable and its application to approximate Reasoning, Information Sciences, 8:199-249.

RT&A, No.3 (74)
Volume 18, September 2023

A New Class of Sin-G Family of Distributions with Applications to Medical Data

LAXMI PRASAD SAPKOTA, PANKAJ KUMAR AND VIJAY KUMAR

•

Department of Mathematics and Statistics, DDU Gorakhpur University, Gorakhpur, UP, India.
lpsapkota75@gmail.com, pankajagadish@gmail.com, vkgkp@rediffmail.com

Abstract

This article is dedicated to the study of the new class of distributions and one of its particular members. Based on the ratio of CDF $G(x)$ and $1+G(x)$ of the baseline distribution, we have developed the new trigonometric family of distributions by transforming the sine function, and we named it the new class sin-G (NCS-G) family of distributions. The general properties of the suggested family of distributions are provided. Using the inverted Weibull distribution as a baseline distribution, we have introduced a member of the suggested family having a reverse-j or increasing, or inverted bathtub-shaped hazard function. Some statistical properties of this NCS-IW distribution are explored. The associated parameters of the new distribution are estimated through the MLE method. To assess the estimation procedure, we conducted a Monte Carlo simulation and found that even for small samples, biases and mean square errors decreased as the size of the sample increased. Two real medical data sets are considered for the application of the NCS-IW distribution. Using some criteria for model selection and goodness of fit test statistics, we empirically proved that the suggested model performs better than six other existing models (most of which have more parameters).

Keywords: Sine-G distribution, Inverse Weibull distribution, Maximum Likelihood Estimation, Entropy, Quantile Function

1. INTRODUCTION

Statistical distributions are frequently used to investigate real-world phenomena. The theory of statistical distributions is extensively studied, as are new developments in their application. Several families of distributions have been developed to describe various real-world phenomena. In reality, this new development in distribution theory is a continuing practice. Many probability distributions proposed in the literature have a large number of parameters to make the model more versatile. However, obtaining estimates for these parameters can be challenging using numerical resources, as per some authors Marshall and Olkin [17]. Hence, it is better to create models with fewer parameters and greater flexibility for modeling actual data. To achieve this objective, a group of researchers searched for new distributions employing trigonometric functions. In the last few years, researchers have been attracted to trigonometric models due to their flexibility and mathematical tractability. Among the various trigonometric G-family members, Kumar et al. [15] have defined a new class of distribution using the sine trigonometric function and defined the sin-exponential model as its member. The cumulative distribution function (CDF) of this family is given by

$$F(x;\chi) = \sin\left\{\frac{\pi}{2}K(x;\chi)\right\} \quad x \in R, \tag{1}$$

where $K(x;\chi)$ is the CDF of any base continuous distribution. Instantaneously, Souza [24] introduced another trigonometric model based on the sine function and Gomez-Deniz and Caldern-Ojeda [9] define the arc-tan-G family of distributions using the arctangent function. Gomez-Deniz and Caldern-Ojeda [9] demonstrated the new distribution family that was used to characterize Norwegian fire insurance data.

This distribution family was introduced for an underlying Pareto distribution and a new model named the Pareto arctan distribution, and it was discovered that when compared to other well-known distributions, this distribution offers an excellent fit. Similarly, the hyperbolic cosine-F families of distributions were defined using a hyperbolic trigonometric function by Kharazmi and Saadatinik [14], and the hyperbolic cosine Rayleigh distribution was defined by Sakthivel and Rajkumar [22]. Using a similar technique as used in sin-G, the Cos-G family of distributions was introduced by Souza et al.[25] who also introduce the Cos-Weibull distribution as a member of Cos-G class. Similarly, Souza et al. [26] have introduced another sin-G class as defined by Kumar et al. [15] with bathtub-shaped or reverse-j, or increasing failure rate function, and studied the Sine inverse Weibull distribution as a particular member. The CDF of the Sin-G class of distribution is

$$F(x;\omega) = \int_0^{\frac{\pi}{2}K(x;\omega)} \cos(t)dt = \sin\left[\frac{\pi}{2}K(x;\omega)\right]; x \in R \tag{2}$$

where $K(x;\omega)$ is the CDF of any parent distribution and $\omega > 0$ is the vector of parameters of the parent distribution. Also, Mahmood et al. [16] have developed the new sin-G family and analyzed the sin-inverse Weibull model in particular. Chesneau and Jamal [6] have defined the sine Kumaraswamy-G family of distributions as having two extra parameters to this family. Muhammad et al. [19] have defined the exponentiated sine-G family and analyzed the particular distribution as an exponentiated sine-Weibull distribution. Another trigonometric function-related probability model introduced by Chaudhary et al.[3] is called Arctan generalized exponential distribution. Using the sine-G family of distributions, Isa et al. [12] have developed a new two-parameter model called the sine Burr XII distribution. Hence, we have noticed that the simple functions are associated with trigonometric distributions and are mathematically tractable (see [15], [26]). Further, the sine transformation can remarkably enhance the flexibility of $G(x)$ without any additional parameters Chesneau and Jamal [6]. Due to these pleasant features, we are motivated towards the sine transformation family. In this study, we have developed a new family of trigonometric models using the sine function, and we called it the "new class of sine-G family" (NCS-G) of distributions. The other parts of this study are organized as follows: Section 2 introduces the model development methodology as well as some key functions of the distribution family. Some general properties and parameter estimation of the NCS-G family are presented in Sections 3 and 4 respectively. In Section 5, a particular member of the NCS-G family is introduced. A detailed study and application of this model are also presented in this section. Finally, we present the conclusion in Section 6.

2. The NCS-G Family of Distribution (NCS-G FD)

Using the T-X approach proposed by Alzaatreh et al. [1], this study proposes a new family of distributions known as the NCS-G family of distributions. Let $G(x;\xi)$ be a baseline CDF of a continuous random variable X and $\xi > 0$ be a vector of associated parameters, and then the CDF $F(x;\xi)$ of the NCS-G FD is defined as

$$F(x;\xi) = \int_0^{\pi\left(\frac{G(x;\xi)}{1+G(x;\xi)}\right)} \cos(t)dt = \sin\left[\pi\frac{G(x;\xi)}{1+G(x;\xi)}\right]; x \in R. \tag{3}$$

Differentiating the CDF defined in Equation (3), the PDF $f(x;\xi)$ of the family is expressed as

$$f(x;\xi) = \pi\cos\left[\pi\frac{G(x;\xi)}{1+G(x;\xi)}\right]\frac{g(x;\xi)}{(1+G(x;\xi))^2}; x \in R. \tag{4}$$

2.1. Reliability Function

The Reliability function of NCS-G FD is given by

$$R(x;\xi) = 1 - \sin\left[\pi\frac{G(x;\xi)}{1+G(x;\xi)}\right]; x \in R. \tag{5}$$

2.2. Hazard Function

The Hazard function of NCS-G FD is given as

$$H(x;\xi)=\pi\cos\left[\pi\frac{G(x;\xi)}{1+G(x;\xi)}\right]\frac{g(x;\xi)}{(1+G(x;\xi))^2}\left[1-\sin\left(\pi\frac{G(x;\xi)}{1+G(x;\xi)}\right)\right]^{-1};x\in R \tag{6}$$

2.3. The Quantile Function (QF)

The p^{th} quantile can be calculated by solving, $Q(p)=F^{-1}(p)$. Now the QF of NCS-G FD is given by

$$Q_X(p;\xi)=G^{-1}\left[\frac{\sin^{-1}(p)}{\pi-\sin^{-1}(p)}\right], \tag{7}$$

where p has $U(0,1)$ distribution.

2.4. Random Deviate Generation

Random deviate for the NCS-G FD can be generated

$$x=G^{-1}\left[\frac{\sin^{-1}u}{\pi-\sin^{-1}u}\right] \tag{8}$$

where $u\in U(0,1)$ distribution.

2.5. Skewness and Kurtosis

Bowley's measure of skewness was defined by Kenney and Keeping [13] as,

$$S_k(B)=\frac{Q(3/4;\xi)+Q(1/4;\xi)-2Q(1/2;\xi)}{Q(3/4;\xi)-Q(1/4;\xi)} \tag{9}$$

and the coefficient of Moor's kurtosis defined by Moors [18] is given by

$$K_u(M)=\frac{Q(0.875;\xi)-Q(0.625;\xi)+Q(0.375;\xi)-Q(0.125;\xi)}{Q(3/4;\xi)-Q(1/4;\xi)}. \tag{10}$$

3. General Properties of NCS-G FD

3.1. Linear form

Using the following Taylor series expansions, we can express the density function of NCS-G FD in a linear form as

$$\cos x=\sum_{n=0}^{\infty}(-1)^{2n}\frac{x^{2n}}{(2n)!}=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}-\cdots;-\infty<x<\infty. \tag{11}$$

$$(1+x)^c=\sum_{n=0}^{\infty}\binom{c}{n}x^n=1+\frac{c}{1!}x+\frac{c(c-1)}{2!}x^2+\frac{c(c-1)(c-2)}{3!}x^3+\cdots;|x|<1. \tag{12}$$

The PDF of NCS-G FD is

$$f(x;\xi)=g(x;\xi)\sum_{i=0}^{\infty}\frac{\pi^{2i+1}(-1)^{2i}}{(2i)!}(1+G(x;\xi))^{2(i-1)}(G(x;\xi))^{2i}. \tag{13}$$

Further expanding Equation (13) using generalized binomial series expansion. The expression for $f(x;\xi)$ becomes

$$f(x;\xi)=g(x;\xi)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}T_{ij}\{G(x;\xi)\}^{2i+j};x\in R \tag{14}$$

here

$$T_{ij}=\frac{\pi^{2i+1}(-1)^{2i}}{(2i)!}\binom{2(i-1)}{j}. \tag{15}$$

3.2. Moments

The r^{th} order non-central moment ($\mu_r^{'}$) for the NCS-G FD is

$$\begin{aligned}\mu_r^{'} &= E(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} T_{ij} \int_{-\infty}^{\infty} x^r \left(G(x;\xi)\right)^{2i+j} g(x;\xi)dx\end{aligned} \tag{16}$$

Further moments can also be calculated using the quantile function for more detail (see Balakrishnan and Cohen [2]). Let $G(x;\xi) = p \Rightarrow g(x;\xi)dx = dp; 0 \leqslant p \leqslant 1$, then r^{th} moment can be computed using

$$\mu_r^{'} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} T_{ij} \int_0^1 p^{2i+j} Q_G^r(p)dp; \;\; 0 < p < 1. \tag{17}$$

where $Q_G(p)$ is the QF of any distribution.

3.3. Moment Generating Function

The MGF ($M_X(t)$) for the NCS-G FD is

$$\begin{aligned}M_X(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!}\mu_r^{'} \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{t^k}{k!} T_{ij} \int_{-\infty}^{\infty} x^r g(x;\xi) \left(G(x;\xi)\right)^{2i+j} dx.\end{aligned} \tag{18}$$

Using the quantile function, MGF can be expressed as

$$M_X(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{t^k}{k!} T_{ij} \int_0^1 p^{2i+j} Q_G^r(p)\, dp, \quad 0 < p < 1. \tag{19}$$

where $Q_G(p)$ is the QF of any distribution.

3.4. Incomplete Moment

The incomplete moment can be defined as $M_r(y) = \int_0^y x^r f(x)dx$. Therefore incomplete moment for NCS-G FD is given by

$$M_r(y) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \int_{-\infty}^{y} T_{ij} x^r g(x;\xi) \left\{G(x;\xi)\right\}^{2i+j} dx \tag{20}$$

Alternately, $M_r(y)$ may be expressed in terms of QF as

$$M_r(y) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} T_{ij} \int_0^{G(y)} p^{2i+j} Q_G^r(p)dp; \;\; 0 < p < 1 \tag{21}$$

3.5. Mean Residual Life (MRL)

The MRL of the random variable X is defined as

$$\bar{M}(y) = \frac{1}{F(y)} \left[\mu - \int_{-\infty}^{y} x f(x)dx\right] - y. \tag{22}$$

Therefore, MRL for NCS-G FD is given by

$$\bar{M}(y) = \frac{1}{F(y)} \left[\mu - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} T_{ij} \int_{-\infty}^{y} x g(x;\xi) \{G(x;\xi)\}^{2i+j} dx \right] - y. \tag{23}$$

Alternately, $\bar{M}(y)$ can be expressed in terms of QF as

$$\bar{M}(y) = \frac{1}{F(y)} \left[\mu - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} T_{ij} \int_{0}^{G(y)} p^{2i+j} Q_G(p) dp \right] - y. \tag{24}$$

3.6. Inequality Measure

In several fields, including insurance, econometrics, and reliability, we can employ Lorenz and Bonferroni curves to measures such as income, poverty, etc.

i) Lorenz Curve

The function of the Lorenz curve is written as hence Lorenz curve for NCS-G FD is given by

$$L_{F(y)} = \frac{1}{\mu} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} T_{ij} \int_{-\infty}^{y} x g(x;\xi) \left(G(x;\xi)\right)^{2i+j} dx. \tag{25}$$

Alternatively, it can be written in terms of QF as

$$L_{F(y)} = \frac{1}{\mu} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} T_{ij} \int_{-\infty}^{G(y)} p^{2i+j} Q_G(p) dp. \tag{26}$$

ii) Boneferroni Curve

The Boneferroni curve can be calculated using $B_{F(y)} = \frac{L_{F(y)}}{F(y)}$. From Equation (25), the Boneferroni curve for the NCS-G FD is calculated as

$$B_{F(y)} = \frac{1}{\mu F(y)} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} T_{ij} \int_{-\infty}^{y} x g(x;\xi) \left(G(x;\xi)\right)^{2i+j} dx \tag{27}$$

3.7. Entropy

Entropy quantifies the uncertainty or variation of a random variable. Its application spans numerous disciplines, including econometrics, probability theory, engineering, and life sciences in general. There are several types of entropy, some of which are as follows:

i) Renyi's Entropy

Entropy is used as a measure of uncertainty or variation in a random variable in many disciplines, including engineering, econometrics, insurance, etc. Renyi [20] introduced entropy measures, which can be used to calculate the variability of uncertainty.

$$R_{\rho}(X) = \frac{1}{1-\rho} \log \int_{-\infty}^{\infty} \{f(x)\}^{\rho} dx \tag{28}$$

and $\rho \neq 1$. The PDF of NCS-G FD $[f(x,\xi)]^{\rho}$ can be defined in the form of

$$[f(x;\xi)]^{\rho} = \pi^{\rho} \left(g(x;\xi)\right)^{\rho} \left[\cos\left(\pi \frac{G(x;\xi)}{1+G(x;\xi)} \right) \right]^{\rho} \left(1+G(x;\xi)\right)^{-2\rho} \tag{29}$$

By considering the Taylor series of the function

$$\left[\cos\left(\pi \frac{G(x;\xi)}{1+G(x;\xi)} \right) \right]^{\rho} \tag{30}$$

at the point s=1/4, we can write

$$[\cos(\pi s)]^{\rho} = \sum_{k=0}^{\infty}\sum_{r=0}^{k} a_k \binom{k}{r} (-1)^{k-r} \left(\frac{1}{4}\right)^{k-r} s^r \tag{31}$$

where $a_k = \frac{1}{k!}\left[\{\cos(\pi s)\}^{\rho}\right]^{(k)}\Big|_{s=\frac{1}{4}}$ using this relation Equation (29) becomes

$$[f(x;\xi)]^{\rho} = \pi^{\rho}\,(g(x;\xi))^{\rho} \sum_{k=0}^{\infty}\sum_{r=0}^{k} a_k \binom{k}{r} (-1)^{k-r} \left(\frac{1}{4}\right)^{k-r} (G(x;\xi))^{r} (1+G(x;\xi))^{-(2\rho+r)} \tag{32}$$

Further expanding Equation (32) using generalized binomial series expansion. The expression for $[f(x;\xi)]^{\rho}$ becomes

$$[f(x;\xi)]^{\rho} = \pi^{\rho} \sum_{k=0}^{\infty}\sum_{r=0}^{k}\sum_{m=0}^{\infty} (-1)^{m+k-r} a_k \binom{k}{r} \left(\frac{1}{4}\right)^{k-r} \binom{(2\rho+r)+m-1}{m} (G(x;\xi))^{r+m} (g(x;\xi))^{\rho} \tag{33}$$

Substituting $[f(x,\xi)]^{\rho}$ into the expression defining Equation (28), Renyi's entropy for NCS-G FD is given by

$$R_{\rho}(X) = \frac{1}{1-\rho}\log\left[\sum_{k=0}^{\infty}\sum_{r=0}^{k}\sum_{m=0}^{\infty} Z_{krm} \int_{-\infty}^{\infty} (g(x;\xi))^{\rho}\,(G(x;\xi))^{r+m}\,dx\right] \tag{34}$$

where $Z_{krm} = (-1)^{m+k-r}\,\pi^{\rho} a_k \binom{k}{r} \left(\frac{1}{4}\right)^{k-r} \binom{(2\rho+r)+m-1}{m}$.

ii) q-Entropy

The q-entropy is given by

$$H(\rho) = \frac{1}{1-\rho}\log\left[1 - \int_{-\infty}^{\infty} \{f(x)\}^{\rho}\,dx\right]. \tag{35}$$

$\rho > 0$ and $\rho \neq 1$. Substituting $[f(x;\xi)]^{\rho}$ from Equation (32) into the expression for $H(\rho)$, the q-Entropy for NCS-G FD is given by

$$H(\rho) = \frac{1}{1-\rho}\log\left[1 - \sum_{k=0}^{\infty}\sum_{r=0}^{k}\sum_{m=0}^{\infty} Z_{krm} \int_{-\infty}^{\infty} (g(x;\xi))^{\rho}\,(G(x,\xi))^{r+m}\,dx\right] \tag{36}$$

where $\rho > 0$ and $\rho \neq 1$.

iii) Shannon's Entropy

When $\rho \uparrow 1$, Shannon's entropy for a random variable X with PDF $f(x)$ is a particular case of Renyi's entropy. Shannon entropy is defined as $\eta_X = E(-\log f(x))$. For the NCS-G FD is given by

$$\eta_X = E\left[-\log\left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} T_{ij}\,g(x;\xi)\,(G(x;\xi))^{2i+j}\right\}\right]. \tag{37}$$

4. Estimation Method NCS-G FD

4.1. Maximum Likelihood Estimation (MLE)

In this section, the parameters of the NCS-G FD are estimated using the MLE method. Given a random sample $x_1,...,x_n$ of size n with parameters vector ξ from the NCS-G FD, we can compute the MLEs. Let $u = \xi^T$ be $(p\times 1)$ parameter vectors, the log density and total log-likelihood function, respectively, are given by

$$L(x;\xi) = \log\pi + \log\left[\cos\left\{\pi\frac{G(x;\xi)}{1+G(x;\xi)}\right\}\right] - 2\log(1+G(x;\xi)) + \log g(x;\xi), \tag{38}$$

and

$$l(\underline{x},\xi) = n\log\pi + \sum_{i=1}^{n}\log\left[\cos\left\{\pi\frac{G(x_i;\xi)}{1+G(x_i;\xi)}\right\}\right] - 2\sum_{i=1}^{n}\log\left(1+G(x_i;\xi)\right) + \sum_{i=1}^{n}\log g(x_i;\xi). \tag{39}$$

Partially differentiating the Equation (39) with respect to ξ gives the score function's components of the $V(u) = \left(\frac{\partial l}{\partial \xi}\right)^T$ as follows

$$\frac{\partial l}{\partial \xi} = -\pi\sum_{i=1}^{n}\tan\left\{\pi\frac{G(x_i;\xi)}{1+G(x_i;\xi)}\right\}\frac{G_k^{'}(x_i;\xi)}{\left(1+G(x_i;\xi)\right)^2} - 2\sum_{i=1}^{n}\frac{G_k^{'}(x_i;\xi)}{\left(1+G(x_i;\xi)\right)} + \sum_{i=1}^{n}\frac{g_k^{'}(x_i;\xi)}{g(x_i;\xi)},$$

where $g_k^{'}(x_i;\xi) = \frac{dg(x_i;\xi)}{d\xi}$, $g_k^{'}(x_i;\xi) = \frac{d^2g(x_i;\xi)}{d^2\xi}$, $G_k^{'}(x_i;\xi) = \frac{dG(x_i;\xi)}{d\xi}$ and $G_k^{'}(x_i;\xi) = \frac{d^2G(x_i;\xi)}{d^2\xi}$.

4.2. Method of Least Square Estimation (LSE)

Another method of estimation was introduced by Swain et al. [27] named the ordinary LSE and weighted LSE to estimate the distribution parameters. Consider $x_{(1)},...,x_{(n)}$ be order statistics of the random sample of size n from $F(x,\xi)$. The LSE for the NCS-G FD can be obtained by minimizing

$$K(X;\xi) = \sum_{i=1}^{n}\left[F(x_{(i)};\xi) - \frac{i}{n+1}\right]^2. \tag{40}$$

with respect to ξ. The least-square estimates for the NCS-G FD also become

$$K(X;\xi) = \sum_{i=1}^{n}\left[\sin\left[\pi\frac{G(x_{(i)};\xi)}{1+G(x_{(i)};\xi)}\right] - \frac{i}{n+1}\right]^2. \tag{41}$$

Now differentiating Equation (41) with respect to ξ we get

$$\frac{\partial K}{\partial \xi} = 2\pi\sum_{i=1}^{n}\left[\sin\left[\pi\frac{G(x_{(i)};\xi)}{1+G(x_{(i)};\xi)}\right] - \frac{i}{n+1}\right]\cos\left[\pi\frac{G(x_{(i)};\xi)}{1+G(x_{(i)};\xi)}\right]\frac{G_k^{'}(x_{(i)};\xi)}{\left(1+G(x_{(i)};\xi)\right)^2}. \tag{42}$$

where $G_k^{'}(x_i;\xi) = \frac{dG(x_{(i)};\xi)}{d\xi}$. By solving $\frac{dK}{d\xi} = 0$, we will get the LSEs.

4.3. Cramer-von Mises Minimum Distance Estimator (CVME)

Cramer-von Mises estimators (CVMEs) are specific types of statistical estimators that minimize the difference between the estimated and the empirical CDF. These estimators are considered to have a lower bias compared to other minimum distance estimators. In the context of estimating parameters for the NCS-G FD distribution, CVMEs can be used to obtain more accurate estimates by minimizing

$$C(X;\xi) = \frac{1}{12n} + \sum_{i=1}^{n}\left[F(x_{(i)};\xi) - \frac{2i-1}{2n}\right]^2. \tag{43}$$

with respect to ξ . The CVMEs for the NCS-G FD also become

$$C(X;\xi) = \sum_{i=1}^{n}\left[\sin\left[\pi\frac{G(x_{(i)};\xi)}{1+G(x_{(i)};\xi)}\right] - \frac{2i-1}{2n}\right]^2 \tag{44}$$

Now differentiating Equation (44) with respect to ξ we get

$$\frac{\partial C}{\partial \xi} = 2\pi\sum_{i=1}^{n}\left[\sin\left[\pi\frac{G(x_{(i)};\xi)}{1+G(x_{(i)};\xi)}\right] - \frac{2i-1}{2n}\right]\cos\left[\pi\frac{G(x_{(i)};\xi)}{1+G(x_{(i)};\xi)}\right]\frac{G_k^{'}(x_{(i)};\xi)}{\left(1+G(x_{(i)};\xi)\right)^2}, \tag{45}$$

where $G_k^{'}(x_{(i)};\xi) = \frac{dG(x_{(i)};\xi)}{d\xi}$. By solving $\frac{dC}{d\xi} = 0$, we will get the CVMEs.

5. SPECIAL MEMBER OF NCS-G FD

Generalization of several distributions can be made using the NCS-G FD. Here we have considered the inverse Weibull (IW) distribution as a parent distribution to introduce a special member.

5.1. A New Class Sin Inverse Weibull (NCS-IW) Distribution

The CDF and PDF of the IW distribution are respectively given by

$$G(x;\delta,\theta) = exp(-\theta x^{-\delta}); x>0, \delta>0, \theta>0$$

and

$$g(x;\delta,\theta) = \delta\theta x^{-(\delta+1)} exp(-\theta x^{-\delta}).$$

The CDF and PDF of the NCS-IW distribution are given by

$$F(x;\theta,\delta) = \sin\left[\pi \frac{\exp(-\theta x^{-\delta})}{1+\exp(-\theta x^{-\delta})}\right]; x>0. \tag{46}$$

$$f(x;\theta,\delta) = \pi\theta\delta x^{-(\delta+1)} \cos\left[\pi \frac{\exp(-\theta x^{-\delta})}{1+\exp(-\theta x^{-\delta})}\right] \frac{\exp(-\theta x^{-\delta})}{\left(1+\exp(-\theta x^{-\delta})\right)^2}; x>0. \tag{47}$$

The reliability and hazard functions, respectively, are given by

$$R(x;\theta,\delta) = 1-\sin\left[\pi \frac{\exp(-\theta x^{-\delta})}{1+\exp(-\theta x^{-\delta})}\right]; x>0. \tag{48}$$

and

$$h(x;\theta,\delta) = \pi\theta\delta x^{-(\delta+1)} \frac{\exp(-\theta x^{-\delta})}{\left(1+\exp(-\theta x^{-\delta})\right)^2} \cos\left[\pi \frac{\exp(-\theta x^{-\delta})}{1+\exp(-\theta x^{-\delta})}\right] \left[1-\sin\left(\pi \frac{\exp(-\theta x^{-\delta})}{1+\exp(-\theta x^{-\delta})}\right)\right]^{-1}; x>0. \tag{49}$$

The possible shapes of PDF and HRF of NCS-IW distribution are shown in Figure (1) and it is observed that HRF can have reverse-j, or inverted bathtub or increasing hazard function. The quantile function and random deviate generation for the NCS-IW distribution, respectively, are given by

$$Q_X(p) = \left[-\frac{1}{\theta}\log\left(\frac{\sin^{-1}p}{\pi-\sin^{-1}p}\right)\right]^{-\frac{1}{\delta}}. \tag{50}$$

and

$$x = \left[-\frac{1}{\theta}\log\left(\frac{\sin^{-1}u}{\pi-\sin^{-1}u}\right)\right]^{-\frac{1}{\delta}}. \tag{51}$$

5.2. Linear Expansion

Using Equation (14), Equation (47) can be expressed in linear form as

$$f(x;\xi) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} B_{ij} x^{-(\delta+1)} \exp\left\{-(2i+j+1)\theta x^{-\delta}\right\} \tag{52}$$

where $B_{ij} = \frac{(-1)^{2i}\theta\delta\pi^{2i+1}}{(2i)!}\binom{2(i-1)}{j}$.

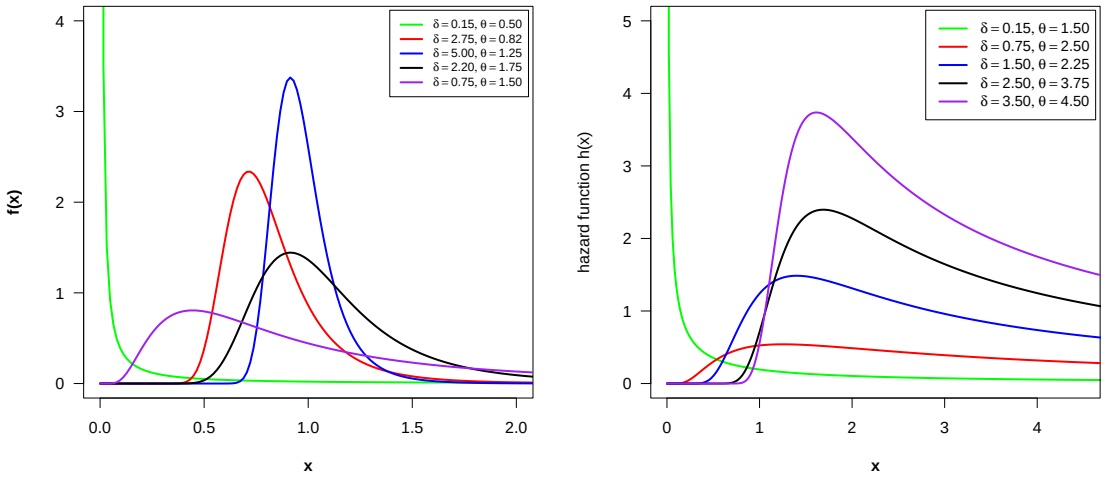


Figure 1: *Shapes of PDF and HRF of NCS-IW distribution*

5.3. Moments

Using the PDF defined in Equation (52), the r^{th} order non-central moment ($\mu_r^{'}$) for the NCS-IW distribution can be presented as

$$\mu_r^{'} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} B_{ij}^{*} \frac{\Gamma\left(\frac{\delta-r}{\delta}\right)}{[\theta\{(2i+j)+1\}]^{\frac{\delta-r}{\delta}}}; \quad \forall \delta > r, \tag{53}$$

where $B_{ij}^{*} = \frac{(-1)^{2i}\theta\pi^{2i+1}}{(2i)!}\binom{2(i-1)}{j}$ and $\Gamma(.)$ is the gamma function.

5.4. Moment Generating Function (MGF)

The MGF ($M_X(t)$) for the NCS-IW distribution is

$$M_X(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \frac{t^k B_{ij}^{*}}{k!} \frac{\Gamma\left(\frac{\delta-r}{\delta}\right)}{[\theta\{(2i+j)+1\}]^{\frac{\delta-r}{\delta}}}; \quad \forall \delta > r. \tag{54}$$

5.5. Incomplete moment

The incomplete moment for NCS-IW distribution is presented as

$$\begin{aligned} M_r(y) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} B_{ij} \int_0^y x^{r-(\delta+1)} \exp\left\{-(2i+j+1)\theta x^{-\delta}\right\} dx \\ &= \frac{1}{\delta}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} B_{ij} \frac{\gamma\left(\frac{\delta-r}{\delta}, (2i+j+1)\theta y^{-\delta}\right)}{\{(2i+j+1)\theta\}^{\frac{\delta-r}{\delta}}}, \end{aligned} \tag{55}$$

where $\gamma(.)$ incomplete gamma function.

5.6. Mean Residual Life

The MRL for NCS-IW distribution is given by

$$\begin{aligned}\bar{M}(y) &= \frac{1}{F(y)}\left[\mu - \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} B_{ij}\int_0^y x^{-\delta}\exp\left\{-(2i+j+1)\theta x^{-\delta}\right\}dx\right] - y \\ &= \frac{1}{F(y)}\left[\mu - \frac{1}{\delta}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} B_{ij}\frac{\gamma\left(\frac{\delta-1}{\delta},(2i+j+1)\theta y^{-\delta}\right)}{\{(2i+j+1)\theta\}^{\frac{\delta-1}{\delta}}}\right] - y,\end{aligned} \tag{56}$$

where $\gamma(.)$ is the incomplete gamma function. Using the Equations (9 and 10) for NCS-IW distribution, we have plotted the graphs of skewness and kurtosis in Figure (2) for different values of the parameters δ and θ.

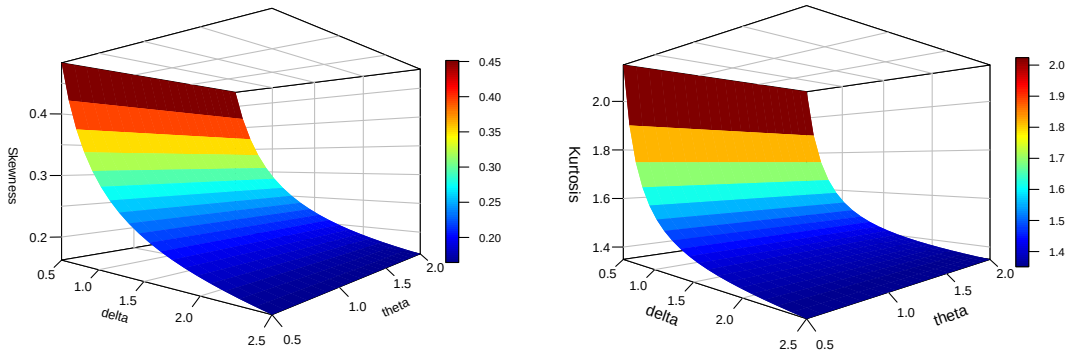


Figure 2: *Skewness and Kurtosis plots of NCS-IW distribution.*

5.7. Entropy

i) Renyi's Entropy
Renyi's entropy for NCS-IW distribution is given by

$$\begin{aligned}R_\rho(X) &= \frac{1}{1-\rho}\log\left[\sum_{k=0}^{\infty}\sum_{r=0}^{k}\sum_{m=0}^{\infty} Z_{krm}(\delta\theta)^\rho\int_0^\infty x^{-\rho(\delta+1)}\exp(-(r+m+\rho)\theta x^{-\delta})dx\right] \\ &= \frac{1}{1-\rho}\log\left[\sum_{k=0}^{\infty}\sum_{r=0}^{k}\sum_{m=0}^{\infty} Z_{krm}\frac{(\delta\theta)^\rho}{\delta}\frac{\Gamma\left\{\frac{(\rho-1)(\delta+1)}{\delta}+1\right\}}{\{(r+m+\rho)\theta\}^{\frac{(\rho-1)(\delta+1)}{\delta}+1}}\right],\end{aligned} \tag{57}$$

where $Z_{krm} = (-1)^{m+k-r}\pi^\rho a_k\binom{k}{r}\left(\frac{1}{4}\right)^{k-r}\binom{(2\rho+r)+m-1}{m}$.

ii) q-Entropy

The q-Entropy for NCS-IW distribution is given by

$$\begin{aligned} H(\rho) &= \frac{1}{1-\rho}\log\left[1 - Z_{krm}\,(\delta\theta)^{\rho}\int_0^{\infty} x^{-\rho(\delta+1)}\exp(-(r+m+\rho)\theta x^{-\delta})dx\right] \\ &= \frac{1}{1-\rho}\log\left[1 - Z_{krm}\frac{(\delta\theta)^{\rho}}{\delta}\frac{\Gamma\left\{\frac{(\rho-1)(\delta+1)}{\delta}+1\right\}}{\{(r+m+\rho)\theta\}^{\frac{(\rho-1)(\delta+1)}{\delta}+1}}\right] \end{aligned} \tag{58}$$

where $\rho > 0$ and $\rho \neq 1$. where $Z_{krm} = (-1)^{m+k-r}\,\pi^{\rho} a_k \binom{k}{r}\left(\frac{1}{4}\right)^{k-r}\binom{(2\rho+r)+m-1}{m}$.

iii) Shannon's Entropy

The Shannon entropy for the NCS-IW distribution is given by

$$\eta_X = E\left[-\log\left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\pi^{2i+1}(-1)^{2i}}{(2i)!}\binom{2(i-1)}{j}x^{-(\delta+1)}\exp(-\theta(2i+j+1)x^{-\delta})\right\}\right]. \tag{59}$$

5.8. Inequality Measure

i) Lorentz Curve

The Lorenz curve for NCS-IW distribution is given by

$$\begin{aligned} L_{F(y)} &= \frac{\delta\theta}{\mu}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}B_{ij}\int_0^{y} x^{-\delta}\exp(-\theta(2i+j+1)x^{-\delta})dx \\ &= \frac{\theta}{\mu}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}B_{ij}\frac{\gamma\left(\frac{\delta-1}{\delta},(2i+j+1)\theta y^{-\delta}\right)}{\{(2i+j+1)\theta\}^{\frac{\delta-1}{\delta}}}. \end{aligned} \tag{60}$$

where $\gamma(.)$ is the incomplete gamma function.

ii) Boneferroni Curve

The Boneferroni curve for the NCS-IW distribution is given by

$$\begin{aligned} B_{F(y)} &= \frac{1}{\mu F(y)}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}B_{ij}\int_0^{y} x^{-\delta}\exp(-\theta(2i+j+1)x^{-\delta})dx \\ &= \frac{1}{\delta\mu F(y)}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}B_{ij}\frac{\gamma\left(\frac{\delta-1}{\delta},(2i+j+1)\theta y^{-\delta}\right)}{\{(2i+j+1)\theta\}^{\frac{\delta-1}{\delta}}}. \end{aligned} \tag{61}$$

where $\gamma(.)$ is the incomplete gamma function.

5.9. Estimation MLE for NCS-IW distribution

We now investigate the MLE for estimating the parameters of the NCS-IW model. As a result, we intend to compute MLEs for the parameters δ and θ. Let $X = (x_1,...,x_n)^T$ be a vector of size n of independent random variables from the NCS-IW distribution. Then, the log-likelihood is given by

$$l(x;\delta,\theta) = n\log(\pi\theta\delta) - (\delta+1)\sum_{i=1}^{n}\log x_i + \sum_{i=1}^{n}\log\cos\left[\pi\frac{\exp(-\theta x_i^{-\delta})}{1+\exp(-\theta x_i^{-\delta})}\right] - 2\sum_{i=1}^{n}\log\left(1+\exp(-\theta x_i^{-\delta})\right) - \theta\sum_{i=1}^{n}x_i^{-\delta} \tag{62}$$

Partially differentiating the Equation (62) with respect to δ and θ gives the score function's components of $V(u) = \left(\frac{\partial l}{\partial \delta}, \frac{\partial l}{\partial \theta}\right)^T$ as,

$$\frac{\partial l}{\partial \delta} = \frac{n}{\delta} - \sum_{i=1}^{n} \log x_i + \pi\theta \sum_{i=1}^{n} \frac{x_i^{-\delta} \log(x_i) \exp(-\theta x_i^{-\delta})}{\left(1+\exp(-\theta x_i^{-\delta})\right)^2} \tan\left[\pi \frac{\exp(-\theta x_i^{-\delta})}{1+\exp(-\theta x_i^{-\delta})}\right] + 2\theta \sum_{i=1}^{n} \frac{x_i^{-\delta} \log(x_i) \exp(-\theta x_i^{-\delta})}{\left(1+\exp(-\theta x_i^{-\delta})\right)} + \theta \sum_{i=1}^{n} x_i^{-\delta} \log(x_i) \tag{63}$$

and

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta} - \pi \sum_{i=1}^{n} \frac{x_i^{-\delta} exp(-\theta x_i^{-\delta})}{\left(1+\exp(-\theta x_i^{-\delta})\right)^2} \tan\left[\pi \frac{\exp(-\theta x_i^{-\delta})}{1+\exp(-\theta x_i^{-\delta})}\right] - 2\sum_{i=1}^{n} \frac{x_i^{-\delta} exp(-\theta x_i^{-\delta})}{\left(1+\exp(-\theta x_i^{-\delta})\right)} - \sum_{i=1}^{n} x_i^{-\delta} \tag{64}$$

The MLEs of δ and θ are obtained by maximizing $l(x;\delta,\theta)$ in δ and θ , which can be done by solving simultaneously the equations $\frac{\partial l}{\partial \delta} = 0$ and $\frac{\partial l}{\partial \theta} = 0$.

5.10. Simulation

Using the maxLik R package introduced by Henningsen and Toomet [10], we generated samples from the quantile function defined in Equation (50) for various parameter combinations of the NCS-IW distribution and calculated the MLEs for each sample using the `maxLik()` function with the BFGS algorithm. This allows us to test parameter estimation problems such as the sharpness or flatness of the likelihood function, as well as estimate the size and direction (underestimate or overestimate) of the MLEs bias. Sample sizes of 20, 30, 40, 50, and 75 are used in the simulation. The procedure is repeated 10,000 times, and the average estimate value, bias, and mean square error (MSE) are calculated. The experiment is summarized in Table 1, which shows the average estimate, bias, and MSEs for each parameter. As can be seen, the MLE method consistently overestimates the parameter δ and underestimates the parameter θ, but as sample size increases, MLEs gradually approach the actual values of δ and θ.

Table 1: *The estimated values, Biases, and MSEs based on 10000 simulations of NCS-IW distribution.*

	Actual values		MLEs		Bias		MSEs	
n	delta	theta	$\hat{\delta}$	$\hat{\theta}$	$\hat{\delta}$	$\hat{\theta}$	$\hat{\delta}$	$\hat{\theta}$
	0.25	0.50	0.268	0.4796	0.018	-0.0204	0.0029	0.0219
20	0.50	0.75	0.5372	0.7263	0.0372	-0.0237	0.0115	0.0318
	0.75	1.00	0.805	0.9816	0.055	-0.0184	0.026	0.0415
	0.25	0.50	0.2621	0.4869	0.0121	-0.0131	0.0016	0.0142
30	0.50	0.75	0.5241	0.7343	0.0241	-0.0157	0.0066	0.0215
	0.75	1.00	0.7874	0.9848	0.0374	-0.0152	0.0154	0.0277
	0.25	0.50	0.2593	0.4889	0.0093	-0.0111	0.0012	0.0109
40	0.50	0.75	0.5175	0.7377	0.0175	-0.0123	0.0046	0.0157
	0.75	11.00	0.7768	0.9911	0.0268	-0.0089	0.0103	0.0201
	0.25	0.50	0.257	0.4919	0.007	-0.0081	0.0009	0.0087
50	0.50	0.75	0.5146	0.7398	0.0146	-0.0102	0.0037	0.0129
	0.75	1.00	0.7696	0.992	0.0196	-0.008	0.0078	0.0159
	0.25	0.50	0.2546	0.4943	0.0046	-0.0057	0.0006	0.0059
75	0.50	0.75	0.5089	0.7444	0.0089	-0.0056	0.0022	0.0084
	0.75	1.00	0.7646	0.994	0.0146	-0.006	0.0052	0.0105

5.11. Application

Employing two real data sets, we exhibit the application of the NCS-IW distribution in this section. The data sets employed for the application of the suggested distribution are given as follows
i) Data set
Data set 1 (cancer data):
The data set contains information on the survival times of 44 patients. These patients who received radiotherapy have head and neck cancer, and this data set was reported by Efron [8].
"12.20, 23.56, 23.74, 25.87, 31.98, 37, 41.35, 47.38, 55.46, 58.36, 63.47, 68.46, 78.26, 74.47, 81.43, 84, 92, 94, 110, 112, 119, 127, 130, 133, 140, 146, 155, 159, 173, 179, 194, 195, 209, 249, 281, 319, 339, 432, 469, 519, 633, 725, 817, 1776".
Data set 2 (relief time data):
The real data set is considered from Clark and Gross [7], which provides the relief times of 20 patients receiving an analgesic. The data are:
"1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2, 1.4, 3, 1.7, 2.3, 1.6, 2.0".
ii) Model Analysis
We calculate some well-known goodness-of-fit statistics to analyze data sets 1 and 2 and the fitted models are evaluated using the log-likelihood value (-2logL), Akaike information criterion (AIC), Hannan-Quinn information criterion (HQIC), Anderson-Darling (AD), Kolmogrov-Smirnov (KS) with p-values, and Cram'er-von Mises (CVM) for more detail (see Chen and Balakrishnan [5]). All the essential computations are carried out in R-software. For the comparison of fitting capability, we have selected some models such as inverse Weibull (IW), transformed sine Weibull (TSW) Sakthivel and Rajkumar [23], arctan generalized exponential (AGE) Chaudhary et al. [3], arctan Lomax (ALomx) Chaudhary and Kumar [4], arcsine exponential (ASE) Rahman [21], and arcsine exponentiated Weibull (ASEW) He et al. [11]. The PDFs of candidate models are as follows

$$f_{IW}(x;\delta,\theta) = \delta\theta x^{-\delta-1}e^{-\theta x^{-\delta}}, x,\delta,\theta > 0.$$

$$f_{TSW}(x;\alpha,\beta,\lambda) = \frac{\pi}{2}\alpha\beta x^{\beta-1}e^{-\alpha x^{\beta}}\left[\frac{\pi\lambda}{2}\left(1-e^{-\alpha x^{\beta}}\right)\cos\left(\frac{\pi}{2}e^{-\alpha x^{\beta}}\right)-(1-\lambda)\sin\left(\frac{\pi}{2}e^{-\alpha x^{\beta}}\right)\right], x,\alpha,\beta,\lambda > 0.$$

$$f_{AGE}(x;\alpha,\beta,\lambda) = \frac{\alpha\beta\lambda e^{-\lambda x}\left(1-e^{-\lambda x}\right)^{\beta-1}}{arc\tan(\alpha)\left[1+\left\{\alpha\left(1-\left(1-e^{-\lambda x}\right)^{\beta}\right)\right\}^{2}\right]}; x,\alpha,\beta,\lambda > 0.$$

$$f_{ALomx}(x;\alpha,\beta,\lambda) = \frac{\alpha\beta\lambda}{arc\tan(\alpha)}\frac{(1+\beta x)^{-\lambda-1}}{\left[1+\left\{\alpha(1+\beta x)^{-\lambda}\right\}^{2}\right]}; x,\alpha,\beta,\lambda > 0.$$

$$f_{ASE}(x;\alpha) = \frac{\sqrt{e^{-x/\alpha}}}{\pi\alpha\sqrt{1-e^{-x/\alpha}}}; x,\alpha > 0.$$

$$f_{ASEW}(x;\alpha,\beta,\lambda) = \frac{2}{\pi}\alpha\beta\lambda x^{\alpha-1}e^{-\lambda x^{\alpha}}\frac{\left(1-e^{-\lambda x^{\alpha}}\right)^{\beta-1}}{\sqrt{1-\left(1-e^{-\lambda x^{\alpha}}\right)^{2\beta}}}; x,\alpha,\beta,\lambda > 0.$$

In Tables 2 and 3, we have presented the estimated values of the parameters and their associated standard error (SE in parentheses) of the models under study using the MLE method for cancer and relief time data. Similarly, in Tables 4 and 5, we have presented the model selection and goodness of fit statistics like log-likelihood, HQIC, AIC, KS, AD, and CVM for both data sets. It has been observed that the suggested model has the least statistics as compared to IW, AGE, ALomx, ASE, ASEW, and TSW. Hence NCS-IW is more flexible (even four trigonometric distributions having three parameters) and provides a good fit. Also, we have displayed the graphical illustrations of the fitted models in Figures 5 and 6. These figures also verified that the NCS-IW model can perform well as compared to candidate models

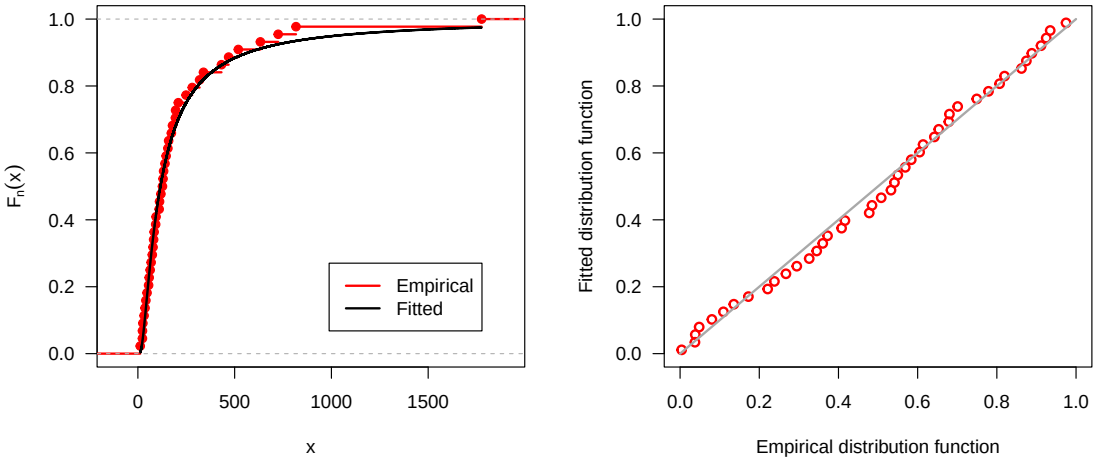


Figure 3: *KS and P-P plots (data-I).*

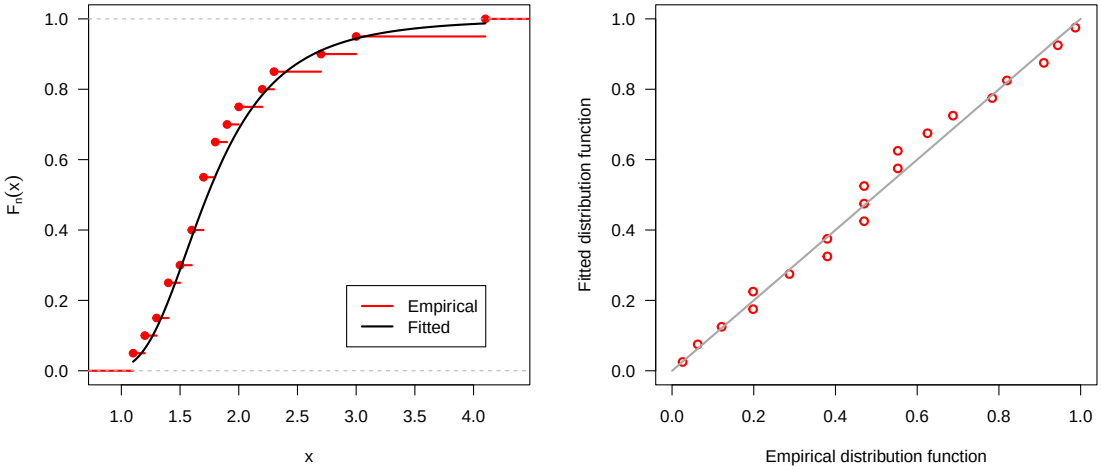


Figure 4: *KS and P-P plots (data-II).*

Table 2: *MLEs with SE (in parentheses) (data-I).*

Distribution	Parameter(SE)	Parameter(SE)	Parameter(SE)
NCS-IW(δ,θ)	0.6317(0.0508)	32.4048(6.9827)	–
IW(δ,θ)	0.9985(0.0393)	75.557(4.4651)	–
AGE(α,β,λ)	0.0179(0.5939)	1.0688(0.2216)	0.0047(9.00E-04)
ALomx(α,β,λ)	27.525(5.8997)	0.0640(0.0335)	1.5273(0.2833)
ASE(α)	341.8104(4.1943)	–	–
ASEW(α,β,λ)	0.4578(0.0133)	13.1876(4.5521)	0.4031(0.0919)
TSW(α,β,λ)	0.0039(0.0025)	0.9742(0.1073)	0.1327(0.1312)

Table 3: *MLEs with SE (in parentheses) (data-II).*

Distribution	Parameter(SE)	Parameter(SE)	Parameter(SE)
NCS-IW(δ,θ)	2.3934(0.4249)	6.0185(1.3910)	–
IW(δ,θ)	4.0175(0.706)	6.0224(2.0083)	–
AGE(α,β,λ)	29.0366(6.6483)	2.9010(3.1180)	2.5293(0.567)
ALomx(α,β,λ)	187.9197(5.1477)	0.2891(0.3043)	12.8568(11.0058)
ASE(α)	127.8946(4.8432)	–	–
ASEW(α,β,λ)	1.0488(0.1284)	104.561(19.0921)	3.1656(0.1303)
TSW(α,β,λ)	0.0811(0.0398)	2.9331(0.4532)	0.1297(0.1388)

Table 4: *Some selection criteria and goodness-of-fit statistics (data-I).*

Distribution	**-2logL**	**AIC**	**HQIC**	KS(p-value)	CVM(p-value)	AD(p-value)
NCS-IW	556.1646	560.1646	561.4879	0.0706(0.9698)	0.0302(0.9768)	0.2106(0.9873)
IW	559.1617	563.1617	564.4851	0.0916(0.8218)	0.0806(0.6906)	0.5084(0.7373)
AGE	563.9111	569.9111	571.8961	0.1496(0.2518)	0.1963(0.2753)	1.0219(0.3455)
ALomx	556.8248	562.8248	564.8098	0.0532(0.9990)	0.0142(0.9998)	0.1482(0.9988)
ASE	587.1019	589.1019	589.7636	0.2771(0.0018)	1.0569(0.0017)	5.3298(0.0020)
ASEW	554.7389	560.7389	562.7239	0.0634(0.9896)	0.0214(0.9959)	0.1351(0.9994)
TSW	561.8178	567.8178	569.8028	0.1126(0.5930)	0.1014(0.5799)	0.6674(0.5857)

Table 5: *Some selection criteria and goodness-of-fit statistics (data-I).*

Distribution	**-2logL**	**AIC**	**HQIC**	KS(p-value)	CVM(p-value)	AD(p-value)
NCS-IW	31.0171	35.0171	35.4059	0.0975(0.9913)	0.0254(0.9906)	0.1594(0.9979)
IW	30.8174	34.8174	35.2062	0.1020(0.9854)	0.0266(0.988)	0.1545(0.9984)
AGE	36.8149	42.8149	43.398	0.1193(0.9385)	0.0577(0.8338)	0.5597(0.6847)
ALomx	35.4117	41.4117	41.9949	0.1136(0.9587)	0.0565(0.8416)	0.4783(0.7670)
ASE	154.7472	156.7472	156.9416	0.8863(0.0000)	5.1247(0.0000)	31.4397(0.0000)
ASEW	31.1885	37.1885	37.7716	0.1170(0.9470)	0.0363(0.9551)	0.2096(0.9877)
TSW	39.7066	45.7066	46.2898	0.1694(0.6147)	0.1415(0.4194)	0.8932(0.4170)

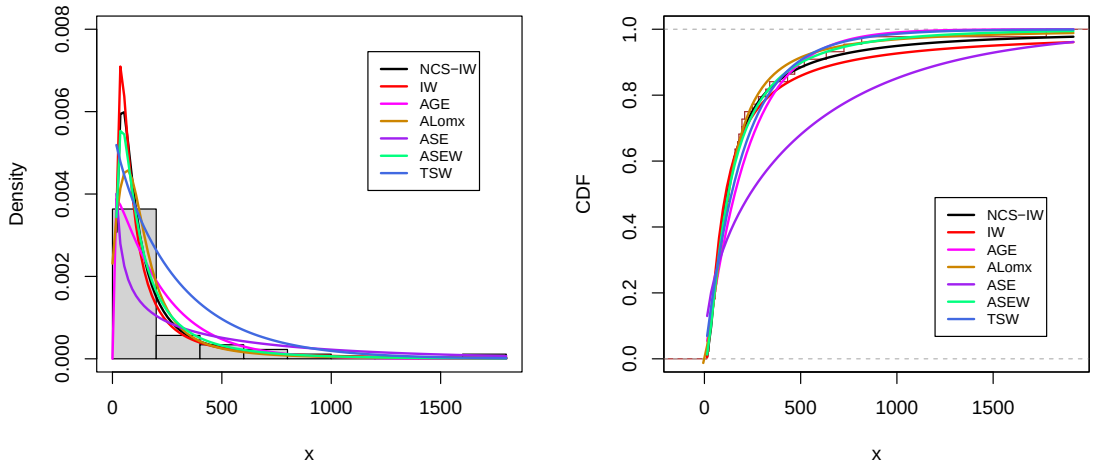


Figure 5: *Estimated PDF (left) and empirical vs estimated CDF (right) (data-I).*

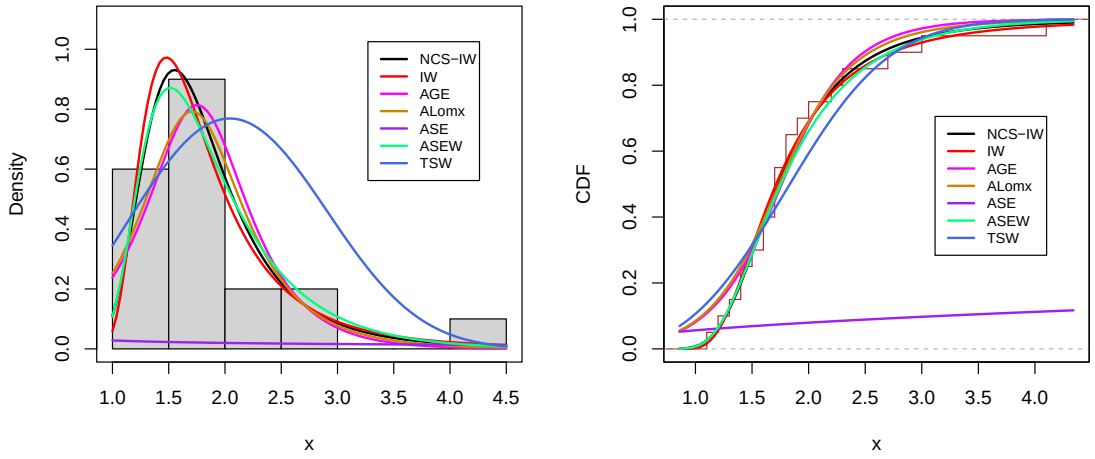


Figure 6: *Estimated PDF (left) and empirical vs estimated CDF (right) (data-II).*

6. Conclusion

Based on the ratio of CDF $G(x)$ and $1+G(x)$ of baseline distribution, we developed the new trigonometric family of distributions by transforming the sine function and we named it the new class sin-G family of distributions. General properties of the suggested family of distributions are provided. Using Inverse Weibull distribution as a baseline distribution, we have introduced a member of the suggested family having reverse-j or increasing or inverted bathtub-shaped hazard function. Some statistical characteristics of this NCS-IW distribution are explored. The associated parameters of the new distribution are estimated through the MLE method. To assess the estimation procedure, we conducted a Monte Carlo simulation and found that even for small samples, biases and mean square errors decreased as the size of the sample increased. Two real medical data sets are considered for the application of the NCS-IW distribution. Using some model selection criteria and goodness of fit test statistics, we empirically proved that the suggested model performs better than six other existing models (most of which have more parameters). Hence, we expect that the suggested family and its member distribution can be used in broader areas like medical science, reliability engineering, survival analysis, etc., and one can generate a new model using this family of distributions in the future.

References

[1] Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. Metron, 71(1), 63-79.

[2] Balakrishnan, N., & Cohen, A. C. (1991). *Order statistics & inference: estimation methods*. Academic Press, London.

[3] Chaudhary, A. K., Sapkota, L. P. & Kumar, V. (2021). Some properties and applications of arctan generalized exponential distribution. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)*, 10(1), 456-468.

[4] Chaudhary, A. K., & Kumar, V. (2021). The ArcTan Lomax distribution with properties and applications. International Journal of Scientific Research in Science, Engineering and Technology, 4099, 117-125.

[5] Chen, G., & Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology*, 27(2), 154-161.

[6] Chesneau, C., & Jamal, F. (2020). The sine Kumaraswamy-G family of distributions. *Journal of Mathematical Extension*, 15.

[7] Clark, V. A., & Gross, A. J. (1975). *Survival distributions: reliability applications in the biomedical sciences*. New York, John Wiley Sons.

[8] Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association*, 83(402), 414-425.
[9] Gomez-Deniz, E., & Calderin-Ojeda, E. (2015). Modelling insurance data with the Pareto ArcTan distribution. *ASTIN Bulletin: The Journal of the IAA*, 45(3), 639-660.
[10] Henningsen, A., & Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. Computational Statistics, 26, 443-458.
[11] He, W., Ahmad, Z., Afify, A. Z., & Goual, H. (2020). The arcsine exponentiated-X family: validation and insurance application. *Complexity*, 1-18.
[12] Isa, A. M., Ali, B. A., & Zannah, U. (2022). Sine Burr XII Distribution: Properties and Application to Real Data Sets. *AJBAR*, 1(3), 48-58.
[13] Kenney, J. F. & Keeping, E. S. (1962). *Mathematics of Statistics*, 3 edn, Chapman and Hall Ltd, New Jersey.
[14] Kharazmi, O. & Saadatinik, A. (2016). Hyperbolic cosine-F families of distributions with an application to exponential distribution. *Gazi University Journal of Science*, 29(4), 811-829.
[15] Kumar, D., Singh, U., & Singh, S. K. (2015). A new distribution using sine function-its application to bladder cancer patients data. *Journal of Statistics Applications & Probability*, 4(3), 417.
[16] Mahmood, Z., Chesneau, C., & Tahir, M. H. (2019). A new sine-G family of distributions: properties and applications. Bull. *Comput. Appl. Math.*, 7(1), 53-81.
[17] Marshall, A. W., & Olkin, I. (2007). *Life distributions* (Vol. 13). Springer, New York.
[18] Moors, J. J. A. (1988). A quantile alternative for kurtosis. Journal of the Royal Statistical Society: *Series D (The Statistician)*, 37(1), 25-32.
[19] Muhammad, M., Alshanbari, H. M., Alanzi, A. R., Liu, L., Sami, W., Chesneau, C., & Jamal, F. (2021). A new generator of probability models: the exponentiated sine-G family for lifetime studies. *Entropy*, 23(11), 1394.
[20] Renyi, A. (1961). *On measures of entropy and information*. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics (Vol. 4, pp. 547-562). University of California Press.
[21] Rahman, M. M. (2021). Arcsine-G Family of Distributions. *J. Stat. Appl. Pro. Lett.* 8(3), 169-179.
[22] Sakthivel, K. M. and Rajkumar, J. (2020). Hyperbolic cosine Rayleigh distribution and its application to breaking stress of carbon fibers. *Journal of Indian Society and Probability Statistics*, 21(2), 471-485.
[23] Sakthivel, K. M., & Rajkumar, J. (2021). Transmuted sine-G family of distributions: theory and applications. *Statistics and Applications*,(Accepted: 10 August 2021).
[24] Souza, L. (2015). *New trigonometric classes of probabilistic distributions (Doctoral dissertation, Thesis*, Universidade Federal Rural de Pernambuco).
[25] Souza, L., Junior, W. R. D. O., de Brito, C. C. R., Ferreira, T. A., & Soares, L. G. (2019a). General properties for the Cos-G class of distributions with applications. *Eurasian Bulletin of Mathematics* (ISSN: 2687-5632), 63-79.
[26] Souza, L., Junior, W., De Brito, C., Chesneau, C., Ferreira, T., & Soares, L. (2019b). On the Sin-G class of distributions: theory, model and application. *Journal of Mathematical Modeling*, 7(3), 357-379.
[27] Swain, J. J., Venkatraman, S., & Wilson, J. R. (1988). Least-squares estimation of distribution functions in Johnson's translation system. *Journal of Statistical Computation and Simulation*, 29(4), 271-297.

STUDY OF THE FUNCTIONING OF A MULTI-COMPONENT AND MULTI-PHASE QUEUING SYSTEM ON THE EXAMPLE OF A VEHICLE REPAIR ENTERPRISE

M.D. Katsman

•

The Joint-Stock Company of Railway Transport of Ukraine "Ukrzaliznytsia", Kyiv, Ukraine
mdkatsman@gmail.com

V.I. Matsiuk

•

National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
vimatsiuk@gmail.com

V.K. Myronenko

•

State University of Infrastructure And Technologies, Kyiv, Ukraine
pupil7591@gmail.com

Abstract

The purpose of the work is to build, on the basis of multi-component and multi-phase models of queuing systems (QS), mathematical models of maintenance and repair of vehicles by repair enterprises to increase the efficiency of their use. Results. The article considers multi-component and multi-stage mathematical models of QSs with the distribution of the arrival flow simultaneously between the system components, which consist of a certain number of service channels and waiting places in the queue. The same service channels can have different performance depending on the type of customers which they serve. Customers go through several stages of service and waiting. Considered are service of customers without a lack of time to stay in the service channel and waiting and with a lack of such time. The service process in the QS of each component consists of several (k_e) stages with the corresponding duration, the full-service period will be equal to the sum of such time intervals. Stage durations have certain probability distributions with appropriate parameters, then the total duration of the service process will have a generalized Erlang distribution with parameters of probability distributions of stages of order k_e. The number of components and their parameters correspond to the similar characteristics of the production divisions of the repair enterprise. The study of the effectiveness of the repair enterprise operation as a multi-component and multi-stage QS consists in determining the probability of service and the probability of failure of QS components and the system as a whole, the number of service channels, the number of customers in components, the number of customers in component queues, the duration of maintenance of customers in components and the system, the duration of being customers in queues of components and QSs, duration of stay of requirements in QSs and duration of customer waiting in QS queues. The model is implemented using Any Logic University Researcher. The AnyLogic University Researcher development environment allowed to combine the principles of system dynamics with the paradigms of agent and discrete-event modeling. In addition, thanks to the built-in Java SE compiler, a library of ready-made solutions is available, including generators of random variables, which significantly expands the possibilities of developing and implementing experiments. In particular, experiments on optimization (relative to a defined criterion), sensitivity of the model, stability of the model, etc. are available.

Keywords: vehicle repair enterprise, maintenance and repair, queuing theory application, multi-component and multi-phase queuing system, simulation, modeling, reliability

I. Introduction

The problem of ensuring the resource and reliability of vehicles is part of the general problem of transport safety and the efficiency of the use of such vehicles. Ensuring high reliability primarily depends on the effectiveness of the maintenance and repair strategy and the quality of work.

Maintenance and repair of railway, including special rolling stock, is carried out in accordance with the requirements of the Regulations [1] at repair plants, depots, track engineering stations, workshops and maintenance points.

Requirements for the condition of rolling stock, the procedure for its maintenance and repair, sending it to repair bodies, as well as technical instructions and typical technological processes for maintenance and repair of rolling stock are determined by Ukrzaliznytsia [1].

The aircraft maintenance system is designed to maintain and restore the airworthiness and serviceability of aircraft and prepare them for flight. Technical operation is carried out by operators, aviation and technical bases, maintenance and repair enterprises, repair enterprises, aviation and technical services of airports [2].

Aircraft maintenance is carried out during major and other repairs (or during equivalent works), inspections, modifications, upgrade, elimination of defects, which are carried out by aircraft repair enterprises both individually and collectively in the relevant workshops, production divisions and areas, laboratories, stands, etc. [3].

Maintenance and repair of motor vehicles and their components is performed in order to maintain them in proper condition and ensure the technical characteristics established by the manufacturer for use, storage or maintenance during the period of operation [4]. Requirements for maintenance and repair of motor vehicles and the services provided by them (work to be performed) are established by technical regulations. The work is carried out in workshops of motor vehicle enterprises and car repair enterprises.

The work [5] is devoted to the creation and introduction into the practice of air transport of information and advisory systems for the maintenance of passenger aircraft based on modern computer technologies and mathematical methods of information processing.

In [6], the structure of the methodical apparatus for ensuring a given level of serviceability of on-board equipment products, in particular optoelectronic sighting systems of military aircraft of the Air Force of Ukraine, is proposed.

Methodical approaches to the structural and parametric determination of general requirements for ground flight maintenance facilities are considered in [7], which can also be used to develop a methodology for conducting tests and assessing the quality of modern weapons systems and military equipment at all stages of the life cycle.

Based on the analysis of the existing methods of calculating the durability indicators of the radio-electronic system of the aircraft, the factors affecting its reliability were identified in [8], and measures were also proposed to improve the existing scientific and methodological apparatus for calculating such indicators.

Works [9] and [10] are devoted to the analysis of the causes of failure situations at the airport. The aircraft maintenance system was analyzed, it was shown that ensuring uninterrupted operation of the airport, execution of the daily flight plan in extraordinary situations is possible only by introducing into the control circuit of the aircraft ground handling system an intelligent decision support system for dispatchers, which will take into account the positive experience of their actions in typical, emergency and failure situations. This will allow, in particular, to reduce the time to get out of a malfunctioning situation and to optimize the operational planning of the ground maintenance of aircraft, considering the available equipment and special technical means.

In work [11], organizational measures are given with the help of which it is possible to

minimize the lack of transport aviation during the transportation of general and oversized cargo. Data on incidents related to aircraft ground handling are given, the causes of the events are indicated. The main ways of eliminating the problems of standardization of airfield technical support in the conditions of interaction with NATO and in the processes of international integration are defined.

The work [12] is devoted to the solution of the problem of minimizing the risks of import substitution in the process of factory repair of military aviation equipment in the conditions of a special period, the issue of providing post-repair military aviation equipment by adjusting the production of necessary component parts by domestic enterprises in the process of import substitution is analyzed.

The work [13] presents the results of the quality of repair of aircraft equipment at aircraft repair enterprises. A significant proportion of the failures detected during the operation of aviation equipment after major (medium) repairs are the result of manufacturing defects of components (parts) that were installed on aircraft. Technological methods for ensuring sufficient repair quality and significantly reducing the risks of production defects are proposed.

The work [14] is devoted to the problem of mathematical modeling of the processes of technical operation of military aircraft. The results of the analysis show that the most acceptable modeling method in terms of the compliance of the models with the proposed requirements is the simulation modeling method, and the more accepted class for creating a stochastic model of aircraft maintenance and repair processes is the class of semi-Markov models.

In work [15], a three-dimensional model of an aircraft skin element with riveted seams was built using the Sold Works software, wind load simulation was carried out in the ANSYS software package, which allowed to determine the stress-strain state of the aircraft skin elements in the presence of multifocal damage to the riveted seams.

Modern methods and approaches to modeling technological systems are considered in [16]. Basic definitions and concepts are given. New approaches to solving problems that arise during the development of models of mechanisms, systems and processes of machine-building production are proposed.

In work [17], it is proposed to consider the production of car service enterprises as an open multi-channel QS, in which random processes occur due to the combined action of random variables. As a result of the experimental study, information was obtained about the indicators characterizing maintenance and repair, as well as affecting the change in the parameters of these processes. The developed model makes it possible to take into account the specifics of managing car maintenance stations.

The paper [18] considered a model for assessing the technical condition of radio-electronic elements of water transport vehicles using control and diagnostic equipment as a QS with a limited number of channels and a storage of customers to be served. On the basis of various optimization criteria, it is possible to establish a rational system for assessing the technical condition of such elements, to determine the expediency (rational, optimal) of developing a number of different types of control and diagnostic equipment and the effectiveness of new assessment methods.

The work [19] is devoted to the development of a simulation model of the influence of an accurate assessment of the readiness factor of mobile control and diagnostic complexes on the reliability of control of radio-electronic systems of marine transport.

In the paper [20], a new model of the task of managing the processes of diagnosis and monitoring of automation tools is proposed for the objects of rail-water transport connection, compiled on the basis of the experimental research results and mathematical description using Markov chains with an informative parameter in the form of damage intensity, aimed at increasing the efficiency of forecasting the technical condition of automation equipment.

The work [21] describes workshops for repairing locomotives in the form of multi-channel QS

with a limited queue. A simulation model of such a workshop as a QS was developed, which allowed rational use of equipment, labor force, as well as distribution of repair work time.

In work [22], the issue of modeling the processes of maintenance and repair of technical systems of a distributed information system is considered. The model is based on a joint presentation of the serviced system and its technical operation system in the form of a closed non-homogeneous QS consisting of two types of QS. The QS of the first type simulates the functioning processes of repair bodies to meet the received requirements.

In work [23], a study of the actions of emergency units of railway transport as processes of functioning of mass service systems was carried out. The authors established quantitative relationships between the intensity of the influence of dangerous factors of a railway emergency situation, the time of arrival, deployment and productivity of actions of emergency aftermath liquidation units and the effectiveness of liquidation works due to the implementation of the principles of network-centric management of complex dynamic hierarchical transport systems.

Thus, to improve the management processes of material, human, financial and informational resources during the maintenance and repair of aviation equipment, in particular on-board power supply systems, a wide range of operations research methods, the theory of mass service systems and simulation modeling are currently used.

II. Methods

The on-hand practical experience of the organization of maintenance and repair of vehicle equipment indicates that certain types of their technical systems that require various types of repair work, modifications, upgrade, inspections, elimination of defects, etc. are sent to specific production divisions of the repair enterprise, which, according to their purpose, carry out the necessary types of work according to the specified technologies.

To simulate the processes of maintenance and repair of vehicle, which are carried out by the production divisions of the repair enterprise, it is advisable to use multi-component and multi-phase QSs, which can be of both Markov and non-Markov types, capable of serving the arriving flows of non-priority, in general, heterogeneous (mixed) customers. At the same time, the system can have an arbitrary number of common service channels of the same type, and each component can also have an arbitrary number of places in the queue.

The same service channels can have different performance depending on the type of requirements for which they are involved: when the j-component of the system receives uniform requirements with the rate λj determined by the overall rate λ of the source, in the general case, of mixed customers. The magnitude of the source of mixed customers entering the system has an intensity (arrival rate) of

$$\lambda = \sum_{j=1}^{L} \lambda_{j,} \qquad j = \overline{1, L,}$$

where L is the number of components in the QS.

The service process in each component of the QS consists of several stages (phases) with the corresponding duration T_i, then the full-service period T_s is equal to

$$T_s = \sum_{i=1}^{K_p} T_i$$

where K_p is the number of such phases.

All T_i durations have certain probability distributions with the appropriate parameters, then T_s will have a generalized Erlang distribution with the parameters of the probability distributions of order K_p.

The number of components and their parameters correspond to similar characteristics of the repair enterprise.

The study of the operation effectiveness of a vehicle repair enterprise as a multi-component QS will consist in determining the probability and time characteristics of each component and the QS as a whole.

Let's consider several examples.

Example 1. Two-component QS with $M/E_4/2/3$ in the first component and $M/E_3/1/2$ in the second component without restrictions.

The graph of states of such a QS is presented in fig. 1.

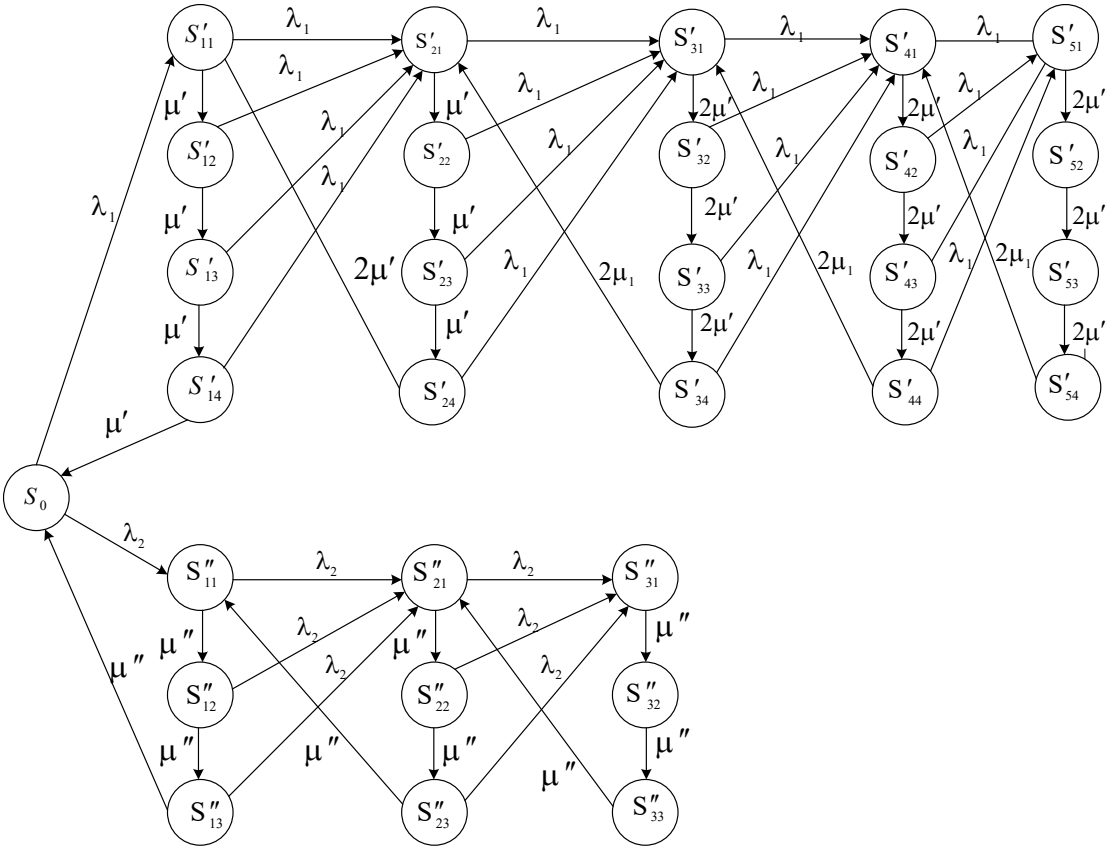


Figure 1: *State graph of QS of $M/E_4/2/3$ type in the first component and $M/E_3/1/2$ type in the second component without constraints*

There are no constraints in the QS because customers do not leave the service channel during service and the queue during the service waiting period due to the lack of time for them to be in service and in the queue.

Using the well-known algorithm for solving the Kolmogorov equations, we obtain the probabilities of the system states P'_{1i}, P'_{2i}, P'_{3i}, P'_{4i} P'_{5i} $(i=\overline{1,4})$ та P''_{1i}, P''_{2i}, P''_{3i} $(i=\overline{1,3})$.

Service in the QS of the first component consists of four phases of duration T'_1, T'_2, T'_3 and T'_4, the full-service time T'_s has a generalized Erlang distribution of the 4th order with a mathematical expectation of $1/\mu_1$.

$$T'_s = \sum_{i=1}^{4} T'_i,$$

where T'_i have an exponential distribution with parameter $\mu' = 4\mu_1$.

Service in the second component of the QS consists of three phases of durationT'_1, T'_2 and T'_3, the full-service time T''_s has a generalized Erlang distribution of the 3rd order with a mathematical expectation of $1/\mu_2$.

$$T''_s = \sum_{i=1}^{3} T''_i,$$

where T''_i have an exponential distribution with parameter $\mu'' = 3\mu_1$.

The QS states of the first component are characterized by the following probabilities [24]:

$$P'_c = \sum_{j=1}^{4} P'_{cj}, c = \overline{1,5}; j = \overline{1,4},$$

where:

$P'_1 = \sum_{j=1}^{K'_p} P'_1 j$ is the probability of one server busy in the QS component (1 customer in component 1);

$P'_2 = \sum_{j=1}^{K'_p} P'_2 j$ is the probability of two servers busy in the QS component (2 customers in component 1);

$P'_3 = \sum_{j=1}^{K'_p} P'_3 j$ is the probability of 3 customers being in the component, of which 2 are served, one is in the queue;

$P'_4 = \sum_{j=1}^{K'_p} P'_4 j$ is the probability of 4 customers being in the component, of which 2 are served, 2 are in the queue;

$P'_5 = \sum_{j=1}^{K'_p} P'_5 j$ is the probability of 5 customers being in the component, of which 2 are served, 3 are in the queue.

Similarly, for the QS states of the second component:

$$P''_c = \sum_{j=1}^{3} P''_{cj}, c = \overline{1,3}; j = \overline{1,3},$$

where:

$P''_1 = \sum_{j=1}^{K''_p} P''_1 j$ is the probability of one customer being served in the component;

$P''_2 = \sum_{j=1}^{K''_p} P''_2 j$ is the probability of two customers being in the component, one served, one in the queue;

$P''_3 = \sum_{j=1}^{K''_p} P''_3 j$ is the probability of three customers being in the component, one served, two in the queue.

The number of busy service channels in the component:

$$\overline{k_1} = P'_1 + 2\sum_{c=2}^{5} P'_c \Big/ 4; \quad \overline{k_2} = \sum_{c=1}^{3} P''_c \Big/ 3.$$

Probability of service in the first component:

$$\mathrm{P}'_s = 1 - \mathrm{P}'_{ls} - \sum_{c=1}^{(n+m)"} P_c^{"} = 1 - \mathrm{P}'_{fl} - \sum_{c=1}^{(n+m)"} P_c^{"},$$

where P'_{ls} is the probability of loss of a customer;

P'_{fl} is the probability of failure of serving a customer.

Here, "failure" means the impossibility of servicing a customer (client) in a system component (for example, due to insufficient service rate compared to the arrival flow rate). "Loss" means leaving the queue for service by a customer (loss of a client by the system) due to the impossibility of waiting for service. Example, cars needing gasoline and service arrive to the gas or car service station. However, if the station already is being used, fully busy with servicing other cars, these potential customers may balk to another service station.

$$\mathrm{P}_{\mathrm{c}}^{"} = \sum_{j=1}^{k_{\mathrm{E}}^{"}} P_{cj}^{"}, c = \overline{1,(n+m)}; \quad \mathrm{P}'_{fl} = P'_{(n+m)'} = \sum_{c=2}^{(n+m)'} P'_{cj} \Big/ k'_{\mathrm{E}};$$

$$P'_{bo} = \sum_{\substack{d=2 \\ j=2}}^{k_{\mathrm{E}}} (d-1) P'_{(b-1)j}, \qquad b = \overline{2,(n+m)'}$$

$$P'_{ls} = P'_5 - \sum_{b=5}^{5} P'_{\mathrm{bo}} \Big/ 4,$$

$$P'_{20} = P'_{12} + 2P'_{13} + 3P'_{14}; \; P'_{30} = P'_{22} + 2P'_{23} + 3P'_{24};$$

$$P'_{40} = P'_{32} + 2P'_{33} + 3P'_{44}; \; P'_{50} = P'_{42} + 2P'_{43} + 3P'_{44};$$

$$P_1^{"} = P_{11}^{"} + P_{12}^{"} + P_{13}^{"}; \quad P_2^{"} = P_{21}^{"} + P_{22}^{"} + P_{23}^{"}; \quad P_3^{"} = P_{31}^{"} + P_{32}^{"} + P_{33}^{"}.$$

Then

$$\mathrm{P}'_s = 1 - \mathrm{P}'_5 + \sum_{b=2}^{5} P'_{\mathrm{bo}} \Big/ 4 - \sum_{i=1}^{3} P_c^{"}. \quad P_s^{"} = 1 - P_{ls}^{"} - \sum_{\mathrm{c}=1}^{5} P'_c.$$

The probability of service in the second component $P_s^{"}$ is determined considering the fact that

$$P_{ls}^{"} = P_{fl}^{"} = P_3^{"} - \frac{P_{20}^{"} + P_{30}^{"}}{3}; \quad P_{20}^{"} = P_{12}^{"} + 2P_{13}^{"}; P_{30}^{"} = P_{22}^{"} + 2P_{23}^{"}.$$

Whence

$$P_s^{"} = 1 - P_3^{"} + \frac{P_{20}^{"} + P_{30}^{"}}{3} - \sum_{c=1}^{5} P'_c.$$

The average number of customers $\overline{N}^{(i)}$ in the i-component:

$$\overline{N}^{(1)} = \sum_{i=1}^{(n+m)'} i\, P'_i / k'_E = (1 \cdot \mathrm{P}'_1 + 2 \cdot \mathrm{P}'_2 + 3 \cdot \mathrm{P}'_3 + 4 \cdot \mathrm{P}'_4 + 5 \cdot \mathrm{P}'_5)/4,$$

$$\overline{N}^{(2)} = \sum_{i=1}^{(n+m)"} i\, P'_i / k_E^{"} = (1 \cdot \mathrm{P}'_1 + 2 \cdot \mathrm{P}'_2 + 3 \cdot \mathrm{P}'_3/3.$$

The average number of requests $\overline{N}_q^{(i)}$ that are in the queue and waiting for service in the i-component:

$$\overline{N}_q^{(1)} = \sum_{q=1}^{m'} qP'_{(n+q)} / k'_E = (1 \cdot \mathrm{P}'_3 + 2 \cdot \mathrm{P}'_4 + 3 \cdot \mathrm{P}'_5)/4:$$

$$\overline{N}_q^{(1)} = \sum_{q=1}^{m''} qP''_{(n+q)} / k''_E = (1 \cdot \mathrm{P}''_2 + 2 \cdot \mathrm{P}''_3)/3.$$

Duration of waiting time for the customer in the queue for the i-component equals:

$$\overline{W}_q^{(1)} = \frac{\overline{N}_q^{(1)}}{\lambda_1};\ \overline{W}_q^{(2)} = \frac{\overline{N}_q^{(2)}}{\lambda_2}.$$

Customer service time in the QS:

$$\overline{t}_{sqs} = \frac{\lambda_1}{\lambda_1+\lambda_2}\overline{t}_s^{(1)} + \frac{\lambda_2}{\lambda_1+\lambda_2}\overline{t}_s^{(2)},$$

$$t_s^{(1)} = \frac{\overline{N}^{(1)}}{\lambda_1};\quad t_s^{(2)} = \frac{\overline{N}^{(2)}}{\lambda_2}.$$

Duration of waiting for customers in QS queues:

$$\overline{w}_{wqs} = \frac{\lambda_1}{\lambda_1+\lambda_2}\overline{w}_q^{(1)} + \frac{\lambda_2}{\lambda_1+\lambda_2}\overline{w}_q^{(2)}.$$

Probability of QS failure:

$$P_{fl}^{qs} = \frac{\lambda_1}{\lambda_1+\lambda_2}\mathrm{P}_{fl_1} + \frac{\lambda_1}{\lambda_1+\lambda_2}\mathrm{P}_{fl_2}.$$

It should be noted that in the L-component QS in a steady mode, the value of the probability of customer service in the j-component can be determined as follows [24]:

$$\mathrm{P}_{s_j} = 1 - \mathrm{P}_{ls_j} - \sum_{\psi \neq j}^{L}\left(\sum_{i=1}^{n+m} P_i\right)_{\psi},$$

$$\psi = \overline{1, L} = 1,2,3 \ldots, j \ldots, L;$$

$$i = \overline{0, (n+m)};$$

$$\mathrm{P}_{S_{qs}} = \sum_{j=1}^{L}\frac{\lambda_j}{\lambda}\mathrm{P}_{S_j},$$

$$\lambda = \sum_{j=1}^{L}\lambda_j.$$

During the maintenance and repair of aircraft, force majeure circumstances may arise, related to the time limitations of customer service for maintenance and waiting in the queue, so let's consider the following example.

Example 2. Two-component QS of M/E4/2/3 type in the first component and of M/E3/1/2 in the second component with restrictions on the time spent in the service period $\beta_{1,2}$ and waiting $\gamma_{1,2}$.

The graph of the states of this QS coincides with the graph of the states of the QS presented in fig. 1.

At each service stage of the first channel, the service duration has an exponential distribution with the parameter $\mu'+\beta$, in the second service channel $2\mu'+2\beta$, in queues $2\mu'+2\gamma_1$.

In the second component, the parameter $\mu''+\beta_2$ at the service stages, and $\mu''+\gamma_2$ at the stages of waiting in the queue.

The expressions for the probabilities of the QS states of the first and second components are similar to the QS considered above.

Service probability for the first component [24]:

$$P'_S = 1 - P'_{ls} - \sum_{i=1}^{3} P''_c,$$
$$P'_{ls} = P'_{fl} + P'_{\underset{s}{lv}} + P'_{\underset{q}{lv}},$$

where

$P'_{\underset{s}{lv}}$ is the probability of the customer leaving the system in the service channel;

$P'_{\underset{q}{lv}}$ is the probability of the customer leaving the system in the queue.

$$P'_{fl} = P'_5 - \textstyle\sum_{b=1}^{5} P'_{bo}/5,$$
$$P'_{\underset{s}{lv}} = \frac{\beta_1}{\lambda_1}\overline{k_1}.$$

The average number of customers in the queue

$$\overline{N}_q^{(1)} = \sum_{q=1}^{m\prime} qP'_{(n+q)} \Big/ k'_E = (1 \cdot P'_3 + 2 \cdot P'_4 + 3 \cdot P'_5)/4.$$

If $\beta_1 = \gamma_1$, then

$$P'_{\underset{q}{lv}} = \frac{\beta_1}{\lambda_1}\overline{N}_q^{(1)}.$$

Hence

$$P'_{ls} = P'_5 - \sum_{b=2}^{k'_\epsilon} P'_{bo} \Big/ k'_E + \frac{\beta_1}{\lambda_1}\overline{k_1} + \frac{\gamma_1}{\lambda_1}\overline{N}_q^{(1)},$$
$$P'_S = P'_{ls} - \sum_{c=1}^{3} P''_c.$$

The probability of customer service in the second component is

$$P''_S = 1 - P''_{ls} - \sum_{c=1}^{5} P'_c,$$

where

$$P_{ls}^{"} = P_{fl}^{"} + P_{lv}^{"}{}_{s} + P_{lv}^{"}{}_{q},$$
$$P_{fl}^{"} = P_{3}^{"} - \frac{P_{20}^{"} + P_{30}^{"}}{\lambda_2},$$
$$P_{lv}^{"}{}_{q} = \frac{\beta_2}{\lambda_2}\overline{k}_2,$$
$$\overline{N}_q^{(2)} = \frac{1 \cdot P_2^{"} + 2 \cdot P_3^{"}}{3}.$$

Then

$$P_{lv}^{"}{}_{q} = \frac{\gamma_2}{\lambda_2}\overline{N}_q^{(2)}.$$

Provided that $\beta_2 = \gamma_2$,

$$P_{lv}^{"}{}_{q} = \frac{\beta_2}{\lambda_2}\overline{N}_q^{(2)}.$$

When applying the proposed mathematical models, it is advisable to consider the following:

- in multi-component QSs, the performance of any component decreases compared to a single-component system at the same rates of service stages. With the same values of the parameters of each component of the QS, the performance of multi-component and single-component systems will be the same;

- if one of the components is a QS with a queue, and the second component is a QS with failures, then the QS with a queue has a higher performance, simultaneously reducing the performance of the second component;

- with small values ($0 \le P_S \le 0.1$), the impact on the system as a whole or on a separate component of the intensities of customers leaving the system during the service period and being in the queue is insignificant. When these intensities change, the P_S value will fluctuate relative to its average value.

Example 3. Consider a two-component QS of $M/E_2/1/2$ type in each component.

Factors of the effectiveness of the functioning of such a QS include: service probabilities P_{S_1} and P_{S_2}, failure probabilities P_{fl_1}, P_{fl_2} and $P_{fl}^{(QS)}$, the number of service channels $\overline{k}_1$ and $\overline{k}_2$, the total number of customers in components $\overline{N}^{(1)}$ and $\overline{N}^{(2)}$, the number of customers in component queues $\overline{N}_q^{(1)}$ and $\overline{N}_q^{(2)}$, service time in components $\overline{t}_1$ та $\overline{t}_2$, waiting time in component queues $\overline{w}_1$ and $\overline{w}_2$ and the customers service time duration in channels and waiting time in the queues of the QS.

III. Results of experiments and discussion of the results

The values of these factors are determined according to the formulas given above.

Graphs of dependencies of some factors' impact on the QS functioning effectiveness are constructed using AnyLogic production process simulation software and presented in fig. 2 – 8 for one of the 15 series of computer experiments.

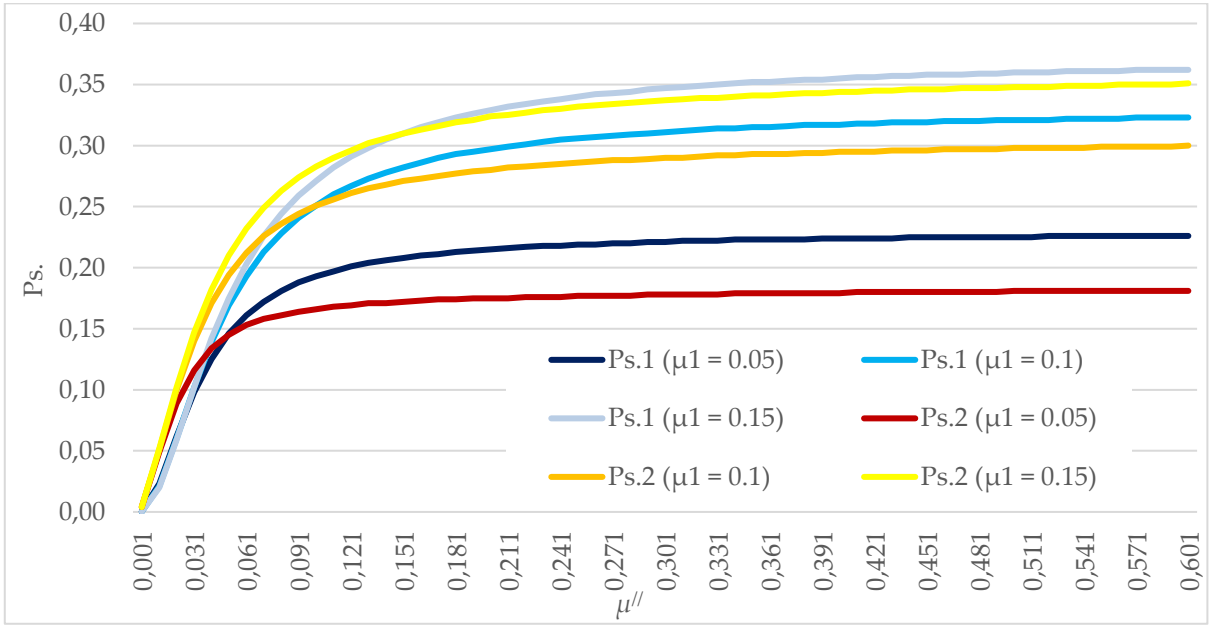


Figure 2: *Dependence of probabilities* P_{S_1} *and* P_{S_2} *on service rate* μ'' *at* $\mu' = 0.05; 0.1; 0.15$

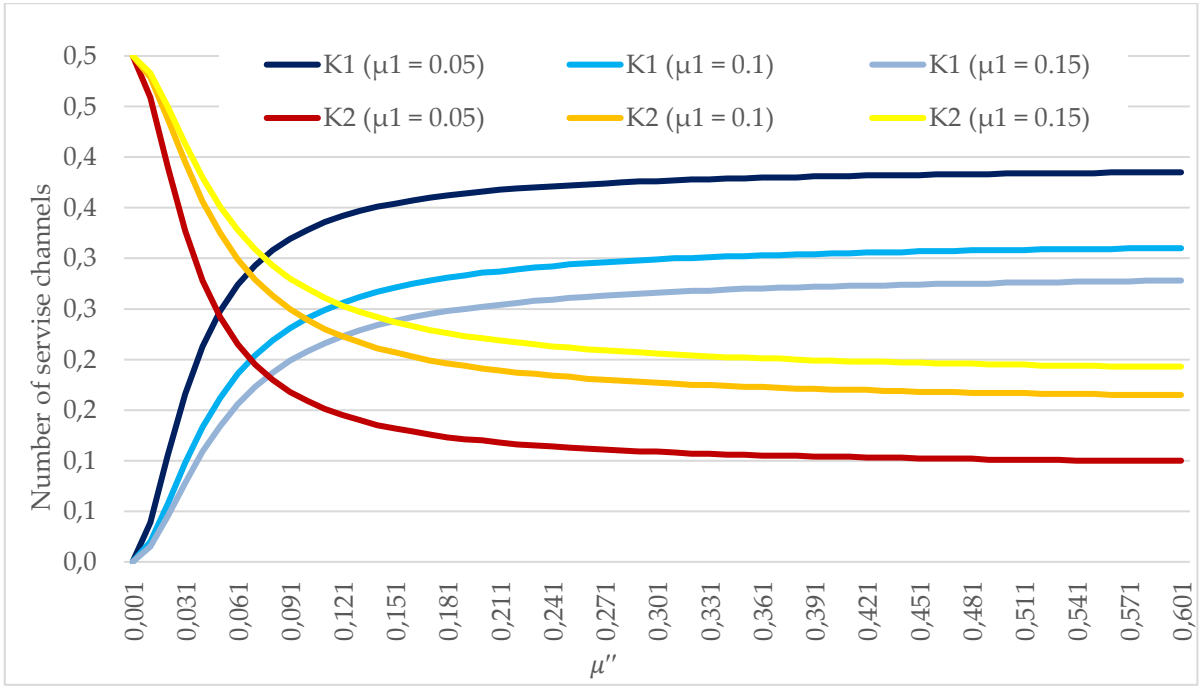


Figure 3: *Dependence of number of service channels* $\overline{k}_1$ *and* $\overline{k}_2$ *on service rate* μ'' *at* $\mu' = 0.05; 0.1; 0.15$

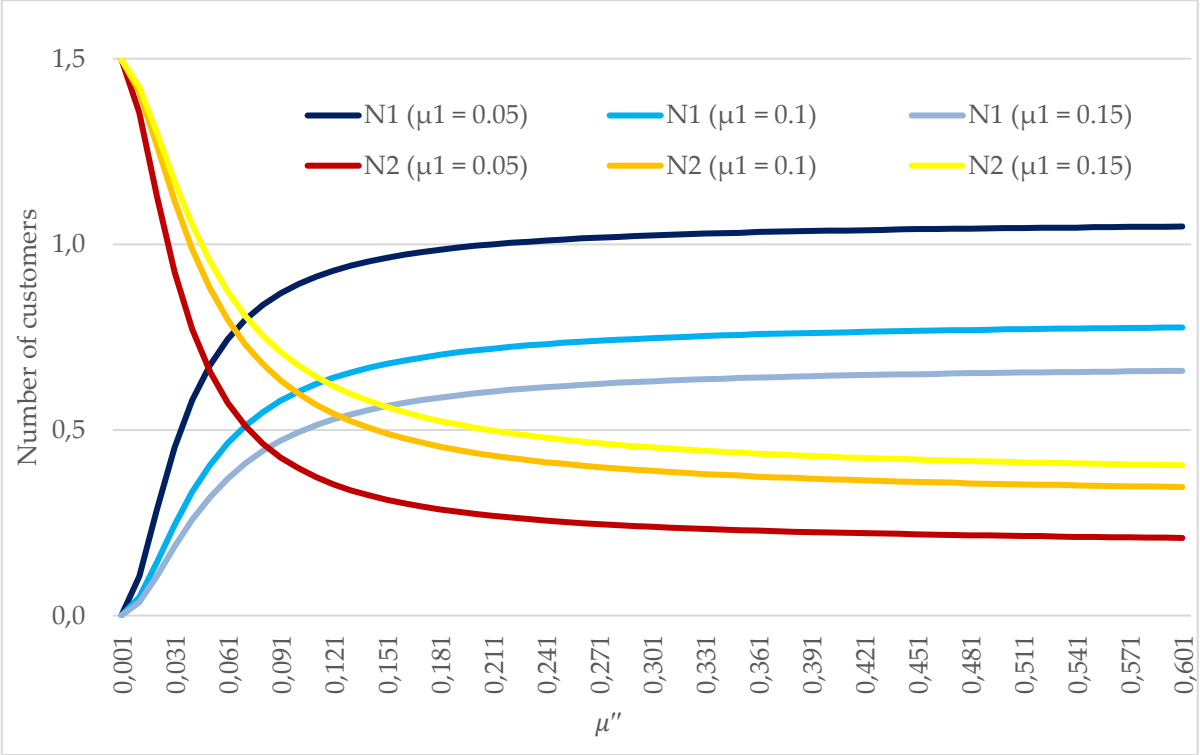


Figure 4: *Dependence of number of customers $\overline{N}^{(1)}$ and $\overline{N}^{(2)}$ in components on service rate μ'' at $\mu' = 0.05; 0.1; 0.15$*

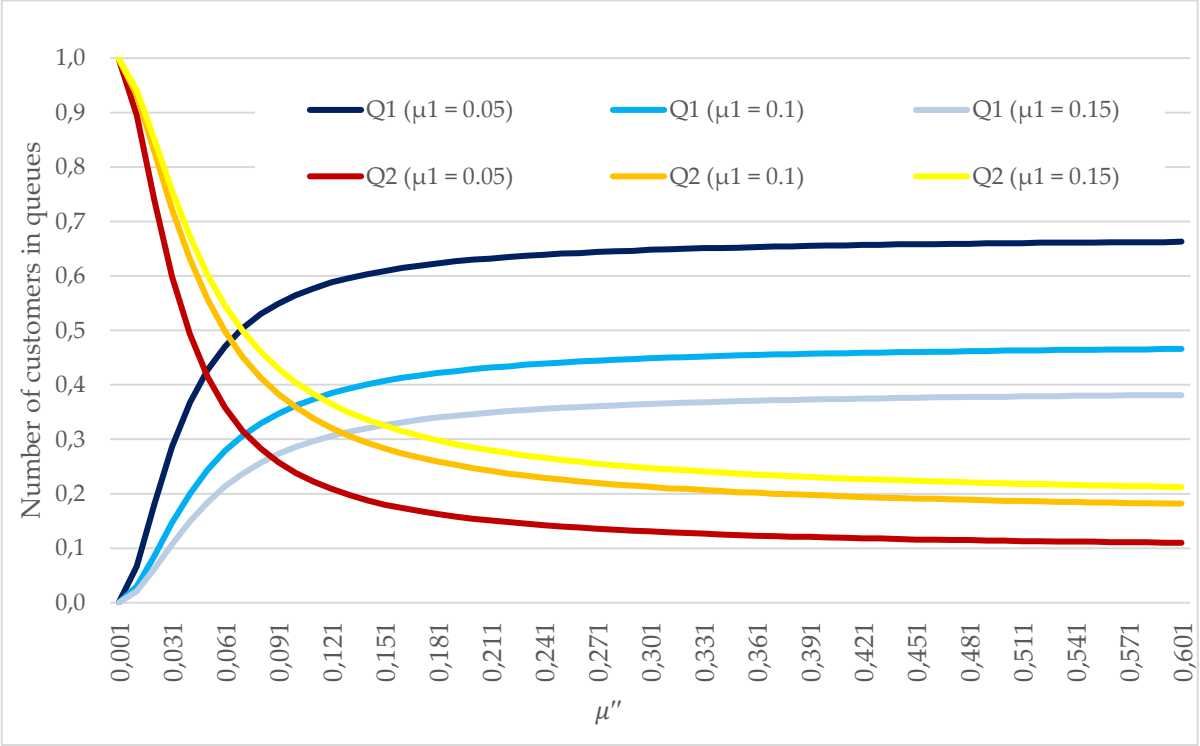


Figure 5: *Dependence of number of customers in queues by component $\overline{N}_q^{(1)}$ and $\overline{N}_q^{(2)}$ on service rate μ'' at μ' = 0.05; 0.1; 0.15*

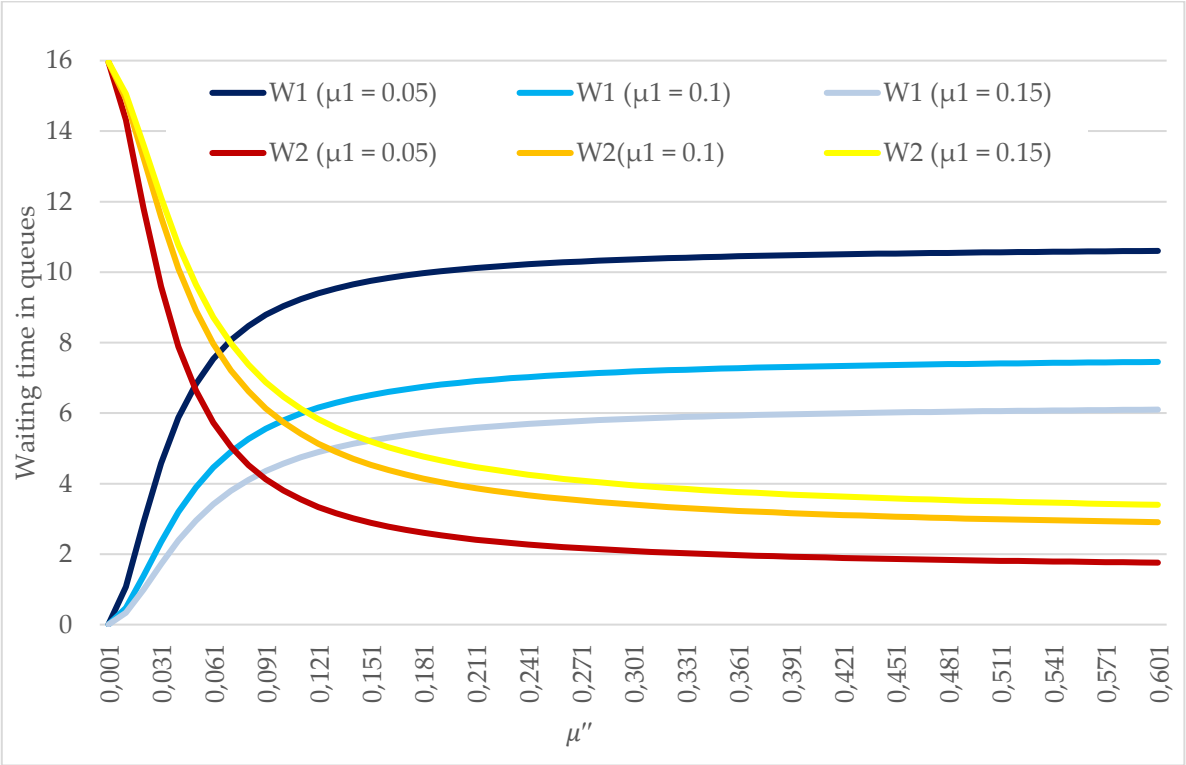


Figure 6: *Dependence of waiting time in queues in components $\overline{w}_1$ and $\overline{w}_1$ on service rate μ'' at μ' = 0.05; 0.1; 0.15*

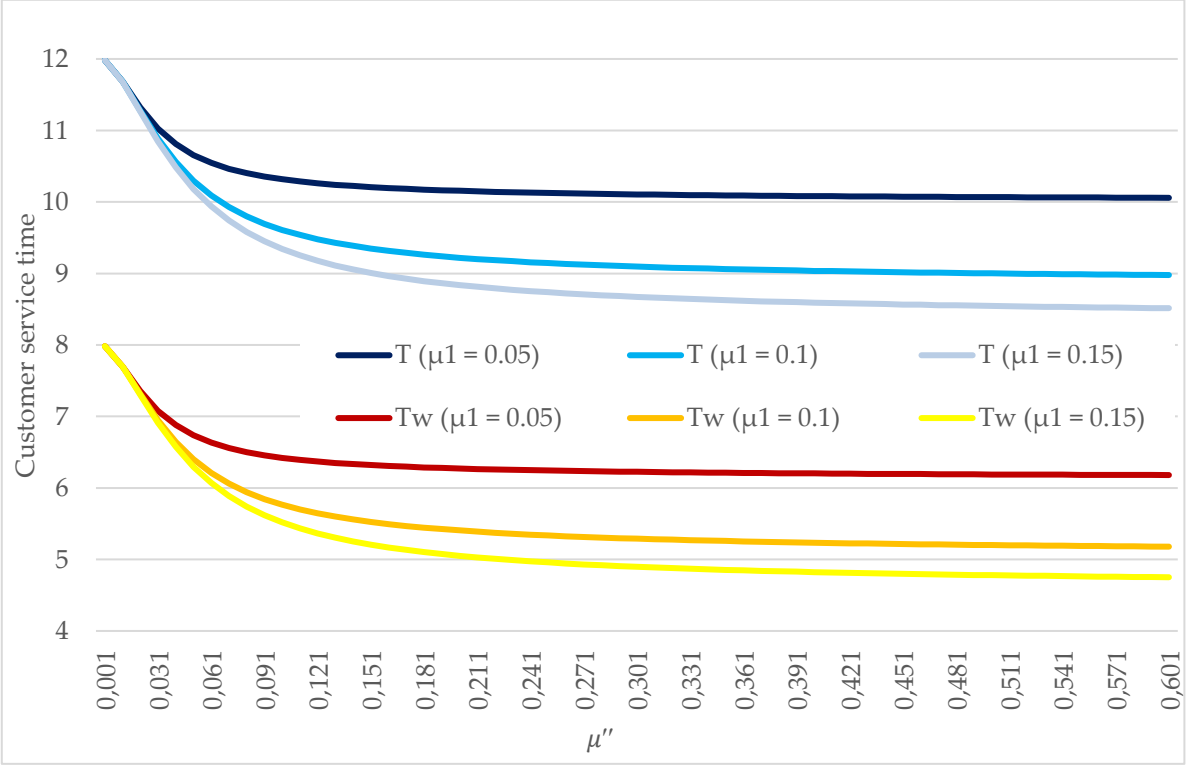


Figure 7: *Dependence of customer service time T and waiting time in QS queues $\overline{W}$ on service rate μ'' at μ' = 0.05; 0.1; 0.15*

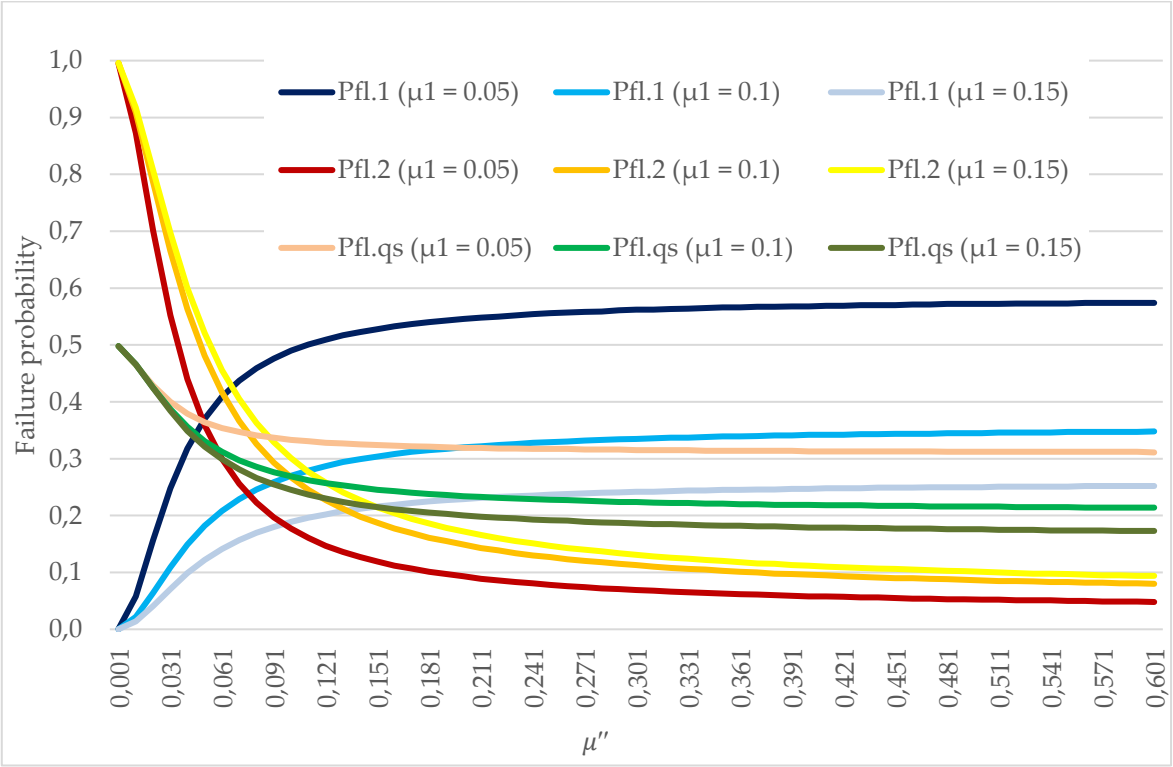


Figure 8: *Dependence of failure probabilities* P_{fl_1} *ma* P_{fl_2} *in components and in QS* $P_{fl_{QS}}$ *on service rate* μ'' *at* $\mu' = 0.05; 0.1; 0.15$

The AnyLogic University Researcher development environment allows you to combine the principles of system dynamics with the paradigms of agent and discrete-event modeling. In addition, thanks to the built-in Java SE compiler, a library of ready-made solutions is available, including generators of random variables, which significantly expands the possibilities of developing and implementing experiments. In particular, experiments on optimization (relative to a defined criterion), model sensitivity, model stability, etc. are available [23], [24], [25], [26], [27].

From the graphs presented in fg. 2 – 8, it can be seen that with a time interval between the arriving customers of 8 hours ($\lambda_1 = \lambda_2 = 0.125$), with an increase in μ', the values of P_{S_1} and P_{S_2} increase from 0.24 to 0.38 and from 0.19 to 0.37, respectively, but are at an insufficient level.

The average number of busy service channels of the first component $\overline{k}_1$ decreases from 0.4 to 0.28, and the values of $\overline{k}_2$ increase from 0.1 to 0.19, which indicates a weak system utilisation.

The total number of customers in the first component $\overline{N}^{(1)}$ decreases from 2.1 to 1.48 due to small values of μ', and $\overline{N}^{(2)}$, on the contrary, increases from 0.48 to 1.6.

The average number of customers in the queue of the first component $\overline{N}^{(1)}$ decreases from 1.4 to 0.75, and in the second component $\overline{N}^{(2)}$ increases from 0.26 to 0.48.

The duration of waiting time $\overline{w}_1$ in the queue of the first component decreases from 10.1 hours to 6 hours, and the value of $\overline{w}_2$ increases from 2.2 hours to 4 hours.

The duration of service time $\overline{t}_{qs}^{s}$ in the QS is reduced from 10 h to 8.5 h, as well as the duration of waiting time in queues $\overline{w}_{qs}$ from 6.1 h to 3 h.

The probability of failure P_{fl} decreases from 0.34 to 0.19.

As we can see, the availability of the presented mathematical model contributes to the determination of measures aimed at increasing the operation effectiveness of the QS, which simulates the work of a vehicle maintenance and repair enterprise, that will ultimately allow to organize more efficient work of such an enterprise.

Conclusions

The theoretical approach proposed by the authors is implemented on the example of the modeling of vehicle maintenance and repair processes by production divisions of a repair enterprise as a multi-component and multi-phase queuing system (QS) and allows to determine the effectiveness of the functioning of such a QS, identify "bottlenecks", unreliable components and obtain arguments for improving the efficiency of the enterprise in rapidly changing conditions.

The presented mathematical apparatus and used simulation modeling tools show their relevance to real processes and can be applied to improve the performance of not only vehicle repair enterprises, but also a greater variety of objects that can be described as QSs of various types, according to conditions of their functioning.

The conducted comprehensive research and its results make it possible to increase the reliability and efficiency of the functioning of a wide class of objects and systems, as they allow to evaluate the quantitative, qualitative and probabilistic characteristics of various technological processes and organizational measures as processes in mass service QSs, and therefore to minimize delays, failures in maintenance and related losses and optimize production, transport-logistics, military and other complex systems.

References

[1] Aviatsiini pravyla Ukrainy «Pidtrymannia lotnoi prydatnosti povitrianykh suden ta aviatsiinykh vyrobiv, komponentiv i obladnannia ta skhvalennia orhanizatsii i personalu, zaluchenykh do vykonannia tsykh zavdan». Nakaz Derzhavnoi aviatsiinoi sluzhby Ukrainy vid 06.03.2019 № 256. Zareiestrovano v Ministerstvi yustytsii Ukrainy 29 bereznia 2019 r. za № 316/33287 {iz zminamy, vnesenymy nakazom Derzhavnoi aviatsiinoi sluzhby Ukrainy № vid 26.11.20}.

[2] Pravyla inzhenerno-aviatsiinoho zabezpechennia derzhavnoi aviatsii Ukrainy. Nakaz Ministerstva oborony Ukrainy vid 05.07.2016 № 343. Zareiestrovano v Ministerstvi yustytsii Ukrainy 08 serpnia 2016 r. za № 1101/29231 {iz zminamy, vnesenymy nakazom Ministerstva oborony Ukrainy № 223 vid 03.08.2021}.

[3] Tamarhazin O.A., M.V. Oleh. Metodolohichni osnovy systemy upravlinnia mobilnoiu aviatsiinoiu transportnoiu systemoiu. Naukoiemni tekhnolohii. 2013. № 1 S. 33-36. Rezhym dostupu: http//gov.ua/UJRN/Nt_ 2013_1_9.

[4] Khromchenko V.A., Vozniuk M.M., Pereskokov V.O. Metodychnyi pidkhid shchodo pidtrymannia nalezhnoho rivnia spravnosti optyko-elektronnykh prytsilnykh system litakiv Povitrianykh syl Zbroinykh Syl Ukrainy. Zbirnyk naukovykh prats Derzhavnoho naukovo-doslidnoho instytutu aviatsii. 2022. Vyp. 18 (25), S. 110-116.

[5] Zhdaniuk M.M., Cherednyk M.M., Makarov S.M., Motiakov Yu.M., Shvets S.A. Systemnyi pidkhid do rozrobky zahalnykh vymoh do zasobiv obsluhovuvannia polotiv. Zbirnyk naukovykh prats Derzhavnoho naukovo-doslidnoho instytutu vyprobuvan i sertyfikatsii ozbroiennia ta viiskovoi tekhniky. 2021. Vyp. 4(10). S. 45-55.

[6] S. V. Haievskyi, S. M. Balakireva, D. V. Komarov, V. O. Yavtushenko. Analiz radioelektronnoi systemy litaka yak obiekta prodovzhennia terminu ekspluatatsii. // Zbirnyk naukovykh prats. «Systemy upravlinnia, navihatsii ta zviazku».Poltava 2020. Vyp. 1(59). S. 15-20.

[7] Medynskyi D.V. Optymizatsiia zabezpechennia tekhnolohichnykh protsesiv nazemnoho obsluhovuvannia povitrianykh korabliv aviatsiinoiu nazemnoiu tekhnikoiu dlia pereshkodzhannia zbiinykh sytuatsii v aeroportu. Vcheni zapysky TNU im. V.I. Vernadskoho. Seriia: Tekhnichni nauky Tom 32(71) ch.2. № 1. S. 113-122.

[8] D.O. Shevchuk, D.V. Medynskyi. Porivnialnyi analiz prychyn vynyknennia zbiinykh sytuatsii v aeroportu. Vcheni zapysky TNU imeni V.I. Vernadskoho. Seriia: Tekhnichni nauky. 2020. Tom 31 (70) № 5. S. 254-260.

[9] S.M.Novichonok, O.M. Babich, I.V.Terentieva Problemy standartyzatsii aerodromno-tekhnichnoho zabezpechennia derzhavnoi aviatsii Ukrainy ta shliakhy yikh vyrishennia v umovakh vzaiemodii z NATO ta mizhnarodnoi intehratsii. – Systemy ozbroiennia i viiskova tekhnika. 2020 № 2(62). S. 24-34.

[10] Ye.Iu. Ilenko, M.B. Sushak, P.M. Steshenko Vyrishennia zadachi minimizatsii ryzykiv importozamishchennia v protsesi zavodskoho remontu viiskovoi aviatsiinoi tekhniky v umovakh osoblyvoho periodu. Zbirnyk naukovykh prats Kharkivskoho natsionalnoho universytetu Povitrianykh Syl im. I. Kozheduba. 2020. №3(65) S. 43-49.

[11] M.B. Sushak, M.M. Derevianko, S.O. Fokin. Analiz tekhnolohichnykh sposobiv pokrashchennia yakosti remontu vyrobiv aviatsiinoi tekhniky. Zbroinykh Syl Ukrainy. Zbirnyk naukovykh prats Derzhavnoho naukovo-doslidnoho instytutu aviatsii. 2022. Vyp 18(25). S. 187-196.

[12] V.M. Hryshyn, O.I. Skliar. Vybir klasu modeli i metodu matematychnoho modeliuvannia protsesu tekhnichnoi ekspluatatsii litalnykh aparativ viiskovoho pryznachennia. Zbirnyk naukovykh prats Derzhavnoho naukovo-doslidnoho instytutu aviatsii. 2022. Vyp. 18 (25). S. 134-140.

[13] S.A.Tsybulnyk, B.A.Okhota. Ymytatsyonnoe modelyrovanye ɜlementa obshyvky samoleta pry mnohoochahovom povrezhdenyy / / Visnyk NTU «KPI».Seriia: Pryladobuduvannia. 2018. - Vyp. 55. - S. 93-100.

[14] Matviichuk V.A., Veselovska N.R., Sharhorodskyi S.A. Matematychne modeliuvannia novitnikh tekhnolohichnykh system: Monohrafiia/ V.A. Matviichuk , N.R. Veselovska, S.A.Sharhorodskyi. Vinnytsia. 2021. 193 s.

[15] Subochev O.I. Modeliuvannia vyrobnytstva servisnykh pidpryiemstv vantazhnykh avtomobiliv/ O.I. Subochev, V.Iu. Malyshchuk, O.Ie. Sichko// Zahalnoderzhavnyi mizhvidomchyi naukovo-tekhnichnyi zbirnyk «Konstruiuvannia ta ekspluatatsiia silskohospodarskykh mashyn». Kropyvnytskyi: TsNTU. 2019. Vyp. 49 S. 221-232.

[16] Pavlenko M.A., Rublova R.I., Trofymenko A.O. Model otsinky tekhnichnoho stanu radioelektronnykh elementiv zasobiv vodnoho transportu. Zbirnyk naukovykh prats ΛΌΗΟΣ. 2020. S.89-92. Rezhym dostupu http: // doi.org/ 10.36074/18.09.2020.v1.33.

[17] Bohomiia V.I., Belobrova T.A., Demianenko S.K. Imitatsiina model dostovirnosti kontroliu radioelektronnykh system zasobiv vodnoho transportu. Naukoiemni tekhnolohii. 2019. № 1(41) S. 82-87.

[18] Yeholnikova O.O. Osoblyvosti tekhnichnoi ekspluatatsii i monitorynhu zasobiv avtomatyky na obiektakh zaliznychno-vodnoho spoluchennia. Vcheni zapysky TNU im. V.I.Vernadskoho. Seriia: Tekhnichni nauky. 2020. Tom 31(70) № 5. S. 195-201.

[19] Bondar B.E., Ochkasov A.B., Bondar E.B., Hryshechkyna T.S. Ocheretniuk M.V. Modelyrovanye orhanyzatsyy remonta lokomotyvov metodamy teoryy system massovoho obsluzhyvanyia. Visnyk Dnipropetrovskoho natsionalnoho universytetu zaliznychnoho transportu. 2018 № 5(77). Rezhym dostupu: http://eadnurt.diit.edu.ua>bistream.pdf.

[20] Huzenko V.L., Myronov E.A., Shestopalova O.L. Modelyrovanye protsessov tekhnycheskoho obsluzhyvanyia y remonta tekhnycheskykh sredstv raspredelennoi ynformatsyonnoi systemы. Ynformatsyonno-ɜkonomycheskye aspektы standartyzatsyy y tekhnycheskoho rehulyrovanyia: Nauchnыi ynternet-zhurnal. 2014. № 6(22). Rezhym dostupu:http:// iea.gostimjo.ru>_66_11.pdf.

[21] Upravlinnia ekolohichnoiu bezpekoiu na zaliznychnomu transporti (prykladni aspekty):monohrafiia/ M.D.Katsman, O.I. Zaporozhets, V.K.Myronenko, V.I.Matsiuk, O.V.Tretiakov. Kyiv FOP: Lukianenko V.V., TPK «Orkhideia», 2011. 380 s.

[22] M. D. Katsman, V. K. Myronenko, and V. I. Matsiuk, “Mathematical models of ecologically hazardous rail traffic accidents,” Reliab. Theory Appl., vol. 10, no. 1 (36), 2015, Accessed: nov. 14, 2020. [Online]. Available: https://www.gnedenko.net/Journal/2015/012015/RTA_1_2015-03.pdf.

[23] V. Matsiuk, N. Ilchenko, O. Pryimuk, D. Kochubei, and A. Prokhorchenko, “Risk

assessment of transport processes by agent-based simulation," AIP Conf. Proc., vol. 2557, no. 1, p. 080003, Oct. 2022, doi: 10.1063/5.0105913.

[24] M. D. Katsman, V. K. Myronenko, V. I. Matsiuk, and P. V. Lapin, "Approach to determining the parameters of physical security units for a critical infrastructure facility," Reliab. Theory Appl., vol. 16, no. 1, pp. 71–80, 2021, doi: 10.24412/1932-2321-2021-161-71-80.

[25] V. Matsiuk et al., "Improvement of efficiency in the organization of transfer trains at developed railway nodes by implementing a 'flexible model,'" Eastern-European J. Enterp. Technol., 2019, doi: 10.15587/1729-4061.2019.162143.

[26] A. Prokhorchenko, L. Parkhomenko, A. Kyman, V. Matsiuk, and J. Stepanova, "Improvement of the technology of accelerated passage of low-capacity car traffic on the basis of scheduling of grouped trains of operational purpose," in Procedia Computer Science, 2019, vol. 149, pp. 86–94.

[27] M.D. Katsman, V.I. Matsiuk, V.K. Myronenko MODELING THE RELIABILITY OF TRANSPORT UNDER EXTREME CONDITIONS OF OPERATION AS A QUEUING SYSTEM WITH PRIORITIES. Reliability: Theory & Applications. 2023, June 2(73): 24-38. https://doi.org/10.24412/1932-2321-2023-273-167-179.

MATHEMATICAL MODELING OF AVERAGE WEIGHTED RENYI'S ENTROPY MEASURE

Savita, Rajeev Kumar
•
Department of Mathematics, Maharishi Dayanand University, Rohtak.
savitanain25@gmail.com, profrajeevmdu@gmail.com

Abstract

A weighted entropy measure of information is provided by a probabilistic experiment whose basic events are described by their objective probabilities and some qualitative (objective or subjective) weights. Weighted entropy has also been applied to equity the amount of information and degree of homogeneity related with a partition of data in classes. These measures have tremendous applications and are found to be quite helpful in many fields. In the present paper, a new weighted Renyi's entropy measure is proposed for the discrete distributions when probabilities are unknown and weights are known. The various characteristics of the measure are investigated. The measure is also studied taking into a particular case. In the last, numerical computation and graphical analysis is also done. Based on the graphical analysis, it is concluded that the proposed measure varies with values of weights and is concave in nature. The developed weighted information measure is useful for the discrete distribution when probabilities are unknown and weights are known.

Keywords: Shannon entropy, Renyi entropy, Weighted entropy, Symmetry, Concavity.

I. Introduction

Shannon's [22] entropy is widely prevalent in the study of probabilistic phenomena pertaining to abroad spectrum of problems. It was given by Shannon as a mathematical function to measure the uncertainty involved in a probabilistic experiment. Shannon entropy for a discrete source is defined as follows:

Let X be a probabilistic experiment with sample space x and probability distribution P, where $p(x_i)$ or p_i is the probability of outcome $x_i \in X$. Then the average amount of information is given by

$$H(P) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) = -\sum_{i=1}^{n} p_i \ln p_i \tag{1}$$

Renyi [19] introduced a flexible extension of Shannon entropy and also, it is one parameter generalization of Shannon entropy. In the analysis of quantum systems and measure of randomness, Renyi entropy is mostly used. Renyi defined entropy as

$$H_\alpha(P) = \frac{1}{1-\alpha}\log\frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i}, \quad \alpha \neq 1, \alpha > 0 \tag{2}$$

which is also known as Renyi's entropy measure of order α . Notice that Shannon's entropy measure is the limiting case of Renyi's entropy measure when $\alpha \to 1$.

The probabilistic measure of entropy (1) acquires a number of interesting properties. After the development of this measure, researchers found the potential of the application of this measure as an important quantity in many fields, from probability theory to engineering, ecology, and neuroscience. On the basis of this a large number of other information theoretic measures have also been derived.

Occasionally, standard distributions in statistical modelling are not appropriate for our data and we need to study weighted distribution. This concept has been applied in a variety of statistical fields, including family size analysis, human heredity, world life population study, renewal theory, biomedical and statistical ecology.

Fisher [5] originated the concept of a weighted distribution from his research into the impact of ascertainment methods on frequency estimates. Rao [17,18] expanded on Fisher's core ideas by discussing the need for a unifying idea and identifying several sample scenarios that can be replicated by what he called weighted distributions. The particular case of weighted distribution is the Size-biased distribution. These distributions naturally appear in real-world situations when observations from a sample are recorded with unequal probability. The utility distribution $W = (w_1, w_2, ...w_n)$, where each w_i is a non-negative real number, is proposed by Belis and Guiasu [1] to measure utility aspect of the outcomes.

The applications of the measure to the theory of questionnaires were given by Guiasu and Picard [7]. Longo [12] applied this useful measure to coding theory. Moreover, in many situations, there is need of a measure of uncertainty of a distribution whenever probabilities are unknown and however weights for each value of the random variable are known, i.e. if X be a discrete random variable having values $x_1, x_2, ...x_n$ having weights $w_1, w_2, ...w_n$ respectively but $p_1, p_2, ...p_n$ are unknown. Patsakis et al. [16] gave the applications of the measure in security quantification.

The suitable generalization of classical entropy is the weighted entropy, which has been proposed by Belis and Guiasu [1], Guiasu [6]. A detailed discussion on weighted entropies have been made by Suhov and sekeh [24]. Mahdy [13] studied the weighted entropy measures and its application in Reliability theory and stochastic. Using two weighted entropy measures, Singh et al. [23] provided the applications of Holder's inequality to coding theory. Some other substantial measures of weighted entropy introduced by Kapur [9] are as under:

- $$H_\alpha(P:W) = -\sum_{i=1}^{n} w_i p_i^\alpha \ln p_i, \quad 1/2 \le \alpha \le 1 \tag{3}$$
- $$H_a(P:W) = -\sum_{i=1}^{n} w_i p_i \ln p_i + 1/a \sum_{i=1}^{n} w_i (1+ap_i)\ln(1+ap_i) - w_{min.} \sum_{i=1}^{n} (1+a)\ln(1+a)p_i \tag{4}$$

where $w_{min.} = \min(w_1, w_2, ...w_3)$

Some characterizations and generalizations of the weighted measure have been provided by Longo [12], Hooda and Tuteja [8], Taneja and Hooda [25], Parkash and Taneja [15], Taneja and Tuteja [26], Kapur [9], Kumar et.al. [10, 11], Endo and Kudo [4], Mohammadi [14], Savita and Kumar [21], Bhat and Pundir [2], Sahni and Kumar [20] etc. Thus, weighted measures of information find tremendous applications and are quite helpful to the researchers in many fields.

II. Average Weighted Entropy Measure

Average weighted Renyi's entropy measure is proposed as under:

$$H_1(w) = \frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{n}(w_i)^{\alpha} / \sum_{i=1}^{n} w_i\right) \tag{5}$$

where $w_i > 0$ and $\sum_{i=1}^{n} w_i = w$ (constant).

The important characteristics of the measure (5) are investigated as under:

- It is a continuous and non -increasing function of α.
- It is permutationally symmetric function of $w_1, w_2, ...w_n$, i.e., it does not change when $w_1, w_2, ...w_n$ are permuted among themselves.
- $H_1(w)$ is non-negative for $\alpha < 0$ and negative for $\alpha > 0$.
- Expansible property: This property is satisfied for the measure (5) which states that entropy does not change by the addition of weight with zero value.

  $$H_1(w_1, w_2, ..., w_n, 0) = H_1(w_1, w_2, ..., w_n)$$

  Here we use convention $0^{\alpha}=0$ for all real values of α.
- The maximum value of the function can be obtained by considering following Lagrangian:

  $$L = \frac{1}{1-\alpha}\ln\left(\sum_{i=1}^{n}(w_i)^{\alpha} / \sum_{i=1}^{n} w_i\right) - \lambda\left(\sum_{i=1}^{n} w_i - w\right)$$

  $$\text{Now } \frac{\partial L}{\partial w_1} = \frac{1}{(1-\alpha)}\left[\frac{\alpha w_1^{\alpha-1}}{\sum_{i=1}^{n} w_i^{\alpha}}\right] - \lambda$$

  $$\frac{\partial L}{\partial w_2} = \frac{1}{(1-\alpha)}\left[\frac{\alpha w_2^{\alpha-1}}{\sum_{i=1}^{n} w_i^{\alpha}}\right] - \lambda$$

  Continuing like this

  $$\frac{\partial L}{\partial w_n} = \frac{1}{(1-\alpha)}\left[\frac{\alpha w_n^{\alpha-1}}{\sum_{i=1}^{n} w_i^{\alpha}}\right] - \lambda$$

  $$\text{Now } \frac{\partial L}{\partial w_i} = 0 \Rightarrow \frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial w_3} = ... = \frac{\partial L}{\partial w_n} = 0$$

  $$\frac{1}{(1-\alpha)}\left[\frac{\alpha w_1^{\alpha-1}}{\sum_{i=1}^{n} w_i^{\alpha}}\right] - \lambda = \frac{1}{(1-\alpha)}\left[\frac{\alpha w_2^{\alpha-1}}{\sum_{i=1}^{n} w_i^{\alpha}}\right] - \lambda = ... = \frac{1}{(1-\alpha)}\left[\frac{\alpha w_n^{\alpha-1}}{\sum_{i=1}^{n} w_i^{\alpha}}\right] - \lambda$$

  $$w_1^{\alpha-1} = w_2^{\alpha-1} = w_3^{\alpha-1} = ... = w_n^{\alpha-1}$$

which is possible only if $w_1 = w_2 = w_3 = \ldots\ldots\ldots = w_n$

Also, $\sum_{i=1}^{n} w_i = w \Rightarrow nw_i = w \Rightarrow w_i = \frac{w}{n}$

Thus, the maximum value of function will exist at $w_i = \frac{w}{n}$ and is given by

$$H_1(w) = \frac{1}{(1-\alpha)} \ln \frac{\sum_{i=1}^{n} \left(\frac{w}{n}\right)^{\alpha}}{w} = \frac{\ln w^{\alpha-1} - \ln\left(n^{\alpha} - 1\right)}{(1-\alpha)}$$

- Additive property: The measure (5) is additive in nature since when

$$H_{n1}(w) = \frac{1}{1-\alpha} \ln\left(\sum_{i=1}^{n} (w_i)^{\alpha} / \sum_{i=1}^{n} w_i\right) \quad \text{and} \; H_{m1}(w') = \frac{1}{1-\alpha} \ln\left(\sum_{j=1}^{m} (w_j')^{\alpha} / \sum_{j=1}^{m} w_j'\right)$$

$$\text{Then} \quad H_{n1+m1}(w \cup w') = \frac{1}{(1-\alpha)} \ln\left(\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{m} (w_i w_j')^{\alpha}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{m} w_i w_j'}\right) = H_{n1}(w) + H_{m1}(w')$$

- Concave property: Since for the measure (5),

$$H_1(w) = \frac{1}{1-\alpha} \ln\left(\frac{\sum_{i=1}^{n} (w_i)^{\alpha}}{\sum_{i=1}^{n} w_i}\right)$$

$$H_1'(w) = \frac{1}{(1-\alpha)} \frac{\alpha\left(\sum_{i=1}^{n} (w_i)^{\alpha-1}\right)}{\sum_{i=1}^{n} w_i^{\alpha}}$$

$$H_1''(w) = -\frac{\alpha\left(\sum_{i=1}^{n} (w_i)^{\alpha-2}\right)}{\left(\sum_{i=1}^{n} w_i^{\alpha}\right)} - \frac{\alpha^2\left(\sum_{i=1}^{n} (w_i)^{\alpha-1}\right)^2}{(1-\alpha)\left(\sum_{i=1}^{n} w_i^{\alpha}\right)^2}$$

Thus, the measure (5) is concave upward for $\alpha > 1$.

I. Particular Cases

- when $w_i = p_i, \quad 0 \le p_i \le 1$

then measure (5) becomes Renyi type entropy measure.

- when $w_i = p_i, \quad 0 \le p_i \le 1$ and $\alpha \to 1$

then measure (5) becomes Shannon type entropy measure.

- when $\alpha \to 1$

then measure (5) reduces to average weighted Shannon's entropy measure.

III. Numerical Computation and Graphical Analysis

For computation of various values of the measure $H_1(w)$ given in (5), three values of weights w_1, w_2 and w_3 such that $w = w_1 + w_2 + w_3$ are considered. Total weights w are assumed to take values 50, 100 and 150 and $\alpha = 3$. The computed values of these cases are presented in table 1. Various graphs are plotted for the obtained values of $H_1(w)$ w.r.t. different weights to study the behavior of the information measure (5).

Table 1: *Average weighted entropy measure*

W=150				W=100				W=50			
w1	w2	w3	$H_1(w)$	w1	w2	w3	$H_1(w)$	w1	w2	w3	$H_1(w)$
5	50	95	-4.3936	5	30	65	-4.0061	5	10	35	-3.3900
10	50	90	-4.3241	10	30	70	-3.8999	10	10	30	-3.1815
15	50	85	-4.2536	15	30	55	-3.7923	15	10	25	-2.9957
20	50	80	-4.1832	20	30	50	-3.6889	20	10	20	-2.9145
25	50	75	-4.1148	25	30	45	-3.5993	25	10	15	-2.9957
30	50	70	-4.0508	30	30	40	-3.5366	30	10	10	-3.1815
35	50	65	-3.9948	35	30	35	-3.5139	35	10	5	-3.3900
40	50	60	-3.9505	40	30	30	-3.5366				
45	50	55	-3.9219	45	30	25	-3.5993				
50	50	50	-3.9120	50	30	20	-3.6889				
55	50	45	-3.9219	55	30	15	-3.7923				
60	50	40	-3.9505	60	30	10	-3.8999				
65	50	35	-3.9948	65	30	5	-4.0061				
70	50	30	-4.0508								
75	50	25	-4.1148								
80	50	20	-4.1832								
85	50	15	-4.2536								
90	50	10	-4.3241								
95	50	5	-4.3936								

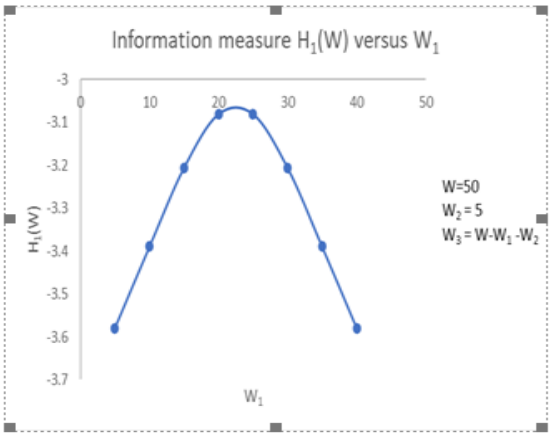


Figure 1: *Information measure* $H_1(w)$ *w.r.t.* w_1 *for* $w = 50$ *and* $w_2 = 5$

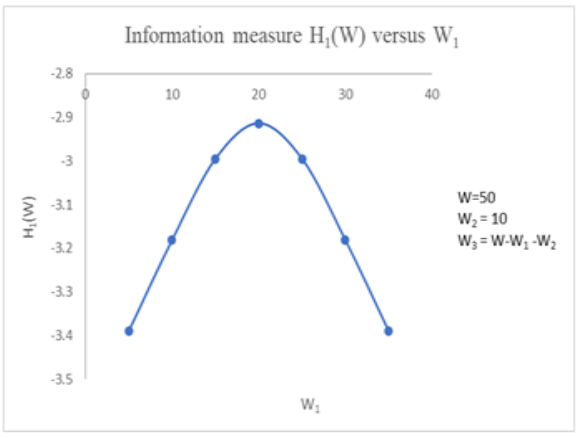


Figure 2: *Information measure* $H_1(w)$ *w.r.t.* w_1 *for* $w = 50$ *and* $w_2 = 10$

Figure 1 and Figure 2 depict the graph between $H_1(w)$ and w_1 for different values of w_i, total weight w = 50 and we have fixed w_2 = 5,10. From the graphs, it can be seen that $H_1(w)$ increases as weights increases up to $w_i = \frac{w}{n}$, and thereafter it decreases.

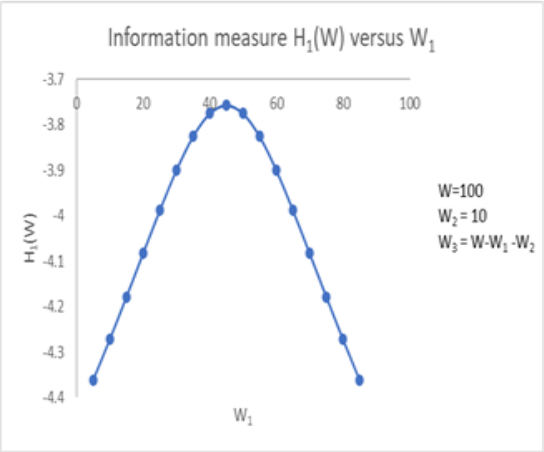


Figure 3: *Information measure* $H_1(w)$ *w.r.t.* w_1 *for* $w = 100$ *and* $w_2 = 10$

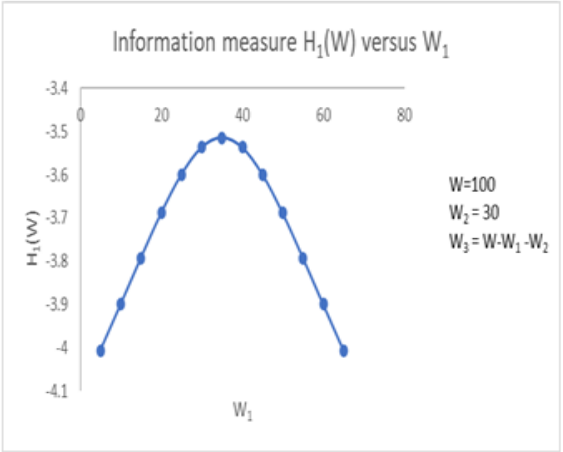


Figure 4: *Information measure* $H_1(w)$ *w.r.t.* w_1 *for* $w = 100$ *and* $w_2 = 30$

Figure 3 and Figure 4 depict the graph between $H_1(w)$ and w_1 for different values of w_i, total weight w = 100 and we have fixed w_2 = 10, 20, 30. From the graphs, it can be seen that $H_1(w)$ increases as weights increases up to $w_i = \frac{w}{n}$, and thereafter it decreases.

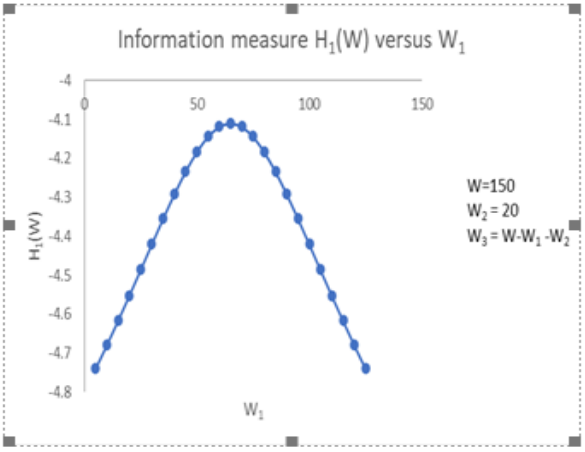


Figure 5: *Information measure* $H_1(w)$ *w.r.t.* w_1 *for* $w = 150$ *and* $w_2 = 20$

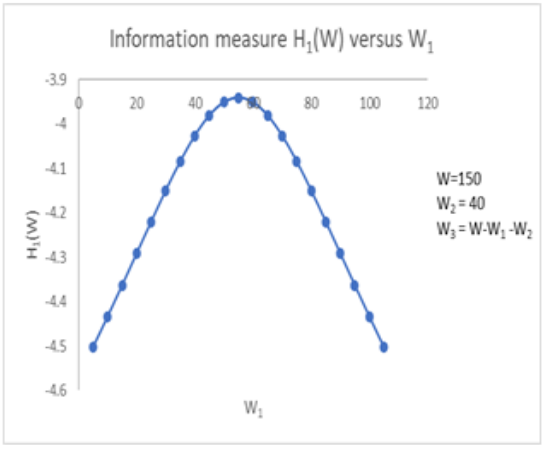


Figure 6: *Information measure* $H_1(w)$ *w.r.t.* w_1 *for* $w = 150$ *and* $w_2 = 40$

Figure 5 and Figure 6 depict the graph between $H_1(w)$ and w_1 for different values of w_i, total weight w = 150 and we have fixed w_2 = 10, 20, 30, 40, 50. From the graphs, it can be seen that $H_1(w)$ increases as weights increases up to $w_i = \frac{w}{n}$, and thereafter it decreases.

IV. Conclusion

The proposed weighted Renyi's measure varies with values of weights and is concave in nature. In case, weights of the distribution are their probabilities, the average weighted Renyi entropy measure reduces to Renyi 's entropy. In the weighted sense, this is in fact generalizations of the Shannon [22] entropy and Burg [3] entropy. The developed weighted information measure is useful for the discrete distributions when probabilities are unknown and weights are known.

References

[1] Belis, M. and Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems, *IEEE Transactions on Information Theory*, 14:593-594. https://doi.org/10.1109/TIT.1968.1054185

[2] Bhat, V.A. and Pundir, S. (2023). Weighted intervened exponential distribution as a lifetime distribution. *Reliability: Theory & Applications.* 2(73): 24-38. https://doi.org/10.24412/1932-2321-2023-273-253-266

[3] Burg, J.P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. *In: Childrers, D.G. (ed), Modern Spectral Analysis*,130-131.

[4] Endo, T. and Kudo, M. (2013). Weighted naïve Bayes classifiers by Renyi entropy. *J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.), Springer-Verlag Berlin Heidelberg, CIARP, Part I, LNCS 8258,* 149–156.

[5] Fisher, R.A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *The Annals of Eugenics*, 6: 13-25.

[6] Guiasu, S. (1986). Grouping data by using the weighted entropy. *Journal of Statistical Planning and Inference*, 15:63-69. https://doi.org/10.1016/0378-3758(86)90085-6

[7] Guiasu, S. and Picard, C.F. (1971). Borne in fericturde la longuerur utile de certains code. *Comptes Rendus Mathematique Academic Des Sciences Paris*, 273: 248-251.

[8] Hooda, D.S. and Tuteja, R.K. (1981). Two generalized measures of useful information. *Information Sciences*, 23: 11-24.

[9] Kapur. J.N. (1995). Measures of information and their applications. *Wiley Eastern,* New York.

[10] Kumar, R., Kumar, S. and Kumar, A. (2010). A new measure of probabilistic entropy and its properties. *Applied Mathematical Sciences*, 4(28): 1387-1394**.**

[11] Kumar, R., Kumar, S. and Kumar, A. (2010). Some new information theoretic measures based upon statistical constants. *Aryabhatta Journal of Mathematics and Informatics*, 2(2): 329-338.

[12] Longo, G. (1972). Quantitative-qualitative measure of information. *Springer-Verlag*, New York.

[13] Mahdy, M. (2018). Weighted entropy measure: a new measure of information with its properties in reliability theory and stochastic orders. Journal of Statistical Theory and Applications. 17(4):703-718. DOI:10.2991/jsta.2018.4.17.11

[14] Mohammadi, U. (2015). Weighted entropy function as an extension of the Kolmogorov-Sinai entropy. *U.P.B. Sci. Bull.,* Series A, 77(4): 117-122.

[15] Parkash, O. and Taneja, H.C. (1986). Characterization of quantitative-qualitative measure of inaccuracy for discrete generalized probability distributions. *Comm. Statistics theory and Methods*, 15(12): 3763-3772.

[16] Patsakis, C., Mermigas, D., Pirounias, S. and Chondrokoukis, G. (2013). The role of weighted entropy in security quantification. *International Journal of Information and Electronics Engineering,* 3(2):156-159.

[17] Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment, in Classical and Contagious Discrete Distributions. *G.P. Patil, ed., Pergamon Press and Statistical Publishing Society, Calcutta*,320–332.

[18] Rao, C.R. (1985). Weighted distributions arising out of methods of ascertainment, in *A* Celebration of Statistics, *A.C. Atkinson & S.E. Fienberg, eds, Springer-Verlag*, New York, Chapter 24:543–569.

[19] Renyi, A. (1961). On measures of entropy and information. *Proceedings IV Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 20-30 June 1961*,1: 547-561.

[20] Sahni, P. and Kumar, R. (2023). Huntsberger type shrinkage entropy estimator for variance of normal distribution under linex loss function. *Reliability: Theory & Applications.* 1(72): 491-499. https://doi.org/10.24412/1932-2321-2023-172-491-499

[21] Savita and Kumar, R. (2019). On A New Weighted Entropy Measure. *Journal of Advanced research in dynamical and control systems.* 11(01)-special issue:1264-1271.

[22] Shannon, C.E (1948). A mathematical theory of communication, *Bell System Technical Journal*, 27:379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

[23] Singh, R.P., Kumar, R. and Tuteja, R.K. (2003). Applications of holder's inequality in information theory. *Information Sciences*, 152: 145-154.

[24] Suhov, Y., Sekeh, S.Y. (2015). Weighted cumulative entropies: An extension of CRE and CE. *arXiv: 1507.0705*[v][cs.IT]. https://doi.org/10.48550/arXiv.1507.07051

[25] Taneja, H.C. and Hooda, D.S. (1983). On characterization of generalized measure of useful information, *Sooch. Jr. Math*, 9: 221-230.

[26] Taneja, H.C. and Tuteja, R.K. (1986). Characterization of quantitative-qualitative measure of inaccuracy, *Kybernetika,* 22: 393-402.

RT&A, No.3 (74)
Volume 18, September 2023

REDUCTION IN WAITING TIME OF SINGLE SERVER MARKOVIAN QUEUING ENCOURAGED ARRIVAL MODEL

Ismailkhan Enayathulla Khan, Rajendran Paramasivam*

•

Department of Mathematics, School of Advanced Sciences,
Vellore Institute of Technology, Vellore, India - 632014.
Ismailkhan.e@vit.ac.in
Correspondence email: prajendran@vit.ac.in

Abstract

There are several other methods for improving efficiency in a control chart. The use of a control chart alone is not advised. Other process improvement methods should always be used in addition to control charts. To trace the evolution of a process variable across time, use a control chart. The variable is applicable to all industries, including service, manufacturing, non-profit, and healthcare. It illustrates how a process variable changes over time and provides information on the kinds of variations that deal with ongoing improvement. Having a solid understanding of variation is necessary for effective control chart usage. Queuing models with constant or variable sizes are extensively used in the modeling of road and transport systems, sophisticated information and computer systems, and inventory replenishment systems. The control chart technique helps in tracking the performance of these queues, because of the single-server Markovian queue with encouraged arrival (SSMQEA model) the company which is running with fewer customers can increase the number of customers and hence the company finance level increase also this SSMQEA method will improve the points in share market. The major measurable performance characteristics of any queuing system are average queue length and average waiting time. Control limits are defined in this study for the $M^{[X]}/M/1$ encouraged arrival queuing model where the batch size follows a geometric distribution. To highlight its uses, numerical observations are also included. Little' s law is also satisfied.

Keywords: Encouraged arrival, batch size, central limit, upper control limit, Lower control limit, Little's law

1. Introduction

A control chart has plenty of other techniques for increasing efficiency. A control chart should not be used in isolation. Along with control charts, other process improvement techniques should always be employed. A control chart is used to track the progression of a process variable across time. The variable may occur in any sort of business or organisation, encompassing service, manufacturing, non-profit and healthcare. It depicts the process variable over time and informs the sort of variation that are dealing with continuous improvement. Understanding variance is essential for efficiently using control charts.

In general, queueing models assume that customers arrive at the service facility individually. Unfortunately, this assumption is broken in many real-world queueing scenarios. People arriving at a post office, ships coming in a convoy at a port, people

attending a wedding reception are all instances of queuing scenarios in which consumers come in groups [1].

The bulk arrival queuing model from a Bayesian point of view is considered in [2]. Statistical process control is a quality control technique that was first established to monitor manufacturing operations. Investigation on server abandonment dynamics in [3]. For quality control in industrial enterprises, [4] provides a number of Shewhart control chart solutions. For the M/M/S queuing model's random queue length, [5] built a control chart.

While [6] created Shewhart control charts for the G/G/S queuing system utilizing. The M/M/1 queuing model's random queue length was controlled using a control chart made using the stacked variance method [7].

The number of customers in a M/Ek/1 wait was examined by [8] using a control chart approach. M/M/1/N queuing systems with encouraged arrival were examined by [9 and 10]. The control diagram for the M^x/M/1 queuing system was explored in [11].

2. Model Recitation

Now we describe the single-server Markovian queue with encouraged arrival (SSMQEA model) as follows:

- The arrivals occur one at a time in line with the Poisson process with parameter λ (1+η), where indicates the percentage change in the number of customers calculated from preceding or clear vision. For example, if a business previously gave discounts and the percentage change in number of consumers was 10% or 30%, then η = 0.1 to η = 0.3, respectively.
- Let p_n be the probability that the system now contains n customers.
- Let d_n represent the batch size distribution.

The steady-state equations that control this model are as follows:

$0=-(\lambda(1+\eta)+\mu)p_n+\mu p_{n+1}+\lambda(1+\eta)\sum_{\varepsilon=1}^{n}p_{n-\varepsilon}d_\varepsilon \quad (n\geq 1),$

$0=-(\lambda(1+\eta)p_0+\mu p_1.$ (1)

The generating function technique may be used to solve the system of equations (1).

Define the generating functions for the steady state probability and the batch size distribution as follows:

$P(a)=\sum_{n=0}^{\infty}p_n a^n, \quad |a|\leq 1,$

$D(a)=\sum_{n=0}^{\infty}d_n a^n, \quad |a|\leq 1,$

Equation (1) is obtained by summing and multiplying by the necessary powers of z.

$0=-\lambda(1+\eta)\sum_{n=0}^{\infty}p_n a^n-\mu\sum_{n=1}^{\infty}p_n a^n+\frac{\mu}{a}\sum_{n=1}^{\infty}p_n a^n+\lambda(1+\eta)\sum_{n=1}^{\infty}\sum_{\varepsilon=0}^{\infty}p_{n-\varepsilon}d_\varepsilon a^n,$ (2)

Contemplate$\sum_{n=1}^{\infty}\sum_{\varepsilon=0}^{\infty}p_{n-\varepsilon}d_\varepsilon a^n=\sum_{\varepsilon=1}^{\infty}d_\varepsilon a^\varepsilon\sum_{n=\varepsilon}^{\infty}p_{n-\varepsilon}a^{n-\varepsilon}=d(a)p(a).$ (3)

Equation (2) becomes Equation (3).

$0=-\lambda(1+\eta)p(a)-\mu(p(a)-p_0)+\lambda(1+\eta)d(a)p(a).$

Solving for p(a), we get

$P(a)=\frac{\mu p_0(1-a)}{\mu(1-a)-\lambda(1+\eta)a(1-d(a))},|a|\leq 1.$ (4)

The complimentary batch size possibilities' production function P(X>X)=1-$d_x=d_x^{\sim}$ is given by $d_x^{\sim}=\sum_{n-1}^{\infty}d_x^{\sim}a^n=\frac{1-d(a)}{1-a}$.

We took R=$\frac{\lambda(1+\eta)}{\mu}$, equation (4) yields

P(a)=$\frac{p_0}{1-Rad_x^{\sim}(a)}$.

Clearly, $d_x^{\sim}(1)=E(x) and\ d_x^{\sim\prime}(1)=\frac{E(x(x-1))}{2}$.

Making use of the normalising conditions, we got $p_0=1-\rho$, where $\rho=\frac{\lambda(1+\eta)}{\mu}E(x)$.

If K_s and K_q are the number of consumers in the system and the queue, respectively, then

$K_s = \frac{\frac{\lambda(1+\eta)}{\mu}E(x)+RE(x^2)}{2(1-\frac{\lambda(1+\eta)}{\mu}E(x))}$,

and $K_q = K_s - \frac{\lambda(1+\eta)}{\mu}E(x)$.

Assume that the number of consumers in any arriving batch is geometrically distributed with parameter β. The probability mass function of batch size is then calculated

$d_x = (1-\beta)\beta^{x-1}, 0 < \beta < 1.$

Then d(a)=$\frac{a(1-\beta)}{1-\beta a}$,

and

E(x)=$\frac{1}{1-\beta}$ $with\ \rho = \frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}$. (5)

From equation (1) to (5), we get

P(a)=$\left((1-\frac{\lambda(1+\eta)}{\mu}E(x)) \sum_{n=0}^{\infty}\left(\beta+(1-\beta)\frac{\lambda(1+\eta)}{\mu}E(x)\right)^n - \sum_{n=0}^{\infty}\left(\beta+(1-\beta)\frac{\lambda(1+\eta)}{\mu}E(x)\right)^n z^{n+1}\right)$

"a" on both sides when performance is compared to results in

$p_n = \frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})(1-\beta)(\beta+(1-\beta)\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})^{n-1}, n > 0.$

Thus, the value A_0^{-1} is evaluated.

To obtained the equations mean and variance

Let $K_s$represent the total number of consumers in the system (both in queue and in service).

The anticipated number of clients in the system is then

$$E(K_s) = \left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})} \right). \quad (6)$$

And the variance of the number of consumers in the system is,

$$Var(K_s) = \left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}(1+\beta\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right))}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})^2*(1-\beta)^2} \right). \quad (7)$$

According to the concept that the number of consumers in the system maintains a normal distribution, the parameters of the control chart are given by

$$\text{Upper Control Limit=E}\left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})} \right) + 3\sqrt{\left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}(1+\beta\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right))}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})^2*(1-\beta)^2} \right)}, \quad (8)$$

$$\text{Central Limit=E}\left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})} \right), \quad (9)$$

$$\text{Lower Contral Limit= E}\left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})} \right) - 3\sqrt{\left(\frac{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}(1+\beta\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right))}{(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta})^2*(1-\beta)^2} \right)}. \quad (10)$$

Using (6) and (7) in (8), the control chart parameters for the $M^{[X]}$/M/1 queuing model are determined.

$$\text{Central Limit}= \frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\Bigg/\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right)(1-\beta),$$

$$\text{Upper Central Limit} = \frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}+3\sqrt{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\left(1+\beta\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right)\right)}\Bigg/(1-\beta)\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right),$$

$$\text{Lower Central Limit}= \frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}-3\sqrt{\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\left(1+\beta\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right)\right)}\Bigg/(1-\beta)\left(1-\frac{\frac{\lambda(1+\eta)}{\mu}}{1-\beta}\right).$$

The anticipated number of waiting units

i. $L_q = \frac{\rho\lambda(1+\eta)}{\mu-\lambda(1+\eta)}$.

The average number of occupied (serviced) units

ii. $L_s = \frac{\lambda(1+\eta)}{\mu-\lambda(1+\eta)}$.

The Service time estimate.

iii. $W_s = \frac{1}{\mu-\lambda(1+\eta)}$.

The expected Waiting time in line is

iv. $W_q = \frac{\rho}{\mu-\lambda(1+\eta)}$.

3. Numerical Illustration

The situation that a system encountered when an organization announced incentives and discounts gave origin to the expression "encouraged arrivals." The modern application of queuing theory to consumer behavior encouraged arrivals. The "essential component" that distinguishes control charts from a conventional line graph or run chart is control limits. Your data is used to calculate control limits. They are sometimes mistaken with specification restrictions offered by your customer. A UCL can be described as an acceptable range of values for particular parameters.

The performance of the queuing system is analyzed numerically with reference to the parameters, and LCL values are negative for the specified parameter values $\lambda(1+\eta)$, μ *and* β they are treated as zero and are not displayed as a distinct column in the table.η represent discounts values 10% to 30% of the table and figure.

The table displays the parameters for the control chart and traffic intensity for the number of customers in the queuing system for various values of $\lambda(1+\eta)$, μ and β

Table 1: *We provide Encouraged arrival 10% discount $M^{[X]}/M/1$ Control chart in the queuing system*

S.NO	$\lambda(1+\eta)$	μ	β	ρ	CL	UCL
1	4.4	10	0.15	0.5177	1.2630	6.6994
2	4.4	10	0.16	0.5238	1.3095	6.9399
3	4.4	10	0.17	0.5301	1.3592	7.1789
4	4.4	10	0.18	0.5366	1.4125	7.4329
5	4.4	15	0.15	0.3447	0.6188	3.9329
6	4.4	15	0.16	0.3488	0.6377	4.0411
7	4.4	15	0.17	0.353	0.6575	4.1549
8	4.4	15	0.18	0.3573	0.6779	4.2721
9	4.4	20	0.15	0.2588	0.4108	2.9647
10	4.4	20	0.16	0.2619	0.4224	3.0409
11	4.4	20	0.17	0.2651	0.4347	3.1206
12	4.4	20	0.18	0.2683	0.4472	3.2024
13	6.6	10	0.15	0.7765	4.0874	18.2359
14	6.6	10	0.16	0.7857	4.3647	19.3911
15	6.6	10	0.17	0.7951	4.6772	20.6823
16	6.6	10	0.18	0.8049	5.0313	22.1489
17	6.6	15	0.15	0.5177	1.2628	6.7521
18	6.6	15	0.16	0.5238	1.3095	6.9399
19	6.6	15	0.17	0.5301	1.3592	7.1793
20	6.6	15	0.18	0.5366	1.4125	7.4329
21	6.6	20	0.15	0.3882	0.7466	4.5029
22	6.6	20	0.16	0.3929	0.7705	4.6329
23	6.6	20	0.17	0.3976	0.7952	4.7676
24	6.6	20	0.18	0.4024	0.8213	4.9086

Table 2: *We provide encouraged arrival 10% discounts Little's Law verification table*

S.NO	$\lambda(1+\eta)$	μ	ρ	L_s	W_s	$L_s = \lambda(1+\eta)W_s$
1	4.4	10	0.44	0.7857	0.1785	0.7857
2	4.4	15	0.29	0.415	0.0943	0.415
3	4.4	20	0.22	0.282	0.0641	0.282
4	6.6	10	0.66	1.9411	0.2941	1.9411
5	6.6	15	0.44	0.7857	0.119	0.7857
6	6.6	20	0.33	0.4925	0.0746	0.4925

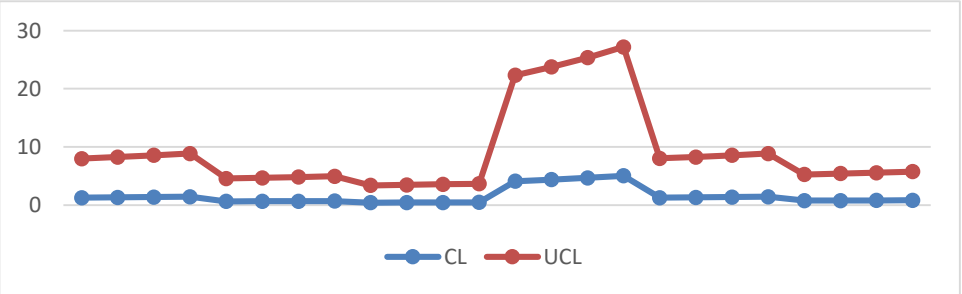


Figure 1: *We provide Pictorial representation encouraged arrival 10% discounts Control Chart.*

From the figure 1, with 10% discounts of encouraged arrival, the number of arrival is more the Poisson arrival Process [11].

Table 3: *We provide Encouraged arrival 20% discount $M^{[X]}/M/1$ Control chart in the queuing system.*

$\boldsymbol{\lambda}$	$\boldsymbol{\mu}$	$\boldsymbol{\beta}$	$\boldsymbol{\eta}$
$\mathbf{4, 6}$	10,15,20	0.15,0.16,0.17,0.18	$\mathbf{0.2}$ ***or*** $\mathbf{20\%}$

S.NO	$\lambda(1+\eta)$	μ	β	ρ	CL	UCL
1	4.8	10	0.15	0.5647	1.5262	7.8148
2	4.8	10	0.16	0.5714	1.5874	8.0995
3	4.8	10	0.17	0.5783	1.6522	8.3998
4	4.8	10	0.18	0.5854	1.7219	8.7209
5	4.8	15	0.15	0.3765	0.7104	4.3422
6	4.8	15	0.16	0.3809	0.7326	4.4656
7	4.8	15	0.17	0.3855	0.7559	4.594
8	4.8	15	0.18	0.3902	0.7804	4.7286
9	4.8	20	0.15	0.2824	0.4629	3.2137
10	4.8	20	0.16	0.2857	0.4762	3.2975
11	4.8	20	0.17	0.2892	0.4902	3.3855
12	4.8	20	0.18	0.2927	0.5047	3.4759
13	7.2	10	0.15	0.8471	6.5177	28.0118
14	7.2	10	0.16	0.8572	7.1463	30.5638
15	7.2	10	0.17	0.8675	7.8877	33.5764
16	7.2	10	0.18	0.878	8.7765	37.1868
17	7.2	15	0.15	0.5647	1.5263	7.8149
18	7.2	15	0.16	0.5714	1.5871	8.0982
19	7.2	15	0.17	0.5783	1.6522	8.4005
20	7.2	15	0.18	0.5854	1.7219	8.7209
21	7.2	20	0.15	0.4235	0.8642	5.0166
22	7.2	20	0.16	0.4286	0.8929	5.1677
23	7.2	20	0.17	0.4337	0.9227	5.3242
24	7.2	20	0.18	0.439	0.9544	5.4888

Table 4: *We provide encouraged arrival 20% discounts Little's law verification table*

S.NO	$\lambda(1+\eta)$	M	P	L_s	W_s	$L_s = \lambda(1+\eta)W_s$
1	4.8	10	0.48	0.923	0.1923	0.923
2	4.8	15	0.32	0.4705	0.098	0.4705
3	4.8	20	0.24	0.3157	0.0657	0.3157
4	7.2	10	0.72	2.5714	0.3571	2.5714
5	7.2	15	0.48	0.923	0.1282	0.923
6	7.2	20	0.36	0.5625	0.0781	0.5625

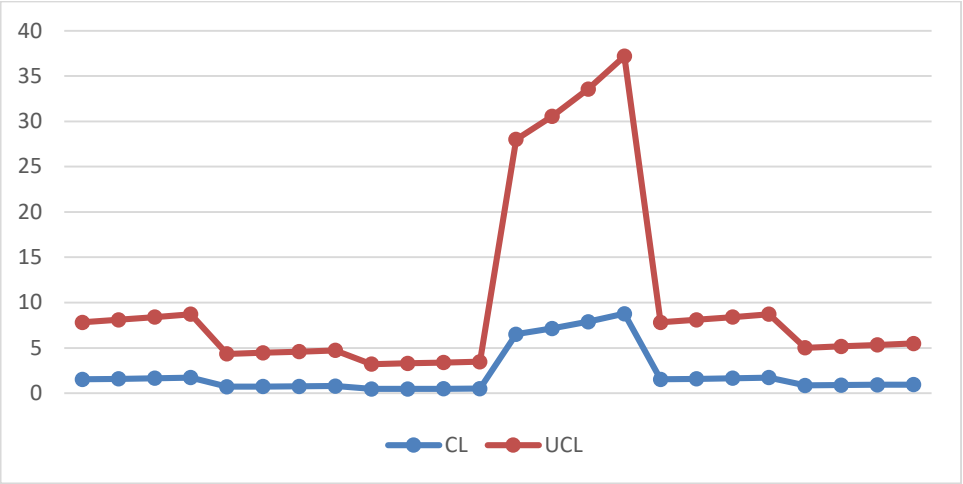


Figure 2: *We provide Pictorial representation encouraged arrival 20% discounts Control Chart.*

From the figure 2, with 20% discounts of encouraged arrival, the number of arrival is more the Poisson arrival Process [11].

Table 5: *We provide Encouraged arrival 30% discount* $M^{[X]}/M/1$ *Control chart in the queuing system*

λ	μ	β	η
4, 6	10,15,20	0.15,0.16,0.17,0.18	**0.3 *or* 30%**

S.NO	$\lambda(1+\eta)$	μ	β	ρ	CL	UCL
1	5.2	10	0.15	0.6118	1.8541	9.1694
2	5.2	10	0.16	0.6191	1.9353	9.5331
3	5.2	10	0.17	0.6265	2.0209	9.9201
4	5.2	10	0.18	0.6289	2.0667	10.1416
5	5.2	15	0.15	0.4082	0.8115	4.7875
6	5.2	15	0.16	0.4131	0.8379	4.9283
7	5.2	15	0.17	0.4181	0.8656	5.0759
8	5.2	15	0.18	0.4232	0.8946	5.2298
9	5.2	20	0.15	0.3059	0.5185	3.4736
10	5.2	20	0.16	0.3095	0.5336	3.5752
11	5.2	20	0.17	0.3132	0.5495	3.6622
12	5.2	20	0.18	0.317	0.5662	3.7628
13	7.8	10	0.15	0.9177	13.1175	54.4512
14	7.8	10	0.16	0.9286	15.4818	63.9578
15	7.8	10	0.17	0.9397	18.7756	77.1462
16	7.8	10	0.18	0.9512	23.7756	97.2152
17	7.8	15	0.15	0.6118	1.8541	9.1694
18	7.8	15	0.16	0.6191	1.9353	9.5331
19	7.8	15	0.17	0.6265	2.0209	9.9201
20	7.8	15	0.18	0.6289	2.0667	10.1416
21	7.8	20	0.15	0.4588	0.9974	5.5907
22	7.8	20	0.16	0.4643	1.0318	5.7658
23	7.8	20	0.17	0.4699	1.0679	5.9479
24	7.8	20	0.18	0.4756	1.106	6.1394

Table 6: *We provide encouraged arrival 30% discounts Little's law verification table*

S.NO	$\lambda(1+\eta)$	μ	P	L_s	W_s	$L_s = \lambda(1+\eta)W_s$
1	5.2	10	0.52	10.833	0.2083	10.833
2	5.2	15	0.34	0.5306	0.102	0.5306
3	5.2	20	0.26	0.3513	0.0675	0.3513
4	7.8	10	0.78	3.545	0.4545	3.545
5	7.8	15	0.52	1.0833	0.1388	1.0833
6	7.8	20	0.39	0.6393	0.0819	0.6393

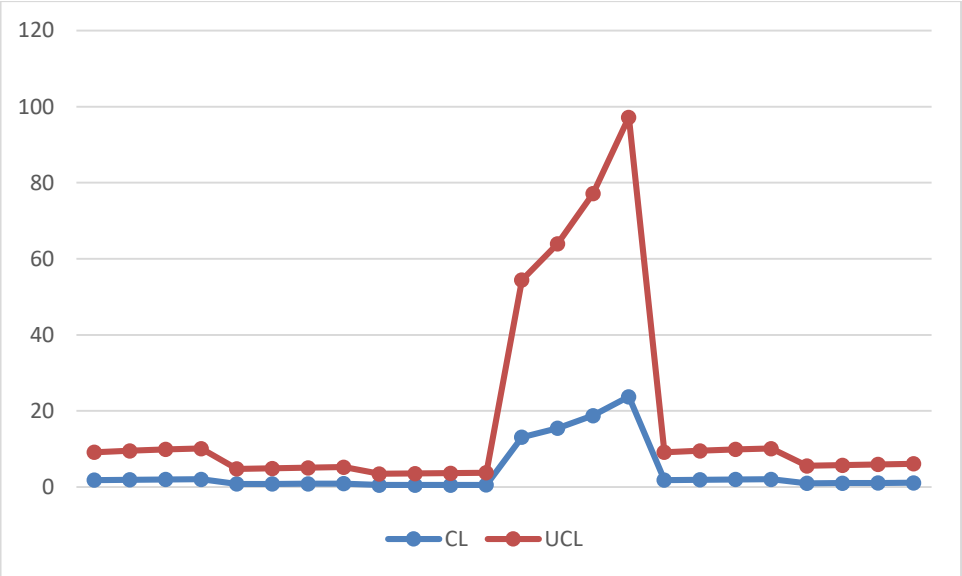


Figure 3: *We provide Pictorial representation encouraged arrival 30% discounts Control Chart.*

From the figure 3, with 30% discounts of encouraged arrival, the number of arrivals is more the Poisson arrival Process [11].We have identified that as the discounts rate increases the arrival rate also increases with SSMQEA model.

4. Result and Discussion

In this model, encouraged arrival, with a maximum 30% discount applicable in this model. Because the system size maximum has increased in this research model, we can expect a maximum profit when we apply this research concept to share markets. The average waiting time and the anticipated maximum waiting time both decrease with an increase in service rate and a constant encouraged arrival rate. The average wait time and anticipated maximum wait time are reduced with an increase in servers.

5. Conclusion

The model provided here has practical applications in systems such as manufacturing, telephony, share markets, and computer networks. From the figure 3, with 30% discounts of encouraged arrival, the number of arrivals is more the Poisson arrival Process [11]. Because the maximum system size in this research model has risen. In this SSMQEA model the company which are running with less customers can increase the number of customers and hence the company finance level increase also this SSMQEA method will improve the points in share market.

References

[1] Gross, D. and Harris, M. Fundamentals of queueing theory, 5th edition, John Wiley and Sons, Inc, 1998.
[2] Armero, C. and Conesa, D. (2000). Prediction in Markovian Bulk arrival queues. *Queueing systems*, 34:327-350.
[3] Down, D.G. Koole, D. and Lewis, M.E. (2011). Dynamic control of a single server with abandonments. *Queueing systems*, 67:63-90.
[4] Montgomery, D.C. Introduction to statistical quality control, 5th edition, John Wiley & Sons, Inc, 2005.
[5] Shore, H. (2000). General control charts for attributes. *IIE transactions*, 32:1149-1160.
[6] Shore, H. (2006). Control charts for the queue length in a G/G/S System. *IIE Transactions*, 38:1117-113.
[7] Khaparde, V. and Dhabe, S.D. (2010). Control chart for random queue length N for (M/M/1): (∞/FCFS) queueing model. *International Journal of Agricultural and Statistical sciences*, 1:319-334.
[8] Poongodi, T. and Muthulakshmi, S. (2012). Random queue length control chart for (M/Ek/1): (∞/FCFS) queueing model. *International Journal of Mathematical Archive*, 3:3340-3344.
[9] Som, B.K. and Seth, S. (2017). An M/M/1/N Queuing system with Encouraged Arrivals. *Global Journal of Pure and Applied Mathematics*, 13:3443-3453.
[10]Khan,I.E. Paramasivam, R. (2022). Reduction in waiting time in an M/M/1/N encouraged arrival queue with feedback, balking and maintaining of reneged customers. *Symmetry*, 14:1743.
[11] Poongodi, T. and Muthulakshmi, S. (2014). Control chart for number of customers in the system $M^{[X]}$/M/1 queuing system. *International Journal of Innovative Research in Science Engineering and Technology*, 3:1-3.

THEORY AND APPLICATIONS OF THE ALPHA POWER TYPE II TOPP-LEONE- GENERATED FAMILY OF DISTRIBUTIONS

Jacob C. Ehiwario[1]
John N. Igabari[2]
Peter E. Ezimadu[3]

[1] Department of Statistics, University of Delta, Agbor, Delta State, Nigeria.
[2,3] Department of Mathematics, Delta State University, Abraka, Nigeria.

jacobehiwario@gmail.com[1]
jn_igabari@delsu.edu.ng[2]
peterezimadu@yahoo.com[3]

Abstract

This paper introduces a composition of two single parameter generalized family of distributions: the alpha power transform and type II Topp-Leone-G families of distributions. Some basic mathematical treatments of the family of distributions are studied. The parameter estimates of the proposed family of distributions are derived via maximum likelihood estimation method and a Monte Carlo simulation study was conducted to examine the asymptotic behaviour of the parameter estimates of sub-model belonging to the proposed family of distributions. To illustrate the applicability of the proposed family of distributions in real world data fittings, two data sets consisting of the daily recovery and mortality rates of Covid-19 patients in Nigeria, from May 1 to June 30, 2020, was employed. The APTIITLK distribution arising from the proposed family of distributions, alongside with some bounded non-nested distributions was used to fit the two data sets and results obtained from the analysis clearly revealed that the APTIITLK distribution outperformed all the non-nested distributions used in fitting the two data sets. Some informative graphical plots for goodness of fit test were investigated to further validate the flexibility of the APTIITLK distribution over the competing distributions.

Keywords: Alpha Power Transformation; Type II Topp-Leone Generated; Quantile; Simulation Study

1. INTRODUCTION

The theory of statistical analysis has received a reasonable attention in the area of developing lifetime distributions. several lifetime distributions have been proposed to analyze real world phenomena in literature. Its utility has found tremendous applications in research fields such as engineering,

biological sciences, machine learning, actuarial sciences, demography, agricultural sciences, etc. Regardless the numerous lifetime distributions in literature, an insatiable quest to develop more flexible and tractable models have evolved among researchers in the field of statistical distribution theory. It is noteworthy that many existing lifetime distributions have failed in providing good fit for certain complex datasets, thus, the drive to develop new ones. Several novel methodologies have been introduced to expand the utility of existing lifetime distributions. Thanks to [1] who developed the exponentiated Weibull family of distributions, [2] introduced the Marshall-Olkin extended family, [3] studied the beta-*G* class of distributions, [4] proposed the transmuted-G method, [5] used the idea of [3] to introduce the Kumaraswamy-*G* method, [6] proposed the transformed-transformer (*T-X*) method, and [7] developed the Weibull-*G* method.

Recently, [8] have suggested a new method of adding extra parameter to an existing lifetime distribution which they called "alpha-power transformation method". Let $G(t)$ denote the cdf of any continuous random variable *T*, [8] defined the alpha-power transformation of $G(t)$ as

$$F_{APT}(t,\alpha)=\begin{cases}\dfrac{\alpha^{G(t)}-1}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\[2ex] G(t), & if \quad \alpha=1\end{cases} \tag{1}$$

The pdf associated to (1) is defined as

$$f_{APT}(t,\alpha)=\begin{cases}\dfrac{\log\alpha}{\alpha-1}\,g(t)\,\alpha^{G(t)}, & if \quad \alpha>0,\ \alpha\neq 1\\[2ex] g(t), & if \quad \alpha=1\end{cases} \tag{2}$$

The methodology defined in (1) and (2) has been adopted by researchers to generalize existing lifetime distributions. Such generalizations include the alpha-power Raleigh distribution by [9], alpha-power transformed Lindley distribution by [10], alpha-power transformed power Lindley distribution by [11], alpha-power inverse Lomax distribution by [12], alpha-power Tessier distribution by [13], alpha-power Topp-Leone distribution by [14], etc.

Another tractable method of generalization is the type II Topp-Leone-G family of distributions proposed by [15]. They adopted the idea of [16] to generalize the Topp-Leone distribution with the cdf defined by

$$\begin{aligned}G_{TIITL-G}(t,\gamma,\xi)&=1-2\gamma\int_0^{1-F(t,\xi)}t^{\gamma-1}(1-t)(2-t)^{\gamma-1}\,dt,\\&=1-\left(1-F^2(t,\xi)\right)^{\gamma},\end{aligned} \tag{3}$$

and pdf obtained as

$$g_{TIITL-G}(t,\gamma,\xi)=2\gamma f(t,\xi)F(t,\xi)\left[1-F^2(t,\xi)\right]^{\gamma-1},\quad t>0,\gamma>0. \tag{4}$$

The one-parameter special case of the Topp-Leone distribution developed by [17] happens to be the simplest (single parameter) distribution with a bathtub hazard rate property and this unique feature has also motivated researchers to study different modification of the distribution to enhance its flexibility in data fitting. [18] developed the Topp-Leone inverse Weibull distribution, [19] proposed the Topp-Leone Weibull distribution, [20] discussed the Topp-Leone generated Weibull distribution, [21] studied the Topp-Leone power Lindley distribution, [22] developed the transmuted version of the Marshall-Olkin Topp-Leone distribution studied in [23], etc.

Inspired by the idea of [24], we construct a novel and more suitable two-parameter generalized class of distributions by considering the cdf defined in (3) as the new baseline distribution in (1). The cdf of the new two-parameter generalized class of distributions is thus, defined as

$$F_{APTIITL-G}(t,\alpha,\gamma,\xi)=\begin{cases}\dfrac{\alpha^{\left\lfloor 1-\left(1-F^2(t,\xi)\right)^{\gamma}\right\rfloor}-1}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ 1-\left(1-F^2\left(t,\xi\right)\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{5}$$

the density function associated to (5) is obtained as

$$f_{APTIITL-G}(t,\alpha,\gamma,\xi)=\begin{cases}\dfrac{\log\alpha}{\alpha-1}2\gamma f(t,\xi)F(t,\xi)\left(1-F^2\left(t,\xi\right)\right)^{\gamma-1}\alpha^{\left\lfloor 1-\left(1-F^2(t,\xi)\right)^{\gamma}\right\rfloor}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ 2\gamma f(t,\xi)F(t,\xi)\left(1-F^2\left(t,\xi\right)\right)^{\gamma-1}, & if \quad \alpha=1\end{cases}. \tag{6}$$

The random variable T in (5) and (6) is said to follow the alpha power type II Topp-Leone-G family of distributions (APTIITL-G for short). The survival function (sf) and hazard rate function (hrf) of the APTIITL-G family are, respectively, defined as

$$S_{APTIITL-G}(t,\alpha,\gamma,\xi)=\begin{cases}\dfrac{\alpha\left(1-\alpha^{-\left(1-F^2(t,\xi)\right)^{\gamma}}\right)}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ \left(1-F^2\left(t,\xi\right)\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{7}$$

and

$$h_{APTIITL-G}(t,\alpha,\gamma,\xi)=\begin{cases}\dfrac{\log(\alpha)2\gamma f(t,\xi)F(t,\xi)\left(1-F^2\left(t,\xi\right)\right)^{\gamma-1}\alpha^{-\left(1-F^2(t,\xi)\right)^{\gamma}}}{\left(1-\alpha^{-\left(1-F^2(t,\xi)\right)^{\gamma}}\right)}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ \dfrac{2\gamma f(t,\xi)F(t,\xi)}{1-F^2\left(t,\xi\right)}, & if \quad \alpha=1\end{cases}. \tag{8}$$

The basic objectives for developing the APTIITL-G family in practice are:

(i) to capture distributions with exponentially decreasing (reversed-J), negatively-skewed, positively-skewed, symmetric shaped property;

(ii) to construct distributions that span various forms of hazard rate property;

(iii) to produce distributions with consistently better fits than existing nested and non-nested distributions.

The rest of this paper is structured into the following sections. In Section 2 presents the materials and method. In detail, we derive the linear representation of the APTIITL-G density function, introduce some special sub-models generated from the APTIITL-G family. Some statistical properties of the APTIITL-G family are studied and the parameter estimation of the APTIITL-G family are obtained via the maximum likelihood method. A simulation study is conducted to investigate the asymptotic behavior of the parameter estimates. In Section 3, two data sets are used to illustrate the potential of sub-model from the APTIITL-G family. Section 4 concludes the paper.

2. MATERIALS AND METHOD

2.1 The density function of APTIITL-G family: linear representation

Most generalized distributions lack closed form expression for some of their statistical properties, thus limiting their utility in data analysis. Statistical properties such as moments, moment generating function, probability weighted moments, etc., are derived from the density function of the distribution. Hence, there is a clear need to obtain the series representation of the density function. To obtain the series representation of the density function of APTIITL-G family, we consider the following useful expansions.

$$\alpha^{t} = \sum_{k=0}^{\infty} \frac{\left(\log(\alpha)\right)^{k} t^{k}}{k!}, \tag{9}$$

$$\left(1-t\right)^{n} = \sum_{q=0}^{\infty} \binom{n}{q} (-1)^{q} t^{q}. \tag{10}$$

(See [25], pg. 26, 2007).

Using (9) and (10) in (6), we have

$$\alpha^{\left[1-\left(1-F^{2}(t,\xi)\right)^{\gamma}\right]} = \sum_{j=0}^{\infty} \frac{\left(\log(\alpha)\right)^{j}}{j!} \left[1-\left(1-F^{2}\left(t,\xi\right)\right)^{\gamma}\right]^{j},$$

$$\left[1-\left(1-F^{2}\left(t,\xi\right)\right)^{\gamma}\right]^{j} = \sum_{k=0}^{j} \binom{j}{k} (-1)^{k} \left(1-F^{2}\left(t,\xi\right)\right)^{\gamma k},$$

$$\left(1-F^{2}\left(t,\xi\right)\right)^{\gamma(k+1)-1} = \sum_{m=0}^{\gamma(k+1)-1} \binom{\gamma(k+1)-1}{m} (-1)^{m} F\left(t,\xi\right)^{2m},$$

By inserting into (6), we have

$$f_{APTIITL-G}\left(t,\alpha,\gamma,\xi\right) = \sum_{j,=0}^{\infty} \sum_{k=0}^{j} \sum_{m=0}^{\gamma(k+1)-1} \psi_{j,k,m}\, \pi_{2(m+1)}\left(t,\alpha,\gamma,\xi\right) \tag{11}$$

where,

$$\psi_{j,k,m} = \frac{\gamma\left(\log(\alpha)\right)^{j+1}}{j!(m+1)(\alpha-1)} \binom{j}{k} \binom{\gamma(k+1)-1}{m} (-1)^{k+m}$$

and

$$\pi_{2(m+1)}\left(t,\alpha,\gamma,\xi\right) = 2\left(m+1\right) f\left(t,\xi\right)\left[F(t,\xi)\right]^{2(m+1)-1}.$$

The pdf of APTIITL-G family defined in (11) is expressed as an infinite linear combination of exp-G

densities with power parameter $2(m+1)$. Whereas, the cdf of APTIITL-G family is expressed as a linear combination of the exp-G cdfs as

$$F_{APTIITL-G}(t,\alpha,\gamma,\xi)=\sum_{j,=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1}\psi_{j,k,m}\,\Pi_{2(m+1)}\left(t,\alpha,\gamma,\xi\right). \tag{12}$$

Where $\Pi_{2(m+1)}\left(t,\alpha,\gamma,\xi\right)$ is the exp-G cdf with power parameter $2(m+1)$.

2.2 Some special sub-models of APTTIITL-G family

In this section, the authors introduced five special sub-models from APTIITL-G family by allowing the baseline distribution in (5) to follow Kumaraswamy, Weibull, log-logistic, Lindley and Bur XII distributions.

2.2.1 The alpha power type II Topp-Leone Kumaraswamy (APTIITLK) distribution

The Kumaraswamy distribution is a bounded lifetime distribution developed by [26], with cdf and pdf, respectively, defined by

$$F\left(t,\beta,\lambda\right)=1-\left(1-t^{\beta}\right)^{\lambda},\quad \beta,\lambda>0,\ 0<t<1, \tag{13}$$

and

$$f\left(t,\beta,\lambda\right)=\lambda\beta t^{\beta-1}\left(1-t^{\beta}\right)^{\lambda-1},\quad \beta,\lambda>0,\ 0<t<1. \tag{14}$$

By inserting (13) into (5), the authors defined the cdf of alpha power type II Topp-Leone Kumaraswamy (APTIITLK) distribution by

$$F_{APTIITLK}(t)=\begin{cases}\dfrac{\alpha^{\left[1-\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}\right]}-1}{\alpha-1}, & if\quad \alpha>0,\ \alpha\neq 1\\[2ex] 1-\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}, & if\quad \alpha=1\end{cases}, \tag{15}$$

and the associated pdf defined as

$$f_{APTIITLK}(t)=\begin{cases}\dfrac{\log\alpha}{\alpha-1}2\gamma\lambda\beta t^{\beta-1}\left(1-t^{\beta}\right)^{\lambda-1}\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma-1}\alpha^{\left[1-\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}\right]}, & if\quad \alpha>0,\ \alpha\neq 1\\[2ex] 2\gamma\lambda\beta t^{\beta-1}\left(1-t^{\beta}\right)^{\lambda-1}\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma-1}, & if\quad \alpha=1\end{cases} \tag{16}$$

The sf and hrf of APTIITLK distribution are obtained, respectively, as

$$S_{APTIITLK}(t)=\begin{cases}\dfrac{\alpha\left(1-\alpha^{-\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}}\right)}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}, & if \quad \alpha=1\end{cases} \tag{17}$$

and

$$h_{APTIITLK}(t)=\begin{cases}\dfrac{\log(\alpha)2\gamma\lambda\beta t^{\beta-1}\left(1-t^{\beta}\right)^{\lambda-1}\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma-1}}{\alpha^{\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}}\left(1-\alpha^{-\left(1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}\right)^{\gamma}}\right)}, & if \quad \alpha>0,\ \alpha\neq 1\\ \dfrac{2\gamma\lambda\beta t^{\beta-1}\left(1-t^{\beta}\right)^{\lambda-1}\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]}{1-\left[1-\left(1-t^{\beta}\right)^{\lambda}\right]^{2}}, & if \quad \alpha=1\end{cases}. \tag{18}$$

The plots of the pdf and hrf of APTIITLK distribution are shown in Figure 1.

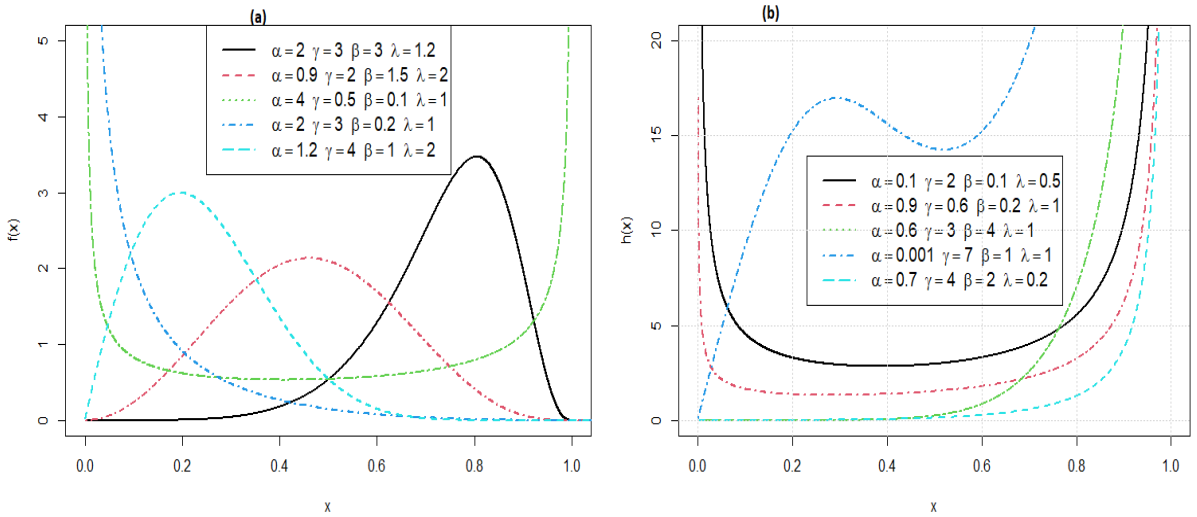


Figure 1: The pdf plot (a) and hrf plot (b) of APTIITLK distribution for different parameter value.

Figure 1 reveals that the pdf of APTIITLK distribution exhibits a decreasing (reserved J-shape), negatively-skewed, positively-skewed, symmetric and bathtub shapes, whereas, the hrf plots indicate an increasing, bathtub and inverted bathtub hazard properties.

2.2.2 The alpha power type II Topp-Leone Weibull (APTIITLW) distribution

Suppose the baseline distribution in (5) follow the Weibull distribution with $F(t,\lambda)=1-e^{-t^{\lambda}}$ and $f(t,\lambda)=\lambda t^{\lambda-1}e^{-t^{\lambda}}$, where $\lambda>0$ is the shape parameter, the authors defined the cdf of alpha power type II Topp-Leone Weibull (APTIITLW) distribution by

$$F_{APTIITLW}(t)=\begin{cases}\dfrac{\alpha^{\left[1-\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}\right]}-1}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ 1-\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{19}$$

and the pdf of APTIITLW distribution is obtained as

$$f_{APTIITLW}(t)=\begin{cases}\dfrac{\log\alpha}{\alpha-1}2\gamma\lambda t^{\lambda-1}e^{-t^{\lambda}}\left[1-e^{-t^{\lambda}}\right]\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma-1}\alpha^{\left[1-\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}\right]}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ 2\gamma\lambda t^{\lambda-1}e^{-t^{\lambda}}\left[1-e^{-t^{\lambda}}\right]\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma-1}, & if \quad \alpha=1\end{cases}. \tag{20}$$

The sf and hrf of the APTIITLW distribution are obtained, respectively, as

$$S_{APTIITLW}(t)=\begin{cases}\dfrac{\alpha\left(1-\alpha^{\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}}\right)}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ \left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{21}$$

and

$$h_{APTIITLW}(t)=\begin{cases}\dfrac{\log(\alpha)2\gamma\lambda t^{\lambda-1}e^{-t^{\lambda}}\left[1-e^{-t^{\lambda}}\right]\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma-1}}{\alpha^{\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}}\left(1-\alpha^{\left(1-\left[1-e^{-t^{\lambda}}\right]^2\right)^{\gamma}}\right)}, & if \quad \alpha>0,\ \alpha\neq 1\\ \\ \dfrac{2\gamma\lambda t^{\lambda-1}e^{-t^{\lambda}}\left[1-e^{-t^{\lambda}}\right]}{1-\left[1-e^{-t^{\lambda}}\right]^2}, & if \quad \alpha=1\end{cases}. \tag{22}$$

Figure 2 presents the pdf and hrf plots of the APTIITLW distribution for some selected values of the parameters.

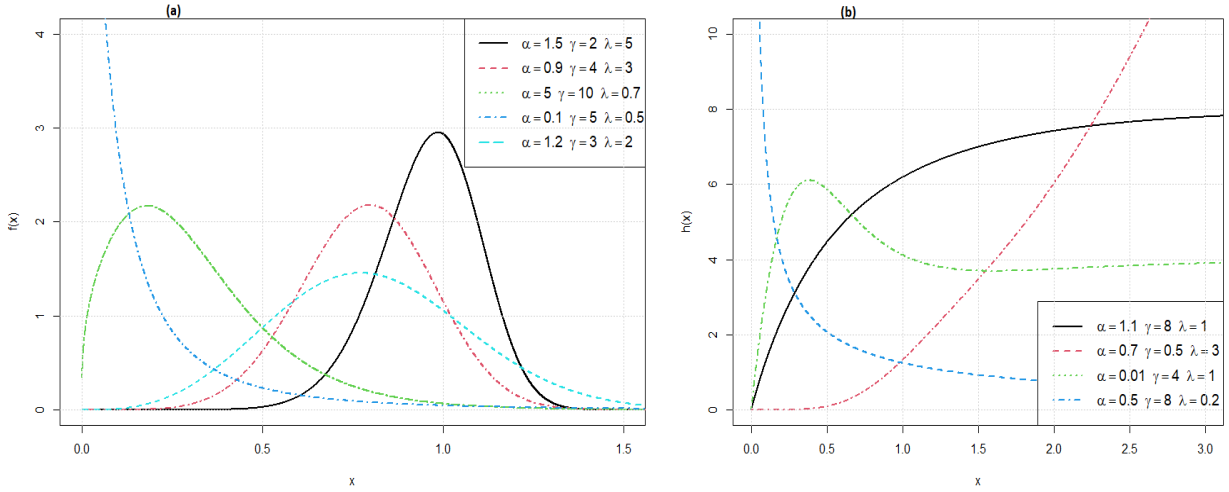


Figure 2: The pdf plot (a) and hrf plot (b) of the APTIITLW distribution for varying choices of parameter.

Clearly, the pdf plot in Figure 2 indicates a decreasing (reserved J-shape), negatively-skewed, positively-skewed, and symmetric shapes, whereas, the hrf plot indicate a decreasing, increasing, and inverted bathtub hazard properties.

2.2.3 The alpha power type II Topp-Leone log-logistic (APTIITL^3) distribution

Let T be a random variable having the log-logistic cumulative distribution function (cdf), $F(t,\lambda)=1-\left(1+t^{\lambda}\right)^{-1}$ and density function (pdf), $f(t,\lambda)=\lambda t^{\lambda-1}\left(1+t^{\lambda}\right)^{-2}$. It is easy to define the cdf and pdf of a new distribution from (5) and (6), respectively, as

$$F_{APTIITL^3}(t)=\begin{cases}\dfrac{\alpha^{\left[1-\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}\right]}-1}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ 1-\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{23}$$

and

$$f_{APTIITL^3}(t)=\begin{cases}\dfrac{\log(\alpha)2\gamma\lambda t^{\lambda-1}}{(\alpha-1)\left(1+t^{\lambda}\right)^2}\left[1-\left(1+t^{\lambda}\right)^{-1}\right]\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma-1}\alpha^{\left[1-\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}\right]}, & if \quad \alpha>0,\ \alpha\neq 1\\ \dfrac{2\gamma\lambda t^{\lambda-1}}{\left(1+t^{\lambda}\right)^2}\left[1-\left(1+t^{\lambda}\right)^{-1}\right]\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma-1}, & if \quad \alpha=1\end{cases}. \tag{24}$$

The cdf and pdf of the APTIITL^3 distribution are readily defined by (23) and (24). The sf and hrf associated to (23) and (24) are obtained, respectively, as

$$S_{APTIITL^3}(t)=\begin{cases}\dfrac{\alpha\left(1-\alpha^{-\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}}\right)}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{25}$$

and

$$h_{APTIITL^3}(t)=\begin{cases}\dfrac{\log(\alpha)2\gamma\lambda t^{\lambda-1}\left(1+t^{\lambda}\right)^{-2}\left[1-\left(1+t^{\lambda}\right)^{-1}\right]\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma-1}}{\alpha^{\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}}\left(1-\alpha^{-\left(1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2\right)^{\gamma}}\right)}, & if\ \alpha>0,\ \alpha\neq 1\\ \dfrac{2\gamma\lambda t^{\lambda-1}\left(1+t^{\lambda}\right)^{-2}\left[1-\left(1+t^{\lambda}\right)^{-1}\right]}{1-\left[1-\left(1+t^{\lambda}\right)^{-1}\right]^2}, & if \quad \alpha=1\end{cases}. \tag{26}$$

The pdf and hrf plots of APTIITL³ distribution for selected values of the parameters are displayed in Figure 3.

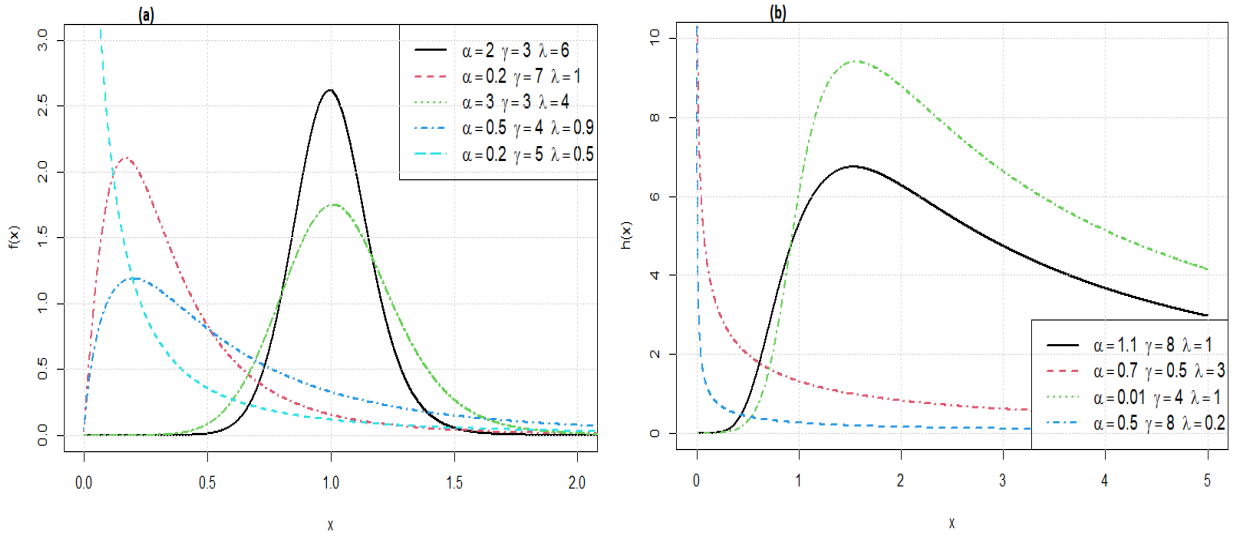


Figure 3: The pdf plot (a) and hrf plot (b) of APTIITL³ distribution for varying choices of parameter.

The pdf plots in Figure 3 indicates a decreasing (reserved J-shape), positively-skewed, and symmetric shapes, whereas, the hrf plots indicate a decreasing and inverted bathtub hazard properties.

2.2.4 The alpha power type II Topp-Leone Lindley (APTIITLL) distribution

The one-parameter Lindley distribution proposed by [27] is defined by the cdf and pdf, respectively, as

$$F\left(t,b\right)=1-\left(1+\frac{bt}{1+b}\right)e^{-bt},\qquad b>0,\quad t>0, \tag{27}$$

and

$$f\left(t,b\right)=\frac{b^2}{1+b}\left(1+t\right)e^{-bt},\qquad b>0,\ \ t>0. \tag{28}$$

By inserting (27) and (28) into (5) and (6), the authors obtained the cdf and pdf of alpha power type II Topp-Leone Lindley (APTIITLL) distribution, respectively, as

$$F_{APTIITLL}\left(t\right)=\begin{cases}\dfrac{\alpha^{\left[1-\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}\right]}-1}{\alpha-1}, & if\quad \alpha>0,\ \alpha\neq 1\\ 1-\left(1-\left[1-\left(1+\dfrac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}, & if\qquad \alpha=1\end{cases}, \tag{29}$$

and

$$f_{APTIITLL}\left(t\right)=\begin{cases}\dfrac{\log(\alpha)2\gamma b^2}{(\alpha-1)(1+b)}\left(1+t\right)e^{-bt}\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma-1}\alpha^{\left[1-\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}\right]}, & if\quad \alpha>0,\ \alpha\neq 1\\ \dfrac{2\gamma b^2}{(1+b)}\left(1+t\right)e^{-bt}\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma-1}, & if\qquad \alpha=1\end{cases} \tag{30}$$

The sf and hrf of APTIITLL distribution are obtained, respectively, as

$$S_{APTIITLL}\left(t\right)=\begin{cases}\dfrac{\alpha\left(1-\alpha^{\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}}\right)}{\alpha-1}, & if\quad \alpha>0,\ \alpha\neq 1\\ \left(1-\left[1-\left(1+\dfrac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}, & if\qquad \alpha=1\end{cases}, \tag{31}$$

and

$$h_{APTIITLL}(t)=\begin{cases}\dfrac{\dfrac{\log(\alpha)2\gamma b^2}{(1+b)}(1+t)e^{-bt}\left[1-\left(1+\dfrac{bt}{1+b}\right)e^{-bt}\right]\left(1-\left[1-\left(1+\dfrac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma-1}}{\alpha^{\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}}\left(1-\alpha^{-\left(1-\left[1-\left(1+\frac{bt}{1+b}\right)e^{-bt}\right]^2\right)^{\gamma}}\right)}, & if \quad \alpha>0,\ \alpha\neq 1\\ \dfrac{\dfrac{2\gamma b^2}{(1+b)}(1+t)e^{-bt}\left[1-\left(1+\dfrac{bt}{1+b}\right)e^{-bt}\right]}{1-\left[1-\left(1+\dfrac{bt}{1+b}\right)e^{-bt}\right]^2}, & if \quad \alpha=1\end{cases}. \tag{32}$$

Figure 4 displays the pdf and hrf plots of the APTIITLL distribution for selected parameter values.

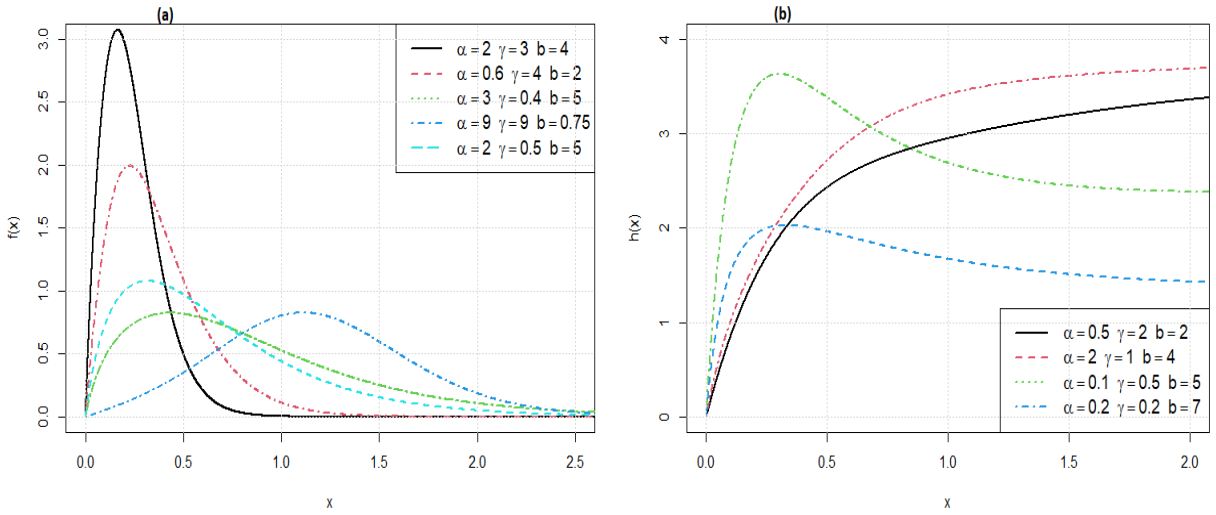


Figure 4: The pdf plot (a) and hrf plot (b) of APTIITLL distribution for varying choices of parameter.

From Figure 4, we observe that the pdf plots of APTIITLL distribution accommodates a positively-skewed and symmetric shapes, whereas, the hrf plots exhibit an increasing and inverted bathtub hazard properties.

2.2.5 The alpha power type II Topp-Leone Burr XII (APTIITLBXII) distribution

The Burr XII distribution is one of the most commonly used models among the twelve (12) special models introduced by [28]. The cdf and pdf of Burr XII distribution are defined, respectively, as

$$F(t,a,b)=1-\left(1+t^{a}\right)^{-b}, \quad a,b>0, \quad t>0, \tag{33}$$

and

$$f(t,a,b)=abt^{a-1}\left(1+t^{a}\right)^{-(b+1)}, \quad a,b>0, \quad t>0. \tag{34}$$

Utilizing the cdf defined in (33) as the baseline distribution in (5), the authors obtained the cdf of alpha power type II Topp-Leone Burr XII (APTIITLBXII) distribution by

$$F_{APTIITLBXII}(t)=\begin{cases}\dfrac{\alpha^{\left[1-\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}\right]}-1}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ 1-\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{35}$$

and

$$f_{APTIITLBXII}(t)=\begin{cases}\dfrac{\log(\alpha)2\gamma abt^{a-1}}{(\alpha-1)\left(1+t^{a}\right)^{(b+1)}}\left[1-\left(1+t^{a}\right)^{-b}\right]\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma-1}\alpha^{\left[1-\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}\right]}, & if \quad \alpha>0,\ \alpha\neq 1\\ \dfrac{2\gamma abt^{a-1}}{\left(1+t^{a}\right)^{(b+1)}}\left[1-\left(1+t^{a}\right)^{-b}\right]\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma-1}, & if \quad \alpha=1\end{cases}. \tag{36}$$

The sf and hrf of APTIITLBXII distribution are obtained, respectively, as

$$S_{APTIITLBXII}(t)=\begin{cases}\dfrac{\alpha\left(1-\alpha^{-\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}}\right)}{\alpha-1}, & if \quad \alpha>0,\ \alpha\neq 1\\ \left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}, & if \quad \alpha=1\end{cases}, \tag{37}$$

and

$$h_{APTIITLBXII}(t)=\begin{cases}\dfrac{\dfrac{\log(\alpha)2\gamma abt^{a-1}}{\left(1+t^{a}\right)^{(b+1)}}\left[1-\left(1+t^{a}\right)^{-b}\right]\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma-1}}{\alpha^{\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}}\left(1-\alpha^{-\left(1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}\right)^{\gamma}}\right)}, & if \quad \alpha>0,\ \alpha\neq 1\\ \dfrac{\dfrac{2\gamma abt^{a-1}}{\left(1+t^{a}\right)^{(b+1)}}\left[1-\left(1+t^{a}\right)^{-b}\right]}{1-\left[1-\left(1+t^{a}\right)^{-b}\right]^{2}}, & if \quad \alpha=1\end{cases}, \tag{38}$$

Some useful pdf and hrf plots of the APTIITLBXII distribution are displayed in Figure 5.

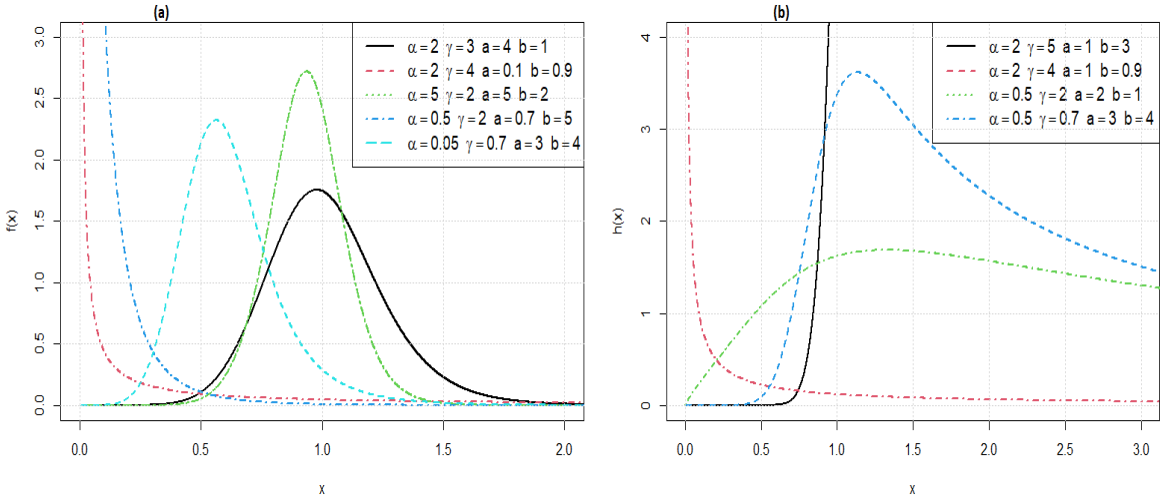


Figure 5: The pdf plot (a) and hrf plot (b) of APTIITLBXII distribution for different values of the parameters.

The density plots of the APTIITLBXII distribution displayed in Figure 5, shows a decreasing, positively-skewed and symmetric shapes, whereas, the hrf plots exhibit a decreasing, increasing and inverted bathtub hazard properties.

2.3 Statistical Properties

This section is devoted to derivation of some statistical properties of APTIITL-G family. In particular, the quantiles, r^{th}-moments, moment generating function, probability weighted moments (PWMs), Renyi entropy and order statistics are derived.

2.3.1 Quantile Function

The quantile function of APTIITL-G family of distributions is obtained as

$$Q_T(u) = F^{-1}\left\{\sqrt{1-\left[1-\frac{\log\left(u(\alpha-1)+1\right)}{\log(\alpha)}\right]^{1/\gamma}}\right\}, \quad u \in (0,1). \tag{39}$$

By inserting $u = 0.5$ in (39), we obtain the median of APTIITL-G family as

$$Q_T(0.5) = F^{-1}\left\{\sqrt{1-\left[1-\frac{\log\left(\alpha+1\right)-\log\left(2\right)}{\log(\alpha)}\right]^{1/\gamma}}\right\}. \tag{40}$$

The utility of (39) is most essential in generating random sample from the distribution.

2.3.2 Moments and Incomplete Moments

Let T be a random variable having the density function of the APTIITL-G family, then from (11), the r^{th} moments of T is defined by

$$E(T^r) = \mu_r^{'} = \sum_{j,=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1} \psi_{j,k,m}\int_{-\infty}^{\infty} t^r \pi_{2(m+1)}(t,\alpha,\gamma,\xi)\,dt, \quad r=1,2,3,4,... \tag{41}$$

The integral part of (41) can be expressed as $E\left[Y_{2(m+1)}^{r}\right]$, which is the r^{th} moments of the exp-G family with power parameter $2(m+1)$.

The mean $\left(\mu_1^{'}\right)$ of the APTIITL-G family is obtained from (41) when $r=1$. The variance $\left(\sigma^2\right)$, skewness $\left(S_k\right)$ and kurtosis $\left(K_s\right)$ are obtained as

$$\text{variance}\left(\sigma^2\right) = \mu_2^{'} - \left(\mu_1^{'}\right)^2, \qquad \text{skewness}\left(S_k\right) = \frac{\mu_3^{'} - 3\mu_2^{'}\mu_1^{'} + 2\left(\mu_1^{'}\right)^3}{\left(\mu_2^{'} - \left(\mu_1^{'}\right)^2\right)^{\frac{3}{2}}},$$

$$\text{kurtosis}\left(K_s\right) = \frac{\mu_4^{'} - 4\mu_3^{'}\mu_1^{'} + 6\mu_3^{'}\left(\mu_1^{'}\right)^2 - 3\left(\mu_1^{'}\right)^4}{\left(\mu_2^{'} - \left(\mu_1^{'}\right)^2\right)^2}.$$

furthermore, we deduce the r^{th} lower incomplete moment of APTIITL-G family from (31) as

$$\varphi_r(q) = \sum_{j,=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1} \psi_{j,k,m}\int_{-\infty}^{q} t^r \pi_{2(m+1)}(t,\alpha,\gamma,\xi)\,dt. \tag{42}$$

Table 1 holds numerical values of the mean $\left(\mu_1^{'}\right)$, variance $\left(\sigma^2\right)$, measures of skewness $\left(S_k\right)$ and kurtosis $\left(K_s\right)$ of alpha power type II Topp-Leone Kumaraswamy (APTIITLK) distribution for some selected values of the parameters. Observations from the table reveal that APTIITLK distribution is negatively-skewed, positively-skewed, symmetric, platykurtic, leptokurtic as well as exhibiting a mesokurtic properties.

Table 1: *The r^{th}-moments of APTIITLK distribution for* $(\beta=2,\ \gamma=3)$

α	λ	$\mu_1^{'}$	σ^2	S	K
0.2	2	0.4423	0.0200	0.1322	3.0658
	4	0.3248	0.0123	0.3264	3.3290
	6	0.2686	0.0089	0.4609	3.1114
1.5	2	0.5236	0.0208	-0.2296	2.9934
	4	0.3896	0.0137	0.0230	2.7440
	6	0.3238	0.0101	0.0834	2.6767
3.0	2	0.5513	0.0198	-0.3418	3.0006
	4	0.4121	0.0134	-0.1413	2.7469
	6	0.3431	0.0099	-0.0618	2.5109

2.3.3 Moment generating function (mgf) and probability weighted moments (PWMs)

The mgf of APTIITL-G family is obtained as

$$M_T(q) = E\left[e^{qT}\right] = \int_{-\infty}^{\infty} e^{qt} f(t)dt,$$
$$= \sum_{j,n=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1} \psi_{j,k,m,n}^{*} E\left[Y_{2(m+1)}^{n}\right], \tag{43}$$

where,

$$\psi^{*}_{j,k,m,n} = \frac{\gamma q^{n}\left(\log(\alpha)\right)^{j+1}}{j!n!(m+1)(\alpha-1)}\binom{j}{k}\binom{\gamma(k+1)-1}{m}(-1)^{k+m},$$

and $E\left[Y^{n}_{2(m+1)}\right]$ is the n^{th} moment of the exp-G family with power parameter $2(m+1)$.

The PWMs of a random variable T as defined in [29] is given as

$$\rho_{q,r} = E\left[T^{r}F^{q}(t)\right] = \int_{-\infty}^{\infty} t^{r} f(t)F^{q}(t)dt. \tag{44}$$

By inserting (5) and (6) into (44), the authors obtained the $(q,r)^{th}$ PWMs of APTIITL-G family as

$$f(t,\alpha,\gamma,\xi)F^{q}(t,\alpha,\gamma,\xi) = \sum_{l=0}^{\infty}(-1)^{q-l}\binom{q}{l}\frac{\log\alpha}{(\alpha-1)^{q+1}}2\gamma f(t,\xi)F(t,\xi)\left(1-F^{2}(t,\xi)\right)^{\gamma-1}\alpha^{(1+l)\left[1-\left(1-F^{2}(t,\xi)\right)^{\gamma}\right]}. \tag{45}$$

Further simplification of (45) and substituting into (44), yields

$$\rho_{q,r} = \sum_{l,j=0}^{\infty}\psi^{**}_{k,m}E\left[Y^{r}_{2(m+1)}\right], \tag{46}$$

where,

$$\psi^{**}_{k,m} = \sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1}\frac{\gamma(1+l)^{j}\left(\log(\alpha)\right)^{j+1}}{j!(m+1)(\alpha-1)^{q+1}}\binom{q}{l}\binom{j}{k}\binom{\gamma(k+1)-1}{m}(-1)^{q-l+k+m}.$$

2.3.4 Renyi Entropy

The Renyi entropy of a random variable T with a known pdf, $f(t)$ is given by

$$\tau_{R}(\omega) = \frac{1}{1-\omega}\log\int_{-\infty}^{\infty} f^{\omega}(t)\,dt, \qquad \omega>0, \omega\neq 1. \tag{47}$$

Applying (6) in (47), the Renyi entropy of APTIITL-G family is defined as follows.

$$\tau_{R}(\omega) = \frac{1}{1-\omega}\log\left[\left[\frac{\log\alpha}{\alpha-1}\right]^{\omega}(2\gamma)^{\omega}\int_{-\infty}^{\infty} f^{\omega}(t,\xi)F^{\omega}(t,\xi)\left(1-F^{2}(t,\xi)\right)^{\omega(\gamma-1)}\alpha^{\omega\left[1-\left(1-F^{2}(t,\xi)\right)^{\gamma}\right]}dt\right]. \tag{48}$$

Employing (11) and (12) into (48), yields

$$\tau_{R}(\omega) = \frac{1}{1-\omega}\log\left[\sum_{j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+\omega)-\omega}\frac{(2\gamma)^{\omega}\left(\log(\alpha)\right)^{j+\omega}\omega^{j}}{j!(\alpha-1)^{\omega}}\binom{j}{k}\binom{\gamma(k+\omega)-\omega}{m}(-1)^{k+m}\int_{-\infty}^{\infty} f^{\omega}(t,\xi)F^{2m+\omega}(t,\xi)dt\right]. \tag{49}$$

2.3.5 Order Statistics

Suppose that $T_1, T_2, ..., T_n$ are random samples generated from a known probability distribution. Let $T_{r:n}$ denote the r^{th} order statistic, then the pdf of $T_{r:n}$ is defined as

$$f_{r:n}(t) = \frac{1}{B(r,n-r+1)}\sum_{p=0}^{n-r}\binom{n-r}{p}(-1)^{p} f(t)F(t)^{r+p-1}, \tag{50}$$

By inserting (5) and (6) into (50), the authors obtained the pdf of APTIITL-G r^{th} order statistics as follows.

$$f(t)F(t)^{r+p-1} = \sum_{l=0}^{\infty}(-1)^{r+p-l-1}\binom{r+p-1}{l}\frac{\log\alpha}{(\alpha-1)^{r+p}}2\gamma f(t,\xi)F(t,\xi)\left(1-F^{2}(t,\xi)\right)^{\gamma-1}\alpha^{(1+l)\left[1-\left(1-F^{2}(t,\xi)\right)^{\gamma}\right]} \tag{51}$$

Employing similar approach in (45), (51) is further simplified as

$$f(t)F(t)^{r+p-1}=\sum_{l,j=0}^{\infty}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1}\frac{2\gamma(1+l)^{j}(\log(\alpha))^{j+1}}{j!(\alpha-1)^{r+p}}\binom{r+p-1}{l}\binom{j}{k}\binom{\gamma(k+1)-1}{m}(-1)^{r+p+k+m-l-1}f(t,\xi)F^{2m+1}(t,\xi), \tag{52}$$

so that (50) now becomes

$$f_{r:n}(t)=\frac{1}{B(r,n-r+1)}\sum_{l,j=0}^{\infty}\varpi_{p,k,m}\,\pi_{2(m+1)}(t), \tag{53}$$

where,

$$\varpi_{p,k,m}=\sum_{p=0}^{n-r}\sum_{k=0}^{j}\sum_{m=0}^{\gamma(k+1)-1}\frac{\gamma(1+l)^{j}(\log(\alpha))^{j+1}}{j!(m+1)(\alpha-1)^{r+p}}\binom{r+p-1}{l}\binom{n-r}{p}\binom{j}{k}\binom{\gamma(k+1)-1}{m}(-1)^{r+2p+k+m-l-1}.$$

Whereas, the s^{th} moment of APTIITL-G r^{th} order statistic can be expressed as

$$E(T_r^s)=\frac{1}{B(r,n-r+1)}\sum_{l,j=0}^{\infty}\varpi_{p,k,m}\,E\left[Y_{2(m+1)}^{s}\right], \tag{54}$$

where $E\left[Y_{2(m+1)}^{s}\right]$ is the s^{th} moment of exp-G family with power parameter $2(m+1)$.

2.4 Parameter Estimation and Simulation Study

2.4.1 Maximum Likelihood Estimation

The method of maximum likelihood estimation is adopted to estimate the unknown parameters of APTIITL-G family of distributions. Suppose $(t_1,t_2,...,t_n)$ are random samples generated from APTIITL-G family, then the log-likelihood function is given as

$$\begin{aligned}\ell(t_i,\psi)&=n\ln(\ln\alpha)-n\ln(\alpha-1)+n\ln(2\gamma)+\sum_{i=1}^{n}\ln\left(f(t_i,\xi)\right)+\sum_{i=1}^{n}\ln\left(F(t_i,\xi)\right)\\&+(\gamma-1)\sum_{i=1}^{n}\ln\left(1-F^2(t_i,\xi)\right)+\ln\alpha\sum_{i=1}^{n}\left(1-\left(1-F^2(t_i,\xi)\right)^{\gamma}\right),\qquad \psi=(\alpha,\gamma,\xi)^T\end{aligned} \tag{55}$$

The score function $U(t_i,\psi)=\left[\frac{\partial\ell(t_i,\psi)}{\partial\alpha},\frac{\partial\ell(t_i,\psi)}{\partial\gamma},\frac{\partial\ell(t_i,\psi)}{\partial\xi}\right]^T$ associated with the log-likelihood function in (55) is obtained by taking the first derivative of (55) with respect to the parameters. These are expressed as

$$\frac{\partial\ell(t_i,\psi)}{\partial\alpha}=\frac{n}{\alpha\ln\alpha}-\frac{n}{\alpha-1}+\frac{1}{\alpha}\sum_{i=1}^{n}\left(1-\left(1-F^2(t_i,\xi)\right)^{\gamma}\right),$$

$$\frac{\partial\ell(t_i,\psi)}{\partial\gamma}=\frac{n}{\gamma}+\sum_{i=1}^{n}\ln\left(1-F^2(t_i,\xi)\right)+\ln\alpha\sum_{i=1}^{n}\ln\left(1-F^2(t_i,\xi)\right)\left(1-F^2(t_i,\xi)\right)^{\gamma},$$

$$\frac{\partial\ell(t_i,\psi)}{\partial\xi_j}=\sum_{i=1}^{n}\frac{f'(t_i,\xi)}{f(t_i,\xi)}+\sum_{i=1}^{n}\frac{f(t_i,\xi)}{F(t_i,\xi)}-2(\gamma-1)\sum_{i=1}^{n}\frac{f(t_i,\xi)}{1-F^2(t_i,\xi)}+2\gamma\ln\alpha\sum_{i=1}^{n}\left(1-F^2(t_i,\xi)\right)^{\gamma-1}F(t_i,\xi)f(t_i,\xi),$$

where $f'(t_i,\xi)=\frac{\partial f(t_i,\xi)}{\partial\xi_j}$, and $\partial\xi_j$ is the j^{th} element of the vector of parameter ξ.

The maximum likelihood estimates (MLEs) of ψ say $\hat{\psi}=\left(\hat{\alpha},\hat{\lambda},\hat{\xi}\right)$, are obtained by solving the system of nonlinear equation $U(t_i,\psi)=0$. Statistical packages such as ***bbmle*** and ***optim*** in R software can be used to numerically compute the parameter estimates

2.4.2 Simulation Study

Again, taking the Kumaraswamy distribution as the generator, the study investigates the performance of the parameter estimates of the APTIITLK distribution via a Monte Carlo simulation study. Random samples of size $n=(100,200,500,800,1000)$ are generated from the APTIITLK distribution at two distinct sets of parameter values $(\alpha=0.2,\beta=0.8,\gamma=3,\lambda=2)$ and $(\alpha=0.2,\beta=0.8,\gamma=3,\lambda=2)$. At each case, the simulation is repeated 3000 times and the following quantities are computed:

i) mean estimate $(\bar{\psi})=\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i$,

ii) average bias $=\frac{1}{N}\sum_{i=1}^{N}(\hat{\psi}_i-\bar{\psi})$,

iii) root mean square error (RMSE) $=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\psi}_i-\bar{\psi})^2}$.

Tables 2 and 3 display the mean estimate, average bias and root mean square errors of the estimates of APTIITLK distribution.

Table 2: *Simulation results of APTIITLK distribution for* $(\alpha=0.7,\beta=0.8,\gamma=0.5,\lambda=2)$

Parameters	N	Mean	Bias	RMSE
	100	0.6672	0.6648	1.0672
	200	0.6708	0.5707	0.9121
α	500	0.6827	0.4825	0.3039
	800	0.6902	0.2401	0.0751
	1000	0.7121	0.1361	0.0022
	100	0.7646	0.1846	0.0721
	200	0.7920	0.0920	0.0352
β	500	0.8014	0.0614	0.0191
	800	0.8022	0.0355	0.0075
	1000	0.8155	0.0252	0.0013
	100	0.4635	0.0233	2.0535
	200	0.4669	0.0151	1.8012
γ	500	0.4702	-0.5750	0.9453
	800	0.4847	-1.2353	0.0413
	1000	0.5177	-1.3023	0.0085
	100	1.8647	0.4106	1.8287
	200	1.8707	0.2651	0.9518
λ	500	1.9237	0.1291	0.4737
	800	2.1101	0.0120	0.2042
	1000	2.1261	0.0014	0.0134

Table 3: *Simulation results of APTIITLK distribution for* $(\alpha = 0.2, \beta = 2, \gamma = 0.2, \lambda = 0.5)$

Parameters	n	Mean	Bias	RMSE
	100	0.1789	0.7149	2.0312
	200	0.1820	0.4812	1.3635
α	500	0.2002	0.2102	0.2738
	800	0.2016	0.1650	0.0225
	1000	0.2106	0.0567	0.0061
	100	1.8981	0.3981	0.3534
	200	1.9017	0.1517	0.2197
β	500	1.9157	0.8573	0.0536
	800	2.1195	0.0995	0.0129
	1000	2.2014	0.0418	0.0031
	100	0.1845	0.4459	0.5416
	200	0.1886	0.1325	0.3162
γ	500	0.2051	0.0640	0.1916
	800	0.2073	-0.9782	0.0177
	1000	0.2105	-1.1086	0.0042
	100	0.4749	0.8951	0.6725
	200	0.4850	0.6232	0.2524
λ	500	0.5002	0.3589	0.1626
	800	0.5160	0.1242	0.0884
	1000	0.5167	0.0253	0.0152

From the results in Tables 2 and 3, the following remarks were observed:

(i) the mean estimate for the parameters approaches the true parameter value as n increases;

(ii) parameter estimates α, β and λ are positively biased, while parameter estimate γ can both be positively and negatively biased;

(iii) the bias and root mean square error for all the parameters decrease as n increases.

These remarks are consistent with the properties of a good estimator.

3. DATA ANALYSIS, RESULTS AND DISCUSSIONS

In this Section, we illustrate the potential of alpha power type II Topp-Leone Kumaraswamy distribution (APTIITLKD) belonging to the APTIITL-G family of distributions using two real data sets. The data sets are concerned with the recovery and mortality rates of Covid-19 patients in Nigeria, covering a duration of two (2) months (May 1 to June 30, 2020).

The flexibility of APTIITLK distribution in data fittings is investigated by comparing its fit with the ones obtained from existing bounded non-nested models. The density function of these competitor distributions is defined as follows:

1. Odd log-logistic Kumaraswamy distribution (OLLKD) studied by [30];

$$f(x,a,b,\alpha) = \frac{ab\alpha x^{a-1}\left(1-x^{a}\right)^{b\alpha-1}\left[1-\left(1-x^{a}\right)^{b}\right]^{\alpha-1}}{\left[\left(1-\left(1-x^{a}\right)^{b}\right)^{\alpha}+\left(1-x^{a}\right)^{b\alpha}\right]^{2}};$$

2. Unit-Burr XII distribution (UBXIID) developed by [31];

$$f(x,\alpha,\beta) = \alpha\beta x^{-1}\left(-\log x\right)^{\beta-1}\left(1+\left(-\log x\right)^{\beta}\right)^{-(\alpha+1)};$$

3. Unit-Burr III distribution (UBIIID) proposed by [32];

$$f(x,\lambda,\beta) = \lambda\beta x^{-2}\left(x^{-1}-1\right)^{\beta-1}\left(1+\left(x^{-1}-1\right)^{\beta}\right)^{-(\lambda+1)};$$

4. Beta distribution reported in [33];

$$f(x,a,b) = \frac{x^{a-1}\left(1-x\right)^{b-1}}{B\left(a,b\right)}, \qquad B\left(a,b\right)=\frac{\Gamma\left(a\right)\Gamma\left(b\right)}{\Gamma\left(a+b\right)};$$

5. Kumaraswamy distribution (KwD) developed by [26];

$$f(x,a,b) = abx^{a-1}\left(1-x^{a}\right)^{b-1}.$$

Data set 1: This data set comprises of the daily recovery rate of Covid-19 patients in Nigeria within the period of 2 months (May 1 to June 30, 2020). The data set is obtained from the ratio of total daily recovery and the total confirmed cases. The data is presented as follows:
0.1617512, 0.1469849, 0.1563722, 0.1488223, 0.1630508, 0.1697933, 0.1704481, 0.1735685, 0.1794748, 0.1768584, 0.1943547, 0.2003342, 0.2152484, 0.2285936, 0.2422018, 0.2618751, 0.2674945, 0.2662348, 0.2708952, 0.2755729, 0.2718073, 0.2764082, 0.2888653, 0.2886848, 0.2864403, 0.2858341, 0.2863850, 0.2907459, 0.2899376, 0.2898021, 0.2959063, 0.2951409, 0.2994731, 0.2981372, 0.3069642, 0.3120567, 0.3127606, 0.3170751, 0.3156003, 0.3123886, 0.3136308, 0.3087811, 0.3221790, 0.3252774, 0.3245260, 0.3211070, 0.3279100, 0.3364533, 0.3412879, 0.3437092, 0.3391559, 0.3398044, 0.3398346, 0.3433625, 0.3457312, 0.3458919, 0.3542364, 0.3582257, 0.3666300, 0.3740898, 0.3793103

Data set 2: This data set holds the daily records of mortality rate of Covid-19 patients in Nigeria within the same time frame in the first data set. It is computed from the ratio of daily death cases and the total confirmed cases. The data is given as follows: 0.003225806, 0.004187605, 0.005863956, 0.0007137759, 0.002033898, 0.001589825, 0.001418037, 0.001022495, 0.002409058, 0.002500568, 0.003232062, 0.001462294, 0.001609334, 0.001162340, 0.0005504587, 0.000711617, 0.0008390670, 0.0009716599, 0.001406030, 0.0001497679, 0.001425314, 0.001239499, 0.0009301090, 0.0003827019, 0.0006197323, 0.0008389262, 0.001832131, 0.0005608525, 0.0005375188, 0.0002029427, 0.001180870, 0.001323502, 0.001109160, 0.001343364, 0.00008683571, 0.0006754475, 0.0008174610, 0.0007208073, 0.0009374268, 0.0005199049, 0.0002883298, 0.001168064, 0.0003293591, 0.0002550695, 0.0009325459, 0.0008404370, 0.0002332634, 0.001747956, 0.0007575758, 0.0003133650, 0.0006058158, 0.0009385497, 0.0005736412, 0.0003275467, 0.0003633061, 0.0003979836, 0.0003004550, 0.0002076671, 0.0001628200, 0.0002785183, 0.0003113567.
Model selection criteria such as the maximized log-likelihood (Log-Lik), Akaike information criterion (AIC), and some goodness of fit test statistics including the Komolgorov-Smirnov (*K-S*) and Crammer von Mises (W^*) test statistics with their corresponding *p*-value are employed for model comparison. Tables 4 and 5 present the summary statistics of the recovery and mortality rate of Covid-19 patients in Nigeria, respectively.

Table 4: *Summary Statistics of the Covid-19 Recovery Data set*

Models	Estimates	Log-Lik	AIC	*K-S* *(p-value)*	W^* *(p-value)*
APTIITLKD	$\alpha = 3.5166$ $\beta = 2.7647$ $\gamma = 11.4660$ $\lambda = 10.7149$	85.3897	-165.7794	0.1223 **(0.2963)**	0.2019 **(0.2641)**
OLLKD	$a = 0.5458$ $b = 0.9851$ $\alpha = 6.7200$	78.4359	-150.8717	0.1366 **(0.1867)**	0.2883 **(0.1458)**
UBXIID	$\alpha = 0.0802$ $\beta = 50.7525$	84.7801	-165.5602	0.1370 **(0.1848)**	0.2756 **(0.1586)**
UBIIID	$\lambda = 0.0496$ $\beta = 20.8434$	40.7476	-77.4952	0.4037 **($1.711e^{-9}$)**	2.6963 **($2.644e^{-7}$)**
Beta	$a = 12.4278$ $b = 31.8699$	79.0322	-154.0643	0.1897 **(0.0214)**	0.4963 **(0.0404)**
Kumaraswamy	$a = 5.6206$ $b = 785.6500$	84.5196	-165.0392	0.1375 **(0.1815)**	0.2452 **(0.1948)**

Table 5: *Summary Statistics of the Covid-19 Mortality Data set*

Models	Estimates	Log-Lik	AIC	*K-S* *(p-value)*	W^* *(p-value)*
APTIITLKD	$\alpha = 68.5111$ $\beta = 0.4399$ $\gamma = 8.7893$ $\lambda = 12.7477$	369.706	-721.412	0.0549 **(0.9880)**	0.0284 **(0.9819)**
OLLKD	$a = 0.2645$ $b = 4.2375$ $\alpha = 5.0911$	360.7049	-715.4097	0.0726 **(0.8811)**	0.0349 **(0.9586)**
UBXIID	$\alpha = 0.0651$ $\beta = 7.7576$	216.3316	-428.6632	0.5653 **($1.665e^{-15}$)**	5.3930 **($2.2e^{-16}$)**
UBIIID	$\lambda = 0.0636$ $\beta = 2.1849$	257.1764	-510.3529	0.5169 **($1.887e^{-15}$)**	4.7144 **($2.2e^{-16}$)**
Beta	$a = 1.5592$ $b = 1446.418$	359.0777	-714.1554	0.0725 **(0.8821)**	0.0639 **(0.7913)**
Kumaraswamy	$a = 1.2125$ $b = 3639.133$	357.7849	-711.5698	0.0843 **(0.7469)**	0.0844 **(0.6686)**

3.1 Discussion of Results

By way of discussing the results in Tables 4 and 5, it is well known that the most appropriate model in fitting any real data set, corresponds to the one having the maximum value of log-likelihood and the minimum value in respect to AIC, *K-S* and W^* with the highest *p*-value. Clearly, from these tables we observed that the APTIITLK distribution satisfying the conditions, outperformed the rest competitor distributions. Thus, becoming the appropriate model in fitting the two data sets considered. Furthermore, we illustrate the flexibility of the APTIITLK distribution over the competitor distribution through graphical plots such as the density and distribution fits of the distributions for each data set as shown in Figures 6 and 7, respectively.

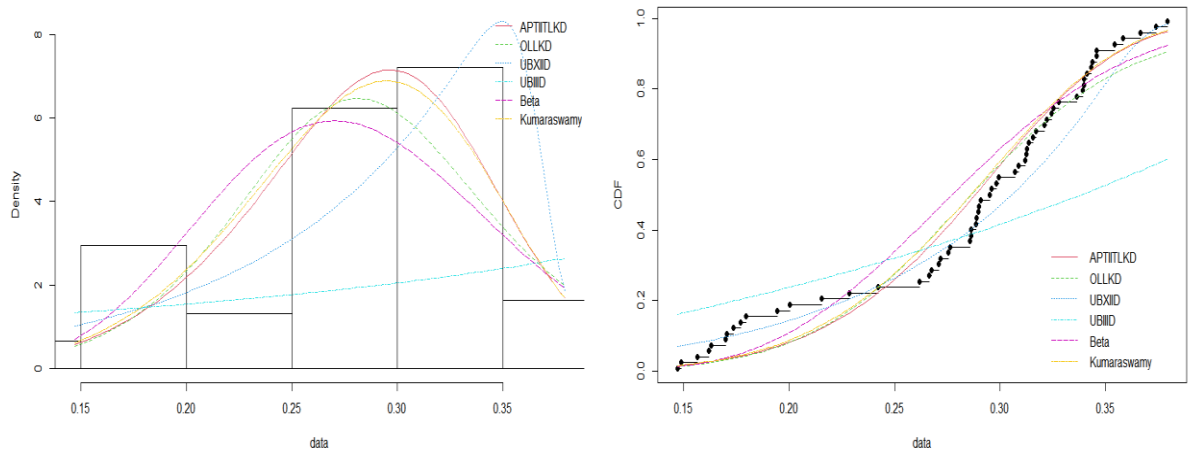


Figure 6: The fitted pdf and cdf of the distributions for Covid-19 recovery data set

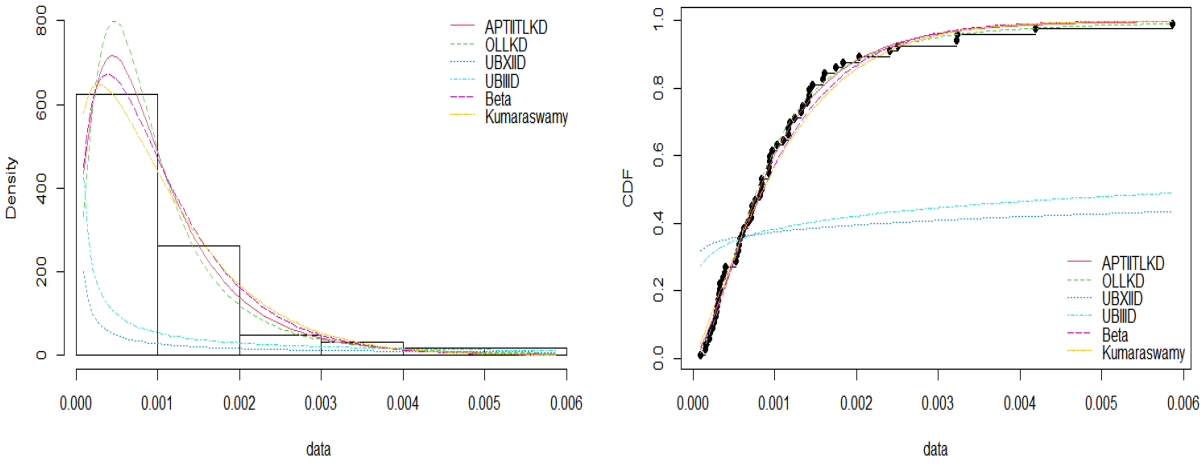


Figure 7: The fitted pdf and cdf of the distributions for Covid-19 mortality data set

4. CONCLUSION

In this paper, we have developed a new family of distributions called "Alpha Power Type II Topp-Leone-generated family of distributions" and some of its mathematical properties were derived. The maximum likelihood estimation method was adopted to obtain the parameter estimate of the family

of distributions. A Monte Carlo simulation study was conducted in other to investigated the performance of the parameter estimates of sub-model belonging to the proposed family of distributions. Two data sets comprising of the daily recovery and mortality rates of Covid-19 patients in Nigeria, from May 1 to June 30, 2020, was employed to illustrate the potential of the proposed family in real world data fittings. Results obtained from the analysis clearly revealed that the APTIITLK distribution from the proposed family performed reasonably better than the compared non-nested distributions in analyzing the two Covid-19 datasets under study.

References

[1] Abbas, S., Taqi, S. A., Mustafa, F., Murtaza, M. and Shahbaz, M. Q. (2017). Topp-Leone Inverse Weibull Distribution: Theory and Application. *European Journal of Pure and Applied Mathematics*, 10: 1005-1022.

[2] Alzaartreh, A., Lee, C., and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1): 63-79.

[3] Aryal, G. R., Ortega, E. M., Hamedani, G. G. and Yousof, H. M. (2017). The Topp-Leone Generated Weibull Distribution: Regression Model, Characterizations and Applications. *International Journal of Statistics and Probability*, 6: 126-141.

[4] Bourguignon, M., Silva, R. B., and Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12: 53-68.

[5] Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics*, 13: 215-232.

[6] Cordeiro, G. M., and de Castro, M. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81(7): 883-898.

[7] Dey, S., Ghosh, I. and Kumar, D. (2018). Alpha-power transformed Lindley distribution: properties and associated inference with application to earthquake data. *Annals of Data Science*, 1-28.

[8] Eghwerido, J. T. (2021). The alpha power Teissier distribution and its applications. *Afrika Statistika*, 16: 2731-2744.

[9] Ehiwario, J. C., Igabari, J. N. and Ezimadu, P. E. (2023). The alpha power Topp-Leone distribution: properties, simulations and applications. *Journal of Applied Mathematics and Physics*, 11: 316-331.

[10] Elbatal, I., Elgarhy, M. and Kibria, B. M. G. (2021). Alpha power transformed Weibull-G family of distributions: theory and applications. *Journal of Statistical Theory and Applications*, 20: 340-354 https://doi.org/10.2991/jsta.d.210222.002

[11] Elgarhy, M., Nasir, M. A., Jamal, F. F. and Ozel, G. (2018). The type II Topp-Leone generated family of distributions: Properties and applications. *Journal of Statistics and Management Systems*, 21: 1529-1551.

[12] Eugene, N. Lee, C. and Famoye, F. (2002). The beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31: 497-512.

[13] Gradshteyn, I. and Ryzhik, I. (2007). Table of Integrals. Series and Products, Elsevier/Academic Press.

[14] Greenwood, J. A, Landwehr, J. M. and Matalas, N. C. (1979). Probability weighted moments: Definitions and relations of parameters of several distributions expressible in inverse form. *Water Resources Research*, 15: 1049-1054.

[15] Hassan, A. S., Elgarhy, M., Mohamd, R. E. and Alrajhi, S. (2019). On the alpha power transformed power Lindley distribution. *Journal of Probability and Statistics*, Article ID8024769.

[16] Korkmaz, M. and Chesneau, C. (2021). On the unit Burr-XII distribution with the quantile regression modeling and applications. *Computational and Applied Mathematics*, 40: 1-26 https://doi.org/10.1007/s40314-021-01418-5

[17] Kumaraswamy, P. (1980). A generalized probability density function for doubly bounded random process. *Journal of Hydrology*, 46: 79-88.

[18] Lindley, D. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society*, **20:** 102-107.

[19] Mahdavi, A. and Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics - Theory and Methods*, 46: 6543-6557.

[20] Malik, A. S., Ahmad, S. P. (2017). Alpha power Rayleigh distribution and its application to life time data. *International Journal of Enhanced Research in Management and Computer Applications*, 6: 212-219.

[21] Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with applications to the exponential and Weibull families, *Biometrika*, 84: 641-652.

[22] Modi, K. and Gill, V. (2020). Unit Burr-III distribution with application. *Journal of Statistics and Management Systems*, 23: 579-592.

[23] Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure rate data. *IEEE Transactions on Reliability*, 42: 299-302.

[24] Opone, F. C. and Ekhosuehi, N. (2017). A study on the moments and performance of the maximum likelihood estimates (mle) of the beta distribution. *Journal of the Mathematical Association of Nigeria (Mathematics Science Series)*, 44: 148-154.00

[25] Opone, F. C., Ekhosuehi, N. and Omosigho, S. E. (2022). Topp-Leone Power Lindley Distribution (TLPD): Its properties and application. *Sankhya A*, 84(2): 597-608. https://doi.org/10.1007/s13171-020-00209-0

[26] Opone, F. C. and Iwerumor, B. N. (2021). A new Marshall-Olkin extended family of distributions with bounded support. *Gazi University Journal of Science*, 34: 899-914.

[27] Opone, F. C. and Osemwenkhae, J. E. (2022). The transmuted Marshall-Olkin extended Topp-Leone Distribution. *Earthline Journal of Mathematical Sciences*, 9 (2022), 179-199.

[28] Opone, F. C. Ubaka, O. N. and Karakaya, K. (2023). Statistical analysis of Covid-19 data using the odd log-logistic Kumaraswamy distribution. *Statistics, Optimization and Information Computing* (To appear)

[29] Ristic, M. M. and Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 82: 1191-1206.

[30] Shaw, W. and Buckley, I. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions and a skew-kurtoticnormal distribution from a rank transmutation map. arXivprereprint, arXiv, 0901.0434.

[31] Topp, C. W. and Leone, F. C. (1955). A family of J-shaped frequency functions. *Journal of the American Statistical Association*, 50: 209-219.

[32] Tuoyo, D., Opone, F. C. and Ekhosuehi, N. (2021). The Topp-Leone Weibull Distribution: Its Properties and Application. *Earthline Journal of Mathematical Sciences*, 7: 381-401. https://doi.org/10.34198/ejms.7221.381401

[33] ZeinEldin, R. A., Ahsan ulHaq, M., Hashmi, S. and Elsehety, M. (2020). Alpha power transformed inverse Lomax distribution with different methods of estimation and applications. *Complexity*, Article ID 1860813. (2020) https://doi.org/10.1155/2020/1860813.

NOVEL DISTRIBUTION FOR MODELING UNCENSORED AND CENSORED SURVIVAL TIME DATA AND REGRESSION MODEL

Adubisi O. D.[1]
Adubisi C. E.[2]

•

[1]Department of Mathematics and Statistics, Federal University Wukari, Nigeria.
[2]Department of Physics, University of Ilorin, Nigeria.

adubisiobinna@fuwukari.edu.ng[1]
chidiadubisi@gmail.com[2]

Abstract

This work proposes a new one-parameter model titled the type II Topp Leone half logistic (TIITL$_{HL}$) model which is characterized by an increasing and decreasing hazard rate function quite dependent on the shape parameter. Some structural properties and basic functions used in reliability analysis are derived. Simulations are carried out for both uncensored and censored samples. The uncensored simulation results indicated that the estimators perform quite well in producing good parameter estimates at finite sample sizes. However, the Anderson Darling estimator (ADE) average estimate tend to the true parameter value faster than other methods with minimum bias. More so, simulation based on censored samples using different censoring proportions showed that the bias, MSE and MRE values decrease as the sample size increases for the different censoring proportions. Two uncensored and censored datasets from the medical and environmental sciences were analysed to show the relevance, flexibility and adaptability of the TIITL$_{HL}$ model, and the new model achieved the best performance when compared with six other competing lifetime models. In addition, the log-TIITL$_{HL}$ regression model constructed and compared with two existing models showed that this model will be a useful option in survival investigation.

Keywords: Half-logistic distribution, Classical estimation methods, Monte-Carlo simulation, type II censoring, Type-II-Topp-Leone-G class

1. Introduction

The standard half-Logistic (HL) model pioneered by [1] has gained a lot of popularity as a significant model given its extensive applicability in lifetime modeling and reliability analysis. The cumulative density function (CDF) of the standard HL model is

$$G(k) = \frac{1 - e^{-k}}{1 + e^{-k}}, k > 0, \tag{1}$$

and the corresponding probability density function (PDF) to (**??**) is

$$g(k) = \frac{2e^{-k}}{(1 + e^{-k})^2}, k > 0, \tag{2}$$

Several authors have pioneered various extensions of the HL model such as the type I HL family of distributions by [2], inverse HL model by [3], Poisson HL model by [4], type II HL family of

distributions by [5], Kumaraswamy HL model by [6], extended HL model by [7], Transmuted HL model by [8], new type I HL model by [9], odd Lindley HL model by [10], weighted HL model by [11], modified HL model by [12], Poisson-logarithmic HL model by [13], extended type I HL family of distributions by [14] and Gamma power HL model by [15].

The type II Topp Leone-G (TIITL-G) class of distributions was pioneered by [16]. The TIITL-G class has just one parameter which imply that the proposed extended HL model in this study will have a single shape parameter. The CDF and PDF of the TIITL-G class are

$$F(k) = 1 - \left[1 - G^2(k)\right]^{\tau}, \tag{3}$$

and

$$f(k) = 2\tau g(k) G(k) \left[1 - G^2(k)\right]^{\tau-1}, \tag{4}$$

where $\tau > 0$ is a shape parameter, $G(k)$ and $g(k)$ are considered as the CDF and PDF of the baseline model. The novelty of this study is the creation of a new one-parameter lifetime model titled the TIITL_{HL} model, investigation of six different estimation methods for the new model with applicability to uncensored and censored survival time datasets, and introduction of a new log-TIITL_{HL} regression model for analysing censored response variable.

The remaining parts are outlined like this: Part 2 introduces the CDF and PDF of the TIITL_{HL} model. Part 3 presents reliability analysis and several important structural properties of the TIITL_{HL} model. Six classical estimation approaches are discussed in Part 4 to appreciate the parameters of the TIITL_{HL} model for uncensored sample. The maximum likelihood estimator based on the type-II right censored scheme is presented in Part 5. The finite sample performance of the TIITL_{HL} estimators is presented in Part 6 using Monte Carlo experiments. Part 7 deals with a new TIITL_{HL} regression model. The applications and empirical results are presented in Part 8. Finally, Part 9 presents the conclusion.

2. Model Genesis

This part introduces a new one-parameter model called the TIITL_{HL} model by inserting Eqs (1) and (2) into Eqs (3) and (4), then the CDF, PDF, survival function and hazard rate function (HRF) of the TIITL_{HL} model are

$$F(k,\tau) = 1 - \left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau}, k > 0, \tau > 0, \tag{5}$$

$$f(k,\tau) = \frac{4\tau e^{-k}}{(1+e^{-k})^2}\left(\frac{1-e^{-k}}{1+e^{-k}}\right)\left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau-1}, k > 0, \tau > 0, \tag{6}$$

Figure 1 depicts the graphical shapes of the TIITL_{HL} density function (PDF) with selected values for τ. The density function (PDF) is uni-modal, right-skewness, and heavy-tailed. The survival (Reliability) function (SF) and hazard (failure) rate (HRF) of the TIITL_{HL} model, take the forms

$$S(k,\tau) = \left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau}, \tag{7}$$

and

$$h(k,\tau) = \frac{4\tau e^{-k}}{(1+e^{-k})^2}\left(\frac{1-e^{-k}}{1+e^{-k}}\right)\left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{-1}, \tag{8}$$

More so, the reversed HRF of the TIITL_{HL} model is

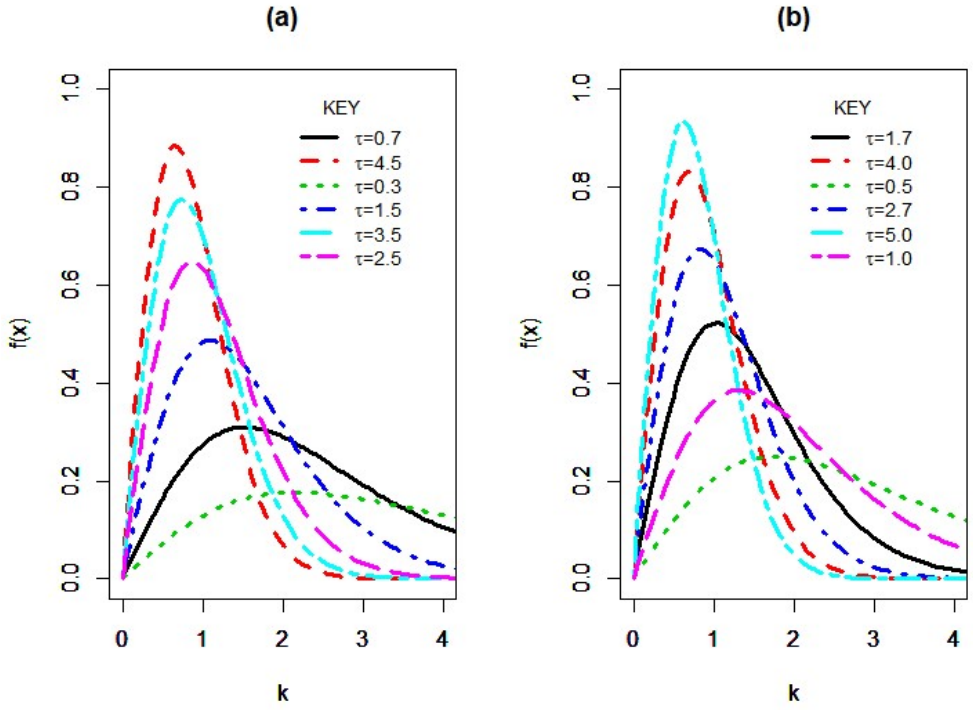


Figure 1: *The density function (PDF) plots of the TIITL$_{HL}$ model.*

$$r(k,\tau) = \frac{4\tau e^{-k}}{(1+e^{-k})^2}\left(\frac{1-e^{-k}}{1+e^{-k}}\right)\left[1-\left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau-1}\left\{1-\left[1-\left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau}\right\}^{-1}, \quad (9)$$

and the cumulative HRF takes the form

$$H(k,\tau) = -\tau\log\left[1-\left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]. \quad (10)$$

The graphical shapes of the HRF for TIITL$_{\text{HL}}$ model with various selected values of τ are depicted in Figure 2. The model is characterized by an increasing-decreasing HRF.

3. Structural Properties

This part describes the statistical properties of the TIITL$_{\text{HL}}$ model..

3.1. Quantile function, Bowley's skewness and Moor's kurtosis

If the random variable (r.v) $K \sim \text{TIITL}_{\text{HL}}(\tau)$, then the quantile function by inverting Eq (5) takes the form

$$k = -\log\left\{\frac{1-[1-(1-u)^{\frac{1}{\tau}}]^{\frac{1}{2}}}{1+[1-(1-u)^{\frac{1}{\tau}}]^{\frac{1}{2}}}\right\}, \quad (11)$$

where $u \sim uniform(0,1)$. By setting $u = 0.5$ in Eq (11), the median (M) of the TIITL$_{\text{HL}}$ model takes the form

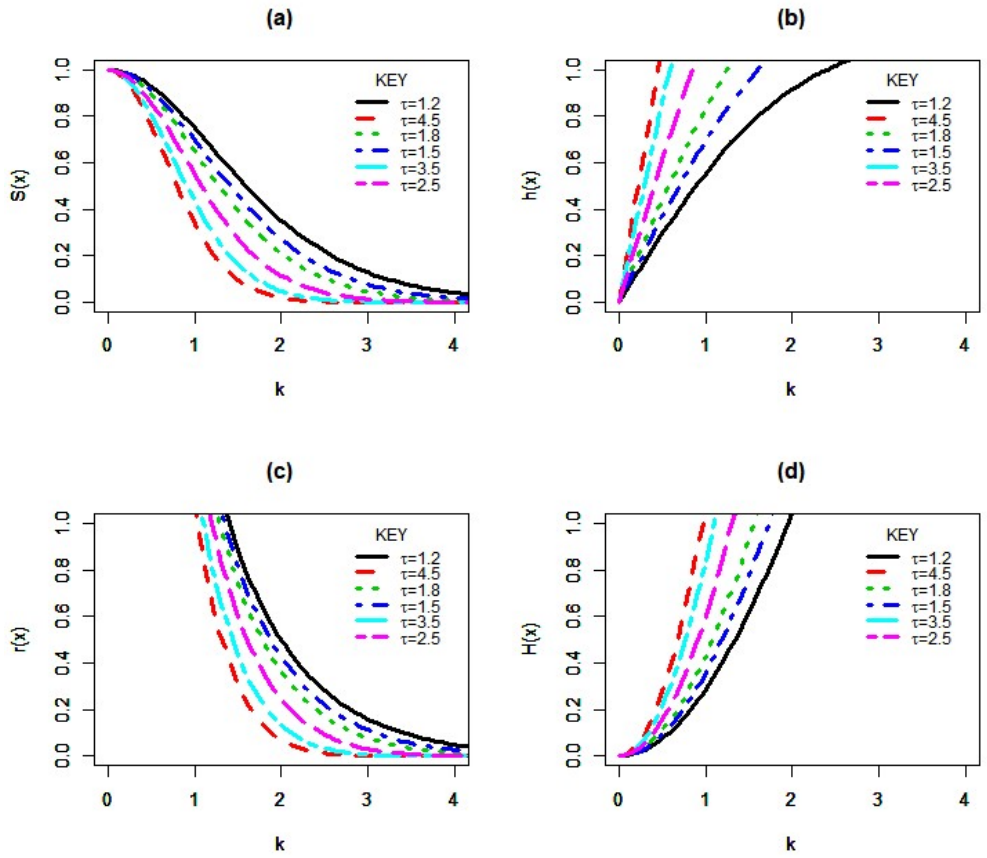


Figure 2: *Survival function plot (a), hazard rate function plot (b), reversed HRF plot (c) and cumulative HRF plot (d) of the TIITL$_{HL}$ model.*

$$k = -\log\left\{\frac{1-[1-(0.5)^{\frac{1}{\tau}}]^{\frac{1}{2}}}{1+[1-(0.5)^{\frac{1}{\tau}}]^{\frac{1}{2}}}\right\}. \tag{12}$$

The Bowley's skewness [17] and Moor's kurtosis [18] are found using the following expressions, respectively.

$$S_k = \frac{Q\left(\frac{3}{4};\tau\right) - 2Q\left(\frac{1}{2};\tau\right) + Q\left(\frac{1}{4};\tau\right)}{Q\left(\frac{3}{4};\tau\right) - Q\left(\frac{1}{4};\tau\right)}. \tag{13}$$

$$K_u = \frac{Q\left(\frac{7}{8};\tau\right) - Q\left(\frac{5}{8};\tau\right) - Q\left(\frac{3}{8};\tau\right) + Q\left(\frac{1}{8};\tau\right)}{Q\left(\frac{6}{8};\tau\right) - Q\left(\frac{2}{8};\tau\right)}. \tag{14}$$

where $Q(.)$ is the quantile function.

3.2. Dispersion index and Coefficient of variation

The dispersion index (DI) tells when a model is suitable for modeling equi-dispersed $(DI = 1)$, under-dispersed $(DI < 1)$ and over-dispersed $(DI > 1)$. The coefficient of variation (CV) is a relative measure of variability and a high CV value shows higher variability. The expressions for the DI and CV functions are

$$DI = \frac{Var(X)}{E(X)} = \frac{\frac{Q\left(\frac{3}{4};\tau\right)-Q\left(\frac{1}{4};\tau\right)}{1.35}}{\frac{Q\left(\frac{3}{4};\tau\right)+Q\left(\frac{1}{2};\tau\right)+Q\left(\frac{1}{4};\tau\right)}{3}}. \tag{15}$$

and

$$CV = \frac{(Var(X))^{\frac{1}{2}}}{E(X)} = \frac{(\frac{Q\left(\frac{3}{4};\tau\right)-Q\left(\frac{1}{4};\tau\right)}{1.35})^{\frac{1}{2}}}{\frac{Q\left(\frac{3}{4};\tau\right)+Q\left(\frac{1}{2};\tau\right)+Q\left(\frac{1}{4};\tau\right)}{3}}. \tag{16}$$

where $Q(.)$ is the quantile function.

Table 1 reports the numerical values of the mean (ME), variance (VAR), standard deviation (STD), median (M), skewness (S_k), kurtosis (K_u), Dispersion index (DI) and Coefficient of variation (CV) for the $TIITL_{HL}$ model using selected values of τ.

Table 1: *The numerical values of ME, VAR, STD, M, S_k, K_u, DI and CV*

τ	ME	VAR	STD	M	S_k	K_u	DI	CV
0.2	5.282	17.356	4.166	4.836	0.238	0.629	3.286	0.789
0.5	2.784	3.531	1.879	2.634	0.177	0.484	1.268	0.675
1.0	1.832	1.293	1.137	1.763	0.135	0.370	0.706	0.621
1.5	1.459	0.769	0.877	1.412	0.117	0.321	0.527	0.601
2.0	1.247	0.543	0.737	1.212	0.108	0.294	0.436	0.591
2.5	1.107	0.419	0.647	1.078	0.102	0.277	0.378	0.584
3.0	1.006	0.341	0.584	0.980	0.097	0.266	0.339	0.581
3.5	0.928	0.287	0.536	0.905	0.095	0.257	0.310	0.578
4.0	0.865	0.248	0.498	0.845	0.092	0.251	0.287	0.576
4.5	0.814	0.218	0.467	0.795	0.091	0.246	0.268	0.574
5.0	0.771	0.194	0.441	0.753	0.089	0.242	0.252	0.572

The ME, STD, S_k and K_u values of the $TIITL_{HL}$ model decrease as the selected values of τ increase. The $TIITL_{HL}$ model is positively skewed and beneficial for over-and-under dispersed datasets. Figure 3 depict the plots of the ME, VAR, S_k and K_u of the $TIITL_{HL}$ for selected values of τ and support the conclusion reached using Table 1.

3.3. Moments and Moment generating function

The r^{th} raw moment of the $TIITL_{HL}$ model is given as

$$\mu'_r = \sum_{a=0}^{\tau-1}\sum_{b,c=0}^{\infty} \vartheta_{a,b,c}\,(b+c+1)^{-r-1}\,\Gamma\,(r+1)\,. \tag{17}$$

Proof. The r^{th} raw moment of the $TIITL_{HL}$ model is found using

$$\begin{aligned} \mu'_r &= \textstyle\int_0^{\infty} k^r f\,(k;\tau)\,dk, \\ &= 4\tau \textstyle\int_0^{\infty} k^r \frac{e^{-k}}{(1+e^{-k})^2}\left(\frac{1-e^{-k}}{1+e^{-k}}\right)\left[1-\left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau-1} dk, \end{aligned} \tag{18}$$

By utilising Taylor series expansions in Eq (**??**), we have

$$\mu'_r = \sum_{a=0}^{\tau-1}\sum_{b,c=0}^{\infty} \vartheta_{a,b,c}\int_0^{\infty} k^r e^{-k(b+c+1)}dk, \tag{19}$$

where $\vartheta_{a,b,c} = 4\tau\,(-1)^{a+b+c}\begin{pmatrix} \tau-1 \\ a \end{pmatrix}\begin{pmatrix} 2a+1 \\ b \end{pmatrix}\begin{pmatrix} -2\,(1+a)-1 \\ c \end{pmatrix}$.

Let

$z = k\,(b+c+1) \Rightarrow k = \frac{z}{(b+c+1)}$,

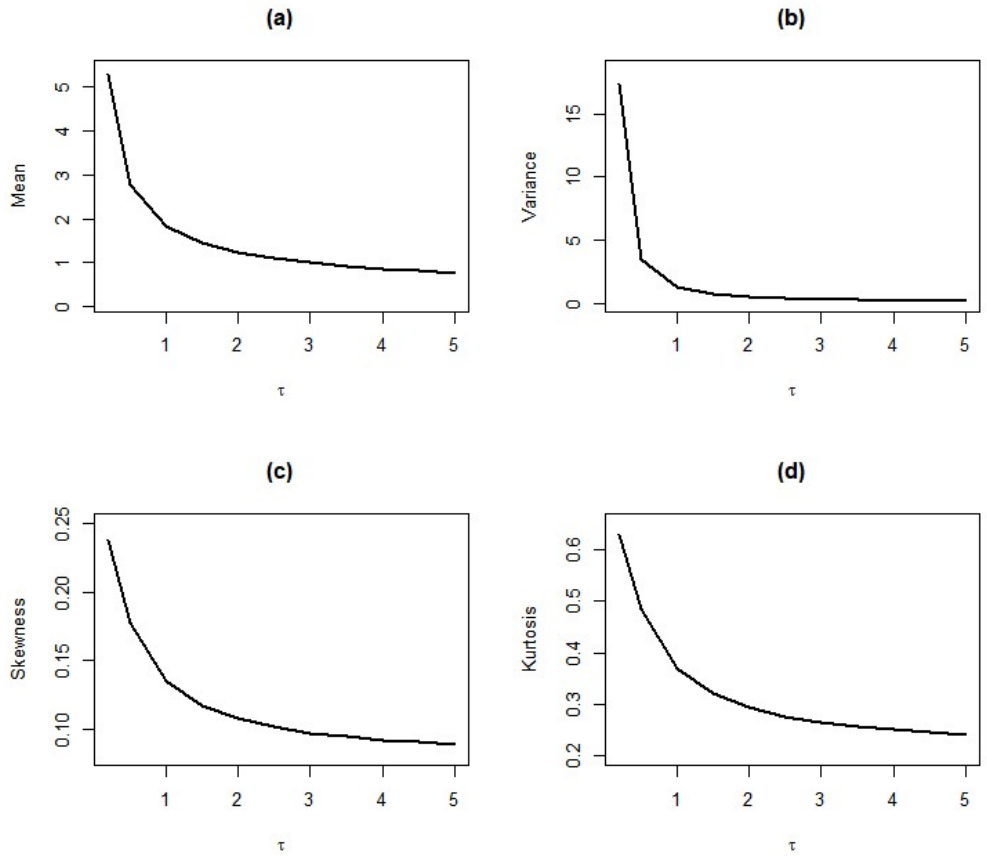


Figure 3: *The Mean, Variance, skewness and kurtosis plots of the TIITL$_{HL}$ model.*

$\frac{dk}{dz} = \frac{1}{(b+c+1)} \Rightarrow dk = \frac{dz}{(b+c+1)}$.
Hence,

$$\mu'_r = \sum_{a=0}^{\tau-1} \sum_{b,c=0}^{\infty} \vartheta_{a,b,c} \int_0^{\infty} \left(\frac{z}{(b+c+1)} \right)^r e^{-z} \frac{dz}{(b+c+1)}, \tag{20}$$

$$\mu'_r = \sum_{a=0}^{\tau-1} \sum_{b,c=0}^{\infty} \vartheta_{a,b,c} \left(b+c+1\right)^{-r-1} \int_0^{\infty} z^r e^{-z} dz, \tag{21}$$

By utilising the gamma integral function $\Gamma\left(\alpha+1\right) = \int_0^{\infty} z^{\alpha} e^{-z} dz$. The rth raw moments of the TIITL$_{\text{HL}}$ model takes the form

$$\mu'_r = \sum_{a=0}^{\tau-1} \sum_{b,c=0}^{\infty} \vartheta_{a,b,c} \left(b+c+1\right)^{-r-1} \Gamma\left(r+1\right). \tag{22}$$

The first four moments are found by inserting $r = 1, 2, 3, 4$ into Eq (22), respectively. ■
The moment generating function (MGF) of the TIITL$_{\text{HL}}$ model is given as

$$M_K\left(t\right) = \sum_{r=0}^{\infty} \sum_{a=0}^{\tau-1} \sum_{b,c=0}^{\infty} \frac{t^r \vartheta_{a,b,c}}{r!} \left(b+c+1\right)^{-r-1} \Gamma\left(r+1\right), \tag{23}$$

Proof.
The MGF of the TIITL$_{\text{HL}}$ model, say $M_K\left(t\right)$ is found using

$$M_K(t) = \mathrm{E}\left(e^{tk}\right) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^{\infty} K^r f(k;\tau)\, dk = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu'_r, \tag{24}$$

By inserting Eq (7) into Eq (24), the MGF takes the form

$$M_K(t) = \sum_{r=0}^{\infty} \sum_{a=0}^{\tau-1} \sum_{b,c=0}^{\infty} \frac{t^r \vartheta_{a,b,c}}{r!} (b+c+1)^{-r-1} \Gamma(r+1). \tag{25}$$

■

3.4. Order statistics

If $k_1, k_2, \ldots, k_n$ be a random sample from the $\mathrm{TIITL_{HL}}$ model with $k_{1:n} < k_{2:n} < \ldots < k_{n:n}$ as the order statistics (O.S). The pdf of the $\mathrm{p^{th}}$ O.S of the $\mathrm{TIITL_{HL}}$ model is

$$f_{p:n}(k) = \frac{4\tau e^{-k}(1+e^{-k})^{-2}(\varphi)\, n!}{(p-1)!\,(n-p)!} \left\{1 - \left[1-(\varphi)^2\right]^{\tau}\right\}^{p-1} \left[1-(\varphi)^2\right]^{\tau[(n-p)+1]-1}. \tag{26}$$

Proof. The pdf of the $\mathrm{p^{th}}$ O.S can be found using

$$f_{p:n}(k) = \frac{n!}{(p-1)!\,(n-p)!} g(k) [G(k)]^{p-1} [1-G(k)]^{n-p}, \tag{27}$$

where $\mathrm{B}(.,.)$ is the beta function. By inserting Eqs (5) and (6) into Eq (27), the pdf of the $\mathrm{p^{th}}$ O.S of the $\mathrm{TIITL_{HL}}$ model after some simplification takes the form

$$f_{p:n}(k) = \frac{4\tau e^{-k}(1+e^{-k})^{-2}(\varphi)\, n!}{(p-1)!\,(n-p)!} \left\{1 - \left[1-(\varphi)^2\right]^{\tau}\right\}^{p-1} \left[1-(\varphi)^2\right]^{\tau[(n-p)+1]-1}, \tag{28}$$

where $\varphi = \left(\frac{1-e^{-k}}{1+e^{-k}}\right)$. By substituting $p = 1$ and $p = n$ into Eq (28), the lowest and highest order statistics are obtained. ■

4. Methods of estimation for uncensored sample

In this part, the parameter of the $\mathrm{TIITL_{HL}}$ model is estimated via the maximum likelihood estimation (MLE), maximum product spacing estimation (MPSE), Anderson Darling estimation (ADE), least square estimation (LSE), weighted least square estimation (WLSE), and Cramer Von Mises estimation (CVME).

4.1. The MLE

If $k_1, k_2, \ldots, k_n$ be the random observed values from $\mathrm{TIITL_{HL}}$ model. Then, the MLE function $\mathrm{L}(\tau)$ takes the form

$$L(\tau) = (4\tau)^n \prod_{i=1}^{n} \frac{e^{-k_i}}{\left(1+e^{-k_i}\right)^2} (\varphi_i) \left[1-(\varphi_i)^2\right]^{\tau-1} \tag{29}$$

where $\varphi_i = \left(\frac{1-e^{-k_i}}{1+e^{-k_i}}\right)$. The log-likelihood function of the $\mathrm{TIITL_{HL}}$ model takes the form

$$\log(L(\tau)) = n\log(4\tau) - \sum_{i=1}^{n} k_i - 2\sum_{i=1}^{n} \log\left(1+e^{-k_i}\right) + \sum_{i=1}^{n} \log(\varphi_i) + (\tau-1) \sum_{i=1}^{n} \log\left[1-(\varphi_i)^2\right], \tag{30}$$

The first derivative of Eq (30) with respect to τ is

$$\frac{\partial \log\left(L\left(\tau\right)\right)}{\partial \tau} = \frac{n}{\tau} + \sum_{i=1}^{n} \log\left[1 - \left(\varphi_i\right)^2\right]. \tag{31}$$

The R (*optim function*) is employed to estimate the TIITL_{HL} parameter using numerical approaches.

4.2. The LSE and WLSE

Minimizing with respect to τ, the LS estimate $\hat{\tau}_{\text{LS}}$ can be found using

$$LS\left(\tau\right) = \sum_{i=1}^{n} \left[1 - \left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau} - \frac{i}{n+1}\right]^2. \tag{32}$$

Likewise, minimizing with respect to τ, the WLS estimates $\hat{\tau}_{\text{WLS}}$ can be found using

$$WLS\left(\tau\right) = \sum_{i=1}^{n} \frac{\left(n+2\right)\left(n+1\right)^2}{i\left(n-i+1\right)} \left[1 - \left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau} - \frac{i}{n+1}\right]^2. \tag{33}$$

4.3. The MPSE

The MPS for the TIITL_{HL} model with ordered sample $k_{(1:n)}, k_{(2:n)}, \ldots, k_{(n:n)}$ is given as follows

$$GM\left(\tau | k_{n:n}\right) = \left[\prod_{i=1}^{n+1} D_i\left(k_i, \tau\right)\right]^{\frac{1}{n+1}}, \tag{34}$$

where $D_i\left(k_i, \tau\right) = F\left(k_{(i:n)}\,|\tau\right) - F\left(k_{(i-1:n)}\,|\tau\right);\ i = 1, 2, \ldots, n+1.$ and $F\left(k, \tau\right)$ is given in Eq (5).

4.4. The ADE

Minimizing with respect τ, the AD estimate $\hat{\tau}_{\text{AD}}$ can be found using

$$AD\left(\tau\right) = -n - \frac{1}{n}\sum_{i=1}^{n}\left(2i-1\right)\left[\log F\left(k_{(i:n)}\,|\tau\right) + \log \bar{F}\left(k_{(n+1-i:n)}\,|\tau\right)\right], \tag{35}$$

where $\bar{F}\left(k, \tau\right) = 1 - F\left(k, \tau\right)$ and $F\left(k, \tau\right)$ is given in Eq (5).

4.5. The CVME

Minimizing with respect τ, the CVM estimates $\hat{\tau}_{\text{CVM}}$ can be found using

$$CVM\left(\tau\right) = \frac{1}{12} + \sum_{i=1}^{n} \left[1 - \left[1 - \left(\frac{1-e^{-k}}{1+e^{-k}}\right)^2\right]^{\tau} - \frac{2(i-1)+1}{2n}\right]^2. \tag{36}$$

5. MLE for type II right censoring

Given that a fixed number of failed units have been observed, a life testing experiment is concluded. Then the remaining units are designated as type-II-censored. Let $k_{(1)}, k_{(2)}, \ldots, k_{(p)}, p \leq n$ denote the ordered values of a random sample (r.s) $k_1, k_2, \ldots, k_n$ (failure times) and observations cease after the p^{th} unsuccessful unit occurs, then the likelihood function is given

$$L\left(\tau; k\right) = \frac{n!}{(n-p)!}\left[R\left(k_p; \tau\right)\right]^{n-p} \prod_{i=1}^{p} f\left(k_i; \tau\right). \tag{37}$$

If $k_1, k_2, \ldots, k_n$ be r.s from the TIITL_{HL} (τ), then the likelihood function is

$$L(\tau;k) = \frac{n!}{(n-p)!}\left[1-\left(\frac{1-e^{-k_p}}{1+e^{-k_p}}\right)^2\right]^{\tau(n-p)} \prod_{i=1}^{p}\left\{\frac{4\tau e^{k_i}\left(\frac{1-e^{-k_i}}{1+e^{-k_i}}\right)}{(1+e^{-k_i})^2}\left[1-\left(\frac{1-e^{-k_i}}{1+e^{-k_i}}\right)^2\right]^{\tau-1}\right\}. \tag{38}$$

The log-likelihood function without the constant term is

$$\begin{aligned} l(\tau;k) \propto\ & p\left[\log(4)+\log(\tau)\right]+\tau(n-p)\log\left[1-\left(\tfrac{1-e^{-k_p}}{1+e^{-k_p}}\right)^2\right]-\textstyle\sum_{i=1}^{p}k_i \\ & -2\textstyle\sum_{i=1}^{p}\log(1+e^{-k_i})+\sum_{i=1}^{p}\log\left(\tfrac{1-e^{-k_i}}{1+e^{-k_i}}\right)+(\tau-1)\sum_{i=1}^{p}\log\left[1-\left(\tfrac{1-e^{-k_i}}{1+e^{-k_i}}\right)^2\right]. \end{aligned} \tag{39}$$

Setting $\frac{\partial}{\partial\tau}l(\tau;k)=0$. The MLE $(\hat{\tau})$ can be found as solution of

$$\frac{p}{\tau}+(n-p)\log\left[1-\left(\frac{1-e^{-k_p}}{1+e^{-k_p}}\right)^2\right]+\sum_{i=1}^{p}\log\left[1-\left(\frac{1-e^{-k_i}}{1+e^{-k_i}}\right)^2\right] \tag{40}$$

Using the R (*optim function*), the non-linear equation in Eq (40) is solved numerically to obtain the MLE $\hat{\tau}$.

6. Simulation

The Monte Carlo simulations for uncensored and censored samples are executed for the TIITL_{HL} parameter (Pa.).

6.1. Simulation based on uncensored sample

The simulations using MLE, LSE, WLSE, MPSE, ADE, and CVME approaches for the TIITL_{HL} parameter are presented in this subpart. The simulation is carried out as follows:

- Set the parameter value τ = 0.5, 2.5 for the Monte Carlo simulation process.
- Random samples of sizes n $= 20, 70, 150, 250, 350$ with replicates $N = 5000$ generated using Eq (11).
- The MLE, LSE, WLSE, MPSE, ADE, and CVME processes are executed to find the estimates of parameter (τ).
- Compute the average estimate (AVEs), absolute biases (ABs), mean square errors (MSEs) and mean relative error (MREs) using the information in the preceding step.

Tables 2 and 3 reports the AVEs, ABs, MSEs and MREs for the MLE, LSE, WLSE, MPSE, ADE, and CVME methods with different sample sizes. The results are graphically summarized in Figures 4 and 5. As seen from these graphs, the ABs, MSEs and MREs tend to zero as n increases for the six estimation methods. However, the ADE average estimate tend to the true parameter value faster than other estimation methods with minimum AB.

Table 2: *The six estimators AVE for* $\tau = 0.5$ *based on uncensored sample.*

n	Measures	MLE	MPSE	ADE	LSE	WLSE	CVME
20	AVE	0.527	0.489	0.518	0.522	0.520	0.526
70	AVE	0.507	0.492	0.504	0.505	0.505	0.506
150	AVE	0.503	0.495	0.502	0.502	0.502	0.503
250	AVE	0.502	0.496	0.501	0.501	0.501	0.501
350	AVE	0.501	0.497	0.500	0.500	0.500	0.500

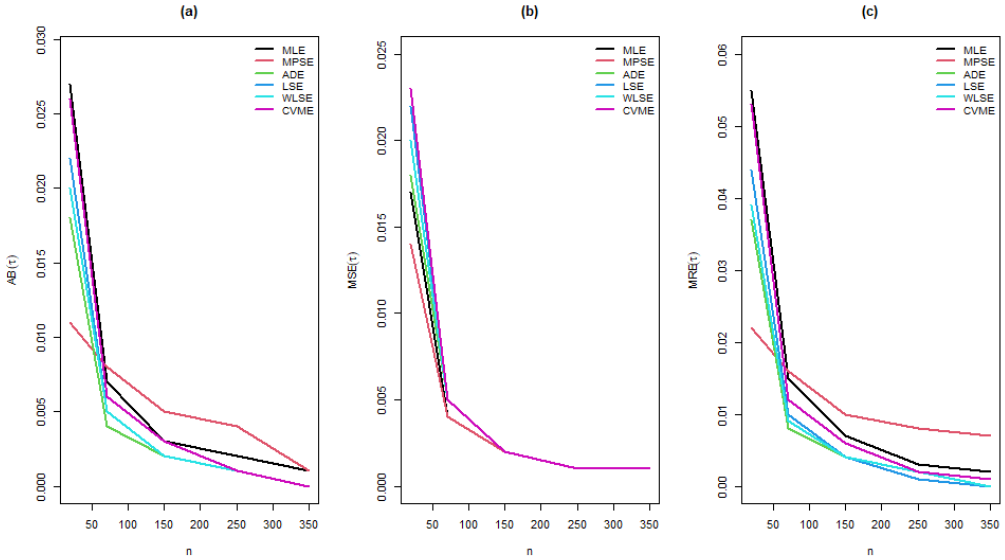


Figure 4: *AB, MSE and MRE of Estimators in Table 2.*

Table 3: *The six estimators AVE for* $\tau = 2.5$ *based on uncensored sample.*

n	Measures	MLE	MPSE	ADE	LSE	WLSE	CVME
20	AVE	2.637	2.444	2.591	2.610	2.598	2.632
70	AVE	2.536	2.460	2.521	2.525	2.524	2.531
150	AVE	2.517	2.475	2.509	2.511	2.511	2.514
250	AVE	2.508	2.480	2.503	2.504	2.504	2.505
350	AVE	2.504	2.483	2.500	2.501	2.501	2.502

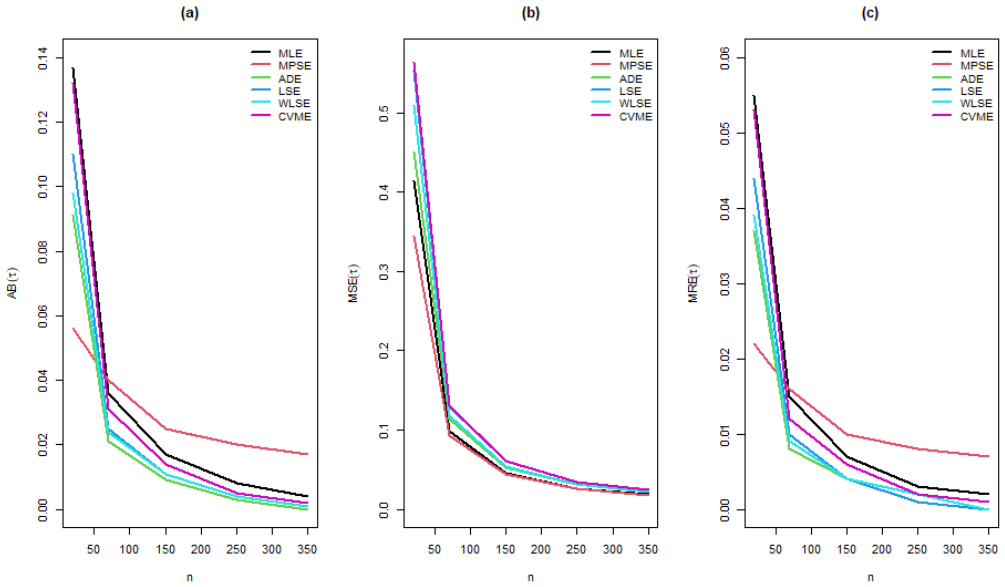


Figure 5: *AB, MSE and MRE of Estimators in Table 3.*

6.2. Simulation based on type-II-right censored sample

The simulation is executed for the MLE using different random sample sizes n = $(20, 70, 150, 250, 350)$ generated with Eq (11). The length of censored sample test is given by $p = nm$, where (m) is the censoring proportion $(0 < m < 1)$. Table 4 reports that the bias, MSE and MRE values decrease as the sample size increases for the different censoring proportions considered.

Table 4: *The MLE, Bias, MSE and MRE based on censored sample.*

n	m	τ	$\hat{\tau}$	Bias	MSE	MRE
20	0.3	0.5	0.881	0.381	0.145	0.762
		2.5	4.405	1.905	3.629	0.762
	0.5	0.5	0.752	0.252	0.063	0.503
		2.5	3.758	1.258	1.583	0.503
	0.7	0.5	0.781	0.281	0.079	0.562
		2.5	3.906	1.406	1.976	0.562
70	0.3	0.5	0.611	0.111	0.012	0.223
		2.5	3.057	0.557	0.310	0.223
	0.5	0.5	0.650	0.150	0.022	0.299
		2.5	3.248	0.748	0.560	0.299
	0.7	0.5	0.657	0.157	0.025	0.315
		2.5	3.287	0.787	0.620	0.315
150	0.3	0.5	0.551	0.051	0.003	0.101
		2.5	2.753	0.253	0.064	0.101
	0.5	0.5	0.554	0.054	0.003	0.108
		2.5	2.771	0.271	0.073	0.108
	0.7	0.5	0.573	0.073	0.005	0.145
		2.5	2.863	0.363	0.132	0.145
250	0.3	0.5	0.567	0.066	0.004	0.132
		2.5	2.831	0.331	0.110	0.132
	0.5	0.5	0.587	0.087	0.008	0.175
		2.5	2.936	0.436	0.190	0.175
	0.7	0.5	0.579	0.079	0.006	0.158
		2.5	2.896	0.396	0.157	0.158
350	0.3	0.5	0.485	-0.015	2E-04	0.030
		2.5	2.424	-0.076	0.006	0.030
	0.5	0.5	0.453	-0.047	0.002	0.095
		2.5	2.264	-0.236	0.056	0.095
	0.7	0.5	0.424	-0.076	0.006	0.153
		2.5	2.118	-0.382	0.146	0.153

7. The log-TIITL$_{\text{HL}}$ regression model

Let K denotes a random variable which follows the TIITL$_{\text{HL}}$ model with parameter τ. Utilizing the transformation $Y = \log(K)$ with location and scale parameters added, the density of Y is

$$f_{TIITL_{HL}}(y,\tau,\mu,\sigma) = \frac{4\tau}{\sigma}\exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right]\left\{1+\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{-2} \times\left\{\frac{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{1+\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right\}\left(1-\left\{\frac{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{1+\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right\}^{2}\right)^{\tau-1}, \tag{41}$$

where $\tau, \sigma > 0$, $y, \mu \in \Re$. The random variable Y has the log-TIITL$_{\text{HL}}$ (LTIITL$_{\text{HL}}$) model with location μ and scale σ parameters, respectively. The survival function to Eq (41) is

$$S_{TIITL_{HL}}(y,\tau,\mu,\sigma) = \left(1-\left\{\frac{1-\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}{1+\exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right]}\right\}^{2}\right)^{\tau}. \tag{42}$$

By inserting $z = (y-\mu)/\sigma$, $z \in \Re$ into Eq (41). The standardized log-TIITL$_{\text{HL}}$ density takes the form

$$f_{TIITL_{HL}}(z,\tau) = \frac{4\tau}{\sigma}\exp\left[(z)-\exp(z)\right]\left\{1+\exp\left[-\exp(z)\right]\right\}^{-2} \times\left\{\frac{1-\exp\left[-\exp(z)\right]}{1+\exp\left[-\exp(z)\right]}\right\}\left(1-\left\{\frac{1-\exp\left[-\exp(z)\right]}{1+\exp\left[-\exp(z)\right]}\right\}^{2}\right)^{\tau-1}. \tag{43}$$

Let $\boldsymbol{K}_i = (k_{i1}, \ldots, k_{im})^T$ be the explanatory vector associated with the i^{th} response variable y_i for $i = 1, \ldots, n$. A regression model based on the TIITL_{HL} density function is given by

$$y_i = \boldsymbol{K}_i^T \boldsymbol{\beta} + \sigma z_i, \; i = 1, \ldots, n, \tag{44}$$

where z_i is the random error which follows the density function Eq (43), $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^T, \sigma > 0, \; \tau > 0$ are unknown parameters and $\boldsymbol{K}_i$ is modeling $\mu_i = \boldsymbol{K}_i^T \boldsymbol{\beta}$. The density and survival functions of y_i are

$$\begin{aligned} f_{TIITL_{HL}}(y_i; \tau, \sigma, \boldsymbol{\beta}^T) = \tfrac{4\tau}{\sigma} \exp\left[z_i - \exp\left(z_i\right)\right] \left\{1 + \exp\left[-\exp\left(z_i\right)\right]\right\}^{-2} \\ . \qquad \times \left\{\tfrac{1-\exp[-\exp(z_i)]}{1+\exp[-\exp(z_i)]}\right\} \left(1 - \left\{\tfrac{1-\exp[-\exp(z_i)]}{1+\exp[-\exp(z_i)]}\right\}^2\right)^{\tau-1}, \end{aligned} \tag{45}$$

and

$$S_{TIITL_{HL}}(y_i; \tau, \sigma, \boldsymbol{\beta}^T) = \left(1 - \left\{\frac{1 - \exp\left[-\exp\left(z_i\right)\right]}{1 + \exp\left[-\exp\left(z_i\right)\right]}\right\}^2\right)^{\tau}, \tag{46}$$

Let F and C denote the sets of units for which y_i is the log-lifetime or log-censoring, respectively. The log-likelihood for $\boldsymbol{\vartheta} = \left(\tau, \sigma, \boldsymbol{\beta}^T\right)^T$ from Eq (44) is

$$\begin{aligned} l\left(\boldsymbol{\vartheta}\right) = r \log\left(\tfrac{4\tau}{\sigma}\right) + \textstyle\sum_{i \in F}\left[z_i - \exp\left(z_i\right)\right] - 2\sum_{i \in F} \log\left\{1 + \exp\left[-\exp\left(z_i\right)\right]\right\} \\ + \textstyle\sum_{i \in F} \log\left\{\tfrac{1-\exp[-\exp(z_i)]}{1+\exp[-\exp(z_i)]}\right\} + (\tau - 1)\sum_{i \in F} \log\left(1 - \left\{\tfrac{1-\exp[-\exp(z_i)]}{1+\exp[-\exp(z_i)]}\right\}^2\right) \\ + \tau \textstyle\sum_{i \in C} \log\left(1 - \left\{\tfrac{1-\exp[-\exp(z_i)]}{1+\exp[-\exp(z_i)]}\right\}^2\right). \end{aligned} \tag{47}$$

where $z_i = \left(y_i - \boldsymbol{K}_i^T \boldsymbol{\beta}\right)/\sigma$ and r is the number of uncensored observations (failures). The MLE $\hat{\vartheta}$ of ϑ can be obtained by maximizing Eq (47) using the R (*optim function*).

8. Applications

The potentiality of the introduced model is illustrated by means of five applications.

8.1. Applications to uncensored data

The first dataset consist of survival time of 72 Guinea pigs infected with virulent tubercle bacilli. The data was initially reported by [19], and analysed by [6] and [20]. The second dataset, discussed by [21] consists of 30 observations of March precipitation (in inches) in Minneapolis/St Paul.

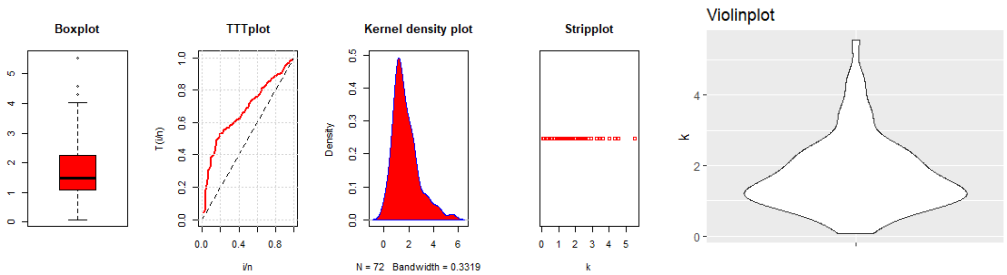


Figure 6: *Box, TTT, kernel-density, strip, and Violin plots of first uncensored data.*

Figures 6 and 7 depict the box plot, TTT plot, kernel density plot, strip plot, and Violin plot of the first and second datasets to check for outliers and symmetric nature. As depicted, both

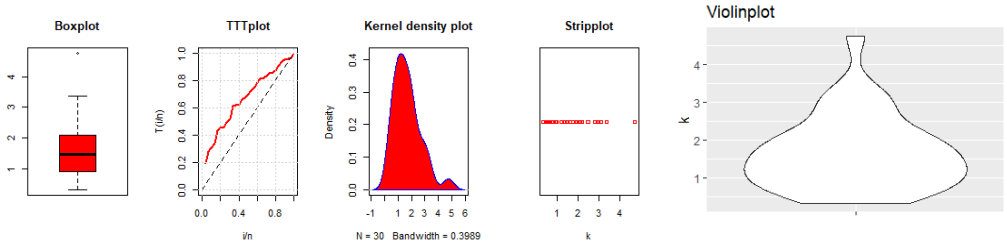


Figure 7: *Box, TTT, kernel-density, strip, and Violin plots of second uncensored data.*

datasets have certain outliers and are asymmetrical in nature. More so, the total test time (TTT) plot depicts that the utilized datasets have increasing HRF which means that the TIITL_{HL} model can be used to model these datasets.

We fit the datasets with the TIITL_{HL}, odd Lindley HL (OLIHL), Generalized HL (GHL), Inverse Lindley (ILIN), Rayleigh (R), Inverse Rayleigh (IR), and Muth (M) models. The MLEs for all the models are computed in R-software via the BFGS method (*optim* function). The W^*, A^* and KS measures (we used abbreviations) for model comparisons. The MLEs and their standard errors (SEs) in parentheses, log-likelihood (LL), and the information criteria are presented in Tables 5 and 6 for both datasets, respectively. The measures reveal that the TIITL_{HL} model provides an appropriate fit to both datasets (with lowest values of the goodness-of-fit statistics).

Table 5: *The MLEs and SEs of the fitted models with goodness-of-fit measures for first uncensored data.*

Model	MLEs (SEs)	LL	W*	A*	CAIC	AIC	BIC	HQIC	KS(P-Value)
TIITLHL	1.257(0.148)	-96.83	0.074	0.498	195.7	195.7	197.9	196.6	0.088(0.636)
OLIHL	0.289()0.024	-157.28	0.891	4.940	316.6	316.6	318.8	317.5	0.393(0.000)
GHL	0.786(0.093)	-104.94	0.080	0.531	211.9	211.9	214.2	212.8	0.218(0.002)
ILIN	1.359(0.124)	-129.20	0.929	5.634	260.5	260.4	262.7	261.3	0.232(0.001)
R	1.442(0.085)	-98.96	0.133	0.814	200.0	199.9	202.2	200.8	0.107(0.376)
IR	0.480(0.028)	-204.82	1.632	9.214	411.7	411.6	413.9	412.5	0.623(0.000)
M	0.480(0.028)	-123.46	0.118	0.735	249.0	248.9	251.2	249.8	0.434(0.000)

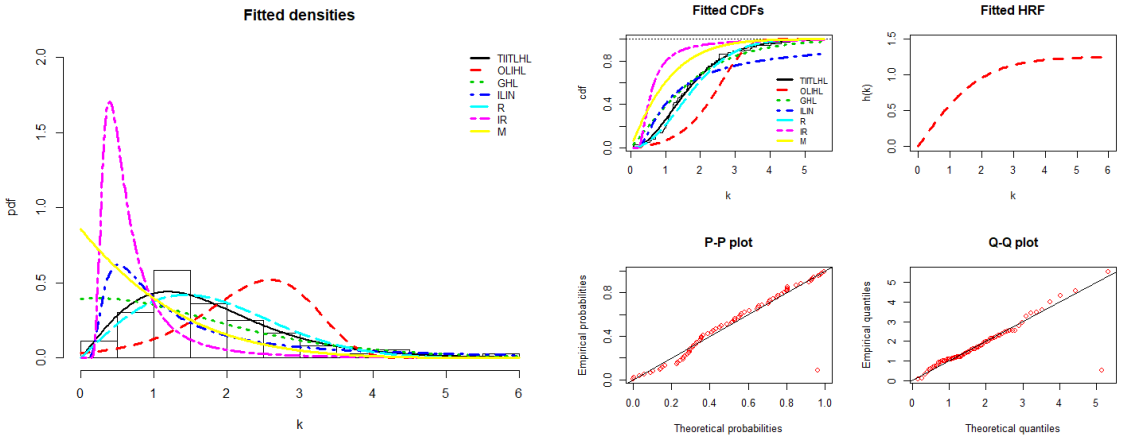


Figure 8: *The fitted pdfs (left panel) and cdfs (top-left,right panel) of competing models, fitted HRF (top-right,right panel), P-P plot (bottom-left,right panel) and Q-Q plot (bottom-right,right panel) of the TIITL_{HL} model for first uncensored data.*

Table 6: *The MLEs and SEs of the fitted models with goodness-of-fit measures for second uncensored data.*

Model	MLEs (SEs)	LL	W*	A*	CAIC	AIC	BIC	HQIC	KS(P-Value)
TIITLHL	1.351(0.247)	-38.34	0.01483	0.1155	78.83	78.68	80.08	79.13	0.059(0.999)
OLIHL	0.364(0.048)	-56.22	0.25812	1.6009	114.58	114.44	115.84	114.89	0.307(0.007)
GHL	0.828(0.151)	-42.43	0.01530	0.1195	87.01	86.87	88.27	87.32	0.182(0.274)
ILIN	1.583(0.227)	-45.22	0.09750	0.6014	92.59	92.44	93.84	92.89	0.228(0.089)
R	1.376(0.125)	-38.92	0.02607	0.1979	79.99	79.85	81.25	80.30	0.084(0.985)
IR	-0.927(0.085)	-44.14	0.16293	0.9881	90.42	90.27	91.67	90.72	0.240(0.064)
M	0.185(0.091)	-48.72	0.02471	0.1898	99.59	99.45	100.85	99.89	0.367(0.001)

Figures 8 and 9 display the histogram and Kaplan-Meier empirical cdf in conjunction with fitted pdfs and cdfs of the TIITL_{HL} and competing models for both datasets. Figures 8 and 9, also depict the fitted HRF of the TIITL_{HL} model with corresponding probability-probability (P-P) and quantile-quantile (Q-Q) plots. The superiority of the TIITL_{HL} model is supported by these figures.

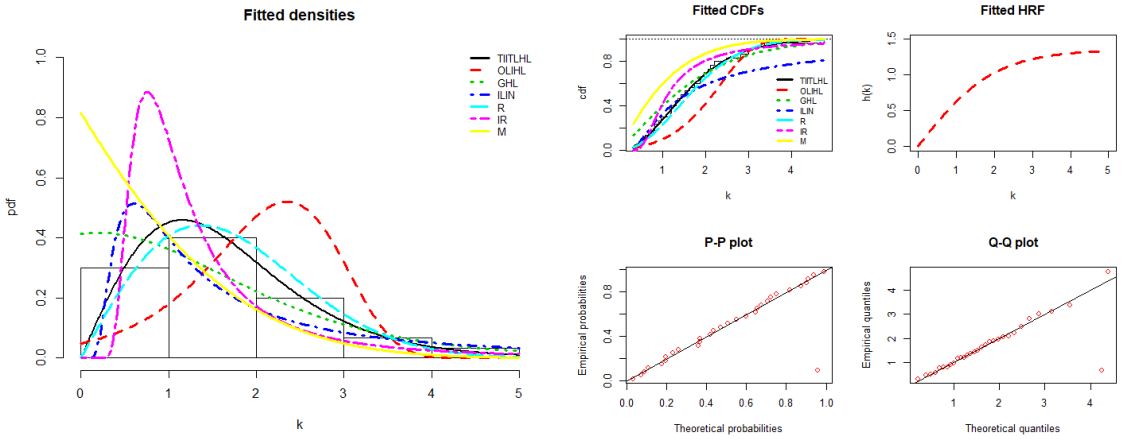


Figure 9: *The fitted pdfs (left panel) and cdfs (top-left,right panel) of competing models, fitted HRF (top-right,right panel), P-P plot (bottom-left,right panel) and Q-Q plot (bottom-right,right panel) of the TIITL_{HL} model for second uncensored data.*

Tables 7 and 8 report the estimation of the TIITL_{HL} parameter based on six estimation methods for the both uncensored datasets. From the results in Tables 7, the MPSE and WLSE are considered as the appropriate methods with smaller KS and larger P-values than the other methods. Likewise, the results in Table 8 attest that the MLE, ADE and WLSE are the appropriate methods with smaller KS and larger P-values than the other methods.

Table 7: *The results of the six estimation methods for first uncensored data.*

Estimate↓Method→	MLE	MPSE	CVME	ADE	LSE	WLSE
τ	1.257	1.215	1.279	1.262	1.266	1.229
KS	0.088	0.081	0.091	0.089	0.089	0.083
P-Value	0.600	0.700	0.600	0.600	0.610	0.700

Table 8: *The results of the six estimation methods for second uncensored data.*

Estimate↓Method→	MLE	MPSE	CVME	ADE	LSE	WLSE
τ	1.351	1.266	1.354	1.347	1.011	1.351
KS	0.059	0.073	0.060	0.059	0.155	0.059
P-Value	0.999	0.999	0.999	0.999	0.470	0.999

8.2. Applications to censored data

The first dataset represents the relief times (in minutes) of twenty patients receiving an analgesic. The data was initially reported by [22] and the complete sample analysed by [23] and [24]. The censored sample (number of failures) p, is chosen as 50% (censoring scheme). The MLE, KS and P-Value for the TIITL_{HL} model are reported in Table 9.

Table 9: *The MLE and performance measure for the first censored data.*

Models	MLE	KS	P-Value
TIITL_{HL}	$\tau = 0.877$	0.230	0.240

The second dataset represents the survival time of 72 Guinea pigs infected with virulent tubercle bacilli. The complete sample was analysed by [6] and [20]. The censored sample (number of failures) p, is chosen as 70% (censoring scheme). The MLE, KS and P-Value for the TIITL_{HL} model are reported in Table 10. It is evident that the TIITL_{HL} appropriately fits the two survival time censored datasets.

Table 10: *The MLE and performance measure for the second censored data.*

Models	MLE	KS	P-Value
TIITL_{HL}	$\tau = 1.174$	0.077	0.790

It is evident that the TIITL_{HL} appropriately fits the two censored samples.

8.3. Regression model application

The usefulness of log-TIITL_{HL} regression model is demonstrated by means of a real data analysis. The log-TIITL_{HL} regression model is compared with log-exponential (LE) and log-Burr-Hatke-exponential (LBHE) regression models [25]. The utilized dataset contains 100 individuals having HIV+ obtained from the **Bolstad2** package in R-software. The observed survival times (y_i), in months, with censoring indicator (0 = alive and 1 = death) is analysed with two explanatory variables: k_{i1}, (0 = no and 1 = yes) represent the history of drug usage and k_{i2} represent the ages of patients . The proposed regression model is

$$y_i = \beta_0 + \beta_1 k_{i1} + \beta_2 k_{i2} + \sigma z_i \tag{48}$$

where z_i has density Eq (43). The MLE method is utilized in estimating the unknown parameters of log-TIITL_{HL}, LE and LBHE regression models. Table 11 reports the regression models estimated parameters, -LL and performance measures (AIC, BIC, AICc and HIQC values). The results provided in Table 11 indicates that the $\text{LTIITL}_{\text{HL}}$ regression model has the lowest value of -LL and performance measures values, respectively. Hence, it is concluded that log-TIITL_{HL} regression model provides appropriate fit than LE and LBHE regression models. More so, the estimated regression parameters β_0, β_1 and β_2 are statistically significant at 5% level of significance.

Table 11: *The regression models estimated parameters and performance measures.*

	LBHE			L-E			Log-TIITL$_{HL}$		
Parameters	Estimates	SE	P-Value	Estimates	SE	P-Value	Estimates	SE	P-Value
τ	1.508	13.659	-	1.599	13.783	-	26.968	48.251	-
σ	0.778	0.067	-	0.839	0.072	-	1.684	0.143	-
β_0	6.883	7.064	0.330	6.542	7.256	0.367	2.303	0.130	<0.001
β_1	-0.091	0.014	<0.001	-0.091	0.014	<0.001	-0.023	0.009	0.020
β_2	-1.021	0.193	<0.001	-1.049	0.189	<0.001	-0.261	0.109	0.017
$-LL$		128.059			128.502			128.051	
AIC		266.12			267.00			266.10	
BIC		279.14			280.03			279.13	
AICc		266.76			267.64			266.74	
HQIC		271.39			272.28			271.37	

9. Conclusion

This work introduced a new one-parameter model titled the TIITL$_{HL}$ model and provided some of its properties. The consistency of the maximum likelihood estimator and five other estimators are proven by uncensored and censored simulation studies. Applications to real medical and environmental sciences datasets revealed its flexibility and adaptability. The log-TIITL$_{HL}$ regression model constructed and fitted to HIV+ data, and compared with other existing models showed that the model will be a useful choice in survival investigation for practitioners. Overall, the five applications showed the usefulness of the new model for asymmetric, uncensored and censored data. In future works, the Bayesian analysis of TIITL$_{HL}$ accelerated failure time model, the TIITL$_{HL}$–G family of distributions and the discrete case of the TIITL$_{HL}$ model will be addressed.

References

[1] Balakrishnan N. (1992). *Handbook of the Logistic Distribution. Statistics: A Series of Textbooks and Monographs, 123*, Marcel Dekker, New York, USA.

[2] Cordeiro, G. M., Alizadeh, M. and Pedro, R. D. M (2016). The type I half-logistic family of distributions. *Journal of Statistical Computation and Simulation*, 86: 707–728.

[3] Rao, R. S., Mamidi, P. L. and Kantam, R. R. (2016). Modified maximum likelihood estimation: Inverse half logistic distribution. *Journal on Mathematics*, 5: 11–19.

[4] Alaa, H. A. H. (2016). Properties, estimations and predictions for a Poisson-half-logistic distribution based on progressively type-II censored samples. *Applied Mathematical Modelling*, 40: 7164–7181.

[5] Soliman, A. H., Elgarhy, M. A. E., and Shakil, M. (2017). Type II Half Logistic Family of Distributions with Applications. *Pakistan Journal of Statistics and Operation Research*, 13: 245–264.

[6] Usman, R. M., Haq, A. M., and Talib, J. (2017). Kumaraswamy half-logistic distribution: properties and applications. *Journal of Statistics Applications & Probability*, 6: 597–609.

[7] Altun, E., Khan, M. N., Alizadeh, M., Ozel, G., and Butt, N. S (2018). Extended Half-Logistic Distribution with Theory and Lifetime Data Application. *Pakistan Journal of Statistics and Operation Research*, 14: 319–331.

[8] Samuel, A. F., and Kehinde, O. A. (2019). A Study on Transmuted Half Logistic Distribution: Properties and Application. *International Journal of Statistical Distributions and Applications*, 5: 54–59.

[9] Alizadeh, M., Nematollahi, A., Altun, E., and Rasekhi, M (2020). A Study on A New Type I Half-Logistic Family of Distributions and Its Applications. *Statistics, Optimization and Information Computing*, 8: 934–949.

[10] Eliwa, M. S., Altun, E., Alhussain, Z. A., Ahmed, E. A., Salah, M. M., Ahmed, H. H., and El-Morshedy, M. (2021). A new one-parameter lifetime distribution and its regression model with applications. *PLoS ONE* 16: e0246969.

[11] Hashempour, M., and Alizadeh, M. (2021). A New Weighted Half-Logistic distribution: Properties, applications and different method of estimations. *Statistics, Optimization & Information Computing*, 11: 554–569.
[12] Mohammad, G. S. (2022). A new two-parameter modified half-logistic distribution: Properties and Applications. *Stat., Optim. Inf. Comput.*, 10: 589—605.
[13] Hashem, A. F., Ku?, C., Pekg?r, A., and Alaa, H. A. H. (2022). Poisson–logarithmic half-logistic distribution with inference under a progressive-stress model based on adaptive type-II progressive hybrid censoring. *J Egypt Math Soc* 30: 1–33.
[14] Majid, H (2022). An extended type I half-logistic family of distributions: Properties, applications and different method of estimations. *Mathematica Slovaca*, 72: 745–764.
[15] Arshad, R. M. I., Tahir, M. H., Chesneau, C., Khan, S. and Jamal, F. (2022). The gamma power half-logistic distribution: theory and applications. *S?o Paulo J. Math. Sci.*, 2022.
[16] Elgarhy, M., Nasir, M. A., Jamal, F., and Ozel, G. (2018). The type II Topp-Leone generated family of distributions: Properties and applications. *Journal of Statistics and Management Systems*, 21: 1529–1551.
[17] Kenney, J. F. Mathematics of statistics, D. Van Nostrand, 1939.
[18] Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37: 25–32.
[19] Bjerkedal, T. (1960). Acquisition of resistance in guinea pies infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene*, 72: 130–148.
[20] Adubisi, O. D., Abdulkadir, A. and Chiroma, H. (2021). A Two Parameter Odd Exponentiated Skew-T Distribution With J-Shaped Hazard Rate Function. *Journal of Statistical Modeling & Analytics (JOSMA)*, 3(1): 26–46.
[21] Haq, M. A. (2016). Kumaraswamy exponentiated inverse Rayleigh distribution. *Mathematical Theory and Modeling*, 6: 93–104.
[22] Gross, A. J., and Clark, V. A. (1975). *Survival Distributions: Reliability Applications in the Biometrical Sciences*, John Wiley, New York.
[23] Shanker, R., Hagos, F. and Sujatha, S. (2015). On modeling of Lifetimes data using exponential & Lindley distributions. *Biometrics & Biostatistics International Journal*, 2: 1–9.
[24] Shafi, S., Wani, S. A., and Shafi, S. (2020). A new three parameter weighted distribution applicable to relief times, waiting times and carbon fiber. *J. Stat. Appl.*, 10: 167–184.
[25] Yadav, A. S., Altun, E., and Yousof, H. M. (2019). Burr-Hatke exponential distribution: A decreasing failure rate model. *Statistical Inference and Applications. Annals of Data Science*, 1-–20.

MOVING AVERAGE AND DOUBLE MOVING AVERAGE CONTROL CHARTS FOR PROCESS VARIABILITY USING AUXILIARY INFORMATION

Vikas Ghute[1] and Sarika Pawar[2]

•

[1,2]Department of Statistics
Punyashlok Ahilyadevi Holkar
Solapur University Solapur (MS)-413255, India
vbghute_stats@rediffmail.com, ssgulame@ gmail.com

[1]Corresponding Author

Abstract

The memory type control charts based on auxiliary information have been introduced in the literature for improved monitoring of the process parameters for normally distributed process. In this paper, we design moving average and double moving average control charts based on auxiliary information for efficient monitoring the shifts in the process variability. Regression estimator of process variance in the form of auxiliary and study variables is considered to construct charting statistics for the proposed charts. The average run length (ARL) and standard deviation of run length (SDRL) performance of the proposed charts is investigated using simulation study and is compared with the originally proposed Shewhart control charts based on auxiliary information and without auxiliary information. The proposed auxiliary information based moving average and double moving average charts are found to be efficient for monitoring the process variance of normally distributed process. An illustrative example based on simulated data set is provided to show the implementation of the proposed charts in detecting shifts in the process standard deviation.

Keywords: Control chart, average run length, auxiliary information, moving average, double moving average.

1. Introduction

Statistical process control (SPC) is a powerful statistical technique used to determine the performance of the process accurately. It has been widely used in manufacturing and service industries. A control chart is one of the most widely known tools in SPC which is extensively used to monitor process quality. It is designed to identify and detect timely assignable causes in the process. In general, control chart is used to detect changes in the process parameters. Two types of control charts are generally used to monitor production processes namely the location chart and the variability chart. The location chart is used to monitor process mean and the variability chart is used to monitor process variability. It is a standard practice to use Shewhart $\bar{X}$ chart for monitoring the process mean and R or S charts for monitoring the process variability. Some practioners recommend a control chart based directly on the sample variance S^2 control chart for monitoring process variability. A major disadvantage of Shewhart type control charts is that they use only

information of last sample observation and ignores the past information of the process which makes it insensitive to small shifts in process parameters.

Memory type control charts are most commonly used in process monitoring to detect small to moderate shifts in the process parameters. They are constructed using past information regarding the production process and are more sensitive to monitor the small and moderate shifts in the process parameters. These control charts includes cumulative sum (CUSUM), exponentially weighted moving average (EWMA) and moving average (MA). Relative to CUSUM chart, the EWMA and MA charts are quite basic. The EWMA chart uses a weighted average as the chart statistic while the time weighted MA chart is based on simple moving average. The moving average statistics of width w is simply the average of the w most recent observations and are more sensitive to monitor the small and moderate shifts in the process parameters. Wong *et al*. [1] developed simple procedures for the design of an individual MA chart and a combined MA-Shewhart scheme. Khoo and Yap [2] proposed the use of single MA chart for joint monitoring of the process mean and variance by combining $\bar{X}$ and S charts into a single chart. Adeoti and Olaomi [3] proposed a moving average control chart based on sample standard deviation for detecting small shifts in process variability. Ghute and Rajmanya [4] developed moving average control chart based on Downton's D statistic and Gini's mean difference G statistic for detecting small shifts in the process standard deviation. The proposed moving average control charts are found to be more efficient for monitoring process variability. In multivariate setup, Ghute and Shirke [5] developed a multivariate moving average T^2 control chart for monitoring mean vector of multivariate process. It was shown that the proposed chart performs better than the Hotelling's T^2 chart in the detection of small to moderate shifts in the process mean vector. Ghute and Shirke [6] also developed a multivariate moving average $|S|$control chart for monitoring covariance matrix of multivariate normally distributed process. It was shown that proposed chart performs better than Shewhart-type $|S|$ chart in the detection of small to moderate changes in the process covariance matrix.

The double moving average (DMA) control charts have been proposed in the literature to further improve the performance of moving average (MA) control charts. Khoo and Wong [7] introduced the DMA chart by computing moving averages twice for early detection of small to moderate shifts in the process mean. It was shown that the DMA control chart performs better than the MA chart for the detection of small to moderate shifts in the mean. Adeoti et al. [8] proposed a DMA control chart based on sample standard deviation for detecting shifts in the process variability. Sukparungsee et al. [9] developed a mixed Tukey-Double moving average control chart for monitoring process mean of symmetric and asymmetric processes. They compared ARL performance of the proposed chart with the existing charts. It was shown that the proposed chart is an effective competitor to the existing counterparts. Taboran and Sukparungsee [10] designed a new EWMA-DMA control chart for detecting change of mean of the process with normal, Laplace, exponential and gamma distributions. They compared ARL performance of the proposed chart with other existing charts. It was shown that the proposed chart has best detection ability for certain level of shifts in process mean.

In order to increase the sensitivity of the traditional control charts many new modifications and improvements in the control charting procedure have been suggested in the SPC literature. One of such modifications is the development of auxiliary-information-based (AIB) control charts which have an excellent speed in detecting shift in the process parameters than those based without it. Such control charts are based on a statistic that utilizes information from both the study and auxiliary variables. The information on auxiliary variable is generally known prior to the sampling procedure and it assists in estimating the study variable with increased accuracy. There are many AIB control charts available in literature with Shewhart type and memory type charting structures for monitoring process mean and process variability. Riaz [11] first suggested AIB-

Shewhart that is based on regression-type mean estimator for monitoring shifts in process mean. It was shown that the AIB-Shewhart chart using the regression estimator performs better than the Shewhart chart for detecting shifts in the process mean. Riaz and Does [12] developed a Shewhart-type variability chart using ratio-type variance estimator for the Phase I quality control. It was shown that AIB Shewhart variability chart is more powerful than the existing Shewhart variance chart. Riaz [13] proposed a Shewhart-type control chart for an improved monitoring of mean level of quality characteristic of interest Y using the information on a single auxiliary characteristic X on product difference pattern. Riaz et al. [14] suggested new AIB-Shewhart chart based on regression estimator for monitoring the process variability. They have shown that the Shewhart chart using regression estimator outperforms the other Shewhart charts when detecting increase in the process variability. Abbas et al. [15] introduced the EWMA chart with the auxiliary information, using regression estimator for monitoring location of the process. It was shown that proposed chart performs better than its existing control charts. Abbasi and Riaz [16] made dual use of auxiliary information to propose new chart for process location. Riaz et al. [17] made dual use of auxiliary information to propose a chart for process variability. Sanusi et al. [18] studied CUSUM charts using different estimators based on auxiliary information. Haq [19] proposed new EWMA charts using auxiliary information for efficiently monitoring the process dispersion. Recently, Amir et al. [20] developed auxiliary information based moving average control chart (denoted as AB-MA chart) for effective monitoring of shifts in the process location parameter. They compared the performance of proposed chart with existing control chart and shown that chart outperforms in detecting small and medium shifts in the process location parameter. Amir et al. [21] designed auxiliary information based double moving average control chart (denoted as ADMA chart) for effective monitoring of the process location parameter. They compared the performance of the proposed ADMA chart with its memory-type counterparts and shown that the proposed ADMA chart performs uniformly better than the EWMA and CUSUM charts when correlation between auxiliary variable and study variable is high.

Most of the studies on AIB memory-type charts have been concentrated on monitoring process mean. Often, monitoring shifts in the variance of related study variable is also important. The purpose of this paper is to develop auxiliary information based MA and DMA control charts for efficient monitoring of process variability of normally distributed process in Phase II. By getting motivation of improved performance of auxiliary information based MA and DMA charts for monitoring process mean recently proposed by Amir et al. [20] and Amir et al. [21] respectively, in the present paper, we develop new MA and DMA charts using auxiliary information for efficiently monitoring the process variability in phase II. We expect that the proposed MA and DMA control charts will be more sensitive for efficiently monitoring process variability. The regression estimator of the process variance in the form of auxiliary and study variables is considered to construct charting statistics for the proposed MA and DMA charts. The performance of the proposed charts is evaluated in terms of the average run length (ARL) and standard deviation of run length (SDRL) criteria. Monte Carlo simulations are used to study the run length profiles of the proposed AIB-MA-V and AIB-DMA-V charts. The performance of the proposed charts is compared with its Shewhart-type counterparts.

Rest of the paper is organized as follows: In Section 2, traditional S^2 chart for monitoring process variability is discussed. The auxiliary information based Shewhart-type control chart for monitoring process variability is presented in Section 3. The design procedure of proposed auxiliary information based moving average and double moving average control charts are presented in Section 4 and Section 5 respectively. In Section 6, a simulation study is conducted to evaluate the performance of proposed control charts in comparison to that of Shewhart-type charts. An illustrative example is presented in Section 7 to demonstrate the implementation of the proposed charts. Finally, Section 8 concludes the findings of the paper.

2. Shewhart control chart for process variability.

In this Section, we discuss the traditional S^2control chart for monitoring process variability that has been constructed without using the auxiliary information. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots Y_{in})\ i - 1,2, \ldots$ be independent random samples of size n $(n \geq 2)$ from a normally distributed process with mean μ_y and variance σ_y^2. Here our objective is to monitor the changes in the process variance. We assume that $\mu_y = \mu_{y,0}$ and $\sigma_y = \sigma_{y,0}$ are known, even if, in practice, these parameters have to be estimated from an in-control population. It is assumed that the process is in-control with variance $\sigma_{y,0}^2$. When shift in process variance $\sigma_{y,0}^2$ occurs, we have change from in-control value $\sigma_{y,0}^2$ to the out-of-control value$\sigma_{y,1}^2$. Let $\lambda = \sigma_{y,1}/\sigma_{y,0} (0 < \lambda \leq 1)$ denotes the amount of shift in the in-control process standard deviation$\sigma_{y,1}$. When$\lambda = 1$, the process is considered to be in-control, otherwise the process is considered to be out-of-control. Let $\bar{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}$ the sample mean at stage i. The S^2 chart can be constructed by plotting sample variances

$$S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(Y_{ij} - \bar{Y}\right)^2, i = 1,2, \ldots. \quad (1)$$

With lower and upper control limits set as

$$LCL = \frac{\sigma^2}{n-1}\chi^2_{n-1,1-\alpha/2}, UCL = \frac{\sigma^2}{n-1}\chi^2_{n-1,\alpha/2} \quad (2)$$

Where $\chi^2_{\alpha/2,}\ \chi^2_{1-\alpha/2,}$ denotes the upper and lower $\alpha/2$percentage points of the Chi-square distribution with $n-1$degrees of freedom. The S^2chart for monitoring process variability gives a signal if S_i^2exceeds the control limits.

3. AIB Shewhart-type V Chart for Process variability

In this Section, we discuss AIB Shewhart-type control chart using regression estimator of the variance for monitoring shifts in process variability. Assume that a process has a quality characteristic of interest Y which is correlated with an auxiliary variable X. The pairs $(Y_i,\ X_i)$are assumed to follow bivariate normal distribution with $N_2(\mu_y, \mu_x, \sigma_y^2, \sigma_x^2, \rho)$. Here ρ represents the correlation coefficient between study variable Y and auxiliary variable X. The observations of Y and X are obtained in the paired form for each sample and the population parameters are assumed to be known.

Let $(Y_1, X_1), (Y_2, X_2), \ldots . (Y_n, X_n)$ represents a sample of size n from the bivariate normal distribution. The auxiliary information based unbiased regression estimator of population variance σ_y^2 of study variable Y using a single auxiliary variable X for a bivariate random sample of size n is given by (Haq,(2017))

$$V = S_y^2 + \rho^2 \frac{\sigma_y^2}{\sigma_x^2}(\sigma_x^2 - S_x^2) \quad (3)$$

Where $S_y^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_i - \bar{Y})^2,\ S_x^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_i - \bar{X})^2$ represent sample variances of Y and X respectively and $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_i, \bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_i$ represent sample means of Y and X respectively.
The mean and variance of V are as follows:

$$E(V) = \sigma_y^2 \text{ and } Var(V) = \sigma_V^2 = \frac{2\sigma_y^4}{n-1}(1 - \rho^4)$$

Riaz et al. (2014) suggested AIB-Shewhart chart based on regression estimator V of process variance σ_y^2 for monitoring process variability. We denote the chart as AIB-Shewhart-V chart. In the construction of AIB-Shewhart-V control chart for monitoring the process variability, the

statistic V is plotted on the chart against the sample number. The control limits of the AIB-Shewhart-V chart are

$$LCL = \sigma_y^2 - L\sigma_y^2\sqrt{\frac{2(1-\rho^4)}{n-1}} \text{ and } UCL = \sigma_y^2 + L\sigma_y^2\sqrt{\frac{2(1-\rho^4)}{n-1}} \quad (4)$$

Where the value of L determines the in-control ARL of the AIB-Shewhart-V chart.

4. AIB Moving Average Control Chart for Process Variability

In this Section, we develop moving average control chart for detecting changes produced in the process variability. The proposed moving average chart is based on auxiliary information based V statistic and chart is denoted as AIB-MA-V chart. The construction of the chart is based on computing the moving averages of V statistics given in Eq. (3). The moving average statistic of span w at time i for a sequence of V statistics are computed as

$$MA_i = \frac{V_i+V_{i-1}+\cdots+V_{i-w+1}}{w}, for\ i \geq w \quad (5)$$

For periods $i < w$ we compute the average of available charting statistic. In other words, average of all V observations up to period i defines moving average. For$i > w$, mean and variance of MA_istatistic for in-control process are given as

$$E(MA_i) = \sigma_y^2 \text{ and } Var(MA_i) = \frac{2\sigma_y^4}{w\,(n-1)}(1-\rho^4)$$

The control limits of MA-V chart are as follows:

$$UCL/LCL = \begin{cases} \sigma_y^2 \pm L\sigma_y^2\sqrt{\frac{2(1-\rho^4)}{w(n-1)}}, & for\ i \geq w \\ \sigma_y^2 \pm L\sigma_y^2\sqrt{\frac{2(1-\rho^4)}{i(n-1)}}, & for\ i < w \end{cases} \quad (6)$$

Where L is a constant chosen to specify in-control ARL for the AIB-MA-V chart. The AIB-MA-V chart is constructed by plotting the MA_i statistics on the chart against the sample number i. An out-of-control signal is issued when $MA_i$is smaller than LCL or larger than the UCL.

5. AIB Double Moving Average Control Chart for Process Variability

In this Section, we present the design of double moving average control chart for detecting changes produced in the process variability. The proposed chart is denoted as AIB-DMA-V chart. The double moving average (DMA) statistic is based on the twice the subgroup average of the MA statistic. The moving average statistic for sequence of subgroup variances with time i and width w is given in Eq. (5). For interval $i < w$ the DMA statistic can be calculated as the mean of all subgroup variances up to interval i while for interval $i \geq w$, the plotting statistic of AIB-DMA-V chart is given by

$$DMA_i = \frac{MA_i+MA_{i-1}+\cdots+MA_{i-w+1}}{w}, for\ i \geq w \quad (7)$$

where w represents the span at time i for computing DMA_istatistic. The in-control mean of DMA_istatistic calculated for $i \geq w$ is given by,

$$E(DMA_i) = \frac{1}{w}E\left(\sum_{j=i-w+1}^{i} MA_j\right) = \frac{1}{w}\left(w\sigma_y^2\right) = \sigma_y^2$$

The in-control variance of DMA_i is given by

$$Var(DMA_i) = \begin{cases} \frac{\sigma_y^4 2(1-\rho^4)}{i^2(n-1)} \sum_{j=1}^{i} \frac{1}{j} & , i \le w \\ \frac{\sigma_y^4 2(1-\rho^4)}{w^2(n-1)} \sum_{j=i-w+1}^{w-1} \frac{1}{j} + (i-w+1)\frac{1}{w} & , w < i < 2w-1 \\ \frac{\sigma_y^4 2(1-\rho^4)}{w^2(n-1)}, & i \ge 2w-1 \end{cases} \quad (8)$$

For the $w = 2$, the variance of the DMA is calculated using the 1st and 3rd lines of the above Eq. (8). The control limits of the proposed AIB-DMA-V chart are given as

$$UCL/LCL = \begin{cases} \sigma_y^2 \pm \frac{L\,\sigma_y^2}{i} \sqrt{\frac{2(1-\rho^4)}{n-1} \sum_{j=1}^{i} \frac{1}{j}} & , i \le w \\ \sigma_y^2 \pm \frac{L\,\sigma_y^2}{w} \sqrt{\frac{2(1-\rho^4)}{n-1} \sum_{j=i-w+1}^{w-1} \frac{1}{j} + (i-w+1)\frac{1}{w}} & , w < i < 2w-1 \\ \sigma_y^2 \pm \frac{L\,\sigma_y^2}{w} \sqrt{\frac{2(1-\rho^4)}{n-1}} & , i \ge 2w-1 \end{cases} \quad (9)$$

The control limits for $w = 2$ are calculated based on the using the 1st and 3rd lines of the above Eq. (9). L is the constant that is set according to the desired in-control ARL for the AIB-DMA-V control chart.

6. Performance Evaluation and Comparison

Performance of a control chart is typically measured in terms of average run length (ARL) and standard deviation of run length ($SDRL$). The ARL is the average number of sample points that is plotted on a chart before the first out-of-control signal is detected, whereas, the $SDRL$ measures the spread of the run length distribution. When the process is out-of-control, it is desirable to have small values of ARL and $SDRL$. The performance of the control chart is measured in terms of in-control ARL (denoted asARL_0) and out-of-control ARL (denoted asARL_1). To compare the efficiency of two control charts for detecting the shift in process parameter, the general practice is to adjust their control limits so that their ARL_0values become same and then compare ARL_1values at various shifts in process parameter. A chart with smaller ARL_1 is considered to be more efficient to detect pre-assigned shift in the process parameter.

In this Section, we evaluate the performance of the proposed AIB-MA-V and AIB-DMA-V charts for detecting shifts in Phase II monitoring. Monte Carlo simulation approach is used to evaluate the run length performance of the proposed charts in comparison with that of the AIB-Shewhart-V and S^2charts. Simulation study based on sample of size $n = 10,15,20$ with $\rho =$ 0.3,0.6,0.9 is carried out to assess the performance of the proposed AIB-MA-V, AIB-DMA-V and AIB-Shewhart-V charts. It is assumed that the in-control process is a bivariate normal distribution. Without loss of generality a standard bivariate normal distribution (i.e.$N_2(0,0,1,1,\rho)$ is considered as in-control process distribution. The out-of-control process is a bivariate normal with the same means of both auxiliary and study variables with changed variance of study variable$Y$. That is, we consider out-of-control procedure as$N_2(0,0,\lambda,1,\rho)$. Using the simulation, the control limit constant L of the AIB-MA-V and AIB-DMA-V charts are obtained for $w = 2,3$ and 4 so thatARL_0 of the chart is approximately 200. The ARL and $SDRL$ values of the chart are found with 50000 simulations for each shift of magnitude λ in the process variance of the study variable Y. The magnitude of shift in the standard deviation of the study variable is considered as$\lambda = 1.0(0.1)2.0,2.5,3.0$. To compare the performance of the proposed AIB-MA-V and AIB-DMA-V charts with AIB-Shewhart-V chart and Shewhart S^2chart, each chart is designed so that ARL_0 is approximately 200. The control limits, ARL and $SDRL$ values of the AIB-Shewhart-V chart are obtained using simulation. The exact ARL values of the traditional S^2chart are computed using Excel.

Tables 1 to 3 represent the ARL and $SDRL$ (shown in parenthesis) values of the proposed control chart with $n = 10,15,20$ according to the variance shift when the correlation coefficient is

0.3, 0.6 and 0.9 respectively. Here control limit constant L is chosen so that the value of $ARL_0$is close to 200.

Table 1: *Run length profiles of the charts with* $n = 10$ *and* $ARL_0 = 200$

Shift λ	S^2chart	AIB-V chart	AIB-MA-V chart			AIB-DMA-V chart		
			$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
				$\rho = 0.3$				
1.0	200.65	199.95	199.03	200.54	200.93	199.90	200.73	200.38
	(197.87)	(197.86)	(200.71)	(200.13)	(200.79)	(199.54)	(200.53)	(198.33)
1.1	72.98	46.54	36.11	31.68	29.23	34.65	30.36	29.01
	(72.48)	(45.94)	(35.64)	(30.92)	(28.56)	(33.80)	(29.01)	(26.72)
1.2	25.23	16.68	12.11	10.38	9.46	11.69	10.40	10.25
	(24.18)	(16.25)	(11.42)	(9.56)	(8.65)	(10.68)	(8.76)	(8.01)
1.3	11.00	8.03	5.89	5.18	4.83	5.89	5.55	5.80
	(10.40)	(7.48)	(5.22)	(4.46)	(4.04)	(4.89)	(4.04)	(3.76)
1.4	6.09	4.66	3.62	3.29	3.12	3.75	3.76	4.03
	(5.52)	(4.11)	(2.99)	(2.57)	(2.42)	(2.76)	(2.43)	(2.37)
1.5	3.92	3.17	2.58	2.40	2.28	2.76	2.87	3.12
	(3.34)	(2.61)	(1.93)	(1.72)	(1.61)	(1.84)	(1.72)	(1.82)
1.6	2.83	2.38	2.02	1.91	1.85	2.22	2.34	2.52
	(2.26)	(1.82)	(1.37)	(1.23)	(1.19)	(1.36)	(1.35)	(1.48)
1.7	2.20	1.92	1.70	1.62	1.58	1.87	1.99	2.12
	(1.66)	(1.33)	(1.05)	(0.95)	(0.91)	(1.07)	(1.12)	(1.23)
1.8	1.81	1.63	1.49	1.44	1.41	1.64	1.73	1.83
	(1.22)	(1.02)	(0.82)	(0.75)	(0.72)	(0.88)	(0.95)	(1.04)
1.9	1.57	1.46	1.36	1.32	1.29	1.48	1.55	1.62
	(0.96)	(0.82)	(0.67)	(0.62)	(0.59)	(0.74)	(0.81)	(0.88)
2.0	1.42	1.33	1.26	1.23	1.21	1.35	1.40	1.46
	(0.77)	(0.67)	(0.55)	(0.51)	(0.49)	(0.63)	(0.69)	(0.74)
2.5	1.10	1.06	1.06	1.05	1.05	1.09	1.11	1.13
	(0.33)	(0.24)	(0.26)	(0.24)	(0.23)	(0.31)	(0.34)	(0.37)
3.0	1.03	1.33	1.02	1.02	1.01	1.03	1.03	1.04
	(0.17)	(0.66)	(0.13)	(0.13)	(0.12)	(0.17)	(0.18)	(0.20)
L	---	3.431	3.090	2.909	2.789	3.742	4.045	4.342
				$\rho = 0.6$				
1.0	200.65	200.97	200.80	199.06	200.21	200.76	199.56	200.79
	(197.87)	(201.34)	(201.53)	(200.18)	(201.40)	(200.75)	(198.82)	(198.20)
1.1	72.98	42.69	33.12	28.76	26.55	31.50	27.92	26.79
	(72.48)	(42.50)	(32.96)	(28.04)	(25.97)	(30.78)	(26.32)	(24.47)
1.2	25.23	14.79	10.66	9.29	8.55	10.37	9.35	9.21
	(24.18)	(14.34)	(9.95)	(8.54)	(7.63)	(9.36)	(7.75)	(6.93)
1.3	11.00	6.94	5.22	4.59	4.36	5.26	5.08	5.33
	(10.40)	(6.43)	(4.52)	(3.84)	(3.62)	(4.23)	(3.60)	(3.34)
1.4	6.09	4.14	3.26	2.98	2.83	3.40	3.45	3.77
	(5.52)	(3.59)	(2.61)	(2.30)	(2.12)	(2.42)	(2.15)	(2.18)

Table 1 continued

Shift λ	S^2chart	AIB-V chart	AIB-MA-V chart			AIB-DMA-V chart		
			$w=2$	$w=3$	$w=4$	$w=2$	$w=3$	$w=4$
				$\rho=0.6$				
1.5	3.92	2.83	2.34	2.19	2.11	2.52	2.68	2.92
	(3.34)	(2.29)	(1.68)	(1.52)	(1.42)	(1.62)	(1.55)	(1.69)
1.6	2.83	2.14	1.87	1.78	1.72	2.05	2.19	2.36
	(2.26)	(1.56)	(1.22)	(1.11)	(1.06)	(1.22)	(1.25)	(1.38)
1.7	2.20	1.76	1.58	1.52	1.49	1.74	1.86	1.97
	(1.66)	(1.16)	(0.92)	(0.84)	(0.81)	(0.96)	(1.03)	(1.13)
1.8	1.81	1.52	1.40	1.37	1.34	1.54	1.63	1.72
	(1.22)	(0.89)	(0.72)	(0.67)	(0.64)	(0.80)	(0.88)	(0.96)
1.9	1.57	1.37	1.29	1.26	1.24	1.39	1.47	1.53
	(0.96)	(0.72)	(0.59)	(0.56)	(0.53)	(0.66)	(0.74)	(0.80)
2.0	1.42	1.27	1.21	1.19	1.18	1.29	1.35	1.40
	(0.77)	(0.58)	(0.49)	(0.46)	(0.45)	(0.57)	(0.63)	(0.68)
2.5	1.10	1.05	1.05	1.04	1.04	1.07	1.09	1.10
	(0.33)	(0.22)	(0.23)	(0.21)	(0.20)	(0.28)	(0.30)	(0.33)
3.0	1.03	1.26	1.01	1.01	1.01	1.02	1.02	1.03
	(0.17)	(0.57)	(0.12)	(0.11)	(0.10)	(0.14)	(0.16)	(0.17)
L	---	3.360	3.045	2.877	2.774	3.680	4.015	4.332
				$\rho=0.9$				
1.0	200.65	200.46	200.25	200.41	199.09	199.40	199.88	199.83
	(197.87)	(199.21)	(198.55)	(199.40)	(200.01)	(198.33)	(199.47)	(198.62)
1.1	72.98	24.31	17.98	15.48	14.28	17.01	14.96	14.38
	(72.48)	(23.72)	(17.24)	(14.62)	(13.42)	(16.04)	(13.29)	(11.85)
1.2	25.23	6.72	5.02	4.51	4.20	5.00	4.86	5.16
	(24.18)	(6.19)	(4.30)	(3.68)	(3.33)	(3.91)	(3.26)	(3.08)
1.3	11.00	3.17	2.60	2.42	2.33	2.76	2.93	3.21
	(10.40)	(2.62)	(1.92)	(1.69)	(1.59)	(1.78)	(1.66)	(1.75)
1.4	6.09	2.04	1.78	1.72	1.68	1.97	2.14	2.32
	(5.52)	(1.47)	(1.11)	(1.02)	(0.97)	(1.11)	(1.15)	(1.28)
1.5	3.92	1.56	1.44	1.41	1.39	1.59	1.71	1.83
	(3.34)	(0.94)	(0.75)	(0.71)	(0.68)	(0.82)	(0.89)	(0.98)
1.6	2.83	1.33	1.26	1.24	1.23	1.37	1.46	1.53
	(2.26)	(0.66)	(0.55)	(0.52)	(0.51)	(0.63)	(0.70)	(0.76)
1.7	2.20	1.20	1.16	1.15	1.14	1.24	1.29	1.35
	(1.66)	(0.48)	(0.42)	(0.40)	(0.39)	(0.50)	(0.56)	(0.61)
1.8	1.81	1.13	1.11	1.09	1.09	1.16	1.20	1.23
	(1.22)	(0.38)	(0.33)	(0.31)	(0.31)	(0.40)	(0.46)	(0.49)
1.9	1.57	1.08	1.07	1.06	1.06	1.10	1.13	1.16
	(0.96)	(0.30)	(0.26)	(0.25)	(0.25)	(0.33)	(0.37)	(0.41)
2.0	1.42	1.05	1.04	1.04	1.04	1.07	1.09	1.11
	(0.77)	(0.24)	(0.21)	(0.21)	(0.20)	(0.27)	(0.30)	(0.33)
2.5	1.10	1.01	1.01	1.01	1.01	1.01	1.02	1.02
	(0.33)	(0.10)	(0.08)	(0.08)	(0.07)	(0.11)	(0.12)	(0.14)
3.0	1.03	1.05	1.00	1.00	1.00	1.00	1.00	1.00
	(0.17)	(0.24)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)
L	---	3.176	2.952	2.842	2.763	3.579	3.984	4.333

Table 2: *Run length profiles of the charts with* $n = 15$ *and* $ARL_0 = 200$

Shift λ	S^2chart	AIB-V chart	AIB-MA-V chart			AIB-DMA-V chart		
			$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
				$\rho = 0.3$				
1.0	200.49	199.97	199.21	199.29	199.17	200.77	199.40	200.93
	(201.15)	(200.70)	(200.38)	(198.87)	(199.22)	(199.73)	(198.08)	(199.18)
1.1	59.44	37.14	27.73	23.65	21.79	25.84	22.80	21.72
	(58.56)	(36.65)	(27.04)	(22.94)	(20.94)	(24.94)	(21.19)	(19.33)
1.2	17.34	11.87	8.41	7.20	6.64	8.03	7.33	7.48
	(16.74)	(11.39)	(7.75)	(6.38)	(5.71)	(6.92)	(5.71)	(5.20)
1.3	7.29	5.39	4.02	3.60	3.41	4.10	4.06	4.38
	(6.76)	(4.82)	(3.32)	(2.84)	(2.63)	(3.08)	(2.56)	(2.51)
1.4	3.95	3.16	2.52	2.34	2.24	2.69	2.83	3.10
	(3.38)	(2.63)	(1.85)	(1.64)	(1.52)	(1.71)	(1.60)	(1.72)
1.5	2.59	2.19	1.87	1.77	1.73	2.04	2.20	2.40
	(2.02)	(1.62)	(1.19)	(1.07)	(1.02)	(1.17)	(1.21)	(1.33)
1.6	1.93	1.69	1.52	1.46	1.44	1.68	1.80	1.94
	(1.34)	(1.08)	(0.84)	(0.77)	(0.74)	(0.88)	(0.96)	(1.05)
1.7	1.56	1.43	1.33	1.29	1.28	1.45	1.54	1.63
	(0.93)	(0.78)	(0.63)	(0.58)	(0.56)	(0.70)	(0.78)	(0.84)
1.8	1.35	1.27	1.21	1.19	1.18	1.30	1.37	1.43
	(0.70)	(0.58)	(0.48)	(0.45)	(0.44)	(0.57)	(0.63)	(0.68)
1.9	1.23	1.17	1.14	1.12	1.11	1.20	1.25	1.29
	(0.52)	(0.45)	(0.38)	(0.36)	(0.34)	(0.47)	(0.51)	(0.55)
2.0	1.15	1.11	1.09	1.08	1.07	1.14	1.17	1.20
	(0.41)	(0.35)	(0.30)	(0.28)	(0.27)	(0.38)	(0.42)	(0.46)
2.5	1.02	1.01	1.01	1.01	1.01	1.02	1.03	1.03
	(0.14)	(0.12)	(0.10)	(0.10)	(0.09)	(0.13)	(0.16)	(0.18)
3.0	1.00	1.11	1.00	1.00	1.00	1.00	1.00	1.00
	(0.05)	(0.35)	(0.04)	(0.04)	(0.03)	(0.06)	(0.06)	(0.07)
L	---	3.268	2.991	2.837	2.750	3.608	3.964	4.299
				$\rho = 0.6$				
1.0	200.49	200.57	200.93	200.57	200.71	199.40	199.46	199.77
	(201.15)	(198.84)	(199.00)	(198.47)	(199.68)	(198.47)	(198.53)	(198.65)
1.1	59.44	33.38	24.94	21.55	19.84	23.27	20.92	19.95
	(58.56)	(33.15)	(24.29)	(20.67)	(18.89)	(22.06)	(19.16)	(17.47)
1.2	17.34	10.23	7.41	6.44	5.97	7.13	6.66	6.84
	(16.74)	(9.67)	(6.74)	(5.60)	(5.08)	(6.03)	(5.03)	(4.57)
1.3	7.29	4.69	3.61	3.25	3.10	3.68	3.74	4.05
	(6.76)	(4.16)	(2.95)	(2.49)	(2.32)	(2.64)	(2.28)	(2.25)
1.4	3.95	2.77	2.28	2.13	2.08	2.46	2.65	2.90
	(3.38)	(2.22)	(1.61)	(1.43)	(1.36)	(1.52)	(1.46)	(1.59)
1.5	2.59	1.96	1.71	1.64	1.61	1.89	2.06	2.23
	(2.02)	(1.37)	(1.03)	(0.95)	(0.92)	(1.05)	(1.10)	(1.23)
1.6	1.93	1.55	1.42	1.39	1.36	1.57	1.70	1.81
	(1.34)	(0.92)	(0.73)	(0.68)	(0.65)	(0.80)	(0.88)	(0.97)
1.7	1.56	1.34	1.26	1.24	1.23	1.37	1.46	1.53
	(0.93)	(0.67)	(0.54)	(0.52)	(0.50)	(0.63)	(0.70)	(0.76)

Table 2 continued

Shift λ	S^2chart	AIB-V chart	AIB-MA-V chart			AIB-DMA-V chart		
			$w=2$	$w=3$	$w=4$	$w=2$	$w=3$	$w=4$
				$\rho=0.6$				
1.8	1.35	1.21	1.16	1.15	1.14	1.25	1.30	1.36
	(0.70)	(0.50)	(0.42)	(0.40)	(0.39)	(0.51)	(0.57)	(0.62)
1.9	1.23	1.13	1.11	1.09	1.09	1.16	1.20	1.24
	(0.52)	(0.38)	(0.33)	(0.31)	(0.31)	(0.41)	(0.46)	(0.50)
2.0	1.15	1.08	1.06	1.06	1.06	1.10	1.13	1.16
	(0.41)	(0.30)	(0.25)	(0.25)	(0.24)	(0.33)	(0.37)	(0.41)
2.5	1.02	1.01	1.01	1.01	1.01	1.01	1.02	1.02
	(0.14)	(0.10)	(0.09)	(0.08)	(0.08)	(0.11)	(0.13)	(0.15)
3.0	1.00	1.08	1.00	1.00	1.00	1.00	1.00	1.00
	(0.05)	(0.30)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.06)
L	---	3.206	2.958	2.826	2.746	3.563	3.950	4.289
				$\rho=0.9$				
1.0	200.49	199.36	199.00	200.65	199.10	199.60	200.81	200.19
	(201.15)	(199.21)	(198.65)	(200.28)	(198.94)	(200.21)	(199.91)	(198.89)
1.1	59.44	17.49	12.61	10.81	9.83	11.87	10.41	10.13
	(58.56)	(16.93)	(11.79)	(9.81)	(8.69)	(10.73)	(8.60)	(7.65)
1.2	17.34	4.44	3.42	3.10	2.94	3.49	3.58	3.92
	(16.74)	(3.93)	(2.70)	(2.27)	(2.11)	(2.40)	(2.04)	(2.05)
1.3	7.29	2.15	1.85	1.78	1.74	2.05	2.25	2.47
	(6.76)	(1.58)	(1.15)	(1.05)	(1.00)	(1.13)	(1.15)	(1.27)
1.4	3.95	1.48	1.37	1.34	1.33	1.53	1.66	1.78
	(3.38)	(0.84)	(0.67)	(0.62)	(0.61)	(0.74)	(0.82)	(0.900
1.5	2.59	1.23	1.18	1.17	1.16	1.28	1.35	1.43
	(2.02)	(0.53)	(0.43)	(0.41)	(0.40)	(0.53)	(0.59)	(0.65)
1.6	1.93	1.11	1.09	1.08	1.08	1.15	1.19	1.24
	(1.34)	(0.34)	(0.30)	(0.29)	(0.28)	(0.38)	(0.44)	(0.49)
1.7	1.56	1.05	1.05	1.04	1.04	1.08	1.10	1.13
	(0.93)	(0.23)	(0.22)	(0.20)	(0.19)	(0.28)	(0.32)	(0.36)
1.8	1.35	1.03	1.02	1.02	1.02	1.04	1.06	1.07
	(0.70)	(0.16)	(0.15)	(0.14)	(0.14)	(0.20)	(0.24)	(0.27)
1.9	1.23	1.01	1.01	1.01	1.01	1.02	1.03	1.04
	(0.52)	(0.12)	(0.11)	(0.10)	(0.10)	(0.15)	(0.18)	(0.20)
2.0	1.15	1.01	1.01	1.01	1.00	1.01	1.02	1.02
	(0.41)	(0.09)	(0.08)	(0.07)	(0.07)	(0.11)	(0.13)	(0.15)
2.5	1.02	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.14)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)
3.0	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00
	(0.05)	(0.09)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
L	---	3.061	2.895	2.809	2.741	3.505	3.934	4.299

Table 3: *Run length profiles of the charts with* $n = 20$ *and* $ARL_0 = 200$

Shift λ	S^2chart	AIB-V chart	AIB-MA-V chart			AIB-DMA-V chart		
			$w = 2$	$w = 3$	$w = 4$	$w = 2$	$w = 3$	$w = 4$
				$\rho = 0.3$				
1.0	199.83	200.49	199.47	199.24	200.78	200.14	200.10	200.53
	(200.31)	(200.63)	(200.17)	(200.61)	(200.42)	(199.60)	(198.72)	(198.39)
1.1	49.93	30.87	22.43	19.05	17.72	20.97	18.30	17.63
	(49.41)	(30.30)	(21.83)	(18.29)	(16.74)	(19.91)	(16.66)	(15.13)
1.2	13.04	9.05	6.38	5.56	5.17	6.21	5.80	6.07
	(12.55)	(8.54)	(5.67)	(4.68)	(4.24)	(5.08)	(4.19)	(3.79)
1.3	5.26	4.03	3.09	2.80	2.70	3.21	3.32	3.64
	(4.72)	(3.50)	(2.38)	(2.06)	(1.92)	(2.20)	(1.90)	(1.96)
1.4	2.91	2.40	2.00	1.89	1.85	2.18	2.36	2.60
	(2.41)	(1.83)	(1.33)	(1.18)	(1.12)	(1.25)	(1.25)	(1.39)
1.5	1.95	1.72	1.53	1.48	1.46	1.70	1.84	1.99
	(1.38)	(1.12)	(0.84)	(0.76)	(0.75)	(0.88)	(0.95)	(1.04)
1.6	1.53	1.39	1.30	1.27	1.25	1.42	1.52	1.62
	(0.89)	(0.73)	(0.59)	(0.55)	(0.53)	(0.66)	(0.74)	(0.80)
1.7	1.29	1.21	1.17	1.15	1.14	1.25	1.32	1.39
	(0.62)	(0.51)	(0.42)	(0.40)	(0.39)	(0.51)	(0.57)	(0.63)
1.8	1.17	1.12	1.09	1.08	1.08	1.15	1.20	1.24
	(0.44)	(0.36)	(0.31)	(0.29)	(0.29)	(0.39)	(0.45)	(0.49)
1.9	1.10	1.07	1.05	1.05	1.05	1.09	1.12	1.15
	(0.33)	(0.27)	(0.23)	(0.22)	(0.21)	(0.30)	(0.35)	(0.39)
2.0	1.05	1.04	1.03	1.03	1.03	1.05	1.07	1.09
	(0.24)	(0.20)	(0.18)	(0.17)	(0.16)	(0.23)	(0.27)	(0.30)
2.5	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01
	(0.06)	(0.05)	(0.04)	(0.04)	(0.01)	(0.06)	(0.07)	(0.08)
3.0	1.00	1.04	1.00	1.00	1.00	1.00	1.00	1.00
	(0.02)	(0.20)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
L	---	3.173	2.934	2.808	2.735	3.545	3.926	4.279
				$\rho = 0.6$				
1.0	199.83	199.36	199.39	199.37	199.88	199.44	199.19	200.34
	(200.31)	(199.22)	(197.10)	(199.03)	(198.13)	(198.89)	(196.75)	(198.49)
1.1	49.93	27.68	19.98	17.16	15.97	18.90	16.74	16.07
	(49.41)	(27.23)	(19.26)	(16.30)	(15.05)	(17.97)	(14.94)	(13.58)
1.2	13.04	7.80	5.61	4.94	4.63	5.49	5.28	5.57
	(12.55)	(7.26)	(4.86)	(4.10)	(3.73)	(4.38)	(3.64)	(3.36)
1.3	5.26	3.53	2.76	2.57	2.45	2.90	3.07	3.38
	(4.72)	(2.99)	(2.07)	(1.81)	(1.67)	(1.89)	(1.70)	(1.78)
1.4	2.91	2.13	1.84	1.75	1.71	2.03	2.21	2.41
	(2.41)	(1.56)	(1.14)	(1.04)	(0.99)	(1.13)	(1.16)	(1.27)
1.5	1.95	1.57	1.43	1.39	1.37	1.59	1.73	1.86
	(1.38)	(0.94)	(0.72)	(0.68)	(0.66)	(0.80)	(0.87)	(0.96)
1.6	1.53	1.30	1.23	1.21	1.20	1.35	1.43	1.52
	(0.89)	(0.62)	(0.50)	(0.48)	(0.46)	(0.60)	(0.67)	(0.73)
1.7	1.29	1.16	1.13	1.12	1.11	1.20	1.26	1.32
	(0.62)	(0.43)	(0.36)	(0.35)	(0.33)	(0.46)	(0.51)	(0.57)
1.8	1.17	1.09	1.07	1.06	1.06	1.12	1.15	1.19
	(0.44)	(0.31)	(0.26)	(0.25)	(0.25)	(0.34)	(0.39)	(0.44)

Table 3 continued

Shift λ	S^2chart	AIB-V chart	AIB-MA-V chart			AIB-DMA-V chart		
			$w=2$	$w=3$	$w=4$	$w=2$	$w=3$	$w=4$
				$\rho=0.6$				
1.9	1.10	1.05	1.04	1.03	1.03	1.07	1.09	1.12
	(0.33)	(0.22)	(0.20)	(0.18)	(0.18)	(0.26)	(0.30)	(0.34)
2.0	1.05	1.03	1.02	1.02	1.02	1.04	1.05	1.07
	(0.24)	(0.16)	(0.15)	(0.14)	(0.14)	(0.20)	(0.23)	(0.26)
2.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.06)	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.06)	(0.07)
3.0	1.00	1.03	1.00	1.00	1.00	1.00	1.00	1.00
	(0.02)	(0.16)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
L	---	3.118	2.906	2.799	2.731	3.510	3.919	4.277
				$\rho=0.9$				
1.0	199.83	200.67	199.75	199.31	200.17	199.85	200.14	199.57
	(200.31)	(200.68)	(197.77)	(198.84)	(199.80)	(200.61)	(197.25)	(196.52)
1.1	49.93	13.74	9.66	8.19	7.45	9.04	8.07	7.99
	(49.41)	(13.20)	(8.88)	(7.18)	(6.32)	(7.85)	(6.29)	(5.47)
1.2	13.04	3.33	2.62	2.43	2.36	2.78	2.97	3.27
	(12.55)	(2.77)	(1.90)	(1.62)	(1.53)	(1.70)	(1.54)	(1.61)
1.3	5.26	1.69	1.52	1.48	1.46	1.71	1.89	2.05
	(4.72)	(1.07)	(0.82)	(0.75)	(0.72)	(0.85)	(0.92)	(1.01)
1.4	2.91	1.25	1.20	1.19	1.17	1.31	1.41	1.51
	(2.41)	(0.56)	(0.46)	(0.44)	(0.42)	(0.55)	(0.63)	(0.69)
1.5	1.95	1.09	1.08	1.07	1.07	1.14	1.19	1.24
	(1.38)	(0.32)	(0.28)	(0.27)	(0.26)	(0.37)	(0.43)	(0.48)
1.6	1.53	1.04	1.03	1.03	1.03	1.06	1.08	1.11
	(0.89)	(0.20)	(0.17)	(0.16)	(0.16)	(0.24)	(0.28)	(0.32)
1.7	1.29	1.01	1.01	1.01	1.01	1.02	1.04	1.05
	(0.62)	(0.12)	(0.11)	(0.10)	(0.10)	(0.15)	(0.19)	(0.22)
1.8	1.17	1.01	1.00	1.00	1.00	1.01	1.01	1.02
	(0.44)	(0.08)	(0.07)	(0.06)	(0.06)	(0.10)	(0.12)	(0.15)
1.9	1.10	1.00	1.00	1.00	1.00	1.00	1.01	1.01
	(0.33)	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)	(0.08)	(0.10)
2.0	1.05	1.00	1.00	1.00	1.00	1.00	1.00	1.01
	(0.24)	(0.04)	(0.02)	(0.03)	(0.02)	(0.04)	(0.06)	(0.07)
2.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.06)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
3.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.02)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
L	---	3.002	2.868	2.790	2.733	3.471	3.917	4.280

Observations from Tables 1-3:

- For any range of shifts in process standard deviation, the proposed AIB-DMA-V and AIB-MA-V charts consistently produces smaller out-of-control *ARL* values than the AIB-Shewhart-V chart and Shewhart S^2 chart. For example, in Table 1, with $\lambda = 1.3$, the traditional S^2chart requires on an average 11 samples to signal, the Shewhart AIB-V chart requires 8.03 samples to signal, the *ARL* reduces to 5.89 with $w = 2$, 5.18 with $w = 3$ and 4.83 with $w = 4$ for proposed AIB-MA-V chart and ARL reduces to 5.89 with $w = 2$, 5.55 with $w = 3$ and 5.80 with $w = 4$ for proposed AIB-DMA-V chart. That means, proposed

AIB-DMA-V and AIB-MA-V charts early detect shifts in process standard deviation earlier than the other two charts.

- The sensitivity of the proposed control charts increases as the span of moving average increases. For example, in Table 3, with $\lambda = 1.2$ and $w = 2$, the proposed AIB-DMA-V chart requires on an average 6.21 samples to signal, the ARL reduces to 5.80 and 6.07 with $w = 3$ and $w = 4$ respectively. The similar performance is observed for AIB-MA-V chart.
- The out-of-control ARL values of proposed AIB-DMA-V chart decreases when the correlation between study and auxiliary variables increases. For example, from Table 1, with fixed $\lambda = 1.3$and $w = 2$, the proposed AIB-DMA-V chart requires on an average 5.89 samples to signal when $\rho = 0.3$, the ARL reduces to 5.26 and 2.76 when $\rho = 0.6$ and $\rho = 0.9$ respectively. The similar performance is observed for AIB-MA-V and Shewhart-AIB-V charts.
- The performance of the proposed AIB-DMA-V and AIB-MA-V charts keeps improving with an increase in sample size n. For example, from Tables 1-3, with fixed $\lambda = 1.2$, $w = 3$ and $\rho = 0.6$, the proposed AIB-DMA-V chart requires on an average 9.35 samples to signal when $n = 10$, the ARL reduces to 6.66 and 5.28 when $n = 15$ and $n = 20$ respectively.

7. An Example

In this Section, we provide an illustrative example in order to demonstrate the practical application of AIB-MA-V, AIB-DMA-V charts for monitoring process variability using auxiliary information. Here we consider a simulated dataset to present implementation of AIB-Shewhart-V, AIB-MA-V and AIB-DMA-V control charts. To identify the performance of these control charts, the in-control ARL value is set as$ARL_0 = 200$. We have considered the paired information on (Y, X) where X is used as auxiliary variable and Y as the study variable of interest. The bivariate data set in the form of 15 sub-groups each of size $n = 10$ are simulated from$N_2(\mu_y, \mu_x, \sigma_y^2, \sigma_x^2, \rho)$ distribution. For in-control state, the first 7 samples are generated from$N_2(0,0,1,1,0.6)$. Thus the process is stable with respect to process variability for first 7 samples. We add 8 samples to simulate an out-of-control process. Starting from sample 8, new samples are generated from a process by introducing a shift $\lambda = 1.5$ in σ_y. Based on 15 subgroups, the values of the control chart statistic for AIB-MA-V and AIB-DMA-V control charts for span $w = 3$ and AIB-Shewhart-V chart are displayed in Table 4. The control limits of AIB-Shewhart-V chart are computed using Eq. (4) while those of AIB-MA-V and AIB-DMA-V charts are computed using Eq. (6) and Eq. (9) respectively. Implementation of the said charts is presented in Figure 1.

From Figure 1, it can be seen that the process remains in-control at the first seven samples. For detecting shift of size $\lambda = 1.5$ in process standard deviation, the AIB-Shewhart-V chart does not produce any out-of-control signal for detecting the shift. So AIB-Shewhart-V chart fail to detect a shift in process standard deviation when the shift occur. The AIB-MA-V chart shows first out-of-control signal at point 13. AIB-DMA-V chart shows first out-of-control signal at point 11 which is earlier than that of AIB-MA-V chart. Hence the proposed AIB-DMA-V chart is effective in detecting shifts in process standard deviation than AIB-MA-V and AIB-Shewhart-V charts.

Table 4: *Chart statistics based on simulated data*

Sample Number	AIB-V	AIB-MA-V	AIB-DMA-V
1	0.91	0.91	0.91
2	0.63	0.77	0.84
3	1.56	1.03	0.91
4	1.65	1.28	1.03
5	0.74	1.31	1.21

Table 4 continued

6	0.95	1.11	1.23
7	0.56	0.75	1.06
8	1.88	1.13	1.00
9	2.30	1.58	1.15
10	0.80	1.66	1.46
11	2.07	1.72	1.65
12	1.71	1.53	1.64
13	2.02	1.93	1.73
14	1.54	1.76	1.74
15	1.03	1.53	1.74

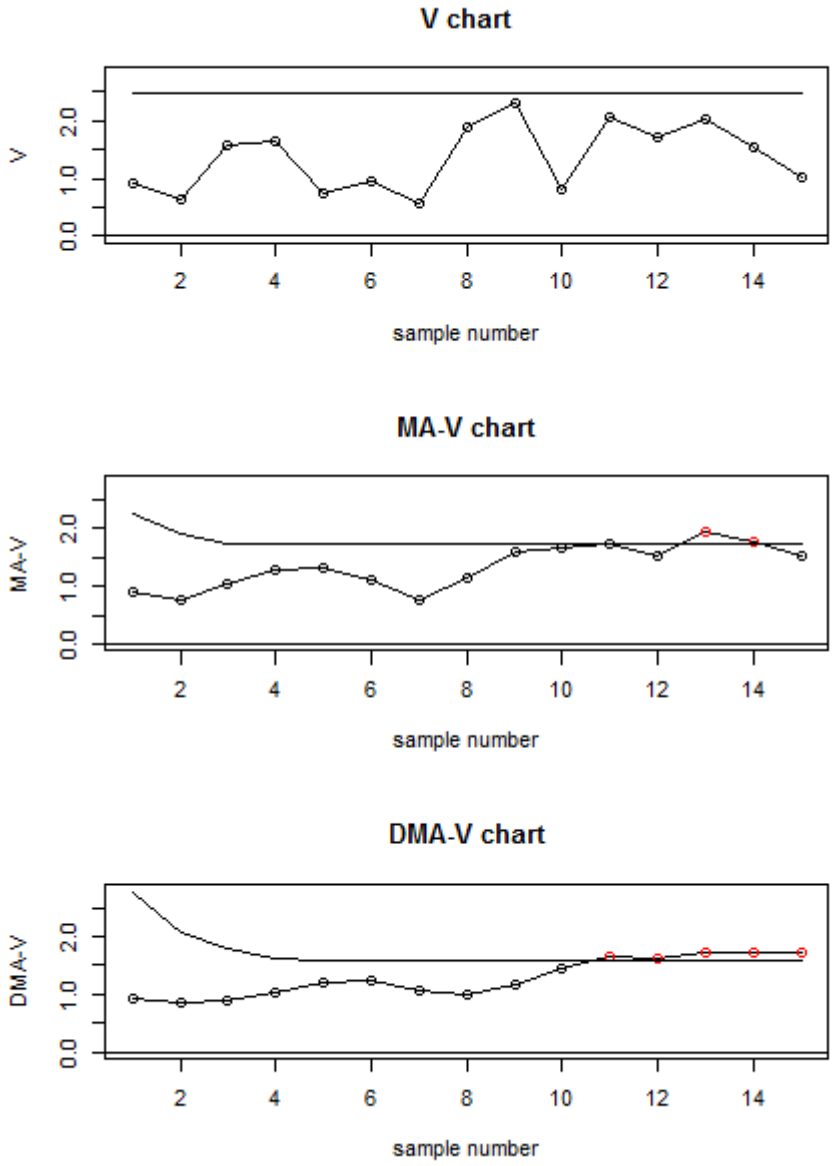


Figure 1: *AIB-Shewhart, AIB-MA and AIB-DMA Charts for process variability*

8. Conclusions

In this paper, we have proposed the AIB-MA and AIB-DMA control charts to efficiently monitoring the variability of normally distributed process. These charts are based on regression estimator of the process variance that utilizes information on a study variable as well as any related auxiliary variable. The construction, performance assessment and an illustrative example of proposed charts are presented in this paper. Using extensive Monte Carlo simulations, ARL and SDRL of the proposed AIB-MA-V and AIB-DMA-V chart has been computed with various choices of correlation coefficient ρ and span w. From the simulation results it is observed that with an increase in the value of ^{w} , the performance of the AIB-MA-V and AIB-DMA-V charts is improved. The performance of the proposed charts keep improving with an increase in the values of sample size n, level of correlation between study variable and auxiliary variable ρ and size of shift λ in process standard deviation at a fixedARL_0. The SDRL values of the charts are approximately the same as ARL values. The performance of the proposed charts is also compared to its existing counterparts incorporated in this study. It has been found that AIB-DMA-V and AIB-MA-V charts perform uniformly better than the AIB-Shewhart-V andS^2 charts for different kind of shifts in the process variability. Among AIB-DMA-V perform better than the AIB-MA-V chart for a span of $w = 2, 3$; while for span of $w = 4$, performance of both proposed charts is similar.

References

[1] Wong, H.B., Gan, F.F. and Chang, T.C. (2005). Design of Moving average control chart, *Journal of Statistical Computation and Simulation*, 74(1):47-62.

[2] Khoo, M.B.C. and Yap, P.W. (2005). Joint monitoring of process mean and variability with a single moving average control chart, *Quality Engineering*, 17(1): 51-65.

[3] Adeoti, O. A. and Olaomi, J. A. (2016). A moving average S control chart for monitoring process variability, *Quality Engineering*, 28(2): 212-219.

[4] Ghute, V.B. and Rajmanya, S.V. (2014). Moving average control charts for process dispersion, *International Journal of Science, Engineering and Technology Research*, 3(7): 1904-1909.

[5] Ghute, V.B. and Shirke, D.T. (2013a). A multivariate moving average control chart for mean vector, *Journal of Academia and Industrial Research*, 1(12): 796-800.

[6] Ghute, V.B. and Shirke, D.T. (2013b). A multivariate moving average control chart for process variability, *International Journal of Statistics and Analysis*, 3(2):97-104.

[7] Khoo, M.B.C. and Wong, V.H. (2008). A double moving average control chart, *Communication in Statistics-Simulation and Computation*, 37: 1696-1708.

[8] Adeoti, O. A., Akomoalfe, A. A. and Adebola, F. B. (2019). Monitoring process variability using double moving average control chart, *Industrial Engineering and Management System*, 18(2): 210-221.

[9] Sukparungsee, S., Saengsura, N., Areepong, Y. Phantu, S. (2021). Mixed Tukey-double moving average for monitoring of process mean, *Thailand Statistician, 19*(4): 855-865.

[10] Taboran, R., & Sukparungsee, S. (2023). On Designing of a New EWMA-DMA Control Chart for Detecting Mean Shifts and Its Application, *Thailand Statistician*, *21*(1): 148-164.

[11] Riaz, M. (2008). Monitoring process mean level using auxiliary information, *Statistica Neerlandica,* 62(4): 458-481.

[12] Riaz, M. and Does, R. J. m. M. (2009). A process variability control chart, Computational Statistics, 24(2): 345-368.

[13] Riaz, M. (2011). An improved control chart structure for process location parameter, *Quality and Reliability Engineering International*, 27: 1033-1041.

[14] Riaz, M., Abbasi, S. A., Ahmad, S. and Zaman, B. (2014). On efficient Phase II process monitoring charts, *The International Journal of Advanced Manufacturing Technology*, 70(9): 2263-2274.

[15] Abbas, N., Riaz, M. and Does, R.J.M.M. (2014). An EWMA-type control chart for monitoring the process mean using auxiliary information, *Communications in Statistics-Theory and Methods*, 43: 3485-3498.

[16] Abbasi, S. A. and Riaz, M. (2016). On dual use of auxiliary information for efficient monitoring, *Quality and Reliability Engineering International*, 32(2): 705-714.

[17] Riaz, M., Mehmood, R., Abbas, N. and Abbasi, S. A. (2016). On effective dual use of auxiliary information in variability control charts, *Quality and Reliability Engineering International*, 32(4): 1417-1443.

[18] Sanusi, R. A., Abbas, N. and Riaz, M. (2017). On efficient CUSUM-type location control charts using auxiliary information, *Quality Technology and Quantitative Management*, 15: 87-105.

[19] Haq, A. (2017). New EWMA control charts for monitoring process dispersion using auxiliary information, *Quality and Reliability Engineering International*, 33(8): 2597-2614.

[20] Amir, M. W., Raza, Z., Abbas, Z., Nazir, H. Z., Akhtar, N., Riaz, M. and Abid, M. (2020). On increasing the sensitivity of moving average control chart using auxiliary variable, *Quality and Reliability Engineering International*, 37(3): 1198-1209.

[21] Amir, M. W., Rani, M., Abbas, Z., Nazir, H. Z., Riaz, M. and Akhtar, N. (2021). Increasing the efficiency of double moving average chart using auxiliary variable, *Journal of Statistical Computation and Simulation*, 91(14): 2880-2898.

RT&A, No.3 (74)
Volume 18, September 2023

A NEW ZERO-INFLATED COUNT MODEL WITH APPLICATIONS IN MEDICAL SCIENCES

Zehra Skinder[1] Peer Bilal Ahmad*[2] Na Elah[3]

•

Department of Mathematical Sciences,
Islamic University of Sciene & Technology-192122, Kashmir[1,2,3]
zehraskinder7@gmail.com
bilalahmadpz@gmail.com*
naaelashah@gmail.com

Abstract

Inflated models are used whenever there are too many frequencies at a given count. In this regard, Poisson moment exponential distribution and a distribution to a point mass at zero are used to create a zero-inflated model namely Zero-Inflated Poisson Moment Exponential Distribution. Its distributional and reliability characteristics are investigated in some detail. A simulation exercise is undertaken to evaluate the effectiveness of the maximum likelihood estimators. The adaptability of the suggested distribution is demonstrated using three real datasets from various domains (e.g., vaccine adverse events, medical science data, epileptic seizure counts). The suggested distribution and the Poisson moment exponential distribution are distinguished by using the two different test procedures.

Keywords: goodness of fit, poisson Moment exponential distribution, likelihood ratio test, zero-inflation, wald test

1. Introduction

Statistical modelling of count observations is an essential part in several areas of scientific research. Frequent zeros in count observations are so common in areas like ecology, epidemiology, public health, engineering, etc. Examples of such data includes the number of foetal movements count per 5 seconds by Leroux *et.al* [16], number of HIV infected patients count by Van den Broek [32] in 1995, number of migrants count at household level by Shukla *et.al*[28] in 2006, number of accidents count due to heavy vehicular traffic for the year 2010 by Sharma *et.al* [27] in 2013, number of suicide cases count due to COVID-19 in India by Rahman *et.al* [23] in 2022, number of antenatal care service visits count by Bekalo *et.al* [4] in 2021. In order to model count observations with frequent zeros, number of Zero-Inflated models have been studied in the literature. The idea of zero-inflation was first given by Neyman [22] in 1939 and feller [9] in 1943 to overcome the situation of more zeros. Zero-inflated Poisson distribution (ZIPD) introduced by Mullahy [20] in 1986 as a mixture between Poisson distribution and a distribution at point mass zero with mixing probability (δ). The probability mass function of the distribution X (δ, ζ) is as follows.

$$P(x,\delta,\zeta) = \begin{cases} \delta + (1-\delta)e^{-\zeta}; x = 0 \\ (1-\delta)\frac{e^{-\zeta}\zeta^x}{x!}; x > 0 \end{cases}$$

where δ is a zero-inflation parameter (0<δ<1), $\zeta \geq 0$ and if δ=0, the distribution reduces to Poisson distribution. Several authors investigated ZIPD such as Singh [30] in 1962, Martin and Katti [17] in 1965, Goralski [10] in 1977, Lambert [14] in 1992, Bohning [5] in 1998 and

sim *et.al* [29] in 2018. Gupta [11] developed a generalized version of Zero-inflated Poisson model called as zero adjusted generalized Poisson model. The parameters of the zero-inflated Poisson model were estimated by Nanjundan and Naika [21] in 2012 by using moment method of estimation and compared with maximum likelihood estimators. Beckett *et.al* [3] used some natural calamities data to study the zero-inflated Poisson model and compared moment method and maximum likelihood method of estimation. Hall [12] in 2000 introduced the zero-inflated binomial distribution. Zero-inflated negative binomial distribution (ZINBD) was distinguished by Suresh *et.al* [31] in 2015. Zero-inflated negative binomial distribution studied by Mwalili *et.al* [19] in 2008 to accommodate extravagant zeros. Ahmad *et.al* [1] in 2014 studied the zero-inflated generalized power series distribution using Bayes estimators of functions of parameters under varied loss functions. Sandhyaa *et.al* [25] in introduced a model called Inflated-parameter Harris Distribution. Several structural properties were explored and characterization on the basis of probability generating function was also given. To check the applicability of the model, real life-data was also considered. Junnumtuam *et.al* [13] introduced a new discrete distribution called the Zero-Inflated Cosine Geometric (ZICG) Distribution for modelling over dispersed data with excessive zeros. Various structural properties like moment generating function, mean and variance were also obtained. Furthermore, confidence interval was also constructed by using the Wald method. The Bayesian method with highest posterior density method was also used to estimate the true confidence intervals. Dara and Ahmad [7] in 2012 introduced the Moment exponential distribution (ME) by weighting the exponential distribution in conformity with Fishers theory (1934). Scollnik [26] in 2022 obtained Bayesian analyses of an exponential-Poisson and related zero augmented type models. Maya *et.al* [18] in 2023 analysed the applications of Poisson moment exponential distribution in the contexts of time series analysis and regression analysis for real world phenomenon. Ahsan-ul-Haq [2] in 2022 introduced the Poisson moment exponential distribution (PMExD) by combining the Moment exponential and Poisson distribution by compounding technique and showed that the model is over dispersed and flexible for statistical data analysis. The probability mass function (p.m.f) of the PMExD is as follows.

$$P(Y = y) = \frac{\zeta^y(1+y)}{(1+\zeta)^{2+y}},$$

Where y=0, 1, 2, 3, ..., and $\zeta > 0$. The PMExD has been found in immense applications in various fields of medical sciences, engineering, entomology and education.
Since in many practical situations the different models like Poisson distribution, zero-inflated Poisson distribution, negative binomial distribution, discrete Weibull distribution, zero-inflated negative binomial distribution etc. are not preferable. In such cases, zero-inflated version of the PMExD provides better fit. For example, in the application section, different real-life datasets are considered. Only the zero-inflated version of PMExD gives best fit in comparison to existing models. So, in this paper we introduce zero-inflated poisson moment exponential distribution (ZIPMExD) along with distributional properties and other important aspects.
This paper is organized as follows. In section 2, we show the derivation of the ZIPMExD, cumulative distribution function. Also, the shapes of probability mass function (p.m.f) and cumulative distribution function (c.d.f) are presented in this section. In section 3, we have obtained the various structural properties along with reliability characteristics and generating functions. In section 4, we discuss the estimation of the parameters of the ZIPMExD by two different methods. A rigorous simulation study is also discussed in this section. In section 5, to check the significance of the inflation parameter, different test procedures are applied for examination. Certain real life data applications are considered in section 6 for highlighting the functionality of the model. Also, it is important to highlight that zero-inflated version of PMExD is not studied yet in the literature.

2. Zero-Inflated Poisson Moment Exponential Distribution

In this part, we present Zero-Inflated Poisson Moment Exponential distribution (ZIPMExD). We have derived the propability mass function of the proposed model along with cumulative distribution function.

Theorem 1. Let Y $\sim$ ZIPMExD (π, δ). Then the probability mass function (p.m.f) of Y is given as

$$P(Y=y)=\begin{cases}\pi+(1-\pi)\frac{1}{(1+\delta)^2} & ;y=0\\ (1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} & ;y=1,2,3,...,\end{cases}$$

Proof: If Y is a random variable of Poisson Moment Exponential distribution, then the probability mass function (p.m.f) of Y can be defined as

$$h(y)=\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} \qquad ;y=0,1,2,...,;\delta>0$$

The Zero-inflated distribution is an extra proportion added to the proportion of zero, then the probability mass function of Zero-inflated distribution is given as

$$p(Y=y)=\begin{cases}\pi+(1-\pi)h(y=0) & ;y=0\\ (1-\pi)h(y) & ;y=1,2,...\end{cases}$$

where $0<\pi<1$.
Then, the p.m.f of the ZIPMExD(π,δ) is obtained by substituting the probability mass function of the Poisson moment exponential random variable into Zero-inflated model. Therefore, it can be written as

$$\begin{cases}\pi+(1-\pi)\frac{1}{(1+\delta)^2} & ;y=0\\ (1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}} & ;y=1,2,3,...,\end{cases} \tag{1}$$

where $0<\pi<1$ and $\delta>0$

The Cumulative Distribution Function of ZIPMExD (π, δ) can be given as

$$F(Y)=P(Y\leq y)$$

$$=\sum_{z=0}^{y}P(Y=z)$$

$$=\pi+\frac{(1-\pi)}{(1+\delta)^2}\sum_{z=0}^{y}\frac{\delta^z(1+z)}{(1+\delta)^z}$$

$$=[1-(1-\pi)(y+\delta+2)\delta^{y+1}(1+\delta)^{-(y+2)}] \tag{2}$$

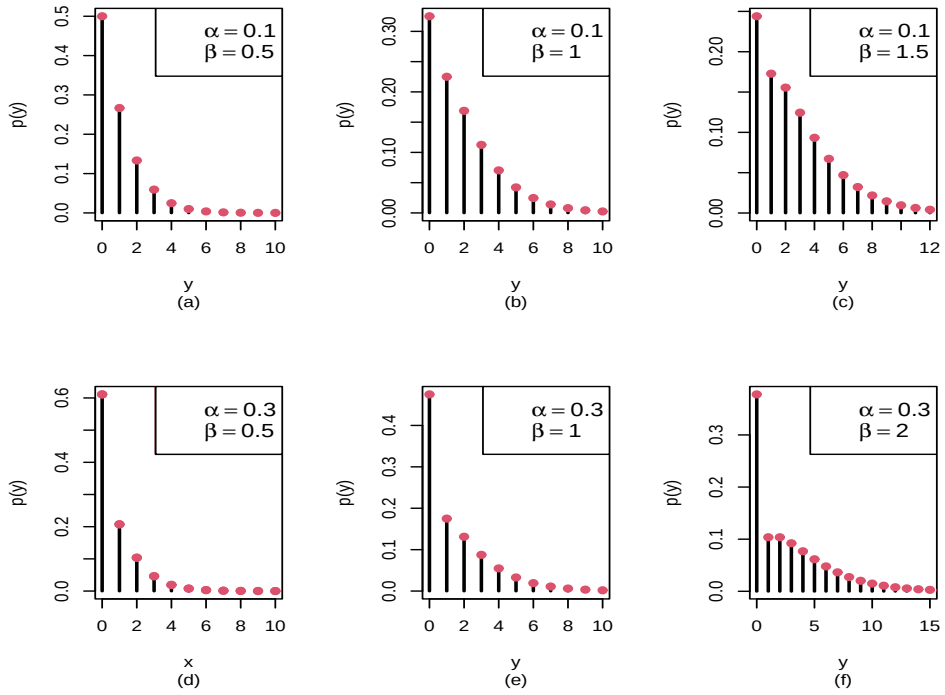


Figure 1: *The Pmf plots of ZIPMExD*

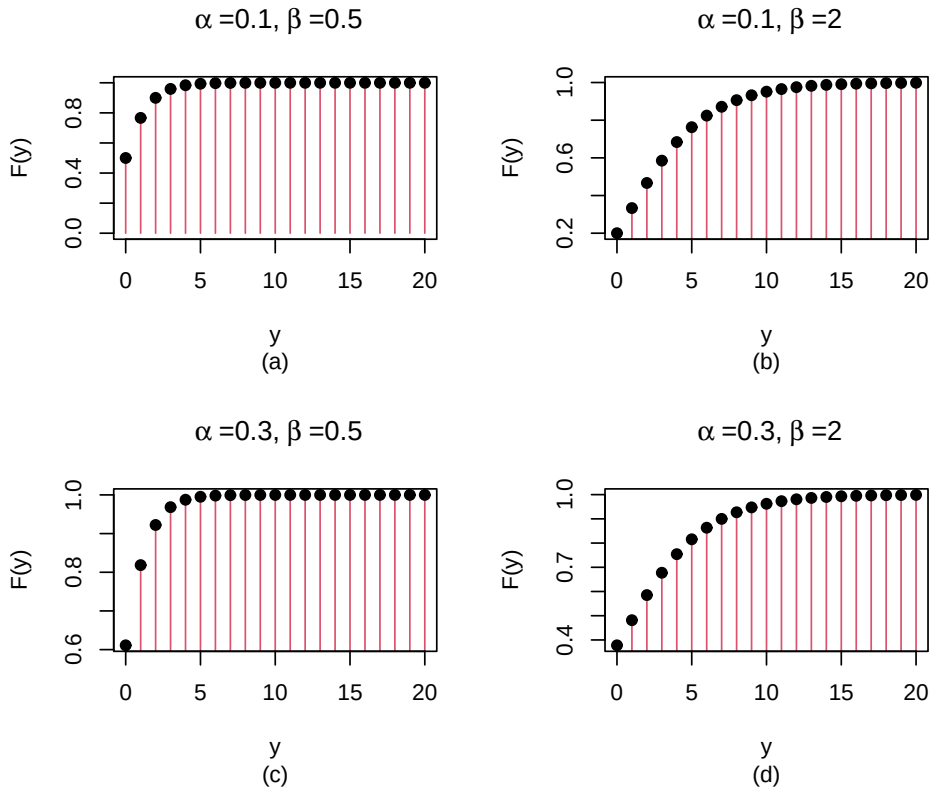


Figure 2: *The Cdf plots of ZIPMExD*

3. Structural properties along with reliability characteristics and generating functions.

In this part, we have obtained the survival function, hazard rate function, generating functions along with associated measures like index of dispersion (ID), skewness (β_1) and kurtosis (β_2) of the ZIPMExD (π, δ).

3.1. Survival Function (SF):

The Survival Function of ZIPMExD (π, δ) is as follows.

$$S(Y) = 1 - F(Y)$$

$$= (1-\pi)\left[\frac{(y+\delta+2)\delta^{y+1}}{(1+\delta)^{y+2}}\right] \tag{3}$$

3.2. Hazard Rate (HR):

Let y_1, y_2, y_3 ,..., y_n be a random sample from ZIPMEx (π, δ) distribution as given by equation (1) Define Z be the number of y_i' s taking the value zero. Then equation (1) can be written as follows

$$P(Y = y_i) = \left[\pi + (1-\pi)\frac{1}{(1+\delta)^2}\right]^Z \left[(1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}}\right]^{1-Z}$$

Now, using S(Y) from equation (3). The Hazard Rate of ZIPMExD (π, δ) is given as

$$H(y) = \frac{P(y)}{S(y)} = \frac{\left[\pi + (1-\pi)\frac{1}{(1+\delta)^2}\right]^Z \left[(1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}}\right]^{1-Z}}{\left[(1-\pi)\frac{(y+\delta+2)\delta^{y+1}}{(1+\delta)^{y+2}}\right]}$$

3.3. Reverse Hazard Rate (RHR):

$$R(y) = \frac{P(y)}{F(y)} = \frac{\left[\pi + (1-\pi)\frac{1}{(1+\delta)^2}\right]^Z \left[(1-\pi)\frac{\delta^y(1+y)}{(1+\delta)^{y+2}}\right]^{1-Z}}{[1-(1-\pi)(y+\delta+2)\delta^{y+1}(1+\delta)^{-(y+2)}]}$$

3.4. Moments and associated measures

3.4.1 Moment Generating Function:

The Moment Generating Function, $M_y(t)$ of ZIPMExD (π, δ) distribution is given as

$$M_y(t) = E(e^{tx}) = \sum_{y=0}^{\infty} e^{ty} P(Y = y)$$

$$= \pi + \frac{(1-\pi)}{(1+\delta)^2} \sum_{y=0}^{\infty} e^{ty} \left[\frac{\delta^y(1+y)}{(1+\delta)^y}\right]$$

$$= \pi + \frac{(1-\pi)}{(1+\delta)^2} \left[\sum_{y=0}^{\infty} \frac{(e^t\delta)^y}{(1+\delta)^y} + \sum_{y=0}^{\infty} y\frac{(e^t\delta)^y}{(1+\delta)^y}\right]$$

$$= \pi + \frac{(1-\pi)}{(1+\delta)^2}\left[\frac{(1+\delta)}{(1+\delta-e^t\delta)^2}\right] \tag{4}$$

Putting $e^t = e^{it}$ in equation (4), the Characteristic Function, $\phi_y(t)$ of ZIPMExD (π, δ) is defined as

$$\phi_y(t) = \pi + \frac{(1-\pi)}{(1+\delta)^2}\left[\frac{(1+\delta)}{(1+\delta-e^{it}\delta)^2}\right] \tag{5}$$

Through MGF, we have derived the first four raw moments of the proposed distribution by differentiating equation (4) at t=0. The first four raw moments of the proposed distribution are as fallows.

$$\mu_1' = (1-\pi)2\delta \tag{6}$$

$$\mu_2' = (1-\pi)2\delta(1+3\delta) \tag{7}$$

$$\mu_3' = (1-\pi)[2\delta(1+9\delta+12\delta^2)] \tag{8}$$

$$\mu_4' = (1-\pi)[2\delta(1+21\delta+72\delta^2+60\delta^3)] \tag{9}$$

3.4.2 Central Moments:

The first four central moments of the proposed distribution are obtained by using the relationship between raw moments and central moments. These are as follows

$$\mu_2 = (1-\pi)2\delta[1+\delta+2\pi\delta] \tag{10}$$

$$\mu_3 = (1-\pi)2\delta[1+3\delta+2\delta^2+6\pi\delta+34\pi\delta^2+8\pi^2\delta^2] \tag{11}$$

$$\mu_4 = (1-\pi)2\delta[1+13\delta+24\delta^2+12\delta^3+24\pi\delta^2+8\pi\delta+24\pi\delta+24\pi\delta^3+24\pi^2\delta^2+24\pi^3\delta^3] \tag{12}$$

Remark 4.1: The ZIPMExD is over dispersed for any $\delta > 0$ and $\pi = [0, 1]$.

proof: Suppose that the ZIPMExD is under dispersed. Then clearly
Mean>Variance, which implies that

$$(1-\pi)2\delta > (1-\pi)2\delta[1+\delta+2\pi\delta]$$

$$\Rightarrow 1 > [1+\delta+2\pi\delta]$$

which shows that $[1+\delta+2\pi\delta] < -1$, which is impossible for any $\delta > 0$ and $\pi = [0, 1]$. Hence the proof. The dispersion index (DI) of the proposed distribution is

$$Dispersion Index = \frac{var(y)}{mean(y)} = \frac{(1-\pi)2\delta[1+\delta+2\pi\delta]}{(1-\pi)2\delta}$$

$$= [1+\delta+2\pi\delta] > 1 \tag{13}$$

Further, Coefficient of variation(CV), Skewness and Kurtosis of the proposed model are given as follows:

$$CV = \frac{SD(y)}{Mean(y)} = \frac{\sqrt{(1-\pi)2\delta[1+\delta+2\pi\delta]}}{(1-\pi)2\delta} \tag{14}$$

$$Skewness(\sqrt{\beta}_1) = \frac{[1+3\delta+2\delta^2+6\pi\delta+8\pi^2\delta^2+34\pi\delta^2]}{\sqrt{(1-\pi)(2\delta)(1+\delta+2\pi\delta)^3}} \tag{15}$$

$$Kurtosis(\beta_2) = \frac{[1+13\delta+24\delta^2+12\delta^3+8\pi\delta+24\pi\delta^2+24\pi\delta^3+24\pi^2\delta^2+24\pi^3\delta^3]}{(1-\pi)2\delta[1+\delta+2\pi\delta]^2} \tag{16}$$

Table 1: ***Behaviour of the model's descriptive statistics for various parameter values.***

	π=0.1					π=0.3					π=0.6				
$\delta\rightarrow$	**0.5**	**1**	**1.5**	**2**	**2.5**	**0.5**	**1**	**1.5**	**2**	**2.5**	**0.5**	**1**	**1.5**	**2**	**2.5**
Mean	0.900	1.800	2.700	3.600	4.500	0.700	1.400	2.100	2.800	3.500	0.400	0.800	1.200	1.600	2.000
Variance	1.440	3.960	7.560	12.240	18.000	1.260	3.640	7.140	11.760	17.5	0.840	2.560	5.160	8.640	13.000
DI	1.600	2.200	2.800	3.400	4.000	1.800	2.600	3.400	4.200	5.000	2.100	3.200	4.300	5.400	6.500
CV	1.300	1.105	1.018	0.971	0.942	1.603	1.362	1.272	1.224	1.195	2.292	2.000	1.892	1.837	1.802
β_1	2.171	2.302	2.432	2.532	2.607	3.281	3.777	4.102	4.324	4.482	5.517	6.421	6.939	7.260	7.488
β_2	6.754	6.136	5.929	5.828	5.769	7.548	6.594	6.250	6.074	5.967	11.795	10.134	9.524	9.204	9.005

From the above table, it can be seen that for different combinations of parameters, the value of dispersion index is greater than one. So, the proposed model is over dispersed. For skewness, it can be seen that model is rightly skewed as the value of skewness increases for different combinations of parameters. Furthermore, from the table, it can be observed that the ZIPMExD is leptokurtic as the value of kurtosis is greater than three for different combinations of parameters.

4. Parametric Estimation

In this part, we have discussed the parametric estimation of the ZIPMExD (π, δ) by moment method of estimation and maximum likelihood method of estimation.

4.1. Moment Method of Estimation (MME)

The parameters π and δ of the proposed model can be obtained using this method as follows: Considering the first two raw moments from equation number (6) and (7)

$$\hat{\delta} = \frac{\mu_1'}{2-2\pi} \tag{17}$$

Now, from equation number (7), we have

$$\mu_2' = (1-\pi)2\delta(1+3\delta) \tag{18}$$

Putting the value of δ from (17) to (18), we get

$$\mu_2' = 2\left[\frac{\mu_1'}{2-2\pi}\right] + 6\left[\frac{\mu_1'}{2-2\pi}\right]^2 - 2\pi\left[\frac{\mu_1'}{2-2\pi}\right] - 6\pi\left[\frac{\mu_1'}{2-2\pi}\right]^2$$

$$2(\mu_2'-\mu_1')\pi^2 - (4\mu_2'-4\mu_1'-3\mu_1^2)\pi + (2\mu_2'-2\mu_1'-3\mu_1^2) = 0 \tag{19}$$

we can get estimated value of π on solving the above quadratic equation and that value of π has been used to estimate the value of δ in equation (17).

4.2. Maximum Likelihood Estimation Method (MLE)

The parameters π and δ of equation (1) can be obtained using this method as follows:
Let y_1, y_2, y_3, ... , y_n be a random sample from ZIPMExD (π, δ) as given by equation (1) and let for i=1, 2, 3, ... , n

$$b_i = \begin{cases} 1; if \quad y_i = 0 \\ 0; \quad otherwise \end{cases}$$

then, for i=1, 2, 3, ..., n equation (1) can be expressed as follows

$$P(Y = y_i) = \left[\pi + \frac{(1-\pi)}{(1+\delta)^2}\right]^{b_i} \left[\frac{(1-\pi)(1+y_i)\delta^{y_i}}{(1+\delta)^{2+y_i}}\right]^{1-b_i}$$

Hence the likelihood function; L=$L(\pi, \delta; y_1, y_2, y_3, ..., y_n)$ will be

$$L = \prod_{i=1}^{n} \left[\pi + \frac{(1-\pi)}{(1+\delta)^2}\right]^{b_i} \left[\frac{(1-\pi)(1+y_i)\delta^{y_i}}{(1+\delta)^{2+y_i}}\right]^{1-b_i}$$

$$= \left[\pi + \frac{(1-\pi)}{(1+\delta)^2}\right]^{n_0} \prod_{i=1}^{n} \left[\frac{(1-\pi)(1+y_i)\delta^{y_i}}{(1+\delta)^{2+y_i}}\right]^{d_i}$$

Where d_i=1-b_i, n_0=$\sum_{i=1}^{n} b_i$. Note that the number of zeros in the sample are represented by n_0. Therefore,

$$\log L = n_0 \log\left[\pi + \frac{(1+\pi)}{(1+\delta)^2}\right] + (n - n_0)\log(1-\pi) + \log\delta \sum_{i=1}^{n} d_i y_i + \log(1+\delta) \sum_{i=1}^{n} d_i (y_i + 2)$$

$$\frac{\partial \log L}{\partial \pi} = \frac{n_0[(1+\delta)^2 - 1]}{(1+\delta)^2\pi + (1-\pi)} - \frac{(n-n_0)}{(1-\pi)} \tag{20}$$

$$\frac{\partial \log L}{\partial \delta} = -\frac{2n_0(1-\pi)}{\pi(1+\delta)^3 + (1-\pi)(1+\delta)} + \frac{1}{\delta}\sum_{i=1}^{n} d_i y_i - \frac{1}{(1+\delta)}\sum_{i=1}^{n} d_i (y_i + 2) \tag{21}$$

$$Let, p = \pi + \frac{(1-\pi)}{(1+\delta)^2} \tag{22}$$

Now, let $\frac{\partial \log L}{\partial \pi}$=0, then from equation (20) and using equation (22),

$$1 - \pi = \frac{p(n-n_0)(1+\delta)^2}{n_0[(1+\delta)^2 - 1]} \tag{23}$$

Now, letting $\frac{\partial \log L}{\partial \delta}$=0, using equation (22), equation (21) reduces

$$-\frac{n_0[2(1-\pi)]}{p(1+\delta)^3} + \sum_{i=1}^{n} \frac{d_i y_i}{\delta} - \sum_{i=1}^{n} \frac{d_i (y_i + 2)}{(1+\delta)} = 0 \tag{24}$$

Now, if we replace p by their sample relative frequencies, i.e., by their sample estimates, the proportion of zeros in the sample, i.e., $\hat{p}$=n_0/n and then Equation (23) reduces to

$$1 - \pi = \frac{(n-n_0)(1+\delta)^2}{n[(1+\delta)^2 - 1]} \tag{25}$$

Now using equation (23), equation (24) can be written as

$$-\frac{[2(n-n_0)(1+\delta)^2]}{[(1+\delta)^2-1]}+\sum_{i=1}^{n}\frac{d_i y_i}{\delta}-\sum_{i=1}^{n}\frac{d_i(y_i+2)}{(1+\delta)}=0$$

$$\Rightarrow M(\delta)=0 \tag{26}$$

Where, $M(\delta)=-\frac{[2(n-n_0)(1+\delta)^2]}{[(1+\delta)^2-1]}+\sum_{i=1}^{n}\frac{d_i y_i}{\delta}-\sum_{i=1}^{n}\frac{d_i(y_i+2)}{(1+\delta)}$

Hence by any numerical means, say Newton Rapson method. Equation (26) can be solved to obtain $\hat{\delta}$ numerically. i.e., M ($\hat{\delta}$)=0
similarly, using equation (22), π can be estimated

$$\hat{\pi}=\frac{1}{n}\left[n_0-\frac{(n-n_0)}{[(1+\delta)^2-1]}\right] \tag{27}$$

Therefore, the maximum likelihood estimates (MLE) of the parameters δ and π are given by solving equation (26) numerically to find $\hat{\delta}$ and $\hat{\pi}$ given by equation (27) respectively.
In order to calculate the asymptotic variance-covariance matrix of the estimates the second order differentiations of the log-likelihood function are given here

$$\frac{\partial \log L}{\partial \pi^2}=-\frac{n_0[(1+\delta)^2-1]^2}{[(1+\delta)^2\pi+(1-\pi)]^2}-\frac{(n-n_0)}{(1-\pi)^2}$$

$$\frac{\partial \log L}{\partial \delta^2}=\frac{2n_o(1-\pi)[2(1+\delta)^2\pi+(1-\pi)]}{[\pi(1+\delta)^3+(1-\pi)(1+\delta)]^2}-\sum_{i=1}^{n}\frac{d_i y_i}{\delta^2}+\sum_{i=1}^{n}\frac{d_i(y_i+2)}{(1+\delta)^2}$$

$$\frac{\partial \log L}{\partial \delta \partial \pi}=\left[\frac{2n_0[\pi(1+\delta)^2+(1-\pi)]-2n_0(1-\pi)[(1+\delta)^2-1]}{[\pi(1+\delta)^2+(1-\pi)]^2}\right]$$

By inverting the Fisher's information matrix (I), the asymptotic variance-covariance matrix of the maximum likelihood estimates of δ and π for ZIPMExD (δ, π) can be obtained as

$$\text{I}=\begin{bmatrix} E\left(-\frac{\partial \log L}{\partial \pi^2}\right) & E\left(-\frac{\partial \log L}{\partial \delta \partial \pi}\right) \\ E\left(-\frac{\partial \log L}{\partial \pi \partial \delta}\right) & E\left(-\frac{\partial \log L}{\partial \delta^2}\right) \end{bmatrix}$$

The ingredients of the above Fisher's information matrix can be obtained as

$$E\left(-\frac{\partial \log L}{\partial \pi^2}\right)=\left(-\frac{\partial \log L}{\partial \pi^2}\right)\Bigg|_{\pi=\hat{\pi},\delta=\hat{\delta}} \tag{28}$$

The asymptotic distribution of the maximum likelihood estimator ($\hat{\delta}$, $\hat{\pi}$) is given by

$$\sqrt{n}\begin{pmatrix}\hat{\delta}\\ \hat{\pi}\end{pmatrix}_{MLE}\xrightarrow{L}AN\left(\begin{pmatrix}\hat{\delta}\\ \hat{\pi}\end{pmatrix},I^{-1}\right),asn\to\infty$$

5. Testing

In this part, we have checked the significance of inflation parameter (π) by likelihood ratio test and score test.

5.1. Likelihood Ratio Test

In order to test the significance of the inflation parameter π of the ZIPMExD, The Likelihood Ratio Test (LRT) is carried out to distinguish between PMExD (ζ) and ZIPMExD (π, δ). Here the null hypothesis is

$$H_o : \pi = 0 Vs H_1 \pi \neq 0$$

In case of LRT, test statistic is given by

$$-2ln\xi = 2(l_1 - l_2), \tag{29}$$

where, $l_1 = lnL(\hat{\theta}; y)$, Where $\hat{\theta}$ is the maximum likelihood estimator for $\theta = (\pi, \delta)$ without limitation, and $l_2 = lnL(\hat{\theta}^*, y)$, in which $\hat{\theta}^*$ is the maximum likelihood estimator for θ under the null hypothesis H_o. The test statistic described in equation (29) is asymptotically distributed as χ^2 with one degree of freedom.

Table 2: *Calculated value of test statistic in case of Likelihood Ratio Test.*

	$lnL(\hat{\theta}^*; y)$	$lnL(\hat{\theta}; y)$	Test statistic
Dataset 1	-6752.66	-6736.52	32.28
Dataset 2	-476.68	-430.85	91.66
Dataset 3	-595.86	-593.64	4.44

Since the critical value at 5% level of significance is 3.84 at one degree of freedom. It can be seen from the above table that the null hypothesis is rejected in all the three data sets. Hence we can say that the additional parameter in the model is significant.

5.2. Wald test

Here for testing the significance of inflation parameter π of ZIPMExD we assess the Wald test. To test the null hypothesis

$$H_0 : \pi = 0 Vs H_1 : \pi \neq 0$$

In case of Wald test, test statistic is given by

$$W_\pi = \frac{\hat{\pi}^2}{Var(\hat{\pi})}, \tag{30}$$

Where $Var(\hat{\pi})$ represents the diagonal element of Fisher information matrix at $\pi = \hat{\pi}$ and $\delta = \hat{\delta}$.The test statistic given in equation (30) is asymptotically distributed as chi^2 with one degree of freedom.

Table 3: *Calculated value of test statistic in case of Wald Test.*

	Test statistic
Dataset 1	39.73
Dataset 2	283.96
Dataset 3	4.72

Since at one degree of freedom, the critical value at 5% level of significance is 3.84. It can be seen from the above table that the null hypothesis is rejected in all the three data sets. Hence we can say that the additional parameter in the model is significant.

5.3. Simulation

In this section, we carry a simulation study to investigate the finite sample behaviour of the maximum likelihood estimators for different sample sizes (n=25,75,100,300,600) on various parameter settings. The procedure was repeated 1000 times for calculation of Bias, Variance, Mean Square Error (MSE) and Coverage Probability and the results are given in Table2. It can be seen from the table, that as the sample size increases, the bias, variance and mean square error decreases and are close to zero for large sample sizes. Also, the coverage probability tends to 0.95 as the sample size increases. These results suggest that maximum likelihood estimates are consistent and therefore can be used in estimating the unknown parameters of the proposed model.

Table 4: *Simulation table of MLE's for proposed model*

Sample		$\pi = 0.3, \delta = 0.9$				$\pi = 0.5, \delta = 2$			
Size(n)	Parameter	Bias	Variance	MSE	Coverage probability (95%)	Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\pi}$	-0.06354	0.02335	0.02739	0.98	-0.03026	0.00832	0.00923	1.00
	$\hat{\delta}$	-0.08737	0.06415	0.07178	0.90	-0.00590	0.24814	0.24818	1.00
75	$\hat{\pi}$	-0.01913	0.00818	0.00854	0.98	-0.01438	0.00422	0.00442	0.96
	$\hat{\delta}$	-0.03034	0.02751	0.02843	0.92	0.00707	0.06000	0.06005	0.98
100	$\hat{\pi}$	0.00948	0.00646	0.00655	0.96	-0.00521	0.00188	0.00191	1.00
	$\hat{\delta}$	0.03178	0.02177	0.02278	0.94	0.00771	0.04532	0.04538	1.00
300	$\hat{\pi}$	-0.01206	0.00241	0.00256	0.92	-0.00828	0.00088	0.00095	1.00
	$\hat{\delta}$	-0.00248	0.00565	0.00565	0.98	-0.03233	0.01900	0.02005	0.96
600	$\hat{\pi}$	-0.01220	0.00065	0.00080	0.98	0.00144	0.00064	0.00064	0.94
	$\hat{\delta}$	-0.00941	0.00263	0.00272	0.96	0.02842	0.01513	0.01594	0.90
Sample		$\pi = 0.5, \delta = 2.5$				$\pi = 0.5, \delta = 0.85$			
Size(n)	Parameter	Bias	Variance	MSE	Coverage probability (95%)	Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\pi}$	-0.00174	0.01119	0.01120	0.94	-0.05718	0.02309	0.02636	0.98
	$\hat{\delta}$	-0.01929	0.62737	0.62775	0.90	-0.00713	0.11008	0.11013	0.94
75	$\hat{\pi}$	-0.01586	0.00481	0.00506	0.96	0.02007	0.00812	0.00853	0.94
	$\hat{\delta}$	-0.03521	0.09355	0.09479	0.94	0.01183	0.03575	0.03589	0.92
100	$\hat{\pi}$	-0.00893	0.00341	0.00349	0.94	0.00874	0.00646	0.00654	0.94
	$\hat{\delta}$	-0.05368	0.09101	0.09389	0.98	0.00454	0.02240	0.02242	0.96
300	$\hat{\pi}$	0.00236	0.00110	0.00111	0.98	-0.00032	0.00243	0.00243	0.92
	$\hat{\delta}$	0.01784	0.03013	0.03045	0.98	0.00635	0.00626	0.00630	0.98
600	$\hat{\pi}$	-0.00198	0.00034	0.00035	1.00	0.00133	0.00097	0.00097	0.94
	$\hat{\delta}$	0.02452	0.01711	0.01771	0.96	-0.00472	0.00378	0.00381	0.96
Sample		$\pi = 0.1, \delta = 0.5$				$\pi = 0.5, \delta = 1.6$			
Size(n)	Parameter	Bias	Variance	MSE	Coverage probability (95%)	Bias	Variance	MSE	Coverage probability (95%)
25	$\hat{\alpha}$	-0.01360	0.00971	0.00989	0.98	-0.02185	0.01302	0.01350	0.98
	$\hat{\theta}$	0.00515	0.05491	0.05494	0.96	-0.14024	0.15844	0.17811	0.90
75	$\hat{\alpha}$	0.00575	0.00632	0.00635	0.98	-0.01570	0.00438	0.00463	0.96
	$\hat{\theta}$	0.01937	0.01274	0.01312	0.98	-0.03387	0.09477	0.09592	0.86
100	$\hat{\alpha}$	0.00114	0.00623	0.00624	0.94	-0.00752	0.00470	0.00476	0.94
	$\hat{\theta}$	0.01932	0.01359	0.01396	1.00	-0.04282	0.05458	0.05642	0.94
300	$\hat{\alpha}$	-0.00765	0.00200	0.00206	1.00	0.00347	0.00127	0.00128	0.96
	$\hat{\theta}$	-0.01014	0.00706	0.00717	0.90	0.02095	0.01841	0.01885	0.96
600	$\hat{\alpha}$	-0.00306	0.00152	0.00153	0.92	-0.00371	0.00066	0.00067	0.96
	$\hat{\theta}$	-0.00216	0.00262	0.00263	0.94	-0.02580	0.01039	0.01105	0.88

6. Applications

In this part, we study the practical significance of Zero-Inflated Poisson Moment Exponential Distribution(ZIPMExD). Three real life data sets are taken to compare Zero-Inflated Poisson Moment Exponential Distribution (ZIPMED) with few other distributions.

6.1. Data set 1

The dataset from Table 5 consists of frequencies regarding the number of vaccine adverse events originally given by Rose *et.al* [24] in 2006. Total number of events were recorded after each of the four injections for the 1005 study participants, which results in 4020 observations. Dar*et.al* [8] recently used Poisson weighted exponential distribution to fit the number of vaccine adverse events data. After analysing data through R software, we can see that our model performs better than other competing models because of highest p-value i.e., (0.928) among all other competing distributions and we can also see that our model favours the criteria i.e., Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) among all other competing models because of lowest values. The other competing models we use in this paper are Poisson Moment distribution (PMD), Zero-inflated Poisson Distribution (ZIPD), Poisson Distribution (PD), Negative Binomial Distribution (NBD), Discrete Weibull Distribution (DWD) and Zero-inflated Negative Binomial Distribution (ZINBD).

Table 5: *Expected frequencies and χ^2 values for fitted models*

Claims	observed count	ZIPMExD	PMExD	ZIPD	PD	NBD	DWD	ZINBD
0	1437	1437	1307	1437	891	1119	1411	1437
1	1010	1009	1124	787	1342	1225	1066	958
2	660	681	724	803	1011	838	668	708
3	428	408	415	546	508	459	393	436
4	236	230	223	279	191	220	223	241
5	122	124	115	114	58	96	123	125
6	62	65	58	39	14	39	66	61
7	34	33	28	11	3	15	35	29
8	14	17	14	3	1	6	18	13
9	8	8	7	1	0	2	9	6
10	4	4	3	0	0	1	5	3
11	4	2	1	0	0	0	2	1
12	1	1	1	0	0	0	1	1
Degrees of Freedom		8	8	5	6	9	8	7
ML Estimates		$\hat{\pi}$=0.0836 $\hat{\delta}$=0.8324	$\hat{\delta}$=0.7534	$\hat{\lambda}$=2.0405 $\hat{\pi}$=0.2614	$\hat{\lambda}$=1.5069	$\hat{p}$=0.5032 $\hat{r}$=1.5267	$\hat{q}$=0.6491 $\hat{\beta}$=1.1469	$\hat{p}$=0.6020 $\hat{r}$=2.6000 $\hat{a}$=0.1229
χ^2-value		3.08	36.64	301.57	1516.9	10.12	8.79	7.23
$p-value$		0.928	< 0.001	< 0.001	< 0.001	0.320	0.359	0.404
$-\hat{log}$		6736.52	6752.66	6868.79	7231.13	6740.60	6739.67	6737.84
AIC		13477.04	13507.32	13741.58	14464.26	13485.21	13483.35	13481.68
BIC		13489.64	13513.62	13754.18	14470.56	13497.80	13495.95	13500.58

6.2. Dataset 2

The dataset from Table 6 has been taken from [15]. The dataset is related to the HIV exposed infant data. The data has been taken from three concerned regions, Nairobi, Kisumu and Mombasa and the data reveals zero-inflation because of the measures that have been put in place to reduce the rate of Mother to Child Transmission (MTCT). A total of 494 samples were collected from 60 health centres in Kenya from these three regions for analysis. From the table, we can see that our model outperforms other competing models because of highest p-value among all other competing distributions and we can also see that our model has lowest criteria i.e., Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) among all other competing models.

Table 6: *Expected frequencies and χ^2-values for fitted models*

Claims	Observed count	ZIPMExD	PMExD	ZIPD	PD	NBD	DWD	ZINBD
0	378	378	323	378	308	354	336	378
1	59	54	123	47	145	86	108	57
2	26	31	35	38	34	31	35	30
3	13	16	9	20	5	13	11	15
4	7	8	2	8	1	5	4	8
5	11	4	0	3	0	2	1	6
Degrees of Freedom		3	2	2	5	2	2	2
ML Estimates		$\hat{\pi}$=0.6215 $\hat{\delta}$=0.2358 $\hat{\delta}$=0.6230		$\hat{\lambda}$=1.6051 $\hat{\pi}$=0.7061	$\hat{\lambda}$=0.4716	$\hat{p}$=0.5145 $\hat{r}$=0.5000	$\hat{q}$=0.3207 $\hat{\beta}$=1.0010	$\hat{p}$=0.6639 $\hat{r}$=2.6000 $\hat{a}$=0.6416
χ^2		2.04	69.90	13.76	172.97	63.53	63.96	3.15
$p-value$		0.564	<0.001	<0.001	<0.001	<0.001	<0.001	0.207
$-\hat{log}$		430.85	476.68	435.39	524.23	438.52	456.12	431.10
AIC		865.70	955.36	874.79	1050.64	881.04	916.25	868.20
BIC		874.10	959.56	883.20	1054.84	889.45	924.66	880.81

6.3. Dataset 3

The dataset from Table7 represents the frequencies of epileptic seizure counts reported in [6]. The measures of goodness-of-fit for all competing distributions are presented and it is evident that the proposed distribution fits well, as it has the highest p-value and lowest AIC and BIC criteria. So, in this regard we see that our model fits better than other fitted models on the given data set.

Table 7: *Expected frequencies and χ^2 values for fitted models*

Claims	Observed count	ZPMExD	PMExD	ZIPD	PD	NBD	DWD	ZINB
0	126	126	112	126	75	120	120	126
1	80	85	97	65	116	93	93	79
2	59	59	64	69	89	59	59	62
3	42	36	37	49	46	35	35	39
4	24	21	20	26	18	20	20	22
5	8	11	11	11	5	11	11	12
6	5	6	5	4	1	6	6	6
7	4	3	3	1	0	3	3	3
8	3	2	1	0	0	2	2	1
Degrees of Freedom		5	6	4	2	5	5	4
ML Estimates		$\hat{\pi}$=0.0959 $\hat{\delta}$=0.8540	$\hat{\delta}$=0.7720	$\hat{\lambda}$=2.1196 $\hat{\pi}$=0.2715	$\hat{\lambda}$=1.5441	$\hat{p}$=0.5009 $\hat{r}$=1.5500	$\hat{q}$=0.6577 $\hat{\beta}$=1.1560	$\hat{p}$=0.6455 $\hat{r}$=3.3845 $\hat{a}$=0.1710
χ^2-value		3.5	9.66	16.68	82.45	6.10	6.10	2.87
$p-value$		0.622	0.139	0.002	<0.001	0.296	0.296	0.579
$-\hat{log}$		575.70	577.92	599.63	636.04	594.94	594.74	576.00
AIC		1155.41	1157.84	1203.27	1274.09	1193.88	1193.49	1156.41
BIC		1162.09	1163.07	1210.99	1277.95	1201.60	1201.22	1167.93

7. Conclusion

A new Zero-inflated version of Poisson moment exponential distribution is introduced in this paper namely Zero-inflated Poisson Moment Exponential Distribution (ZIPMExD). Key statistical properties of the distribution including generating functions, reliability characteristics and moments have been derived. For parametric estimation purpose, two different methods i.e., moment method and maximum likelihood method of estimation have been used. Simulation study has been done for evaluating the proficiency of the estimation measures considered in this paper. Further, the procedure of Log Likelihood ratio test and Wald test are designed for testing the significance of inflation parameter. Three real life data sets are reviewed for demonstrating the practicality of the introduced model juxtapose to the existent models being PMExD, PD, ZIPD, DWD, NBD and ZINBD. We can see that ZIPMExD in terms of Chi-square value and p-value gives best fit as the existent models do not show best fit. The information measures like Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in terms of numerical value reveals that ZIPMExD can be considered as a suitable model in comparison to other models as discussed in this paper.

References

[1] Ahmad, P. B. (2019). Bayesian Analysis of Zero-Inflated Generalized Power Series Distributions Under Different Loss Functions. Bayesian Analysis and Reliability Estimation of Generalized Probability Distributions, 1.

[2] Ahsan-ul-Haq, M. (2022). On Poisson moment exponential distribution with applications. Annals of Data Science, 1-22.

[3] Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., & Pal, N. (2014). Zero-inflated Poisson (ZIP) distribution: Parameter estimation and applications to model data from natural calamities. Involve, a Journal of Mathematics, 7(6), 751-767.

[4] Bekalo, D. B., & Kebede, D. T. (2021). Zero-inflated models for count data: an application to number of antenatal care service visits. Annals of Data Science, 8, 683-708.

[5] B?hning, D. (1998). Zero-inflated Poisson models and CA MAN: A tutorial collection of evidence. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 40(7), 833-843.

[6] Chakraborty, S. (2010). On some distributional properties of the family of weighted generalized Poisson distribution. Communications in Statistics—Theory and Methods, 39(15), 2767-2788.

[7] Dara ST, Ahmad M (2012) Recent advances in moment distribution and their hazard rates. LAP LAMBERT Academic Publishing, Chisinau

[8] Dar, S. A., Hassan, A., Ahmad, P. B., & Wani, S. A. (2021). A new count data model applied in the analysis of vaccine adverse events and insurance claims. Statistics in Transition new series, 22(3), 157-174.

[9] Feller, W. (1943). On a general class of" contagious" distributions. The Annals of mathematical statistics, 14(4), 389-400.

[10] GORALSKI, A. (1977). DISTRIBUTION Z-POISSON.

[11] Gupta, P. L., Gupta, R. C., & Tripathi, R. C. (1996). Analysis of zero-adjusted count data. Computational Statistics & Data Analysis, 23(2), 207-218.

[12] Hall, D. B. (2000). Zero?inflated Poisson and binomial regression with random effects: a case study. Biometrics, 56(4), 1030-1039.

[13] Junnumtuam, S., Niwitpong, S. A., & Niwitpong, S. (2023). Bayesian Computation for the Parameters of a Zero-inflated Cosine Geometric Distribution with Application to COVID-19 Pandemic Data. CMES-COMPUTER MODELING IN ENGINEERING & SCIENCES, 135(2), 1229-1254.

[14] Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34(1), 1-14.
[15] Kibika, S. A. (2020). The Zero Inflated Negative Binomial-Shanker distribution and its application to HIV exposed infant data (Doctoral dissertation, Strathmore University).
[16] Leroux, B. G., & Puterman, M. L. (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. Biometrics, 545-558.
[17] Martin, D. C., & Katti, S. K. (1965). Fitting of certain contagious distributions to some available data by the maximum likelihood method. Biometrics, 21(1), 34-48.
[18] Maya, R., Huang, J., Irshad, M. R., & Zhu, F. (2023). On Poisson Moment Exponential Distribution with Associated Regression and INAR (1) Process. Annals of Data Science, 1-19.
[19] Mwalili, S. M., Lesaffre, E., & Declerck, D. (2008). The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. Statistical methods in medical research, 17(2), 123-139.
[20] Mullahy, J. (1986). Specification and testing of some modified count data models. Journal of econometrics, 33(3), 341-365.
[21] Nanjundan, G., & Naika, T. R. (2012). Asymptotic comparison of method of moments estimators and maximum likelihood estimators of parameters in zero-inflated poisson model.
[22] Neyman, J. (1939). On a new class of" contagious" distributions, applicable in entomology and bacteriology. The Annals of Mathematical Statistics, 10(1), 35-57.
[23] Rahman, T., Hazarika, P. J., Ali, M. M., & Barman, M. P. (2022). Three-Inflated Poisson Distribution and its Application in Suicide Cases of India During Covid-19 Pandemic. Annals of Data Science, 9(5), 1103-1127.
[24] Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. Journal of biopharmaceutical statistics, 16(4), 463-481.
[25] Sandhyaa, E., & Abrahamb, T. L. (2016). Inflated-parameter Harris distribution. JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE-JMCS, 16(1), 33-49.
[26] Scollnik, D. P. (2022). Bayesian analyses of an exponential-Poisson and related zero augmented type models. Journal of Applied Statistics, 49(4), 949-967.
[27] Sharma, A. K., & Landge, V. S. (2013). Zero inflated negative binomial for modeling heavy vehicle crash rate on Indian rural highway. International Journal of Advances in Engineering & Technology, 5(2), 292.
[28] Shukla, K. K., & Yadava, K. N. S. (2006). The Distribution of the Number of Migrants at the Household Level. Journal of Population and Social Studies [JPSS], 14(2), 153-166.
[29] Sim, S. Z., Gupta, R. C., & Ong, S. H. (2018). Zero-inflated Conway-Maxwell Poisson distribution to analyze discrete data. The international journal of biostatistics, 14(1).
[30] Singh, S. N. (1962, January). Note on inflated Poisson-distribution. In ANNALS OF MATHEMATICAL STATISTICS (Vol. 33, No. 3, p. 1210). IMS BUSINESS OFFICE-SUITE 7, 3401 INVESTMENT BLVD, HAYWARD, CA 94545: INST MATHEMATICAL STATISTICS.
[31] Suresh, R., Nanjundan, G., Nagesh, S., & Pasha, S. (2015). On a characterization of Zero-inflated negative binomial distribution. Open Journal of Statistics, 5(06), 511-513.
[32] Van den Broek, J. (1995). A score test for zero inflation in a Poisson distribution. Biometrics, 738-743.