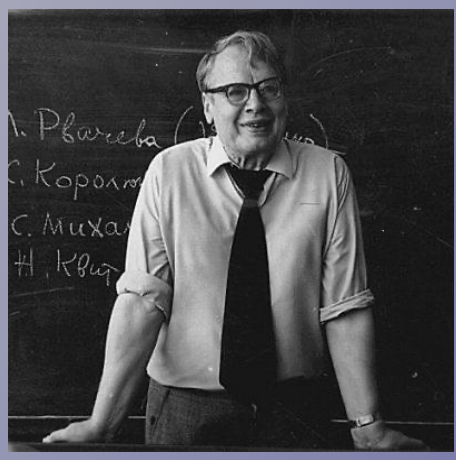
ELECTRONIC JOURNAL
OF INTERNATIONAL
GROUP ON RELIABILITY

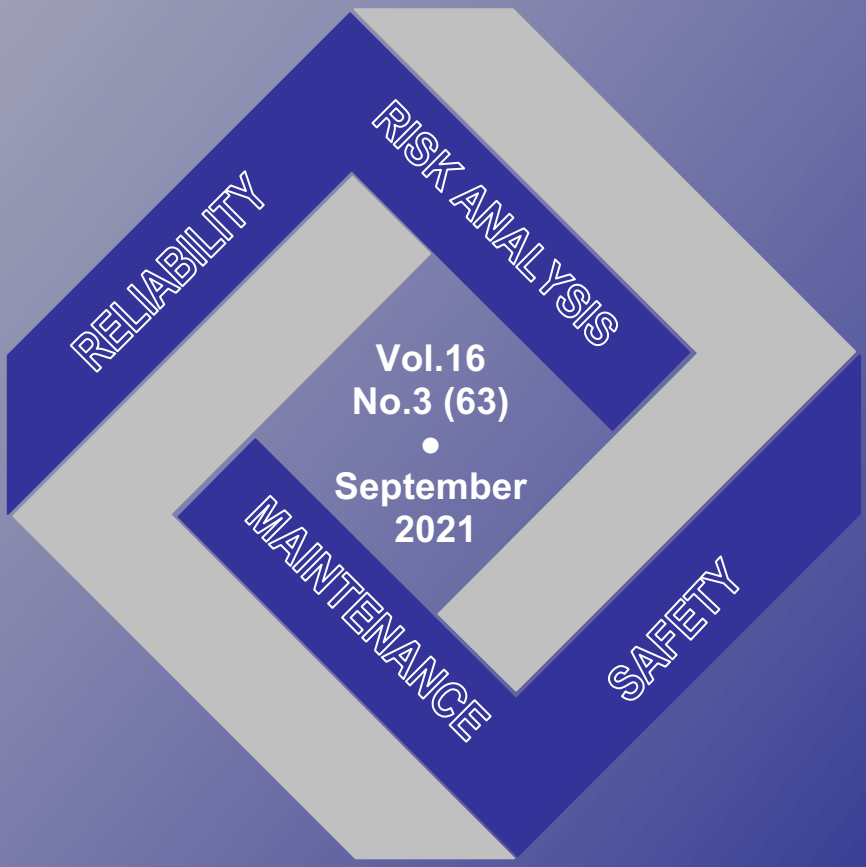
Gnedenko Forum Publications



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

ISSN 1932-2321

RELIABILITY:
THEORY&APPLICATIONS



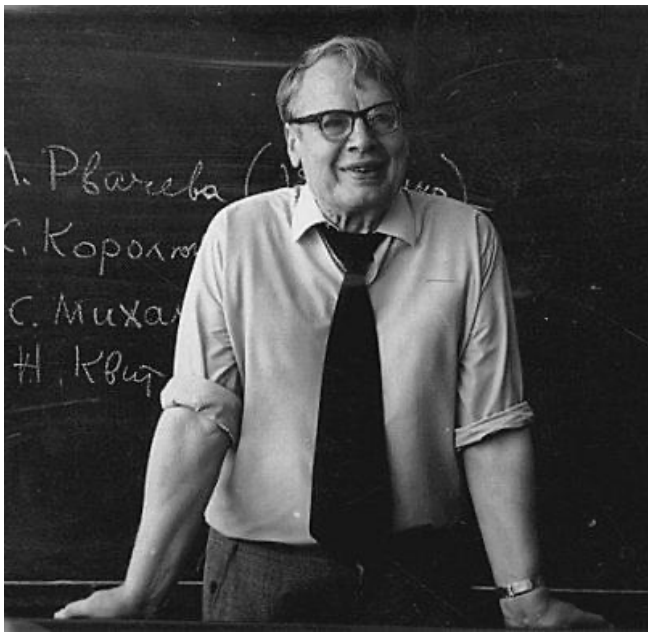
VOL.16 NO.3 (63)
SEPTEMBER, 2021

San Diego

ISSN 1932-2321

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





RELIABILITY:

THEORY & APPLICATIONS

Vol.16 No.3 (63),
September 2021

San Diego
2021

Editorial Board

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

Theoretical papers have to 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 have to 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 has to 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

Performance Analysis of the Water Treatment Reverse Osmosis Plant 16

Amrita Agrawal, Deepika Garg, Arun Kumar, Rakesh Kumar

In this research paper, profit analysis of a Water Treatment Reverse Osmosis (RO) Plant is carried out by using the Regenerative Point Graphical Technique (RPGT) under specific conditions for system parameters. The paper analyzes the behavior of a water treatment RO plant consisting of subunits namely Multimedia filter (MMF), Cartridge filter (CF), High-pressure pump (HPP), RO System (ROS). The system is in a working state when all subunits are in good condition. A repair facility is accessible for all subunits. Availability of the plant, Busy Period of the Server (BPS) and Expected number of inspection by the repairman (ENIR) is calculated by using the RPGT technique. Finally, numerical analysis is carried out for calculating the performance measures and their comparisons.

Reliability Test Plan For The Marshall-Olkin Extended Inverted Kumaraswamy Distribution **26**

Jiju Gillariose, Ehab M. Almetwally, Joshin Joseph, Vishna Devi

This paper mainly interested in studying the wider range behavior of the Marshall-Olkin extended inverted Kumaraswamy distribution. The parameters of model are estimated by various estimation methods. A reliability sampling plan is proposed which can save the test time in practical situations. Some tables are also provided for the new sampling plans so that this method can be used conveniently by practitioners. The developed test plan is applied to ordered failure times of software release to provide its importance in industrial applications.

Cluster Formation In An Acyclic Digraph Adding New Edges 37

Gurami Tsitsiashvili, Marina Osipova

In this paper, we construct an algorithm for converting an acyclic digraph that defines the structure of a complex system into a class of cyclically equivalent vertices by adding several additional edges to the digraph. This addition of the digraph makes it possible to introduce negative feedbacks and, consequently, to stabilize the functioning of the complex system under consideration and so to increase its reliability.To do this, the original digraph is transformed into a bipartite undirected graph, in which only the input and output vertices and the edges between them remain. In the constructed bipartite undirected graph, we search for the minimal edge cover and restore the orientation of the edges in it. Next, we construct an algorithm for adding new edges, based on the search for Hamiltonian (or Eulerian) paths and turning the minimum edge cover into a class of cyclically equivalent vertices. The minimal number of edges to be added is not larger than the number of edges in the minimum edge cover.

Truncated Shukla Distribution: Properties And Applications..... 41

Kamlesh Kumar Shukla, Rama Shanker

In this paper, Truncated Shukla distribution has been proposed. Some statistical properties including moments, coefficient of variation, skewness and index of dispersion have been derived. Survival and Hazard functions are derived and its behaviors are presented graphically. Maximum likelihood method of estimation has been used to estimate the parameter of proposed model. Simulation study of proposed distribution has also been discussed. It has been applied on three data sets and compares its superiority over two parameter Power Lindley, Gamma, Weibull, Shukla distributions and one parameter Truncated Akash, Truncated Lindley, Lindley and Exponential distributions

Dynamical Behavior of an SEIS Epidemic Model with Nonlinear Incidence Rate..... **56**

Garima Saxena, R. K. Sharma, Chandrashekhar Chauhan, Ankit Agrawal

In this paper, an epidemic model with nonlinear incidence rate is studied. The basic reproduction number ( ) is calculated. The local and global stability of the disease free equilibrium and the endemic equilibrium of the model are discussed and also the global asymptotical stability of the disease free equilibrium and endemic equilibrium are discussed. The stability analysis of the model shows that the system is locally asymptotically stable at disease free equilibrium and endemic equilibrium under suitable conditions. Moreover, show that the disease free equilibrium and the unique endemic equilibrium of the system is globally asymptotically stable under certain conditions. Finally, numerical simulations are given to support some of the theoretical results.

Sampled Ready Queue Processing Time Estimation Using Size Measure Information In Multiprocessor Environment..... **63**

Sarla More, Diwakar Shukla

In a multiprocessor computer system, there exist a ready queue of large number of processes waiting for computing resources allocation by the processors. These jobs may have size measure, which are additional information priory known while entry to the ready queue. Suppose the sudden system breakdown occurs and recovery management is required immediately. At this stage, one can find some processes who are completely finished, some partially processed, some blocked by processors and remaining waiting for allocation in the ready queue. Prime act of a system manager is to evaluate the maximum time required to process all the remaining jobs. This paper presents an estimation strategy for such, derived by applying the lottery scheduling, sampling technique and imputation methodology. Expressions for mean squared error of the proposed strategy are derived and optimized for suitable selection of system parameters. Three cases are discussed and compared and consequent results are numerically supported. It is found that at the optimal choice of constants in the estimation methodology, the shortest confidence interval can be predicted estimating the remaining required time. Such findings are useful as a part of disaster management of a cloud based multiprocessor data centre.

Fuzzy Regression Based Patient Life Risk Rate Prediction Using Oxygen Level, Pulse Rate And Respiration Rate In Covid-19 Pandemic (FRPRPS) **81**

Gaurav Kant Shankhdhar, Himanshu Pandey, Atul Kumar Pal, Sumit Mishra

Today, the general situation worldwide is that the hospitals, sanatoriums and medical colleges are running out of beds, oxygen, medical staff, ventilators and other required paraphernalia that is mandatory for the treatment of the vicious pandemic [1]. The requirement is for a system that takes in some input parameters like Oxygen level of the patient, pulse rate and respiration rate and in turn predicts the Life Risk Rate of that patient [2]. The model used here is a fuzzy regression model that gives the prediction of Life Risk Rate between 1 and 10 units. The lower the predicted Life Risk Rate, the better the chances of survival of the Covid patient. But if the predicted Life Risk Rate is more than the mean of the observations of the Risk in the dataset, then immediate emergency is needed. The benefit of this system is that the patients requiring immediate admission and treatment can be filtered and medical aid in hospital be thereby provided for critical patients. Rest may be home quarantined and domestic medical aid may be given to them until in some unfortunate situation their Risk Rate is near alarming. This paper aims to provide some help in this crucial situation.

Reliability, Availability, Maintainability, and Dependability (RAMD) Analysis of Computer Based Test (CBT) Network System 99

Abdullahi Sanusi, Ibrahim Yusuf, Na'fiu Hussain

Computer Based Test System also known as an e-examinations system, is software that can be used to administer examinations for distant or in-house applicants via internet or in an internet. Computer Based Test System/Software comprises of many components. So, it is vital to ensure its smooth operation, which can be achieved by the proper operation of its components/subcomponents. It is necessary to improve components/subcomponents operational availability. For this reason, the present research proposes to explore Computer Based Test System reliability indices using a RAMD technique at the component/subcomponent level. As a result, all subsystem/component transition diagrams are constructed, and the Chapman-Kolmogorov differential equations are formulated using the Markov birth-death process. For various subsystems/components of the system, numerical findings for reliability, availability, maintainability, and dependability, all of which are crucial to system performance, have been obtained and given in tables and figures. Other measurements, such as MTTF, MTBF, dependability ratio, and dependability minimum have also been obtained. Based on the numerical results, the most significant subsystem/component has been determined and the significance of the research has been emphasized.

Performance Evaluation of a Complex Reverse Osmosis Machine System in Water Purification using Reliability, Availability, Maintainability and Dependability Analysis 115

Anas Sani Maihulla, Ibrahim Yusuf, Saminu I. Bala

Today, reverse osmosis (RO) is a critical technique in the production of fresh water all over the world. As a result, downtimes due to repairing operations (after breakdowns, membrane blockage, pressure losses, etc.) or preventative maintenance (cleaning of membranes, component replacements, etc.) must be kept to a minimum in duration and frequency to guarantee optimum availability. Indeed, enhancing the availability (or dependability) of the RO plant as a whole system leads to a significant decrease in operating and maintenance expenses. We look at a recursive technique for reliability, availability, maintainability, and dependability in this study (RAMD. In addition, the efficacy of a RO unit, mean time to failure (MTTF), mean time to repair (MTTR), and dependability ratio were evaluated. The primary goal is economic optimization. For the method's validation, we utilized data from a RO unit that had a repair rate and a failure rate during a one-year period. It was demonstrated that all subsystems (pretreatment, dosage, etc.) had high availability. The high-pressure pump has a somewhat lower availability. For example, 0.59113 was the lowest availability for all subsystems, and it is for the RO membrane, which is where the majority of the purifications take place. A sensitivity analysis was performed to identify the essential components for the RO plant's availability. The collected findings demonstrate that the availability, reliability, dependability, and maintainability of the high-pressure pump have a significant impact on the overall system availability. As a result, special care should be given in the selection and maintenance of the high-pressure pump.

Review of Performance Factors of Emotional Speaker Recognition System: Features, Feature Extraction Approaches and Databases **132**

Satish Kumar Das, Uttpal Bhattacharjee, Amit Kumar Mandal

Emotion is a conscious mental reaction accompanied by physiological and behavior changes in human body.In speaker authentication system, emotional state of the speaker plays a vital role. Recently, the field of speaker recognition in emotional context attracts more and more attention of many research focuses. However, to implement more realistic and intelligent emotional speaker recognition system it is interesting to study this system under real life conditions. Speech emotion recognition is a system in which speech signals are processed to classify the embedded emotions. In recent past, speaker emotion recognition has gained a lot of attention from different researchers as it has many applications. In this regards, study of prior works is useful for further research in the field of speaker verification in emotional context. So, performance and reliability of Emotional Speaker Recognition System depend on the proper selection of features to characterize different emotional states, feature extraction approaches and databases. In this paper we briefly discuss about different features, feature extraction approaches and emotion recognition and speaker verification databases.

Parameter estimation for progressive censored data under accelerated life test with k levels of constant stress **149**

Mustafa Kamal

Accelerated life testing (ALT) is a time-saving technique that has been used in a variety of sectors to get failure time data for test units in a relatively short time it takes to test them under regular operating circumstances. One of the primary goals of ALT is to estimate failure time functions and reliability under typical use. In this article, an ALT with k increasing stress levels that is stopped by a type II progressive censoring (TIIPC) scheme is considered. At each stress level, it is assumed that the failure times of test units follow a generalized Pareto (GnP) distribution. The link between the life characteristic and stress level is considered to be log-linear. The maximum likelihood estimation (MLE) method is used to obtain inferences about unknown parameters of the model. Furthermore, the asymptotic confidence intervals (ACIs) are obtained by utilizing the inverse of the fisher information matrix. Finally, a simulation exercise is presented to show how well the developed inferential approaches performed. The performance of MLEs is assessed in terms of relative mean square error (RMSE) and relative absolute bias (RAB), whereas the performance of ACIs is assessed in terms of their length and coverage probability (CP).

Validation Of DNAFIDs Model Through Finite State Machine 160

Yogesh Pal, Santosh Kumar, Madhulika Singh, Shweta Dwivedi

The assurance of quality and reliability of process models and workflows is essential for model driven software development. There are numerous ways to achieve these objectives. One is model checking through which it can be verified that a model satisfies specific logical rules. The model to be checked is usually given as finite state machine. Rules have to be specified at the level required by the model checker. In this work, we develop a model for validating the DNA profiling through finite state. This enables the research/business process professionals to use model checking techniques and to produce higher quality research/business models for subsequent software development. The approach is demonstrated by validating event-driven process chains.

Design of One-sided Modified S Control Charts for Monitoring a Finite Horizon Process..... 168

Mei Tuan Teng, Sin Yin Teh, Khai Wah Khaw, XinYing Chew, Wai Chung Yeong

Control charting techniques are widely used in the manufacturing industry. One of the common charts that are used to monitor process variability is the S control chart. Finite horizon process monitoring has received great attention in the last decade. In the current literature, no attempt has been made to monitor the process variability in a finite horizon process. To fill this gap in research, this paper proposes two one-sided modified S charts for monitoring the standard deviation in a finite horizon process. The performance of the proposed charts is evaluated in terms of the truncated average run length and truncated standard deviation of the run-length criteria. The numerical performances of the proposed charts are shown with the selection of numerous process shifts. The effect of the sample sizes, the number of inspections and the process shifts are studied.

Availability And Performance Analysis Of Computer Network With Dual-Server Using Gumbel-Hougaard Family Copula Distribution 183

Ismail Tukur, Kabiru H. Ibrahim, Muhammad Salihu Isa, Ibrahim Yusuf

The determination of this paper is to study reliability measures and routine analysis of computer network, which is a combination of four subsystems A, B, C and D and all the subsystem connected in series parallel pattern, the subsystem A is client, the subsystem B is load balancer and subsystem C is servers which is divided in to two subsystem (i.e. subsystem C1 and subsystem C2) and C1andC2 served as computer servers together with two unit each and working 1-out-of-2: G policy, and subsystem D is centralized server. The system has two types of failure, degraded and complete failure. The system can completely fail due to failure of one of the following subsystems A, B, C and D. The system is at partial failed state if at least one unit is working in either subsystem C1 or subsystem C2.The system is examined using supplementary variables techniques and Laplace transform. General distribution and copula family are employed to restore degraded and complete failed state respectively. Calculated results have been highlighted by the means of tables and graphs.

Reliability Optimization Using Heuristic Algorithm In Pharmaceutical Plant..... **195**

Tripti Dahiya, Deepika Garg, Sarita Devi, Rakesh Kumar

In this paper, reliability of the liquid medicine manufacturing system of pharmaceutical plant named as Yaris Pharmaceuticals is enhanced on solving a redundancy allocation problem with the help of three algorithms HASL1 (Heuristic algorithm with selection factor 1), HASL2 (Heuristic Algorithm with Selection factor 2) and HASL3 (Heuristic Algorithm with Selection factor 3). It is ensured that redundancy is allocated within given cost constraints to maximize system reliability. Post allocation of redundancy the results of these algorithms are analyzed with the help of graphs, it has been found that the reliability of the system is optimized.

On Estimating Standby Redundancy System in a MSS Model with GLFRD Based on Progressive Type II Censoring Data..... 206

Marwa KH. Hassan

Redundancy is an approach to improve the reliability system. There are three main models of redundancy. In a system with standby redundancy, there are number of components only one of which works at a time and the ther remain as standbys. When an impact of stress exceeds the strength of the active component, for the first time, it fails and another from standbys, if there is any, is activated and faces the impact of stresses, not necessarily identical as faced by the preceding component and the system fails when all the components have failed. In This paper, we consider the problem of estimation the reliability of a multicomponent stress- strength system called N-M- cold -standby redundancy. This system includes N- subsystem consisting of M- independent distributed strength components only one of which works under the impact of stress. The system fails when all the components have failed. Assuming the stress and strength random variables have the generalized linear failure rate distribution with common scale parameters and different shape parameter. The reliability estimated based on progressive type II data. Simulation study is used to compare the performance of the estimators. Finally, real data set is used the proposed model in practice.

Critical Review Of Rams Tools And Techinques For The Analysis Of Multi Component Complex Systems 220

Shanti Parkash, P.C. Tewari

This work provides the critical review of usefulness of Reliability, Availability, Maintainability and Safety (RAMS) approaches in complex mechanical systems. A broad range of research works available such as articles, conference proceedings and books covering RAMS approaches in industries as well as in the field of research is critically reviewed. These include different tools, techniques and methods which may be helpful in qualitative as well as in quantitative analysis. It provides the informations about the past and current scenario of RAMS practices in industries as well as in research. In this work the authors look for certain articles which included two or more aspects of RAMS. Limited work is reported in the field of safety.

Effect Of Activation Function In Speech Emotion Recognition On The Ravdess Dataset 228

Komal D. Anadkat, Dr. Hiteishi M. Diwanji

Since last decade, Speech Emotion recognition has attracted extensive research attention to identify emotions by user's pitch and voice. Many research has been done in this field to recognize emotions using different machine learning as well as deep learning approaches. In this paper, we tried three different machine learning algorithms named SVM, Logistic regression and Random forest which take four different features named MFCC, Chroma, Mel-scale spectrogram and tonnetz as an input on RAVDESS dataset where SVM is more accurate than others. As deep learning approaches are more capable to identify hidden patterns and classify the data more accurately, we tried popular algorithm like MLP, CNN and LSTM. In deep learning approach, activation function is one of the most dominant parameters which a designer can choose to make classification more accurate. In this paper, we tried to show the effect of different activation functions on the overall accuracy of the model and analyzed the results.

On the Use of Entropy as a Measure of Dependence of Two Events **237**

Valentin Vankov Iliev

We define degree of dependence of two events A and B in a probability space by using Boltzmann-Shannon entropy function of an appropriate probability distribution produced by these events and depending on one parameter (the probability of intersection of A and B) varying within a closed interval I. The entropy function attains its global maximum when the events A and B are independent. The important particular case of discrete uniform probability space motivates this definition in the following way. The entropy function has a minimum at the left endpoint of I exactly when one of the events and the complement of the other are connected with the relation of inclusion (maximal negative dependence). It has a minimum at the right endpoint of I exactly when one of these events is included in the other (maximal positive dependence). Moreover, the deviation of the entropy from its maximum is equal to average information that carries one of the binary trials A [Ac and B [Bc with respect to the other. As a consequence, the degree of dependence of A and B can be expressed in terms of information theory and is invariant with respect to the choice of unit of information. Using this formalism, we describe completely the screening tests and their reliability, measure efficacy of a vaccination, the impact of some events from the financial markets to other events, etc. A link is available for downloading an Excel program which calculates the degree of dependence of two events in a sample space with equally likely outcomes.

Statistical Analysis of Marshall-Olkin inverse Maxwell Distribution: Estimation and Application to Real Data 249

C. P. Yadav, Jitendra Kumar, M. S. Panwar

In this paper, Marshall-Olkin inverse Maxwell distribution is proposed by generalizing the inverse Maxwell distribution under the Marshall-Olkin family of distribution that leads to greater flexibility in modeling various new data types. The basic statistical properties for the proposed distribution including moments, quantile function, median, skewness, kurtosis, and stochastic ordering are derived. Point estimates for the parameters are obtained by using two well known methods maximum likelihood and maximum spacing methods. The confidence intervals are used by using asymptotic properties of maximum likelihood estimators and boot-p methods. We have applied the proposed distribution under different real-life scenarios such as record value problem, system lifetime distributions, stress-strength reliability and random censored problems. For illustration purposes, simulation and real data results are established.

On Discrete Scheduled Replacement Model of a Series-Parallel System 273

Tijjani A. Waziri, Ibrahim Yusuf

This paper investigated the properties of discrete scheduled replacement model of a series-parallel system, with six units. The six units of the system formed three subsystems, which are subsystems A, B and C. Subsystem A is having three parallel units, subsystem B is having a single unit and subsystem C is having two parallel units. It is assumed that, the repairable system is subjected to two categories of failures (Category I and Category II). The mathematical expressions for both reliability function and failure rates, and an elementary renewal theorem were used based on some assumptions in constructing the discrete scheduled replacement model for a series-parallel system. A simple illustrative numerical example where made available, so as to study the properties of the replacement model constructed.

A New Generalization of Exponential Distribution for Modelling reliability Data **284**

V. Jilesh, Aifoona Ahammed P.M

In this paper, a new generalization of the exponential distribution is proposed. Different properties, important reliability measures and special cases of this distribution are investigated. Unknown parameters are estimated using the maximum likelihood method of estimation. A simulation study is carried out to assess the accuracy of the maximum likelihood estimates. Two real data sets are successfully modelled with the proposed distribution.

Analysis on Dual Supply Inventory Model having Negative Arrivals and Finite Life Time Inventory 295

M. L. Soujanya, P. Vijaya Laxmi

In this paper the impact of dual supply chain on a perishable inventory model with negative arrivals is evaluated. The perishable and replenishment rates of dual suppliers are distributed exponentially. Arrival process follows Poisson distribution and the probability for an ordinary customer is p and for the negative customer is q. Limiting distribution of the assumed model is obtained. Numerical results are presented for cost function and various system performance parameters. The impact of dual suppliers on the optimal reorder points will be useful in developing strategies for handling various perishable inventory problems with replenishment rates.

Harris Extended Two Parameter Lindley Distribution and Applications in Reliability 302

Sophia P.Thomas, Lishamol Tomy, K.K.Jose

This paper introduces a new generalization of the two parameter Lindley distribution distribution namely, Harris extended two parameter Lindley distribution. Various structural properties of the new distribution are derived including moments, quantile function, Renyi entropy, and mean residual life. The model parameters are estimated by maximum likelihood method. The usefulness of the new model is illustrated by means of two real data sets on Wheaton river flood and bladder cancer. Also, we derive a reliability test plan for acceptance or rejection of a lot of products submitted for inspection with lifetimes following this distribution. The operating characteristic functions of the sampling plans are obtained. The producer's risk, minimum sample sizes and associated characteristics are computed and presented in tables. The results are illustrated using two data sets on ordered failure times of products as well as failure times of ball bearings.

An Optimum Resource Score Estimation Method Using Bipartite Graph Model And Single Node Systematic Sampling..... 322

Deepika Rajoriya, D. Shukla

Consider a graphical population of vertices (nodes) and edges, where edges are connected with vertices to form a Bipartite graph. A complete Bipartite graph has vertices that can be partitioned into two subsets such that no edge has both endpoints in the same subset, and every possible edge connected to vertices in different subsets is a part of graph. In real life, there may hundreds of cities where at least one possible way exists reaching source to destinations. Several tourist places and small towns are the examples where the road transportation is available between origin and destination and these roads constitute Bipartite graph when they are like edges. The travel needs resource consumption who could be measured through resource-score. Walking at the hill station needs more energy consumption than at the plane area. This paper suggests an example to estimate the resource consumption by the values of score. Further, paper proposes a sample based methodology for calculating the average resource consumption between a pair of small town (city) and tourist place. Bipartite graph is used as a model tool. A single-node systematic sampling procedure is proposed under the Bipartite graph setup which is found useful for solution. The suggested estimation strategy is optimum at specific choice of parametric values. For quick selection, ready-reckoner tables are prepared who provide immediate optimum choice of constant. Results are numerically supported by the empirical study and proved by the calculation of confidence intervals.

Performance Analysis of the Water Treatment Reverse Osmosis Plant

Amrita Agrawal[1], Deepika Garg [1, #], Arun Kumar[1]

•

[1]School of Engineering and Sciences,
GD Goenka University, Gurugram (INDIA)
Corresponding author - [#]deepika.garg@gdgoenka.ac.in

Rakesh Kumar[2]

•

[2]Namibia University of Science and Technology, Windhoek (NAMIBIA)
rkumar@nust.na

Abstract

In this research paper, profit analysis of a Water Treatment Reverse Osmosis (RO) Plant is carried out by using the Regenerative Point Graphical Technique (RPGT) under specific conditions for system parameters. The paper analyzes the behavior of a water treatment RO plant consisting of subunits namely Multimedia filter (MMF), Cartridge filter (CF), High-pressure pump (HPP), RO System (ROS). The system is in a working state when all subunits are in good condition. A repair facility is accessible for all subunits. Availability of the plant, Busy Period of the Server (BPS) and Expected number of inspection by the repairman (ENIR) is calculated by using the RPGT technique. Finally, numerical analysis is carried out for calculating the performance measures and their comparisons.

Keywords: Regenerative Point Graphical Technique, Profit Analysis, Availability, Water Treatment Reverse Osmosis (RO) Plant.

I. Introduction

Reliability performance measures have incredible importance in the modern system such as the bread-making system, power plants and engineering systems. For making the system more significant, it is necessary to keep reliability measures up in the framework. In the majority of the systems, significant levels are kept up by giving skilled repair facility and upkeep activities. In some cases, redundant standby units are introduced to obtain the highest significant level.

In today's scenario, 3% of water is fresh on earth out of which 2.5% is unapproachable as it is in the form of glaciers, polar ice caps, atmosphere and soil, so only 0.5% of the water is accessible as freshwater. With only 0.5% water available, it's crucial to have Water Treatment Plant (WTP) to treat the wastewater and provide us freshwater for our daily use. For continuous working of these resources, it is essential to have timely maintenance of these systems to reduce the failure rate and keep the machines up and running. For upgrading and maintaining the efficiency of WTP's, unproductive time due to servicing (breakdown, jam of membrane, low pressure etc.) have to be minimized and assure maximum availability. Generally, the fundamental problem in the WTP is the low maintenance and poor quality material of the components used at the time of manufacturing. The solution to these problems is the regular use of safety measures and

maintenance techniques.

Thus Reliability, Availability and Maintainability (RAM) analysis of WTP's become a thoughtful issue for making the system more efficient and productive. Water treatment RO plant comprises of the following components which include Raw Water Forwarding Pump (RWFP), Flow Indicators (FI), Pressure Indicators (PI), Multi-Media Filter (MMF), Cartridge Filter (CF), Antiscalant Dosing pump with Tank (ASD), High-Pressure Pump (HPP), RO System (ROS), Product Water Storage Tank (PWST), Reject Water Storage Tank (RWST) and ancillary elements such as valves and gauges. The sub-system will fail if the primary and standby redundant units fail, thus producing total system failure. Cold standby excess units are switched in with the help of a perfect switch over the frameworks, which distinguishes the failure unit and switched in redundant standby unit.

Asi et al. (2021) studied a relative investigation of five productive dependability techniques to drive common rules for probabilistic evaluation of bridge pier. Li et al. (2020) discussed the time-dependent analysis with testing in practical engineering applications. Four models are developed to exhibit the effectiveness and exactness of the Improved Composite Limit state (ICLS) technique for the time-subordinate dependability analysis. Kumar et al. (2019) studied the behavior of the washing units in the paper industry by using the RPGT technique and noticing the framework's performance having all kinds of failures and test the workability of replacement of the breakdown structure. Kumar et al. (2018, 2017) have studied the behavior of a bread system and edible oil refinery plant. Zhai et al. (2015) developed an analytical technique based on a multi-valued verdict diagram to analyze the reliability of the system. Kumar et al. (2019) analyzed maintenance for a cold reserve framework that contains two identical subunits with server failure by using RPGT. Rajbala et al. (2019) studied the analysis and modeling: a case study EAEP industrial plant. Garg et al. (2009) analyzed the performance of a screw plant by using MATLAB Tool and cattle feed plant. Garg et al. (2010) articulated the crank availability of the component of the automobile industry taking the failure/repair rate of units as independent and solved the problem by using probability consideration and supplementary technique. Garg et al. (2010) discussed redundancy allocation in the pharmaceutical Plant. Wang et al. (2012) used some of the non-protective variables of distributions to demonstrate uncertainty, which was generally considered as stochastic factors for reliable models.

The main motive of this paper is to find the significant and critical parameters for the behavior and profit analysis of the water treatment RO plant by using the RPGT technique. For this purpose, State transition probabilities, availability, busy period of the server (BPS), maintenance specialist and profit analysis are evaluated. Finally, the numerical analysis is carried out for comparisons and comparing the results for making the system more efficient and productive.

II. Problem Description and Assumptions

I. System Description

The process diagram of the water treatment RO plant is shown in Figure 1.

- Multi-Media Filter (A):- It filters macro particles from the feed water. It consists of graded quartz and anthracite.
- Cartridge Filter (B):- This is a five-micron filter that filters micro particles from the feed water to enhance the membrane life by minimizing fouling on the membranes.
- High-Pressure Pump(C):- This pump creates the pressure above the osmotic pressure for reverse osmosis to take place.
- RO System (D):-It consists of RO Pressure vessels and RO Membranes.
    - RO Pressure Vessels (D1):- These are vessels that can take the load of the high-pressure created by the high-pressure pump and are also used to house

the RO membranes.

- RO Membranes (D2):- This is the heart of the system and the purification of the water is done by reverse osmosis process. The feed water is split into two streams; one is the stream of low TDS water called permeate and the other is the stream of high TDS water called Reject.

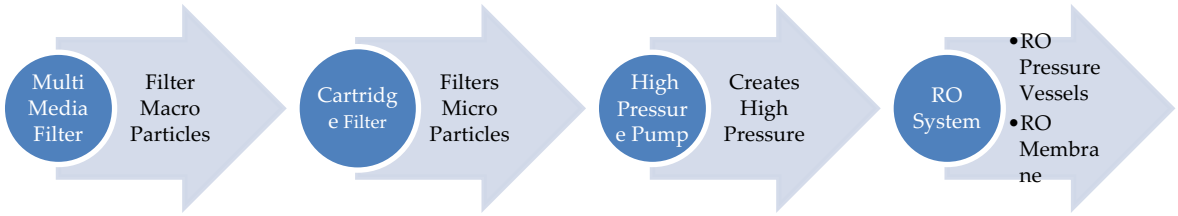


Figure 1: *Process Diagram of the Water Treatment RO Plant*

II. Notations

A, B, C, D : Working states
a, b, c, d : Failed states of A, B, C, D respectively
D_1, D_2 : Cold standby redundant D unit
s_i/r_i : Repair/Failure rates respectively; i = 1,2,3,4
$q_{i,j}(t)$: Probability distribution function from state S_i to S_j
$p_{i,j}$: Transition probability from state S_i to S_j
Ri(t) : Reliability of the system at time t, for the regenerative state S_i
μ_i : Mean sojourn time consumed in state S_i, before going in any other states
* : Laplace transform
T_0 : Mean Time to System Failure
A_0 : Availability of the System
B_0 : Mean Busy Period of the Server
V_0 : Expected Number of Inspections by the Repairman
P_0 : Profit Function
D_1 : Revenue per unit up-time of the system
D_2 : Cost per unit time in which system is under repair
D_3 : Cost due to inspection by the repairman

III. Assumptions

- The repair process begins soon after a unit fails.
- Failure and repair events are all statistically independent.
- The Repair unit is a new one.

IV. State Transition Diagram

S_1 : Initial Working state when all the four units are working; so system is working
S_5, S_9 : Reduced working states when units A, B, C are working; unit D is down and under repair; cold standby redundant units D_1, D_2 are working in place of unit D
S_2; S_6; S_{10} : Failed states when unit A fails and units B,C,D; D_1 ; D_2 are working
S_3; S_7; S_{11} : Failed states when unit B fails and units B, C, D; D_1 ; D_2 are working
S_4; S_8; S_{12} : Failed states when unit C fails and units B,C,D; D_1 ; D_2 are working
S_{13} : Failed state when unit D fails and units B, C, D are working

State S_1 is taken as the base state. By considering all the above annotations and assumptions, the

State Transition Diagram of the framework is shown in Figure 2.

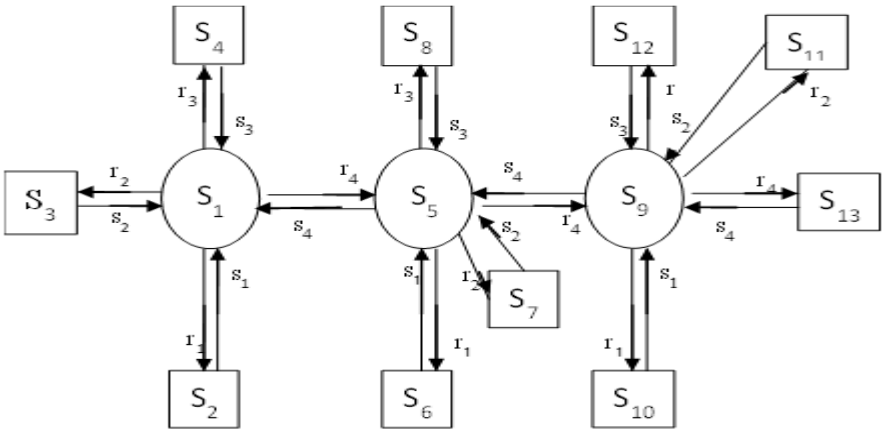


Figure 2: *Transition Diagram of the system*

S_1 = ABCD, S_2 = aBCD, S_3 = AbCD, S_4 = ABcD,
S_5 = ABCD$_1$, S_6 = aBCD$_1$, S_7 = AbCD$_1$, S_8 = ABcD$_1$,
S_9 = ABCD$_2$, S_{10} = aBCD$_2$, S_{11} = AbCD$_2$, S_{12} = ABcD$_2$,
S_{13} = ABCd

V. Transition Probabilities and Mean Sojourn Times (MST)

Table 1 and Table 2 represents the Transition probabilities and MST for the states i, j respectively.

Table 1: *Transition Probabilities*

$\mathbf{q_{i,j}(t)}$	$\mathbf{p_{ij} = q^*_{i,j}(0)}$
$q_{1,i}(t) = r_j e^{-(r_1+r_2+r_3+r_4)t}$; i =2,3,4,5 & j = 1,2,3,4	$p_{1,i} = r_j/(r_1+r_2+r_3+r_4)$ i =2,3,4,5 & j = 1,2,3,4
$q_{2,1} = s_1 e^{-s_1 t}$	$p_{2,1} = 1$
$q_{3,1} = s_2 e^{-s_2 t}$	$p_{3,1} = 1$
$q_{4,1} = s_3 e^{-s_3 t}$	$p_{4,1} = 1$
$q_{5,1}(t) = s_4 e^{-(r_1+r_4+r_3+r_2+s_4)t}$ $q_{5,i}(t) = r_j e^{-(r_1+r_4+r_3+r_2+s_4)t}$ i =6,7,8,9 & j = 1,2,3,4	$p_{5,1} = s_4/(r_1+r_3+r_2+r_4+s_4)$ $p_{5,i} = r_j/(r_1+r_3+r_2+r_4+s_4)$ i =6,7,8,9 & j = 1,2,3,4
$q_{6,5} = s_1 e^{-s_1 t}$	$p_{6,5} = 1$
$q_{7,5} = s_2 e^{-s_2 t}$	$p_{7,5} = 1$
$q_{8,5} = s_3 e^{-s_3 t}$	$p_{8,5} = 1$

$q_{9,5}(t) = s_4 e^{-(r_1+r_3+r_2+r_4+s)t}$	$p_{9,5}= s_4/(r_1+r_4+r_3+r_2+s_4)$
$q_{9,i}(t) = r_j e^{-(r_1+r_2+r+r_4+s_4)t}$	$p_{9,i}= r_j/(r_1+r_3+r_2+r_4+s_4)$
i =10,11,12,13 & j = 1,2,3,4	i =10,11,12,13 & j = 1,2,3,4
$q_{10,9}= s_1 e^{-s_1 t}$	$p_{10,9}= 1$
$q_{11,9}= s_2 e^{-s_2 t}$	$p_{11,9}= 1$
$q_{12,9}= s_3 e^{-s_3 t}$	$p_{12,9}= 1$
$q_{13,9}= s_4 e^{-s_4 t}$	$p_{13,9}= 1$

Table 2: *Mean Sojourn Time (MST)*

$R_i(t)$	$\mu_i=R_i^*(0)$
$R_1(t)= e^{-(r_1+r_3+r_2+r_4)t}$	$\mu_1 = 1/(r_1+r_3+r_2+r_4)$
$R_2(t)= e^{-s_1 t}$	$\mu_2= 1/s_1$
$R_3(t)= e^{-s_2 t}$	$\mu_3= 1/s_2$
$R_4(t)= e^{-s_3 t}$	$\mu_4= 1/s_3$
$R_5(t)= e^{-(r_1+r_3+r_2+r_4+s_4)t}$	$\mu_5= 1/(r_1+r_3+r_2+r_4+s_4)$
$R_6(t)= e^{-s_1 t}$	$\mu_6= 1/s_1$
$R_7(t)= e^{-s_2 t}$	$\mu_7= 1/s_2$
$R_8(t)= e^{-s_3 t}$	$\mu_8= 1/s_3$
$R_9(t)= e^{-(r_1+r_3+r_2+r_4+s_4)t}$	$\mu_9= 1/(r_1+r_3+r_2+r_4+s_4)$
$R_{10}(t)= e^{-s_1 t}$	$\mu_{10}= 1/s_1$
$R_{11}(t)= e^{-s_2 t}$	$\mu_{11}= 1/s_2$
$R_{12}(t)= e^{-s_3 t}$	$\mu_{12}= 1/s_3$
$R_{13}(t)= e^{-s_4 t}$	$\mu_{13}= 1/s_4$

III. Evaluation of Path Probabilities

Implementing the RPGT technique and considering 'S_1' as the starting state of the framework. Path Probabilities from state 'S_1' to various vertices are stated below:

$V_{1,1} = 1$ (1)

$V_{1,i} = (1,i) = p_{1,i;}$ where i = 2,3,4 (2)

$V_{1,5}= p_{1,5}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})\{(1-p_{5,9}p_{9,5})/(1-p_{9,10}p_{10,9})(1-p_{9,11}p_{11,9})(1-p_{9,12}p_{12,9})(1-p_{9,13}p_{13,9})\}$ (3)

$V_{1,i}= p_{1,5}p_{5,i}/(1-p_{5,6}p_{5,6})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})\{(1-p_{5,9}p_{9,5})/(1-p_{9,10}p_{10,9})(1-p_{9,11}p_{11,9})(1-p_{9,12}p_{12,9})(1-p_{9,13}p_{13,9})\};$
i = 6, 7, 8 (4)

$V_{1,9}= p_{1,5}p_{5,9}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})(1-p_{9,10}p_{10,9})(1-p_{9,11}p_{11,9})(1-p_{9,12}p_{12,9})$
$(1-p_{9,13}p_{13,9})\{(1-p_{5,9}p_{9,5})/(1-p_{9,10}p_{10,9})(1-p_{9,11}p_{11,9})(1-p_{9,12}p_{12,9})(1-p_{9,13}p_{13,9})\}$ (5)

$V_{1,i} = p_{1,5}p_{5,9}p_{9,i}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})(1-p_{9,10}p_{10,9})(1-p_{9,11}p_{11,9})(1-p_{9,12}p_{12,9})$

$(1-p_{9,13}p_{13,9})\{(1-p_{5,9}p_{9,5})/(1-p_{9,10}p_{10,9})(1-p_{9,11}p_{11,9})(1-p_{9,12}p_{12,9})(1-p_{9,13}p_{13,9})\}$; where i = 10,11,12,13 (6)

Path Probabilities from state S9′ to various vertices are stated below:

$V_{9,1}= p_{9,5}p_{5,1}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})(1-p_{1,2}p_{2,1})(1-p_{1,3}p_{3,1})(1-p_{1,4}p_{4,1})$
$\{(1-p_{5,1}p_{1,5})/(1-p_{1,2}p_{2,1})(1-p_{1,3}p_{3,1})(1-p_{1,4}p_{4,1})\}$ (7)

$V_{9,i} = p_{9,5}p_{5,1}p_{1,i}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})(1-p_{1,2}p_{2,1})(1-p_{1,3}p_{3,1})(1-p_{1,4}p_{4,1})$
$\{(1-p_{5,1}p_{1,5})/(1-p_{1,2}p_{2,1})(1-p_{1,3}p_{3,1})(1-p_{1,4}p_{4,1})\}$; where i = 2, 3, 4 (8)

$V_{9,5} = p_{9,5}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})\{(1-p_{5,1}p_{1,5})/(1-p_{1,2}p_{2,1})(1-p_{1,3}p_{3,1})(1-p_{1,4}p_{4,1})\}$ (9)

$V_{9,i} = p_{9,5}p_{5,i}/(1-p_{5,6}p_{6,5})(1-p_{5,7}p_{7,5})(1-p_{5,8}p_{8,5})\{(1-p_{5,1}p_{1,5})/(1-p_{1,2}p_{2,1})(1-p_{1,3}p_{3,1})(1-p_{1,4}p_{4,1})\}$; where i = 6,7,8 (10)

$V_{9,9} = 1$ (11)

$V_{9,i} = p_{9,i}$; where i = 10,11, 12, 13 (12)

IV. Evaluation of System Parameters

The MTSF and other parameters are evaluated under steady-state conditions by using S_1 as the base state.

- Mean Time to System Failure (T_0): Regenerative working states to which the framework can transit (primary state 'S_1'), before arriving any failed state are 'i' = 1, 5, 9.
  $T_0 = (V_{1,1}\mu_1+V_{1,5}\mu_5+V_{1,9}\mu_9)/\{1-V(1,5,1)\}(1-p_{1,5}p_{5,1})$ (13)
- Availability of the System (A_0): Regenerative state at which framework is accessible are 'j' = 1, 5, 9, ; 'i' = 1 to 13.
  $A_0 = \left[\sum_j V_{\xi,j}, f_j, \mu_j\right]/\left[\sum_i V_{\xi,i}, f_j, \mu_i^1\right]$ (14)
  $A_0 = (V_{9,1}\mu_1+V_{9,5}\mu_5+V_{9,9}\mu_9)/D$ (15)
  Where $D = V_{1,i}\mu_i$, $\xi = 0$; $1 \le i \le 13$
- Busy Period of the Server (B_0): Regenerative positions where server is busy are j = 2 to 13; 'i' = 1 to 13. Considering $\xi = 0$
  $B_0 = \left[\sum_j V_{\xi,j}, n_j\right]/\left[\sum_i V_{\xi,i}, \mu_i^1\right]$ (16)
  $B_0 = (V_{1,j}\mu_j)/D$; $2 \le j \le 13$. (17)
- Expected Number of Inspections by the Repairman (V_0): Regenerative positions where the technician visit is j = 2 to 13; i = 0 to 13. Considering $\xi = 0$
  $V_0 = \left[\sum_j V_{\xi,j}\right]/\left[\sum_i V_{\xi,i}, \mu_i^1\right]$ (18)
  $V_0 = (V_{1,j})/D$; $2 \le j \le 13$. (19)

V. Results and Discussions

Particular Cases:- $s_i = s$ $(0 \le i \le 4)$, $r_i = r$ $(0 \le i \le 4)$

I. Mean Time to System Failure (MTSF) (T_0)

Table 3 shows the values of T_0 for varying repair/failure rates. Figure 3 displays the increasing decreasing trend of T_0 for varying repair/failure rates.

Table 3: *Mean Time to System Failure (MTSF)*

Repair rates/ Failure rates	s = .50	s = .60	s = .70
r = .10	2.86	2.80	2.79
r = .20	1.66	1.53	1.47
r = .30	0.52	0.47	0.42

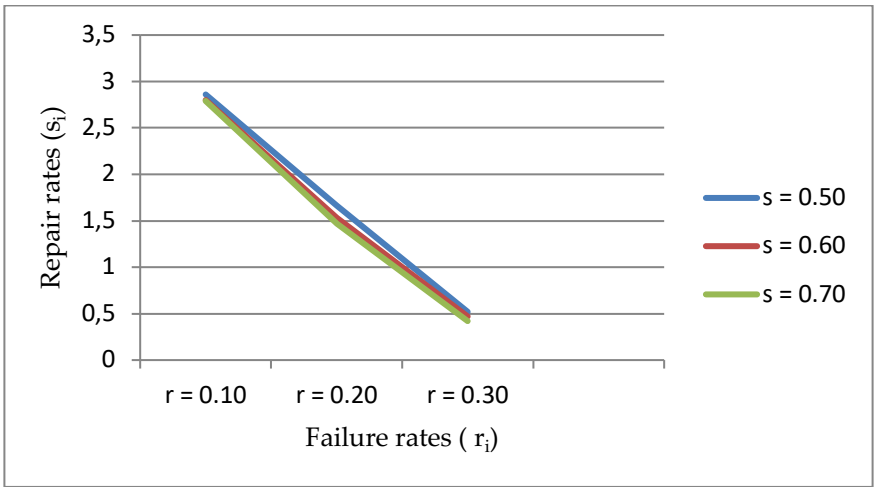


Figure 3: *Mean Time to System Failure (MTSF)*

II. Availability of the System (A_0)

Table 4 presents the values of A_0 for varying repair/failure rates. Figure 4 displays the increasing decreasing trend of A_0 for changing repair/failure rates.

Table 4: *Availability of the System (A_0)*

Repair rates/ Failure rates	s = .50	s = .60	s = .70
r = .10	.66	.70	.73
r = .20	.48	.51	.56
r = .30	.31	.40	.51

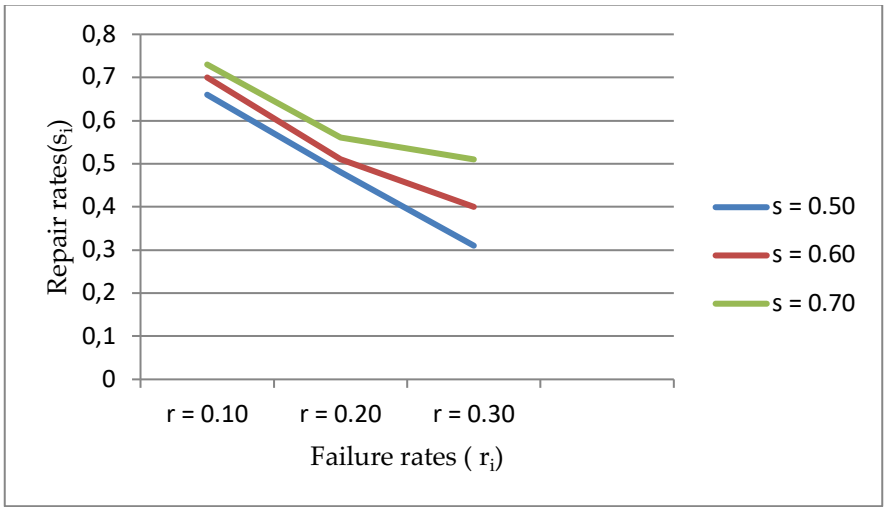


Figure 4: *Availability of the System (A_0)*

III. Busy Period of the Server (BPS) (B_0)

Table 5 shows the values of B_0 for varying repair/failure rates. Figure 5 displays the increasing decreasing trend of B_0 for varying repair/failure rates.

Table 5: *Busy Period of the Server (BPS)*

Repair rates/ Failure rates	s = .50	s = .60	s = .70
r = .10	.33	.28	.24
r = .20	.53	.47	.41
r = .30	.79	.63	.55

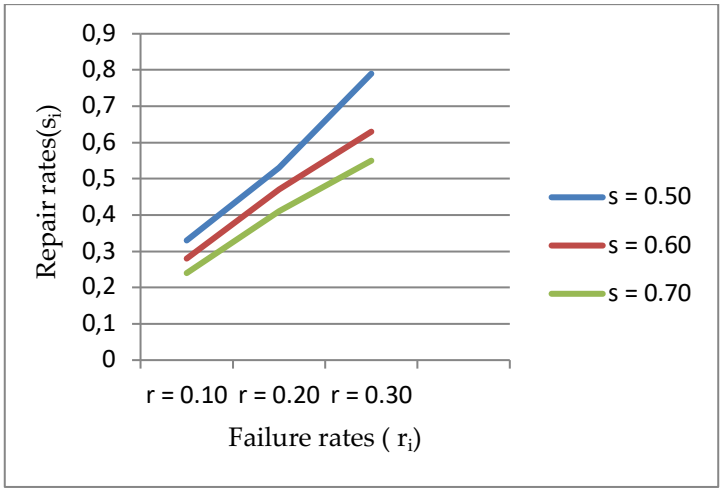


Figure 5: *Busy Period of the Server (BPS)*

IV. Expected Number of Inspection by the Repairman (ENIR) (V_0)

Table 6 shows the values of V_0 for varying repair/failure rates. Figure 6 displays the increasing decreasing trend of V_0 for changing repair/failure rates.

Table 6: *Expected Number of Inspection by Repairman*

Repair rates/ Failure rates	s = .50	s = .60	s = .70
r = .10	.12	.16	.19
r = .20	.14	.19	.23
r = .30	.21	.28	.33

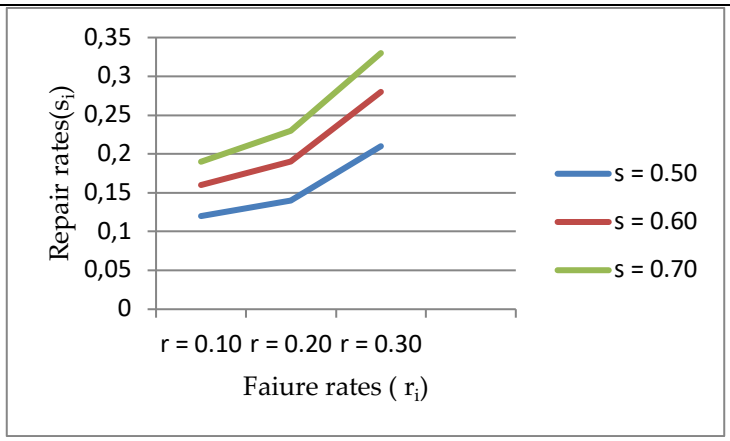


Figure 6: *Expected Number of Inspection by Repairman*

V. Profit Function

Profit analysis of the framework is calculated by applying the profit function given below

$P_0 = D_1A_0 – D_2B_0 - D_3V_0$ (20)

Assuming $D_1 = 2000$, $D_2 = 50$, $D_3 = 100$

Table 7 represents the values of profit function for varying repair/failure rates. Figure 7 shows the increasing decreasing trend of the profit function for varying repair/failure rates.

Table 7: *Profit Function*

Repair rates/ Failure rates	s = .50	s = .60	s = .70
r = .10	1291.5	1370.0	1429.0
r = .20	919.5	977.5	1076.5
r = .30	559.5	740.5	959.5

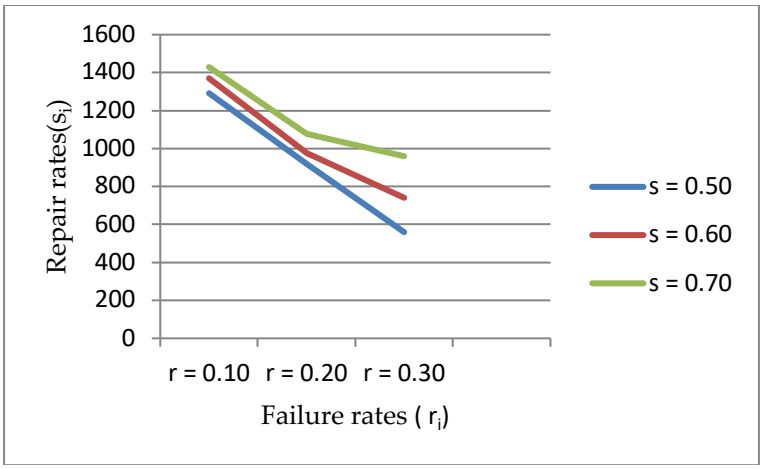


Figure 7: *Profit Function*

VI. Conclusion

Reliability, Availability and Maintainability (RAM) analysis of WTP's becomes an essential aspect for making the system more efficient and productive. The above calculations and graphs conclude that the availability of the system and the profit function reduces with the rise in failure rate and

increases with the rise in repair rate. It is also observed that the expected no. of inspections by the repairman increases with the rise in failure rate while BSP and MTSF reduce with the rise in repair rates. Thus the effectiveness and the reliability of the plant can be improved by increasing the repair rate and decreasing the failure rate.

References

[1] Asi, J. J., Seghier, M. E. A. B., Ohadi, S., Dong, Y. andPlevris, V. (2021). A Comparative Study on the Efficiency of Reliability Methods for the Probabilistic Analysis of Local Scour at a Bridge Pier in Clay-Sand-Mixed Sediments. *MDPI*, 2:63-77.

[2] Garg, D., Kumar, K. and Pahuja, G. L. (2010). Redundancy-Allocation in Pharmaceutical Plant. International Journal of Engineering Science and Technology, 2(5):1088-1097.

[3] Garg, D., Kumar, K. and Singh, J. (2010). Availability Analysis of a Cattle Feed Plant Using Matrix Method. International Journal of Engineering, 3(2):201-219.

[4] Garg, D., Singh, J. and Kumar, K. (2009). Performance Analysis of Screw Plant Using Matlab Tool. International Journal of Industrial Engineering Practice, 1(2):155-159.

[5] Garg, S., Singh, J. and Singh, D. V. (2010). Availability Analysis of Crank-Case Manufacturing in a Two-Wheeler Automobile Industry. Applied Mathematical Modeling, 34(6):1672-1683.

[6] Garg, S., Singh, J. and Singh, D. V. (2010). Availability and Maintenance Scheduling of a Repairable Block-Board Manufacturing System. International journal of reliability and safety, 4(1): 104-118.

[7] Kumar, A., Garg, D. and Goel, P. (2019). Mathematical Modeling and Behavioral Analysis of a Washing Unit in Paper Mill. International Journal of System Assurance Engineering and Management, 10(6):1639-1645.

[8] Kumar, A., Garg, D. and Goel, P. (2017). Mathematical Modeling and Profit Analysis of an Edible Oil Refinery Industry.Airo International Research Journal, 13:1-14.

[9] Kumar, A., Garg, D. and Goel, P. (2019). Sensitivity Analysis of a Cold Standby System with Priority for Preventive Maintenance.Journal of Advances and Scholarly Researches in Allied Education, 16:253-258.

[10] Kumar, A., Garg, D. and Goel, P. (2018). Behaviour Analysis of a Bread Making System. International Journal of Statistics and Applied Mathematics, 3(6):56-61

[11] Li, J., Chen, J. and Chen, Z. (2020). Developing an Improved Composite Limit State Method for Time-Dependent Reliability Analysis. Quality Engineering, 32(3):298-311.

[12] Rajbala, and Garg, D. (2019). Behaviour Analysis of Alloy Wheel Plant. International Journal of Engineering and Advanced Technology (IJEAT), 9(2):319-327.

[13] Wang, Z., Huang, H. Z., Li, Y., Pang, Y. and Xiao, N. C. (2012). An Approach to System Reliability Analysis with Fuzzy Random Variables. Mechanism and Machine Theory, 52:35-46.

[14] Zhai, Q., Xing, L., Peng, R. and Yang, J. (2015). Multi-Valued Decision Diagram-Based Reliability Analysis of $ k $-out-of-$ n $ Cold Standby Systems Subject to Scheduled Backups. IEEE Transactions on Reliability, 64(4):1310-1324.

Reliability Test Plan For The Marshall-Olkin Extended Inverted Kumaraswamy Distribution

Jiju Gillariose[1], Ehab M. Almetwally[2],Joshin Joseph[3],Vishna Devi[4]

[1]*Department of Statistics, CHRIST (Deemed to be University), Hosur Road, Bangalore, Karnataka– 560029, India*
Email: jijugillariose@yahoo.com
[2]*Department of Statistics, Delta University of Science and Technology, Belkas, Egypt Email:ehabxp2009@hotmail.com*
[3]*Department of Finance, KMML a PSU Under Government of Kerala, Sankaramangalam, Kollam-691583,India*
Email:joshinpariyath@gmail.com
[4]*Department of Mathematics, KE College, Mananam, Kerala-686561, India, Email: vishnadevid@gmail.com*

Abstract

This paper mainly interested in studying the wider range behavior of the Marshall-Olkin extended inverted Kumaraswamy distribution. The parameters of model are estimated by various estimation methods. A reliability sampling plan is proposed which can save the test time in practical situations. Some tables are also provided for the new sampling plans so that this method can be used conveniently by practitioners. The developed test plan is applied to ordered failure times of software release to provide its importance in industrial applications

Keywords: Reliability Test Plan, Kumaraswamy Distribution, Marshall-Olkin Family, Method of Maximum Likelihood, Method of Percentiles

I. Introduction

In statistical literature, there are numerous distributions but still remain many important problems where the real data does not follow any of the existing probability models. Because of this, significant strive has been taken in the development of generalizations of standard probability distributions along with relevant statistical methodologies. Kumaraswamy distribution introduced by Kumaraswamy (1980) is derived from beta distribution after fixing some parameters posses a closed-form cdf (cumulative density function) which is invertible. This distribution is applicable to many natural phenomena related to which outcomes have lower and upper bounds. The inverted Kumaraswamy model is the probability distribution of a random variable whose reciprocal has a Kumaraswamy distribution proposed by Abd AL-Fattah et al. (2017). Further, Iqbal et al. (2017) derived generalized form of inverted Kumaraswamy distribution by inserting another parameter to inverted Kumaraswamy distribution.

The method of addition of parameters has used to enhance the properties of existing family of distributions. This added new parameter improves the goodness-of-fit of the generated family. Parameters can be introduced by various methods, then we have new families such as exponentiated family of distributions (Gupta et al., 1998), transformed-

transformer (T-X) family of distributions (Alzaatreh, 2011), Kumaraswamy family distributions (Cordeiro and Castro, 2011), geometric exponential-Poisson family family of distributions (Nadarajah et al., 2013), etc. Many researchers used the Marshall-Olkin method introduced by Marshall-Olkin (1997) to propose new distributions and established their distinct properties and characteristics.

This paper mainly focus on different methods of estimation and the reliability test plan for the Marshall-Olkin extended inverted Kumaraswamy distribution. The paper is organized as follows: Section 2 deals with the basic concepts of Marshall-Olkin Extended (MOE) inverted Kumaraswamy distribution (Tomy and Gillariose, 2017). Different methods of estimation discussed in Section 3. The reliability test plan is conducted in Section 4. This work is concluded in Section 5.

.

II. Inverted Kumaraswamy Distribution

The cdf and probability density function (pdf) of MOE inverted Kumaraswamy (MOEIKum) distribution, respectively, are given by

$$G(x,\alpha,\beta,\gamma)=\frac{(1-(1+x)^{-\gamma})^{\beta}}{(\alpha+(1-\alpha)(1-(1+x)^{-\gamma})^{\beta})},\ x>0,\alpha,\beta,\gamma>0 \tag{2.1}$$

and

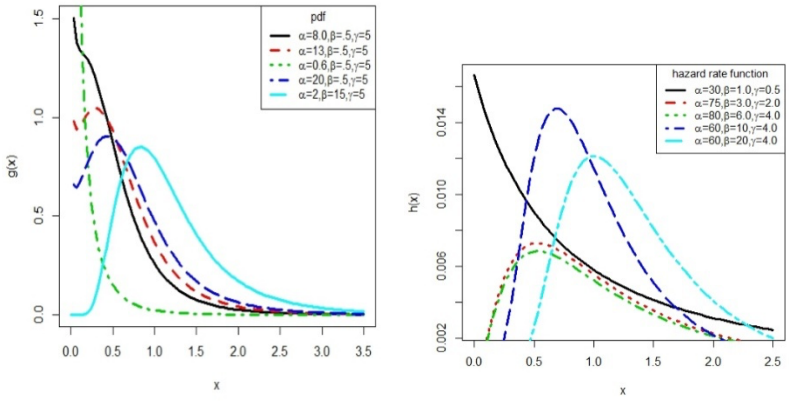
$$g(x,\alpha,\beta,\gamma)=\frac{\alpha\beta\gamma(1+x)^{-(\gamma+1)}(1-(1+x)^{-\gamma})^{\beta-1}}{[\alpha+(1-\alpha)(1-(1+x)^{-\gamma})^{\beta}]^{2}},\ x>0,\alpha,\beta,\gamma>0 \tag{2.2}$$

The hazard rate function of the MOEIKum distribution is given by the following equation

$$h(x,\alpha,\beta,\gamma)=\frac{\beta\gamma(1+x)^{-(\gamma+1)}(1-(1+x)^{-\gamma})^{\beta-1}}{[\alpha+(1-\alpha)(1-(1+x)^{-\gamma})^{\beta}][1-(1-(1+x)^{-\gamma})^{\beta}]},\ x>0,\alpha,\beta,\gamma>0$$

The different shapes of the pdf and hazard rate function of the MOEIKum distribution are displayed in Figure 1 for selected values of α, β and γ. From the figure we can see that hazard rate function accommodates increasing, decreasing, and unimodal shaped forms, that depend basically on the values of the shape parameters. This distribution can be expressed as a limiting case of some existing distributions and also from this distribution we can derive a number of sub-models for example, Lomax distribution, MOE Lomax distribution, log-logistic distribution etc. Consider the following theorem, which establish behavior of the MOEIKum distribution.

Figure 1: Graphs of pdf and hazard rate function of the MOEIKum distribution for different values of α, β and γ.



Theorem: 1*Let* $\{X_i,\ i \geq 1\}$ *be a sequence of i.i.d. random variables with common survival function* $\overline{F}(x)$*. Let N be a geometric random variable independently distributed of* $\{X_i,\ i \geq 1\}$ *such that* $P(N = n) = p(1-p)^{n-1}, n = 1,2,\ldots, 0 < p < 1$*. Let* $U_N = min_{1 \leq i \leq n} X_i$*. Then* $\{U_N\}$ *is distributed as MOEIKum*(p, β, γ) *iff* $\{X_i\}$ *follows IKum*(β, γ) .

Proof: The survival function of the random variable U_N is

$$\overline{W}(x) = P(U_N > x)$$
$$= \sum_{n=1}^{\infty} P(U_n > x)P(N = n)$$
$$= \sum_{n=1}^{\infty} [\overline{F}(x)]^n p(1-p)^{n-1}$$
$$= \frac{p\overline{F}(x)}{1-(1-p)\overline{F}(x)}$$

$$W(x) = \frac{(1-(1+x)^{-\gamma})^{\beta}}{(p+(1-p)(1-(1+x)^{-\gamma})^{\beta})}$$

which is cdf of a random variable with MOEIKum(p, β, γ) distribution.

Remark: 1 *Let* $\{X_i,\ i \geq 1\}$ *be a sequence of i.i.d. random variables with common survival function* $\overline{F}(x)$*. Let N be a geometric random variable independently distributed of* $\{X_i,\ i \geq 1\}$ *such that* $P(N = n) = p(1-p)^{n-1}, n = 1,2,\ldots, 0 < p < 1$*. Let* $V_N = max_{1 \leq i \leq n} X_i$*. Then* $\{V_N\}$ *is distributed as MOEIKum*$(\frac{1}{p}, \beta, \gamma)$ *iff* $\{X_i\}$ *follows IKum*(β, γ) *distribution.*

III. Estimation

This section describes different estimation methods for estimating the parameters α, β, and γ of the MOEIKum distribution.

I. Method of Maximum Likelihood

Let $X_1, X_2, \ldots, X_n$ is a random sample of size n from MOEIKum(α, β, γ), then the log likelihood function is given by

$$logL(\alpha, \beta, \gamma) = nlog(\alpha\beta\gamma) - (\gamma+1)\sum_{i=1}^{n} log(x_i) + (\beta-1)\sum_{i=1}^{n} log(1-(1+x_i)^{-\gamma})$$
$$-2\sum_{i=1}^{n} log(1-\overline{\alpha}[1-[1-(1+x_i^{-\gamma})]^{\beta}])$$

The partial derivative of the log likelihood functions with respect to the parameters are

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} - 2\sum_{i=1}^{n} \frac{1-[1-(1+x_i^{-\gamma})]^{\beta}}{1-\overline{\alpha}[1-[1-(1+x_i^{-\gamma})]^{\beta}]}$$

$$\frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^{n} log(1-(x_i+1)^{-\gamma}) - 2\overline{\alpha}\sum_{i=1}^{n} \frac{[1-(1+x_i^{-\gamma})]^{\beta} log(1-[1+x_i^{-\gamma}])}{1-\overline{\alpha}[1-[1-(1+x_i^{-\gamma})]^{\beta}]}$$

$$\frac{\partial \log L}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^{n} log(x_i) + (\beta-1)\sum_{i=1}^{n} \frac{[1-(1+x_i^{-\gamma})] log(1-(1+x_i^{-\gamma}))}{[1-(1+x_i^{-\gamma})]}$$
$$-2\sum_{i=1}^{n} \frac{[1-(1+x_i^{-\gamma})] log(1-(1+x_i^{-\gamma}))}{1-\overline{\alpha}[1-[1-(1+x_i^{-\gamma})]^{\beta}]}$$

The maximum likelihood estimates can be numerically obtained by solving the equations $\frac{\partial \log L}{\partial \alpha} = 0, \frac{\partial \log L}{\partial \beta} = 0, \frac{\partial \log L}{\partial \gamma} = 0.$

II. Methods of Ordinary and Weighted Least-Squares

A regression based method estimators of the unknown parameters suggested by Swain et al. (1988) to estimate the parameters of beta distributions. Let $X_1, X_2, \ldots, X_n$ is a random sample of size n from a distribution function with cdf G(x) and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics of the observed sample. It is well-known that $G(X_{(i)})$ behaves like the i^{th} order statistics of a sample of size n from U(0,1), therefore we have

$$E\big(G(X_{(j)})\big) = \frac{j}{n+1},\ Var\big(G(X_{(j)})\big) = \frac{j(n-j+1)}{(n+1)^2(n+2)},\ for i < j \qquad (3.1)$$

Therefore, from equation (3.1) the least squares estimators of the unknown parameters α, β, and γ of MOEIKum(α, β, γ), can be obtained by minimizing

$$\sum_{i=1}^{n} \left(G(X_{(i)}) - \frac{j}{n+1}\right)^2 = \sum_{i=1}^{n} \left(\frac{(1-(1+X_{(i)})^{-\gamma})^{\beta}}{(\alpha+(1-\alpha)(1-(1+X_{(i)})^{-\gamma})^{\beta})} - \frac{j}{n+1}\right)^2$$

with respect to α, β, and γ.

By equation (3.1), the weighted least squares estimators of the unknown parameters of
MOEIKum(α, β, γ) can be obtained by minimizing

$$\sum_{i=1}^{n} \frac{1}{Var\big(G(X_{(i)})\big)}\left(G(X_{(i)}) - \frac{j}{n+1}\right)^2 = \sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{j(n-j+1)}$$
$$\left(\frac{(1-(1+X_{(i)})^{-\gamma})^{\beta}}{(\alpha+(1-\alpha)(1-(1+X_{(i)})^{-\gamma})^{\beta})} - \frac{j}{n+1}\right)^2$$

with respect to α, β, and γ.

$$x = \left\{1 + \left[F(x;\alpha,\beta,\gamma)\alpha - \frac{\alpha}{1+\alpha}\right]^{\frac{1}{\beta}}\right\}^{-\frac{1}{\gamma}} - 1.$$

Let $X_{(i)}$ denoted as the i^{th} order statistic $X_{(1)} < X_{(2)} <\ldots< X_{(n)}$. If $p_i = \frac{i}{n+1}$ denotes some estimate of $F(x;\alpha,\beta,\gamma)$, then the estimates of α, β, and γ can be obtained by minimizing

$$\sum_{i=1}^{n}\left[X_{(i)} - \left\{1+\left[F(x;\alpha,\beta,\gamma)\alpha - \frac{\alpha}{1+\alpha}\right]^{\frac{1}{\beta}}\right\}^{-\frac{1}{\gamma}} + 1\right]^2$$

with respect to α, β, and γ. These estimates of α, β, and γ were obtained by using R in method of maximum likelihood, methods of ordinary and weighted least-squares and Method of Percentiles.

III. Acceptance Sampling Plans

Reliability sampling plans are used for determining the acceptability of any product. In this section, we develop reliability test plan with the life time governed by an IMOEKu distribution with cdf

$$G(x,\alpha,\beta,\gamma) = \frac{(1-(1+x)^{-\gamma})^{\beta}}{(\alpha+(1-\alpha)(1-(1+x)^{-\gamma})^{\beta})}, x>0, \alpha,\beta,\gamma>0 \quad (4.1)$$

If a scale parameter $\theta > 0$ is introduced, the distribution function of IMOEKu is given by

$$G(x,\alpha,\beta,\gamma,\theta) = \frac{(1-(1+x/\theta)^{-\gamma})^{\beta}}{(\alpha+(1-\alpha)(1-(1+x/\theta)^{-\gamma})^{\beta})}, x>0, \alpha,\beta,\gamma>0 \quad (4.2)$$

A common practice in life testing is to terminate the life test by a pre-determined time *t* and note the number of failures (assuming that a failure is well-defined). One of the objectives of these experiments is to set a lower confidence limit on the average life. It is then desired to establish a specified

average life with a given probability of at least p^*. The decision to accept the specified average life occurs if and only if the number of observed failures at the end of the fixed time *t* does not exceed a given number c called the acceptance number. The test may get terminated before the time *t* is reached when the number of failures exceeds c in which case the decision is to reject the lot. For such a truncated life test and the associated decision rule, we are interested in obtaining the smallest sample sizes necessary to achieve the objective. Here, it is assume that α, β and γ are known while θ is unknown. So, average life time depends only on θ. A sampling plan consists of

- the number of units n on test,
- the acceptance number c,
- the maximum test duration t, and
- the ratio $\frac{t}{\theta_0}$ where θ_0 is the specified average life.

The consumer's risk, i.e., the probability of accepting a bad lot (the one for which the true average life is below the specified life θ_0) not to exceed $1-p^*$, so that p^* is a minimum confidence level with which a lot of true average life below θ_0 is rejected, by the

sampling plan. For a fixed p^* our sampling plan is characterized by (n, c, $\frac{t}{\theta_0}$). Here, it is consider sufficiently large lots so that the binomial distribution can be applied. The problem is to determine for given values of p^*, $(0 < p^* < 1)$, θ_0 and c the smallest positive integer n such that

$$\sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \leq 1 - p^* \tag{4.3}$$

holds where $p = G(x, \alpha, \beta, \gamma, \theta_0)$ is given by (5.2) indicates the failure probabilities before time t which depends only on the ratio t/θ_0 it is sufficient to specify this ratio for designing the experiment. If the number of observed failures before t is less than or equal to c, from (5.3), we have:

$$G(t, \theta) \leq G(t, \theta_0) \Leftrightarrow \theta \geq \theta_0 \tag{4.4}$$

The minimum values of n satisfying the inequality (5.4) are obtained and displayed in Table1 for p^*=0.75, 0.90, 0.95 and t =1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 3.0, 3.5, 4.0 and $\alpha = \beta = \gamma = 2$. If $p = G(x, \alpha, \beta, \gamma, \theta_0)$ is small and n is large (as is true in some cases of our present work), the binomial probability may be approximated by Poisson probability with parameter $\lambda = np$ so that the left side of (4.3) can be written as

$$\sum_{i=0}^{c} \frac{e^{-\lambda}\lambda^x}{x} \leq 1 - p^* \tag{4.5}$$

where $p = G(x, \alpha, \beta, \gamma, \theta_0)$. The minimum values of n satisfying (4.5) are obtained for the same combination of p values as those used for (4.3). The results are given in Table 2. The operating characteristic function of the sampling plan (n, c, t/θ_0) gives the probability L(p) of accepting the lot with:

$$L(p) = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \tag{4.6}$$

where $p = G(x, \alpha, \beta, \gamma, \theta)$ is considered as a function of θ, i.e., the lot quality parameter. It can be seen that the operating characteristic is an increasing function of θ. For given p^*, t/θ_0 the choice of c and n is made on the basis of operating characteristics. Values of the operating characteristics as a function of θ/θ_0 for a few sampling plans are given in Table 3.

The producer's risk is the probability of rejecting lot when $\theta > \theta_0$. We can compute the producer's risk by first finding $p = F(t, \theta)$ and then using the binomial distribution function. For a given value of the producer's risk say 0.05, one may be interested in knowing what value of θ/θ_0 will ensure a producer's risk less than or equal to 0.05 if a sampling plan under discussion is adopted. It should be noted that the probability p may be obtained as function ofθ/θ_0, as

$$p = F(\frac{t}{\theta_0}\frac{\theta_0}{\theta}) \tag{4.7}$$

The value θ/θ_0 is the smallest positive number for which the following inequality hold:

$$\sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i} \geq .95 \tag{4.8}$$

For a given sampling plan (n, c, t/θ_0) and specified confidence level p^*. the minimum values of θ/θ_0 satisfying the inequality (4.8) are given in Table 4.

Example: Consider the following ordered failure times of the release of a software given in terms of hours from the starting of the execution of the software denoting the times at which the failure of the software is experienced (Wood, 1996). This data can be regarded as an ordered sample of size 10 with observations
(xi , i = 1, . . . , 10) = 519, 968, 1430, 1893, 2490, 3058, 3625, 4422, 5218, 5823

Let the specified average life be 1000 hrs and the testing time be 1250 hrs, this leads to ratio of $t/\theta = 1.25$ with corresponding n and c as 10, 2 from Table 4.1 for $p^* = 0.9$. Therefore, the sampling plan for the above sample data is (n=10, c=2, $t/\theta_0 = 1.25$). Based on the observations, we have to decide whether to accept the product or reject it. We accept the product only, if the number of failures after 1250 hrs is less than or equal to 2. However the confidence level is assured by the sampling plan only if the given life times follow an MOEIKum distribution. In order to confirm that the given sample is generated by lifetimes following at least approximately the inverse Raleigh distribution, we have compared the sample quantiles and the corresponding population quantiles and found a satisfactory agreement. Thus, the adoption of the decision rule of the sampling plan seems to be justified. We see that in the sample of 10 failures there are 2 failures at 519 and 968 hrs before 1250 hrs. Therefore we accept the product.

Table 1: Minimum sample sizes necessary to assert the average life to exceed a given value t/θ_0 with probability p^* and the corresponding acceptance number c, $\alpha = \beta = \gamma = 2$ using Binomial probabilities.

					t/θ_0						
p^*	c	1	1.25	1.5	1.75	2	2.25	2.5	3	3.5	4
	0	3	3	2	2	2	2	2	1	1	1
	1	6	5	4	4	4	3	3	3	3	3
	2	9	8	7	6	5	5	5	4	4	4
	3	12	10	9	8	7	7	6	6	5	5
	4	15	12	11	10	9	8	8	7	7	6
.75	5	18	15	13	11	11	10	9	8	8	8
	6	21	17	15	13	13	11	11	10	9	9
	7	24	19	17	15	15	13	12	11	11	10
	8	27	22	19	17	16	14	14	12	12	11
	9	29	24	21	19	17	16	15	14	13	12
	10	32	26	23	20	19	17	16	16	14	14
	0	5	4	3	3	3	2	2	2	2	1
	1	9	7	6	5	5	4	4	4	3	2

	2	12	10	8	7	7	6	5	5	5	3
	3	16	13	11	9	9	7	7	7	6	5
	4	19	15	13	11	10	8	9	9	8	6
.90	5	22	18	15	13	12	11	11	10	9	7
	6	25	20	17	15	14	13	12	11	10	8
	7	28	23	19	17	16	15	14	12	12	9
	8	31	25	21	19	17	16	15	14	13	10
	9	34	27	24	21	19	18	17	15	14	11
	10	37	30	26	23	21	19	18	17	16	12
	0	7	5	4	4	3	3	3	2	2	1
	1	11	8	7	6	4	5	5	4	4	3
	2	15	11	10	8	8	7	6	6	5	4
	3	18	13	12	11	10	9	8	7	7	5
	4	21	17	14	13	11	11	10	9	8	6
.95	5	25	20	17	15	13	12	11	10	10	7
	6	28	22	19	17	15	14	13	12	11	8
	7	31	25	21	19	17	16	15	13	12	9
	8	34	27	24	21	19	17	16	15	14	10
	9	37	30	26	23	21	19	18	16	15	11
	10	40	32	28	24	22	21	19	18	16	12

Table 2: Minimum sample sizes necessary to assert the average life to exceed a given value t/θ_0 with probability p^* and the corresponding acceptance number c, α=β=γ=2 using Poisson probabilities. probabilities.

					t/θ_0						
p*	C	1	1.25	1.5	1.75	2	2.25	2.5	3	3.5	4
	0	4	3	3	4	4	3	3	2	2	2
	1	7	6	5	7	6	5	5	4	4	4
	2	11	9	8	10	8	7	7	6	6	5
	3	14	11	10	12	11	9	8	8	7	7
	4	17	14	12	14	13	11	10	10	9	8
.75	5	19	16	14	17	15	13	12	11	10	10
	6	22	19	16	19	17	15	14	13	11	11
	7	25	21	18	21	19	17	15	14	13	12
	8	28	23	21	24	21	18	17	16	14	14
	9	31	26	22	26	23	20	19	18	16	15
	10	34	28	24	28	24	22	20	19	18	16
	0	6	5	5	4	4	4	4	3	3	3
	1	10	9	8	7	6	6	6	5	5	5
	2	14	12	10	9	9	8	8	7	7	7
	3	18	15	13	12	11	10	10	9	9	8
	4	21	17	15	14	13	12	12	11	10	10
.90	5	24	20	18	16	15	14	14	12	12	11
	6	28	23	20	18	17	16	16	14	13	13
	7	31	25	22	20	19	17	17	16	15	14
	8	34	28	24	22	21	19	19	17	16	16
	9	37	30	27	24	22	21	21	19	18	17
	10	40	33	29	26	24	23	22	20	19	19

	0	8	7	6	5	5	5	5	4	4	4
	1	13	10	9	8	8	7	6	7	6	5
	2	17	14	12	11	10	10	8	9	8	7
	3	20	17	15	13	12	12	10	10	10	8
	4	24	20	17	16	15	13	11	12	12	10
.95	5	27	23	20	18	17	16	13	14	13	11
	6	31	26	22	20	19	18	14	16	15	13
	7	34	29	25	23	21	19	16	17	17	14
	8	38	31	27	24	23	21	17	19	18	15
	9	41	34	29	27	25	23	19	21	20	17
	10	44	34	32	29	26	25	20	22	21	18

Table 3: Operating characteristic values of the sampling plan (n, c, t/θ_0 for given p^* and $\alpha = \beta = \gamma = 2$ under MOEIKum probabilities.

						$\boldsymbol{\theta/\theta_0}$			
$\boldsymbol{p^*}$	N	c	t/θ_0	2	4	6	8	10	12
	9	2	1	0.7833	0.97365	0.9965	0.9991	0.9997	0.9998
	8	2	1.25	0.7053	0.9653	0.9934	0.9982	0.9994	0.9997
	7	2	1.5	0.6622	0.9538	0.9905	0.9973	0.999	0.9996
	6	2	1.75	0.6589	0.9496	0.9891	0.9968	0.9988	0.9995
.75	5	2	2	0.6974	0.9546	0.9899	0.997	0.9989	0.9995
	5	2	2.25	0.6207	0.9332	0.984	0.995	0.99813	0.9991
	5	2	2.5	0.547	0.907	0.9765	0.9992	0.997	0.9988
	4	2	3	0.6169	0.9212	0.979	0.9929	0.9971	0.9987
	4	2	3.5	0.518	0.879	0.9642	0.9871	0.9946	0.9975
	4	2	4	0.4315	0.8306	0.9448	0.979	0.9909	0.9956
	12	2	1	0.6208	0.9543	0.9916	0.9978	0.9992	0.9997
	10	2	1.25	0.5583	0.9357	0.987	0.9964	0.9987	0.9995
	8	2	1.5	0.5688	0.9329	0.9857	0.9959	0.9985	0.9994
	7	2	1.75	0.5458	0.9217	0.9822	0.9947	0.998	0.9992
.9	7	2	2	0.4389	0.8807	0.9701	0.9905	0.9964	0.9984
	6	2	2.25	0.4723	0.887	0.9709	0.9906	0.9964	0.9984
	5	2	2.5	0.547	0.9067	0.9762	0.9922	0.997	0.9986
	5	2	3	0.4158	0.8456	0.9546	0.981	0.9934	0.997
	5	2	3.5	0.3107	0.7739	0.9251	0.9715	0.9877	0.9941
	3	2	4	0.7214	0.994	0.9993	0.9939	0.9974	0.9988
	15	2	1	0.4664	0.91911	0.9841	0.9957	0.9985	0.9994
	11	2	1.25	0.4892	0.917	0.9828	0.9952	0.9983	0.9993
	10	2	1.5	0.4088	0.8809	0.9724	0.9918	0.997	0.9987
	8	2	1.75	0.4419	0.8886	0.9733	0.9918	0.997	0.9987
.95	8	2	2	0.3338	0.8341	0.9558	0.9857	0.9945	0.9976
	7	2	2.25	0.3465	0.8324	0.9587	0.9846	0.994	0.9973
	6	2	2.5	0.3919	0.8473	0.9574	0.9856	0.9943	0.99747
	6	2	3	0.2633	0.7533	0.9214	0.9709	0.9878	0.9943
	5	2	3.5	0.3107	0.7739	0.9251	0.9715	0.9877	0.9941
	4	2	4	0.4315	0.8303	0.9448	0.979	0.9909	0.9995

Table 4: Minimum ratio of true θ and required θ_0 for the acceptability of a lot with producer's risk of 0.05 for $\alpha = \beta = \gamma = 2$ under MOEIKum probabilities.

p^*	c	1	1.25	1.5	1.75	2	2.25	2.5	3	3.5	4
	0	9.83	12.29	11.45	13.36	15.26	17.18	19.08	15.45	18.025	20.025
	1	4.56	5.19	5.33	6.22	7.11	6.36	7.07	8.48	9.89	11.31
	2	3.56	4.18	4.71	4.52	4.6	5.17	5.74	5.88	6.86	7.84
	3	3.14	3.22	3.23	4.33	4.03	4.54	4.52	5.43	5.32	6.08
	4	2.69	2.87	3.32	3.53	3.72	3.88	4.31	4.64	5.42	5.46
.75	5	2.58	2.77	3.02	3.16	3.62	3.79	3.87	4.16	4.86	5.55
	6	2.47	2.6	2.86	3.02	3.03	3.41	3.79	4.23	4.42	5.05
	7	2.47	2.52	2.78	2.95	3.37	3.36	3.53	3.91	4.56	4.84
	8	2.29	2.45	2.65	2.82	3.09	3.29	3.53	3.79	4.42	4.59
	9	2.14	2.38	2.59	2.77	2.92	3.12	3.31	3.79	4.12	4.43
	10	2.14	2.32	2.52	2.66	2.93	3.02	3.2	3.79	3.97	4.54
	0	12.76	15.13	15.89	18.54	21.19	17.17	19.08	22.9	26.72	20.6
	1	5.81	6.36	6.84	7.27	8.31	8.01	8.89	10.68	9.9	8.07
	2	4.15	4.78	4.71	5.2	5.95	5.81	5.74	6.89	8.04	5.86
	3	3.56	3.93	4.24	4.33	4.95	4.54	5.04	6.05	6.33	6.08
	4	3.33	3.37	3.72	3.88	4.92	3.97	4.9	5.88	6.18	5.37
.9	5	2.97	3.22	3.32	3.53	3.81	3.97	4.41	4.94	5.31	4.91
	6	2.82	2.97	3.12	3.43	3.62	3.88	4.03	4.55	4.86	4.59
	7	2.58	2.77	3.02	3.16	3.53	3.79	4.03	4.16	4.86	4.32
	8	2.47	2.68	2.86	3.09	3.22	3.48	3.72	4.16	5.42	4.18
	9	2.38	2.6	2.86	3.02	3.16	3.48	3.65	3.96	4.42	4.05
	10	2.38	2.52	2.78	2.95	3.16	3.29	3.47	4.03	4.42	3.39
	0	15.5	16.9	17.31	20.2	19.66	22.12	24.58	22.9	26.72	20.6
	1	6.87	6.77	7.54	7.98	7.06	8.95	9.95	10.6	12.37	13.34
	2	3.85	3.92	4.46	4.94	5.16	6.69	6.45	7.74	8.04	7.43
	3	3.86	3.93	4.46	4.94	5.16	5.57	5.54	6.05	7.06	6.07
	4	3.37	3.71	3.87	4.33	4.32	4.86	5.04	5.57	6.04	5.37
.95	5	3.21	3.53	3.71	3.88	4.03	4.29	4.52	4.94	5.77	4.91
	6	2.92	3.16	3.38	3.62	3.86	4.07	4.26	4.833	5.23	4.59
	7	2.78	3.03	3.2	3.44	3.63	3.91	4.13	4.45	4.89	4.33
	8	2.63	2.82	3.12	3.31	3.52	3.69	3.87	4.45	4.89	4.18
	9	2.53	2.78	3.01	3.22	3.43	3.59	3.8	4.17	4.61	4.04
	10	2.46	2.68	2.89	3.03	3.22	3.52	3.63	4.17	4.42	3.93

IV. Conclusion

In this paper, a comprehensive description of properties of MOEIKum distribution are provided with the hope that it will attract wider applications in the area of research. Different methods for estimating unknown parameters of MOEIKum distribution are derived. Additionally, acceptance sampling plan is developed based on the truncated life test when the life distribution of the test items follows an MOEIKum distribution

References

[1] Abd AL-Fattah, A.M., EL-Helbawy, A.A. and AL-Dayian, G.R. (2017). IvertedKumaraswamy Distribution: Properties and Application. Pakistan Journal of Statistics, 33, 37–61.

[2] Alzaatreh, A. (2011). A New Method for Generating Families of Continuous Distributions. Ph.D. Thesis, Central Michigan University, Mount Pleasant, Michigan, USA. (Unpublished)

[3] Cordeiro, G.M, and Castro, M. (2011). A New Family of Generalized Distributions. Journal of Statistical Computation and Simulation, 81, 883–898.

[4] Gupta, R.C., Gupta, P.L. and Gupta, R.D., (1998). Modeling Failure Time Data by Lehman Alternatives. Communications in Statistics Theory and Methods, 27, 887–904.

[5] Iqbal Z., Tahir, M.M., Riaz, N., Ali, S.A., and Ahmad, M. (2017). Generalized Inverted Kumaraswamy Distribution: Properties and Application. Open Journal of Statistics, 7, 645–662.

[6] Kao, J.H.K. (1958). Computer methods for estimating Weibull parameters in reliability studies. Transactions of IRE-Reliability and Quality Control, 13, 15–22.

[7] Kao, J.H.K. (1959). A graphical estimation of mixed Weibull parameters in life testing electron tube. Technometrics, 1, 389–407.

[8] Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. Journal of Hydrology, 46, 79–88.

[9] Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrica, 84, 641–652.

[10] Nadarajah, S., Cancho, V.G., and Ortega E.M.M. (2013). The geometric exponential Poisson distribution. Statistical Methods and Applications, 22, 355–380.

[11] Swain, J., Venkatraman, S. and Wilson, J. (1988). Least Squares Estimation of Distribution Function in Johnson's Translation System, Journal of Statistical Computation and Simulation, 29, 271–297.

[12] Tomy L and Gillariose J (2017). The Marshall-Olkin IKum Distribution. Biometrics Biostatistics International Journal, 7, 1–6.

[13] Wood, A. (1996). Predicting software reliability. IEEE Transactions on Software Engineering, 22, 69–77.

Cluster Formation In An Acyclic Digraph Adding New Edges

Gurami Tsitsiashvili, Marina Osipova

Institute for Applied Mathematics,
Far Eastern Branch of Russian Academy Sciences
guram@iam.dvo.ru, mao1975@list.ru

Abstract

In this paper, we construct an algorithm for converting an acyclic digraph that defines the structure of a complex system into a class of cyclically equivalent vertices by adding several additional edges to the digraph. This addition of the digraph makes it possible to introduce negative feedbacks and, consequently, to stabilize the functioning of the complex system under consideration and so to increase its reliability.To do this, the original digraph is transformed into a bipartite undirected graph, in which only the input and output vertices and the edges between them remain. In the constructed bipartite undirected graph, we search for the minimal edge cover and restore the orientation of the edges in it. Next, we construct an algorithm for adding new edges, based on the search for Hamiltonian (or Eulerian) paths and turning the minimum edge cover into a class of cyclically equivalent vertices. The minimal number of edges to be added is not larger than the number of edges in the minimum edge cover.

Keywords: cluster, maximal matching, increasing alternating path, minimal edge cover, bipartite graph, star, Hamiltonian path, Eulerian path.

I. Problem statement and solution idea

Suppose that a complex system, such as a protein network, is represented by an acyclic digraph G without isolated vertices. In particular, using the algorithm built in [1] for identifying cyclic equivalence classes (clusters) in a digraph, the original digraph is transformed into an acyclic digraph whose vertices are clusters.

Let's call the vertices of the digraph G, from which only the edges come out, input and denote the set of input vertices U_*.Let's call the vertices that only the edges come in, output and denote the set of output vertices U^*. As the digraph G has not isolated vertices so from any input vertex there is a path to some output vertex and to any output vertex there is a path from some input vertex.

Our task is to add new directed edges to the digraph G so that there is a path from any vertex of the resulting digraph to any other vertex. This addition of the digraph makes it possible to introduce negative feedbacks and, consequently, to stabilize the functioning of the complex system under consideration and so to increase its reliability. In a sense, such a problem is the inverse of the digraph clustering problem considered in [1].

In this paper, this is achieved in two stages. At the first stage, we construct a bipartite digraph with vertex fractions U_*, U^* and edges (u_*,u^*), $u_* \in U_*$, $u^* \in U^*$ that are entered if there is a path from

vertex u_* to vertex in u^* the digraph G. We remove the orientation of the edges (u_*,u^*), $u_*\in U_*, u^*\in U^*$, from the bipartite graph and find the minimum edge cover [2], [3]. Its connectivity components are star graphs (connected graphs where all edges originate from a single vertex). In the minimal edge cover, we restore the orientation of the edges and denote the resulting bipartite digraph $\hat{G}$.

At the second stage, a minimal set of edges is introduced into the digraph $\hat{G}$, which turns all the vertices into a cluster. To do this, we first add a minimal set of edges in the constructed star graphs that generate a Hamiltonian or Eulerian path with a starting and ending vertices. Then edges are added that connect these paths in a Hamiltonian or Eulerian cycle. All the edges entered in the digraph $\hat{G}$ are added to the original digraph G, turning it into a cluster.

II. Finding feedbacks in a digraph $\hat{G}$

Consider a bipartite digraph $\hat{G}$, represented by a collection of unrelated stars. Let the star G_1 (Figure 1), have a vertex $1_*\in U_*$, as its root, leaves $1^*,\ldots,m^*$ and edges $1_*\to 1^*,\ldots,1_*\to m^*$. Let's add a minimal set of $m-1$ edges $1^*\to 2^*, 2^*\to 3^*,\ldots,(m-1)^*\to m^*$ to this star (coming out of the vertices $1^*,2^*,3^*,\ldots,(m-1)^*$), building a Hamiltonian path in it (a simple path that passes through all the vertices once):

$$1_*\to 1^*\to 2^*\to\ldots\to(m-1)^*\to m^*.$$

Let's call the vertex 1_* the starting point and the vertex m^* the ending point in this path.

This star can also be supplemented with a minimal set of $m-1$ edges $1^*\to 1_*, 2^*\to 1_*,\ldots,$ $(m-1)^*\to 1_*$ (coming out of the vertexes $1^*, 2^*, 3^*,\ldots,(m-1)^*$), building an Euler path in it (a path that passes through all the edges once):

$$1_*\to 1^*\to 1_*\to 2^*\to 1_*\to\ldots\to(m-1)^*\to 1_*\to m^*.$$

Let us call the vertex 1_* the starting point, and the vertex m^* the ending point in this path.

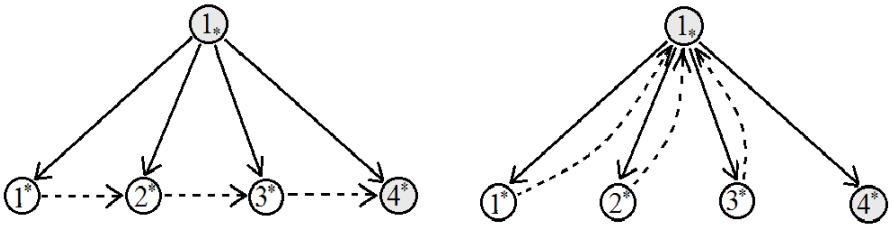


Figure 1. The Hamiltonian (left) and Eulerian (right) paths for the star G_1, $m=4$.

Let the star G_2, (Fig. 2), have a vertex $1^*\in U^*$, leaves $1_*,\ldots,n_*$ and edges $1_*\to 1^*,\ldots,n_*\to 1^*$. Let`s add this star by minimal set of $n-1$ edges $1^*\to 2_*, 2_*\to 3_*,\ldots,(n-1)_*\to n_*$, (included in the vertices $2_*,3_*,\ldots,n_*$), building a Hamiltonian path in it:

$$1_*\to 1^*\to 2_*\to 3_*\to\ldots\to(n-1)_*\to n_*.$$

Let's call the vertex 1_* as the starting point and the vertex n^* as the ending point in this path.

This star can also be supplemented with a minimal set of $n-1$ edges $1^*\to 2_*, 1^*\to 3_*,\ldots,1^*\to n_*$ (included in the vertices $2_*,3_*,\ldots,n_*$) building an Eulerian path in it:

$$1_*\to 1^*\to 2_*\to 1^*\to 3_*\to\ldots\to n_*\to 1^*.$$

Let us call the vertex 1_* as the starting point and the vertex 1^* as the ending point in this path.

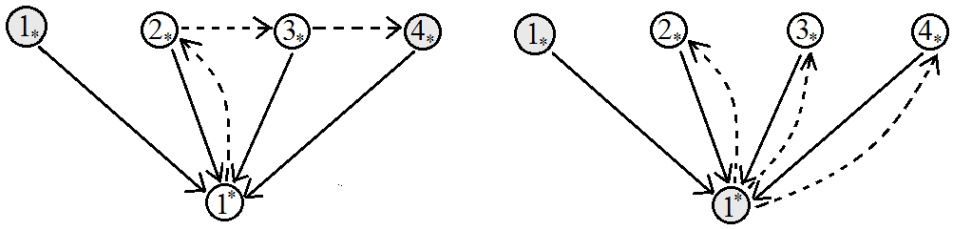


Figure 2. Hamiltonian (left) and Eulerian (right) paths for the star G_2, $n=4$.

Suppose now that the bipartite digraph $\hat{G}$ consists of several stars with roots from the set U_* and with roots from the set U^*. We connect by additional edge the final vertex of the Hamiltonian (Eulerian) path in the first star with the initial vertex of the Hamiltonian (Eulerian) path in the second star, etc. Then the final vertex of the Hamiltonian (Eulerian) path constructed for the last star we connect with the initial vertex of the Hamiltonian (Eulerian) path constructed for the first star.

As a result, we get a Hamiltonian (Eulerian) cycle passing through all the vertices of a bipartite digraph $\hat{G}$ (for an example, see the Hamiltonian cycle in Fig.3). In this case, the number of additional edges is equal to the number of edges in the bipartite digraph $\hat{G}$, which is the minimal edge covering of a bipartite digraph G. In addition to the Hamiltonian or Eulerian cycle, you can build a mixed-type cycle by connecting the Hamiltonian and Eulerian paths in series. Denote $n(\hat{G})$ number of edges in digraph $\hat{G}$, and $N(\hat{G})$ minimal number of new edges, the introduction of which in the digraph $\hat{G}$ leads to the formation of cycles containing all the vertices of the digraph $\hat{G}$ and consequently $N(\hat{G}) \le n(\hat{G})$. If all the stars in the minimal edge cover $\hat{G}$ are of the same type, then

$$N(\hat{G}) = n(\hat{G}). \tag{1}$$

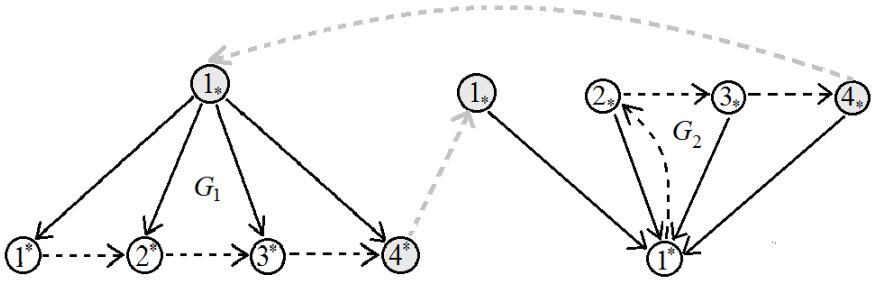


Figure 3. The Hamiltonian cycle for stars G_1, $m=4$, and G_2, $n=4$.

Indeed, let the digraph $\hat{G}$ consists of isolated stars of the type G_1. Then the total number of added edges (marked with dotted lines in Fig. 1 -- 3) coming out of the leaves of these stars in the cluster is equal to the total number of these leaves and cannot be less and so $N(\hat{G}) \ge n(\hat{G})$. Connecting the inequalities $N(\hat{G}) \le n(\hat{G})$ and $N(\hat{G}) \ge n(\hat{G})$ we obtain the equality (1).

Similarly, let the digraph $\hat{G}$ consists of isolated stars of type G_2. Then the total number of added edges included in the cluster leaves is the same as the total number of these leaves, so $N(\hat{G}) \ge n(\hat{G})$. Connecting this inequality with the opposite inequality, $N(\hat{G}) \le n(\hat{G})$ we obtain the equality (1).

But if the digraph $\hat{G}$ consists of isolated stars of type G_1 and stars of type G_2 then it is possible to decrease the number of added edges to make this digraph a cluster (see Figure 4). Therefore, in the general case, the ratio between $N(\hat{G})$ and $n(\hat{G})$ is as follows $N(\hat{G}) \le n(\hat{G})$.

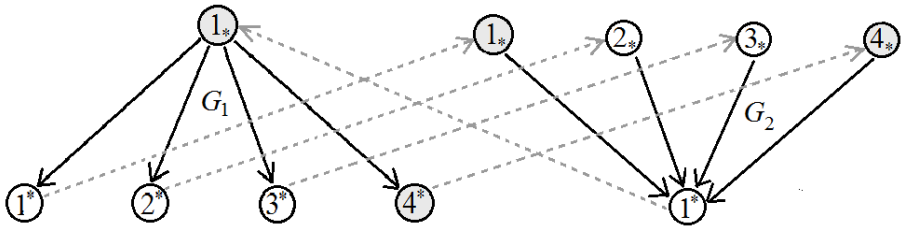


Figure 4. Example when $N(\hat{G}) < n(\hat{G})$ for stars G_1, $m = 4$, and G_2, $n = 4$.

III. Algorithms for finding the minimum edge cover in a bipartite digraph

Following [2] -- [4] to determine the minimum edge cover in an undirected bipartite graph $\overline{G}$, we first find the maximum matching, i.e. the maximum volume set of non-adjacent edges. For each vertex that does not belong to the maximum matching, some edge is selected that connects this vertex to the maximum matching. The maximum matching, together with the so-chosen edges, forms the minimum edge covering. The maximum matching in an undirected bipartite graph can be constructed in the following ways.

One way is to find the maximum flow in the graph $\overline{G}$. By adding the source s and edges from s to all the vertices from U_*, the drain t and edges from all the vertices of the fraction U^* to t. We assign each edge a throughput of one and find the maximum flow between the vertices and sequentially determine the paths that increase the flow. Then the edges between U_* and U^*, on which the flow is equal to one, form the maximum matching.

Another way to find the maximum matching is based on the construction of increasing alternating paths. Let some matching in the graph $\overline{G}$ be given (for example, one edge). We will call the edges of the matching strong, and the other edges of the graph weak. A vertex is called free if it does not belong to a matching. An alternating path is a simple path in which strong and weak edges alternate (i.e., a strong edge is followed by a weak one, and a weak one is followed by a strong one). An alternating path is called an increasing path if it connects two free vertices. If there is such a path relative to a given match, then you can build a larger match. By turning weak edges into strong ones, and strong edges into weak ones, we increase the number of matching edges by one. A match is maximal if and only if there are no increasing alternating paths relative to it.

References

[1]. Tsitsiashvili G. Sh. (2013). Sequential algorithms of graph nodes factorization. *Reliability: Theory and Applications*, 8 (4): 30-33.

[2]. Alekseev V. E., Zakharova D. V. Graph Theory: A textbook. Nizhny Novgorod: Nizhny Novgorod State University, 2017. (In Russian).

[3]. Cormen T. H., Leiserson Ch. E., Rivest R. L., Stein Cl. Introduction to Algorithms, 3rd Edition. The MIT Press, 2009.

[4]. Ford L. R., Fulkerson D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, 8: 399–404.

Truncated Shukla Distribution: Properties And Applications

Kamlesh Kumar Shukla[1] and Rama Shanker[2]

•

[1] Department of Community Medicine, Noida International Institute of Medical Science, NIU, Gautam Buddh Nagar, India, kkshukla22@gmail.com

[2]Department of Statistics, Assam University, Silchar, India
Shankerrama2009@gmail.com

Abstract

In this paper, Truncated Shukla distribution has been proposed. Some statistical properties including moments, coefficient of variation, skewness and index of dispersion have been derived. Survival and Hazard functions are derived and its behaviors are presented graphically. Maximum likelihood method of estimation has been used to estimate the parameter of proposed model. Simulation study of proposed distribution has also been discussed. It has been applied on three data sets and compares its superiority over two parameter Power Lindley, Gamma, Weibull, Shukla distributions and one parameter Truncated Akash, Truncated Lindley, Lindley and Exponential distributions

Keywords: Truncated distribution, Shukla distribution, Maximum likelihood Estimate

I. Introduction

In the recent decades, life time modeling has been becoming popular in distribution theory, where many statisticians are involved in introducing new models. Some of the life time models are very popular and applied in biological, engineering and agricultural areas, such as Lindley distribution suggested by Lindley [1], weighted Lindley distribution introduced by Ghitany et al. [2], Akash distribution of Shanker [3], Ishita distribution proposed by Shanker and Shukla [4], Pranav distribution of Shukla [5] etc, are some among others and extension of these lifetime distributions has also been capturing the attention of researchers in different areas of statistics, medical statistics, biomedical engineering, etc. Since each distribution is based on certain assumption and when these assumptions are not satisfied in the stochastic nature of the data, there is a need for another distribution. In search of a new distribution to fit the nature of some data sets where previously introduced lifetime distribution did not gives good fit, recently Shukla and Shanker [6] have introduced a lifetime distribution named Shukla distribution which is convex combination of exponential (θ) and gamma (α,θ) distributions and is defined by its probability density function (pdf) and cumulative distribution function (cdf) as

$$f(x;\theta,\alpha)=\frac{\theta^{\alpha+1}}{\theta^{\alpha+1}+\Gamma(\alpha+1)}\left(\theta+x^{\alpha}\right)e^{-\theta x};x>0,\theta>0,\alpha\geq 0 \tag{1.1}$$

$$F(x;\theta,\alpha)=1-\frac{\theta^{\alpha}\left(\theta+x^{\alpha}\right)e^{-\theta x}+\alpha\Gamma(\alpha,\theta x)}{\theta^{\alpha+1}+\Gamma(\alpha+1)};x>0,\theta>0,\alpha\geq 0 \tag{1.2}$$

Shukla and Shanker [6] have discussed in details about its mathematical and statistical properties, estimation of parameters and application to model lifetime.

Truncated type of distribution are more effective in application to modeling life time data because its limits used as bound either upper or lower or both according to given data. Truncated normal distribution is well explained in Johnson et al [7]. It has wide application in economics and statistics. Many researchers have proposed truncated type of distribution and applied it in different areas of statistics, especially in censor data such as truncated Weibull distribution by Zange and Xie [8] , truncated Lomax distribution by Aryuyuen and Bodhisuwan [9], truncated Pareto distribution by Janinetti and Ferraro[10], truncated Lindley distribution by Singh et al [11], are some among others. Truncated version of distribution can be defined as

Definition 1. Let X be a random variable that is distributed according to some pdf $g(x;\theta)$ and cdf $G(x;\theta)$, where θ is a parameter vector of X.

Let X lies within the interval $[a,b]$ where $-\infty<a\leq x\leq b<\infty$ then the conditional on $a\leq x\leq b$ is distributed as truncated distribution and thus the pdf of truncated distribution as reported by Singh et al [11]is defined by

$$f(x;\theta)=g(x/a\leq x\leq b;\theta)=\frac{g(x;\theta)}{G(b;\theta)-G(a;\theta)} \tag{1.3}$$

where $f(x;\theta)=g(x;\theta)$ for all $a\leq x\leq b$ and $f(x;\theta)=0$ elsewhere.

Notice that $f(x;\theta)$ in fact is a pdf of X on interval $[a,b]$. We have

$$\int_a^b f(x;\theta)dx=\frac{1}{G(b;\theta)-G(a;\theta)}\int_a^b g(x;\theta)dx$$

$$=\frac{1}{G(b;\theta)-G(a;\theta)}G(b;\theta)-G(a;\theta)=1 \tag{1.4}$$

The cdf of truncated distribution is given by

$$F(x;\theta)=\int_a^x f(x;\theta)dx=\frac{G(x;\theta)-G(a;\theta)}{G(b;\theta)-G(a;\theta)} \tag{1.5}$$

The main objectives of this paper are (i) to propose new truncated distribution using Shukla distribution, which is called as Truncated Shukla distribution (TSD) (ii) to know statistical performance and its suitability, it has been compared with classical distributions of two- parameter as well as one parameter using three lifetime datasets. It has been divided in eight sections. Introduction about the paper is described in the first section. In the second section, Truncated Shukla distribution has been derived. Mathematical and statistical properties including its moment have been discussed in third section. Behavior of hazards rate has been presented mathematically as well as graphically in fourth section. Moments and its related expression have been discussed in fifth section. Simulation study of the presented distribution has been discussed to check estimation parameters using Bias and Mean square error in sixth section. Estimation of parameter of proposed distribution has been discussed in seven section where its applications and comparative study of other classical two parameter life time distributions as well as one parameter distributions have been illustrated using various fields of life time data. In the last, conclusions have been drawn according to studied of behavior and properties of Truncated Shukla distribution (TSD).

.

II. Truncated Shukla Distribution

In this section, pdf and cdf of new truncated distribution is proposed and named Truncated Shukla distribution , using (1.3) & (1.4) of definition1 and from (1.1) & (1.2) , which is defined as :

Definition 2: Let X be random variable which is distributed as Truncated Shukla distribution (TSD) with scale parameter a, b and θ , and shape parameter α, will be denoted by TSD (a,b,θ,α). The pdf and cdf of X are respectively:

$$f(x;\theta,\alpha)=\frac{\theta^{\alpha+1}(x^{\alpha}+\theta)e^{-\theta x}}{\theta^{\alpha}\left((\theta+a^{\alpha})e^{-\theta a}-(\theta+b^{\alpha})e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)} \tag{1.6}$$

$$F(x;\theta,\alpha)=\frac{\theta^{\alpha}\left((\theta+a^{\alpha})e^{-\theta a}-(\theta+x^{\alpha})e^{-\theta x}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta x)\right)}{\theta^{\alpha}\left((\theta+a^{\alpha})e^{-\theta a}-(\theta+b^{\alpha})e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)} \tag{1.7}$$

Where $-\infty < a \le x \le b < \infty$, and $\theta > 0$

Performance of pdf of TSD for varying values of parameter has been illustrated in the figure 1.From the pdf plots of TSD, it is quite obvious that TSD is suitable for datasets of various nature.

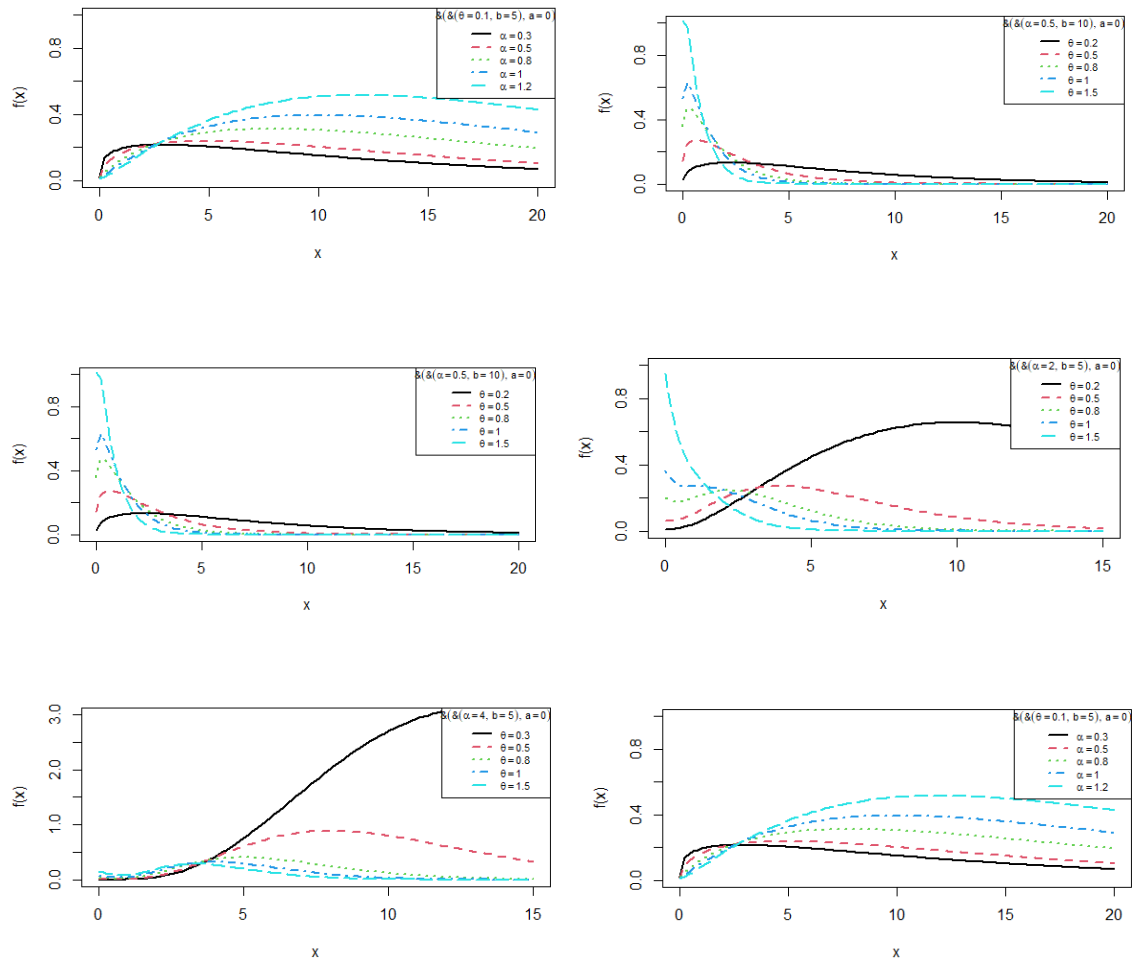


Figure 1: *pdf plots of TSD for varying values of parameters continued…*

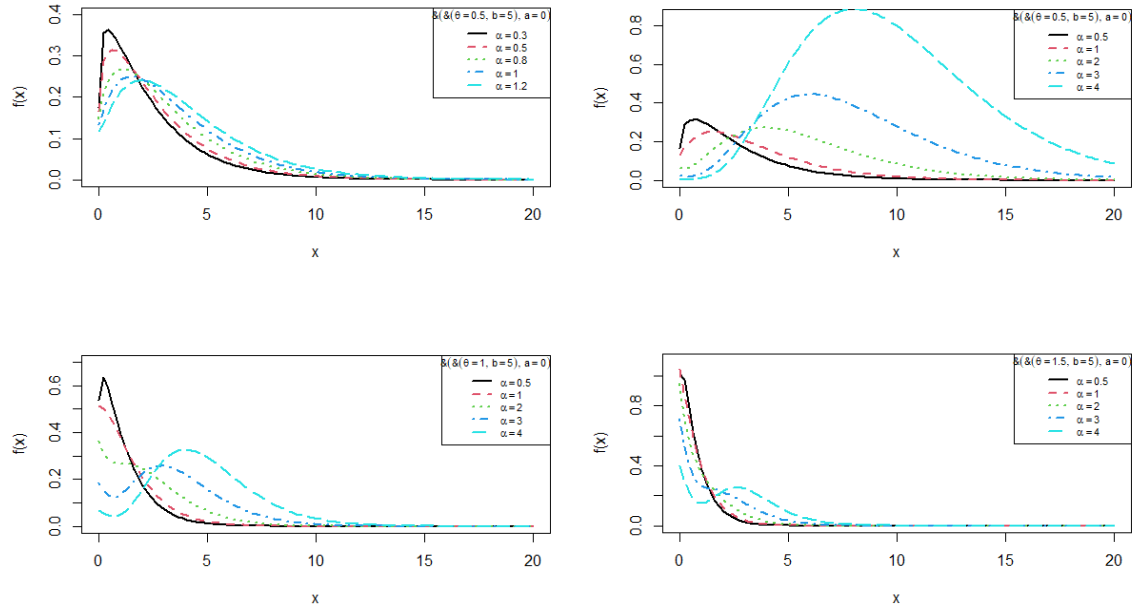


Figure 1: *pdf plots of TSD for varying values of parameters*

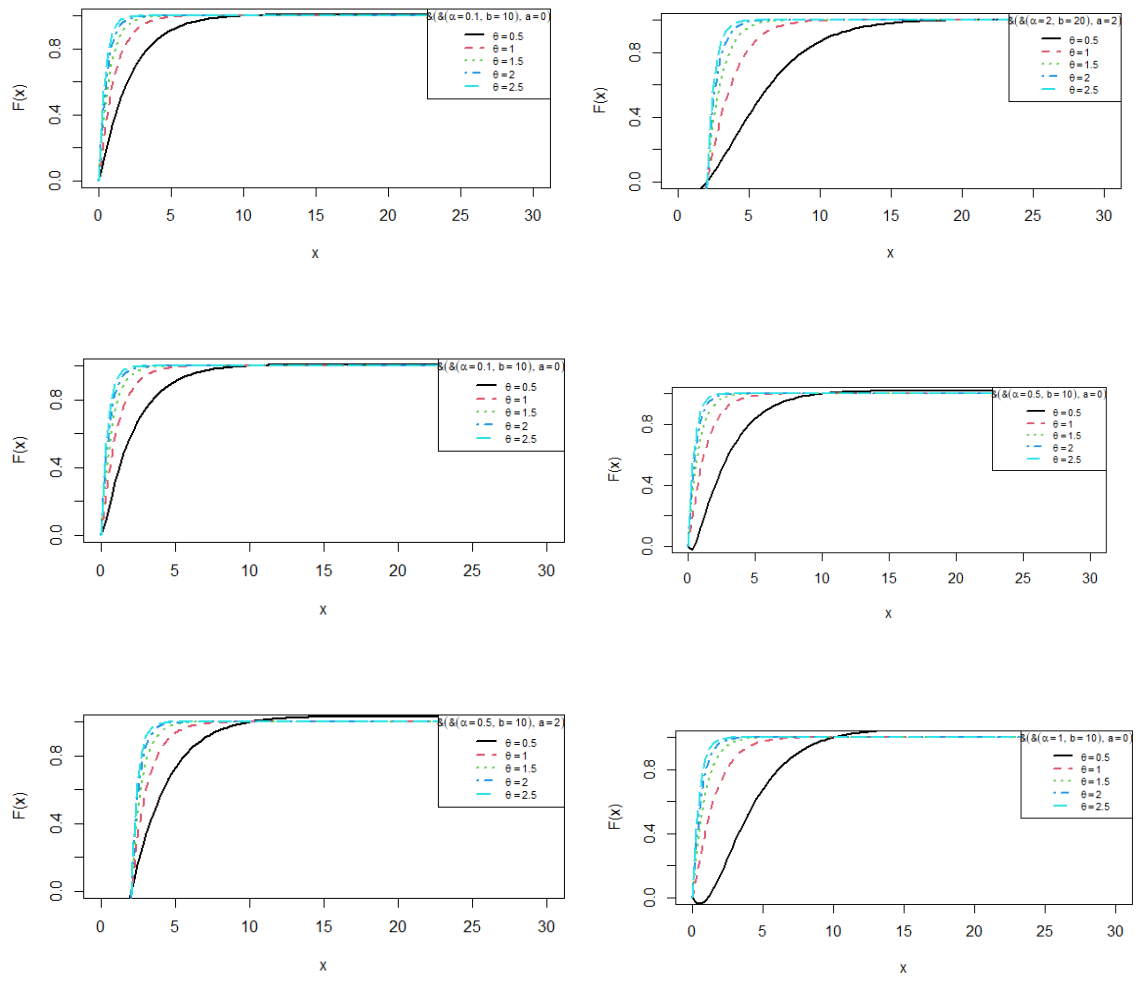


Figure 2: *cdf plots of TSD for varying values of parameters continued…*

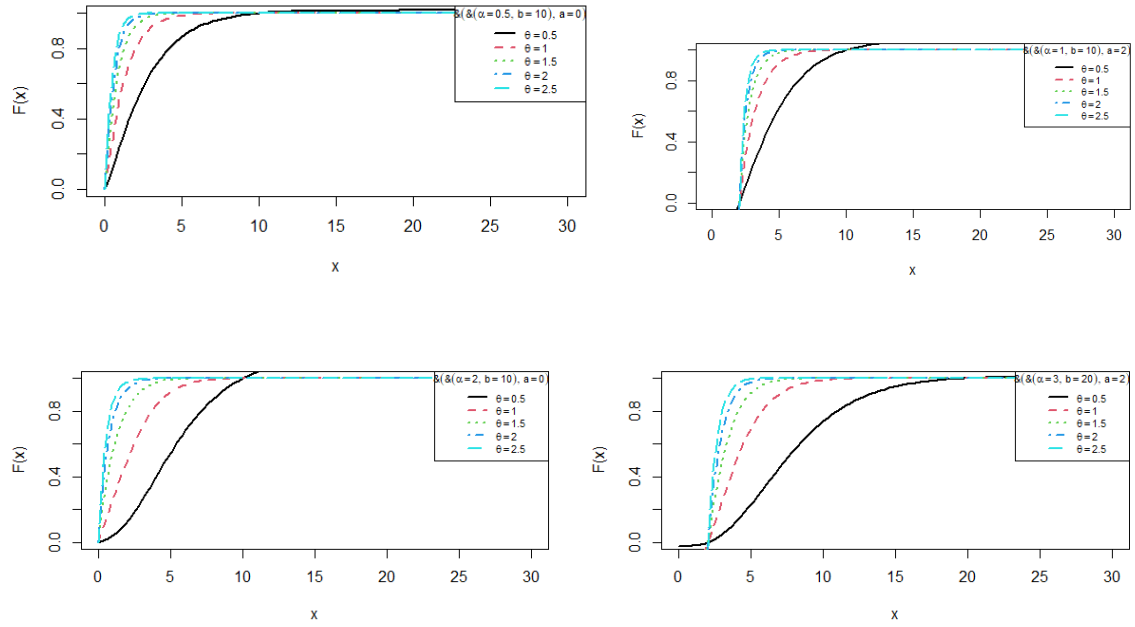


Figure 2: *cdf plots of TSD for varying values of parameters*

III. Survival and hazard function

The survival function $S(x) = S(x;\theta,\alpha)$ and hazard function $h(x) = h(x;\theta,\alpha)$ of TSD are respectively, obtained as

$$S(x;\theta,\alpha) = 1 - F(x;\theta,\alpha) = \frac{\theta^{\alpha}\left((\theta + x^{\alpha})e^{-\theta x} - (\theta + b^{\alpha})e^{-\theta b}\right) + \alpha\left(\Gamma(\alpha,\theta x) - \Gamma(\alpha,\theta b)\right)}{\theta^{\alpha}\left((\theta + a^{\alpha})e^{-\theta a} - (\theta + b^{\alpha})e^{-\theta b}\right) + \alpha\left(\Gamma(\alpha,\theta a) - \Gamma(\alpha,\theta b)\right)}$$

$$h(x) = \frac{f(x;\theta,\alpha)}{S(x;\theta,\alpha)} = \frac{\theta^{\alpha+1}(x^{\alpha} + \theta)e^{-\theta x}}{\theta^{\alpha}\left((\theta + x^{\alpha})e^{-\theta x} - (\theta + b^{\alpha})e^{-\theta b}\right) + \alpha\left(\Gamma(\alpha,\theta x) - \Gamma(\alpha,\theta b)\right)}$$

It is quite obvious that the hazard rate function $h(x)$ is independent of parameter 'a'. Behavior of hazard function of TSD for varying values of parameter is presented in figure 3.

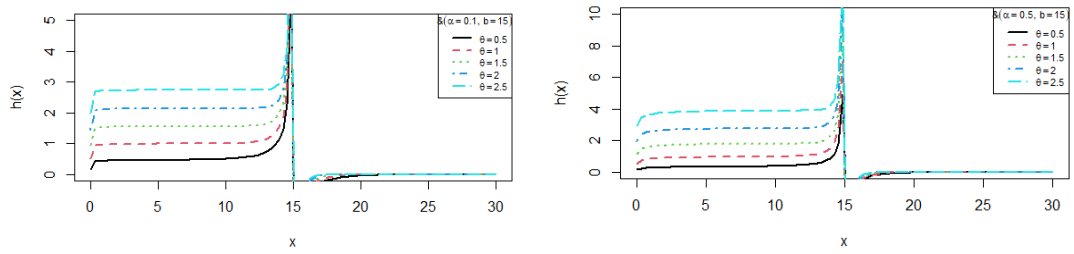


Figure 3: *h(x) plots of TSD for varying values of parameters continued…*

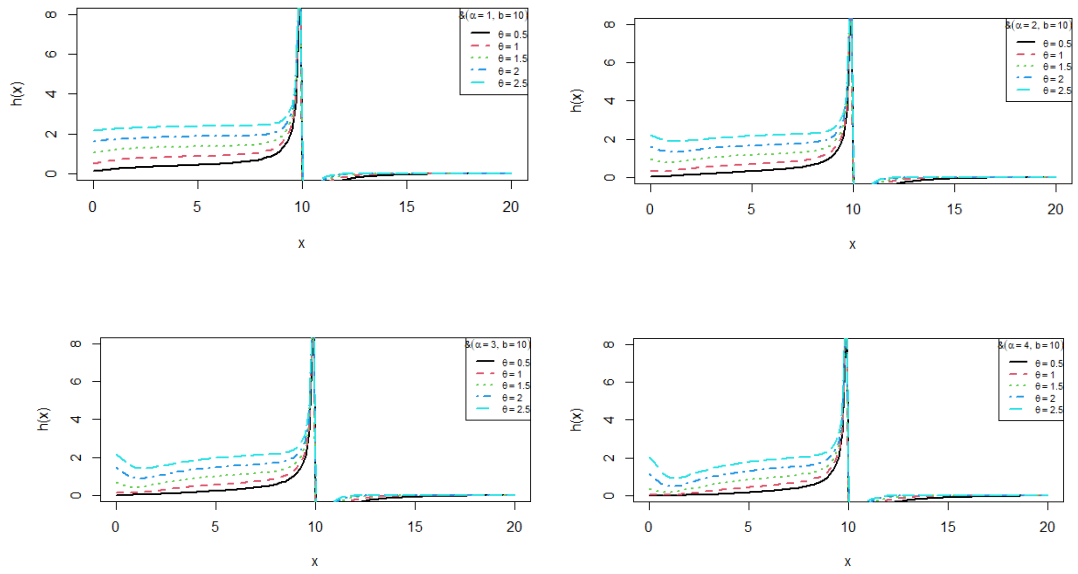


Figure 3: *h(x) plots of TSD for varying values of parameters*

IV. Moments and Mathematical Properties

The rth moment about origin of TSD is defined as

$$\mu_r' = \frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+r+1)\right]}{\theta^r\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}};\ r=1,2,3,\ldots$$

The first four moments about origin of TSD can thus be given by

$$\mu_1' = \frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+2)\right]}{\theta\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}}$$

$$\mu_2' = \frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+3)\right]}{\theta^2\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}}$$

$$\mu_3' = \frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+4)\right]}{\theta^3\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}}$$

$$\mu_4' = \frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+5)\right]}{\theta^4\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}}.$$

The variance of TSD can be obtained as

$$\mu_2 = \left(\frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+3)\right]}{\theta^2\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}}\right)$$

$$-\left(\mu_1' = \frac{\left[\theta^{\alpha+1}\Gamma(\alpha+1)+\Gamma(\alpha+2)\right]}{\theta\left\{\theta^\alpha\left((\theta+a^\alpha)e^{-\theta a}-(\theta+b^\alpha)e^{-\theta b}\right)+\alpha\left(\Gamma(\alpha,\theta a)-\Gamma(\alpha,\theta b)\right)\right\}}\right)^2$$

Expressions of other central moments are not being given here because they have lengthy

expressions. However, they can be easily obtained. Natures of the mean and the variance for varying values of parameters of TSD are presented in figures 4 and 5 respectively.

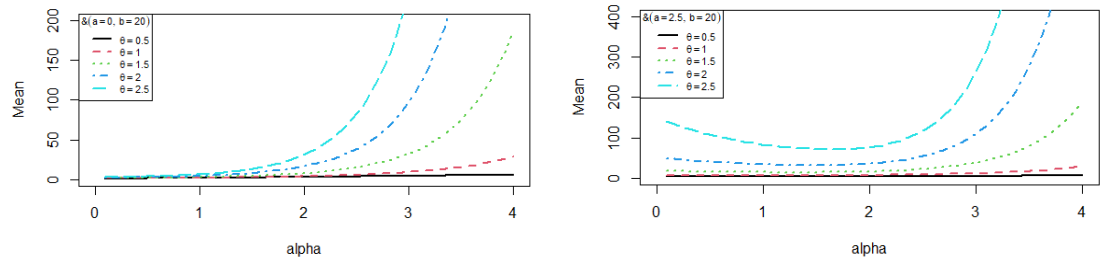


Figure 4: *Mean of TSD on varying value of parameters*

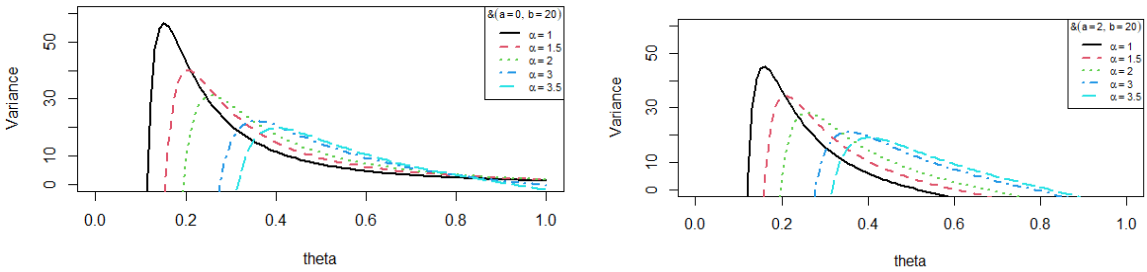


Figure 5: *Variance of TSD on varying value of parameters*

V. Maximum Likelihood Estimation

Let $(x_1, x_2, x_3, \ldots, x_n)$ be a random sample of size n from (1.6). The likelihood function, L of TSD is given by

$$L = \left(\frac{\theta^{\alpha+1}}{\theta^{\alpha}\left((\theta+a^{\alpha})e^{-\theta a} - (\theta+b^{\alpha})e^{-\theta b}\right) + \alpha\left(\Gamma(\alpha,\theta a) - \Gamma(\alpha,\theta b)\right)} \right)^{n} \prod_{i=1}^{n} \left(\theta + x_i^{\alpha}\right) e^{-n\theta\bar{x}}$$

and its log likelihood function is thus obtained as

$$\ln L = n \ln \left(\frac{\theta^{\alpha+1}}{\theta^{\alpha}\left((\theta+a^{\alpha})e^{-\theta a} - (\theta+b^{\alpha})e^{-\theta b}\right) + \alpha\left(\Gamma(\alpha,\theta a) - \Gamma(\alpha,\theta b)\right)} \right) + \sum_{i=1}^{n} \ln\left(\theta + x_i^{\alpha}\right) - n\theta\bar{x}$$

Taking $\hat{a} = Min(x_1, x_2, x_3, \ldots, x_n)$, $\hat{b} = Max(x_1, x_2, x_3, \ldots, x_n)$, the maximum likelihood estimate $\hat{\theta}$ of parameter θ is the solution of the log-likelihood equation $\frac{\partial \ln L}{\partial \theta} = 0$. It is obvious that $\frac{\partial \ln L}{\partial \theta} = 0$ will not be in closed form and hence some numerical optimization technique can be used to solve the equation for θ. In this paper the nonlinear method available in R software has been used to find the MLE of the parameter θ.

.

VI. Simulation Study

In this section, simulation of study of (1.6) has been carried out. Acceptance and Rejection method has been used to generate random number. Bias Error and Mean square Error have been calculated for varying values parameters θ and α whereas parameter a and b kept constant. It is obvious that as the sample size increases and values of parameters increases, the bias error and the mean square errors of both parameters decrease.

Table 1. *Simulation of TSD at a=0, b=20 and* $\theta = 0.1$ *&* $\alpha = 0.5$

Sample Size (n)	θ	α	Bias Error(θ)	MSE(θ)	Bias Error(α)	MSE(α)
40	0.1	0.5	0.02709	0.02935	0.071393	0.203882
	0.5	1.0	0.017091	0.011685	0.058893	0.138738
	1.0	2.0	0.004591	0.000843	0.033893	0.045951
	1.5	3.0	-0.007908	0.002501	0.008893	0.003163
60	0.1	0.5	0.005875	0.009437	0.027591	0.0456763
	0.5	1.0	0.017091	0.002071	0.019257	0.022251
	1.0	2.0	-0.002458	0.0003625	0.002591	0.0004028
	1.5	3.0	-0.010791	0.006987	-0.014075	0.0118871
80	0.1	0.5	0.007277	0.004236	0.015552	0.0193501
	0.5	1.0	0.0022772	0.000414	0.0093024	0.0069227
	1.0	2.0	-0.003972	0.0012626	-0.00319	0.0008179
	1.5	3.0	-0.010222	0.0083604	-0.015697	0.019713
100	0.1	0.5	0.0060863	0.0037043	0.013617	0.0185445
	0.5	1.0	0.0020863	0.0004352	0.0086178	0.0074267
	1.0	2.0	-0.002913	0.0008489	-0.001382	0.0001910
	1.5	3.0	-0.007913	0.0062625	-0.011382	0.0129553

Table 2. *Simulation of TSD at a=0, b=20 and* $\theta = 0.5$ *&* $\alpha = 1$

Sample Size (n)	θ	α	Bias Error(θ)	MSE(θ)	Bias Error(α)	MSE(α)
40	0.1	0.5	0.02900024	0.0336405	0.07806883	0.2437896
	0.5	1.0	0.01900024	0.0144403	0.06556883	0.1719708
	1.0	2.0	0.00650024	0.0016901	0.04056883	0.0658331
	1.5	3.0	-0.00599976	0.0014398	0.01556883	0.0096955
60	0.1	0.5	0.01367068	0.0112132	0.03260302	0.0637774
	0.5	1.0	0.00700401	0.0029433	0.02426968	0.0353410
	1.0	2.0	-0.0013293	0.0001060	0.00760302	0.0034683
	1.5	3.0	-0.0096626	0.0056020	-0.00906366	0.0049289
80	0.1	0.5	0.00681314	0.0037135	0.01293939	0.0133942
	0.5	1.0	0.00181314	0.0002629	0.00668939	0.0035798
	1.0	2.0	-0.00443685	0.0015748	-0.00581060	0.0027010
	1.5	3.0	-0.01068685	0.0091367	-0.01831060	0.0268222
100	0.1	0.5	0.00630281	0.0039725	0.01477387	0.0218267
	0.5	1.0	0.00230281	0.0005302	0.00977387	0.0095528
	1.0	2.0	-0.00269718	0.0007274	-0.00022612	0.0000511
	1.5	3.0	-0.00769718	0.0059246	-0.01022613	0.0104573

Table 3. *Simulation of TSD at a=0, b=20 and* $\theta = 1$ *&* $\alpha = 2$

Sample Size (n)	θ	α	Bias Error(θ)	MSE(θ)	Bias Error(α)	MSE(α)
40	0.1	0.5	0.02307633	0.0213006	0.04443256	0.0789700
	0.5	1.0	0.01307633	0.0068396	0.03193256	0.0407875
	1.0	2.0	0.000576330	0.0000132	0.00693256	0.0019224
	1.5	3.0	-0.0119236695	0.0056869	-0.01806743	0.0130572
60	0.1	0.5	0.01594010	0.0152452	0.03033473	0.0552117
	0.5	1.0	0.00927343	0.0051597	0.02200140	0.0290436
	1.0	2.0	0.00094010	0.0000530	0.00533473	0.0017075
	1.5	3.0	-0.00739322	0.0032795	-0.01133193	0.0077047
80	0.1	0.5	0.00993921	0.0079030	0.01602242	0.0205374
	0.5	1.0	0.00493921	0.0019516	0.00977242	0.0076400
	1.0	2.0	-0.00131078	0.0001374	-0.00272757	0.0005951
	1.5	3.0	-0.00756078	0.0045732	-0.01522757	0.0185503
100	0.1	0.5	0.00943759	0.0089068	0.01685769	0.0284181
	0.5	1.0	0.00543759	0.0029567	0.01185769	0.0140604
	1.0	2.0	0.00043759	0.0000191	0.00185769	0.0003451
	1.5	3.0	-0.00456240	0.0020815	-0.00814230	0.0066297

Table 4. *Simulation of TSD at a=0, b=20 and* $\theta = 1.5$ *&* $\alpha = 3$

Sample Size (n)	θ	α	Bias Error(θ)	MSE(θ)	Bias Error(α)	MSE(α)
40	0.1	0.5	0.04484480	0.08044228	0.09300719	0.34601349
	0.5	1.0	0.03484480	0.04856643	0.08050719	0.0407875
	1.0	2.0	0.022344809	0.01997162	0.05550719	0.25925630
	1.5	3.0	0.009844809	0.00387681	0.03050719	0.03722754
60	0.1	0.5	0.02788600	0.046657753	0.05440362	0.17758525
	0.5	1.0	0.02121933	0.027015617	0.04607029	0.12734829
	1.0	2.0	0.01288600	0.009962946	0.02940362	0.05187438
	1.5	3.0	0.00455267	0.001243609	0.01273696	0.00973380
80	0.1	0.5	0.01902251	0.028948472	0.03411478	0.09310547
	0.5	1.0	0.01402251	0.015730463	0.02786478	0.06211569
	1.0	2.0	0.00777251	0.004832953	0.01536478	0.01888612
	1.5	3.0	0.00152251	0.000185443	0.00286478	0.00065655
100	0.1	0.5	0.01418562	0.02012320	0.02460816	0.06055620
	0.5	1.0	0.01018562	0.002689072	0.01960816	0.03844803
	1.0	2.0	0.00518562	0.000093457	0.00960816	0.00923169
	1.5	3.0	0.00018562	0.0020815	-0.00039183	0.00015353

VII. Applications on life time data

In this section, TSD has been applied on three real lifetime datasets, where maximum likelihood method of estimation has been used for the estimation of its parameter. Booth parameters and θ and α are estimated using maximum likelihood estimation whereas another parameters a, and b are considered as lowest and highest values of data. i. e. *a=max(x)* and *b=max(x),* where x is data set. Goodness of fit has been decided using Akaike information criteria (AIC), Bayesian Information criteria (BIC) and Kolmogorov Simonov test (KS) values respectively, which are calculated for each distribution and also compared with p-value. As we know that best goodness of fit of the distribution can be decide on the basis minimum value of KS, AIC and BIC.

The datasets considered for testing goodness of fit of the TSD are as follows:

Data Set 1: The data is given by Birnbaum and Saunders [12] on the fatigue life of 6061 – T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The data set consists of 100 observations with maximum stress per cycle 31,000 psi. The data ($\times 10^{-3}$)are presented below (after subtracting 65**).**

5	25	31	32	34	35	38	39	39	40	42
43	43	43	44	44	47	47	48	49	49	49
51	54	55	55	55	56	56	56	58	59	59
59	59	59	63	63	64	64	65	65	65	66
66	66	66	66	67	67	67	68	69	69	69
69	71	71	72	73	73	73	74	74	76	76
77	77	77	77	77	77	79	79	80	81	83
83	84	86	86	87	90	91	92	92	92	92
93	94	97	98	98	99	101	103	105	109	136
147										

Data Set 2: This data set is the strength data of glass of the aircraft window reported by Fuller *et al* [13]:

18.83	20.8	21.657	23.03	23.23	24.05	24.321	25.5	25.52	25.8	26.69	26.77
26.78	27.05	27.67	29.9	31.11	33.2	33.73	33.76	33.89	34.76	35.75	35.91
36.98	37.08	37.09	39.58	44.045	45.29	45.381					

Data Set 3: The following data represent the tensile strength, measured in GPa, of 69 carbon fibers tested under tension at gauge lengths of 20mm (Bader and Priest[14]):

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	3.858								

The values of MLE's, Standard Errors, - 2ln L, AIC, K-S and p-values for the fitted distributions for the three datasets are presented in tables 5, 6 and 7, respectively. It crystal clear that the TSD gives much closer fit than the considered distributions.

Table 5: *MLE's, Standard Errors, - 2ln L, AIC, K-S and p-values of the fitted distributions for data set-1*

Distributions	**ML Estimates**	$-2\ln L$	**AIC**	**BIC**	**K-S**	**p-value**
TSD	$\hat{\theta}=0.10677$ $\hat{\alpha}=6.36414$	**914.72**	**918.72**	**923.93**	**0.095**	0.320
Power Lindley	$\hat{\theta}=0.00266$ $\hat{\alpha}=1.55583$	925.41	929.41	934.62	0.155	**0.015**
Gamma	$\hat{\theta}=0.11400$ $\hat{\alpha}=7.78985$	915.77	919.77	924.98	0.097	**0.281**
Weibull	$\hat{\theta}=0.00272$ $\hat{\theta}=1.39558$	989.35	993.35	998.56	0.294	**0.000**
Shukla	$\hat{\theta}=0.11253$ $\hat{\alpha}=6.6892$	915.76	919.76	924.97	0.099	**0.271**
TAD	$\hat{\theta}=0.03917$	939.13	941.13	942.05	0.153	**0.017**
TLD	$\hat{\theta}=0.02199$	958.88	960.88	962.31	0.186	**0.001**
Lindley	$\hat{\theta}=0.02886$	983.10	985.10	986.54	0.252	**0.000**
Exponential	$\hat{\theta}=0.01463$	**1044.87**	**1046.87**	**1048.30**	**0.336**	**0.000**

Table 6: *MLE's, Standard Errors, - 2ln L, AIC, K-S and p-values of the fitted distributions for data set-2*

Distributions	**ML Estimates**	$-2\ln L$	**AIC**	**BIC**	**K-S**	**p-value**
TSD	$\hat{\theta} = 0.21682$ $\hat{\alpha} = 5.93067$	**201.61**	**205.61**	**208.48**	**0.106**	0.834
Power Lindley	$\hat{\theta} = 0.00243$ $\hat{\alpha} = 1.9439$	220.14	224.14	226.13	0.198	**0.152**
Gamma	$\hat{\theta} = 0.61482$ $\hat{\alpha} = 18.9433$	208.22	212.22	216.05	0.134	**0.578**
Weibull	$\hat{\theta} = 0.00203$ $\hat{\theta} = 1.80566$	241.61	245.61	247.61	0.353	**0.000**
Shukla	$\hat{\theta} = 0.6144$ $\hat{\theta} = 17.9299$	208.23	212.23	216.05	0.134	**0.580**
TAD	$\hat{\theta} = 0.08776$	201.96	203.96	205.58	0.112	**0.786**
TLD	$\hat{\theta} = 0.05392$	202.18	204.18	205.61	0.117	**0.738**
Lindley	$\hat{\theta} = 0.06299$	253.98	255.98	256.98	0.365	**0.000**
Exponential	$\hat{\theta} = 0.032452$	**274.52**	**276.52**	**277.52**	**0.458**	**0.000**

Table 7: *MLE's, Standard Errors, - 2ln L, AIC, K-S and p-values of the fitted distributions for data set-3*

Distributions	**ML Estimates**	$-2\ln L$	**AIC**	**BIC**	**K-S**	**p-value**
TSD	$\hat{\theta}=8.0439$ $\hat{\alpha}=18.8613$	**99.71**	**103.71**	**106.58**	**0.061**	0.985
Power Lindley	$\hat{\theta}=0.05447$ $\hat{\alpha}=3.75881$	101.52	105.52	110.73	0.055	**0.988**
Gamma	$\hat{\theta}=9.2873$ $\hat{\alpha}=22.8030$	101.97	105.97	111.18	0.057	**0.977**
Weibull	$\hat{\theta}=0.00643$ $\hat{\theta}=5.16972$	103.47	107.47	112.68	0.065	**0.928**
Shukla	$\hat{\theta}=5.9922$ $\hat{\theta}=17.1611$	184.35	188.35	193.56	0.290	**0.000**
TAD	$\hat{\theta}=0.70314$	110.76	112.76	114.68	0.152	**0.079**
TLD	$\hat{\theta}=0.28986$	112.19	114.19	115.63	0.157	**0.065**
Lindley	$\hat{\theta}=0.65450$	238.38	240.38	241.37	0.401	**0.000**
Exponential	$\hat{\theta}=0.40794$	**261.73**	**263.73**	**264.73**	**0.448**	**0.000**

Fitted pots of considered distributions for the datasets 1, 2 and 3 are given in the figures 6, 7 and 8 respectively which also supports the claim that TSD is the best distribution among the considered datasets.

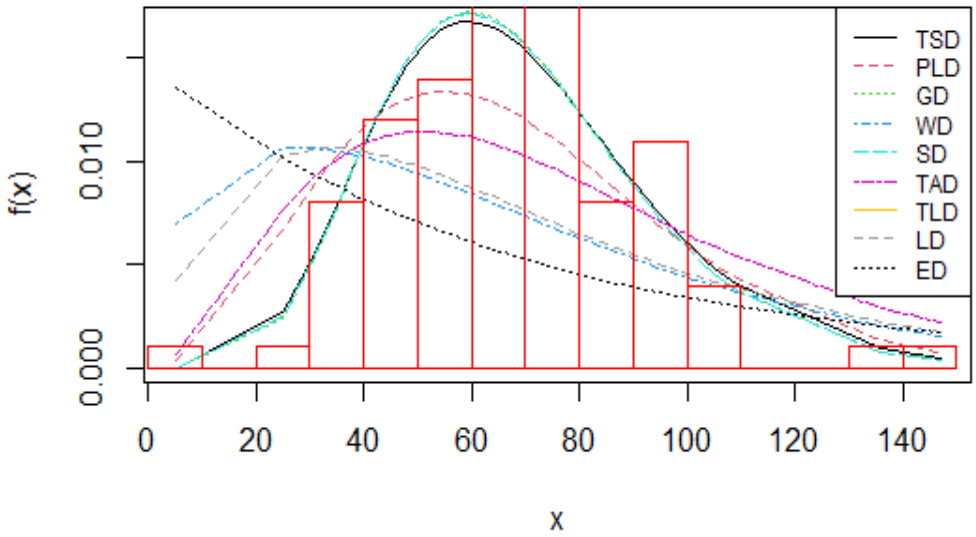


Figure.6. *Fitted plots of distributions for the dataset-1*

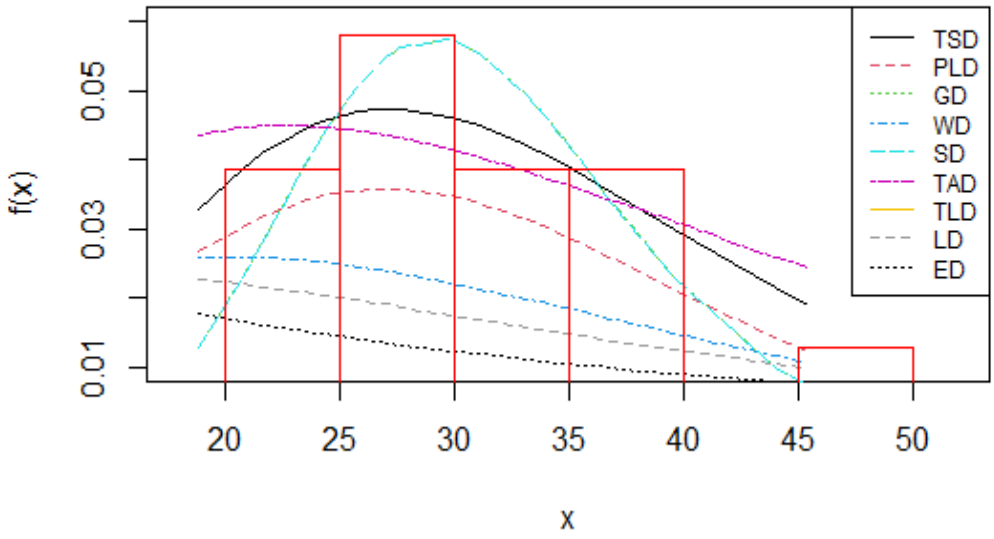


Figure.7. *Fitted plots of distributions for the dataset-2*

VIII. Conclusions

In this paper, Truncated Shukla distribution (TSD) has been proposed. Its mathematical and statistical properties have been discussed. Maximum likelihood method has been used for the estimation of its parameters. Goodness of fit of TSD has been discussed with three life time data sets and compared with Gamma, Weibull, Power Lindley distribution (PLD), truncated Lindley, Truncated distribution (TLD), truncated Akash distribution (TAD), Lindley and exponential distribution. It has been observed that TSD gives good fit over the considered distributions on all the data sets.

,

References

[1] Lindley, D.V. (1958), Fiducial distributions and Bayes' Theorem, *Journal of the Royal Statistical Society, Series B*, 20:102 – 107.

[2] Ghitany ME, Atieh B, Nadarajah S. (2008), Lindley distribution and its Applications. *Mathematics Computing and Simulation*, 78: 493– 506.

[3] Shanker, R. (2015), Akash distribution and its application, *International Journal of Probability and Statistics.* 4(3):65–75.

[4] Shanker,R. and Shukla, K.K. (2017), Ishita distribution and its applications, *Biometrics & Biostatistics International Journal*, 5 : 1-9.

[5] Shukla, K. K. (2018), Pranav distribution with Properties and Applications, *Biometrics and Biostatistics International Journal*,7(3):244-254.

[6] Shukla, K.K. and Shanker, R. (2019), Shukla distribution and its Application, *Reliability: Theory and Application, No-3*, (54): 46 - 55

[7] Johnson, N. L., Kotz,S. and Balakrishnan,N., Continuous Univariate Distributions, Wiley, New York (1994).

[8] Zange,Z., and Xie, M., (2011),On the upper truncated Weibull distribution and its reliability implications, *Reliability Engineering & System Safety*, 96:194-200.

[9] Aryuyuen, S. and Bodhisuwan, W. (2018). The truncated power Lomax distribution: Properties and applications, *Walailak Journal of Science and Technology*, 16: 655-668.

[10] Zaninetti,L. and Ferraro,M. (2008), On the truncated Pareto distribution with applications, *Central European Journal of Physics*, 6 :1-6.

[11] Singh,S. K., Singh,U. and Sharma, V.K. (2014), The truncated Lindley distribution: Inference and application, *Journal of Statistics Applications & Probability*, 3:219-228.

[12] Birnbaum, Z.W., Saunders, S.C. (1969), Estimation for a family of life distribution with applications to fatigue, *Journal of applied probability*, 6(2):328-347.

[13] Fuller, E.J., Frieman, S., Quinn, J., Quinn, G., and Carter, W., Fracture mechanics approach to the design of glass aircraft windows: A case study, SPIE Proc(1994) 2286:419- 430.

[14] Bader,M.G., Priest, A. M. (1982): Statistical aspects of fiber and bundle strength in hybrid composites, In; hayashi, T., Kawata, K. Umekawa, S. (Eds), Progressin Science in Engineering Composites, ICCM-IV, Tokyo, 1129 – 1136.

Dynamical Behavior of an SEIS Epidemic Model with Nonlinear Incidence Rate

Garima Saxena[1], R. K. Sharma[2], Chandrashekhar Chauhan[3] & Ankit Agrawal[4]

•

[1,3]Institute of Engineering & Technology, DAVV, Indore (M.P.), India
[1]rimasa.1907@gmail.com, [3]cchauhan@ietdavv.edu.in
[2]Department of Mathematics, Govt. Holkar Science College, DAVV, Indore (M.P.), India
[2]raj_rma@yahoo.co.in
[4]Department of Mathematics, Govt. Girls P.G. College, Ujjain (M.P.), India
[4]ankit_agrawal458@yahoo.com

Abstract

In this paper, an SEIS epidemic model with nonlinear incidence rate is studied. The basic reproduction number (R_0) is calculated. The local and global stability of the disease free equilibrium E_0 and the endemic equilibrium E^* of the model are discussed and also the global asymptotical stability of the disease free equilibrium and endemic equilibrium are discussed. The stability analysis of the model shows that the system is locally asymptotically stable at disease free equilibrium E_0 and endemic equilibrium E^* under suitable conditions. Moreover, show that the disease free equilibrium and the unique endemic equilibrium of the system is globally asymptotically stable under certain conditions. Finally, numerical simulations are given to support some of the theoretical results.

Keywords: epidemic models, equilibrium, local and global stability.

I. Introduction

Mathematical models describing the population dynamics of infectious diseases have been playing an important role in disease control for a long time. In recent years, various epidemic models have been proposed and explored to prevent and control the spread of the infectious diseases, such as measles, tuberculosis, and flu (see e.g., [2, 12]). Simple mass action bilinear incidence rate βSI was introduced Kermack –Mcendrick [5] in 1927. In many epidemic models, bilinear incidence rate βSI is frequently used [1, 6, 13, 14, 15]. Moreover, nonlinear incidence rates of the form $\beta I^p S^q$ were investigated by Liu, Hethcote and Levin [9], Liu, Levin and Iwasa [10]. In this paper, we consider an $SEIS$ epidemic model with nonlinear incidence rate $\beta I^p S^q$ taking $p = 2$ and $q = 1$ that is βSI^2 have similar repertoires of dynamical behaviors, much wider than of bilinear incidence rate models, and we study the existence and stability of the equilibriums of the $SEIS$ epidemic model.

This manuscript is organized as follows: In Sect. 2, $SEIS$ model is presented. In Sect. 3, basic properties of solutions are discussed. In Sect. 4, we determine all possible equilibria of model. In Sect. 5, we discuss and analyze the local stability of the equilibriums. In Sect. 6, we discuss and analyze the global stability of the equilibriums. We present in Sect. 7, some numerical examples of the dynamics of the model. Finally, in Sect. 8, we discussed the conclusion.

II. Model Formulation

By a standard nonlinear incidence rate βSI^2, we consider an $SEIS$ epidemic model which consists of three compartments corresponding to three population classes, namely, susceptible (S), exposed (but not yet infectious) (E), infectious (I) and the total population (N).
The model is given as follows:

$$\left.\begin{aligned}\frac{dS}{dt} &= A-\beta I^2 S-dS+\gamma I\\ \frac{dE}{dt} &= \beta I^2 S-(d+\varepsilon)E\\ \frac{dI}{dt} &= \varepsilon E-(d+\gamma+\alpha)I\end{aligned}\right\} \tag{2.1}$$

whose state space is the first quadrant $R_3^+=\{(S,E,I): S\geq 0, E\geq 0, I\geq 0\}$ and subject to the initial conditions $S(0)=S_0\geq 0$, $E(0)=E_0\geq 0$, $I(0)=I_0\geq 0$. It is assumed that all the parameters are positive. From the model, the parameters can be summarized in the following list:
A is the recruitment rate of the population, d is the natural death rate of the population, α is the constant rate of infectious hosts suffer an extra disease related death, β is the transmission or contact rate, ε and γ is the transfer rates among the corresponding classes.

III. Basic Properties of the Model

Summing up the four equations of model (2.1) and denoting

$$N(t)=S(t)+E(t)+I(t),$$

Having, $N'(t)=A-dN-\alpha I$. If disease is not present that is $I=0$, then $N'(t)=A-dN$. This shows that population size $N\to\frac{A}{d}$ as $t\to\infty$. Since the spread of the disease in the population will lead to the decrease of N, it follows that $N\in[0,\frac{A}{d}]$. It follows that the solutions of model (2.1) exists in the region defined by

$$\Omega=\left\{(S,E,I)\in R_3^+: S,E,I\geq 0, S+E+I\leq\frac{A}{d}\right\}.$$

This gives the following lemma which shows that the solutions of model (2.1) are bounded, continuous for all positive time and lie in a compact set.

IV. Existence of Equilibria

In this section, we obtain the existence of the disease-free equilibrium E_0 and the endemic equilibrium E^* of model (2.1).
Set the right sides of model (2.1) equal zero, that is,

$$\begin{cases}\frac{dS}{dt} = A-\beta I^2 S-dS+\gamma I=0\\ \frac{dE}{dt} = \beta I^2 S-(d+\varepsilon)E=0\\ \frac{dI}{dt} = \varepsilon E-(d+\gamma+\alpha)I=0\end{cases} \tag{4.1}$$

The model (2.1) always has the disease-free equilibrium point $E_0(A/d,0,0)$. Solving (4.1) we also get a unique positive, endemic equilibrium point $E^*(S^*,E^*,I^*)$of the model (2.1), where $S_*=\frac{(d+\varepsilon)(d+\gamma+\alpha)}{\beta\varepsilon I}$, $E_*=\frac{(d+\gamma+\alpha)}{\varepsilon}I$ and I^* is given as a root of the quadratic equation

$$\Omega_1 I^2 + \Omega_2 I + \Omega_3 = 0,$$

where, $\Omega_1 = \beta(d+\varepsilon)(d+\alpha+\gamma) - \beta\varepsilon\gamma$, $\Omega_2 = -A\beta\varepsilon$ and $\Omega_3 = d(d+\varepsilon)(d+\alpha+\gamma)$.

Now, $I^* = \dfrac{A\beta\varepsilon + \sqrt{\Delta}}{2\beta(d+\varepsilon)(d+\alpha+\gamma) - 2\beta\varepsilon\gamma}$,

where, $\Delta^2 = 4\beta\varepsilon\gamma d(d+\varepsilon)(d+\alpha+\gamma) + 4\beta d(d+\varepsilon)^2(d+\alpha+\gamma)^2[R_0 - 1]$.

Obviously $\Delta_1 > 0$ when $R_0 > 1$.

According to a direct computation, define the basic reproduction number as follows:

$$R_0 = \frac{A^2\beta^2\varepsilon}{4d(d+\varepsilon)^2(d+\gamma+\alpha)^2}.$$

It means the average new infections caused by a single infected individual in a whole susceptible population.

V. Local Stability Analysis

In this section, we study the local stability of the disease-free equilibrium E_0 and the endemic equilibrium E^* of model (2.1).

Theorem 5.1 If $R_0 < 1$, the disease-free equilibrium E_0 of model (2.1) is locally asymptotically stable. If $R_0 > 1$, the disease-free equilibrium E_0 is unstable.

Proof. The Jacobian matrix of model (2.1) at the disease-free equilibrium E_0 is

$$J(E_0) = \begin{pmatrix} -d & 0 & \gamma \\ 0 & -(d+\varepsilon) & 0 \\ 0 & \varepsilon & -(d+\gamma+\alpha) \end{pmatrix}$$

The characteristic equation of $J(E_0)$ is $(d+\lambda)(d+\varepsilon+\lambda)(d+\gamma+\alpha+\lambda) = 0$.

This equation has the following roots: $\lambda_1 = -d$, $\lambda_2 = -(d+\varepsilon)$ and $\lambda_3 = -(d+\gamma+\alpha)$ are always negative. Hence E_0 is locally asymptotically stable for $R_0 < 1$, while it is unstable for $R_0 > 1$.

Theorem 5.2 If $R_0 > 1$, the endemic equilibrium E^* of model (2.1) is locally asymptotically stable.

Proof. The Jacobian matrix of system (2.1) at E^* is

$$J(E^*) = \begin{pmatrix} -(m+d) & 0 & -n+\gamma \\ m & -(d+\varepsilon) & n \\ 0 & \varepsilon & -(d+\gamma+\alpha) \end{pmatrix}$$

where $m = \beta I^{*2}$ and $n = 2\beta S^* I^*$

The characteristic equation of $J(E^*)$ is

$$\lambda^3 + \lambda^2\{(d+\gamma+\alpha)(m+2d+\varepsilon)\} + \lambda\{(d+\gamma+\alpha)(m+2d+\varepsilon) + (m+d)(\varepsilon+d) - \varepsilon n\} + (d+\gamma+\alpha)(m+d)(\varepsilon+d) - \varepsilon(nd+m\gamma) = 0 \quad (5.1)$$

from numerical computation, we realized the real part of equation (5.1) cannot be positive. This indicates that, the steady state(s) will also be stable.

VI. Global Stability Analysis

In this section, we study the global stability of the disease-free equilibrium E_0 and the endemic equilibrium E^* of model (2.1).

Theorem 6.1 If $R_0 < 1$, the disease-free equilibrium E_0 of model (2.1) is globally asymptotically stable.

Proof. We prove the global stability of the model (2.1) at the equilibrium E_0 when $R_0 < 1$. Taking the Lyapunov function

$$V(t) = E(t)$$

Calculating the derivative of $V(t)$ along the positive solution of model (2.1), it follows that

$$\dot{V}(t) = \beta SI^2 - (d+\varepsilon)E = \beta SI^2 - \frac{(d+\varepsilon)(d+\alpha+\gamma)}{\varepsilon} I$$

Since the incidence function

$$\beta SI^2 \le \frac{\beta AI^2}{d} \text{ for } 0 \le S \le \frac{A}{d}.$$

$$\dot{V}(t) \le \left[\frac{\beta AI}{d} - \frac{(d+\varepsilon)(d+\alpha+\gamma)}{\varepsilon} \right] I \le 0$$

Furthermore, $\dot{V} = 0$ only if $I = 0$, so the largest invariant set contained in $\{(S, E, I) \in \Omega : \dot{V} = 0\}$ is the plane $I = 0$. By Lassalle's invariance principle [7], this implies that all solution in Ω approach the plane $I = 0$ as $t \to \infty$. On the other hand, solutions of (2.1) contained in such plane satisfy $\frac{dS}{dt} = A - dS$, $\frac{dE}{dt} = -(d+\varepsilon)E$, which implies that $S \to \frac{A}{d}$ and $E \to 0$ as $t \to \infty$, that is, all of these solutions approach E_0 is globally asymptotically stable in Ω.

Next, we analysis the global stability of an endemic equilibrium E^* by using geometric approach method described by Li and Muldowney in [8]. We first briefly explain the geometric approach method.

Theorem 6.2 (Li & Muldowney [8]). Suppose that the system $x' = f(x)$, with $f : D \subset \tilde{}^n \to \tilde{}^n$, satisfies the following:

(H1) D is a simply connected open set,

(H2) there is a compact absorbing set $K \subset D$,

(H3) x^* is the only equilibrium in D.

Then the equilibrium x^* is globally stable in D if there exists a *Lozinskiĭ* measure η such that

$$\lim_{t\to\infty} \sup \sup_{x_0 \in K} \frac{1}{t} \int_0^t \eta(B(x(s, x_0)))ds < 0,$$

$$B = P_f P^{-1} + PJ^{[2]} P^{-1}$$

and $Q \to Q(x)$ is an $\binom{n}{2} \times \binom{n}{2}$ matrix valued function.

In our case, model (2.1) can be written as $x' = f(x)$ with $f : D \subset \tilde{}^n \to \tilde{}^n$ and D being the interior of the feasible region Ω. The existence of a compact absorbing set $K \subset D$ is equivalent to proving that (2.1) is uniformly persistent (see [8]) and the proof for this in the case when $R_0 > 1$ is similar to that of proposition 4.2 of [8]. Hence, **(H1)** and **(H2)** hold for system (2.1), and by assuming the uniqueness of the endemic equilibrium in D, we can prove its global stability with the aid of *Theorem 6.2*.

Theorem 6.3 If $R_0 > 1$ then the endemic equilibrium E^* of the system (2.1) is globally asymptotically stable in the feasible region Ω..

Proof. Let J be the Jacobian matrix of the system (2.1), i.e.

$$J = \begin{pmatrix} -\beta I^2 - d & 0 & -2\beta IS + \gamma \\ \beta I^2 & -(d+\varepsilon) & 2\beta IS \\ 0 & \varepsilon & -(d+\gamma+\alpha) \end{pmatrix}$$

Then the second additive compound matrix of J is given by

$$J^{[2]} = \begin{pmatrix} -\beta I^2 - 2d - \varepsilon & 2\beta IS & 2\beta IS - \gamma \\ \varepsilon & -\beta I^2 - 2d - \alpha - \gamma & 0 \\ 0 & \beta I^2 & -2d - \alpha - \varepsilon - \gamma \end{pmatrix}$$

Choose the function $P = P(S, E, I) = \text{diag}(1, \frac{E}{I}, \frac{E}{I})$; then $P^{-1} = \text{diag}(1, \frac{I}{E}, \frac{I}{E})$ and $P_f = \text{diag}(0, \frac{E'I - I'E}{I^2}, \frac{E'I - I'E}{I^2})$. Also, $P_f P^{-1} = \text{diag}(0, \frac{E'}{E} - \frac{I'}{I}, \frac{E'}{E} - \frac{I'}{I})$

$$PJ^{[2]}P^{-1}=\begin{pmatrix} -\beta I^2-2d-\varepsilon & \dfrac{2\beta I^2 S}{E} & \dfrac{2\beta I^2 S}{E}-\dfrac{\gamma I}{E} \\ \dfrac{\varepsilon E}{I} & -\beta I^2-2d-\alpha-\gamma & 0 \\ 0 & \beta I^2 & -2d-\alpha-\varepsilon-\gamma \end{pmatrix}$$

The matrix $B=P_f P^{-1}+PJ^{[2]}P^{-1}$ can be written in matrix form

$$B=\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

where

$$B_{11}=-\beta I^2-2d-\varepsilon,\quad B_{12}=\left(\frac{2\beta I^2 S}{E},\frac{2\beta I^2 S}{E}-\frac{\gamma I}{E}\right),\quad B_{21}=\left(\frac{\varepsilon E}{I},0\right)^T,$$

$$B_{22}=\begin{pmatrix} -\beta I^2-2d-\alpha-\gamma+\dfrac{E'}{E}-\dfrac{I'}{I} & 0 \\ \beta I^2 & -2d-\alpha-\varepsilon-\gamma+\dfrac{E'}{E}-\dfrac{I'}{I} \end{pmatrix}$$

Let (u,v,w) be a vector in R^3; its norm $\|.\|$ is defined as $\|(u,v,w)\|=\max\{|u|,|v|+|w|\}$

Let $\mu(B)$ be the Lozinskiĭ measure with respect to this norm.

$$\mu(B)\le \sup\{g_1,g_2\}$$

where

$g_1=\mu_1(B_{11})+|B_{12}|$, $g_2=\mu_1(B_{22})+|B_{21}|$, $|B_{12}|$, $|B_{21}|$ are matrix norm with respect to l_1 vector norm and μ_1 denotes the Lozinskiĭ measure with respect to l_1 vector norm, then $\mu_1(B_{11})=-\beta I^2-2d-\varepsilon$, $|B_{12}|=\max\left(\frac{2\beta I^2 S}{E},\frac{2\beta I^2 S}{E}-\frac{\gamma I}{E}\right)=\frac{2\beta I^2 S}{E}$, $|B_{21}|=\frac{\varepsilon E}{I}$.

Therefore,

$$\left.\begin{aligned} g_1&=\mu_1(B_{11})+|B_{12}|=-\beta I^2-2d-\varepsilon+\frac{2\beta I^2 S}{E} \\ g_2&=\mu_2(B_{22})+|B_{21}|=-2d-\alpha-\gamma+\frac{E'}{E}-\frac{I'}{I}+\frac{\varepsilon E}{I} \end{aligned}\right\}.$$

where

$$\mu_2(B_{22})=\max\{-2d-\alpha-\gamma+\frac{E'}{E}-\frac{I'}{I},-2d-\alpha-\varepsilon-\gamma+\frac{E'}{E}-\frac{I'}{I}\}=-2d-\alpha-\gamma+\frac{E'}{E}-\frac{I'}{I}.$$

From system (2.1) we have $\frac{E'}{E}=\frac{\beta SI^2}{E}-(\varepsilon+d)$ and $\frac{I'}{I}=\frac{\varepsilon E}{I}-(d+\gamma+\alpha)$

Then

$$\begin{aligned} g_1&=-\beta I^2-2d-\varepsilon+\frac{2\beta I^2 S}{E} \\ &\le \frac{\beta I^2 S}{E}-(\varepsilon+2d)\le \frac{E'}{E}-d \\ g_2&=-2d-\alpha-\gamma+\frac{E'}{E}-\frac{I'}{I}+\frac{\varepsilon E}{I}=\frac{E'}{E}-d \end{aligned}$$

Furthermore, obtain

$$\begin{aligned} \mu(B)&\le \sup\{g_1,g_2\} \\ &\le \left\{\frac{E'}{E}-d,\frac{E'}{E}-d\right\}\le \frac{E'}{E}-d \end{aligned}$$

By integrating both sides at the same time,

$$\frac{1}{t}\int_0^t \mu(B)\mathrm{ds}\le \frac{1}{t}\mathrm{In}\frac{E(t)}{E(0)}-d$$

$$q = \lim_{t\to\infty} \sup\sup \frac{1}{t}\int_0^t \mu(B)\mathrm{ds} < -d < 0.$$

Hence, E^* is globally asymptotically stable in Ω.

VII. Numerical Simulations

To see the dynamical behavior of system (2.1) some numerical simulations are given. For this, consider the Hypothetical set of parameter values as the following.

Case I. $A = 10$, $d = 0.2$, $\beta = 0.04$, $\alpha = 1.25, \varepsilon = 1.2$, $\gamma = 0.4$ then the basic reproduction number $R_0 = 0.035777641 < 1$, $S(t)$ approaches to its steady state value while $E(t)$ and $I(t)$ approach zero as time goes to infinity, the disease dies out **(Fig. 1)**.

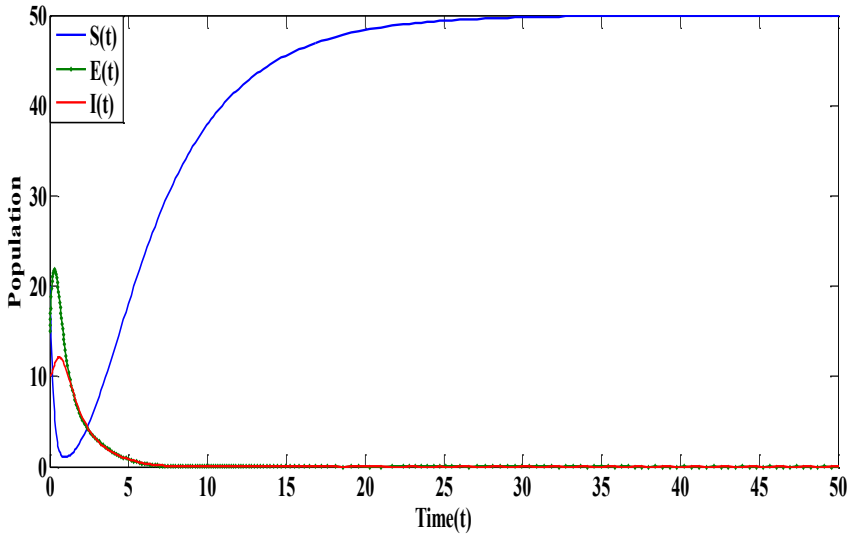


Figure 1: *The figure represents that the disease dies out*

Case II. $A = 10$, $d = 0.2$, $\beta = 1$, $\alpha = 1.25, \varepsilon = 1.2$, $\gamma = 0.4$ then the basic reproduction number $R_0 = 22.36102622 > 1$, all the three component $S(t)$, $E(t)$ and $I(t)$ approach to their steady state values as time goes to infinity, the disease becomes endemic **(Fig. 2)**.

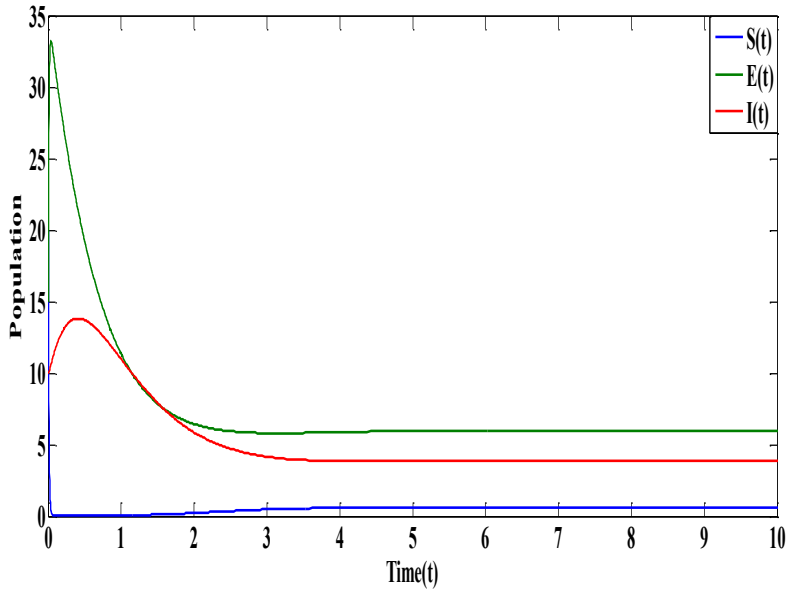


Figure 2: *The figure represents that the disease endemic*

VIII. Conclusion

This paper presents a mathematical study on the dynamics of an $SEIS$ epidemic model that incorporates constant recruitment, exponential natural death as well as the disease related rate, so that the population size may vary in time. The incidence rate is of the non-linear incidences frequently used in the literature. Also, we see that if the basic reproduction number R_0 is less one the disease free equilibrium E_0 is locally and globally asymptotically stable in feasible region Ω and disease always dies out (see **Fig. 1**). If the basic reproduction number R_0 is greater than one the unique endemic equilibrium E^* is locally and globally under certain condition if the interior of Ω. In this case, the disease cannot control easily (see **Fig. 2**).

References

[1] Acedo, L, Gonzalez-Parra, G, Arenas, A. (2010). An exact global solution for the classical epidemic model. *Nonlinear Anal., Real World Appl.,* 11:1819-1825.

[2] Alexander, M. E. and Moghadas, S. M. (2005). Bifurcation analysis of SIRS epidemic model with generalized incidence. *SIAM Journal on Applied Mathematics*, 65 (5):1794-1816.

[3] Fan, M., Michael, Y. Li. and Wang, Ke. (2001). Global stability of an SEIS epidemic model with recruitment and a varying total population size. *J. Math. Biol.* 170:199-208.

[4] Kar T. K, Batabyal, Ashim. (2010). Modeling and analysis of an epidemic model with non-monotonic incidence rate under treatment. *Journal of Mathematics Research*, 2(1):103-115.

[5] Kermack, W. O. and McKendrick A. G. (1927). A Contribution to the Mathematical Theory of Epidemics, *Proc. R. Soc. Lond. A*, 115:700-721.

[6] Jiang, Z, Wei, J, J. (2008). Stability and bifurcation analysis in a delayed SIR model. *Chaos Solitons Fractals*. 35:609-619.

[7] Lasalle, J. P. The stability of dynamical system, (SIAM, Philadelphia, P A; 1976).

[8] Li, M. Y. and Muldowney, J. S. (1996). A Geometric Approach to Global Stability Problem, *SIAM Journal on Mathematical Analysis*. 27(4):1070-1083.

[9]. Liu, W. M., Hethcote, H. W. and Levin, S. A. (1987). Dynamical behavior of epidemiological models with nonlinear incidence rates. *J. Math. Biol.* 25: 359-380.

[10]. Liu, W. M, Levin, S. A. and Iwasa, Y. (1986). Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. *J. Math. Biol.* 23:187-204.

[11]. Song, Y, Pang, T. S. (2013). The SIQR epidemic model with constant input and nonlinear incidence rate. *J. Biomath.* 28:454-460.

[12]. Zhou, X. and Cui , J. (2011). Analysis of stability and bifurcation for an SEIR epidemic model with saturated recovery rate. *Communications in Nonlinear Science and Numerical Simulation.* 16 (11):4438-4450.

[13]. Zhang, F, Li and Zhang, F , Z. (2008). Global stability of an SIR epidemic model with constant infectious period. *Appl. Math. Comput.* 199:285-291.

[14]. Wang, W and Ruan, S. G. (2004). Bifurcation in epidemic model with constant removal rate infective. *J. Math. Anal. Appl.* 291:775-793.

[15]. Wang, J. J, Zhang, J. Z and Jin, Z. (2010). Analysis of an SIR model with bilinear incidence rate. *Nonlinear Anal., Real World Appl.* 11:2390-2402.

[16]. Zhang J, Jia J, Song X. (2014). Analysis of an SEIR epidemic model with saturated incidence and saturated treatment function. *The Scientific World Journal*. 2014:1-11.

[17]. Xiao, D and Runa, S. (2007). Global analysis of an epidemic model with non-monotonic incidence rate. *Mathematical Bioscience.* 208:419-429.

Sampled Ready Queue Processing Time Estimation Using Size Measure Information In Multiprocessor Environment

Sarla More and Diwakar Shukla
•
Dr. Harisingh Gour University, Sagar, MP, India
sarlamore@gmail.com, diwakarshukla@rediffmail.com

Abstract

In a multiprocessor computer system, there exist a ready queue of large number of processes waiting for computing resources allocation by the processors. These jobs may have size measure, which are additional information priory known while entry to the ready queue. Suppose the sudden system breakdown occurs and recovery management is required immediately. At this stage, one can find some processes who are completely finished, some partially processed, some blocked by processors and remaining waiting for allocation in the ready queue. Prime act of a system manager is to evaluate the maximum time required to process all the remaining jobs. This paper presents an estimation strategy for such, derived by applying the lottery scheduling, sampling technique and imputation methodology. Expressions for mean squared error of the proposed strategy are derived and optimized for suitable selection of system parameters. Three cases are discussed and compared and consequent results are numerically supported. It is found that at the optimal choice of constants in the estimation methodology, the shortest confidence interval can be predicted estimating the remaining required time. Such findings are useful as a part of disaster management of a cloud based multiprocessor data centre.

Keywords: Ready Queue, Lottery scheduling, Multiprocessors, Simulation, Sampling, Random, Estimation

I. Introduction

Assume a computer system equipped with several processors having a ready queue, where processes are waiting for allocation of resources. Lottery scheduling is a type of priority scheduling where the computer system resources are allocated randomly to the waiting processes. In this, a bunch of token numbers are assigned to processes and multi-processors used to issue random numbers. A process who contains the token of issued number receives first the desired system resource. Using this, every process has chance of allocation of resources, sooner or later, therefore, probability of starvation vanishes. A process in a ready queue may have a predetermined measure of its size in terms of bytes which is additional information, can be used for better prediction of the remaining mean time of ready queue processing. This paper presents a strategy for effective use of known size measure of processes.

Let $t_1, t_2, t_3 \ldots\ldots t_k$ be the time of k processes and $x_1, x_2, x_3 \ldots\ldots x_k$ be their size measure. Figure 1 shows r processors ($r < k$) and waiting queue.

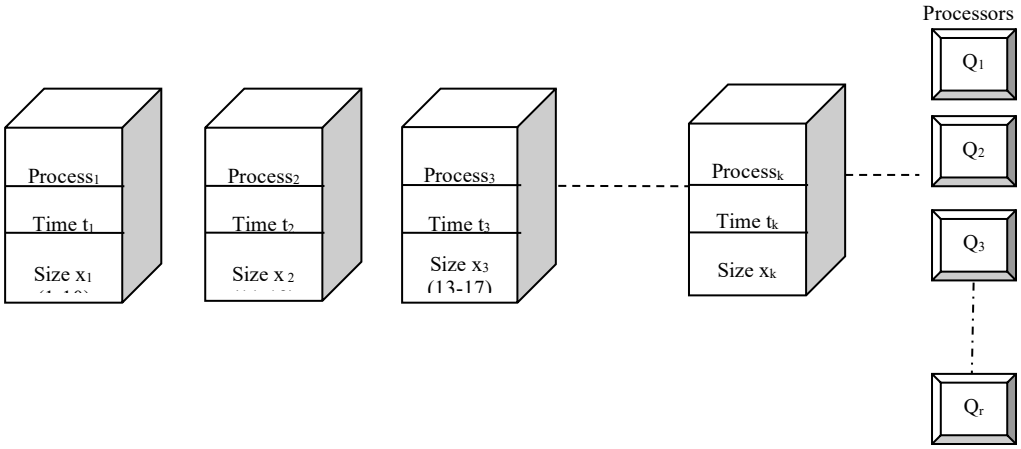


Figure 1: *Ready queue with waiting Processes and Multiprocessor*

Further assume processes A_1, A_2, A_3…….....are of small size with time consumption t_{11}, t_{12}, t_{13}……..... and size measure x_{11}, x_{12}, x_{13}…... (see figure 2). The large size processes areB_1, B_2, B_3……..... with time consumption t_{21}, t_{22}, t_{23}……..... and size measure x_{21}, x_{22}, x_{23}…….....…(see figure 3). All the A_i and B_i are to be processed by r processors, under the size measures. The case of partially processed and completely processed [23] exists when sudden breakdown occurs. One can further think of possibilities as under:

(a) inside multi-processors, some are completely processed,
(b) some are partially processed,
(c) some are blocked, and
(d) size measure of processes are known.

This paper extends approach of [23] in collective presence of (a), (b), (c) & (d) when multiprocessor computer system fails at an instant. The issue of estimation (prediction) of recovery time duration of the remaining is focused.

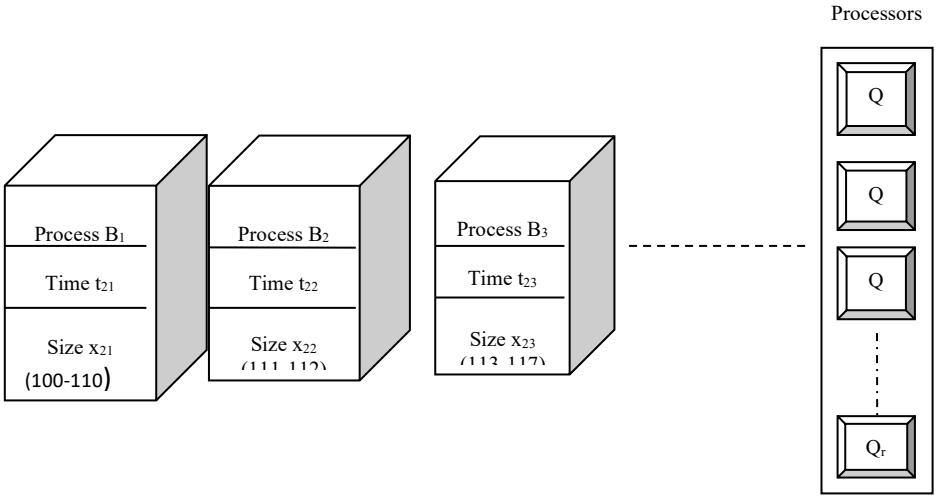


Figure 2: *Small size processes and Multiprocessors*

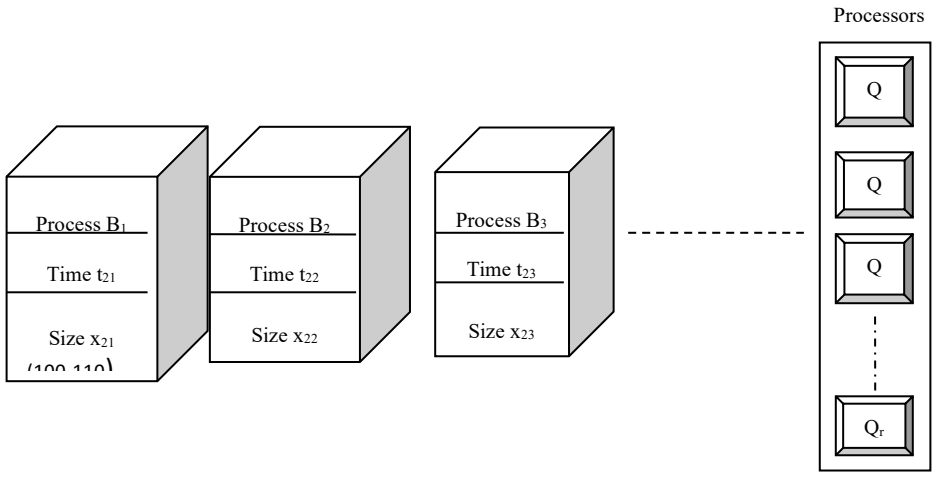


Figure 3: *Big size processes and Multiprocessors*

II. A Review

Lottery scheduling is a resource sharing technique [12] like a particular case of priority scheduling where the processes in ready queue are allotted bunches of ticket numbers. Process who receives maximum count of tickets has highest priority of being allocated the demanded system resources. Lottery scheduling is efficient and effective [22] in the framework of LINUX kernel also. It could be used as a model tool [4], [6] for estimating the mean time of processing of a ready queue where large number of jobs are in waiting but only some have processed. Completed jobs could be used as a sample just like a preliminary source of information for prediction. Concept of grouping of homogeneous jobs together [5] came into existence which has improved the prediction. Units in sample may have additional correlated variables which could be utilized for efficient computing ([7], [8], [9], [10], [11]) along with precise prediction. An exhaustive review [3] on the similar problem contributes few recent aspects of solutions extendable into [1] and [2]. Some authors have extended the lottery scheduling variants [20], [21] in the form of hybrid multi-level structure using Markov chain model along with analysis and chance based prediction.

Sampling techniques are useful tools for parameter estimation and value prediction. Random sampling schemes exist in statistical literature ([13], [15], [16]) who are widely used for parameter evaluation of a finite collection. Some popular schemes are like stratified sampling, cluster sampling, two stage sampling, systematic sampling, successive sampling etc.([14], [17], [18]) useful in varying situations of the aggregate . Moreover, such needs appropriate selection of methods also [19] to provide accurate confidence interval for unknown parameter.

Imputation is a methodology used when one or more values in a sample are found missing (or non-responded). For example, if a processor blocks a process then mean time parameter remains unpredicted using sample from the ready-queue. However, some processes may be blocked after the partial processing. For completely blocked processes, Random Imputation methods ([25], [26], [27]) could be used to recover information. In this, the missing values are selected randomly from the available part of sample values and replaced. Other popular imputation methods are mean imputation, deductive imputation, mean imputation within classes, deductive imputation within class, hot deck imputation, cold deck imputation etc. ([28], [29], [30], [31]). This paper considers the approach of [6] and [23] and extends using [10], [11], assuming situation when one process is blocked, one is partially processed and remaining others in a

processor are completed before the occurrence of sudden breakdown.

I. Remaining Time Estimation Problem

Assume a large number of processes (say N) present in a ready queue of a multiprocessor computer system and only few of them (say n, n<N) have been processed before a fixed time instant. The remaining in the ready queue are (N-n) for whom the expected time computation is required. If sample mean time of those who already processed is Δ then remaining time estimate is δ = [(N-n) Δ] which is an unknown quantity. For any two real numbers 'a' and 'b', if Δ is predicted as $\Delta \in$ (a,b) where (a,b) is an interval containing Δ with very high probability, then δ_1 =[(N-n) a]is lowest, δ_2 =[(N-n) b] is the value of highest expected time. If highest expected time is precisely estimated then it could be used for backup management during system failure. The efficient estimation of this expected range is a problem which is undertaken in this paper for strategy formation in the multiprocessor setup with the consideration of multiple real life possibilities.

II. **Confidence Interval (CI)**

It is a statistical tool for evaluating the precision of mean time estimate. If catches the true unknown value then it is termed as a confidence interval. Let P[A] denotes the probability of happening of event A. In statistical theory, for any two real numbers a', b', the 95% confidence interval is defined as P[a' < true unknown value < b'] = 0.95**.** Define length of CI= l = (b'-a'). Let one confidence interval has length $l_1$obtained through a methodand other has length $l_2$obtained by another method. If $l_2 < l_1$then second one is said to be better than the first in terms of efficient prediction.

III. **Motivation**

Earlier contributions (specially [6], [23]) were under assumption that processes present in a multiprocessors system are completely processed before sudden failure. But this is not a practical reality. While sudden failure, some jobs may complete, some may partially processed and some may blocked by the processors [see figure 4]. The processed and unprocessed case was considered in [23] [see figure (5)]. This paper extends the approach of [23] by applying the tools of random imputation method against the blocked processes.

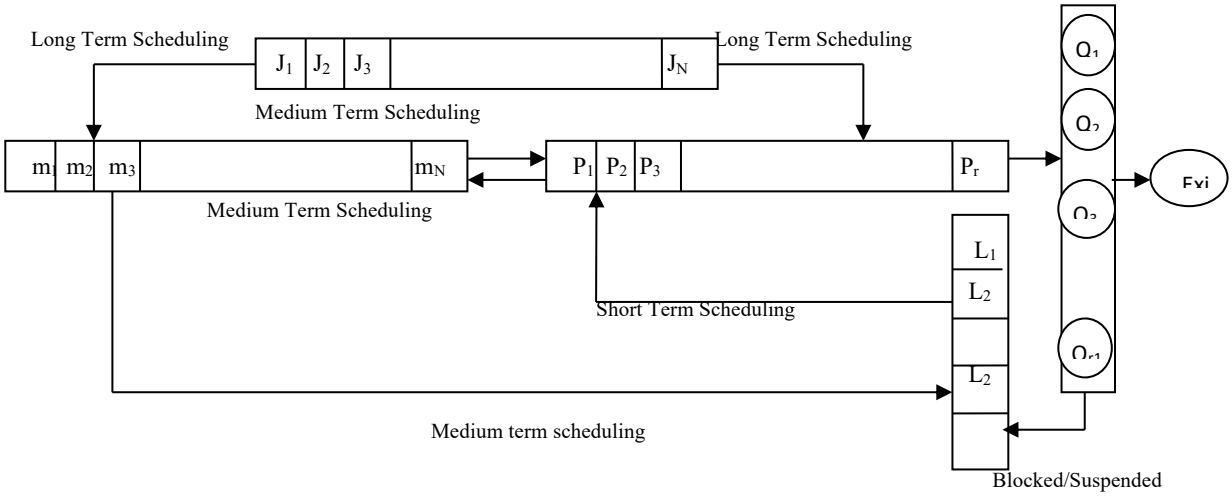


Figure 4: *Ready Queue Processing under Lottery Scheduling (due to [6])*

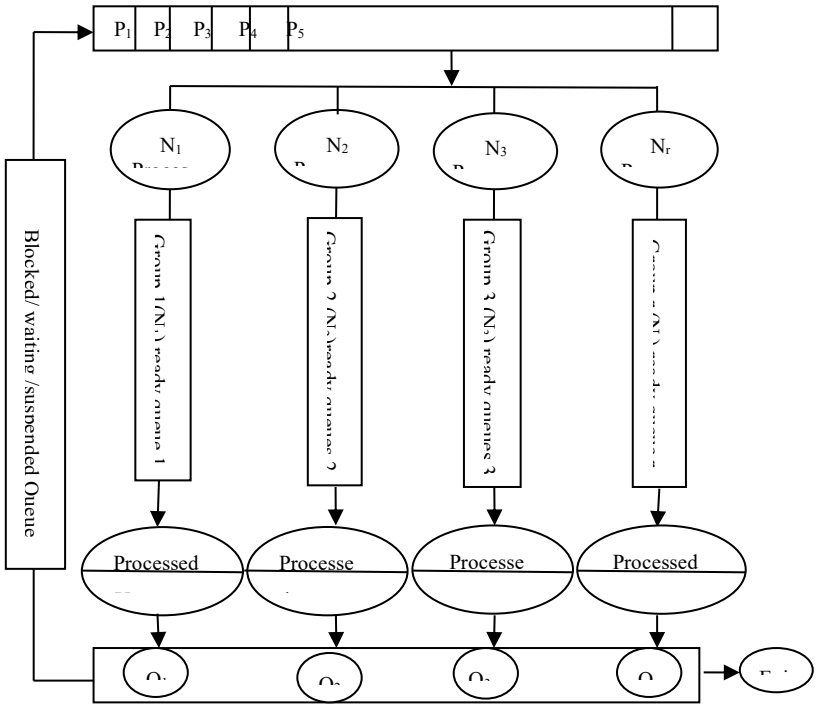


Figure 5: *Setup of ready queue and multiprocessor environment (due to[23])*

III. Proposed Computational Setup

Assume the existence a virtual sampled ready queue in a system of multiprocessors environment. Some jobs are randomly selected using lottery scheduling from the ready queue and placed in the sampled ready queue from top to bottom in the sequential manner of their selection. Processors are assigned processes in the ordered manner from top to bottom ofthe virtual sampled ready queue. Figure 6 shows basic setup of this approach without size measure while figure 5 shows the earlier approaches [4], [5], [6], [23]. Moreover, figure 7 reveals the special case when all sample units processed before the occurrence of breakdown.

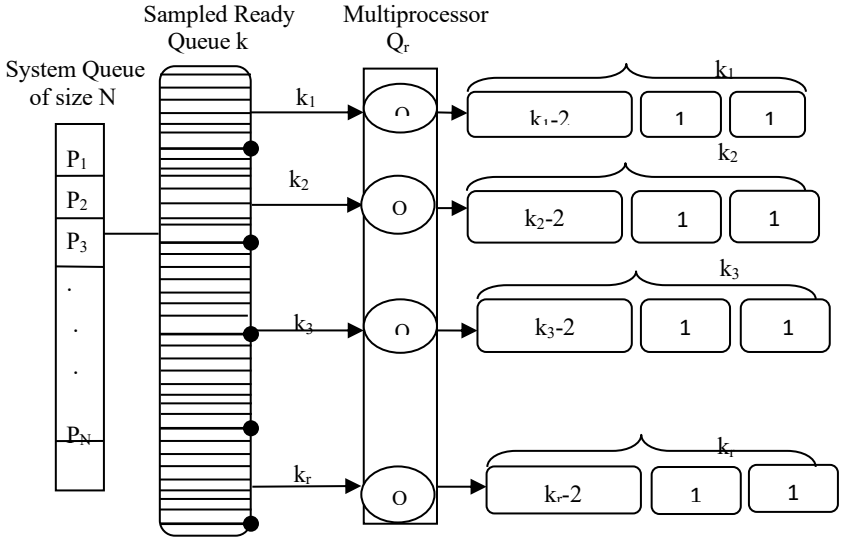


Figure 6: *Sampled Ready Queue Processing Time Estimation setup without size measure*

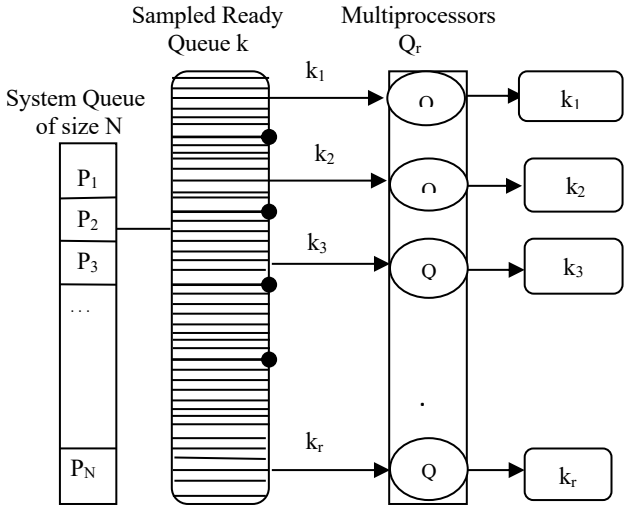


Figure 7: *Sampled Ready Queue Processing Time Estimation when all processed before breakdown*

I. Assumption and Model

In view of figure 6, let the selection of processes is as per priority scheduling, in particular, as per lottery scheduling. The process who selects first is placed at the top of the virtual queue who is segment or group of processes likely to allocate to the multi-processors.

1). Assume r processors in system and a ready queue of N processes denoted as [P_1, P_2, P_3…….....P_N]who are waiting for allocation of resources.

2). The selection of process for resource allocation is on priority basis using lottery scheduling.

3). If all N are processed completely then time consumed by them are [t_1, t_2, t_3,…….t_N] and each process has size measure [x_1, x_2, x_3,…….x_N] who are priory known.

4). Define the whole ready queue mean time$\bar{t}= \frac{1}{N}\sum_{i=1}^{N} t_i$, size measure mean$\bar{x}= \frac{1}{N}\sum_{i=1}^{N} x_i$ and respective mean squares $S_t^2 = \frac{1}{N-1}\sum_{i=1}^{N}(t_i - \bar{t})^2$, $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$
The process P_i of known size X_i consumes time t_i(i = 1,2,3,……N).

5). Hereby denote r multiprocessors as Q_1, Q_2, Q_3…..Q_r, (r < N) and time consumed by the i^{th} process in the j^{th} processor is t_{ij} with corresponding size measures x_{ij} (j = 1,2,3,……r)

6). Total completion time of ready queue is $N\bar{t}$, which is an unknown quantity. This paper is focused to estimate such using sampling methodology. Lottery scheduling is a tool for such estimation where process P_i has a bunch of token numbers and Q_j generates a random number. A process who receives the random number gets the desired resource from Q_j.

7). A virtual ready queue of size k (k < N, k>3r) exists to store sequentially the records of randomly selected k processes from N. The j^{th} segment of virtual sampled queue is k_j(k =$\sum_{j=1}^{r} k_j$), who is allocated to the j^{th} processor Q_j in sequential manner.

8). In sample let sx_{jl} denotes the file size measure and st_{jl} denotes time consumed by i^{th} process in Q_j(l = 1,2,3,...k_j) when all processed completely who are included in the sample of size k.

(i). Sample mean of time $\overline{st}= \frac{1}{k}\sum_{j=1}^{r}\sum_{l=1}^{k_j} st_{jl}$

(ii). Sample mean square of time, $(es)_t^2 = \frac{1}{k-1}\sum_{j=1}^{r}\sum_{l=1}^{k_j}(st_{jl} - \overline{st})^2$

(iii). Sample mean of size, ($\overline{sx}$)= $\frac{1}{K-1}\sum_{j=1}^{k_j}\sum_{l=1}^{k_j}(sx_{jl})$

(iv). Sample mean square of size, $(es)_x^2 = \frac{1}{k-1}\sum_{j=1}^{r}\sum_{l=1}^{k_j}(sx_{jl} - \overline{sx})^2$

The term$\overline{st}$, $\overline{sx}$, $(es)_t^2$, $(es)_x^2$hold when system runs without failure.

9). Assume system breakdown occurs at the time instant T and there are (k_j - 2) processes who are finished in Q_j, but one remain partially processed and one remain unprocessed (blocked). This is an assumed model sh own in fig. 6 and fig.8.

10). Let $(st')_{jl}$ is time consumed by the l^{th} process in the processor Q_j [l =1, 2, 3... (k_j - 2)],who is among those processed completely before the occurrence of T.

11). Some sample mean related measures are:

(i). Sample mean of (k_j - 2) process, $(\bar{st}')_j = \frac{1}{(k_j-2)} \sum_{l=1}^{(k_j-2)} (st'_{jl})$

(ii). Sample mean square, $(es')_j^2 = \frac{1}{(k_j-3)} \sum_{l=1}^{k_j-2} (st'_{jl} - (\bar{st}')_j)^2$

(iii). Similar is for size measure also as (sx'_{jl}) represents size of l^{th} process who is in Q_j before T.

(iv). Sample mean, $(\bar{sx}')_j = \frac{1}{(k_j-2)} \sum_{l=1}^{k_j-2} (sx'_{jl})$

(v). $(\bar{sx})_j = \frac{1}{(k_j)} \sum_{l=1}^{k_j} (sx'_{jl})$ is sample mean of all k_j known values related to x in j^{th} segment of ready queue.

(vi). Sample mean square, $(ex')_j^2 = \frac{1}{(k_j-3)} \sum_{l=1}^{k_j-2} (sx'_{jl} - (\bar{sx}')_j)^2$

(vii). Sample Covariance, $(es'x')_j = \frac{1}{(k_j-3)} \sum_{l=1}^{k_j-2} (st'_{jl} - (\bar{st}')_j)(sx'_{jl} - (\bar{sx}')_j)$

12). Assume t_m^* is partially processed time of a process in Q_j (j = m =1,2,3....r) whose sample mean under T is

$(\bar{t}^*/T) = \frac{1}{r} \sum_{m=1}^{r} t_m^*$

Variance $(\bar{t}^* / T) = V(\bar{t}^* / T) = (\frac{1}{r} - \frac{1}{N-k+r})\, S_T^2$, where S_T^2 is the conditional ready queue mean square of the remaining unsampled part [N-K+r] expressed as:

$S_T^2 = \frac{1}{(N-k+r-1)} \sum_{i=i}^{N-k+r-1} (t_i - \bar{t}_T)^2$ where

$$\bar{t}_T = \frac{1}{N-k+r} \sum_{i=1}^{N-k+r} (t_i)$$

Herein to mention that S_T^2 and $\bar{t}_T$ contain time t_i only from non-sampled processes (N-k) of the main ready queue with the addition of those r who partially processed. For such, the size converts from N into (N-k+r) and only those processes are the part of $\bar{t}_T$ and S_T^2 who are in (N-K+r).

13). The r blocked processes are imputed by Random Imputation Method using random selection of a process among (k_j-2) relating to Q_j. Let for Q_j this imputed time is denoted as t_m^{**}.

(i). Sample mean of imputed time, $\bar{t}^{**} = \frac{1}{r} \sum_{m=1}^{r} t_m^{**}$

(ii). Variance of imputation under T, $V(\bar{t}^{**}/T) = \left(\frac{1}{r} - \frac{1}{k}\right) (es)^2$, r < k.

14). Sample based estimate of $(es)^2$ can be obtained by using all k values of time consumption in sample including the partially processed time t_m^* and imputed time value t_m^{**}. It is denoted as $(es^*)^2$ and mathematically expressed as

$(es^*)^2 = \frac{1}{k-1} \sum_{j=1}^{r} \sum_{l=1}^{k_j} (st^*_{jl} - \bar{st}^*)^2$ where (st^*_{jl}) and $\bar{st}^*$ include completely processed time st^*_{ij}, partially processed t_m^* and imputed t_m^{**}.

15). The sample estimate of S_T^2 is $(es')^2 = \frac{1}{r-1} [\sum_{m=1}^{r} (t_m^* - \bar{t}^*)^2]$

16). Bias of estimation strategy is assumed negligible wherever appears and applicable in mathematical expressions.

IV. Computational Set-up

The objective is to compute the remaining ready queue processing time while occurrence of

sudden failure of system at time instant T. This is subject to condition that r processes are partially processed, r are unprocessed (blocked) and remaining (K-2r) are fully completed. Blocked and partially processed are one each from every Q_j and the available size measures are the part of computation. Some frequently used symbols for process time t and process size measure X are as under:

$\bar{t}= \frac{1}{N}\sum_{i=1}^{N} t_i = \frac{1}{N}\sum\sum t_{ij}$ …(4.1)

$\bar{t}^{*}= \frac{1}{r}\sum_{m=1}^{r} t_m^{*}$ …(4.2)

$\bar{t}^{**}= \frac{1}{r}\sum_{m=1}^{r-1} t_j^{**}$ …(4.3)

$(\bar{st}')_j= \frac{1}{(k_j-2)}\sum_{j=1}^{k_j-2}(st'_{jl})$ …(4.4)

$(\bar{sx}')_j= \frac{1}{(k_j-2)}\sum_{j=1}^{k_j-2}(sx'_{jl})$ …(4.5)

$(\bar{sx})_j= \frac{1}{(k_j)}\sum_{l=1}^{k_j}(sx'_{jl})$ …(4.6)

$(es')_j^2 = 1/(k_j-3)\sum_{l=1}^{k_j-2}(st'_{jl}-(\bar{st}')_j)^2$ …(4.7)

$(ex')_j^2 = 1/(k_j-3)\sum_{l=1}^{k_j-2}(sx'_{jl}-(\bar{sx}')_j)^2$ …(4.8)

$(es'x')_j= \frac{1}{(k_j-3)}\sum_{l=1}^{k_j-2}(st'_{jl}-(\bar{st}')_j)(sx'_{jl}-(\bar{sx}')_j)$ …(4.9)

$(es^{*})^2 = \frac{1}{k-1}\sum_{j=1}^{r}\sum_{l=1}^{k_j}(st^{*}_{jl}-\bar{st}^{*})^2$ …(4.10)

$R_{N_j}=\left[\frac{(\bar{st}')_j}{(\bar{sx}')_j}\right]$ …(4.11)

$\bar{t}_{rj} = [(\bar{st}')_j\left(\frac{(\bar{sx})_j}{(\bar{sx}')_j}\right)\alpha_j]$, α_j being constant, $(0<\alpha_j<\infty)$ …(4.12)

I. Estimation Strategy

The sample based proposed estimation strategy for mean time is:

$(t_{mean}/T) = €_1[\sum_{j=1}^{r} w_j\ (\bar{t}_{rj}/T)] + €_2\ (\bar{t}^{*}/T) + (1-€_1-€_2)\ (\bar{t}^{**}/T)$

with condition that $\sum_{p=1}^{3} €_p = 1$ and $€_p$ denotes constants to be determine suitability and $w_j = (k_j/k)$ is known weight ($\sum w_j = 1$).

With the help of Cochran [16; see page 166, page 27, 29] for t_{mean}, the expected value E[.] is expressed as:

$E[t_{mean}/T] = E[€_1[\sum_{j=1}^{r} w_j\ (\bar{t}_{rj}/T)] + €_2\ (\bar{t}^{*}/T) + (1-€_1-€_2)\ (\bar{t}^{**}/T)]$

$= €_1[\sum_{j=1}^{r} W_j E\ (\bar{t}_{rj}/T)] + €_2 E\ (\bar{t}^{*}/T) + (1-€_1-€_2)E\ (\bar{t}^{**}/T)]$

$\neq \bar{t}$ which shows estimator (t_{mean}/T) is biased.

II. Mean Squared Error

Let MSE (.), V (.) and B (.) denote mean squared error, variance and bias respectively. One can express

$MSE(t_{mean}/T) = Variance(t_{mean}/T) + [Bias(t_{mean}/T)]^2$ which holds in general.

Assume the bias is small, therefore negligible (as in assumption no. 16)

$MSE(t_{mean}/T) = Variance(t_{mean}/T)$

$= €_1^2[\sum_{j=1}^{r} w_j^2\ MSE(\bar{t}_{rj}/T)] + €_2^2 V\ (\bar{t}^{*}/T) + (1-€_1-€_2)V\ (\bar{t}^{**}/T)]$

$= €_1^2[\sum_{j=1}^{r}\left(\frac{1}{(k_j-2)}-\frac{1}{k}\right) w_j^2\{(es')_j^2+\alpha_j^2 R_{Nj}^2(ex')_j^2 - 2\alpha_j R_{Nj}(es'x')_j\}] + €_2^2[\left(\frac{1}{r}-\frac{1}{N-k+r}\right)s_T^2]$

$+(1-€_1-€_2)^2\sum_{j=1}^{r}\left(1-\frac{1}{K_j-2}\right) w_j(es')_j^2$ (as per Cochran[16] page24, page29 and page164)

The expressions P, Q, R are in the sample based estimate form of population parameters

Let $P = \sum_{j=1}^{r} \left(\frac{1}{(k_j-2)} - \frac{1}{k}\right) w_j^2 \{ (es')_j^2 + \alpha_j^2 R_{Nj}^2 (ex')_j^2 - 2\alpha_j R_{Nj} (es'x')_j \}$

$Q = \left(\frac{1}{r} - \frac{1}{N-k+r}\right) S_T^2$

$R = \sum_{j=1}^{r} \left(1 - \frac{1}{K_j - 2}\right) w_j^2 (es')_j^2$

The above expression is re-written as:

$V[t_{mean/T}] = [€_1^2 P + €_2^2 Q + (1 - €_1 - €_2)^2 R]$ ignoring the covariance terms due to independency. For optimum variance, differentiate $V[t_{mean}/T]$ with respect to $€_1$ and $€_2$and equate to zero, one gets

$(€_1)_{opt} = (QR) / [PQ+PR+QR] = QM$

$(€_2)_{opt} = PQ/ [PQ+PR+QR] = PM$ where $M = R/ [PQ+PR+QR]$

One can differentiate the variance expression by $\alpha_j$also to get optimum value which is $(\alpha_j)_{opt} = [(es'x')_j / (R_{Nj} * (ex')_j^2)]$ Substituting optimum choices in expression, the optimum variance is: $V[t_{mean}/T]_{opt} = (€_1)^2_{opt} P + (€_2)^2_{opt} Q + (1 - (€_1)_{opt} - (€_2)_{opt})^2 R]$ with $(\alpha_j)_{opt}$ as above.

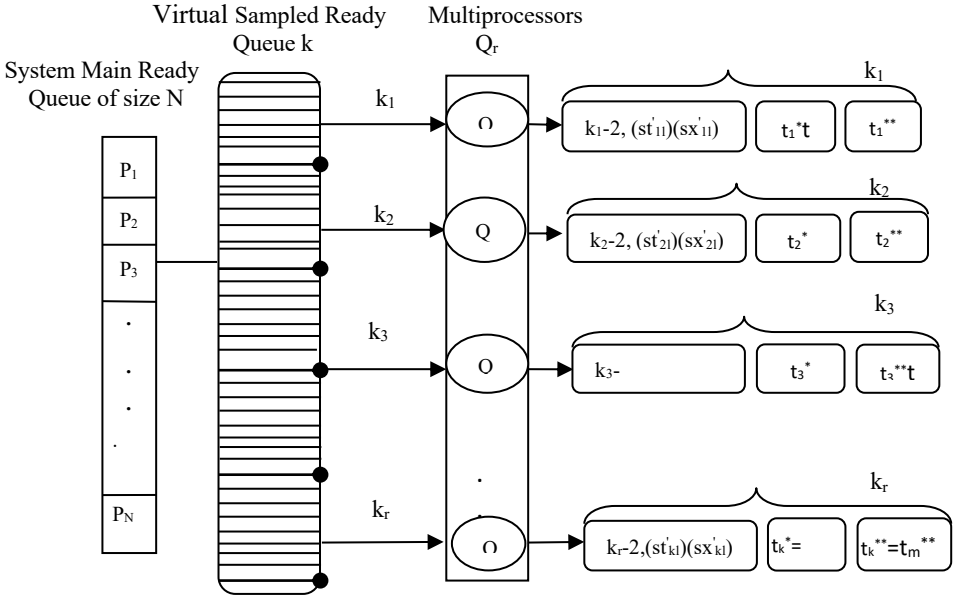


Figure 8: *Proposed Model of Virtual sampled Ready Queue Processing Time Estimation with size and Imputation*

V. Numerical Illustrations

Consider the 150 processes with processed CPU time whose details are in table 1 with assumption that all 150 processes have been completed.

Table 1: *System Ready Queue Processes with time (N = 150)*

Process	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}	J_{15}
CPU Time	30	20	42	45	59	35	25	48	50	60	32	55	62	47	69
Process Size	41	71	103	142	316	82	199	163	220	127	76	192	251	52	133
Process	J_{16}	J_{17}	J_{18}	J_{19}	J_{20}	J_{21}	J_{22}	J_{23}	J_{24}	J_{25}	J_{26}	J_{27}	J_{28}	J_{29}	J_{30}
CPU Time	34	24	44	70	57	65	38	84	101	66	80	90	92	111	85
Process Size	318	202	106	181	242	148	46	252	136	222	261	97	109	271	116
Process	J_{31}	J_{32}	J_{33}	J_{34}	J_{35}	J_{36}	J_{37}	J_{38}	J_{39}	J_{40}	J_{41}	J_{42}	J_{43}	J_{44}	J_{45}
CPU Time	61	52	72	75	89	67	51	78	80	91	63	86	93	77	99
Process Size	172	243	253	262	83	203	183	166	219	193	223	272	281	301	289

Process	J_{46}	J_{47}	J_{48}	J_{49}	J_{50}	J_{51}	J_{52}	J_{53}	J_{54}	J_{55}	J_{56}	J_{57}	J_{58}	J_{59}	J_{60}
CPU Time	64	54	74	100	87	95	68	114	131	96	110	123	122	141	49
Process Size	205	244	223	254	146	263	53	218	273	139	282	302	173	309	290
Process	J_{61}	J_{62}	J_{63}	J_{64}	J_{65}	J_{66}	J_{67}	J_{68}	J_{69}	J_{70}	J_{71}	J_{72}	J_{73}	J_{74}	J_{75}
CPU Time	118	81	102	105	119	97	88	108	110	121	240	113	122	107	129
Process Size	313	194	153	255	225	169	206	264	58	274	283	303	184	291	216
Process	J_{76}	J_{77}	J_{78}	J_{79}	J_{80}	J_{81}	J_{82}	J_{83}	J_{84}	J_{85}	J_{86}	J_{87}	J_{88}	J_{89}	J_{90}
CPU Time	94	73	104	130	117	234	98	237	161	126	143	236	152	171	233
Process Size	207	246	228	360	256	275	217	265	226	195	284	292	304	300	280
Process	J_{91}	J_{92}	J_{93}	J_{94}	J_{95}	J_{96}	J_{97}	J_{98}	J_{99}	J_{100}	J_{101}	J_{102}	J_{103}	J_{104}	J_{105}
CPU Time	120	112	132	135	149	125	115	138	140	150	122	232	152	137	159
Process Size	247	79	208	276	285	257	56	293	266	187	305	178	310	299	215
Process	J_{106}	J_{107}	J_{108}	J_{109}	J_{110}	J_{111}	J_{112}	J_{113}	J_{114}	J_{115}	J_{116}	J_{117}	J_{118}	J_{119}	J_{120}
CPU Time	124	114	134	160	147	155	128	174	191	156	170	180	182	201	175
Process Size	277	286	211	248	227	294	157	258	229	267	196	298	188	306	270
Process	J_{121}	J_{122}	J_{123}	J_{124}	J_{125}	J_{126}	J_{127}	J_{128}	J_{129}	J_{130}	J_{131}	J_{132}	J_{133}	J_{134}	J_{135}
CPU Time	235	142	162	165	179	151	145	168	171	238	152	175	189	167	241
Process Size	287	278	295	197	249	307	268	311	213	350	112	314	259	297	230
Process	J_{136}	J_{137}	J_{138}	J_{139}	J_{140}	J_{141}	J_{142}	J_{143}	J_{144}	J_{145}	J_{146}	J_{147}	J_{148}	J_{149}	J_{150}
CPU Time	154	144	164	190	177	185	158	204	221	186	200	210	212	231	209
Process Size	214	250	260	279	288	296	308	269	312	245	317	198	319	315	239

Table 2: *Descriptive Statistics of Table 1*

S.No.	Parameters Name	Calculated value
1	Number of Processes N	150
2	Mean time ($\bar{t}$)	122.51
3	Total sum of square =$\sum t_i^2$	2697717
4	Mean square St^2	3080.62

Assume that there are three processors Q_1, Q_2, Q_3in the system (r = 3) and a random sample of k = 30 is drawn from N = 150 by lottery scheduling. The sample k = 30 is divided into k_1 = 12, k_2= 10, k_3 = 8 in sequential manner for virtual sampled ready queue. The k_j process are assigned to Q_j(j = 1, 2, 3). Calculation is performed on 10 random samples each of size 30. Computation for only one sample is presented below:

I. CASE I: $\alpha_j = 0$ ($\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$)

Calculation for Sample No. 1 where sample size k=30

k_1:($J_{01,}$ 30,41), ($J_{31,}$61,172), ($J_{61,}$118,313), ($J_{91,}$120,247), ($J_{121,}$235,287), ($J_{63,}$102,153), ($J_{32,}$52,243), ($J_{62,}$81,194), ($J_{92,}$112,79), ($J_{122,}$142,278), ($J_{3,}$42,103), ($J_{33,}$72,253)

Partial Processed = ($J_{03,}$ 42), (Processed= 22, unprocessed=20), Blocked = ($J_{33,}$72),

Blocked replaced α' =($J_{8,}$48)

[$\bar{st}_1'$ = 104.9, from eq.(4.4),$(es')_1^2$= 3317.65, from eq.(4.7)],

[$\bar{sx}_1$=2363/12 = 196.91, from eq. (4.5),$\bar{sx}_1'$=2007/10= 200.7, from eq. (4.6)], $(ex')_1^2$=8158.45,

from eq.(4.8)] [R_{N_1}=$\left[\frac{(\bar{st}')_1}{(\bar{sx}')_1}\right]$ =0.52,from eq.(4.10)],[(es'x')$_1$ = 2982.52,from eq.(4.9)]

k2:(J_{49},100,254), (J_{34},75,262), (J_{64},105,255), (J_{94},135,276), (J_{124},165,197), (J_{135},241,230), (J_{35},89,83), (J_{65},119,225) (J_{95},149,285), (J_{125},179,249)

Partial Processed=(J_{95},149),(Processed=100, unprocessed=49), Blocked=(J_{125},179),

Blocked replaced β' =(J_{38}, 78)

[$\bar{st}_2'$ = 128.62, from eq. (4.4), $(es')_2^2$= 2843.98 from eq.(4.7)]

[$\bar{sx}_2$=2316/10 = 231.6, from eq. (4.5),$\bar{sx}_2' = \frac{1782}{8} =$ 222.75, from eq. (4.6), $(ex')_2^2$=3806.21, from eq.(4.8)]

[$R_{N_2}=\left[\frac{(\bar{st}')_2}{(\bar{sx}')_2}\right]$= 0.57, from eq.(4.10)], [(es'x')$_2$ = 281.75, from eq.(4.9)]

k3:(J_{29}, 111,271), (J_{59}, 141,309), (J_{89}, 171,300), (J_{96},125,257), (J_{119},201,306), (J_{149}, 231,315), (J_{67}, 88,206), (J_{97}, 115, 56)

Partial Processed = (J_{67}, 88), (Processed=40, unprocessed=48), Blocked = (J_{97}, 115),

Blocked replaced γ'=(J_{10}, 60)

[$\bar{st}_3'$ = 163.33, from eq.(4.4), $(es')_3^2$= 2152.66 from eq.(4.7)]

[$\bar{sx}_3$=2020/8=252.5,from eq. (4.5),$\bar{sx}_3' = \frac{1758}{6} =$ 293, from eq. (4.6), $(ex')_3^2$= 547.6, from eq.(4.8)]

[$R_{N_3}=\left[\frac{(\bar{st}')_3}{(\bar{sx}')_3}\right]$=0.55, from eq.(4.10)],[(es'x')$_3$ = 841.2, from eq.(4.9)]

$\bar{t}^*$ = (22+100+40)/3 = 54

$\bar{t}^{**}$ = (α' + β' + γ')/3 = (48+78+60) / 3 = 62

Estimated S_T^2 = 1658 (using point 15)

Let $P = \sum_{i=1}^{r} \left(\frac{1}{(k_j-2)} - \frac{1}{k}\right) w_j^2 \{(es')_j^2 + \alpha_j^2 R_{Nj}^2 (ex')_j^2 - 2\alpha_j R_{Nj} (es'x')_j\}$

$Q = \sum \left(\frac{1}{r} - \frac{1}{N-k+r}\right)$[estimated S_T^2]

$R = \sum \left(1 - \frac{1}{kj-2}\right) w_j (es')_j^2$

$R = \sum \left(1 - \frac{1}{kj-2}\right) w_j (es')_j^2$

Calculation of P, Q, R at $\alpha_1=\alpha_2=\alpha_3 = 0$

$P = \left(\frac{1}{10} - \frac{1}{30}\right)(0.4)^2 \{3317.65\} + \left(\frac{1}{8} - \frac{1}{30}\right)(0.33)^2 \{2843.98\} + \left(\frac{1}{6} - \frac{1}{30}\right)(0.26)^2 \{2152.66\}$

= 0.066 *0.16*3317.65+ 0.092*0.1089*2843.98 +0.133*0.0676*2152.66= 82.88

$Q = \left(\frac{1}{3} - \frac{1}{150-30+3}\right)(1658) = (0.3252 * 1658) = 539.18$

$R = \left(1 - \frac{1}{10}\right)(0.4)^2 * 3317.65 + \left(1 - \frac{1}{8}\right)(0.33)^2 * 2843.98 + \left(1 - \frac{1}{6}\right)(0.26)^2 * 2152.66$

= 0.9 *0.16*3317.65+ 0.875*0.1089*2843.98 +0.833*0.0676*2152.66 = 869.95

Calculation of Mean and Variance V[t_{mean}/T]

$(€_1)_{opt} = (QR) / [PQ+PR+QR] = QM = 539.18*869.95/[82.88*539.18+82.88*869.95+539.18*869.95]$

= 469059.641/585848.3354= 0.8006

$(€_2)_{opt} = PQ/ [PQ+PR+QR] = PM = 82.88*539.18/[82.88*539.18+82.88*869.95+539.18*869.95]$

= 44687.2384/ 585848.3354= 0.0762

$t_{mean}/T = (€_1)_{opt} [\sum_{j=1}^{r} w_j \bar{t}_{rj}] + (€_2)_{opt} (\bar{t}^*) + (1 - (€_1)_{opt} - (€_2)_{opt}) (\bar{t}^{**})$

t_{mean}/T= 0.8006 [0.4*104.7(196.91/200.7)+0.33*128.62(231.6/222.75) +0.26*163.33(252.5/293)]

+0.0762*54 +0.1232*62 = 0.8006 [41.08 +44.13 +36.59] +4.11+1.88= 97.51+4.11+7.63= 109.25

$V[t_{mean}/T] = (€_1)^2_{opt} P + (€_2)^2_{opt} Q + (1 - (€_1)_{opt} - (€_2)_{opt})^2 R]$

$V[t_{mean}/T] = [(0.8006)^2 * 82.88 + (0.0762)^2 * 539.18 + 0.0152 * 869.95] = 53.12+3.13+13.22 = 69.47$

The 95% confidence intervals for $\bar{t}$, $P[(t_{mean}/T) \pm 1.96\sqrt{[V(t_{mean}/T)]}] = 0.95$

$= 109.25 \pm 1.96\sqrt{69.47} = 109.25 \pm 16.33 = (92.92, 125.58)$

Table 3: *Estimated Sample Mean, Variance and Confidence Interval(CI) of Ten Random Samples CASE I: At* $(€_1)_{opt}$, $(€_2)_{opt}$, $\alpha_j = 0$ ($\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 0$)

S.No.	True Mean	Estimated Sample Mean	$V[t_{mean}/T]$	95% Confidence Interval (CI)	CI Length
1	122.51	109.25	69.47	(92.92, 125.58)	32.66
2	122.51	123.30	61.64	(107.92, 138.68)	30.76
3	122.51	107.67	75.92	(90.59, 124.74)	34.15
4	122.51	114	289.87	(80.63, 147.37)	66.74
5	122.51	128.09	285.83	(94.95, 161.22)	66.27
6	122.51	113.82	30.09	(103.07, 124.57)	21.50
7	122.51	119.23	39.79	(106.87, 131.59)	24.72
8	122.51	113.51	185.98	(86.78, 140.23)	53.45
9	122.51	133.73	175.83	(107.74, 159.30)	51.56
10	122.51	111.47	56.65	(96.72, 126.22)	29.5
			Average Length (411.31/10)		41.13

CASE II: $\alpha_j = 1$ ($\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = 1$)

Calculation for Sample No. 1 , k=30, on above sample and P, Q, R at ($\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = 1$)

$P = \left(\frac{1}{10} - \frac{1}{30}\right) (0.4)^2 \{3317.65+1*0.52*0.52*8158.45-2*1*0.52*2982.52\}$
$+\left(\frac{1}{8} - \frac{1}{30}\right) (0.33)^2 \{2843.98+1*0.57*0.57*3806.21-2*1*0.57*281.75\}$
$+\left(\frac{1}{6} - \frac{1}{30}\right) (0.26)^2 \{2152.66+1*0.55*0.55*547.6-2*1*0.55*841.2\}$
= 0.066 *0.16*2421.87+ 0.092*0.1089*3759.42 +0.133*0.0676*1392.98 = 75.76

$Q = (\frac{1}{3} - \frac{1}{150-30+3})$ 1658= 0.3252 *1658= 539.1816

$R = (1 - \frac{1}{10}) (0.4)^2 *3317.65+ \left(1-\frac{1}{8}\right) (0.33)^2 *2843.98+\left(1-\frac{1}{6}\right) (0.26)^2 *2152.66$
= 0.9 *0.16*3317.65+ 0.875*0.1089*2843.98+0.833*0.0676*2152.66 = 869.95

Calculation of Mean and Variance $V[t_{mean}/T]$ at ($\alpha_1 = 1, \alpha_2 = 1, \alpha_3 = 1$)

$(€_1)_{opt}$= (QR) / [PQ+PR+QR] = QM = 539.1816*869.95/[75.76*539.1816+75.76*869.95+539.1816*869.95]
= 469061.03292/575816.842936= 0.8146

$(€_2)_{opt}$= PQ/ [PQ+PR+QR] = PM =75.76*539.1816/[75.76*539.1816+75.76*869.95+539.1816*869.95]
= 40848.398016/ 575816.842936= 0.0709

$(t_{mean}/T) = (€_1)_{opt}[\sum_{j=1}^{r} w_j \bar{t}_{rj}] + (€_2)_{opt} (\bar{t}^*) + (1- (€_1)_{opt} - (€_2)_{opt}) (\bar{t}^{**})$

(t_{mean}/T)= 0.8146[0.4*104.7(196.91/200.7)+0.33*128.62(231.6/222.75) +0.26*163.33(252.5/293)]
+0.0709*54 +0.1145*62 = 0.8146[41.08+44.13 +36.59] +3.82+7.09=99.21+4.44+7.09= 110.74

$V[t_{mean}/T] = (€_1)^2_{opt} P + (€_2)^2_{opt} Q + (1- (€_1)_{opt} - (€_2)_{opt})^2 R]$ with $\alpha_j = 1$for all j = 1,2,3

$V[t_{mean}/T]$ = [(0.8146)² *75.76+ (0.0709)²*539.1816+ 0.0131*869.95] = 50.27+2.71+11.39=64.37

The 95% confidence intervals for $\bar{t}$,P $[(t_{mean}/T) \pm 1.96\sqrt{[V(t_{mean}/T)]}]$ = 0.95
$=110.74 \pm 1.96\sqrt{64.37}$ =110.74±15.73 = (95.01, 126.46)

Table 4: *CASE II: At* $(€_1)_{opt}$, $(€_2)_{opt}$, $\alpha_j = 1$ ($\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 1$)
Sample Mean , Variance and Confidence Interval(CI) of Ten Random Samples

S.No.	True Mean	Estimated Sample Mean	V[t_{mean}/T]	95% Confidence Interval	CI Length
1	122.51	110.74	64.37	(95.01, 126.46)	31.45
2	122.51	125.72	49.78	(111.89, 139.55)	27.66
3	122.51	112.61	54.66	(98.11, 127.10)	28.99
4	122.51	113.87	305.14	(79.63, 148.10)	68.47
5	122.51	127.45	235.98	(97.35, 157.55)	60.2
6	122.51	113.63	62.92	(98.08, 129.18)	31.1
7	122.51	119.64	37.80	(107.58, 131.69)	24.11
8	122.51	144.02	144.34	(120.48, 167.57)	47.09
9	122.51	133.45	171.53	(107.77, 159.12)	51.35
10	122.51	122.85	40.12	(110.44, 135.26)	24.82
				Average Length (395.2/10)	39.52

III. CASE III: $\alpha_j = \alpha_{opt}$ where $(\alpha_{opt})_j = (es'x')_j / (R_{Nj} * (ex')_j^2)$

Calculation for Sample No. 1, sample size k=30, $\alpha_1 = (\alpha_{opt})_1 = 0.70$, $\alpha_2 = (\alpha_{opt})_2 = 0.13$, $\alpha_3 = (\alpha_{opt})_3 = 2.79$and P, Q, R at $\alpha_j = (\alpha_{opt})_j$[$\alpha_1 = (\alpha_{opt})_1$, $\alpha_2 = (\alpha_{opt})_2$, $\alpha_3 = (\alpha_{opt})_3$]

P = $\left(\frac{1}{10}-\frac{1}{30}\right)$ $(0.4)^2$ {3317.65+0.70*0.70*0.52*0.52*8158.45-2*0.70*0.52*2982.52}
+$\left(\frac{1}{8}-\frac{1}{30}\right)$ $(0.33)^2$ {2843.98+0.13*0.13*0.57*0.57*3806.21-2*0.13*0.57*281.75}
+$\left(\frac{1}{6}-\frac{1}{30}\right)$ $(0.26)^2$ {2152.66+2.79*2.79*0.55*0.55*547.6-2*2.79*0.55*841.2}
= 0.066 *0.16*2227.34+ 0.092*0.1089*2823.12 +0.13*0.0676*860.45 = 59.36

Q = $(\frac{1}{3}-\frac{1}{150-30+3})$ 1658= 0.3252 *1658 = 539.1816

R = $(1-\frac{1}{10})$ $(0.4)^2$ *3317.65+ $\left(1-\frac{1}{8}\right)$ $(0.33)^2$ *2843.98 +$\left(1-\frac{1}{6}\right)$ $(0.26)^2$ *2152.66
= 0.9 *0.16*3317.65+ 0.875*0.1089*2843.98 +0.833*0.0676*2152.66 = 869.95

Calculation of Mean and Variance V[t_{mean}/T]at$\alpha = (\alpha_{opt})_j$

$(€_1)_{opt}$= (QR) / [PQ+PR+QR] = QM = 539.1816*869.95/[59.36*539.1816+59.36*869.95+539.1816*869.95]
= 469061.03292/552707.084696= 0.8486

$(€_2)_{opt}$= PQ/ [PQ+PR+QR] = PM =59.36*539.1816/[59.36*539.1816+59.36*869.95+539.1816*869.95]
= 32005.819776/ 552707.084696= 0.0579

(t_{mean}/T)= $(€_1)_{opt}$[$\sum_{j=1}^{r} w_j \bar{t}_{rj}$] + $(€_2)_{opt}$ $(\bar{t}^*)$ + (1- $(€_1)_{opt}$– $(€_2)_{opt}$) $(\bar{t}^{**})$

(t_{mean}/T)= 0.8486[0.4*104.7(196.91/200.7)+0.33*128.62(231.6/222.75) +0.26*163.33(252.5/293)]
+0.0579*54 +0.0935*62 = 0.8486[41.08+44.13 +36.59] +3.13+5.797= 103.35+3.13+5.79= 112.27

V[t_{mean}/T] = $(€_1)^2_{opt}$ P + $(€_2)^2_{opt}$ Q + (1- $(€_1)_{opt}$– $(€_2)_{opt})^2 R$]

V[t_{mean}/T] = [$(0.8486)^2$ *59.36+ $(0.0579)^2$*539.1816+ 0.0087*869.95] = 42.74+1.80+7.56=52.10

The 95% confidence intervals for $\bar{t}$ P [(t_{mean}/T) ± $1.96\sqrt{[\,V\,(t_{mean}/T)]}$ = 0.95
= 112.27± $1.96\sqrt{52.10}$ = 112.27± 14.14 = (98.13, 126.41)

Table 5: *CASE III: At* $(€_1)_{opt}$, $(€_2)_{opt}$, $\alpha_j = (\alpha_{opt})_j$ *(j = 1, 2, 3)*
Estimated Sample Mean, Variance and Confidence Interval (CI) of Ten Random Samples

S.No.	True Mean	Estimated Sample Mean	$V[t_{mean}/T]$	95% Confidence Interval (CI)	CI Length
1	122.51	112.27	52.10	(98.13, 126.41)	28.28
2	122.51	126.91	44.04	(113.90, 139.91)	26.01
3	122.51	110.92	50.71	(96.96, 124.88)	27.92
4	122.51	114.44	32.79	(103.22, 125.66)	22.44
5	122.51	127.37	230.85	(97.59, 157.14)	59.55
6	122.51	114.59	26.79	(104.44, 124.74)	20.30
7	122.51	121.13	30.35	(110.33, 131.93)	21.60
8	122.51	139.34	105.75	(119.18, 159.49)	40.31
9	122.51	129.54	84.13	(111.56, 147.52)	35.96
10	122.51	123.66	31.31	(112.69, 134.62)	21.93
				Average Length (304.3/10)	30.43

IV. CASE IV: At $€_1 = 1, €_2 = 0$, with $\alpha_j = (\alpha_{opt})_j$

It is the case when no imputation used and partially processed situation not considered. But it is away from practical situation.

$(\alpha_{opt})_j = (es'x')_j / (R_{Nj} * (ex')_j^2)$

Calculation for Sample No. 1, sample size k=30, when $€_1 = 1$, $€_2 = 0$ with $\alpha_j = \alpha_{opt}$ and P, Q, R at $(\alpha_{opt})_j$

P = 59.36, Q = 539.1816, R = 869.95

Calculation of Mean and Variance $V[t_{mean}/T]$ at $(\alpha_{opt})_j$ with $(€_1 = 1, €_2 = 0)$

$(t_{mean}/T) = €_1 [\sum_{j=1}^{r} w_j \bar{t}_{rj}] + €_2 (\bar{t}^*) + (1 - €_1 - €_2)(\bar{t}^{**})$

(t_{mean}/T)= 1* [0.4*104.7(196.91/200.7)+0.33*128.62(231.6/222.75) +0.26*163.33(252.5/293)]

= [41.08+44.13 +36.59] +0+0= 121.8

$V[t_{mean}/T] = [(1)^2 * 59.36 + (0)^2 * 539.1816 + 0 * 869.95] = 59.36$

The 95% confidence intervals for $\bar{t}$ are $P[t_{mean}/T \pm 1.96\sqrt{[V(t_{mean}/T)]}] = 0.95$

$= 121.8 \pm 1.96\sqrt{59.36} = 121.8 \pm 15.10 = (106.70, 136.90)$

Table 6: *CASE IV: when*$[€_1 = 1, €_2 = 0\ \alpha_j = (\alpha_{opt})_j]$
Estimated Sample Mean, Variance and Confidence Interval (CI)of Ten Random Samples

S.No.	True Mean	Estimated Sample Mean	$V[t_{mean}/T]$	95% Confidence Interval (CI)	CI Length
1	122.51	121.8	59.36	(106.70, 136.90)	30.20
2	122.51	136.51	50.11	(121.63, 150.38)	28.75
3	122.51	117.57	56.95	(60.62, 132.36)	71.74
4	122.51	119.77	47.64	(106.24, 133.29)	27.05
5	122.51	121.23	44.23	(108.19, 134.27)	26.08
6	122.51	125.01	27.56	(114.72, 135.30)	20.58
7	122.51	127.21	34.62	(115.67, 138.74)	23.07
8	122.51	116.64	58.63	(101.63, 131.64)	30.01
9	122.51	114	41.37	(101.40, 126.60)	25.20
10	122.51	127.23	30.69	(116.37, 138.09)	21.72
				Average Length	30.44

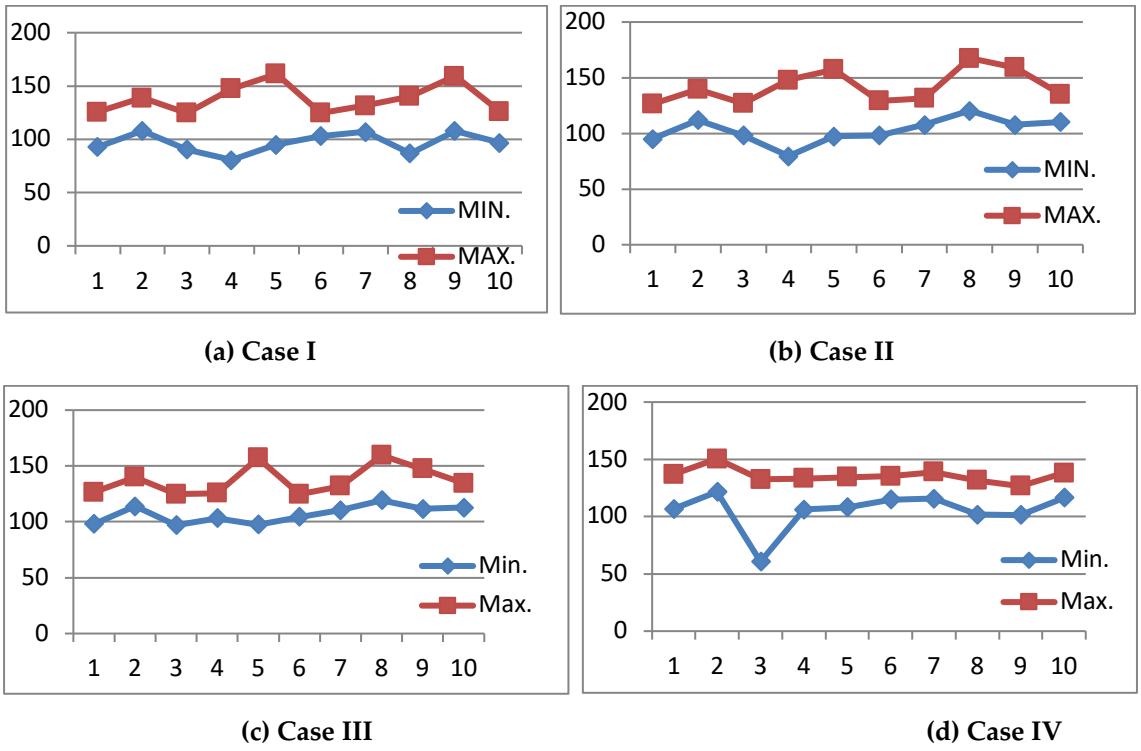


Figure 9: *(a), (b), (c), (d) are graphical representation of Confidence Interval range of Ten Random Samples for four different cases of Table 3,4,5 and 6 (X-axis has sample number as shown in table 3,4,5,6)*

Table 7: *Comparison Between Cases I,II, and III*

S. No.	CASE 1 $\alpha_j = 0$ ($\alpha_1 = \alpha_2 = \alpha_3 = 0$)		CASE II $\alpha_j = 1$ ($\alpha_1 = \alpha_2 = \alpha_3 = 1$)		CASE III $(\alpha)_{j=}(\alpha_{opt})_j^2$		CASE IV [€1=1,€2=0] with $\alpha_j = \alpha_{opt}$	
	95%Confidence Interval	Length	95%Confidence Interval	Length	95%Confidence Interval	Length	95%Confidence Interval	Length
1.	(92.92, 125.58)	32.66	(95.01, 126.46)	31.45	(98.13, 126.41)	28.28	(106.70, 136.90)	30.20
2.	(107.92, 138.68)	30.76	(111.89, 139.55)	27.66	(113.90, 139.91)	26.01	(121.63, 150.38)	28.75
3.	(90.59, 124.74)	34.15	(98.11, 127.10)	28.99	(96.96, 124.88)	27.92	(60.62, 132.36)	71.74
4.	(80.63, 147.37)	66.74	(79.63, 148.10)	68.47	(103.22, 125.66)	22.44	(106.24, 133.29)	27.05
5.	(94.95, 161.22)	66.27	(97.35, 157.55)	60.2	(97.59, 157.14)	59.55	(108.19, 134.27)	26.08
6.	(103.07, 124.57)	21.50	(98.08, 129.18)	31.1	(104.44, 124.74)	20.30	(114.72, 135.30)	20.58
7.	(106.87, 131.59)	24.72	(107.58, 131.69)	24.11	(110.33, 131.93)	21.60	(115.67, 138.74)	23.07
8	(86.78, 140.23)	53.45	(120.48, 167.57)	47.09	(119.18, 159.49)	40.31	(101.63, 131.64)	30.01
9.	(107.74, 159.30)	51.56	(107.77, 159.12)	51.35	(111.56, 147.52)	35.96	(101.40, 126.60)	25.20
10.	(96.72, 126.22)	29.5	(110.44, 135.26)	24.82	(112.69, 134.62)	21.93	(116.37, 138.09)	21.72
Average Length		41.13		39.52		30.43		30.44

VI. Comparison and Discussion

The proposed setup has three parameters $€_1$, $€_2$and α ($0< \alpha < \infty$) whose suitable choices provide the best estimate. The case I has $\alpha_j= 0$ (for all j) which means there is no consideration of size measure in the strategy. Case II considers $\alpha_j= 1$ (for all j) indicating for the presence of size measure x in the estimation strategy but at a particular choice. Case III considers $\alpha_j= (\alpha_{opt})_j$ (for all j) where size measure is at the best (optimal) fractional level incorporated in strategy of prediction. All the three cases (see table 7) are showing the average length of confidence intervals, but smallest average interval length is 30.43 obtained by the case III where choices $(€_1)_{opt}$, $(€_2)_{opt}$ and$(\alpha_{opt})_j$ are used. The ten sample average confidence intervals are in table 8. Fig.9 shows smooth, increasing, condensed and controlled variations of lower and upper limits of CI, best found in the case III which deserved for recommendation.

Table8: *Ten Sample average Confidence Interval & estimated total processing Remaining time for Recovery Management*

	Case I (Without size measure)	Case II (With size measure)	Case III (With size measure)	True Value
Average Interval (Over 10 samples)	(96.8 - 137.9)	(102.6 - 142.1)	(106.8 - 137.2)	122.51
CI Length	41.1	39.5	30.4	
Lowest Predicted Remaining time	(N-k)*96.8 = 11,616 units	(N-k)*102.6 = 12312 units	(N-k)*106.8 = 12816 units	------
Highest Predicted Remaining time	(N-k)*137.9 =16,548 units	(N-k)*142.1 = 17052 units	(N-k)*137.2 = 16464 units	

To note that average intervals (table 8) are producing the same length as shown in table 7.Define relative efficiency measure in terms of percentage as:

Percentage Relative Efficiency (PRE) = $[\frac{[\text{LengthofCIofcaseI}]-[\text{LengthofCIofothercases}]}{\text{LengthofCIofcaseI}}]$ X 100

Table 9: *Percentage Relative efficiency (PRE)*

Case II with respect to Case I	Case III with respect to Case I
PRE = 3.91 %	PRE = 26.01 %

The case III is more efficient (26.02%) than the case II with respect to case I as base where no size measure considered for estimation. In fact, all the sample computed confidence intervals are catching the true value (122.51) which is the strength of the proposed method. The minimal highest predicted time required to process the remaining jobs in ready queue (after breakdown)is 16464 units which is in case III (see table 8).

VII. Conclusion

On recapitulation, the paper considers the practical problem of remaining time estimation of processes in ready queue, after the occurrence of system failure in a multiprocessor computer system. While sudden breakdown how much backup time and computer related infrastructure required? This time duration and maximum time estimation are done using the tools of sampling theory and assumption of lottery scheduling. This scheduling opens avenues for application of random sampling tools and techniques. A concept of virtual ready queue is added as a new feature, who found useful in allocating the processes to multi-processors. The virtual ready queue along with lottery scheduling have created environment for the estimation of remaining time of

main ready queue. The proposed estimation strategy is capable enough to predict for mean time. For efficient estimation $(€_1)_{opt}$, $(€_2)_{opt}$ and (α_{opt}) are used who provide the lowest length confidence interval. The Case III found best estimated and predicted than case I and case II. The case III also provides prediction indicating the minimal highest remaining time to arrange backup accordingly while failure. Imputation has improved the level of estimation and use of additional information (size measure) contributed a lot for higher precision. Such estimates are useful for backup and recovery management while the occurrence of system breakdown. Such findings are useful for risk evaluation and disaster management in setup of cloud computing and data centre.

References

[1] More Sarla, and Shukla Diwakar, Some new methods for ready queue processing time estimation problem in a multiprocessor environment, Social Networking and Computational Intelligence, Lecture notes in Networks and Systems, Springer, Singapore, March 2020; 100: 661-670.

[2] More, Sarla and Shukla Diwakar, Analysis, and extension of methods in ready queue processing time Estimation in Multiprocessor Environment, Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity university Rajasthan, Jaipur-India, February 2019; 1558-1563.

[3] More, Sarla and Shukla Diwakar, "A review on ready queue processing time estimation problem and methodologies used in multiprocessor environment", International Journal of Commuter Science and Engineering, 2018; 6(5): 1186-1191.

[4] Shukla Diwakar, Jain Anjali, and Choudhary Amita, Estimation of ready queue processing time under SL scheduling scheme in multiprocessors environment, International Journal of Computer Science and Security, 2010; 4(1): 74-81.

[5] Shukla Diwakar, Jain Anjali and Choudhary Amita, "Estimation of ready queue Processing time under usual group lottery scheduling (GLS) in multiprocessor environment", International Journal of Commuter Applications, 2010; 8(14): 39-45.

[6] Shukla Diwakar, Jain Anjali and Choudhary Amita, Prediction of ready queue processing time in multiprocessor environment using lottery scheduling (ULS), International Journal of Commuter Internet and Management, 2010; 18 (3):58-65.

[7] Shukla Diwakar, and Jain Anjali, Analysis of ready queue processing time under PPS-LS and SRS-LS scheme in multiprocessing environment, GESJ: Computer Science and Telecommunications, 2012; 33(1): 54-61.

[8] Shukla Diwakar and Jain Anjali, Estimation of ready queue processing time using efficient factor type estimator (E-F-T) in multiprocessor environment, International Journal of Computer Applications, 2012; 48(16): 20-27.

[9] Shukla Diwakar and Jain Anjali, Ready queue mean time estimation in lottery scheduling using auxiliary variables in multiprocessor environment, International Journal of Commuter Applications, 2012; 55(13):13-19.

[10] Jain, Anjali and Shukla, Diwakar, Estimation of ready queue processing time using factor type (F-T) estimator in multiprocessor environment, COMPUSOFT, An International Journal of Advanced Computer Technology, 2013; 2(8): 256-260.

[11] Shukla Diwakar , Jain Anjali and Verma Kapil, Estimation of ready queue processing time using transformed factor- type (T-F-T) estimator in multiprocessor environment, International Journal of Computer Applications, 2013;79(16): 40-48.

[12] Carl A. Waldspurger and E William Weihl, Lottery Scheduling: Flexible proportional share resource management, The 1994 Operating systems design and implementation conference (OSDI '94), Monterey, California.

[13] Johnnie Daniel, Sampling Essentials: Practical Guidelines for Making Sampling Choices, Sage Publication, 2011.
[14] Paul S. Levy and Stanley Lemeshow, Sampling of Populations: Methods and Applications, Volume 543 of Wiley Series in Survey Methodology, Wiley, 2008.
[15] Sampath, S.,Sampling Theory and Methods, Alpha Science International Publication, 2005.
[16] Cochran, W.G, Sampling Technique, Wiley Eastern publication, New Delhi,2005.
[17] Poduri S. R. S. Rao, Sampling Methodologies with Applications, Texts in Statistical Science, Chapman and Hall/CRC Press, 2000.
[18] Ranjan K. Som, Practical Sampling Techniques, Second Edition Statistics: A Series of Textbooks and Monographs, CRC Press, 1995.
[19] Steven K. Thompson, Sampling, Volume 272 of Wiley Series in Probability and Statistics, 1992.
[20] Shweta Ojha, Saurabh Jain and Diwakar Shukla, Hybrid lottery multi-level queue scheduling with a Markovian model, GESJ: Computer Science and Telecommunications, 2011; 3(32): 86-97.
[21] Pradeep kumarJatav, Rahul Singhai, and Saurabh Jain, Analysis of hybrid lottery scheduling algorithm using Markov chain model, International Journal for Research in Engineering Application and Management (IJREAM), 2018; .04 (09): 180-192.
[22] MaríaMejía, Adriana Morales-Betancourt and Tapasya Patki, Lottery scheduler for the Linux kernel, The author; licensee universidad Nacional de Colombia. Dyna 82 (189), Medellín. Printed, Online DOI: http://dx.doi.org/10.15446/dyna.v82n189.43068, 2015; 216-225.
[23] Diwakar Shukla and Sarla More, Modified group lottery scheduling algorithm for ready queue mean time estimation in multiprocessor environment, Reliability: Theory & Applications (RT&A), 2020; 15, 4(59): 69-85.
[24] Carl A. Waldspurger, Lottery and stride scheduling: Flexible proportional-share resource management, Massachusetts Institute of Technology, September 1995.
[25] Singh H.P., Gupta A. and Tailor R., Estimation of population mean using a difference-type exponential imputation method, https://doi.org/10.1007/s42519-020-00151-2, Journal of Statistical Theory and Practice 2021; 15(19).
[26] Kone Dramane, Goore Bi Tra, and Dr. Kimou Kouadio Prosper, New hybrid method for efficient imputation of discrete missing attributes, International Journal of Innovative Science and Research Technology, 2020; 5(11): 983-991.
[27] GaribNath Singh, Mohd. Khalid and Jong-Min Kim, Some imputation methods to deal with the problems of missing data in two-occasion successive sampling, Communications in Statistics-Simulation and Computation, 2019; 50(2):557-580.
[28] Diwakar Shukla, Narendra Singh Thakur and Sharad Pathak, Some new aspects on imputation in sampling, African Journal of Mathematics and Computer Science Research, 2013; 6(1): 5-15.
[29] Diwakar Shukla, D. S. Thakur and N. S. Thakur, Utilization of mixture of $\overline{x}$, $\overline{x}_1$, and $\overline{x}_2$ in imputation for missing data in post-stratification, African Journal of Mathematics and Computer Science Research, 2012; 5(4):78-89.
[30] Diwakar Shukla, Narendra Singh Thakur, Dharmendra Singh Thakur and Sharad Pathak, Linear combination based imputation method for missing data in sample, International Journal of Modern Engineering Research (IJMER) 2011;1(2): 580-596.
[31] Diwakar Shukla and Narendra Singh Thakur, Imputation Methods in Sampling Hardcover, Aman Prakashan, January 2014, available at *Amazon:https://www.amazon.in/Imputation-Methods-Sampling-DiwakarShukla/dp/B07VY2B59D.*

RT&A, No 3 (63)
Volume 16, September 2021

Fuzzy Regression Based Patient Life Risk Rate Prediction Using Oxygen Level, Pulse Rate And Respiration Rate In Covid-19 Pandemic (FRPRPS)

Gaurav Kant Shankhdhar[1], Himanshu Pandey[2], Atul Kumar Pal[3], Sumit Mishra[4]

[1]Department of Basic Sciences, Babu Banarasi Das University, Lucknow, emaiil: g.kant.82@gmail.com

[2]Department of Computer Science, Lucknow University, Lucknow, email: hpandey010@gmail.com

[3]Department of Basic Sciences, Babu Banarasi Das University, Lucknow, email: atulpal892@bbdu.ac.in

[4]Department of Computer Science, BBDEC, Babu Banarasi Das University, Lucknow, email: mishrasumit221@gmail.com

Abstract

Today, the general situation worldwide is that the hospitals, sanatoriums and medical colleges are running out of beds, oxygen, medical staff, ventilators and other required paraphernalia that is mandatory for the treatment of the vicious pandemic [1]. The requirement is for a system that takes in some input parameters like Oxygen level of the patient, pulse rate and respiration rate and in turn predicts the Life Risk Rate of that patient [2]. The model used here is a fuzzy regression model that gives the prediction of Life Risk Rate between 1 and 10 units. The lower the predicted Life Risk Rate, the better the chances of survival of the Covid patient. But if the predicted Life Risk Rate is more than the mean of the observations of the Risk in the dataset, then immediate emergency is needed. The benefit of this system is that the patients requiring immediate admission and treatment can be filtered and medical aid in hospital be thereby provided for critical patients. Rest may be home quarantined and domestic medical aid may be given to them until in some unfortunate situation their Risk Rate is near alarming. This paper aims to provide some help in this crucial situation.

Keywords: Fuzzy, Regression, Covid, Prediction, Oxygen Level, Pulse rate

I. Introduction

I. Literature Review

Fuzzy regression analysis gives a fuzzy functional relationship between dependent and independent variables where vagueness is present in some form. The input data may be crisp or fuzzy. In this chapter the authors consider two types of fuzzy regression. The first is based on possibilistic concepts and the second upon a least squares approach. However, in both the notion of "best fit" incorporates the optimization of a functional associated with the problem. In possibilistic regression, this function takes the form of a measure of the spreads of the estimated output, either as a weighted linear sum involving the estimated coefficients in linear regression, or

as quadratic form in the case of exponential possibilistic regression. These optimization problems reduce to linear programming. For the least squares approach, the functional to be minimized is an L_2 distance between the observed and estimated outputs. This reduces to a class of quadratic optimization problems and constrained quadratic optimization. The method can incorporate stochastic fuzzy input and fuzzy kriging uses covariances to obtain BLUE estimators [3].

In this paper, we propose simple but powerful methods for fuzzy regression analysis for Covid affected patients on the basis of oxygen, pulse rate and Respiration rate using R language. Since neural networks have high capability as an approximator of nonlinear mappings, the proposed methods can be applied to more complex systems than the existing LP based methods. First we propose learning algorithms of neural networks for determining a nonlinear interval model from the given input-output patterns. A nonlinear interval model whose outputs approximately include all the given patterns can be determined by two neural networks. In this paper, next is shown two methods for deriving nonlinear fuzzy models from the interval model determined by the proposed algorithms. Nonlinear fuzzy models whose *h*-level sets approximately include all the given patterns can be derived. Last is shown an application of the proposed methods to a real problem[4].

During the ongoing coronavirus disease (COVID-19) pandemic, reports in social media and the lay press indicate that a subset of patients are presenting with severe hypoxemia in the absence of dyspnea, a problem unofficially referred to as “silent hypoxemia.” To decrease the risk of complications in such patients, one proposed solution has been to have those diagnosed with COVID-19 but not sick enough to warrant admission monitor their arterial oxygenation by pulse oximetry at home and present for care when they show evidence of hypoxemia [5]. Though the ease of use and low cost of pulse oximetry makes this an attractive option for identifying problems at an early stage, there are important considerations with pulse oximetry about which patients and providers may not be aware that can interfere with successful implementation of such monitoring programs [6]. Only a few independent studies have examined the performance of pocket oximeters and smart phone–based systems, but the limited available data raise questions about their accuracy, particularly as saturation falls below 90%. There are also multiple sources of error in pulse oximetry that must be accounted for, including rapid fluctuations in measurements when the arterial oxygen pressure/tension falls on the steep portion of the dissociation curve, data acquisition problems when pulsatile blood flow is diminished, accuracy in the setting of severe hypoxemia, dyshemoglobinemias, and other problems. Recognition of these issues and careful counseling of patients about the proper means for measuring their oxygen saturation and when to seek assistance can help ensure successful implementation of needed monitoring programs [7].

The fundamental differences between fuzzy regression and ordinary regression are identified Here [8]. Fuzzy regression can be used to fit fuzzy data and crisp data into a regression model, whereas ordinary regression can only fit crisp data. Through a comprehensive literature review, three approaches of fuzzy regression are summarized. The first approach of fuzzy regression is based on minimizing fuzziness as an optimal criterion. The second approach uses least-squares of errors as a fitting criterion, and two methods are summarized in this paper. The third approach can be described as an interval regression analysis. For each fuzzy regression method, numerical examples and graphical presentations are used to evaluate their characteristic and differences with ordinary least-squares regression.

II. Fuzzy Regression Model for Covid Risk Prediction

The fight to Covid-19 has produced numerous immature remedies, though not quite effective but some being a ray of hope. The major problems in nearly all Covid affected countries are lack of beds or insufficient oxygen and ventilators [9].

The primary deciders for the level of risk associated with a Covid patient are blood oxygen level, pulse rate and respiratory rate. Blood Pressure and Sugar level do play very important part clinically. But when Correlation Analysis of the sample data was done, both of these factors showed minuscule relatedness with Risk shown in table 1[10]. In addition to average levels of systolic and diastolic BP, blood pressure variability (BPV) has also been positively associated with high risks of morbidity and mortality in patients with hypertension. Recent studies also suggested that high BPV could predict a high risk of organ damage, cardiovascular events, and all-cause and cardiovascular mortality independent of mean Blood Pressure in patients with hypertension or cerebrovascular disease [11]. So, though, there is a direct impact of Blood Pressure clinically on Covid patients, the authors have hardcoded the range of Blood Pressure to Risk Factor. Same is also true for Patients with abnormal Sugar levels [12].

Table1: *Correlation between Risk and Oxygen, Pulse Rate and Respiration Rate.*

	Oxygen	*BP(50-120)*	*Sugar (67-210)*	*Pulse Rate(60-120)*	*Rrate(12-25)*	*RISK*
Oxygen	1					
BP(50-120)	0.047024825	1				
Sugar (67-210)	-0.022852292	-0.103454373	1			
Pulse Rate(60-120)	0.841028127	-0.063381361	0.013401908	1		
Rrate(12-25)	0.766020421	-0.127063601	0.048936429	0.896696141	1	
RISK	-0.784911875	0.11131738	-0.062356984	-0.924711669	-0.941593809	1

As is evident from the above table of correlation coefficients, Oxygen, Pulse Rate and Respiration Rate shown in yellow are significant but Blood Pressure and Sugar shown in red do not statistically contribute much as they are very low in correlation coefficientsand so avoided [13].

I. Generation of Input and output variable data values through programming in RLanguage

The authors through a self-developed program in R Language have used the "FuzzyR" library to pass a "fis" file (Fuzzy Inference System) that contains all the details of the fuzzy system including all Member Functions and also the rules that govern the functioning of the system shown in table 2. [14].

Table 2: *Fuzzy Rules*

Rule	IF Oxygen		AND Pulse Rate	AND Respiration Rate	THEN RISC
1	Low		low	low	High
2	medium		Medium	medium	Medium
3	High		low	medium	Low
4	Low		low	low	High
5	Low		Medium	medium	High
6	medium				Medium
7	Low				High
8	Low		Low		High
9	Low		High		High
10	Low		Low	High	High
11				Medium	Medium

The fis file used here is shown in figure 1. The dataset has been programmatically generated by using the "FuzzyR" library of R language where all input variables were created randomly through loops and so does the output variable. The program is shown in snippetas figure 2 [15].

```
% $Revision: 1.1 $

[System]
Name='COVID_fis.fis'
Ninputs=5
Noutputs=1
Nrules=17
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'

[Input1]
Name='Oxygen Level'
Range=[70 99]
NumMFs=3
MF1='low':'trimf',[70 75 85]
MF2='medium':'trimf',[75 90 100]
MF3='high':'trimf',[85 95 99]
[Input2]
Active='yes'
Name='Pulse Rate'
Range=[60,120]
NumMFs=3
```

```
MF1='low':'trimf',[60 75 90]
MF2='high':'trimf',[90 105 120]
MF3='Medium':'trimf',[70 90 110]
[Input5]
Active='yes'
Name='Respiration Rate'
Range=[9 16]
NumMFs=3
MF1='low':'trimf',[11 12 13]
MF2='high':'trimf',[12.500 14.500 16]

MF3='medium':'trimf',[12 13 14]

[Output1]

Name='RISC'

Range=[0 10]

NumMFs=3

MF1='Low':'trapmf',[0 2 5]

MF2='Medium':'trimf',[2 5 8]

MF3='High':'trapmf',[6 8 10]

[Rules]

1 0 0 0 0, 3 (1) : 1
2 0 0 0 0, 3 (1) : 1
3 0 0 0 0, 1 (1) : 1
0 1 0 0 0, 3 (1) : 1
0 2 0 0 0, 2 (1) : 1
0 0 1 0 0, 3 (1) : 1
0 0 2 0 0, 3 (1) : 1
0 0 0 1 0, 3 (1) : 1
0 0 0 3 0, 2 (1) : 1
0 0 0 0 1, 3 (1) : 1
0 0 0 0 2, 3 (1) : 1
#In this experiment all the possible input MFs (Oxygen, Blood Pressure, Sugar, Pulse Rate and
#Respiration Rate) values were tried by a loop and evalfis() function to compute the output crisp
#values of all possible range of complexities. These all 1000 value sets were recorded
library(ggplot2)
library(FuzzyR)

#the fis file is read in fisStdr variable that contains the fis file.
fisStdr<-readfis("F:\\COVID_fis.fis")
#IL is a matrix that has 1000 rows and number of columns are 5. The sample size, n is taken to be
1000.
IL<-matrix(, nrow = 1000, ncol = 5)
```

```
#44 different combinations of the 5 input variables were run through this fuzzy program and
recorded in IL Matrix.
for(i in 1:45)
{
IL[i,1]<-runif(1, min=70,max=99)#oxygen
IL[i,2]<-runif(1, min=50,max=120)#BP Diastolic
IL[i,3]<-runif(1, min=70,max=180)#sugar level
IL[i,4]<-runif(1, min=60,max=120)#Pulse rate
IL[i,5]<-runif(1, min=9,max=16)#Respiration Rate
}
#inpt variable stores the resultant matrix IL
inpt<-matrix(IL,1000,5)

print(inpt)

#print("Defuzzified Value")
#evalfis function computes the inference of the fuzzy rules and gives the output variable for each set
of the input variables
resMATStdr=evalfis(inpt,fisStdr)
print(resMATStdr)
```

Figure 1: *Fuzzy Program to find Risk*

The output of the program with manual alignment and correction is shown in figure 2, discussed later in the paper.

Sl.No	Oxygen	Pulse Rate	RR	RISK
1	92.75268	90	20	1.44063
2	81.29803	55	14	7.475597
3	93.28156	91	22	1.798155
4	94.75914	89	21	1.670695
5	84.28443	57	12	7.578823
6	82.02732	58	12	7.768271
7	92.73445	90	19	1.3326018
8	82.2335	56.5	13	7.838347
9	97.95146	91	16	2.985029
10	86.9621	61	13	5.139235
11	76.94824	60	10	7.693251
12	98.59557	90	19	1.688301
13	79.19153	55	13	6.45691
14	85.87632	58	13	5.189274
15	97.63994	92	22	1.705292
16	90.54865	55	13	5.63892
17	85.19086	59	13	5.25
18	77.85256	57	12	5.459252
19	82.89129	50	10	8.64634
20	73.35761	56	12	6.206666
21	80.25458	50	11	7.36884
22	83.4767	51	12	6.804132
23	93.2275	90	21	1.859923
24	79.50032	50	10	8.25
25	87.0218	49	10	7.768378
26	71.38359	47	9	8.080668
27	95.48155	95	18	1.943947
28	93.24146	95	18	1.95
29	96.19863	94	19	1.802394
30	80.56792	60	11.131345	5.186457
31	87.22132	50	10	7.811492
32	88.14657	58	12	6.838689
33	72.17375	48	9	8.476195
34	87.03107	61	10	7.312335
35	85.96788	65	10	7.787564
36	88.32188	65	10	7.8
37	88.29243	68	11	7.058466
38	91.9299	91	20	1.789683
39	82.09052	50	11	7.080497
40	79.15857	45	10	8.638569
41	85.30994	67	10	7.7220664
42	78.05531	49	12	8.398254
43	83.3797	52	11	7.058882

Figure 2: *The output of the program with manual alignment and correction*

The Blood Pressure and Sugar Membership Functions were not included due to the reason that they had very less correlation with Patient's Risk but it is hereby pointed that the blood pressure and sugar levels do have strong impact on Covid Patients [16]. So the authors have taken these two factors differently using hard coding system, discussed later in this chapter [24].

II. Membership Functions in Fuzzy Regression Patient Risk Prediction System (FRPRPS)

FRPRPS shown in figure 3 is a combination of Fuzzy and Regression system. In this system, Fuzzy Input variables like Oxygen level of Covid +ve patient, pulse rate and respiratory rate are identified as being significant factors for calculating the Fuzzy output variable Risk factor or Mortality Rate of that patient. Then through the 'R Language' a series of combinations of 1000 tuples or rows containing crisp values are generated as shown in figure 2.

After this, the crisp values are perused and checked to see if the Oxygen level-Pulse rate-Respiratory rate-Risk Factor row values are correct in actuality, that is, near to Real world values. If there is some correction needed then that correction is made by having proper alignment and resetting the said crisp input variable(s).

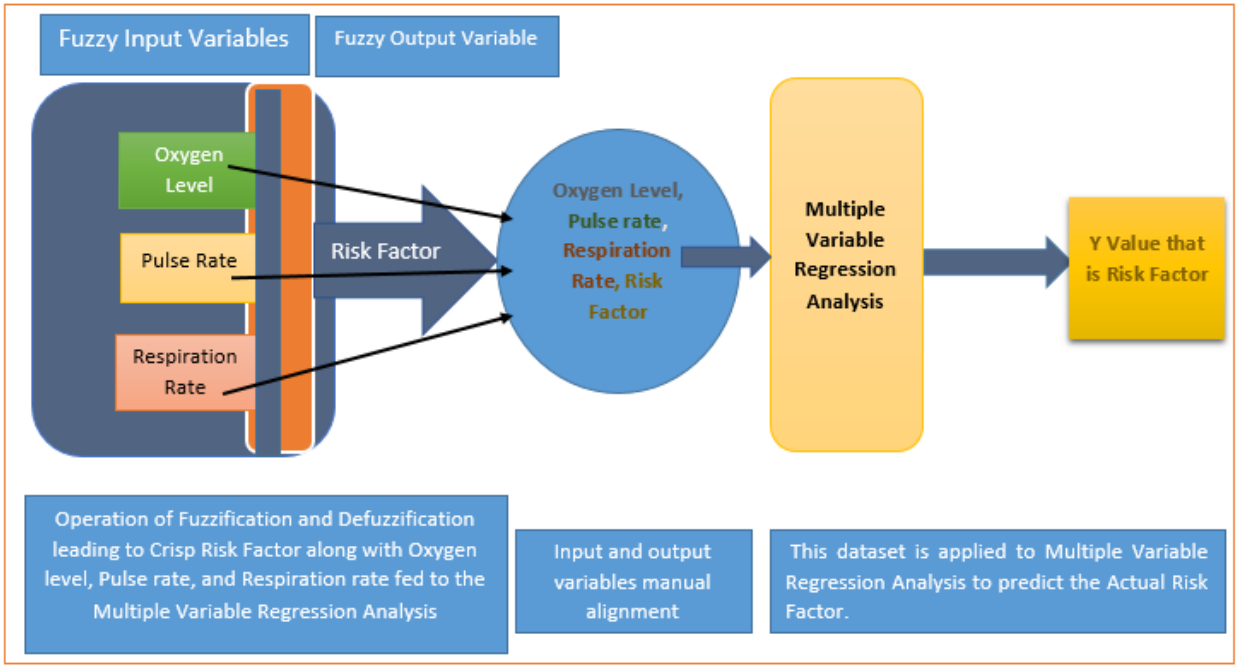


Figure 3: *Membership Functions in Fuzzy Regression Patient Risk Prediction System FRPRPS*

As a final step these crisp values are passed to the regression analysis component of this system where Oxygen level, Pulse rate, Respiratory rate crisp values are fed as independent variables and Risk Factor is predicted (Risk factor / Mortality rate is the dependent variable).The manual correction of seemingly incorrect crisp values of input variables is as in figure 4.The step by step process of FRPRPS is shown in figure 5.

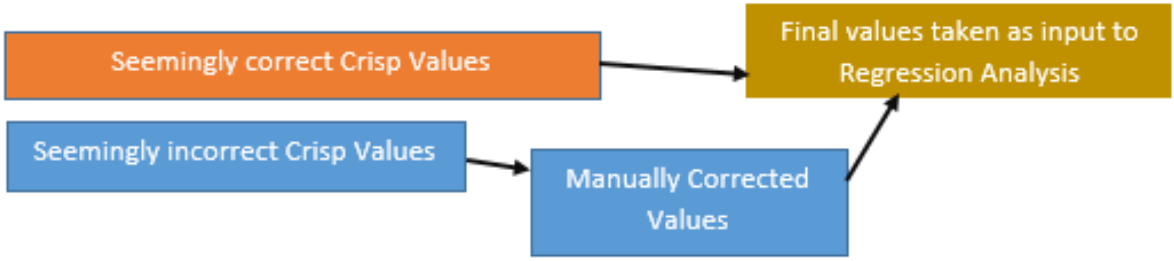


Figure 4: *Manual correction of incorrect Crisp values of SaO_2, Pulse Rate and Respiratory Rate.*

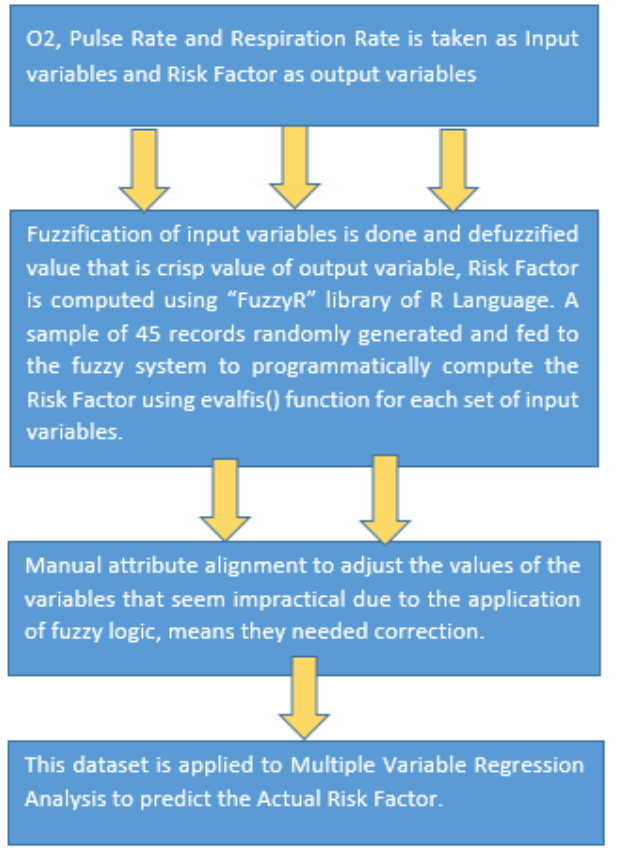


Figure 5. *Step by step process of FRPRPS*

III. Input Membership Functions

The fuzzy system is developed in FISPro software where all membership functions are created, rules written and inference conducted. But it is pertinent to mention over here that after designing the member ship functions and rules, "FuzzyR" library of R Package is utilized to infer 1000 samples of input and output variables iteratively. The FisPro software interface is shown in figure 7[18]

- Oxygen Level (70 to 99)

The Oxygen Level input variable keeps the record of the oxygen level. The second input variable Pulse Rate ranges from 70, which is very low to 99 that is kept on a higher side and in between lies the range of medium. The "Mamdani" inference mechanism is used with conjunction set to "minimum". The shape of the membership functions is decided by trial and error.

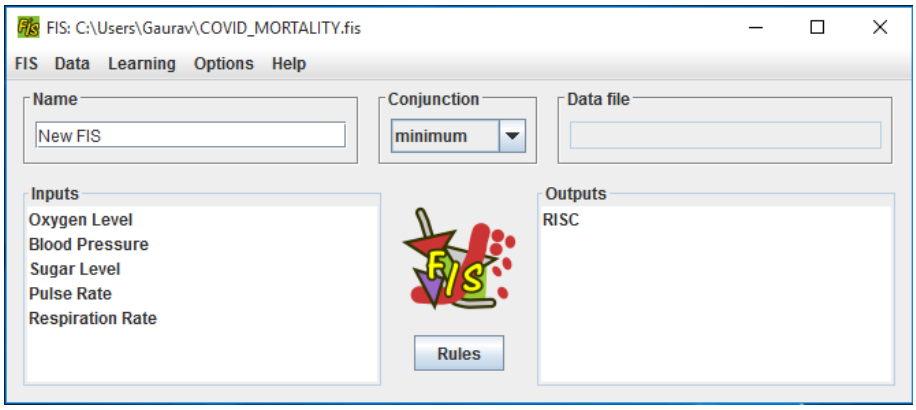


Figure 7: *FisPro Software for Fuzzy Logic implementation*

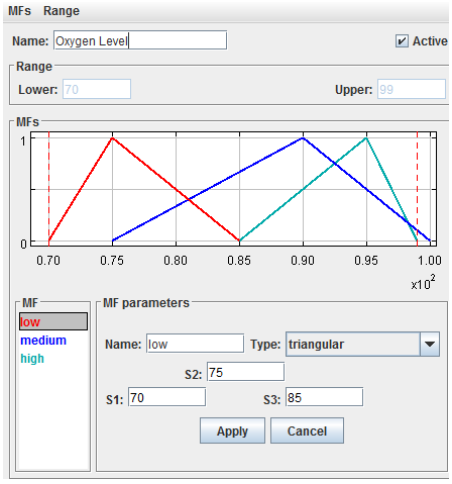


Figure 8a: *Oxygen Level Input Variable*

- Pulse rate (60 to 120)

The second input variable Pulse Rate ranges from 60 to 120. As in case of oxygen input variable membership functions which are low, medium and high are designed using trial and error basis.

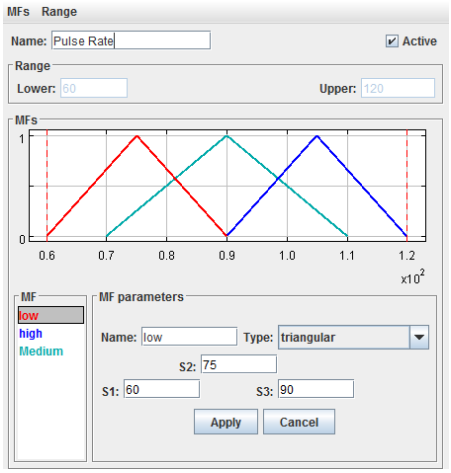


Figure 8b: *Pulse Rate Input Variable*

- Respiration Rate (11 to 20)

The third input variable Pulse Rate ranges from 11 to 20. As in case of before mentioned input variable membership functions of respiration rate are low, medium and high and are designed using trial and error basis.

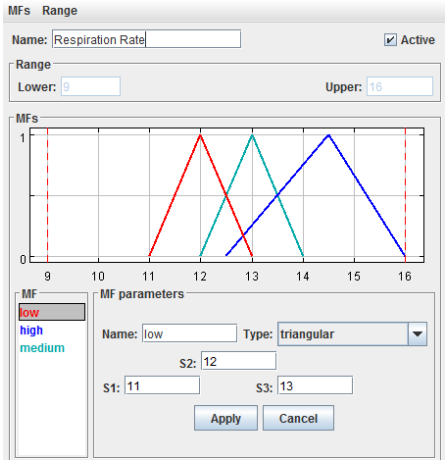


Figure 8c: *Respiration Rate Input Variable*

IV. Output Membership Functions

I. Risk Factor (1 to 10)

The only output variable Risk Factor ranges from 1 to 10. As in case of input variables membership functions of risk factor are low, medium and high and are also designed using trial and error technique.

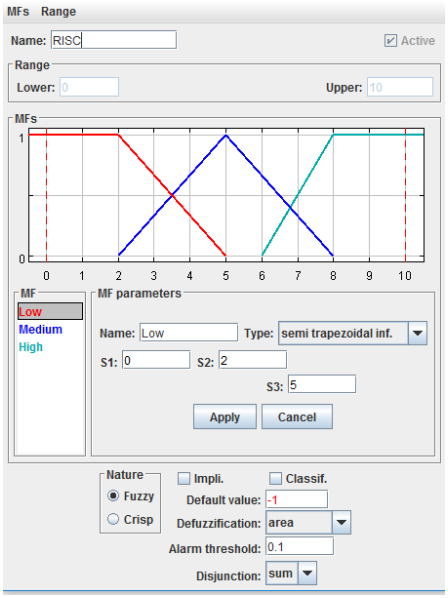


Figure 8d: *Risk Factor Output Variable*

Once the defuzzified values have been generated for Risk Membership Functions, then the full set of the 3 input MFs and 1 output MF is perused to find if any abruption, illogical values of variables, or some discrepancy in the combination of all the variables that look incorrect have been manually aligned and set until all 1000 records look correct [11], [19].

The frequencydistribution graph for Risk factor generated through Ggplot2 in R for Risk Output Variable is shown in figure 9. According to the graph maximum cases are having Risk factor nearly 2 and the next below level is the number of cases having Risk factor of 7.8 [20], [21]. The third maximum cases are having Risk factor of 4.5.

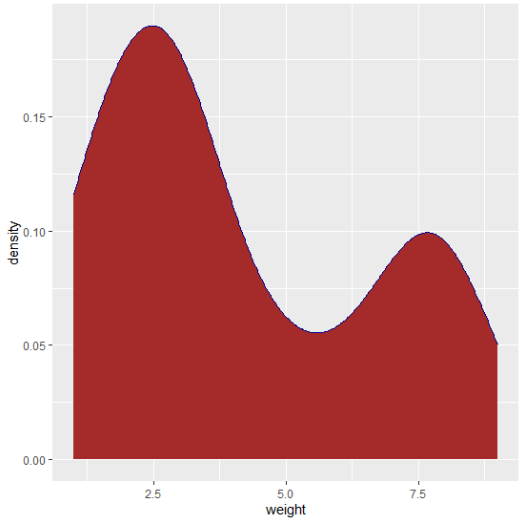


Figure 9: The frequency distribution graph for Risk factor generated through Ggplot2 in R for Risk Output Variable

II. Inference on Fuzzy sets

The data obtained in figure 3 is used as an input to the Regression Analysis. Through the regression analysis conducted over this data, the line of best fit along with coefficients and intercept are established as shown in figure 11a, 11b and 11c [22].

Regression is done in SPSS, Statistical Package for Social Sciences [23]. The Regression details are shown in figure 10[24].

SUMMARY OUTPUT

Regression Statistics	
Multiple F	0.958948
R Square	0.919581
Adjusted I	0.913395
Standard I	0.763177
Observati	43

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	3	259.7439	86.58129	148.6531	2.2E-21
Residual	39	22.71511	0.582439		
Total	42	282.459			

	Coefficients	*ndard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Upper 95.0%*
Intercept	14.75931	2.002229	7.371441	6.66E-09	10.70942	18.8092	10.70942	18.80920106
Oxygen	-0.00049	0.030823	-0.01595	0.987354	-0.06284	0.061854	-0.06284	0.061854182
Pulse Rate	-0.06193	0.01847	-3.35304	0.001788	-0.09929	-0.02457	-0.09929	-0.024571267
RR	-0.36706	0.065726	-5.58468	1.95E-06	-0.5	-0.23412	-0.5	-0.234116598

Figure 10: *Regression Analysis of Fuzzy Data*

As enclosed in the yellow boundary, the coefficient of X1 that is, Oxygen is -0.00049, the coefficient of X2 that is, Pulse Rate is -0.06193, the coffecient of X3, Respiratory Rate is

-0.36706. The intercept c=14.75931.

The equation of Line of Regression is Y=m1X1+m2X2=m3X3……..……(1)

Which turns out to be: Y=-0.00049X1+-0.06193X2+-0.36706X3+14.75931……(2)

As an example X1 is taken to be 92, X2 as 70 and X3 as 12. The result comes out to be 5.9742…

The Regression Line of fit for Oxygen is shown in figure 11. a

The x axis shows the oxygen level of the Covid patient that ranges from 70 to 99. On the Y axis lies the risk factor dependent variable Risk which is dependent on Oxygen Level. Safe Oxygen level lies

above 95[25]. The Risk is interpreted between 1 and 10 where 1 is least risk and 10 denotes highest risk.

Table 3: *Level of attention required on the basis of Risk factor / Mortality Rate*

Risk Factor/ Mortality Rate	Level of Attention Required
1-3	Home Quarantine/Isolation and inhiliation
3-5	Home Quarantine with Oxygen support and inhilation
5-7	Immediate Hospital bed allotment with Oxygen Support, Nebulization and inhilation
7-10	Immergency (Extremely high Mortality rate)

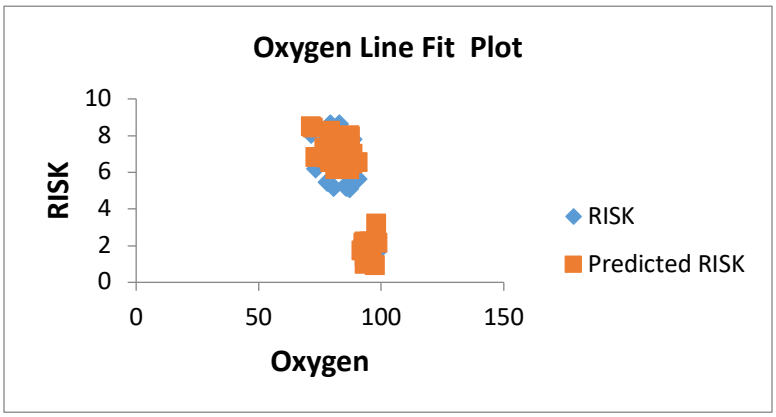


Figure 11.a: Oxygen Line Fit Plot

The Regression Line of fit for Pulse Rate is shown in figure 11.b

The x axis shows the pulse rate of the Covid patient that ranges from 60 to 120.

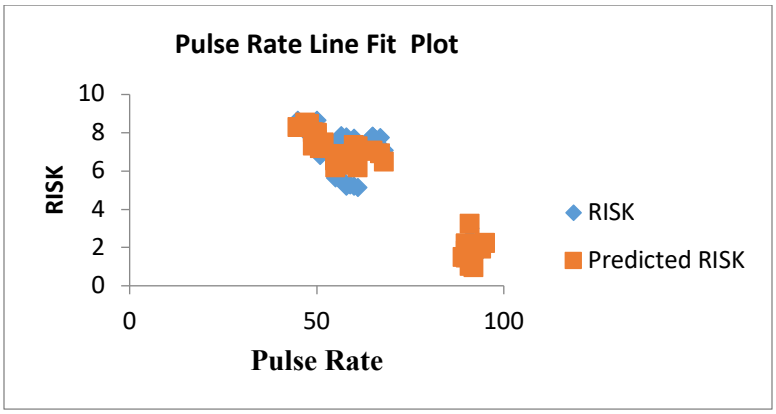


Figure 11b: *Pulse rate Line Fit Plot*

The Regression Line of fit for Respiration Rate is shown in figure 11.c

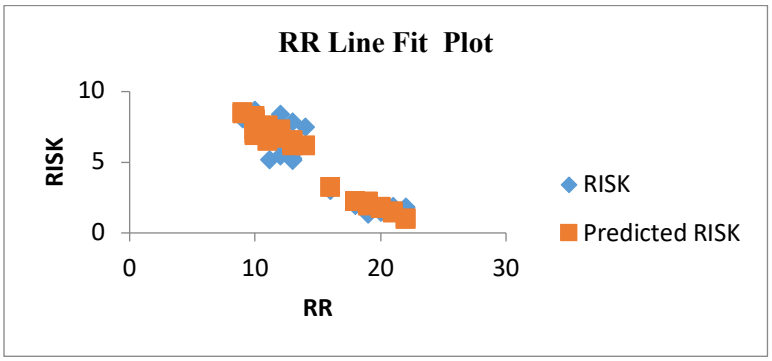


Figure 11 c *Respiration Rate Line Fit Plot*

The x axis shows the respiration rate of the Covid patient that ranges from 11 to 20. The overlapping clearly indicates the level of dependence between RR and Risk.

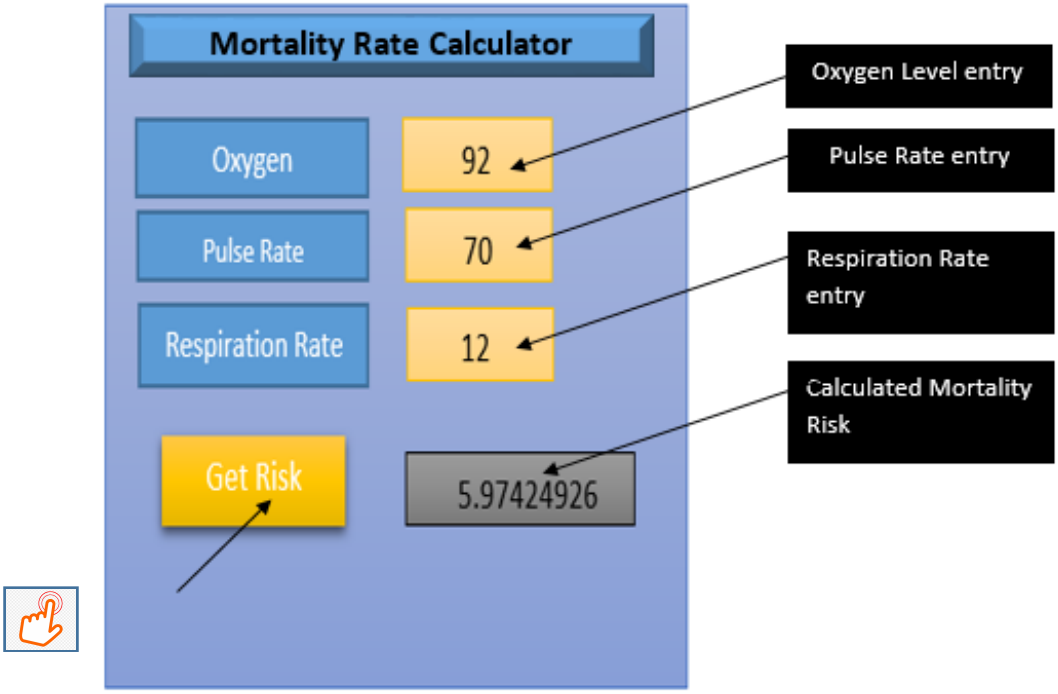


Figure 12.*Implementation of Risk factor Prediction*

The authors have developed a Mortality Rate Calculator which predicts the risk associated to a Covid patient on the basis of his oxygen level, SaO_2 measured in percentage (%), pulse rate in BPM and respiration rate in breaths per minute. The normal reading of Oxygen level is 98. Anything below 95 is alarming. The pulse rate should be between 70 to 100. The normal respiration rate has 12 to 16 breaths per minute.As you can see the three input variables Oxygen level, Pulse rate and Respiration Rate have to be entered to get the Mortality Risk of a Covid patient as in figure 11. Here with the given data, the click on Get Risk button displays the risk associated. [26].

The graph for the data having Oxygen level SaO_2, pulse rate and RR along with Risk factor in % is shown in figure 12.

Normal range of A normal ABG oxygen level for healthy lungs falls between 80 and 100 millimeters of mercury (mm Hg). If a pulse ox measured your blood oxygen level (SpO_2), a normal reading is typically between 95 and 100 percent.[32].

The normal pulse for healthy adults ranges from 60 to 100 beats per minute. The pulse rate may fluctuate and increase with exercise, illness, injury, and emotions. Females ages 12 and older, in general, tend to have faster heart rates than do males. Athletes, such as runners, who do a lot of cardiovascular conditioning, may have heart rates near 40 beats per minute and experience no problems [33]. The mortality rate of a Covid patient is shown as a curved line in figure 13 where SaO_2, PR and RR are depicted as clustered column histogram.

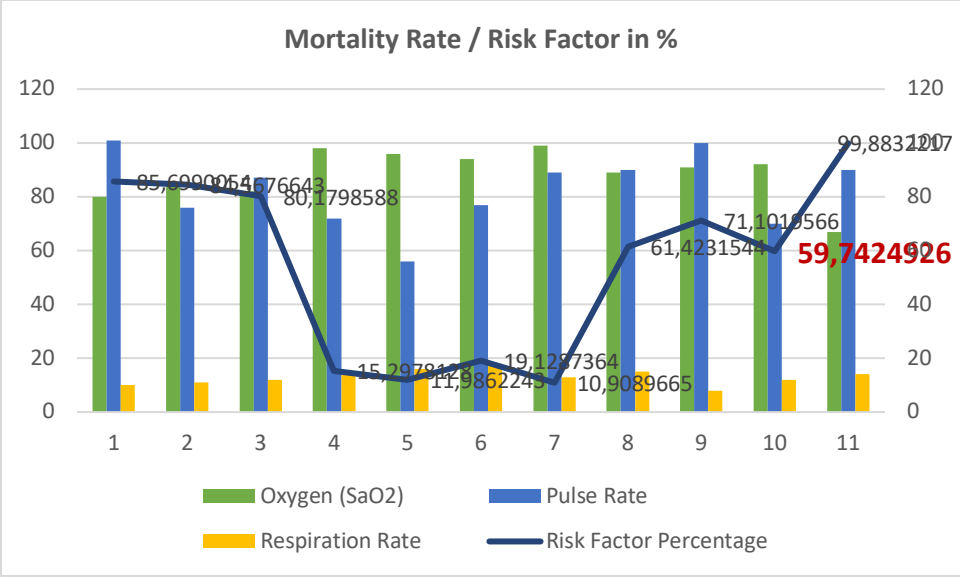


Figure 13. *Clustered column Histogram for Oxygen, pulse rate and Respiration rate with output variable Risk factor measured here in this graph as percentage.*

V. Blood Pressure and Sugar Level

As pointed out earlier that blood pressure and sugar showed very low level of correlation with Risk, but clinically both play a vital role in deciding the risk. The problem with Blood Pressure and sugar level is that these values move in both directions that is Low←--Normal--→High which is not a suitable criteria for fuzzy logic. This was the reason for taking blood pressure and sugar separately as factors for calculating risk [27].
If you have high blood pressure, it's a good idea to take extra care to protect yourself during the coronavirus (COVID-19) outbreak. Early research shows that people with this condition may be more likely to get COVID-19, have worse symptoms or even die from the infection [19].

High Blood Pressure Risks

Growing data shows a higher risk of COVID-19 infections and complications in people with high blood pressure.

Analysis of early data from both China and the U.S. shows that high blood pressure is the most commonly shared pre-existing condition among those hospitalized, affecting between 30% to 50% of the patients. Same also goes for people having moderate to high diabetes.

VI. Conclusion and Future Scope

This paper aims to provide a quick opinion about the degree of Risk involved for a Covid patient on the basis of his Oxygen level, pulse rate and respiration rate. The FRPRPS first applies Fuzzy logic rules on three input variables and a single output variable. This is done iteratively to generate 1000 rows of oxygen level, pulse rate and respiration rate along with the Risk using FuzzyR library. Then, this fuzzy system is manually aligned for correctness if it was needed. The fuzzy system is then passed through a regression model to predict the Risk. Blood Pressure and Sugar were not taken as input variables because they grow in both directions that is normal to low and normal to high. This feature of bidirectional growth cannot be accommodated by fuzzy paradigm [28].

As a future scope, the shape of the membership functions may be altered in order to produce better results [29], [30]. Blood Pressure and Sugar level may be vectored unidirectional in order to be suited as input variables for Fuzzy System to produce results dependent on these two as well [31].

References

[1] Zadeh, Lotfi A. "Fuzzy logic." *Computer* 21.4 (1988): 83-93.
[2] Tanaka, Kazuo. "An introduction to fuzzy logic for practical applications." (1997).
[3] Diamond, Phil, and Hideo Tanaka. "Fuzzy regression analysis." *Fuzzy sets in decision analysis, operations research and statistics*. Springer, Boston, MA, 1998. 349-387.
[4] Ishibuchi, Hisao, and Hideo Tanaka. "Fuzzy regression analysis using neural networks." *Fuzzysets and systems* 50.3 (1992): 257-265.
[5] Schmee, Josef, and Gerald J. Hahn. "A simple method for regression analysis with censored data." *Technometrics* 21.4 (1979): 417-432.
[6] Kaneko, Hiromasa, Masamoto Arakawa, and Kimito Funatsu. "Development of a new regression analysis method using independent component analysis." *Journal of chemical information and modeling* 48.3 (2008): 534-541.
[7] Luks AM, Swenson ER. Pulse Oximetry for Monitoring Patients with COVID-19 at Home. Potential Pitfalls and Practical Guidance. *Ann Am Thorac Soc*. 2020;17(9):1040-1046. doi:10.1513/AnnalsATS.202005-418FR
[8] Usher, Ann Danaiya. "Medical oxygen crisis: a belated COVID-19 response." *The Lancet* 397.10277 (2021): 868-869.
[9] Hayes, Matthew James, and Peter R. Smith. "A new method for pulse oximetry possessing inherent insensitivity to artifact." *IEEE Transactions on Biomedical Engineering* 48.4 (2001): 452-461.
[10] Chang, Yun-Hsi O., and Bilal M. Ayyub. "Fuzzy regression methods–a comparative assessment." *Fuzzy sets and systems* 119.2 (2001): 187-203.
[11] Stein, Felix, et al. "Oxygen provision to fight COVID-19 in sub-Saharan Africa." *BMJ Global Health* 5.6 (2020): e002786.

[12] World Health Organization. *Oxygen sources and distribution for COVID-19 treatment centres: interim guidance, 4 April 2020*. No. WHO/2019-nCoV/Oxygen_sources/2020.1. World Health Organization, 2020.

[13] Daniel, Yvonne, et al. "Haemoglobin oxygen affinity in patients with severe COVID-19 infection." *British Journal of Haematology* 190.3 (2020): e126-e127.

[14] Melin, Patricia, et al. "Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: the case of Mexico." *Healthcare*. Vol. 8. No. 2. Multidisciplinary Digital Publishing Institute, 2020.

[15] Ocampo, Lanndon, and Kafferine Yamagishi. "Modeling the lockdown relaxation protocols of the Philippine government in response to the COVID-19 pandemic: An intuitionistic fuzzy DEMATEL analysis." *Socio-economic planning sciences* 72 (2020): 100911.

[16] Ran, J., Song, Y., Zhuang, Z. *et al.* Blood pressure control and adverse outcomes of COVID-19 infection in patients with concomitant hypertension in Wuhan, China. *Hypertens Res* 43, 1267–1276 (2020). https://doi.org/10.1038/s41440-020-00541-w

[17] Civanlar, M. Reha, and H. Joel Trussell. "Constructing membership functions using statistical data." *Fuzzy sets and systems* 18.1 (1986): 1-13.

[18] Turksen, I. B. "Measurement of membership functions and their acquisition." *Fuzzy sets and systems* 40.1 (1991): 5-38.

[19] Hong, Tzung-Pei, and Chai-Ying Lee. "Induction of fuzzy rules and membership functions from training examples." *Fuzzy sets and Systems* 84.1 (1996): 33-47.

[20] Ali, Omar Adil M., Aous Y. Ali, and Balasem Salem Sumait. "Comparison between the effects of different types of membership functions on fuzzy logic controller performance." *International Journal* 76 (2015): 76-83.

[21] Wang, Hsiao-Fan, and Ruey-Chyn Tsaur. "Insight of a fuzzy regression model." *Fuzzy sets and systems* 112.3 (2000): 355-369.

[22] Bardossy, Andras. "Note on fuzzy regression." *Fuzzy Sets and Systems* 37.1 (1990): 65-75.

[23] Hong, Dug Hun, and Changha Hwang. "Support vector fuzzy regression machines." *Fuzzy sets and systems* 138.2 (2003): 271-281.

[24] Sarvazad, H., et al. "Evaluation of electrolyte status of sodium, potassium and magnesium, and fasting blood sugar at the initial admission of individuals with COVID-19 without underlying disease in Golestan Hospital, Kermanshah." *New Microbes and New Infections* 38 (2020): 100807.

[25] National High Blood Pressure Education Program. *The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents*. No. 5. US Department of Health and Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, National Hight Blood Pressure Education Program, 2005.

[26] Shankhdhar G.K., Sharma R., Darbari M. (2021) SAGRO-Lite: A Light Weight Agent Based Semantic Model for the Internet of Things for Smart Agriculture in Developing Countries. In: Pandey R., Paprzycki M., Srivastava N., Bhalla S., Wasielewska-Michniewska K. (eds) Semantic IoT: Theory and Applications. Studies in Computational Intelligence, vol 941. Springer, Cham. https://doi.org/10.1007/978-3-030-64619-6_12.

[27] Sumit Kumar Mishra, V.K. Singh, Gaurav Kant Shankhdhar, Ontology Development For Wheat Information System, IJRET-International Journal of Research in Engineering and Technology 2015, V0l4/I05.

[28] G. K. Shankhdhar, M. Darbari, R. Sharma and H. Ahmed, "Fuzzy Approach to Select Most Suitable Conflict Resolution Strategy in Multi-Agent System," 2019 International Conference on Cutting-edge Technologies in Engineering (ICon-CuTE), Uttar Pradesh, India, 2019, pp. 9-15.

[29] Gaurav Kant Shankhdhar, Manuj Darbari (2016), Building Custom, Adaptive and Heterogeneous Multi-Agent Systems for Semantic Information Retrieval Using Organizational-Multi-Agent Systems Engineering, O-MaSE, IEEE Explore, ISBN: 978-1-5090-3480-2.

[30] J. Singh and H. Pandey. , "Moving Video Camera Vigilance Using DBSCAN", Reliability: Theory & Applications, No 2 (53) Volume 14, ISSN: 1932-2321, June 2019.

[31] H. Pandey, and M. Darbari., Coalescence of Evolutionary Multi-Objective Decision making approach and Genetic Programming for Selection of Software Quality Parameter, International Journal of Applied Information System (IJAIS), Foundation of Computer Science, New York, USA, Volume 7, No. 11, PP. ISSN: 2249-0868, Nov. 2014.

[32] S. Bansal and H. Pandey, Develop Framework for selecting best Software Development Methodology, International Journal of Scientific and Engineering Research, Volume 5, Issue 4, PP. 1067-1070, ISSN: 2229-5518, Apr. 2014.

[33] M. Srivastava and H. Pandey,A Literature Review of E- Learning Model Based on Semantic Web Technology, International Journal of Scientific and Engineering Research" Volume 5, Issue 10, PP. 174-178, ISSN: 2229-5518, Oct. 2014.

[34] H. Pandey, A New NFA Reduction Algorithm for State Minimization Problem, International Journal of Applied Information Systems (IJAIS), Foundation of Computer Science FCS, New York, USA, Volume 8, No.3, PP. 27-30, ISSN: 2249-0868, Feb. 2015.

[35] H.Pandey, LR Rotation rule for creating Minimal NFA, International Journal of Applied Information Systems (IJAIS), Foundation of Computer Science FCS, New York, USA, Volume 8, No.6, PP. 1-4, ISSN: 2249-0868, Apr. 2015.

Reliability, Availability, Maintainability, and Dependability (RAMD) Analysis of Computer Based Test (CBT) Network System

Abdullahi Sanusi [*1], Ibrahim Yusuf[2] and Na'fiu Hussain[3]

[1]School of Continuing Education, Bayero University, Kano, Nigeria
[2]Department of Mathematical Sciences, Bayero University, Kano, Nigeria
[3]Department of Mathematical Sciences, Bayero University, Kano, Nigeria
asanusi.sce@buk.edu.ng
iyusuf.mth@buk.edu.ng
nafiu_hussaini@yahoo.com

Abstract

Computer Based Test System also known as an e-examinations system, is software that can be used to administer examinations for distant or in-house applicants via internet or in an internet. Computer Based Test System/Software comprises of many components. So, it is vital to ensure its smooth operation, which can be achieved by the proper operation of its components/subcomponents. It is necessary to improve components/subcomponents operational availability. For this reason, the present research proposes to explore Computer Based Test System reliability indices using a RAMD technique at the component/subcomponent level. As a result, all subsystem/component transition diagrams are constructed, and the Chapman-Kolmogorov differential equations are formulated using the Markov birth-death process. For various subsystems/components of the system, numerical findings for reliability, availability, maintainability, and dependability, all of which are crucial to system performance, have been obtained and given in tables and figures. Other measurements, such as MTTF, MTBF, dependability ratio, and dependability minimum have also been obtained. Based on the numerical results, the most significant subsystem/component has been determined and the significance of the research has been emphasized.

Keywords: Reliability, availability, maintainability, dependability, Subsystem/component, Computer Based Test (CBT).

I. Introduction

Examination is one of the most important aspect of educational sectors or institutions for evaluating student performance. One of the elements for determining the standard and efficiency of an educational institution is the regularity and integrity with which examinations are conducted. The computerization of many test operations improves the examination system.

Different types of Computer Based Test (CBT) Software Systems are being used by many educational institutions and examination bodies. The advantages of these technologies over the old technique (Traditional/Convectional) of examination cannot be over stressed as they aid in achieving efficiency and error-free outcomes and computations.

In many nations, there has been an increasing interest in developing and employing computer-based assessments in educational evaluation in recent years. Therefore, it is important to stay up with technological advancements.

Every Computer Based Test Software System must have high levels of reliability, availability, and dependability to be effective. These dimensions mentioned above can be used extensively to assess service quality in a variety of ways. Each subsystem/component of Computer Based Test (CBT) Software System has its own functions or characteristics and it is necessary to analyze the features of these subsystems/components to identify the subsystems/components that mainly influence the performance of the system. To achieve this, RAMD technique is mostly used by system engineers.

RAMD analysis is a critical step in assuring successful operations and production, as it identifies components or subsystems that can be improved. RAMD assesses the system at different stages using various performance modeling methods. RAMD evaluation can be used to derive important performance indicators. These indicators include, MTBF, MTTR, availability, reliability, maintainability, dependability ratio, and dependency minimum. These performance indicators are widely used for planning of maintenance policies to enhance the performance of the system.

Researchers have used different methods to evaluate the performance of various systems in the literature. Aggarwal et al. [1] have used RAMD technique to build a performance model of a dairy plant's skim milk powder production system. Aggarwal et al. [2] applied RAMD analysis to construct a mathematical model for evaluating the performance of serial processes in sugar plant's refining system. Choudhary et al. [3] proposed a way for increasing cement plant reliability. The system's MTBF and MTTR were obtained during a two-year period, and RAMD indices were analyzed. Corvaro et al. [4] have used RAM model to evaluate the operating performance of reciprocating compressors used in the gas and oil industries. De Sanctis et al. [5] provided a methodology for enhancing industry performance and suggested to engineers some maintenance strategies for handling issues such as high costs, safety, and environmental protection. For this, RAMD analysis was performed using equipment from the oil and gas sector as a case study object. Dahiya et al. [6] adopted the RAMD method to evaluate the performance of the sugar industry's A-Pan crystallization system. Garg [8] used a soft computing-based hybridized technique to analyze the performance of an industrial system. Kumar et al. [9] have recently discussed reliability and maintainability investigation of a sewage treatment plant's power producing unit. Kumar et al. [10] have used reliability and availability analysis to estimate the profit of an engineering system with several subsystems arranged in series connection. Malik and Tewari [11] have built a mathematical model for evaluating the performance of a water flow system and suggested some maintenance priorities. Mehta et al. [12] have discussed availability analysis of an industrial system applying supplementary variable technique. Qiu and Cui [13] approved a system reliability performance based on a dependable two-stage failure process, which includes the defect initialization stage and the defect development stage, both of which have competing failures. Based on the cost-free warranty policy, Niwas and Garg [14] proposed a methodology for measuring the reliability and profit of an industrial system. Saini and Kumar [15] analyzed the performance of an evaporation system in the sugar sector via RAMD analysis. Sanusi et al. [16] recently presented performance evaluation of an industrial arranged as series-parallel system. Velmurugan et al. [18] have provided reliability, availability, and maintainability analysis in forming industry. Yusuf, I [19] investigated the availability modelling and evaluation of repairable system subject to minor deterioration under imperfect repairs.

Singh et al. [17] have recently discussed probabilistic assessment of CBT network system consisting four subsystems connected in series using Copula technique. In their work, they evaluated the performance of the CBT network system without taking into account its components or subsystems. They investigated two types of repairs: Copula and General repairs to see how failure and repair affected reliability measures. Their findings showed that when Copula repair is used, the

performance of the CBT network system can be improved. This study therefore addressed the gap left by Singh et al. [17] by investigating Computer Based Test (CBT) Software System reliability measures utilizing a RAMD technique at the component level in order to identify the most critical subsystem/component in the system and to build maintenance plans for this subsystem/component.

This paper is composed of 7 Sections. The first Section contains an introduction and a few brief reviews that are required for this study. The materials and methods are discussed in Section 2. Section 3 is devoted to the description of the system. The results of the RAMD analysis of the system are summarized in Section 4. Numerical simulation is covered in Section 5. The outcome discussion was presented in Section 6, and the paper was concluded in Section 7.

II. Materials and Methods

The tools for computing RAMD measures for the model under consideration are described in this section. All of the data in this study is valid only during a steady-state period when all the failure and repair rates are exponentially distributed and statistically independent.

I. Reliability function

In terms of failure rate, the reliability of a component can be represented as:

$$R(t) = \int_t^{\infty} f(t_0)dt_0. \tag{1}$$

For a component with an exponentially distributed failure rate, equation (1) is reduced to:

$$R(t) = e^{-\lambda t}. \tag{2}$$

II. Availability function

Mathematically, availability is expressed as:

$$A(t) = limA(T) = \frac{MTBF}{MTBF+MTTR}. \tag{3}$$

III. Maintainability

System maintainability can be expressed mathematically as:

$$M(t) = P(T \leq t) = 1 - e^{\left(\frac{-t}{MTTR}\right)} = 1 - e^{-\mu t}. \tag{4}$$

where μ is the constant system's repair rate.

IV. Dependability

The dependability ratio for exponentially distributed random variables is given below:

$$d = \frac{\mu}{\theta} = \frac{MTBF}{MTTR}. \tag{5}$$

The following formula calculates the minimum value of dependability:

$$D_{min} = 1 - \left(\frac{1}{d-1}\right)\left(e^{-Ind/d-1} - e^{-dInd/d-1}\right). \tag{6}$$

V. MTTR

Mean Time To Repair is mathematically expressed as:

$$MTTR = \frac{1}{\alpha}. \quad (7)$$

where α is the system's repair rate.

VI. MTBF

The Mean Time Before Failure for an exponentially dispersed system is as follows:

$$MTBF = \int_0^\infty R(t)\,dt = \int_0^\infty e^{-\mu t}\,dt = \frac{1}{\mu} \quad (8)$$

Where μ is the failure rate.

VII. Exponential distribution.

A random variable X is said to obey an exponential distribution with parameter $\theta > 0$, if its probability density function is given by:

$$f(x,\theta) = \begin{cases} \theta e^{-\theta x}, if\ x \geq 0 \\ 0,\ otherwise \end{cases} \quad (9)$$

VIII. Constant failure rate

The constant hazard rate function can be written as follows:

$$f(t,\theta) = \begin{cases} \theta e^{-\theta t}, if\ t \geq 0 \\ 0,\ otherwise \end{cases} \quad (10)$$

Where θ is constant with probability density function, with $F(t) = 1 - e^{-\theta t}$ & $R(t) = e^{-\theta t}$.

IX. Notations


Failure state of all subsystems.


Operative state of all subsystems.

G, H, I, J, K, L, and M: Represent states in which a subsystem is operating at maximum efficiency.
P, q, r and s: Represent the failure states of subsystem A, B, C, and D, respectively.
$\mu_i, i = 1,2,3,4$: Rate of failure of subsystems A, B, C, and D, respectively.
$\alpha_i, i = 1,2,3,4$: Rate of repair of subsystems A, B, C and D, respectively.
$P_0(t)$: Probability that the system is operating at maximum capacity when it starts up.
$P_i;\ i = 0,1,2,3$: Steady-state probability that the system is in i^{th} state.

X. System description

The Computer Based Test system studied in this research work consists of four distinct subsystems/components. The brief description of the subsystems/components is given below:

Subsystem A (Clients): This subsystem is made up of three active clients. For the system to work, two clients must be operational. When one client of subsystem A fails, the system's capacity is lowered.
Subsystem B (Load Balancer): This subsystem has only one unit, failure of this unit leads to a complete system failure.
Subsystem C (Distributed Database Servers): This subsystem consists of two active servers arranged

in parallel. When one of the two active servers in this subsystem fails, the system operates at a reduced capacity. While failure of the two servers bring the system to a total failure.

Subsystem D (Centralized Distributed Server): There is only one unit of centralized distributed server in this subsystem. When this server fails, the system as a whole fail. Figure 1 below depicts a visual representation of the Computer Based Test system.

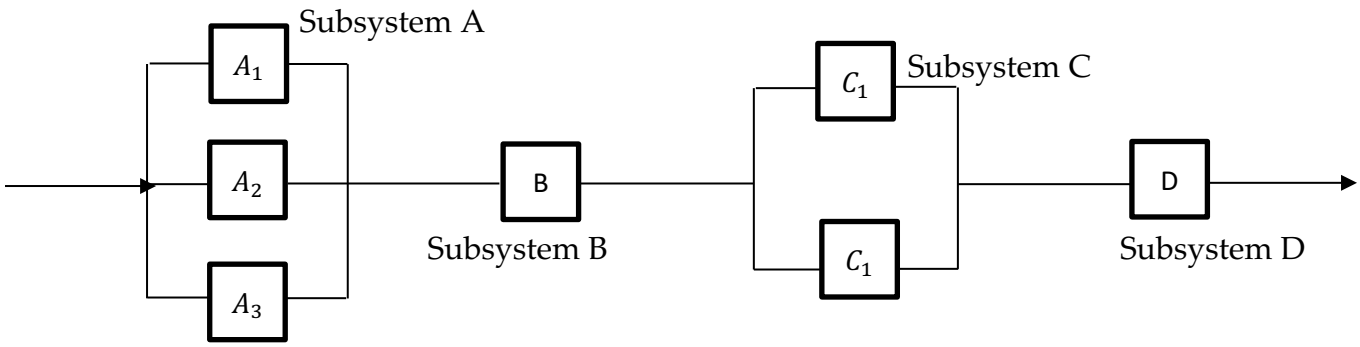


Figure 1: *Reliability block diagram of CBT software system*

III. RAMD analysis of the system

For mathematical modeling of a Computer Based Test Network System, Chapman Kolmogorov differential equations for each subsystem were developed utilizing the Markov birth-death process in this section. Figures 2, 3, 4, and 5 represent transition diagrams for all four subsystems, using the notation from section 2.1 above. System performance indicators such as availability, reliability, maintainability, and dependability have been obtained by solving the appropriate Chapman-Kolmogorov differential equations in a steady-state and employing normalization conditions simultaneously. Table 1 depicts various subsystem failure and repair rates.

Table 1: *Failure and Repair rates of subsystems of CBT network system*

Subsystem	Failure rate	Repair rate
SSA	$\delta_1 = 0.002$	$\eta_1 = 0.35$
SSB	$\delta_2 = 0.0015$	$\eta_2 = 0.40$
SSC	$\delta_3 = 0.005$	$\eta_3 = 0.082$
SSD	$\delta_4 = 0.032$	$\eta_4 = 0.95$

The following are RAMD analysis of subsystems for Computer Based Test (CBT) Network System:

I. RAMD analysis of subsystem A (Clients)

This subsystem has three active clients. The failure rate of all three clients is the same, and the failure of two clients\units causes the entire subsystem to fail. Figure 2 is used to formulate the differential equations for subsystem A, which are stated as follow:

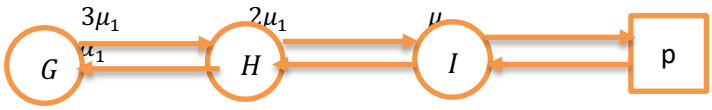


Figure 2: Transition diagram of subsystem A

$$P_0^1(t) = -3\mu_1 P_0 + \alpha_1 P_1, \tag{11}$$

$$P_1^1(t) = -(2\mu_1 + \alpha_1)P_1 + 3\mu_1 P_0 + \alpha_1 P_2, \tag{12}$$

$$P_2^1(t) = -(\mu_1 + \alpha_1)P_2 + 2\mu_1 P_1 + \alpha_1 P_3, \quad (13)$$

$$P_3^1(t) = -\alpha_1 P_3 + \mu_1 P_2. \quad (14)$$

Under steady-state, equations (11)-(14) can be reduced to the following using the initial conditions and taking $t \to \infty$:

$$-3\mu_1 P_0 + \alpha_1 P_1 = 0, \quad (15)$$

$$-(2\mu_1 + \alpha_1)P_1 + 3\mu_1 P_0 + \alpha_1 P_2 = 0, \quad (16)$$

$$-(\mu_1 + \alpha_1)P_2 + 2\mu_1 P_1 + \alpha_1 P_3 = 0, \quad (17)$$

$$-\alpha_1 P_3 + \mu_1 P_2 = 0. \quad (18)$$

Solving equations (15)-(18) recursively and using normalizing condition (i.e., $P_0 + P_1 + P_2 + P_3 = 1$), we have:

$P_0 = \frac{1}{1+3\frac{\mu_1}{\alpha_1}+6\left(\frac{\mu_1}{\alpha_1}\right)^2+6\left(\frac{\mu_1}{\alpha_1}\right)^3}$, $P_1 = 3\frac{\mu_1}{\alpha_1}P_0$, $P_2 = 6\left(\frac{\mu_1}{\alpha_1}\right)^2 P_0$, and $P_3 = 6\left(\frac{\mu_1}{\alpha_1}\right)^3 P_0$.

Now, the steady-state availability is obtained as the summation of all the working state probabilities as:

$$Av_{sysA}(t) = P_0 + P_1 + P_2. \quad (19)$$

Thus, we have the availability of subsystem A as:

$$Av_{sysA}(t) = \frac{\alpha^2+3\alpha_1\mu_1+6\mu_1^2}{\alpha^2+3\alpha_1\mu_1+6\mu_1^2+\frac{6\mu_1^3}{\alpha_1}} = 0.9999. \quad (20)$$

The reliability of the system is given by equation (1). For a component with an exponentially distributed failure rate, equation (1) is reduced to:

$$R(t) = e^{-\mu t}. \quad (21)$$

Thus, the reliability of subsystem A is obtained as:

$$R_{sysA}(t) = e^{-\mu_1 t}, \quad (22)$$

$$R_{sysA}(t) = e^{-0.002t}. \quad (23)$$

Equation (4) calculates the system's maintainability. Thus, the maintainability of subsystem A is presented by equation (25) below.

$$M_{sysA}(t) = 1 - e^{-\alpha_1 t}, \quad (24)$$

$$M_{sysA}(t) = 1 - e^{-0.35t}. \quad (25)$$

Using equations (4), (5), (6), (7), and (8), other performance indicators of subsystem A are given below:

$MTBF = 500h, MTTR = 2.8571h, d = 175.0026, D_{min(sysA)}(t) = 0.9945.$

II. RAMD analysis of subsystem B (Load Balancer)

There is only one unit of load balancer in this subsystem which is connected to the following units in series. Failure of this unit leads to system failure. The differential equations for subsystem B are written in figure 3, and are as follow:

Figure 3: Transition diagram of subsystem B

$$P_0^1(t) = -\mu_2 P_0 + \alpha_2 P_1, \tag{26}$$

$$P_1^1(t) = -\alpha_2 P_1 + \mu_2 P_0, \tag{27}$$

Under steady-state, equations (26) and (27) can be reduced to the following using the initial conditions and taking $t \to \infty$:

$$-\mu_2 P_0 + \alpha_2 P_1 = 0, \tag{28}$$

$$-\alpha_2 P_1 + \mu_2 P_0 = 0, \tag{29}$$

Solving equations (28) and (29) recursively and using normalizing condition (i.e., $P_0 + P_1 = 1$), we have:

$$P_0 = \frac{\alpha_2}{\alpha_2+\mu_2} \; and \; P_1 = \frac{\mu_2}{\alpha_2} P_0,$$

Now, the steady-state availability is obtained as the summation of all the working state probabilities as:

$$Av_{sysB}(t) = P_0. \tag{30}$$

Thus, we have the availability of subsystem B as:

$$Av_{sysB}(t) = \frac{\alpha_2}{\alpha_2+\mu_2} = 0.9963. \tag{31}$$

The reliability of the system is given by equation (1). For a component with an exponentially distributed failure rate, equation (1) is reduced to:

$$R(t) = e^{-\mu t}. \tag{32}$$

Thus, the reliability of subsystem B is obtained as:

$$R_{sysB}(t) = e^{-\mu_2 t}, \tag{33}$$

$$R_{sysB}(t) = e^{-0.0015t}. \tag{34}$$

Equation (4) calculates the system's maintainability. Thus, the maintainability of subsystem B is presented by equation (36) below.

$$M_{sysB}(t) = 1 - e^{-\alpha_2 t}, \tag{35}$$

$$M_{sysB}(t) = 1 - e^{-0.40t}. \tag{36}$$

Using equations (4), (5), (6), (7), and (8), other performance indicators of subsystem B are given

below:

$MTBF = 666.6667h, MTTR = 2.5000h, d = 266.6667, D_{min(sysB)}(t) = 0.9963.$

III. RAMD analysis of subsystem C (Distributed Database Servers)

This subsystem consists of two identical active servers that are connected in a parallel configuration. When one of the subsystem's two active servers fails, the system's capacity is reduced. The failure of the two servers, on the other hand, causes the entire system to fail. Figure 4 shows the differential equations for subsystem C, which are as follow:

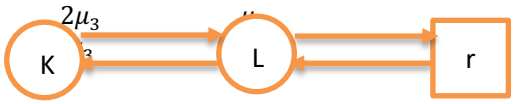


Figure 4: Transition diagram of subsystem C

$$P_0^1(t) = -\mu_3 P_0 + \alpha_3 P_1, \tag{37}$$

$$P_1^1(t) = -(\mu_3 + \alpha_3)P_1 + \mu_3 P_0 + \alpha_3 P_2, \tag{38}$$

$$P_2^1(t) = -\alpha_3 P_2 + \mu_3 P_1. \tag{39}$$

Under steady-state, equations (37)-(39) can be reduced to the following using the initial conditions and taking $t \to \infty$:

$$-\mu_3 P_0 + \alpha_3 P_1 = 0, \tag{40}$$

$$-(\mu_3 + \alpha_3)P_1 + \mu_3 P_0 + \alpha_3 P_2 = 0, \tag{41}$$

$$-\alpha_3 P_2 + \mu_3 P_1. \tag{42}$$

Solving equations (40)-(42) recursively and using normalizing condition (i.e., $P_0 + P_1 + P_2 = 1$), we have:

$P_0 = \frac{1}{\left(1+\frac{\mu_3}{\alpha_3}+\left(\frac{\mu_3}{\alpha_3}\right)^2\right)}, P_1 = \frac{\mu_3}{\alpha_3}P_0$, and $P_2 = \left(\frac{\mu_3}{\alpha_3}\right)^2 P_0$.

Now, the steady-state availability is obtained as the summation of all the working state probabilities as:

$$Av_{sysC}(t) = P_0 + P_1. \tag{43}$$

Thus, we have the availability of subsystem C as:

$$Av_{sysC}(t) = \frac{\alpha_3^2+\alpha_3\mu_3}{\alpha_3^2+\alpha_3\mu_3+\mu_3^2} = 0.9965. \tag{44}$$

The reliability of the system is given by equation (1). For a component with an exponentially distributed failure rate, equation (1) is reduced to:

$$R(t) = e^{-\mu t}. \tag{45}$$

Thus, the reliability of subsystem C is obtained as:

$$R_{sysC}(t) = e^{-\mu_3 t}, \tag{46}$$

$$R_{sysC}(t) = e^{-0.005t}. \tag{47}$$

Equation (4) calculates the system's maintainability. Thus, the maintainability of subsystem C is presented by equation (49) below.

$$M_{sysC}(t) = 1 - e^{-\alpha_3 t}, \tag{48}$$

$$M_{sysC}(t) = 1 - e^{-0.082t}. \tag{49}$$

Using equations (4), (5), (6), (7), and (8), other performance indicators of subsystem C are given below:

$MTBF = 200h, MTTR = 12.1951h, d = 16.4000, D_{min(sysC)}(t) = 0.9492.$

IV. RAMD analysis of subsystem D (Centralized Distributed Server)

There is only one unit in this subsystem. Failure of this unit leads to a complete system failure. The differential equations for subsystem D are written in figure 5, and are as follow:

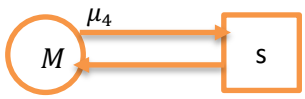


Figure 5: Transition diagram of subsystem D

$$P_0^1(t) = -\mu_4 P_0 + \alpha_4 P_1, \tag{50}$$

$$P_1^1(t) = -\alpha_4 P_1 + \mu_4 P_0. \tag{51}$$

Under steady-state, equations (50) and (51) can be reduced to the following using the initial conditions and taking $t \to \infty$:

$$-\mu_4 P_0 + \alpha_4 P_1 = 0, \tag{52}$$

$$-\alpha_4 P_1 + \mu_4 P_0 = 0. \tag{53}$$

Solving equations (50) and (51) recursively and using normalizing condition (i.e., $P_0 + P_1 = 1$), we have:

$P_0 = \frac{\alpha_4}{\alpha_4+\mu_4}$ and $P_1 = \frac{\mu_4}{\alpha_4} P_0.$

Now, the steady-state availability is obtained as the summation of all the working state probabilities as:

$$Av_{sysD}(t) = P_0. \tag{54}$$

Thus, we have the availability of subsystem D as:

$$Av_{sysD}(t) = \frac{\alpha_4}{\alpha_4+\mu_4} = 0.9674. \tag{55}$$

The reliability of the system is given by equation (1). For a component with an exponentially distributed failure rate, equation (1) is reduced to:

$$R(t) = e^{-\mu t}. \tag{56}$$

Thus, the reliability of subsystem D is obtained as:

$$R_{sysD}(t) = e^{-\mu_4 t}. \tag{57}$$

$$R_{sysD}(t) = e^{-0.032t}. \tag{58}$$

Equation (4) calculates the system's maintainability. Thus, the maintainability of subsystem D is presented by equation (60) below.

$$M_{sysD}(t) = 1 - e^{-\alpha_4 t}. \tag{59}$$

$$M_{sysD}(t) = 1 - e^{-0.95t}. \tag{60}$$

Using equations (4), (5), (6), (7), and (8), other performance indicators of subsystem D are given below:

$MTBF = 31.2500h, MTTR = 1.0526h, d = 29.6884, D_{min(sysD)}(t) = 0.9673.$

V. Numerical simulation

In this section, we present the numerical findings in tables and figures to validate the formulae derived and to provide rapid insight into the system's optimal design.

I. System reliability

Since all four subsystems are connected in a sequential manner. Failure of one subsystem causes the whole system to fail. The following formula calculates the CBT network's overall system reliability:

$$R_{sys}(t) = R_{sysA}(t) \times R_{sysB}(t) \times R_{sysC}(t) \times R_{sysD}(t),$$
$$R_{sysA}(t) = e^{-(0.0405)t} \tag{61}$$

II. System availability

All four subsystems are connected in a sequential manner. One failure causes the entire system to fail. The following formula calculates the CBT network's overall system availability:

$$Av_{sys}(t) = Av_{sysA}(t) \times Av_{sysB}(t) \times Av_{sysC}(t) \times Av_{sysD}(t),$$
$$Av_{sys}(t) = 0.9604. \tag{62}$$

III. System maintainability

All the four subsystems are connected in series; thus, the CBT network's total system maintainability is determined by:

$$M_{sys}(t) = M_{sysA}(t) \times M_{sysB}(t) \times M_{sysC}(t) \times M_{sysD}(t),$$
$$M_{sys}(t) = (1 - e^{-1.7820t}). \tag{63}$$

IV. System dependability

Since all the four subsystems are arranged in series, the CBT network's overall system dependability is given by:

$$D_{min(sys)}(t) = D_{min(sysA)}(t) \times D_{min(sysB)}(t) \times D_{min(sysC)}(t) \times D_{min(sysD)}(t),$$
$$D_{min(sys)}(t) = 0.9097. \tag{64}$$

Table 2 shows a summary of the RAMD results.

Table 2: *RAMD analysis for Computer Based Test (CBT) Network System*

RAMD indices of subsystems	subsystem A	subsystem B	subsystem C	subsystem D
Availability	0.9999	0.9963	0.9965	0.9674
Reliability	$e^{-0.002t}$	$e^{-0.0015t}$	$e^{-0.005t}$	$e^{-0.032t}$
Maintainability	$1-e^{-0.35t}$	$1-e^{-0.40t}$	$1-e^{-0.082t}$	$1-e^{-0.95t}$
Dependability ratio	175.0026	266.6667	16.4000	29.6884
MTBF	$500h$	$666.6667h$	$200h$	$31.2500h$
MTTR	$2.8571h$	$2.5000h$	$12.1951h$	$1.0526h$
Dependability min.	0.9945	0.9963	0.9492	0.9673

Table 3 shows how each subsystem's reliability varies with regard to different time intervals.

Table 3: *Variation in subsystems reliability over time*

Time (t) in (Months)	$R_{sysA}(t)$	$R_{sysB}(t)$	$R_{sysC}(t)$	$R_{sysD}(t)$	$R_{sys}(t)$
0	1.0000	1.0000	1.0000	1.0000	1.0000
10	0.9802	0.9851	0.9512	0.7261	0.6670
20	0.9608	0.9704	0.9048	0.5273	0.4449
30	0.9418	0.9560	0.8607	0.3829	0.2967
40	0.9231	0.9418	0.8187	0.2780	0.1979
50	0.9048	0.9277	0.7788	0.2019	0.1320
60	0.8869	0.9139	0.7408	0.1466	0.0880
70	0.8694	0.9003	0.7047	0.1065	0.0587
80	0.8521	0.8869	0.6703	0.0773	0.0392
90	0.8353	0.8737	0.6376	0.0561	0.0261

Table 4 displays how each subsystem's maintainability varies over time.

Table 4: *Changes in subsystems maintainability over time*

Time (t) in (days)	$M_{sysA}(t)$	$M_{sysB}(t)$	$M_{sysC}(t)$	$M_{sysD}(t)$	$M_{sys}(t)$
0	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.9698	0.9817	0.5596	0.9999	0.9999
20	0.9991	0.9997	0.8060	0.9999	1.0000
30	0.9999	0.9999	0.9146	1.0000	1.0000
40	0.9999	0.9999	0.9623	1.0000	1.0000
50	0.9999	0.9999	0.9834	1.0000	1.0000
60	0.9999	1.0000	0.9927	1.0000	1.0000
70	1.0000	1.0000	0.9968	1.0000	1.0000
80	1.0000	1.0000	0.9986	1.0000	1.0000
90	1.0000	1.0000	0.9986	1.0000	1.0000

Table 5: *Variation in systems reliability as a result of changes in subsystem A's failure rate*

	Subsystem A		System	
Time in (Months)	$\alpha_1 = 0.001$	$\alpha_1 = 0.005$	$\alpha_1 = 0.001$	$\alpha_1 = 0.005$
0	1.00000	1.00000	1.00000	1.00000
10	0.99005	0.95123	0.67368	0.64726
20	0.98020	0.90484	0.45384	0.41895
30	0.97045	0.86071	0.30575	0.27117
40	0.96079	0.81873	0.20598	0.17552
50	0.95123	0.77880	0.13876	0.11361

60	0.94176	0.74082	0.09348	0.07353
70	0.93239	0.70469	0.06298	0.04760
80	0.92312	0.67032	0.04243	0.03081
90	0.91393	0.63763	0.02858	0.01994

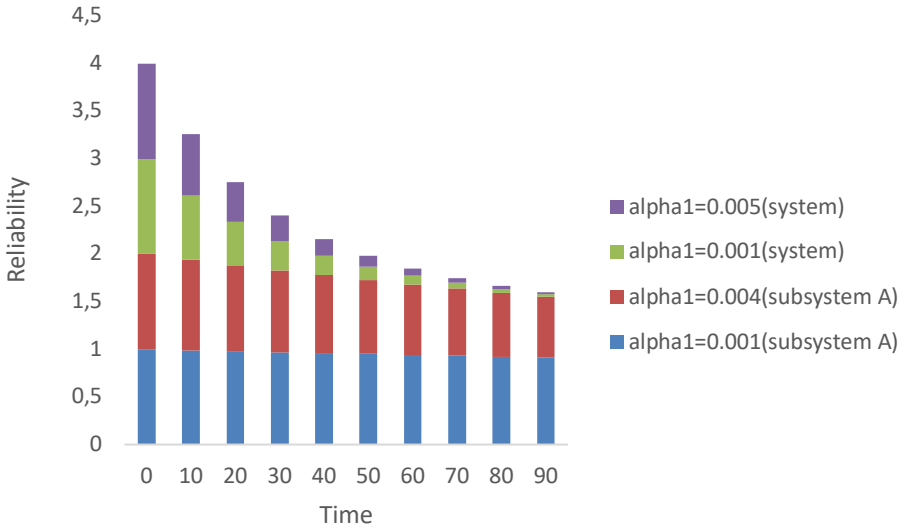


Figure 2: *Effect of α_1 on system reliability and subsystem A reliability*

Table 6: *Variation in systems reliability as a result of changes in subsystem B's failure rate*

	Subsystem B		System	
Time in (Months)	$\alpha_2 = 0.0009$	$\alpha_2 = 0.004$	$\alpha_2 = 0.0009$	$\alpha_2 = 0.004$
0	1.00000	1.00000	1.00000	1.00000
10	0.99104	0.96079	0.67099	0.65051
20	0.98216	0.92312	0.45023	0.42316
30	0.97336	0.88692	0.30210	0.27527
40	0.96464	0.85214	0.20271	0.17907
50	0.95599	0.81873	0.13601	0.11648
60	0.94743	0.78663	0.09126	0.07577
70	0.93894	0.75578	0.06124	0.04929
80	0.93053	0.72615	0.04109	0.03206
90	0.92219	0.69768	0.02757	0.02086

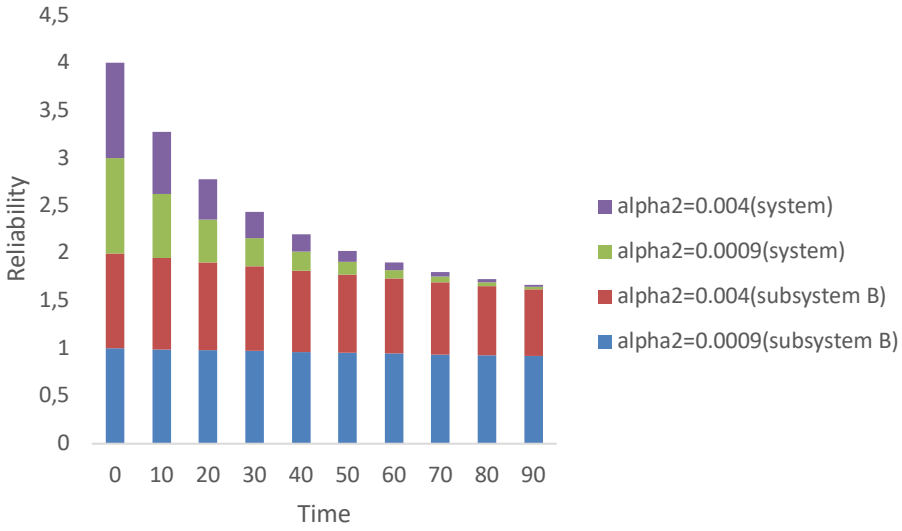


Figure 3: *Effect of α_2 on system reliability and subsystem B reliability*

Table 7: *Variation in systems reliability as a result of changes in subsystem C's failure rate*

	Subsystem C		System	
Time in (Months)	$\alpha_3 = 0.0008$	$\alpha_3 = 0.01$	$\alpha_3 = 0.0008$	$\alpha_3 = 0.01$
0	1.00000	1.00000	1.00000	1.00000
10	0.99203	0.90484	0.69559	0.63445
20	0.98413	0.81873	0.48384	0.40252
30	0.97629	0.74082	0.33655	0.25538
40	0.96851	0.67032	0.23410	0.16203
50	0.96079	0.60653	0.16284	0.10280
60	0.95313	0.54881	0.11327	0.06522
70	0.94554	0.49659	0.07879	0.04138
80	0.93800	0.44933	0.05480	0.02625
90	0.930553	0.40657	0.03812	0.01666

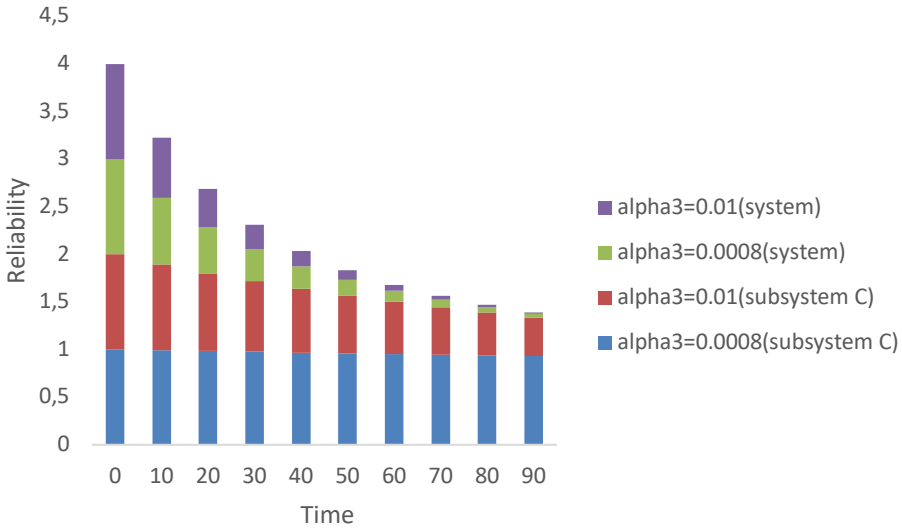


Figure 4: *Effect of α_3 on system reliability and subsystem C reliability*

Table 8: *Variation in systems reliability as a result of changes in subsystem D's failure rate*

	Subsystem D		System	
Time in (Months)	$\alpha_4 = 0.01$	$\alpha_4 = 0.065$	$\alpha_4 = 0.01$	$\alpha_4 = 0.065$
0	1.00000	1.00000	1.00000	1.00000
10	0.90484	0.52205	0.83110	0.47951
20	0.81873	0.27253	0.69073	0.22993
30	0.74082	0.14227	0.57407	0.11025
40	0.67032	0.07427	0.47711	0.05287
50	0.60653	0.03877	0.39653	0.02535
60	0.54881	0.02024	0.32956	0.01255
70	0.49659	0.01057	0.27390	0.00583
80	0.44933	0.00552	0.22764	0.00279
90	0.40657	0.00288	0.18919	0.00134

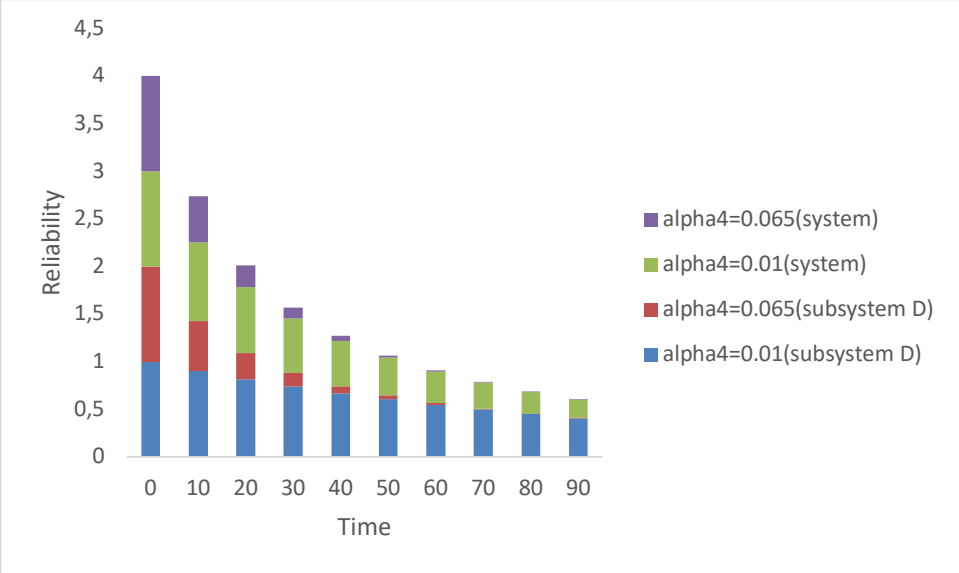


Figure 5: *Effect of α_4 on system reliability and subsystem D reliability*

VI. Result discussion

On the basis of the above analysis, tables 3 and 4 indicate that the reliability and maintainability of the system at time $t = 50$ months are 0.1320 and 1.0000, respectively. At time $t = 50$ months, the system has a comparable value of $R_{sysA}(t) = 0.9048$, $R_{sysB}(t) = 0.9277$, $R_{sysC}(t) = 0.7788$, and $R_{sysD}(t) = 0.2019$. The probability of accomplishing satisfactory maintenance and repair within 50 months is $M_{sys}(t) = 1.0000$, and the maintainability value for the crucial subsystems is $M_{sysA}(t) = 0.9999$, $M_{sysB}(t) = 0.9999$, $M_{sysC}(t) = 0.9834$, and $M_{sysD}(t) = 1.0000$. By exhibiting a form declination, the reliability of the system at time $t = 60$ months is reduced to 0.0880. This is owing to the subsystem D's poor reliability value. This sensitivity study reveals that centralized database server (CDS), which is subsystem D, is the system's most important and sensitive component. This implies that maintaining this subsystem is crucial for increasing overall system reliability. This is supported by this subsystem availability, which is low when compared to the availability of other

subsystems. The importance of maintenance is shown in the value of reliability, i.e. the lower the reliability the necessity of the maintenance. For this reason, system designers and maintenance engineers must devise a strategy for the maintenance of this subsystem. Tables 5-6 show how the reliability of key subsystems and the overall system has changed over time and with varying failure rates. We can see from these tables and figures that the entire system reliability is significantly dependent on the failure rate (α_4) of subsystem D, necessitating close attention to this subsystem. This demonstrates that optimal system reliability can be reached when the overall system's failure rate is low and supporting units are included.

VII. Conclusion

In this paper, the RAMD indices for each subsystem are critically analyzed to find the most sensitive component of the system under consideration. The basic expressions associated with RAMD measurements for each subsystem were obtained and validated through numerical experiment. Table 1 shows the values of failure and repair rates that are assumed for each subsystem. Table 2 presents all the RAMD measurements for each subsystem. The influence of varying failure rates on subsystems and system reliability is shown in tables 3, 4, 5, and 6 and their corresponding figures 2, 3, 4, and 5. Based on the numerical findings for a given case in tables 2-6 and figures 2-5, it is concluded that the centralized database server is the Computer Based Test (CBT) Software System's significant and delicate component. It is widely accepted that system failure during an examination will have negative impact on educational standards, and that there may even be a catastrophe. As a result, if the models/results given in this paper are modified, management will be able to avoid incorrect evaluations and erroneous decision-making, resulting in unnecessary expenditures and drop in educational standards. Furthermore, the established strategy (RAMD technique) for maintenance policies for the model under consideration may be recommended in order to increase system smooth operation and reduce system failure rate. These are the findings of this present research. In our future research, we will incorporate minimal repair at the failure of each subsystem/component.

References

[1] Aggarwal A, Kumar S, Singh V, (2015). Performance modeling of the skim milk powder production system of a dairy plant using RAMD analysis. Int J Qual Reliab Manag 32(2):167–181.

[2] Aggarwal AK, Kumar S, Singh V, (2017). Performance modeling of the serial processes in refining system of a sugar plant using RAMD analysis. Int J Syst Assur Eng Manag 8(2):1910–1922.

[3] Choudhary D, Tripathi M, Shankar R, (2019). Reliability, availability and maintainability analysis of a cement plant: a case study. Int J Qual Reliab Manag. https ://doi.org/10.1108/IJQRM-10-2017-0215.

[4] Corvaro F, Giacchetta G, Marchetti B, and Recanati M, (2017). "Reliability, Availability, Maintainability (RAM) study, on reciprocating compressors API 618 Petroleum", Vol. 3, 266-272.

[5] De Sanctis I, Paciarotti C, Di Giovine O, (2016). Integration between RCM and RAM: a case study. Int J Qual Reliab Manag 33(6):852–880.

[6] Dahiya O, Kumar A, Saini M, (2019). Mathematical modeling and performance evaluation of A-Pan crystallization system in a sugar industry. SN Appl Sci. https://doi.org/10.1007/s42452-019-0348-0.

[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. 1:1507, https://doi.org/10.1007/s42452-019-1544-7.

[8] Harish Garg (2017). Performance analysis of an industrial system using soft computing based hybridized technique. J Braz Soc Mech Sci Eng 39(4):1441–1451.

[9] Kumar A, Goyal D, and Saini M, (2020). Reliability and maintainability analysis of power generating unit of sewage treatment plant. Int. J. of Stat. and Reliab. Eng. 7(1), 41-48.

[10] Kumar A, Pant S, Singh SB, (2017). Availability and cost analysis of an engineering system involving subsystems in series configuration. Int J Qual Reliabil Manag 34(6):879–894.

[11] Qiu, Q.; Cui, L. (2018). Reliability evaluation based on a dependent two-stage failure process with competing failures. Appl. Math. Model. 64, 699–712.

[12] Ram Niwas and Harish Garg (2018). An approach for analyzing the reliability and profit of an industrial system based on the cost-free warranty policy. Journal of the Brazilian Society of Mechanical Sciences and Engineering, Springer, 40, pp. 1-9.

[13] Malik S, Tewari PC, (2018). Performance modeling and maintenance priorities decision for water flow system of a coal based thermal power plant. Int J Qual Reliabil Manag 35(4):996–1010.

[14] Mehta M., Singh J., and Sharma S, (2018). "Availability Analysis of an Industrial System using Supplementary Variable Technique," Jordan journal of mechanical and industrial engineering, vol. 12, no. 4, pp. 245-251, ISSN 1995-6665.

[15] Saini M, and Kumar A, (2019). "Performance analysis of evaporation system in sugar industry using RAMD analysis". J Braz Soc Mech Sci Eng 41:4.

[16] Sanusi, A., Yusuf, I., Mamuda. B.Y. (2020). Performance evaluation of an industrial configured as series-parallel system, Journal of Mathematical and Computational Science. 10, pp. 692-712.

[17] Singh, V.V., Lado Ismail, A.K., Yusuf, I. and Abdullahi, A.H. (2021). Probabilistic Assessment of Computer-Based Test (CBT) Network System Consists of Four Subsystems in Series Configuration Using Copula Linguistic Approach. Journal of Reliability and Statistical Studies.

[18] Velmurugan K, Venkumar P, and Sudhakarapandian R., (2019). Reliability Availability Maintainability Analysis in Forming Industry. International Journal of Engineering and Advanced Technology; 91S4, 822–828.

[19] Yusuf, I. (2016). Reliability modeling of a parallel system with a supporting device and two types of preventive maintenance. International Journal of Operational Research. 25(3), pp. 269-287.

Performance Evaluation of a Complex Reverse Osmosis Machine System in Water Purification using Reliability, Availability, Maintainability and Dependability Analysis

Anas Sani Maihulla
•
Department of Mathematics, Sokoto State University, Sokoto, Nigeria
anasmaihulla@gmail.com

Ibrahim Yusuf
•
Department of Mathematical Sciences, Bayero University, Kano
iyusuf.mth@buk.edu.ng

Saminu I. Bala
•
Department of Mathematical Sciences, Bayero University, Kano
sibala.mth@buk.edu.ng

Abstract

Today, reverse osmosis (RO) is a critical technique in the production of fresh water all over the world. As a result, downtimes due to repairing operations (after breakdowns, membrane blockage, pressure losses, etc.) or preventative maintenance (cleaning of membranes, component replacements, etc.) must be kept to a minimum in duration and frequency to guarantee optimum availability. Indeed, enhancing the availability (or dependability) of the RO plant as a whole system leads to a significant decrease in operating and maintenance expenses. We look at a recursive technique for reliability, availability, maintainability, and dependability in this study (RAMD. In addition, the efficacy of a RO unit, mean time to failure (MTTF), mean time to repair (MTTR), and dependability ratio were evaluated. The primary goal is economic optimization. For the method's validation, we utilized data from a RO unit that had a repair rate and a failure rate during a one-year period. It was demonstrated that all subsystems (pretreatment, dosage, etc.) had high availability. The high-pressure pump has a somewhat lower availability. For example, 0.59113 was the lowest availability for all subsystems, and it is for the RO membrane, which is where the majority of the purifications take place. A sensitivity analysis was performed to identify the essential components for the RO plant's availability. The collected findings demonstrate that the availability, reliability, dependability, and maintainability of the high-pressure pump have a significant impact on the overall system availability. As a result, special care should be given in the selection and maintenance of the high-pressure pump.

Keywords: Reverse osmosis, reliability, failure rate, repair rate.

I. Introduction

Water scarcity is worsening as a result of globalization. The water cycle is being disrupted as a result of the world's significant changes in climatic pattern, Muhammad F.I. [1]. Groundwater, which is either depleted to a certain level or polluted, is a less expensive and more reliable supply of water.

Polluted water may contain biological or inorganic materials as residuals. S.L. Brown et al. [2]. M. Badruzzaman. In 2019 et al. [3] investigated the selection of pretreatment methods for seawater reverse osmosis facilities. F. Saffarimiandoab et al. [4] conducted study on the biofouling behavior of zwitterionic silane covered reverse osmosis membranes contaminated by marine microorganisms. Evita A. et al. [5] conducted the study named a strategy plan for the reuse of treated municipal wastewater for agricultural irrigation on the island of CreteSlvia C Oliveira and Marcos Von Sperling [6] created a reliability study for wastewater treatment plants. Slvia C Oliveira and Marcos Von Sperling did research on reverse osmosis difficulties such as pressure drop, mass transfer, turbulence, and unsteadiness [7Seawater pretreatment for reverse osmosis: Chemistry, pollutants, and coagulation was investigated by James K. Edzwald and Johannes Haarhoff [8]. M.F. Idrees [9] also works with the Performance Analysis and Treatment Technologies of a Reverse Osmosis Plan. C. Li, S. Besarati, and colleagues [10] conducted research on reverse osmosis desalination powered by a low temperature supercritical organic Rankine cycle a few years agoAutomation and dependability are critical components of every advanced reverse osmosis plant in order to fulfill environmental and economic criteria [11]. [12] Developed a computational model based on diffusion and convection transport mechanisms and the concentration polarization concept to predict the performance of a RO membrane using different feed water concentrations, feed flow rates, feed water pressures, membrane specifications, and feed water properties. [13] conducted a study of the concepts and categorization of membrane distillation, with an emphasis on the variables influencing it and ways to improving its efficacy. [14] created a model with five input factors (feed temperature, feed total dissolved solids (TDS), trans-membrane pressure (TMP), feed flow rate, and time) and two output parameters (permeate TDS and flow rate) to estimate the performance of a saltwater reverse osmosis (SWRO) desalination plant It was then used to simulate feed water temperature. [15] compared two hollow fiber module designs (inside/out and outside/in). [16] Experiments were conducted using pure water and NaCl solutions ranging from 15 g/L to 300 g/L, as well as two different fiber materials and structures. Vacuum membrane distillation (VMD) is a method of desalinating saltwater. The two designs were evaluated in terms of pure water permeability and global heat transfer coefficient. It is described how hydrodynamics affects global heat and mass transport coefficients. [17] Investigated the chemical profile, antioxidant and anti-obesity effects of concentrated fractions derived from micro-filtered OMW processed by direct contact membrane distillation (DCMD). Some phenols selected as phytochemical indicators were quantified using ultrahigh performance liquid chromatography (UHPLC). [18] A sequential Direct Contact Membrane Distillation (DCMD) and a Reverse Osmosis (RO) hybrid membrane system were used to treat the pollutants found in olive mill wastewater (total organic carbon (TOC), dissolved organic carbon (DOC), total phosphorus (TP), total nitrogen (TN), and total polyphenols. The effects of permeate flux and pressures on pollutant parameter removals were also investigated. Freshwater shortage has been identified as one of the major issues that humans must solve in the twenty-first century. [19] Investigated how an environmentally friendly, cost-effective, and energy-efficient membrane distillation (MD) process can reduce pollution caused by industrial and domestic wastes. RO is a technique that separates and removes dissolved solids, organics, pyrogens, submicron colloidal debris, color, nitrate, and bacteria from water using semi-permeable spiral wound membranes. Under pressure, feed water is supplied via a semipermeable membrane, where water penetrates the membrane's minute holes and is delivered as filtered water known as permeate water [20]. [21] Investigated different reliability metrics of STP generators using the RAMD method at the component level. For all generator subsystems, mathematical models based on the Markovian birth death process have been developed. These models are extremely useful in analyzing generator reliability, maintainability, and availability. Many years ago, M. D. and Hajeeh Chaudhuri [22] investigated the reliability and availability of reverse osmosis RAMD analysis of complex system is highly beneficial in identifying feasible design modifications for reliability, availability, maintainability, and dependability (RAMD). These changes are necessary to improve the system's dependability, mean time between failures, and availability. The enactment of the industrial system

is dependent on the legislation of the components Monika S. and Ashish K. [26].

Our motivation for exploring the reverse osmosis machine system stems from a serious issue that the water purification industries are encountering due to machine system subsystem failure. And the resulting slow progress in technological advancement in water purification, as well as its importance in the lives of people all over the world Industries are striving hard to keep up with the increasing complexity of machine systems. From the finding of the paper, RAMD analysis used to test the strength, efficiency, and performance improvement of the RO system. Where strength, efficiency and performance improvement of the RO system are determined, the users will be able to serve the cost of medical treatment due to un-pure water. Save from aquatic pollutants. The work is divided into four pieces, including the current introduction. RAMD indices for subsystems are commented on in the second section. The third portion included system description and numerous important definitions; notations are included. Section 4 includes RAMD analysis. Section 5 is devoted to the results' conclusion and implications.

II. RAMD indices for subsystems

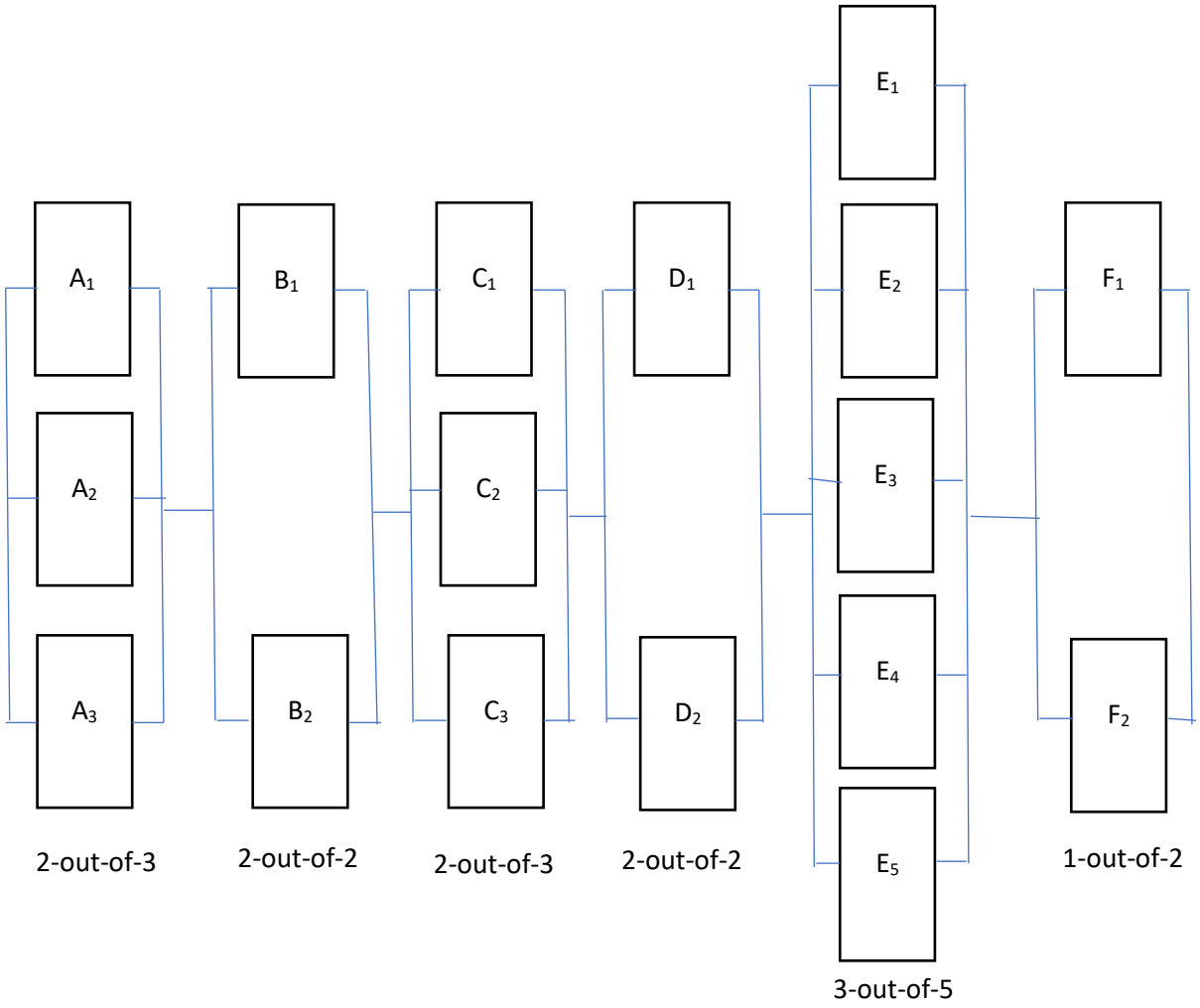


Figure 1: *Block diagram for the RO series-parallel system*

Notations and their meaning

: Represent the system in full capacity state

: Represent the system in reduced but operational state

: Represent the system in failed state

Represent the initial state of the system working in full capacity state.

P_1 : Represent the state in which one parallel unit is failed

P_2 : Represent the state in which two parallel unit is failed

P_3 : Represent the state in which three parallel unit is failed

$K_{i\ \ i=i,2,\ldots,6}$ Represent the failure rates subsystems

$\xi_{j\ \ j=1,2,\ldots,6}$ Represent the repair rates subsystems

$P_x(t)$ Probability to remain at xth state at time t

$\frac{d}{dt}P_x(t)$, x= 0,1,2,3. Represent the derivative with respect to time t

$A_{i,\ i=1,2,3}$ Units in the subsystem 1.

$B_{j,\ j=1,2}$ Units from subsystem 2

$C_{k,\ k=1,2,3}$ Units from subsystem 3

$D_{n,\ n=1,2}$ Units from subsystem 4

$E_{m,\ m=1,2,3,4,5}$ Units from subsystem 5

$F_{x,\ x=1,2}$ Units from subsystem 6.

RAMD indices for subsystem 1 (raw water tank)

Raw water tank: Raw water tanks are used to store raw water temporarily until it is treated. This adaptable tank may be folded entirely and compactly for travel. Because the tank is pop-up, no rods or poles are required. Two out of three parallel subsystem, failure of any two can cause the failure of the entire system.

$$\frac{d}{dt}P_0(t) = -2K_1P_0 + \boldsymbol{\xi_1}P_1K_1 \quad (1)$$

$$\frac{d}{dt}P_1(t) = -(K_1 + \boldsymbol{\xi_1})P_1 + 2K_1P_0 + \boldsymbol{\xi_1}P_2 \quad (2)$$

$$\frac{d}{dt}P_2(t) = -\boldsymbol{\xi_1}P_2 + K_1P_1 \quad (3)$$

Under steady state, the state's probabilities in equation (1) - (3) are as follows

$$P_1 = \frac{2K_1}{\boldsymbol{\xi_1}}P_0 \quad (4)$$

$$P_2 = \frac{2K_1^2\boldsymbol{\xi_1}}{\xi_1^2}P_0 \quad (5)$$

Using normalization condition

$$P_0 + P_1 + P_2 = 1 \quad (6)$$

Substituting (4) and (5) into (6) we have

$$P_0 + \frac{2K_1}{\boldsymbol{\xi_1}}P_0 + \frac{2K_1^2\boldsymbol{\xi_1}}{\xi_1^2}P_0 = 1$$

$$P_0 = \frac{\boldsymbol{\xi_1}}{\boldsymbol{\xi_1}+2K_1^2+2K_1} \quad (7)$$

Availability = $A_{s1} = \left(\frac{0.07}{0.08005}\right) = 0.87445$

Table 4 contains important device output metrics that have been extracted.
Dependability of subsystem 1

$$D_{min} = 1 - (\frac{1}{d-1})(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ ln\ d}{d-1}}) \quad (8)$$

$$d = \frac{\xi}{\mathrm{K}} = \frac{MTBF}{MTTR} \quad (9)$$

$d_1 = \frac{\xi_1}{\mathrm{K}_1} = \frac{0.07}{0.005} = 14.00$
$D_{\min\,(s1)} = 1 - (\frac{1}{14-1})(e^{-\frac{2.6391}{13}} - e^{\frac{36.9468}{13}})$
$D_{\min\,(s1)} = 1 - 0.05831$

The reliability of subsystem 1

$$R_{s1}(t) = e^{-0.005t} \quad (10)$$

Maintainability of subsystem 1

$$M_{s1}(t) = 1 - e^{-0.07t} \quad (11)$$

Other performance measures of system effectiveness of subsystem 1 are as follows
MTBF = 200.00
MTTR = 14.286
Dependability ratio = 14.00

RAMD indices for subsystem 2 (sand filter)

Sand Filter: George Solt CEng, F. IChem E. [23] Sand filters are widely used in water purification and remove suspended matter by a completely different mechanism. Instead of the water passing through small orifices through which particles cannot pass, it runs through a bed of filter medium, typically 0.75 mm sand 750 mm deep. Two out of two series subsystem, failure of any one can cause the failure of the entire system.

$$\frac{d}{dt}P_0(t) = -2\mathrm{K}_2P_0 + \boldsymbol{\xi_2}P_1 \quad (12)$$

$$\frac{d}{dt}P_1(t) = 2\mathrm{K}_2P_0 - \boldsymbol{\xi_2}P_1 \quad (13)$$

Under steady state, equation (8) and (9) reduces to

$$-2\mathrm{K}_2P_0 + \boldsymbol{\xi_2}P_1 = 2\mathrm{K}_2P_0 - \boldsymbol{\xi_2}P_1 \quad (14)$$

and

$$P_1 = \frac{2\mathrm{K}_2}{\boldsymbol{\xi_2}}P_0 \quad (15)$$

Using normalization condition

$$P_0 + P_1 = 1 \quad (16)$$

Substituting (10) into (11) we have

$$P_0 + \frac{2\mathrm{K}_2}{\boldsymbol{\xi_2}}P_0 = 1 \quad (17)$$

$$\mathrm{P}_0 = \frac{\boldsymbol{\xi_2}}{\boldsymbol{\xi_2}+2\mathrm{K}_2} \quad (19)$$

Availability = $A_{s2} = \left(\frac{0.09}{0.11}\right) = 0.81818$

Table 4 contains important device output metrics that have been extracted.
Dependability of subsystem 2

$$D_{min} = 1 - (\frac{1}{d-1})(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ ln\ d}{d-1}}) \quad (20)$$

$$d = \frac{\xi}{\mathrm{K}} = \frac{MTBF}{MTTR} \quad (21)$$

$d_2 = \frac{\xi_2}{\mathrm{K}_2} = \frac{0.09}{0.01} = 9.00$
$D_{\min\,(s2)} = 1 - (\frac{1}{9-1})(e^{-0.27465} - e^{-2.47188})$

$D_{\min\,(s2)} = 0.91557$
The reliability of subsystem 2
$R_{s2}(t) = e^{-0.01t}$

Maintainability of subsystem 2
$M_{s1}(t) = 1 - e^{-0.09t}$
Other performance measures of system effectiveness of subsystem 1 are as follows
MTBF = 100.00
MTTR = 11.111
Dependability ratio = 9.00

RAMD indices for subsystem 3 (activated carbon filter)

Activated carbon filter: Y. K. Siong et al. [24] is used to purify water without leaving any harmful chemicals. Prototype is being made by using activated carbon and ultraviolet radiation system for water treatment. Surface area and porosity analysis. Scanning electron microscopy (SEM) is used to obtain the magnified image of GAC-A and GAC-B for comparison between the surface morphology. Two out of three parallel subsystem of the activated carbon filter were considered, failure of any two can cause the failure of the entire system.

$$\frac{d}{dt}P_0(t) = -2\mathrm{K}_3P_0 + \boldsymbol{\xi_3}P_1 \quad (22)$$

$$\frac{d}{dt}P_1(t) = -(\mathrm{K}_3 + \boldsymbol{\xi_3})P_1 + 2\mathrm{K}_3P_0 + \boldsymbol{\xi_3}P_2 \quad (23)$$

$$\frac{d}{dt}P_2(t) = -\boldsymbol{\xi_3}P_2 + \mathrm{K}_3P_1 \quad (24)$$

Under steady state, equation (13) - (15) reduces to

$$P_1 = \frac{2\mathrm{K}_3}{\boldsymbol{\xi_3}}P_0 \quad (25)$$

Substituting (16) into (15)

$$P_2 = \frac{2\mathrm{K}_3^2}{\xi_3^2}P_0 \quad (26)$$

Using normalization condition

$$P_0 + P_1 + P_2 = 1 \quad (27)$$

Substituting (16) and (17) into (18) we have

$$P_0 + \frac{2\mathrm{K}_3}{\boldsymbol{\xi_3}}P_0 + \frac{2\mathrm{K}_3^2}{\xi_3^2}P_0 = 1 \quad (28)$$

$$\mathrm{P}_0 = \frac{\xi_3^2}{\xi_3^2 + 2\mathrm{K}_3^2 + 2\,\boldsymbol{\xi_3}\mathrm{K}_3} \quad (29)$$

Availability = $A_{s3} = \left(\frac{0.0121}{0.01585}\right) = 0.76341$

Table 4 contains important device output metrics that have been extracted.
Dependability of subsystem 3

$$D_{min} = 1 - \left(\frac{1}{d-1}\right)\left(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ln d}{d-1}}\right) \quad (30)$$

$$d = \frac{\xi}{\mathrm{K}} = \frac{MTBF}{MTTR} \quad (31)$$

$d_3 = \frac{\xi_3}{\mathrm{K}_3} = \frac{0.11}{0.015} = 7.33$
$D_{\min\,(s3)} = 1 - \left(\frac{1}{6.33}\right)\left(e^{-0.31469} - e^{-2.30666}\right)$
$D_{\min\,(s3)} = 0.90041$
The reliability of subsystem 3
$R_{s3}(t) = e^{-0.015t}$

Maintainability of subsystem 1
$M_{s1}(t) = 1 - e^{-0.011t}$
Other performance measures of system effectiveness of subsystem 1 are as follows
MTBF = 66.667
MTTR = 9.091
Dependability ratio = 7.3333

RAMD indices for subsystem 4 (precision filter)

$$\frac{d}{dt}P_0(t) = -2\mathrm{K}_4P_0 + \boldsymbol{\xi_4}P_1 \quad (32)$$

$$\frac{d}{dt}P_1(t) = 2\mathrm{K}_4P_0 - \boldsymbol{\xi_4}P_1 \quad (33)$$

Under steady state, equation (20) and (21) reduces to

$$-2\mathrm{K}_4P_0 + \boldsymbol{\xi_4}P_1 = 2\mathrm{K}_4P_0 - \boldsymbol{\xi_4}P_1 \quad (34)$$

and

$$P_1 = \frac{2\mathrm{K}_4}{\boldsymbol{\xi_4}}P_0 \quad (35)$$

Using normalization condition

$$P_0 + P_1 = 1 \quad (36)$$

Substituting (22) into (23) we have

$$P_0 + \frac{2\mathrm{K}_4}{\boldsymbol{\xi_4}}P_0 = 1 \quad (37)$$

$$\mathrm{P}_0 = \frac{\boldsymbol{\xi_4}}{\boldsymbol{\xi_4}+2\mathrm{K}_4} \quad (38)$$

Availability = $A_{s4} = \left(\frac{0.13}{0.17}\right) = 0.76471$

Table 4 contains important device output metrics that have been extracted.
Dependability of subsystem 4

$$D_{min} = 1 - \left(\frac{1}{d-1}\right)\left(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ ln\ d}{d-1}}\right) \quad (39)$$

$$d = \frac{\boldsymbol{\xi}}{\mathrm{K}} = \frac{MTBF}{MTTR} \quad (40)$$

$d_4 = \frac{\boldsymbol{\xi_4}}{\mathrm{K}_4} = \frac{0.13}{0.02} = 6.50$
$D_{\min\,(s4)} = 1 - \left(\frac{1}{5.5}\right)\left(e^{-0.34033} - e^{-2.21213}\right)$
$D_{\min\,(s4)} = 0.89053$
The reliability of subsystem 4
$R_{s4}(t) = e^{-0.020t}$

Maintainability of subsystem 4
$M_{s4}(t) = 1 - e^{-0.13t}$
Other performance measures of system effectiveness of subsystem 1 are as follows
MTBF = 50.00
MTTR = 7.6923
Dependability ratio = 6.500

RAMD indices for subsystem 5 (RO membrane)

Reverse Osmosis Membrane P. A. Taylor [25]. RO membranes are normally deployed as cross-flow filters, where the high velocity of the wastewater along the filter keeps the flow turbulent which helps control the thickness of the solids on the filter and reduces plugging of the filter. Three out of five parallel subsystem of the RO membrane were considered, failure of any three can cause the failure of the entire system.

$$\frac{d}{dt}P_0(t) = -2K_5P_0 + \xi_5 P_1 \quad (41)$$

$$\frac{d}{dt}P_1(t) = -(2K_1 + \xi_1)P_1 + 3K_5P_0 + \xi_5 P_2 \quad (42)$$

$$\frac{d}{dt}P_2(t) = -(K_5 + \xi_5)P_2 + 2K_5P_1 + \xi_5 P_3 \quad (43)$$

$$\frac{d}{dt}P_3(t) = -\xi_5 P_5 + K_5P_5 \quad (44)$$

Under steady state, equation (25) - (27) reduces to

$$P_1 = \frac{3K_5}{\xi_5}P_0 \quad (45)$$

Substituting (29) into (26)

$$P_2 = \frac{6K_5^2}{\xi_5}P_0 \quad (46)$$

Substituting (30) into (28) we have

$$P_3 = \frac{6K_5^2}{\xi_5^2}P_0 \quad (47)$$

Using normalization condition

$$P_0 + P_1 + P_2 + P_3 = 1 \quad (48)$$

Substituting (30) and (31) into (33) we have

$$P_0 + \frac{3K_5}{\xi_5}P_0 + \frac{6K_5^2}{\xi_5}P_0 + \frac{6K_5^2}{\xi_5^2}P_0 = 1 \quad (49)$$

$$P_0 = \frac{\xi_5^2}{\xi_5^2 + 3K_5\xi_5 + 6\xi_5K_5^2 + 6K_5^2} \quad (50)$$

Availability = $A_{s5} = (\frac{0.0225}{0.0380625}) = 0.59113$

Table 4 contains important device output metrics that have been extracted.
Dependability of subsystem 5

$$D_{min} = 1 - (\frac{1}{d-1})(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ ln\ d}{d-1}}) \quad (51)$$

$$d = \frac{\xi}{K} = \frac{MTBF}{MTTR} \quad (52)$$

$d_5 = \frac{\xi_5}{K_5} = \frac{0.15}{0.025} = 6.00$

$D_{\min(s5)} = 1 - \left(\frac{1}{5}\right)(e^{-0.35835} - e^{-2.15011})$

$D_{\min\ (s5)} = 0.88353$

The reliability of subsystem 5

$R_{s5}(t) = e^{-0.025t}$

Maintainability of subsystem 5

$M_{s5}(t) = 1 - e^{-0.15t}$

Other performance measures of system effectiveness of subsystem 1 are as follows
MTBF = 40.00
MTTR = 6.6667
Dependability ratio = 6.00

RAMD indices for subsystem 6 (Water producing tank)

Water producing Tank: The water is now pure and is pumped into a tank, where it is kept pressured until the faucet is turned on. The tank has two bladders that pressurize the water, allowing it to enter and escape as needed. The tank is constantly under pressure, and water only fills it to around two-thirds of the water inflow pressure. A bladder filled with compressed air sits at the bottom of the tank, and a butyl water bladder, a thick substance comparable to the interior lining of a steel food can, is at the top. When you turn on the faucet, the air pressure sends the water out in a constant stream simultaneously, the intake valve opens to allow more water in, maintaining a constant level

of pressure driving the water out. Two out of two series subsystem for the Water producing tank, failure of any one can cause the failure of the entire system.

$$\frac{d}{dt}P_0(t) = -2K_6P_0 + \xi_6P_1 \quad (53)$$

$$\frac{d}{dt}P_1(t) = 2K_6P_0 - \xi_6P_1 \quad (54)$$

Under steady state, equation (34) and (35) reduces to

$-2K_6P_0 + \xi_6P_1 = 2K_6P_0 - \xi_6P_1$

and

$$P_1 = \frac{2K_6}{\xi_6}P_0 \quad (55)$$

Using normalization condition

$$P_0 + P_1 = 1 \quad (56)$$

Substituting (36) into (37) we have

$$P_0 + \frac{2K_6}{\xi_6}P_0 = 1 \quad (57)$$

$$P_0 = \frac{\xi_6}{\xi_6+2K_6} \quad (58)$$

Availability = $A_{s6} = \left(\frac{0.17}{0.23}\right) = 0.73913$

Table 4 contains important device output metrics that have been extracted.

Dependability of subsystem 6

$$D_{min} = 1 - \left(\frac{1}{d-1}\right)\left(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ ln\ d}{d-1}}\right) \quad (59)$$

$$d = \frac{\xi}{K} = \frac{MTBF}{MTTR} \quad (60)$$

$d_6 = \frac{\xi_6}{K_6} = \frac{0.17}{0.03} = 5.67$

$D_{\min\,(s6)} = 1 - \left(\frac{1}{4.67}\right)\left(e^{-0.37156} - e^{-2.10675}\right)$

$D_{\min\,(s6)} = 0.87837$

The reliability of subsystem 6

$R_{s6}(t) = e^{-0.030t}$ Type equation here.

Maintainability of subsystem 6

$M_{s6}(t) = 1 - e^{-0.17t}$

Other performance measures of system effectiveness of subsystem 1 are as follows

MTBF = 33.333

MTTR = 5.8824

Dependability ratio = 5.6667

System description

The RO system is used to treatment of water. In the RO system, cells are arranged in series-parallel configuration. RO system consists of six subsystems as described below.

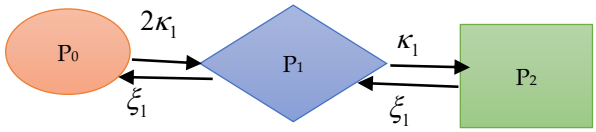


Figure 2: *Transition diagram of raw water tank*

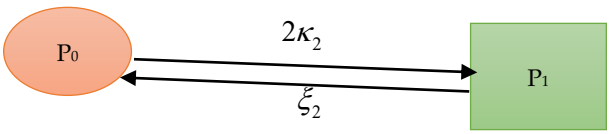


Figure 3: *Transition diagram of sand filter*

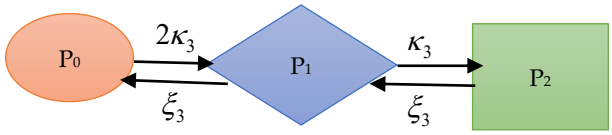


Figure 4: *Transition diagram of activated carbon filter*

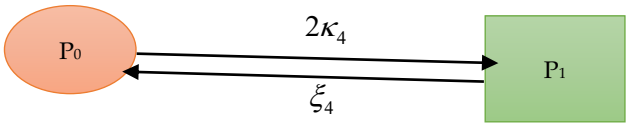


Figure 5: *Transition diagram of precision filter*

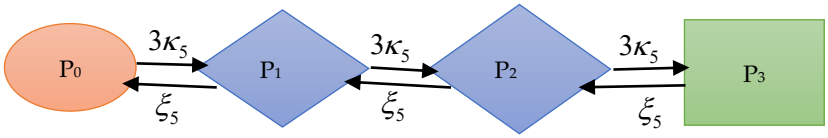


Figure 6: *Transition diagram of RO membrane*

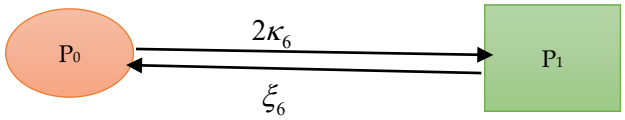


Figure 7: *Transition diagram of water producing tank*

Table1: *Failure and repair rates of component of the RO system*

Subsystem	Failure Rate (K)	Repair Rate (ξ)
S_1	K_1= 0.005	$\xi_1 = 0.07$
S_2	K_2= 0.010	$\xi_2 = 0.09$
S_3	K_3= 0.015	$\xi_3 = 0.11$
S_4	K_4= 0.020	$\xi_4 = 0.13$
S_5	K_5= 0.025	ξ_5=0.15
S_6	K_6= 0.030	ξ_6=0.17

III. Materials and methods

All of the measures discussed in this study are only valid in the steady-state era, when all failure and repair rates are exponentially distributed.

Reliability

The chance that a device will run without failure for a particular period of time is referred to as reliability under the operational conditions indicated.

$$R(t) = \int_t^{\infty} f(x)dx \tag{61}$$

MTBF

Mean Time between Failures (MTBF): The average period of good system functioning is referred to as the mean time between failures. When the failure rate is reasonably consistent over the operating period, the MTBF is the reciprocal of the constant failure rate or the ratio of the test time to the number of failures [26].

$$\text{MTBF} = \int_0^{\infty} R(t)dt = \int_0^{\infty} e^{-\theta t} = \frac{1}{\theta} \tag{62}$$

MTTR

Mean Time between repairs (MTTR): is the reciprocal of the system repair rate.

$$\text{MTTR} = \frac{1}{\xi} \tag{63}$$

Availability

Availability: Availability is a performance criterion for repairable systems that takes into consideration both the system's dependability and maintainability. It is defined as the likelihood that the system will function properly when it is needed [28].

$$\text{Availability} = \frac{Life\ time}{Total\ time} = \frac{Life\ time}{Life\ time + Repair\ time} = \frac{MTTF}{MTTF+MTTR} \tag{64}$$

Maintainability

Maintainability: [28] is a design, installation, and operation feature that is generally stated as the likelihood that a machine can be kept in, or returned to, a given operational condition within a specified time interval when maintenance is necessary.

$$M(t) = 1 - e^{(-\frac{-t}{MTTR})} \tag{65}$$

Dependability [27] dependability was stated as a design element It assesses performance by utilizing average failure and repair rates, as well as dependability and availability. The benefit of dependability is that it allows for the comparison of cost, reliability, and maintainability. The dependability ratio for random variables with exponential distribution is as follows:

$\boldsymbol{\xi} = Repair\ rate$, K = Failure rate

$d = \frac{\beta}{\lambda} = \frac{MTBF}{MTTR}$

The high value of dependability ratio represents the necessity of maintenance. C. Li, S. Besarati, and colleagues [10] mentioned that the dependability value increases if availability is above 0.9 and decrease if availability is less than 0.1. the minimum value of dependability is given by:

$$D_{min} = 1 - (\frac{1}{d-1})(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ ln\ d}{d-1}}) \tag{66}$$

System Reliability

$$R_{sys}(t) = R_{s1}(t) \times R_{s2}(t) \times R_{s3}(t) \times R_{s4}(t) \times R_{s5}(t) \times R_{s6}(t)$$
$$= e^{-(K_1+K_2+K_3+K_4+K_5+K_6)t} \tag{67}$$

System Availability

Arranged in series, failure of one cause the complete failure of the system.

$$A_{sys} = A_{s1} \times A_{s2} \times A_{s3} \times A_{s4} \times A_{s5} \times A_{s6} \tag{68}$$

$$A_{sys} = (\frac{\xi_1}{\xi_1+2K_1^2+2K_1}) \times (\frac{\xi_2}{\xi_2+2K_2}) \times (\frac{\xi_3^2}{\xi_3^2+2K_3^2+2\,\xi_3 K_3}) \times (\frac{\xi_4}{\xi_4+2K_4}) \times (\frac{\xi_5^2}{\xi_5^2+3K_5\xi_5+6\,\xi_5 K_5^2+6K_5^2}) \times (\frac{\xi_6}{\xi_6+2K_6})$$

$$A_{sys} = \left(\frac{0.07}{0.08005}\right) \times \left(\frac{0.09}{0.11}\right) \times \left(\frac{0.0121}{0.01585}\right) \times \left(\frac{0.13}{0.17}\right) \times$$

$$(\frac{0.0225}{0.0380625}) \times \left(\frac{0.17}{0.23}\right)$$

$$A_{sys} = 0.87445 \times 0.81818 \times 0.76341 \times 0.76471 \times 0.59113 \times 0.73913$$

$$A_{sys} = 0.18249$$

System Maintainability

$$M_{sys}(t) = M_{s1}(t) \times M_{s2}(t) M_{s3}(t) \times M_{s4}(t)) \times M_{s5}(t)\ ) \times M_{s6}(t) \tag{69}$$

$$= (1 - e^{-\xi_1(t)}) \times (1 - e^{-\xi_2(t)}) \times (1 - e^{-\xi_3(t)}) \times (1 - e^{-\xi_4(t)})\ ) \times (1 - e^{-\xi_5(t)})) \times (1 - e^{-\xi_6(t)})$$

$$1 - e^{-(\xi_1+\xi_2+\xi_3+\xi_4+\xi_5+\xi_6)t} \tag{70}$$

$$= (1 - e^{-0.07(t)}) \times (1 - e^{-0.09(t)}) \times (1 - e^{-0.11(t)}) \times (1 - e^{-0.13(t)})\ ) \times (1 - e^{-0.15(t)})) \times (1 - e^{-0.17(t)})$$

$$= 1 - e^{-0.72(t)}$$

System dependability

$D_{\min\,(sys)} = D_{\min\,(s1)} \times D_{\min\,(s2)} \times D_{\min\,(s3)} \times D_{\min\,(s4)} \times D_{\min\,(s5)} \times D_{\min\,(s6)}$

$D_{min} = 1 - (\frac{1}{d-1})(e^{-\frac{lnd}{d-1}} - e^{-\frac{d\ln d}{d-1}})$

$d = \frac{\xi}{\mathrm{K}} = \frac{MTBF}{MTTR}$

$d_1 = \frac{\xi_1}{\mathrm{K}_1} = \frac{0.07}{0.005} = 14.00$

$d_2 = \frac{\xi_2}{\mathrm{K}_2} = \frac{0.09}{0.01} = 9.00$

$d_3 = \frac{\xi_3}{\mathrm{K}_3} = \frac{0.11}{0.015} = 7.33$

$d_4 = \frac{\xi_4}{\mathrm{K}_4} = \frac{0.13}{0.02} = 6.50$

$d_5 = \frac{\xi_5}{\mathrm{K}_5} = \frac{0.15}{0.025} = 6.00$

$d_6 = \frac{\xi_6}{\mathrm{K}_6} = \frac{0.17}{0.03} = 5.67$

$D_{\min\,(s1)} = 1 - (\frac{1}{14-1})(e^{-\frac{2.6391}{13}} - e^{\frac{36.9468}{13}})$

$D_{\min\,(s1)} = 1 - 0.05831$

$D_{\min\,(s1)} = 0.94169$

$D_{\min\,(s2)} = 1 - (\frac{1}{9-1})(e^{-0.27465} - e^{-2.47188})$

$D_{\min\,(s2)} = 0.91557$

$D_{\min\,(s3)} = 1 - (\frac{1}{6.33})(e^{-0.31469} - e^{-2.30666})$

$D_{\min\,(s3)} = 0.90041$

$D_{\min\,(s4)} = 1 - (\frac{1}{5.5})(e^{-0.34033} - e^{-2.21213})$

$D_{\min\,(s4)} = 0.89053$

$D_{\min(s5)} = 1 - \left(\frac{1}{5}\right)(e^{-0.35835} - e^{-2.15011})$

$D_{\min\,(s5)} = 0.88353$

$D_{\min\,(s6)} = 1 - (\frac{1}{4.67})(e^{-0.37156} - e^{-2.10675})$

$D_{\min\,(s6)} = 0.87837$

$D_{\min(sys)} = 0.94169 \times 0.91557 \times 0.90041 \times 0.89053 \times 0.88353 \times 0.87837$

$D_{\min(sys)} = 0.53652$

Table 2: *Variation of reliability of subsystems with time*

Time (in days)	$R_{s1}(t)$	$R_{s2}(t)$	$R_{s3}(t)$	$R_{s4}(t)$	$R_{s5}(t)$	$R_{s6}(t)$	$R_{sys}(t)$
0	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
30	0.86071	0.74082	0.63763	0.54881	0.47237	0.40657	0.04285
60	0.74082	0.54881	0.40657	0.30119	0.22313	0.16530	0.00184
90	0.63763	0.40657	0.25924	0.16530	0.10540	0.06721	7.0×10^{-5}
120	0.54881	0.30119	0.16530	0.09072	0.04979	0.02732	3.37×10^{-6}
150	0.47237	0.22313	0.10540	0.04979	0.02352	0.01111	1.44×10^{-7}
180	0.40657	0.16530	0.06721	0.02732	0.01111	0.00452	6.19×10^{-9}
210	0.34994	0.12246	0.04285	0.01500	0.00525	0.00184	2.66×10^{-10}
240	0.25924	0.09072	0.02732	0.00823	0.00248	0.00075	9.83×10^{-12}
270	0.22313	0.06721	0.01742	0.00452	0.00117	0.00030	4.11×10^{-13}
300	0.19205	0.04979	0.01111	0.00248	0.00055	0.00012	1.74×10^{-14}
330	0.16530	0.03688	0.00708	0.00136	0.00026	0.00005	7.63×10^{-16}
360	0.14227	0.02732	0.00452	0.00075	0.00012	0.00002	3.16×10^{-17}

Table 3: *Variation of Maintainability of subsystems with time*

Time (in days)	$M_{s1}(t)$	$M_{s2}(t)$	$M_{s3}(t)$	$M_{s4}(t)$	$M_{s5}(t)$	$M_{s6}(t)$	$M_{sys}(t)$
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
30	0.87754	0.93279	0.96312	0.97976	0.98889	0.99390	0.75917
60	0.98500	0.99548	0.99864	0.99959	0.99988	0.99996	0.97866
90	0.99816	0.99970	0.99995	0.99999	0.99999	1.00000	0.99779
120	0.99978	0.99998	0.99999	1.00000	1.00000	1.00000	0.99975
150	0.99997	0.99999	1.00000	1.00000	1.00000	1.00000	0.99996
180	0.99999	1.00000	1.00000	1.00000	1.00000	1.00000	0.99999
210	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
240	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
270	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
300	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
330	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
360	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Table 4: *RAMD indices for the R.O system*

RAMD indices of Subsystems	Subsystem S_1	Subsystem S_2	Subsystem S_3	Subsystem S_4	Subsystem S_5	Subsystem S_6	System
Reliability	$e^{-0.005t}$	$e^{-0.01t}$	$e^{-0.015t}$	$e^{-0.020t}$	$e^{-0.025t}$	$e^{-0.030t}$	$e^{-0.105t}$
Maintainability	$1-e^{-0.07t}$	$1-e^{-0.09t}$	$1-e^{-0.11t}$	$1-e^{-0.13t}$	$1-e^{-0.15t}$	$1-e^{-0.10t}$	$1-e^{-0.72t}$
Availability	0.87445	0.81818	0.76341	0.76471	0.59113	0.73913	0.18249
MTBF	200.000	100.000	66.667	50.000	40.000	33.333	490.0003
MTTR	14.286	11.111	9.091	7.6923	6.6667	5.8824	54.7294
Dependability	0.94169	0.91557	0.90041	0.89053	0.88353	0.87837	0.53652
Dependability ratio	14.000	9.00000	7.33330	6.50000	6.00000	5.66667	

Table 5: *Variation of reliability of system due to variation in failure rate of raw water tank (subsystem 1)*

	R.O System		raw water tank (Subsystem 1)	
Time (in days)	$K_1 = 0.0001$	$K_1 = 0.0002$	$K_1 = 0.0001$	$K_1 = 0.0002$
0	1.00000	1.00000	1.00000	1.00000
10	0.36751	0.36714	0.99900	0.99800
20	0.13506	0.13480	0.99800	0.99601
30	0.04964	0.04949	0.99700	0.99402
40	0.01824	0.01817	0.99601	0.99203
50	0.00670	0.00667	0.99501	0.98020
60	0.00246	0.00245	0.99402	0.98807
70	0.00091	0.00090	0.99302	0.98610
80	0.00033	0.00029	0.99203	0.98413
90	0.00012	0.00012	0.99104	0.98216
100	0.00004	0.00004	0.99005	0.98020

Table 6: *Variation of reliability of system due to variation in failure rate of sand filter (subsystem 2)*

	R.O System		sand filter (subsystem 2)	
Time (in days)	$K_2 = 0.0003$	$K_2 = 0.0004$	$K_2 = 0.0003$	$K_2 = 0.0004$
0	1.00000	1.00000	1.00000	1.00000
10	0.38558	0.38520	0.99700	0.99601
20	0.14867	0.14838	0.99402	0.99203
30	0.05733	0.05715	0.99104	0.98807
40	0.02210	0.02202	0.98807	0.98413
50	0.00852	0.00848	0.98511	0.98020
60	0.00329	0.00327	0.98216	0.97629
70	0.00127	0.00126	0.97922	0.97239
80	0.00049	0.00048	0.97629	0.96851
90	0.00019	0.00019	0.97336	0.96464
100	0.00007	0.00002	0.97045	0.96079

Table 7: *Variation of reliability of system due to variation in failure rate of activated carbon filter (subsystem 3)*

	R.O System		Act. carbon filter (Subsystem 3)	
Time (in days)	$K_3 = 0.0005$	$K_3 = 0.0006$	$K_3 = 0.0005$	$K_3 = 0.0006$
0	1.00000	1.00000	1.00000	1.00000
10	0.40454	0.40414	0.99501	0.99402
20	0.16365	0.16333	0.99005	0.98807
30	0.06620	0.06601	0.98511	0.98216
40	0.02678	0.02668	0.98020	0.97629
50	0.01083	0.01078	0.97531	0.97045
60	0.00438	0.00436	0.97045	0.96464
70	0.00177	0.00176	0.96561	0.95887
80	0.00072	0.00071	0.96079	0.95313
90	0.00029	0.00029	0.95600	0.94743
100	0.00012	0.00012	0.95123	0.94176

Table 8: *Variation of reliability of system due to variation in failure rate of precision filter (subsystem 4)*

	R.O System		precision filter (Subsystem 4)	
Time (in days)	$K_4 = 0.0007$	$K_4 = 0.0008$	$K_4 = 0.0007$	$K_4 = 0.0008$
0	1.00000	1.00000	1.00000	1.00000
10	0.42443	0.42401	0.99302	0.99203
20	0.18014	0.17978	0.98610	0.98413
30	0.07646	0.07623	0.97922	0.97629
40	0.03245	0.03232	0.97239	0.96851
50	0.01377	0.01370	0.96561	0.96079
60	0.00585	0.00581	0.95887	0.95313
70	0.00248	0.00246	0.95218	0.94554
80	0.00105	0.00104	0.94554	0.93800
90	0.00045	0.00044	0.93894	0.93053
100	0.00019	0.00019	0.93239	0.92312

Table 9: *Variation of reliability of system due to variation in failure rate of R.O membrane (subsystem 5)*

	R.O System		R.O membrane (Subsystem 5)	
Time (in days)	$K_5 = 0.0009$	$K_5 = 0.0010$	$K_5 = 0.0009$	$K_5 = 0.0010$
0	1.00000	1.00000	1.00000	1.00000
10	0.44530	0.44486	0.99104	0.99005
20	0.19829	0.19790	0.98216	0.98020
30	0.08830	0.08804	0.97336	0.97045
40	0.03932	0.03916	0.96464	0.96079
50	0.01751	0.01742	0.95600	0.95123
60	0.00780	0.00775	0.94743	0.94176
70	0.00347	0.00345	0.93894	0.93239
80	0.00155	0.00153	0.93053	0.92312
90	0.00069	0.00068	0.92219	0.91393
100	0.00031	0.00030	0.91393	0.90484

Table 10: *Variation of reliability of system due to variation in failure rate of water producing tank (subsystem 6)*

	R.O System		producing tank (Subsystem 6)	
Time (in days)	$K_6 = 0.0011$	$K_6 = 0.0012$	$K_6 = 0.0011$	$K_6 = 0.0012$
0	1.00000	1.00000	1.00000	1.00000
10	0.46720	0.46673	0.98906	0.98807
20	0.21827	0.21784	0.97824	0.97629
30	0.10198	0.10167	0.96754	0.96464
40	0.04764	0.04745	0.95695	0.95313
50	0.02226	0.02215	0.94649	0.94176
60	0.01040	0.01034	0.93613	0.93053
70	0.00486	0.00482	0.92589	0.91943
80	0.00227	0.00225	0.91576	0.90846
90	0.00106	0.00105	0.90574	0.89763
100	0.00050	0.00049	0.89583	0.88692

IV. Discussion

The reliability and maintenance characteristics of all subsystems are shown in Tables 2 and 3. All of the extra RAMD metrics are listed in Table 4. According to the numerical analysis in table 2, the system's reliability after 60 days of operation is just 0.00184. It happened because of the least reliable subsystems, the R.O membrane and water producing tank (subsystem 5 and 6), whose corresponding reliability are 0.22313 and 0.16530 respectively. In this case, it is recommended that weak performance be given more attention and that suitable maintenance methods be established to enhance their reliability. The maintainability of the system after 60 days is just 0.97866. While the correspondent value for the RO membrane i.e. subsystem 5 is 0.99988. Attention is highly needed to the subsystem by providing more redundant and possible replacement of the affected subsystem i.e. subsystems 5 and 6 although the maintainability sound good. Tables 5, 6, 7, 8, 9, and 10 illustrated the time-dependent reliability behavior of several subsystems as well as the variability in their failure rates. Precision filters and R.O Membrane systems are the most critical, highly sensitive components that necessitate special attention in order to improve system reliability. According to the preceding discussion, regular maintenance plans that properly monitor the failure rates of the Precision filter and R.O Membrane system will surely enhance the efficacy and working time of the reverse osmosis system of water treatment.

V. Conclusion

Through desalination, RO is a significant technique for generating drinkable water from saltwater. The failure behavior of a desalination system's components determines its performance. Because the RO system was designed to be a power-saving system, the dependability of its subsystems must be maintained at a high level by correct design and material selection of these subsystems for continuous plant operation Thus, in this study, the system's availability and reliability were examined, as well as other characteristics such as MTTF, MTTR, dependability analysis, and maintainability.

The design of an integrated RO system is recommended because it is high performing, consumes little power, and is cost effective.

References

[1] Muhammad F. I. (2020) "Performance Analysis and Treatment Technologies of Reverse Osmosis Plant" Case Studies in Chemical and Environmental Engineering,2, https://doi.org/10.1016/j.cscee.2020.100007

[2] Brown,S.L., K.M. Leonard, K.M and Messimer,S.L. (2008) "Evaluation of ozone pretreatment on flux parameters of reverse osmosis for surface water treatment" Ozone: Science & Engineering, 30:2, 152-164, DOI: 10.1080/01919510701864031

[3] Badruzzaman,M, Voutchkov,N, Weinrich, L and Jacangelo, J.G.(2019) "Selection of pretreatment technologies for seawater reverse osmosis plants": a review, Desalination 449 (2019) 78–91, https://doi.org/10.1016/j.desal.2018.10.006.

[4] Saffarimiandoab,F., Gul, B, Y., Erkoc-Ilter, S., Guclu, S., Unal, S., Tunaboylu, B., Menceloglu, Y.Z and Koyuncu, I. (2019) "Evaluation of biofouling behavior of zwitterionic silane coated reverse osmosis membranes fouled by marine bacteria". Prog. Org. Coating Vol. 134 303–311, https://doi.org/10.1016/ j.porgcoat.2019.05.027.

[5] Evita A. and Evan D. (2012) "A strategic plan for reuse of treated municipal wastewater for crop irrigation on the Island of Crete" Agricultural Water Management**,** Volume 105, Pages 57-64. https://doi.org/10.1016/j.agwat.2012.01.002

[6] Oliveira, S.C and Sperling, M.V. (2007) "Reliability analysis of wastewater treatment plants" journal of Comparative study, Water Research, Volume 42, Issues 4–5,1182-1194 https://doi.org/10.1016/j.watres.2007.09.001

[7] Srivathsan,G., Sparrow, E.M and Gorman, J.M. (2014). "Reverse osmosis issues relating to pressure drop, mass transfer, turbulence, and unsteadiness" Desalination, 341,83-86, https://doi.org/10.1016/j.desal.2014.02.021

[8] Edzwald, J.K and Haarhoff, J.(2011) "Seawater pretreatment for reverse osmosis: Chemistry, contaminants, and coagulation" Water Research Volume 45, Issue 17, Pages 5428-5440. https://doi.org/10.1016/j.watres.2011.08.014

[9] Muhammad F. I. (2020) "Performance Analysis and Treatment Technologies of Reverse Osmosis Plant" Case Studies in Chemical and Environmental Engineering,2, https://doi.org/10.1016/j.cscee.2020.100007

[10] Li,C., Besarati,S., Goswami,Y., Stefanakos, E and Chen, H. (2013) "Reverse osmosis desalination driven by low temperature supercritical organic Rankine cycle". Appl. Energy, Vol. 102, 1071–1080. https://doi.org/10.1016/j.apenergy.2012.06.028

[11] Srivastava, S., Vaddadi, S.,Kumar,P and S. Sadistap (2018). "Design and development of reverse osmosis (RO) plant status monitoring system for early fault prediction and predictive maintenance" *Applied Water Science* **vol. 8**, no. 159pp 1-10

https://doi.org/10.1007/s13201-018-0821-8

[12] Sadri, S., Khoshkhoo, R.H and Ameri, M. (2016). "Multi objective optimization of reverse osmosis desalination plant with energy approach" Journal of Mechanical Science andTechnology **vol. 30**, pp 4807–4814. https://doi.org/10.1007/s12206-016-0953-4

[13] Li, Y and Tian, K (2009) "Application of vacuum membrane distillation in water treatment," Journal of Sustainable Development, vol. 2, no. 3, pp. 183–186. **DOI:**10.5539/jsd.v2n3p183

[14] Lee, Y.G., Lee, Y.S., Jeon, J.J., Lee, S.,Yang, D.R., Kim, I.S and HaKim, J.(2009). "Artificial neural network model for optimizing operation of a seawater reverse osmosis desalination plant" Vol. 247, Issues 1–3, Pp 180-189. https://doi.org/10.1016/j.desal.2008.12.023

[15] Wirth,D and C. Cabassu, C (2002). "Water desalination using membrane distillation: comparison between inside/out and outside/in permeation" ScienceDirect Volume 147, Issues 1–3, 10, Pages 139-145.

[16] Ezugbe, E.O and Rathilal, S.(2020). "Membrane Technologies in Wastewater Treatment" Membranes, 10(5), 89; 1-28. ; https://doi.org/10.3390/membranes10050089

[17] Tundis,R., Conidi,C., Loizzo, M. R., Sicari, V., Romeo, R. and Cassano, A. (2021). "Concentration of Bioactive Phenolic Compounds in Olive Mill Wastewater by Direct Contact Membrane Distillation" Molecules, vol. 26, 1808. https://doi.org/10.3390/molecules26061808. 1-13.

[18] Sponza, D, T. (2021). "Treatment of Olive Mill Effluent with Sequential Direct Contact Membrane Distillation (DCMD)/Reverse Osmosis (RO) Hybrid Process and Recoveries of Some Economical Merits" Res Appl. 2021 vol. 1(1): 001-010. Advanced Journal of Physics Research and Applications. Vol. 1(1): 001-010.

[19] Biniaz, P., Ardekani, N. T., Makarem, M. A. and Rahimpour, M. R. (2019). "Water and Wastewater Treatment Systems by Novel Integrated Membrane Distillation (MD), Chemengineering, vol. 3, *3*(1), 8; https://doi.org/10.3390/chemengineering3010008

[20] Garud R. M., Kore S. V., Kore V. S., and Kulkarni G. S. (2011) "A Short Review on Process and Applications of Reverse Osmosis" www.environmentaljournal.org Volume 1, Issue 3: 233-238 .

[21] Gupta, N., Kumar, A and Saini, M. (2021) "Reliability and Maintainability Investigation of Generator in Steam Turbine Power Plant using RAMD analysis" Journal of Physics doi:10.1088/1742-6596/1714/1/012009. No. **1714** 012009

[22] Hajeeh, M and Chaudhuri, D.(2000) "Reliability and availability assessment of reverse osmosis", Desalination, Vol. 130 pp 185–192. https://doi.org/10.1016/S0011-9164(00)00086-2

[23] George, Y. K., Solt C Eng, F I Chem, E. (2002) in Plant Engineer's Reference Book (Second Edition).

[24] Siong, J. Idris a , M. Mazar Atabaki , (2013) "Performance of Activated Carbon in Water Filters" See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/234060484 Article in Water Resources.

[25] Taylor, P.A.(2015)"Physical, chemical, and biological treatment of groundwater at contaminated nuclear and NORM sites", Woodhead Publishing Series in Energy 2015,237-256, https://doi.org/10.1016/B978-1-78242-231-0.00010-7

[26] Monika S. and Ashish K. (2019). "Performance analysis of evaporation system in sugar industry using RAMD analysis" journal of Brazilian society of mechanical science and engineering Vol. 41. (175). https://doi.org/10.1007/s40430-019-1681-3,

[27] Ebeling, A. (2000) An introduction to reliability and maintainability engineering. Tata Mcgraw Hill Company Ltd, New Delhi

[28] Wohl,J.G. (1996) System operational readiness and equipment dependability. IEEE Trans Reliability 15 (1): 1-6. doi: 10.1109/TR.1966.5217582.

Review of Performance Factors of Emotional Speaker Recognition System: Features, Feature Extraction Approaches and Databases

[1]Satish Kumar Das, [2]Uttpal Bhattacharjee, [3]Amit Kumar Mandal

•

Department of Computer Science & Engineering
Rajiv Gandhi University, Arunachal Pradesh, India
[1]satish.das@rgu.ac.in
[2]utpal.bhattacharjee@rgu.ac.in
[3]amitkumar.tu@gmail.com

Abstract

Emotion is a conscious mental reaction accompanied by physiological and behavior changes in human body.In speaker authentication system, emotional state of the speaker plays a vital role. Recently, the field of speaker recognition in emotional context attracts more and more attention of many research focuses. However, to implement more realistic and intelligent emotional speaker recognition system it is interesting to study this system under real life conditions. Speech emotion recognition is a system in which speech signals are processed to classify the embedded emotions. In recent past, speaker emotion recognition has gained a lot of attention from different researchers as it has many applications. In this regards, study of prior works is useful for further research in the field of speaker verification in emotional context. So, performance and reliability of Emotional Speaker Recognition System depend on the proper selection of features to characterize different emotional states, feature extraction approaches and databases. In this paper we briefly discuss about different features, feature extraction approaches and emotion recognition and speaker verification databases.

Keywords: speech emotion recognition, features, feature extraction approaches, databases

I. Introduction:

The human voice is used to expresses the emotions. The speech has potential to be an essential mode of interaction with the computer. Recognition of emotion from human speech is an active research area in signal processing. This is due to the enormous possibilities in the field of human-machine interaction [1].Speech emotion recognition is using in different fields from medical scienceto business, from entertainment to interactive systems.

A speech signal is a logical arrangement of sounds from which our brain acquires the information and knowledge. But how a machine interprets the human speech signals and gains information and knowledge from is in the heart of speech recognition system. The basic goal of speech recognition is to offer an interaction between a person and a machine.

In this paper we are trying to present comprehensive literature review of various features, feature extraction techniques and datasets for emotion recognition and speaker verification. In our previous paper [2], we tried to present comprehensive literature review of various classifiers for speech emotion recognition. The paper is not to compare and evaluate the works by different

authors. Moreover, we are not surveying an exhaustive list of all research works. Also we are not providing justification for choosing particular features, extraction methods and datasets for specific tasks as they can be combined differently which may yield different results.

II. Features of Speech Emotion Recognition:

To efficiently characterize different emotions, a SER system needs to extract suitable features. Performance of classification is affected by the proper extraction and selection of features. There are different speech features for SER, but we can say which one is best. Speech features can be classified as continuous, spectral, qualitative and Teager energy operator based features. Continuous features are pitch, energy, formants; qualitative features are voice qualities such as hoarse, tense, breathy; spectral features are LPC, MFCC, LFPC and TEO-based features are TEO-FM-Var, TEO-Auto-Env, TEO-CB-Auto-Env.

I. Continuous features:

Continuous features can be prosody features and acoustics features. According to many researchers [3, 4, 5] prosody features such as pitch and energy represents the larger portion of the emotional content of utterance. Several studies [6, 7, 8] show that the arousal state of the speaker can affect the overall energy and energy distribution across frequency spectrum. Continuous speech features such as vocal cues [9] have been used by many researchers in SER. Different researchers [5, 10, 11] grouped the acoustic features into pitch-related features, energy-related features, timing features, formants features articulation features.

Some basic global features are fundamental frequency, energy, duration, formants, etc. Different researchers [5, 9, 10, 11, 12, 13, 14, 15, 16] have studied the relationships between these speech features and the basic emotions and shown that prosodic features are good indicators of emotions. But for some researchers [17, 18], prosodic features do not have much impact on emotions.

II. Qualitative features:

Voice quality is strongly related to emotional content of utterance [5, 19, 20]. Many researchers studied the auditory aspects of emotions and tried to define a relation [5, 10, 12]. Voice quality has the subjective impression from the contribution of different phonetic variables [10]. Cowie et al. [5] grouped the acoustic correlates which are related to the voice quality into the voice level, voice pitch, phrase, phoneme, word and feature boundaries and temporal structures.

But role of voice quality of subject in determining emotions has some confusion and problem. First, the term tense, harsh, and breathy can have different interpretations for different researchers based on their understanding [20]. According to Scherer [22], anger, joy, and fear are associated with tense voice while sadness is associated with lax voice. But for Murray [10], both anger and happiness are associated with breathy voice while resonant voice is associated with sadness. Second, it is the difficulty to automatically decide those voice qualities directly from the speech signal.

III. Spectral features:

Spectral features have significant role in SER. Spectral signals convey the speech signal's frequency content and provide complementary information to be used in prosodic features [23]. They are selected as short term representation for speech signal with longer temporal information [24]. Emotional content of a speech signal can affect on the distribution of the spectral energy

across the spectrum of speech [25]. Utterances with happiness imply the high energy at high level of frequency; on the other hand utterances with the sadness imply low energy at the same level of frequency [9, 26].

However, there are limitations of short-term spectral features for speech recognition [27]. Conventional spectral features such as MFCCs convey only the short-term spectral properties of the speech signal, not the temporal behavior information. To overcome the limitation, Wu et al. [23] proposed long-term modulation spectral features (MSFs) using both acoustic and temporal modulation frequency components.

IV. TEO-based features:

Teager Energy Operator is a non-liner operator, which represents the frequency and the changes of the signal amplitude which instantaneous [28]. It was initially proposed for nonlinear speech signals modeling, but later on used widely in the audio signal processing.TEO was introduced by Teager [29] who believed that speech is a function of non-linear flow of air in vocal system [30]. The fundamental frequency changes under stressful conditions. TEO of multi-frequency signals represents the individual frequency components and relationships between these components [31]. So, to detect stress in speech signal, features based on TEO can be used. Caims et al. [32] used the Teager energy of the pitch contour to classify clear, angry, loud, neutral and Lombard effects of speech. Zhou [31] proposed some other TEO-based features, such TEO-FM-Var, TEO-Auto-Env and TEO-CB-Auto-Env for detecting stress speech versus neutral speech. These features were also used to classify the stressed speech into angry, loud, and Lombard.

Proper feature selection in SER system mainly depends on the considered classification task. From review of this section, we may conclude that, continuous features like the fundamental frequency and the pitch are more preferable for high-arousal versus low-arousal emotions classification. While TEO-based features should be used for detecting stress in speech and the spectral features are the more promising for representation of speech. Features are somehow related to each other and so these features can be combined to get better classification results.

III. Feature Extraction

Feature extraction directly implies the accuracy and the performance of the system. Feature extraction is performed to represent a speech signal by a set of predetermined components of the speech signal [33]. Infeature extraction, the speech signal is transformed to a form which is more logical but concise and reliable than the original signal. It is a process of changing the speech waveform to a parametric form for subsequent processing and analysis where data rate is relatively lesser. Some commonly used approaches are:

- Mel Frequency Cepstral Coefficients (MFCC)
- Linear Prediction Coefficients (LPC)
- Linear Prediction Cepstral Coefficients (LPCC)
- Line Spectral Frequencies (LSF)
- Discrete Wavelet Transform (DWT)

I. Mel Frequency Cepstral Coefficients (MFCC):

MFCC is a static feature extraction approach. MFCC is replication of the human's ear system. MFCC is considered as frequency domain feature and designed to represents signal as the short-term power spectrum. It is based on discrete cosine transform of logarithm power spectrum on nonlinear Mel frequency scale [34]. The formula to calculate the mels of frequency is [35,36]:

$$Mel(f) = 2595 * \log_{10}(1 + F/100) \quad (1)$$

Where f is the frequency of signal in Hz. The MFCCis calculated using the following equation [35, 37]:

$$\hat{C}_n = \sum_{n=1}^{k} (\log \hat{S}_k) \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{k}\right] \quad (2)$$

where $\hat{C}_n$ is the final MFCC coefficients, $\hat{C}_n$is the output of filterbank and k is the number of mel cepstrum coefficients.

Continuous Speech →Pre-emphasis → Frame Blocking → Windowing → Fast Fourier Transform → Mel-Scale Filter Bank → Log → Discrete Cosine Transform → MFCC

Figure 1: Steps in MFCC approach

MFCC can effectively be used the low frequency region compared to the high frequency region. So, it can be used to compute low frequency formants and for the vocal tract resonances description [33]. MFCC features are used to distinguish between information of speech and non-speech signals [38]. Features with lower order MFCC can contains phonetic information and with higher order contains non-speech information. In case of stable and consistent source characteristics, MFCC is perfect representation for sound [36]. Information with frequencies at a maximum of 5 kHz, which covers most energy of human generated sounds [39] can be captured by it. But in the presence of background noise [40, 41], MFCC features do not work accurately and might not be well suited for generalization [36].

II. Linear Prediction Coefficient (LPC):

LPC is a static feature extraction method that imitates the human vocal tract [42]. LPC is used to represent the spectral part of the speech in compressed manner, using linear predictive model. LPC is applied to get the filter coefficients which are equivalent to the vocal tract by minimizing the mean square error [43]. The linear predictive model is given [44, 45] as:

$$\hat{S}(n) = \sum_{k=1}^{p} a_k s(n-k) \quad (3)$$

Where $\hat{s}$is the predicted sample, p is the predictor coefficients, s is the speech sample. Then the prediction error is given as [16, 25]:

$$e(n) = s(n) - \hat{s}(n) \quad (4)$$

LPC is derived by the following equation:

$$a_m = \log\left[\frac{1-k_m}{1+k_m}\right] \quad (5)$$

Here a_m is the linear prediction coefficient, and k_m is the reflection coefficient.

Speech signal →Frame blocking → Windowing → Auto Correlation Analysis → LPC Analysis → LPC

Figure 2: Steps in LPC approach

LPC is used for speech reconstruction, coding and synthesis and known for its computation speed and accuracy. LPC helps to encode high quality speech signal at low bit rate [46, 47, 48]. Linear predictive analysis has the ability to efficiently select the vocal tract information from speech [42]. The steady and consistent behaviors are excellently represented by LPC [36]. LPC is very accurate in estimating speech parameters and highly sensitive to quantization noise [49].

III. Linear Prediction Cepstral Coefficients (LPCC):

LPCCs are the coefficients of the Fourier transformation of the log magnitude spectrum [50] of LPC. LPCCs are LPC coefficients presented in cepstral domain and used to represent spectral envelope [49].

LPCC is calculated using the following equation:

$$C_m = a_m + \sum_{k=1}^{m-1} \left[\frac{k}{m}\right] c_k \, a_{m-k} \quad (6)$$

where C_m is the cepstral coefficient and a_m is the linear prediction coefficient.

Speech signal →Frame blocking → Windowing → Auto Correlation Analusis → LPC Analysis → LPC Parameter Conversion → LPCC

Figure 3: Steps in LPCC approach

LPCC has the ability to perfectly represent the speech waveforms and the speech characteristics with limited features [51].So, it is commonly applied in speech processing. LPCC have lower susceptibility to noise [50] and gives lower error rate [51].The higher order cepstral coefficients are sensitive to noise [52] whereas the lower order cepstral coefficients are sensitive to the spectral slope. These problems can be handled using weighted coefficient [53].

IV. Line Spectral Frequencies (LSF):

LSFs are the alternative parameters which are used to represent all-pole spectrum of speech signal [54]. LFSs are used to represent LPCs for transmission over a channel [55]. We can express the linear predictive (LP) polynomial as the mean of palindromic and antipalindromic polynomials [56]. The LP polynomial is given by:

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \quad (7)$$

This polynomial can be expressed in terms of palindromic and antipalindromic polynomials as given below:

$$A(z) = 0.5[P(z) + Q(z)] \quad (8)$$

Where,

$$P(z) = A(z) + z^{-(p+1)}A(z^{-1}) \quad (9)$$

$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1}) \quad (10)$$

Speech signal →Frame blocking → Windowing → Auto Correlation Analusis → LPC Analysis → Decomposition → Root Finder → LPF

Figure 4: Steps in LFS approach

The LSF representation of linear predictive polynomial consists of the location of the roots of the palindromic and antipalindromic polynomials. LSFs are used in coding, recognition and synthesization of speech [57, 58]. LSFs have low sensitivity to quantization noise and can be interpolated [59]. LSFs have relatively uniform spectral sensitivity [60]. Since LSFs have excellent quantization properties, they are important in transmission of vocal tract information from speech coder to decoder system. A near-minimal data set is provided for subsequent classification by the LSP representation in many cases [59].

V. Discrete Wavelet Transform (DWT):

DWT is a time scale representation of the signal. It is a special case of wavelet transform where the wavelets are sampled discretely. DWT composed of a high pass wavelet filter and a low pass scaling filter [61, 62].

$$\phi(t) = \sum_{n=0}^{N-1} h[n]\sqrt{2}\phi(2t - n) \quad (11)$$

$$\psi(t) = \sum_{n=0}^{N-1} g[n]\sqrt{2}\phi(2t - n) \quad (12)$$

Where$\phi(t)$is the scaling function,$\psi(t)$is the wavelet function and h[n] is low-pass filter with impulse response and g[n] is high-pass filter with impulse response.

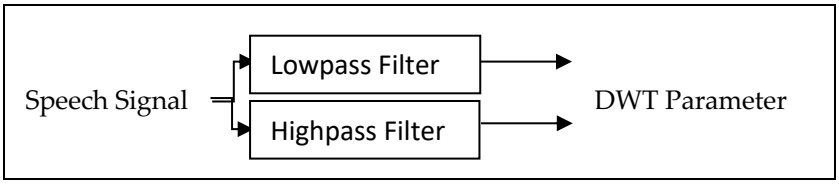


Figure 5: Steps in DWT approach

As it captures both frequency content and temporal content, it outperformed Fourier transformation. The DWT contains different scales of frequency information, thus DWT can help to get the speech information of respective frequency band [63]. Adequate numbers of frequency bands are provided by DWT for effective analysis of speech [64].

In this paper we have just included some commonly used feature extraction approaches. There are other feature extraction approaches for speech processing: Perceptual Linear Predictive (PLP) [65, 66], Relative Spectral Processing (RASTA) [67], Principle Component Analysis (PCA) [68, 69], Gammatone Frequency Cepstral Coefficient (GFCC) [70], Wavelet Packet Decomposition (WPD) [71, 72], Log-frequency power coefficient (LFPC) [25], etc.

IV. Speech Emotion Recognition and Speaker Verification Databases

Suitable speech database is necessary for characterizing emotions for recognition or for verification. Quality of the database is an important issue to evaluate the performance of the system [24]. The proper design of the database is very much important for the classification task. Low-quality databases may yield incorrect conclusions. This section is divided into two sections: speech emotion recognition databases and speaker verification databases. We are not going to discuss all the existing databases; rather we will briefly discuss some interesting to us databases.

I. Speech Emotion Recognition Databases:

DESD: Engberg et al. [73] described Danish Emotional Speech Database for evaluation of emotional state in emotional speech. It was a part of the VAESS project. It was recorder using microphone in Danish language with 4 speakers, out of which 2 were males and 2 were females. Ages of the speakers were between 34 and 52. They recorded the session on DAT tape @48kHz. It consists of 2 words, 9 sentences and 2 passages and five emotions: Neutral, surprise, anger, happiness and sadness.

SUSAS: Hansen et al. [74] developed Speech Under Simulated and Actual Stress (SUSAS) database for analysis and algorithm formulation in the field of speech recognition. The database is partitioned into five domains: a) styles of talking, b) speech with noise, c) speech associated with anxiety, depression, fear, d) actual subject motion-fear tasks and e) dual tracking computer response task. It has records of 32 speakers, ages between 22 and 76, out of which 19 are males and 13 are females and more than 16000 utterances. Samples are 1-channel 16-bit linear PCM with sample rate @8kHz.

KISMET: Breazeal et al. [75] used this database for explore the ability of recognition in robot directed speech. They used 2 female speakers and five communicative intents such as approval, attention, prohibition, soothing, and neutral. They used wireless microphone on Linux based system. There are 726 samples in .wav format with 16-bit single channel and 8 kHz signal.

INTERFACE: Hozjan el al. [76] recorded a database, as a part of the IST project Interface, to study emotional speech and to analysis the emotional characteristics for automatic emotion classification. They recorded for six emotions: anger, disgust, fear, joy, sadness and surprise, in English, French,

Slovenian and Spanish language. They used 2 male and 2 female speakers for English and 1 male and 1 female speaker for other languages. The number of sentences for English was 8928, Slovenian was 6080, French was 5600 and Spanish was 5520. They recorded the samples @16 kHz in linear format using condenser microphones.

ESMBS: New et al [25] used ESMBS database for text independent emotion classification of speech. It consists of 6 speakers (3 male and 3 females) with Burmese language and 6 speakers (3 male and 3 females) with Mandarin language. It has 10 utterances * 6 emotions * 12 speakers = 720 utterances generated by speakers. They recorded by using microphone in a quiet environment.

MPEG-4: Schuller et al [77] used MPEG-4 database for speaker independent speech emotion recognition. The emotion they were used are fear, anger, joy, sadness, disgust, surprise and neutrality. There were 2440 samples from 35 speakers. A condenser microphone was used to record the samples. There were 1,144 phrases and 1,507 utterances with textual contents extracted from 7 US movies.

Berlin Emotional Database: Burkhardt et al [78] described a database of emotional speech in German language. It consisted of 5 male speakers and 5 female speakers who produced 10 utterances, out of which 5 were short and 5 were long. They used seven emotions: neutral, anger, fear, joy, sadness, disgust and boredom and used about 800 sentences. They recorded the samples with frequency @48 kHz and later they downsampled them to 16 kHz.

CLDC: Zhou et al [79] used Chinese Linguistic Data Consortium (CLDC) for speech emotion recognition. It consists of 4 speakers: 2 male and 2 female speakers and six emotions: normal, happiness, surprise, anger, sadness and fear. There were 200 utterances for each emotion which yield 1200 utterances in total. Samples are recorded using 16-bit channel with sample rate of 16 kHz.

KES: Kim et al [80] used KES database designed by Professor C.Y. Lee for emotion recognition. The database contains context independent short, medium, and long sentences and four emotions: neutrality, joy, sadness and anger. The database has 10 speakers compromised of 5 male and 5 female speakers and 5400 sentences. They recorded the data @16kHz and in 32bits format over 30dB SN in a silent experimental environment.

FAU Aibo: Batliner et al [81] designed FAU Aibo Emotion Corpus to collect spontaneous and emotional speech of children. It consists of data from 51 children, compromised of 21 male children and 30 female children, aged ranges 10-13 years. They recorded the sample data with sampling rate @48 kHz in 16-bit format and downsampled them to 16 kHz using wireless headset and DAT-recorder. It has 48401 words and 9.2 hours of recordings.

IITKGP-SESC: Koolagudi et al. [82] introduced IITKGP-SESC, a Emotion Speech Corpus simulated by Indian Institute of Technology, Kharagpur for speech emotion analysis. The database was recorded using 10 speakers where there were 5 males and 5 females, aged ranges from 25 to 40 years, in Telegu language. They recorded for the eight emotions: Neutral, happy, anger, surprise, compassion, disgust, fear and sarcastic. They sampled the signals @16 kHz in 16-bit format using dynamic cardioid microphone. The database has 12000 utterances in total; 1500 utterances per emotion.

TURES: Oflazoglu et al [83] constructed Turkish Emotional Speech (TURES) Database for emotion recognition. They labeled each utterance with happy, surprised, sad, angry, fear, neutral and other emotional states and with 3- dimensional emotional space such as activation, valence and dominance. The database contains 582 speakers, compromised of 394 male speakers and 188 female speakers, and 5100 utterances extracted from movies. The audio channels were saved as mono 16-bit, PCM-wave format, sampled @48 kHz.

EMOVO: Costantini et al. [84] described EMOVO which is the first Italian emotional database. It has six emotional states; they are anger, disgust, fear, joy, neutral, surprise and sadness, and 6 speakers, 3 males and 3 females, age ranges from 23 to 30 years. They recorded the speech samples @48 kHz with 16-bit stereo in .wav format using two microphones and a digital reorder.

BAUM-1: Zhalehpour et al [85] discussed BAUM-1 which is a audio-visual database in Turkish language. It contains six emotional states: happiness, sadness, surprise, disgust, anger, fear, boredom and contempt, and mental states: unsure, thinking, concentrating and bothered. The data were collected from 31 speakers where it compromised of 18 males and 13 females. They recorded 1222 video clips with stereo and mono cameras.

RAVDESS: Livingstone et al. [86] developed an Audio-Visual Database of Emotional Speech and Song known as RAVDESS in North American English. The dataset contains eight emotional states: anger, calm, disgust, fearful, happy, neutral, sad and surprised. It has 24 speakers, compromised of 12 males and 12 females, and 7356 recordings. All speakers were recorded for 60 spoken utterances and 44 sung utterances, in total 104 utterances. They recorded the speech samples @48 kHz sampling rate with 16-bit in .wav format using condenser microphone.

Table 2:*Databases of speechemotion recognition*

Database	Year	Language	#Speakers (male/female)	#Utterance	Emotions
DESD	1996	Da	4(2/2)	10 minutes of speech	An, Ha, Ne, Sa, Su
SUSAS	1999	En	32(19/13)	16000	An, Lo, Lm, Ne
KISMET	2002	En	2(0/2),	1002	Ap, At, Ne, Pr, So
INTERFACE	2002	En, Fr, Sl, Sp	10(5/5)	26128	An, Di, Fe, Jo, Sa, Su
ESMBS	2003	Bu, Ma	12(6/ 6)	720	An, Di, Fe, Jo, Sa, Su
MPEG-4	2005	En	35	2440	An, Di, Fe, Jo, Ne, Sa, Su
Berlin Emotional Database	2005	Ge	10(5/ 5)	10	An, Bo, Di, Fe, Jo, Ne, Sa
CLDC	2006	Ch	4	1200	An, Fe, Jo, Ne, Sa, Su
KES	2007	Ko	10(5/5)	5400	An, Jo, Ne, Sa
FAU Aibo	2008	Ge	51(21/30)		An, Bo, Em, He, Jo, Mo, Ne, Re, Rs Su, To
IITKGP-SESC	2009	Te	10(5/5)	12000	An, Cm, Di, Fe, Ha, Ne, Sr, Su
TURES	2013	Tu	582(394/188)	5100	An, Fe, Ha, Ne, Sa, Su, Other
EMOVO	2014	It	6(3/3)	588	An, Di, Fe, Jo, Ne, Sa, Su
BAUM-1	2017	Tu	31(18/31)		An, Bo, Bt, Co, Cn, Di, Fe, Ha, Sa, Su, Th, Un
RAVDESS	2018	En	24(12/12)	7356	An, Ca, Di, Fe, Ha, Ne, Sa, Su

Abbreviation for languages: Bu: Burmese, Ch: Chinese, Da: Danish, En: English, Fr: French, Ge: German, It: Italian, Ko: Korean, Ma: Mandarin, Pe: Persian, Sl: Slovenian, Sp: Spanish, Te: Telegu, Tu: Turkish.

Abbreviation for emotions: An: Anger, Ap: Approval, At: Attention, Bo: Boredom, Bt: Bothered, Ca: Calm, Cm: Compassion Cn: Concentration, Co: Contempt, Di: Disgust, Em: Emphatic Fe: Fearful, Ha: Happiness, He: Helpless, Jo: Joy, Lm: Lombard, Lo: Loud, Mo: Motherese, Ne: Neutral, Pr: Prohibition, Re: Reprimanding, Rs: Rest, Sa: Sadness,Sr: Sarcastic, So: Soothing,Su: Surprise, Th: Thingking, To: Touchy, Un: Unsure.

II. Speaker Verification Databases:

YOHO: Campbell et al. [87] designed a YOHO database for text-dependent speaker verification based on combination-lock phrase. It was collected by International Telephone & Telegraph under US Government contract. It consisted of 138 speakers where there were 108 males and 30 females. There were 4 enrollment sessions per subject of 24 phrases each and 10 test sessions per subject of 4 phrases each. They used sample rate and sample coding of 8 kHz and 16-bit word respectively. The size of the data was 15 GB and device used foe recording was microphone.

MAT-2000: Wang et al. [88] describe a database, MAT-2000 (Mandarin speech data Across Taiwan), of Mandarin Chinese spoken in Taiwan. It was produced by ACLCLP and Philips Research East-Asia and collected through telephone networks. MAT-2000 consisted of 2232 speakers, out of which 1227 were female and 1005 were male, with 83.7h of recording and 641,936 spoken syllables. Sampled data were recorded in binary format at sampling rate 8 kHz and encoded as 16-bit linear PCM.

BANCA: Bailliere et al. [89] described BANCA database, a multi-modal database for multi-modal verification systems. It contains recordings of 208 subjects in four different European languages in controlled, degraded and adverse scenarios. There are 12 sessions in total, 4 for each scenario, spanned over 3 months. They used a digital camera and a webcam and recorded audio in 16 bit and 12 bit @32 kHz.

VidTIMIT: Sanderson et al. [90] created VidTIMIT database, which is an audio-visual multi-modal database for face and speech recognition, identification and verification. It has 44 files with recordings of 43 peoples, 24 being male and 19 being female, in 3 sessions. They recorded audio with 16 bit mono @32 kHz in WAV format and video with resolution 512 x 384 in JPEG format. It has 10 sentences per person and collected from NTIMIT database.

MIT-MDSVC: Woo al el. [91] discussed about MIT-MDSVC, which was collected for speaker verification research at MIT. It consisted of a set of enrolled users and a set of imposters. Data were collected in two sessions and 54 samples per user were recorded in each session, which resulted in 5,184 examples from enrolled users and 2,700 examples from imposter users. They used 48 speakers in the enrollment set and 40 speakers in imposter set.

BioSecure: Ortega-Garcia et al. [92] described a multimodal database, BioSecure, which is a collection of biometric data such as voice, iris, face, fingerprint, hand and signature modalities. It consists of three datasets: internal, desktop and mobile dataset. Internal dataset has 2 sessions with 971 donors, desktop dataset has 2 sessions with 667 donors and the mobile dataset has 2 sessions with 713 donors.

MOBIO: S. Marcel [93] et al. in 2010 used MOBIO database to evaluate the performance of speaker verification methods in mobile environment. MOBIO consisted of diverse set of bi-modal data i.e. audio and video data. The data were captured from 152 participants in a ratio of 1:2 i.e. 100 male participants and 52 female participants using a mobile phone and a laptop. There were 12 sessions for each client divided in two phases; 6 sessions for each phase. The 1st phase consisted of 21 questions while the 2nd phase consisted of 11 questions.

UNMC-VIER: Wong et al. [94] created UNMC-VIER database for robust audio-visual recognition systems. It contains video as well as audio recordings of 123 subjects, out of which 74 are males

and 49 are females. Recordings were done in both controlled and uncontrolled environments. They used a camcorder and a webcam for audio and video recording. In controlled environment, with camcorder, they recorded audio with 16 bit stereo @48 kHz in MP2 format and with webcam, they recorded audio with 16 bit mono @22 kHz in PCM format. In uncontrolled environment, audio recording was same as in controlled environment for camcorder; but for webcam, it was 16 bit mono @32 kHz with WMV2 format.

AusTalk: Burnham et al. [95] described AusTalk, a Australian speech database that included component of emotional speech. It was recorded in English with 1000 speakers in 3 one-hour each per participant sessions. A total of 322 words and a set of 58 sentences were selected for design the test set.

RSR2015: Larcher [96] et al. from Human Language Technology department, Institute for Infocomm Research designed a database named RSR2015, stands for Robust Speaker Recognition 2015, for text-depended speaker verification. It was based on fixed-size pass-phrases. It has 300 speakers aged between 17 to 42 years where number of male is 157 and number of female is 143. They recorded more than 151 hours of speech data in English language in 9 sessions, each of consisted of 30 short sentences. They used 6 smart phones and 1 tablet and selected the sentences from TIMIT database.

SAS: Wu et al [97] developed SAS (spoofing and anti-spoofing) database for text-independent automatic speaker verification. SAS has two sub datasets: SAS-VCTK based on Voice Cloning Toolkit database and SAS-RSR based on RSR2015 database. It starts with data from VCTK database with 45 males and 61 females. They divided the data into: Part A, Part B, Part C, Part D, Part E with 24 parallel utterances (for spoofing), 20 non-parallel utterances (for spoofing), 50 non-parallel utterances (for verification), 100 non-parallel utterances (for verification) and 200 non-parallel utterances (for verification) per speaker respectively. In Part A and Part B, all signals were sampled to 48 kHz and 16 kHz respectively, while in Parts C, Part D and Part E, all signals are sampled to 16 kHz.

VoxCeleb: Nagrani [98, 99] et al. developed an audio-visual dataset named VoxCeleb that contains short clips of speech collected from YouTube for speaker identification and verification. It was released in two stages: VoxCeleb1 and VoxCeleb2. VoxCeleb1 contains 1251 speakers of which male being 690 and female being 561 and 153516 utterances while VoxCeleb2 contains 6112 speakers of which male being 3761 male and female being 2351 and 1,128,246 utterances. Data are extracted from 22496 videos for VoxCeleb1 and 150480 videos for VoxCeleb2.

DeepMine: Zeinali et al. [100] designed DeepMine database in English and Persian language for text-independent, text-dependent and text-prompted speaker verification. It consists of 1969 speakers with 1149 males and 820 females and more than 540000 hours of recordings with 22742 sessions. It has three parts: fixed phrase for text-dependent verification, random sequence of words for text-prompted verification and random phonetic level transcription phrase for text-independent verification.

HI-MIA: X. Qin [101] et al. in 2020 introduced HI-MIA database for text-dependent speaker verification in far-field conditions in both English and Chinese. It includes two sub databases: AISHELL-wakeup dataset and AISHELL2019B-eval dataset, with utterances from 254 and 86 speakers respectively. The AISHELL-wakeup consisted of 3,936,003 utterances with 131 male and 123 female speakers. They recorded 160 utterances for each speaker with 120 in noisy environment and 40 in clean environment. The AISHELL2019B-eval dataset consisted of with 44 male and 42 female speakers. They set up the dataset with 40 in noisy environment and 120 in clean environment.

Table 2:*Databases for speaker verification*

Database	Year	#Speaker (Male/Female)	Phrase	Age Group	Language	Session	Environment/ Condition
YOHO	1995	138(108/30)	Combination lock	--	En	14	Office
MAT-2000	2000	2232(1005/1227)	Random	--	Ma	--	Noisy
BANCA	2003	208(104/104)	--	--	En, Fr, It, Sp	12	Quiet & Noisy
VidTIMIT	2003	43(24/19)	--	--	En	3	Noisy
MIT-MDSVC	2006	88(49/39)	--	--	En	2	Quiet & Noisy
BioSecure	2008	2351	--	18-75	En	6	Quiet & Noisy
MOBIO	2010	152(100/52)	--	--	En	12	Office
UNMC-VIER	2010	123(74/49)	--	--	En	2	Quiet & Noisy
AusTalk	2011	1000(500/500)	--	<25->50	En	3	Quiet
RSR2015	2012	300 (157/143)	Fixed-size pass	17 to 42	En	9	Office
SAS	2015	106(45/61)	Fixed, random	--	En	--	Clean
VoxCeleb1	2017	1251(690/561)	Random	--	Multi	--	Multi-Media
VoxCeleb2	2018	6112(3761/2351)	Random	--	Multi	--	Multi-media
DeepMine	2019	1969(1149/820)	Fixed, random	10 to 60	En, Pe	22742	--
HI-MIA	2020	340(175/165)	--	10 to 50+	En, Ch	--	Quiet & Noisy

Abbreviation for Language: Ch: Chinese, En: English, Fr: French, It: Italian, Ma: Mandarin, Pe: Persian, Sp:

V. Conclusion:

Emotion recognition from speech signal is quite difficult because speaking styles, speaking rates of the speakers is different from person to person and it also changes from place to place i.e. different for native speakers and non-native speakers. Hence it is more important to select particular speech features which are not affected by the culture, region, and speaking style of the speaker. Frequency analysis of a signal can provide more relevant information than the time domain analysis. There are different spectral, prosodic, and acoustic properties of the signal which contains the information which are helpful in extracting features from speech signal. After feature extraction, feature selection is also very important and then followed by a suitable classifier to recognize the emotions. In this paper, we have briefly discussed some aspects of speaker and speech emotion recognition system. We have discussed features to characterize different emotional state, different feature extraction approaches and some databases for emotion detection and speaker verification. Performance of emotion recognition and speaker verification system depends on the proper selection of these aspects as different combinations of these may produce different results. We have presented some features with some basic merits and demerits of feature extraction approaches, information of emotional and verification databases; but have not provided any comparison among them.

References

[1] Soltani, K. and Ainon, R. N. (2007), Speech emotion detection based on neural networks, *9th International Symposium on Signal Processing and Its Applications*, 1-3.

[2] Mandal, A. K., Das, S. K. and Bhattacharjee, U. (2021), Speech recognition classifiers: a literature review, *Solid State Technology*, 64(2): 5215-5230.

[3] Bosch L. (2003), Emotions, speech and the ASR framework, *Speech Communication.* 40(1-2): 213–225.

[4] Busso, C., Lee, S. and Narayanan S. (2009), Analysis of emotionally salient aspects of fundamental frequency for emotion detection, *IEEE Trans. Audio Speech Language Process*, 17(4): 582–596.

[5] Cowie R., Douglas-Cowie E., Tsapatsoulis, N., Kollias, S., Fellenz, W. and Taylor, J. (2001), Emotion recognition in human–computer interaction, *IEEE Signal Processing Magazine,* 18(1): 32–80.

[6] Cowie R., and Cornelius, R. R. (2003), Describing the emotional states that are expressed in speech, *Speech Communication*, 40 (1–2): 5–32.

[7] Johnstone, T. and Scherer, K. R. (2000), Vocal Communication of Emotion, second ed., Guilford, New York, 226–235.

[8] Williams, C. and Stevens, K. (1981), Vocal correlates of emotional states, *Speech Evaluation in Psychiatry, Grune and Stratton,* 189–220.

[9] Banse, R. and Scherer, K. (1996), Acoustic profiles in vocal emotion expression, *Journal of Personality and Social Psychology*, 70(3): 614–636.

[10] Murray, I. and Arnott, J. (1993), Toward a simulation of emotions in synthetic speech: A review of the literature on human vocal emotion, *Journal of the Acoustic Society of America*, 93(2): 1097–1108.

[11] Lee, C. and Narayanan, S. (2005), Towards detecting emotions in spoken dialogs, *IEEE Trans. Speech Audio Process*ing, 13(2): 293–303.

[12] Cowie, R. and Douglas-Cowie, E. (1996), Automatic statistical analysis of the signal and prosodic signs of emotion in speech, *Proc. Fourth International Conference on Spoken Language*, 3: 1989–1992.

[13] Oster, A. and Risberg, A. (1986), The identification of the mood of a speaker by hearing impaired listeners, *Speech Transmission Lab. Quarterly Progress Status Report 4, Stockholm, 1986*, 79–90.

[14] Beeke, S., Wilkinson, R. and Maxim, J. (2009), Prosody as a compensatory strategy in the conversations of people with agrammatism, *Clinical Linguistic & Phonetics*, 23(2): 133–155.

[15] Borchert,]. M. and Dusterhoft, A. (2005), Emotions in speech—experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments, *Proc. 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, 147–151.

[16] Tao, J., Kang, Y. and Li, A. (2006), Prosody conversion from neutral speech to emotional speech, *IEEE Trans. Audio Speech Language Processing*, 14(4): 1145–1154.

[17] Cahn, J. (1990), The generation of affect in synthesized speech, *Journal of the American Voice I/O Society*, 8: 1–19.

[18] Rabiner, L. and Schafer, R., Digital Processing of Speech Signals, first ed., Pearson Education, 1978.

[19] Davitz, J. R., The Communication of Emotional Meaning, McGraw-Hill, New York, 1964.

[20] Gobl, C. and Chasaide, A. N. (2003), The role of voice quality in communicating emotion, mood and attitude, *Speech Communication*, 40(1–2): 189–212.

[21] Schlosberg, H. (1954), Three dimensions of emotion, *Psychological Rev.*, 61(2): 81–88.

[22] Scherer, K. R. (1996), Vocal affect expression: a review and a model for future research, *Psychological Bull.*, 99(2): 143–165.

[23] Wu, S., Falk, T. H. and Chan, W. (2011), Automatic speech emotion recognition using modulation spectral features, *Speech Communication*, 53(5): 768-785.

[24] Ayadi, M. E., Kamel, M. S. and Karray, F. (2011), Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern Recognition,* 44(3): 572–587.

[25] Nwe, T., Foo, S. and De Silva, L. (2003) Speech emotion recognition using hidden markov models, *Speech Communication*, 41(4): 603–623.

[26] Kaiser, L. (1962), Communication of affects by single vowels, *Synthese*, 14(4): 300–319.

[27] Morgan, N., Zhu, Q., Stolcke, A., Sonmez, K., Sivadas, S., Shinozaki, T., Ostendorf, M., Jain, P., Hermansky, H., Ellis, D., Doddington, G., Chen, B., Cetin, O., Bourlard, H. and Athineos, M. (2005), Pushing the envelope – aside, *IEEE Signal Processing Magazine*, 22(5): 81–88.

[28] Beyramienanlou, H. and Lotfivand, N. (2018), An efficient teager energy operator-based automated QRS complex detection, *Journal of Healthcare Engineering*, 2018: 1-11.

[29] Teager, H. and Teager, S. (1990), Evidence for nonlinear production mechanisms in the vocal tract, *Hardcastle W.J., Marchal A. (eds) Speech Production and Speech Modelling, NATO Advanced Institute*, 55: 241–261.

[30] Teager, H. (1990), Some observations on oral air flow during phonation, *IEEE Trans. Acoustic Speech Signal Processing*, 28(5): 599–601.

[31] Zhou, G., Hansen, J. and Kaiser, J. (2001), Nonlinear feature based classification of speech under stress, *IEEE Trans. Speech Audio Processing*, 9(3): 201–216.

[32] Caims, D. and Hansen, J. (1994), Nonlinear analysis and detection of speech under stressed conditions, *Journal of the Acoustic Society of America*, 96(6): 3392–3400.

[33] Alim, S. A. and Rashid, N. K. A. (2018), Some commonly used speech feature extraction algorithms, *From Natural to Artificial Intelligence - Algorithms and Applications,* R. Lopez-Ruiz, Ed., IntechOpen.

[34] Ravikumar, K. M., Reddy B. A., Rajagopal, R. and Nagaraj, H. C. (2008), Automatic detection of syllable repetition in read speech for objective assessment of stuttered Disfluencies, *Proc. World Academy Science, Engineering and Technology*, 36: 270-273.

[35] Chakroborty, S., Roy, A. and Saha, G. (2006), Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification, *IEEE International Conference on Industrial Technology*, 387-390.

[36] Chu, S., Narayanan, S. and Kuo, C. C. (2008), Environmental sound recognition using MP-based features, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1-4.

[37] Shah, S. A. A., ul Asar A, Shaukat SF. Neural network solution for secure interactive voice response, *World Applied Sciences Journal*, 6(9):1264-1269.

[38] Mubarak, O. M., Ambikairajah, E. and Epps, J. (2005), Analysis of an MFCC-based audio indexing system for efficient coding of multimedia sources, *8th international symposium on signal processing and its applications, Sydney, Australia*.

[39] Corrnaz C, Hunkeler U, Velisavljevic V., An automatic speaker recognition system. Switzerland: Lausanne; 2003. Retrieved from: http://read.pudn.com/downloads60/sourcecode/multimedia/audio/209082/asr_project.pdf

[40] Rao, T. B., Reddy, P. and Prasad, A. (2011), Recognition and a panoramic view of Raaga emotions of singers-application Gaussian mixture model, *International Journal of Research and Reviews in Computer Science*, 2(1): 201-204.

[41] Narang, S. and Gupta, M. D. (2015), Speech feature extraction techniques: a review, *International Journal of Computer Science and Mobile Computing*, 4(3): 107-114.

[42] Al-Sarayreh, K. T., Al-Qutaish, R. E. and Al-Kasasbeh, B. M. (2009), Using the sound recognition techniques to reduce the electricity consumption in highways, *Journal of American Science*, 5(2): 1-12.

[43] Kumar, P. and Chandra, M. (2011), Speaker identification using Gaussian mixture models. *MIT International Journal of Electronics and Communication Engineering,* 1(1): 27-30.

[44] Kumar, P. P., Vardhan. K. and Krishna, K. (2010), Performance evaluation of MLP for speech recognition in noisy environments using MFCC & wavelets, *International Journal of Computer Science & Communication*, 1(2): 41-45.

[45] Agrawal, S., Shruti, A. K. and Krishna, C. R. (2010), Prosodic feature based text dependent speaker recognition using machine learning algorithms, *International Journal of Engineering Science and Technology,* 2(10): 5150-5157.

[46] Kumar, R., Ranjan, R., Singh, S. K., Kala, R., Shukla, A. and Tiwari, R. (2009), Multilingual speaker recognition using neural network, *Proceedings of the Frontiers of Research on Speech and Music*, 1-8.

[47] Paulraj, M. P., Sazali, Y., Nazri, A. and Kumar, S. (2009), A speech recognition system for Malaysian English pronunciation using neural network, *Proc. of the International Conference on Man-Machine Systems*.

[48] Tan, C. L. and Jantan, A. (2004), Digit recognition using neural networks, *Malaysian Journal of Computer Science*, 17(2): 40-54.

[49] Anusuya, M. A. and Katti, S. K. (2009), Speech Recognition by Machine: A Review, *International journal of computer science and Information Security*, 6(3): 181-205.

[50] El Choubassi, M. M., El Khoury, H. E., Alagha, C. E. J, Skaf, J. A. and Al-Alaoui, M. A. (2003), Arabic speech recognition using recurrent neural networks. *Proc. 3rd IEEE International Symposium on Signal Processing and Information Technology*, 543-547.

[51] Wu, Q. Z., Jou, I. C. and Lee, S. Y. (1997), On-line signature verification using LPC cepstrum and neural networks, *IEEE Transactions on Systems, Man, and Cybernetics*, *Part B: Cybernetics,* 27(1): 148-153.

[52] Rabiner, L. and Juang, B., Fundamentals of Speech Recognition, PTR Prentice Hall, NJ, 1993.

[53] Zhu, L. and Yang, Q. (2012), Speaker Recognition System Based on weighted feature parameter, *Physics Procedia*, 25: 1515-1522.

[54] Itakura, F. (1975), Line spectrum representation of linear predictive coefficients of speech signals, The Journal of the Acoustic Society of America, 57(S1): S35.

[55] Sahidullah, Md., Chakroborty, S. and Saha, G. (2010), On the use of perceptual Line Spectral pairs Frequencies and higher-order residual moments for Speaker Identification, *International Journal of Biometrics*. 2(4): 358–378.

[56] Ma, Z., Taghia, J., Kleijn, Bastiaan, W., Leijon, A. and Guo, J. (2015), Line spectral frequencies modeling by a mixture of von Mises–Fisher distributions, *Signal Processing,* 114:, 219–224.

[57] de Alencar V. F. S. and Alcaim A. (2005), Transformations of LPC and LSF Parameters to Speech Recognition Features. *In: Singh S., Singh M., Apte C., Perner P. (eds) Pattern Recognition and Data Mining. ICAPR 2005. Lecture Notes in Computer Science*, 3686. Springer, Berlin, Heidelberg.

[58] Saito, S. (1992), Speech Science and Technology, Ohmsha, Tokyo, Japan, 81–86.

[59] McLoughlin, I. V. (2008), Line spectral pairs, *Signal processing*, 88: 448-467.

[60] Li, J., Chaddha, N. and Gray, R. M. (1999), Asymptotic performance of vector quantizers with a perceptual distortion measure, *IEEE Trans. Information Theory,* 45(May): 1082–1091.

[61] Hao, Y. and Zhu, X. (2000), A new feature in speech recognition based on wavelet transform, *Proc. IEEE 5th Inter Conf on Signal Processing,* 3(3): 1526-1529.

[62] Soman, K. P. and Ramchandran, K. I. (2005), Insight into Wavelets from Theory to Practice, 2nd edition, Prentice-Hall of India, New Delhi.

[63] Nehe, N. S. and Holambe, R. S. (2012), DWT and LPC based feature extraction methods for isolated word recognition, *EURASIP Journal on Audio, Speech, and Music Processing*, 7(2012): 1-7.

[64] Gałka, J. and Ziółko, M. (2009), Wavelet speech feature extraction using mean best basis algorithm, *International Conference on Nonlinear Speech Processing Berlin*, 128-135.

[65] Hermansky, H. (1990), Perceptual linear predictive (PLP) analysis of speech, The Journal of the Acoustic Society of America, 87(4): 1738–1752.

[66] Honig, F., Stemmer, G., Hacker, C. and Brugnara, F. (2005), Revising perceptual linear prediction (PLP), *INTERSPEECH 2005*, 2997-3001.

[67] Sing, R. and Rao, P. (2007), Spectral subtraction speech enhancement with RASTA filtering, *Proc. National Conference on Communications*, Kanpur, India.

[68] Ahmed, A. I., Chiverton, J. P., Ndzi, D. L. and Becerra, V. M. (2019), Speaker recognition using PCA-based feature transformation, *Speech Communication*, 110: 33-46.

[69] Jing, X., Ma, J., Zhao, J. and Yang, H. (2014), Speaker recognition based on principal component analysis of LPCC and MFCC, *IEEE Int. Conf. on Signal Processing, Communications and Computing*, 403-408.

[70] Kumaran, U., Rammohan, R., Nagarajan, S., et al. (2021), Fusion of mel and gammatone frequency cepstral coefficients for speech emotion recognition using deep C-RNN, *Int Journal of Speech Technology*, 24: 303–314.

[71] Kotnik, B. and Kacic, Z. (2007), A comprehensive noise robust speech parameterization algorithm using wavelet packet decomposition-based denoising and speech feature representation techniques, *EURASIP Journal on Advances in Signal Processing*, 1-20.

[72] Mini, P. P., Thomas, T. and Gopikakumari, R. (2021), Wavelet feature selection of audio and imagined/vocalized EEG signals for ANN based multimodal ASR system, *Biomedical Signal Processing and Control*, 63.

[73] Engberg, I. S., Hansen, A. V. (1996), Documentation of the Danish Emotional Speech database (DES), *Internal AAU report, Center for Person Kommunikation, Aalborg Univ., Denmark (Online)*.

[74] Hansen, J. H. L., and Bou-Ghazale, S. E. (1997), Getting started with SUSAS: a speech under simulated and actual stress database, *Proc. EuroSpeech 97*, 4: 1743–1746.

[75] Breazeal, C. and Aryananda, L. (2002), Recognition of affective communicative intent in robot-directed speech, *Autonomous Robots*, 12: 83–104.

[76] Hozjan, V., Kacic, Z., Moreno, A., Bonafonte, A., and Nogueiras, A. (2002). Interface databases: Design and collection of a multilingual emotional speech database, *Proc. 3rd international conference on language*, 2019–2023.

[77] Schuller, B., Reiter, S., Muller, R., Al-Hames, M., Lang, M. and Rigoll, G. (2005) Speaker independent speech emotion recognition by ensemble classification, *IEEE International Conference on Multimedia and Expo*, 864–867.

[78] Burkhardt, F., Paeschke, A. Rolfes, M., Sendlmeier, W. and Weiss, B. (2005), A database of German emotional speech, *Proc. INTERSPEECH 2005*, 1517–1520.

[79] Zhou, J. Wang, G., Yang, Y. and Chen, P. (2006), Speech emotion recognition based on rough set and svm, *5th IEEE International Conference on Cognitive Informatics*, 53–61.

[80] Kim, E., Hyun, K., Kim, S. and Kwak, Y. (2007), Speech emotion recognition using eigen-fft in clean and noisy environments, *16th IEEE International Symposium on Robot and Human Interactive Communication*, 689–694.

[81] Batliner, A., Steidl, S. and Noth, E. (2008), Releasing a thoroughly annotated and processed spontaneous emotional database: the FAU Aibo Emotion Corpus, *Proc. of a Satellite Workshop of LREC 2008 on Corpora for Research on Emotion and Affect*, 28-31.

[82] Koolagudi S. G., Maity S., Kumar V. A., Chakrabarti S. and Rao K. S. (2009), IITKGP-SESC: Speech database for emotion analysis, *Ranka S. et al. (eds) Contemporary Computing. IC3 2009. Communications in Computer and Information Science,* 40: 485-492.

[83] Oflazoglu, C. and Yildirim, S. (2013), Recognizing emotion from Turkish speech using acoustic features. *EURASIP Journal on Audio, Speech, and Music Processing*, 1: 1-11.

[84] Costantini, G. Iaderola, I., Paoloni, A. and Todisco, M. (2014), Emovo corpus: an italian emotional speech database, *Int. Conf. on Language Resources and Evaluation*, 3501–3504.

[85] Zhalehpour, Z., Onder, O., Akhtar, Z. and Erdem, C. E. (2017), Baum-1: A spontaneous audio-visual face database of affective and mental states, *IEEE Transactions on Affective Computing*, 8(3):300–313, 2017.

[86] Livingstone, S. R. and Russo, F. A. (2018), The ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, *PLoS ONE*, 13(5).

[87] Campbell, J. and Higgins, A., YOHO Speaker Verification LDC94S16. Web Download. Philadelphia: Linguistic Data Consortium, 1994.

[88] Wang H. C., Seide F., Tseng C. Y. and Lee L. S. (2000), MAT2000 – design, collection, and validation of a Mandarin 2000-speaker telephone speech database," Proc. ICSLP-2000, 4: 460-463.

[89] Bailly-Bailliére E. et al. (2003) The BANCA database and evaluation protocol, *In: Kittler J., Nixon M.S. (eds) Audio- and Video-Based Biometric Person Authentication. AVBPA 2003. Lecture Notes in Computer Science*, 2688: 625–638.

[90] Sanderson, C. (2002), Automatic person verification using speech and face information", *PhD Thesis, School of Microelectronic Engineering, Griffith University, Brisbane, Australia*.

[91] Woo, R. H., Park, A. and Hazen, T. J. (2006), The MIT mobile device speaker verification corpus: data collection and preliminary experiments, *Proc. of Odyssey, The Speaker & Language Recognition Workshop*, 1-5.

[92] Ortega-Garcia, J., et al. (2010), The multiscenario multienvironment biosecure multimodal database, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6): 1097-1111.

[93] Marcel, S., et al. (2010), Mobile biometry (MOBIO) face and speaker verification evaluation, *ICPR 2010*, 210-255.

[94] Wong, Y., Ch'ng, S., Seng, K., Ang, L., Chin, S., Chew, W. and Lim, K. (2011), A new multi-purpose audio-visual UNMC-VIER database with multiple variabilities, *Pattern Recognition Letters*, 32(13), 1503–1510.

[95] Burnham, D., et al. (2011), Building an audio-visual corpus of Australian English: Large corpus collection with an economical portable and replicable black box, *Proc. ISCA, 2011*.

[96] Larcher, A., Lee, K., Ma, B. and Li, H. (2014), Text-dependent speaker verification: classifiers, databases and RSR2015, *Speech Communication*, 60: 56-77.

[97] Wu, Z., *et al*. (2015), SAS: A speaker verification spoofing database containing diverse attacks, *IEEE International Conference on Acoustics, Speech and Signal Processing, 2015*, 4440-4444.

[98] Nagrani, A., Chung, J. S. and Zisserman, A. (2017), VoxCeleb: A Large-Scale Speaker Identification Dataset, *Proc. INTERSPEECH 2017*, 2616-2620.

[99] Chung, J. S., Nagrani, A. and Zisserman, A. (2018), VoxCeleb2: Deep Speaker Recognition, *Proc. INTERSPEECH 2018*, 1086-1090.

[100] Hossein, Z, Jan, C. and Lukáš, B. (2019), A multipurpose and large scale speech corpus in Persian and English for speaker and speech recognition: the DeepMine database, *Proc. IEEE Automatic Speech Recognition and Understanding Workshop, 2019*, 397-402.

[101] Qin, X., Bu, H. and Li, M. (2020), HI-MIA: a far-field text-dependent speaker verification database and the baselines, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, 7609-7613.

Parameter estimation for progressive censored data under accelerated life test with *k* levels of constant stress

Mustafa Kamal
•
Department of Basic Sciences, College of Science and Theoretical Studies,
Saudi Electronic University, KSA
e-mail: m.kamal@seu.edu.sa

Abstract

Accelerated life testing (ALT) is a time-saving technique that has been used in a variety of sectors to get failure time data for test units in a relatively short time it takes to test them under regular operating circumstances. One of the primary goals of ALT is to estimate failure time functions and reliability under typical use. In this article, an ALT with k increasing stress levels that is stopped by a type II progressive censoring (TIIPC) scheme is considered. At each stress level, it is assumed that the failure times of test units follow a generalized Pareto (GnP) distribution. The link between the life characteristic and stress level is considered to be log-linear. The maximum likelihood estimation (MLE) method is used to obtain inferences about unknown parameters of the model. Furthermore, the asymptotic confidence intervals (ACIs) are obtained by utilizing the inverse of the fisher information matrix. Finally, a simulation exercise is presented to show how well the developed inferential approaches performed. The performance of MLEs is assessed in terms of relative mean square error (RMSE) and relative absolute bias (RAB), whereas the performance of ACIs is assessed in terms of their length and coverage probability (CP).

Keywords: Simulation, type-II progressive censoring, multiple constant stress accelerated life test

I. Introduction

Traditional reliability tests are designed to examine failure time data acquired under normal operating circumstances. However, due to restricted testing time and highly reliable products like as electronics systems, insulating materials, engines, and so on, such life data is not easy to obtain. As a result, the use of traditional reliability tests is inappropriate, time consuming and expensive. Therefore, ALT is widely used in the manufacturing and production industries due to its capacity to provide timely and adequate failure data for product reliability and design assessment. Furthermore, the growing competitiveness of innovation, as well as the desire to shorten product development time, have underlined the use and significance of ALT techniques. It is especially difficult to detect faults in highly reliable items or systems under typical operating conditions in a short period of time. Thus, in the manufacturing business, ALT has become an essential element of the product design and development process. In ALT, samples of test items are exposed to more intense levels of stress, such as temperature, voltage, humidity etc. to cause early failures, and the resultant failure times and the used censoring schemes are recorded. The data is then utilized to create an ALT model for extrapolating the product's reliability under normal operation conditions.

The stress loading in ALT may be applied in a various different way, however the most often utilized stress loadings are constant, step, and progressive stress loadings [1] (abbreviated as

CSALT, SSALT & PSALT). In CSALT, units are tested at two or more constant levels of high stress until they all fail or the test is stopped owing to some censoring scheme or other factors. Many researchers have investigated CSALT models, including [2-9]. In SSALT, the testing units are initially subjected to a starting high level of stress and the failures are noted and then the test items are removed at prespecified time to test at the next level of stress, and so on. Many scholars have looked at the SSALT models, including [10-19]. Progressive stress loading, often known as ramp stress loading, is a method of exposing test units to gradually increasing stress over time. PSALT designs were initially proposed by [20], who took into account both exponential and Weibull life distributions. Since then, several writers have explored PSALT for different distributions with different types of data, including [21-25].

Most items are subjected to constant stress when they are in use in general. The CSALT is straightforward in most tests, making it easier to maintain a consistent stress level and the CSALT mimics actual use of the product. For some materials and products, CSALT models are better designed and empirically validated. Furthermore, data analysis for estimating reliability is well established and automated. However, CSALT is the most commonly utilized ALT method because of its above-mentioned benefits, but the majority of ALT research has concentrated on statistical inference with only two or three levels of constant stress.

So far, there are only a few studies which have considered multi-stress CSALT. [26] discussed the reliability analysis of type-I censored Weibull failure data obtained from a system of multiple components connected in series under CSALT assuming a log-linear relation between scale parameter and the stress variable. [27] used TIIPC data to develop MLEs and Bayes estimates (BEs) of the parameters of the extended exponential distribution under CSALT. [28] examined several estimating methods for the parameters of the exponentiated distributions family, with the exponentiated inverted Weibull regarded as a particular example under CSALT. [29] estimated the parameters of a lower-truncated family of distributions using the MLE approach for simple CSALT under TIIPCS. Assuming mean life as a linear function of the stress, [30] considered hybrid type-I censored data from a CSALT and obtained the MLEs and approximation MLEs of the parameters of a generalized log-location-scale distribution. [31] investigated a multiple stress CSALT and utilized MLE and BE approaches to construct point and interval estimates of Weibull distribution parameters based on TIIP adaptive hybrid censored data. Assuming that failure under arithmetically increasing stress levels of CSALT, [32] employed MLE methods for estimating the Burr-X life distribution parameters. [33] studied CSALT and estimated the doubly truncated Burr-XII parameters using MLE and BE methods. [34] address the problem of statistical inference using MLE and BE approaches for TIIPC data under multiple stress CSALT, assuming that failure times follow the modified Kies exponential distribution and removal follows a binomial distribution.

In this paper, an ALT with k constant stress levels which is stopped by a TIIPC scheme is considered. The following is how the paper is structured. Section 2 provides fundamental terminology, failure distribution, and basic multi-stress CSALT assumptions. In Section 3, the MLE technique is employed to derive estimates of the parameters using TIIPC data. In Section 4, a simulation study with different test setting is conducted to compare the performance of the proposed model. Section 5 concludes the paper with some remarks.

II. Assumptions and procedure for k-stress level ALT

A k levels of CSALT is considered. Let Q_0 be the normal stress level and $Q_i, i = 1,2,\dots k$ are the k levels of applied higher constant stress levels. The following assumptions are used in this paper:

1. Suppose $n_i, i = 1, 2, \dots, k$ are samples containing independent and identical items put on test at the same time at stress levels $Q_i, i = 1,2,\dots k$ in such way that $\mathcal{N} = \sum_{i=1}^{k} n_i$, where $\mathcal{N}$ total number of items assigned on all stress levels to test.

2. The product's life has a GnP distribution under normal stress Q_0 and accelerated stress $Q_i, i = 1,2,\dots k$. The density function, cumulative distribution function, the reliability function and the hazard rate function of GnP distribution are as follows:

$$f(t_i, \phi_i, \psi_i) = \phi_i \psi_i (1 + \psi_i t_i)^{-(\phi_i+1)}, \quad t_i, \phi_i, \psi_i > 0 \tag{1}$$

$$F(t_i, \phi_i, \psi_i) = 1 - (1 + \psi_i t_i)^{-\phi_i}, \qquad t_i, \phi_i, \psi_i > 0 \tag{2}$$

$$R(t_i, \phi_i, \psi_i) = (1 + \psi_i t_i)^{-\phi_i}, \qquad t_i, \phi_i, \psi_i > 0 \tag{3}$$

$$h(t_i, \phi_i, \psi_i) = \frac{\phi_i \psi_i}{(1 + \psi_i t_i)}, \qquad t_i, \phi_i, \psi_i > 0 \tag{4}$$

where ϕ_i is the shape parameter and the scale parameter is ψ_i at stress level $Q_i, i = 1,2,\dots k$.

3. At each increased stress, the product's failure mechanism stays unchanged. Because ϕ_i specifies the failure mechanism, it follows that

$$\phi_0 = \phi_1 = \phi_2 = \dots = \phi_k = \phi \tag{5}$$

4. The parameter ψ_i has a log-linear relationship with the stress variable Q_i and may be described as follows:

$$log\, \psi_i = \alpha + \beta Z_i, \quad i = 0, 1, 2, \dots, k \tag{6}$$

where α and $\beta(> 0)$ are the unknown parameters of the relationship and their values usually depend on true nature of the test items. And $Z_i = Z(Q_i)$ is an increasing function of stress Q_i. Eq. (6) depends on the type of stress used for testing, e. g., if stress is temperature, then, the Arrhenius model is used and can written as $log\, \psi_i = \alpha + \beta/Q_i$, where Q_i is temperature stress. If stress is voltage, then the inverse power model is appropriate to be used and can written as $log\, \psi_i = \alpha + \beta(log(Q_i))$, where Q_i is voltage stress. For weather conditions, exponential model is used and can written as $log\, \psi_i = \alpha + \beta Q_i$. Above defined three well-known models may be converted into the linear form as in eq. (6) by transforming the stress with $Z(Q_i) = 1/Q_i, log(Q_i)$ and Q_i respectively.

Let $t_{ij}, i = 1,2,\dots k;\ j = 1, 2, \dots, n_i$ are observed ordered failures with progressive censoring scheme $w_{ij} = (w_{i1,} w_{i2}, \dots, w_{im_i})$, $i = 1,2,\dots k;\ j = 1, 2, \dots, m_i$ at each Q_i stress level. Based on k stress ALT with TIIPC scheme w_{ij}, n_i units are put under accelerated testing condition Q_i and the experiment will be run until m_i failures at each stress level and the number of failures is prefixed. Now, the TIIPC can be implemented as follows: at each $Q_i, i = 1,2,\dots k$, at first failure time t_{i1}, w_{i1} items are omitted from the remaining $(n_i - 1)$ survivals randomly. Similarly at time t_{i2}, w_{i2} is the number of removed items from $(n_i - 2 - w_{i1})$ remaining survivals and so on until the desired number of failures m_i, $i = 0, 1, 2, \dots, k$ at each stress level obtained and then the test is terminated by removing all the remaining survivals $w_{ij} = n_i - m_i - \sum_{i=1}^{m_i - 1} w_{ij}$ from the test.

III. Parameter estimation

In statistics, MLEs and BE techniques are two of the most significant and commonly used approaches. The MLEs are asymptotically normal and consistent. The BE technique necessitates the selection of previous knowledge of unknown parameters, although this is generally a challenging task in reality. Furthermore, BE method frequently necessitates the use of complicated integral procedures. As a result, in a CSALT, this paper uses an MLE method utilizing TIIPC data.

Let the obtained observed failure TIIPC samples at i^{th} stress level in the considered ALT $t_{i1} \le t_{i2} \le \dots \le t_{im_i}$, $i = 0, 1, 2, \dots, k$, then the likelihood for the observed data under TIIPC scheme can

be obtained in the following form

$$L(\mathcal{y}, \xi, \delta) = \prod_{i=1}^{k} \left\{ C_i \prod_{j=1}^{m_i} \mathbb{f}_{\mathcal{T}_{ij}}(t_{ij}) \left(1 - \mathcal{F}_{\mathcal{T}_{ij}}(t_{ij})\right)^{w_{ij}} \right\} \tag{7}$$

where, $C_i = n_i(n_i - 1 - w_{i1})(n_i - 2 - w_{i1} - w_{i2}) \ldots \sum_{i=1}^{m_i - 1} w_{ij}$. Now, the log likelihood $\ell = L(\mathcal{y}, \xi, \sigma, \delta)$ corresponding to Eq. (7) after substituting the values of $\mathbb{f}_{\mathcal{T}_{ij}}(t_{ij})$ & $\mathcal{F}_{\mathcal{T}_{ij}}(t_{ij})$ and taking log on both sides is obtained as follows:

$$\ell = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \{log(\phi) + log(\psi_i) - (w_{ij}\phi + \phi + 1) log(1 + \psi_i t_{ij})\} \tag{8}$$

Now, from equation (6), we can drive

$$\psi_i = \psi_0 e^{(\beta \mathcal{Z}_i - \beta \mathcal{Z}_0)} = \psi_0 \vartheta^{\Omega_i}, \qquad i = 0, 1, 2, \ldots, k \tag{9}$$

Where, $\psi_0 = \alpha + \beta \mathcal{Z}_0$ represents the GnP distribution's scale parameter at stress $\mathcal{Q}_0$ and $\psi_1/\psi_0 = \vartheta = e^{\beta(\mathcal{Z}_1 - \mathcal{Z}_0)}$ denotes the acceleration factor from $\mathcal{Q}_1$ to $\mathcal{Q}_0$, and

$$\Omega_i = (\mathcal{Z}_i - \mathcal{Z}_0)/(\mathcal{Z}_1 - \mathcal{Z}_0), \qquad i = 0, 1, 2, \ldots, k \tag{10}$$

Because the transformation from (α, β, ϕ) to $(\psi_0, \vartheta, \phi)$ is one-to-one, we can immediately calculate the product's life at $\mathcal{Q}_0$ using the new transformed parameters. As a result, the likelihood function (8) may be rewritten as follows:

$$\ell = \sum_{i=1}^{k} \sum_{j=1}^{m_i} \{log(\phi) + log(\psi_0) + \Omega_i log(\vartheta) - (1 + \phi + w_{ij}\phi) log(1 + \psi_0 \vartheta^{\Omega_i} t_{ij})\} \tag{11}$$

By solving the following likelihood equations, the MLEs of the parameters can now be calculated:

$$\frac{\partial \ell}{\partial \psi_0} = \frac{1}{\psi_0} - \frac{\vartheta^{\Omega_i}(1 + (1 + w_{ij})\phi) t_{ij}}{1 + \psi_0 \vartheta^{\Omega_i} t_{ij}} = 0 \tag{12}$$

$$\frac{\partial \ell}{\partial \vartheta} = \frac{\Omega_i}{\vartheta} - \frac{\psi_0 \vartheta^{\Omega_i - 1} \Omega_i (1 + (1 + w_{ij})\phi) t_{ij}}{1 + \psi_0 \vartheta^{\Omega_i} t_{ij}} = 0 \tag{13}$$

$$\frac{\partial \ell}{\partial \phi} = \frac{1}{\phi} - (1 + w_{ij}) log(1 + \psi_0 \vartheta^{\Omega_i} t_{ij}) = 0 \tag{14}$$

We now have a system of three nonlinear equations with three unknowns, making it difficult to find closed-form solutions manually. Hence, numerical solution to equations is obtained using Newton Raphson iterative approach, the R programming language is used to get the solutions.

By using asymptotic characteristics of the MLEs, the ACIs of the parameters may now be estimated using TIIPC by mathematically inverting Fisher's information matrix. As a result, we can compute the estimates of 95% two-sided ACIs for ψ_0, ϑ and ϕ as follows:

$$\widehat{\psi_0} \pm 1.96 \sqrt{var(\widehat{\psi_0})}; \quad \hat{\vartheta} \pm 1.96 \sqrt{var(\hat{\vartheta})}; \quad \hat{\phi} \pm 1.96 \sqrt{var(\hat{\phi})} \tag{15}$$

Where, $var(\widehat{\psi_0})$, $var(\hat{\vartheta})$ and $var(\hat{\phi})$ are main diagonal entries of inverted Fisher matrix and the elements of the matrix are given by following equations:

$$\frac{\partial^2 \ell}{\partial \phi^2} = -\frac{1}{\phi^2} \tag{16}$$

$$\frac{\partial^2 \ell}{\partial \vartheta^2} = -\frac{\Omega_i(1+\psi_0\vartheta^{\Omega_i}t_{ij}(1-\phi(1+w_{ij})(1+\psi_0\vartheta^{\Omega_i}t_{ij})+\Omega_i(1+\phi+w_{ij}\phi)))}{\vartheta^2(1+\psi_0\vartheta^{\Omega_i}t_{ij})^2} \tag{17}$$

$$\frac{\partial^2 \ell}{\partial \psi_0^2} = -\frac{1}{\psi_0^2} + \frac{\vartheta^{2\Omega_i}(1+(1+w_{ij})\phi)t_{ij}^2}{(1+\psi_0\vartheta^{\Omega_i}t_{ij})^2} \tag{18}$$

$$\frac{\partial^2 \ell}{\partial \phi \partial \vartheta} = \frac{\partial^2 \ell}{\partial \vartheta \partial \phi} = -\frac{\psi_0\Omega_i\vartheta^{\Omega_i-1}(1+w_{ij})t_{ij}}{1+\psi_0\vartheta^{\Omega_i}t_{ij}} \tag{19}$$

$$\frac{\partial^2 \ell}{\partial \phi \partial \psi_0} = \frac{\partial^2 \ell}{\partial \psi_0 \partial \phi} = -\frac{\vartheta^{\Omega_i}(1+w_{ij})t_{ij}}{1+\psi_0\vartheta^{\Omega_i}t_{ij}} \tag{20}$$

$$\frac{\partial^2 \ell}{\partial \psi_0 \partial \vartheta} = \frac{\partial^2 \ell}{\partial \vartheta \partial \psi_0} = -\frac{\Omega_i\vartheta^{\Omega_i-1}(1+\phi+w_{ij}\phi)t_{ij}}{(1+\psi_0\vartheta^{\Omega_i}t_{ij})^2} \tag{21}$$

IV. Simulation Study

In this section, the performance of the considered methodology for estimating the parameters of the GnP distribution based on the CSALT with k stresses for TIIPC data utilizing the log liner association between stress and life characteristic is examined through a Monte-Carlo simulation using the R-package. Two set of initial values ($\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5$), ($\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8$) of parameter with various sample combinations (n_i, m_i)= (20, 10), (20, 15), (30, 15), (30, 20), (30, 25), (40, 20), (40, 25), (40, 30), (50, 25), (50, 30), (50, 35), (60, 35), (60, 40), (60, 45) are selected for simulation. Apart from the normal stress level $Q_0 = 110$, three levels of constant stress are assumed as; $Q_1 = 150$, $Q_2 = 220$ and $Q_3 = 250$ under TIIPC scheme. Additionally, two distinct removal schemes $i. w_{i1}, w_{i2}, \ldots, w_{i(m-1)} = (n_i - m_i)$ & $w_{im_i} = 0$; $ii. w_{i1}, w_{i2}, \ldots, w_{i(m-1)} = 1$ & $w_{im_i} = n_i - m_s + 1$ are used to generate TIIPC samples with various combinations of (n_i, m_i) under three different constant-stress levels. For each test scheme, average RABs and RMSEs for point estimates, as well as lower and upper ACI limits (LACIL, UACIL) and ACIs lengths (ACIsL) of 95% ACIs with corresponding CPs are computed. Following steps are used to perform the simulation study:

Step 1: Set the initial values of the parameters ψ_0, ϑ and ϕ.

Step 2: Set the values of stress levels Q_i, $i = 0, 1, 2, \ldots, k$.

Step 3: Set the values of $(n_i, m_i), i = 0, 1, 2, \ldots, k$ at each stress levels Q_i.

Step 4: Now using the defined values in step 1-3, generate $i = 1, 2, \ldots, k$ random samples of size m_i of TIIPC data from Uniform (0, 1) distribution according to the steps outlined by [35].

Step 5: Using inverse CDF method, for each sample size m_i obtained in step 4, generate TIIPC data from GnP distribution using $(exp(-ln(1-u)/\phi) - 1)/\psi$.

Step 6: For each stress levels along with removal scheme, repeat the steps 1-5 for 10000 times.

Step 7: Compute the average of MLEs of ψ_0, ϑ and ϕ with their respective RABs and RMSEs.

Step 8: Compute LACIL, UACIL, ACIsL of 95% ACIs with corresponding CPs of ψ_0, ϑ and ϕ.

Table 1-6 displays the numerical findings of RMSEs and RABs of MLEs and LACIL, UACIL, and ACIsL, as well as the corresponding CPs of ACIs. Figures 1-3 represent the behavior of the RABs and RMSEs with respect to the sample size.

Table 1: *MLEs with RMSEs, RABs of* ψ_0 *&* *LACIL, UACIL, ACIsL, CPs of ACIs with* $(\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5)$

(n_i, m_i)	CS	MLE	RMSE	RAB	LACIL	UACIL	ACIsL	CP
20, 10	1	1.379815	0.960779	0.439453	-0.503310	3.262942	3.766253	0.950000
20, 15	1	1.439942	0.797518	0.433950	-0.123190	3.003076	3.126269	0.940000
30, 15	1	1.370340	0.705081	0.408689	-0.011620	2.752298	2.763916	0.960000
30, 20	1	1.486785	0.666933	0.357396	0.179597	2.793973	2.614376	0.950000
30, 25	1	1.312451	0.664671	0.392446	0.009696	2.615206	2.605510	0.950000
40, 20	1	1.291834	0.644747	0.383245	0.028131	2.555537	2.527407	0.950000
40, 25	1	1.370246	0.639035	0.389515	0.117737	2.622754	2.505017	0.940000
40, 30	1	1.365194	0.611388	0.358700	0.166874	2.563514	2.396640	0.980000
50, 25	1	1.371536	0.588579	0.350251	0.217920	2.525151	2.307231	0.970000
50, 30	1	1.053091	0.533587	0.382436	0.007259	2.098922	2.091663	0.940000
50, 35	1	1.244513	0.521590	0.330190	0.222196	2.266830	2.044634	0.950000
60, 35	1	1.275332	0.504480	0.344008	0.286551	2.264113	1.977562	0.940000
60, 40	1	1.294818	0.451734	0.283193	0.409419	2.180216	1.770797	0.950000
60, 45	1	1.244503	0.451180	0.266636	0.360191	2.128815	1.768624	0.950000
20, 10	2	1.359424	0.946580	0.432959	-0.495874	3.214721	3.710595	0.980000
20, 15	2	1.418662	0.785732	0.427537	-0.121372	2.958696	3.080068	0.940000
30, 15	2	1.350089	0.694661	0.402649	-0.011446	2.711624	2.723070	0.940000
30, 20	2	1.464813	0.657076	0.352115	0.176943	2.752682	2.575739	0.950000
30, 25	2	1.293055	0.654848	0.386646	0.009553	2.576557	2.567005	0.950000
40, 20	2	1.272743	0.635218	0.377582	0.027715	2.517771	2.490056	0.950000
40, 25	2	1.349996	0.629591	0.383759	0.115997	2.583995	2.467997	0.950000
40, 30	2	1.345019	0.602353	0.353399	0.164408	2.525630	2.361222	0.940000
50, 25	2	1.351267	0.579881	0.345075	0.214700	2.487834	2.273134	0.970000
50, 30	2	1.037528	0.525702	0.376785	0.007152	2.067903	2.060751	0.940000
50, 35	2	1.226121	0.513882	0.325310	0.218912	2.233330	2.014417	0.950000
60, 35	2	1.256485	0.497025	0.338925	0.282316	2.230653	1.948337	0.970000
60, 40	2	1.275682	0.445058	0.279008	0.403369	2.147996	1.744628	0.950000
60, 45	2	1.226111	0.444512	0.262695	0.354868	2.097355	1.742486	0.970000

Table 2: *MLEs with RMSEs, RABs of* ψ_0 *&* *LACIL, UACIL, ACIsL, CPs of ACIs with* $(\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8)$

(n_i, m_i)	CS	MLE	RMSE	RAB	LACIL	UACIL	ACIsL	CP
20, 10	1	1.594526	0.948091	0.458727	-0.263732	3.452785	3.716517	0.930000
20, 15	1	1.556388	0.844726	0.385720	-0.099275	3.212051	3.311326	0.950000
30, 15	1	1.582766	0.812384	0.367365	-0.009506	3.175038	3.184544	0.930000
30, 20	1	1.546824	0.807062	0.381692	-0.035018	3.128666	3.163684	0.970000
30, 25	1	1.189059	0.807045	0.483961	-0.392749	2.770867	3.163616	0.960000
40, 20	1	1.557939	0.806578	0.406983	-0.022955	3.138833	3.161788	0.940000
40, 25	1	1.589465	0.763197	0.372843	0.093600	3.085331	2.991731	0.920000
40, 30	1	1.520138	0.754312	0.365067	0.041687	2.998588	2.956901	0.950000
50, 25	1	1.560208	0.736059	0.399957	0.117533	3.002883	2.885350	0.970000
50, 30	1	1.147427	0.687020	0.415694	-0.199131	2.493985	2.693117	0.970000
50, 35	1	1.581000	0.661877	0.319240	0.283721	2.878278	2.594557	0.950000
60, 35	1	1.466991	0.645759	0.352575	0.201303	2.732678	2.531375	0.940000
60, 40	1	1.582213	0.589180	0.280259	0.427421	2.737005	2.309584	0.950000
60, 45	1	1.350998	0.574095	0.326686	0.225772	2.476224	2.250452	0.970000
20, 10	2	1.496527	0.812236	0.370885	-0.095460	3.088511	3.183967	0.940000
20, 15	2	1.521890	0.781138	0.353235	-0.009140	3.052921	3.062062	0.940000
30, 15	2	1.487331	0.776021	0.367011	-0.033670	3.008333	3.042004	0.950000
30, 20	2	1.143326	0.776005	0.465347	-0.377640	2.664295	3.041939	0.930000
30, 25	2	1.498018	0.775556	0.391330	-0.022070	3.018108	3.040180	0.940000
40, 20	2	1.528332	0.733843	0.358503	0.091020	2.966664	2.876664	0.950000
40, 25	2	1.461671	0.725300	0.351026	0.040084	2.883258	2.843174	0.950000
40, 30	2	1.500200	0.707749	0.384574	0.113013	2.887387	2.774375	0.960000
50, 25	2	1.103295	0.660596	0.399705	-0.191470	2.398063	2.589535	0.960000
50, 30	2	1.520192	0.636420	0.306962	0.272809	2.767575	2.494766	0.960000
50, 35	2	1.410568	0.620922	0.339014	0.193561	2.627575	2.434014	0.950000
60, 35	2	1.521359	0.566519	0.269480	0.410982	2.631736	2.220754	0.950000
60, 40	2	1.299036	0.552014	0.314121	0.217088	2.380985	2.163896	0.960000
60, 45	2	1.429823	0.520777	0.260784	0.409101	2.450546	2.041445	0.970000

Table 3: *MLEs with RMSEs, RABs of* ϑ *&* *LACIL, UACIL, ACIsL, CPs of ACIs with* $(\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5)$

(n_i, m_i)	CS	MLE	RMSE	RAB	LACIL	UACIL	ACIsL	CP
20, 10	1	0.956331	0.146834	0.116788	0.668536	1.244127	0.575591	0.960000
20, 15	1	0.970048	0.144884	0.116523	0.686076	1.254020	0.567944	0.950000
30, 15	1	0.964952	0.142178	0.112322	0.686283	1.243622	0.557339	0.970000
30, 20	1	0.938985	0.138562	0.116477	0.667403	1.210567	0.543164	0.950000
30, 25	1	0.964784	0.127289	0.102683	0.715298	1.214270	0.498972	0.940000
40, 20	1	0.937890	0.120433	0.104640	0.701840	1.173940	0.472099	0.950000
40, 25	1	0.948488	0.119736	0.097341	0.713805	1.183171	0.469366	0.950000
40, 30	1	0.951535	0.117819	0.098871	0.720610	1.182461	0.461852	0.960000
50, 25	1	0.936677	0.115799	0.101152	0.709711	1.163643	0.453932	0.960000
50, 30	1	0.879029	0.112284	0.099044	0.658952	1.099106	0.440154	0.950000
50, 35	1	0.884003	0.110710	0.103224	0.667011	1.100995	0.433983	0.960000
60, 35	1	0.951515	0.102543	0.079229	0.750531	1.152500	0.401969	0.950000
60, 40	1	0.962086	0.094635	0.077911	0.776601	1.147571	0.370970	0.940000
60, 45	1	0.948595	0.092008	0.079129	0.768259	1.128932	0.360673	0.960000
20, 10	2	1.004047	0.169615	0.129299	0.671601	1.336493	0.664892	0.910000
20, 15	2	0.993732	0.160594	0.128606	0.678968	1.308495	0.629528	0.940000
30, 15	2	0.919236	0.159483	0.138557	0.606649	1.231823	0.625174	0.970000
30, 20	2	0.975282	0.152032	0.125374	0.677299	1.273265	0.595966	0.960000
30, 25	2	0.987848	0.149513	0.121256	0.694803	1.280893	0.586090	0.960000
40, 20	2	0.945983	0.138707	0.117505	0.674117	1.217849	0.543732	0.960000
40, 25	2	0.967796	0.136775	0.109025	0.699717	1.235876	0.536159	0.970000
40, 30	2	0.908435	0.136498	0.121973	0.640898	1.175972	0.535073	0.940000
50, 25	2	0.854772	0.134755	0.124194	0.590652	1.118892	0.528240	0.930000
50, 30	2	0.963932	0.132995	0.111667	0.703263	1.224602	0.521339	0.940000
50, 35	2	0.970456	0.129270	0.109791	0.717087	1.223824	0.506737	0.960000
60, 35	2	0.950931	0.121757	0.106660	0.712287	1.189574	0.477287	0.960000
60, 40	2	0.977545	0.116375	0.096155	0.749450	1.205640	0.456190	0.940000
60, 45	2	1.003948	0.101925	0.080405	0.804175	1.203721	0.399546	0.970000

Table 4: *MLEs with RMSEs, RABs of* ϑ *&* *LACIL, UACIL, ACIsL, CPs of ACIs with* $(\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8)$

(n_i, m_i)	CS	MLE	RMSE	RAB	LACIL	UACIL	ACIsL	CP
20, 10	1	0.729710	0.156055	0.171920	0.423842	1.035579	0.611737	0.970000
20, 15	1	0.635912	0.147363	0.168628	0.347080	0.924744	0.577663	0.970000
30, 15	1	0.752896	0.146134	0.156156	0.466473	1.039318	0.572845	0.950000
30, 20	1	0.761058	0.141137	0.146475	0.484429	1.037687	0.553258	0.980000
30, 25	1	0.659879	0.139685	0.167380	0.386096	0.933663	0.547567	0.940000
40, 20	1	0.755815	0.138839	0.148709	0.483691	1.027939	0.544249	0.950000
40, 25	1	0.737196	0.134817	0.147778	0.472956	1.001437	0.528481	0.970000
40, 30	1	0.744293	0.128940	0.144283	0.491570	0.997016	0.505445	0.960000
50, 25	1	0.743150	0.122515	0.130935	0.503021	0.983279	0.480258	0.960000
50, 30	1	0.747007	0.120722	0.130088	0.510392	0.983622	0.473230	0.960000
50, 35	1	0.616800	0.118942	0.144958	0.383673	0.849927	0.466253	0.950000
60, 35	1	0.761865	0.113971	0.122099	0.538481	0.985249	0.446768	0.960000
60, 40	1	0.756613	0.105303	0.110844	0.550219	0.963008	0.412788	0.960000
60, 45	1	0.741342	0.098291	0.106389	0.548692	0.933991	0.385299	0.950000
20, 10	2	0.635545	0.121039	0.186381	0.398309	0.872781	0.474472	0.950000
20, 15	2	0.757236	0.120128	0.156847	0.521786	0.992686	0.470901	0.970000
30, 15	2	0.617029	0.110272	0.174321	0.400897	0.833161	0.432264	0.940000
30, 20	2	0.705092	0.106901	0.148531	0.495566	0.914619	0.419054	0.960000
30, 25	2	0.614613	0.099448	0.154738	0.419694	0.809531	0.389836	0.940000
40, 20	2	0.605555	0.099098	0.157103	0.411322	0.799788	0.388466	0.960000
40, 25	2	0.712096	0.098604	0.106531	0.518833	0.905359	0.386527	0.970000
40, 30	2	0.621327	0.097804	0.154010	0.429632	0.813023	0.383391	0.950000
50, 25	2	0.614977	0.083327	0.128538	0.451657	0.778298	0.326641	0.950000
50, 30	2	0.669440	0.074306	0.107218	0.523801	0.815080	0.291279	0.950000
50, 35	2	0.617426	0.067618	0.102392	0.484895	0.749956	0.265061	0.940000
60, 35	2	0.613535	0.064974	0.104036	0.486187	0.740883	0.254696	0.940000
60, 40	2	0.612077	0.064608	0.106285	0.485446	0.738708	0.253262	0.930000
60, 45	2	0.613641	0.063932	0.104994	0.488334	0.738948	0.250613	0.930000

Table 5: *MLEs with RMSEs, RABs of* ϕ *&* *LACIL, UACIL, ACIsL, CPs of ACIs with* $(\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5)$

(n_i, m_i)	CS	MLE	RMSE	RAB	LACIL	UACIL	ACIsL	CP
20, 10	1	0.356076	0.217232	0.331312	-0.069698	0.781851	0.851548	0.980000
20, 15	1	0.423038	0.170116	0.307633	0.089610	0.756466	0.666856	0.940000
30, 15	1	0.442701	0.103119	0.181128	0.240587	0.644815	0.404228	0.960000
30, 20	1	0.385605	0.096882	0.187231	0.195716	0.575493	0.379777	0.950000
30, 25	1	0.408111	0.095668	0.188336	0.220601	0.595620	0.375019	0.950000
40, 20	1	0.322006	0.092630	0.225161	0.140452	0.503560	0.363108	0.980000
40, 25	1	0.365803	0.080635	0.180866	0.207758	0.523847	0.316089	0.950000
40, 30	1	0.335572	0.078809	0.191962	0.181107	0.490037	0.308930	0.990000
50, 25	1	0.448864	0.077859	0.139534	0.296260	0.601468	0.305208	0.960000
50, 30	1	0.420924	0.064415	0.122822	0.294671	0.547176	0.252505	0.950000
50, 35	1	0.469790	0.062875	0.101525	0.346555	0.593024	0.246469	0.940000
60, 35	1	0.411541	0.053268	0.103181	0.307136	0.515946	0.208809	0.940000
60, 40	1	0.326189	0.046967	0.110134	0.234134	0.418244	0.184110	0.950000
60, 45	1	0.333336	0.035991	0.087729	0.262794	0.403879	0.141085	0.980000
20, 10	2	0.387373	0.124545	0.226787	0.143265	0.631482	0.488217	0.970000
20, 15	2	0.360833	0.105094	0.239976	0.154850	0.566817	0.411967	0.970000
30, 15	2	0.465610	0.104133	0.167382	0.261510	0.669710	0.408200	0.920000
30, 20	2	0.404602	0.101346	0.199987	0.205964	0.603241	0.397277	0.950000
30, 25	2	0.402596	0.094031	0.194124	0.218295	0.586897	0.368602	0.970000
40, 20	2	0.340660	0.093892	0.218835	0.156631	0.524689	0.368058	0.960000
40, 25	2	0.447136	0.078380	0.134325	0.293511	0.600761	0.307250	0.930000
40, 30	2	0.403542	0.077751	0.163517	0.251151	0.555934	0.304783	0.980000
50, 25	2	0.450346	0.073391	0.134952	0.306500	0.594192	0.287693	0.960000
50, 30	2	0.390654	0.061132	0.122158	0.270836	0.510473	0.239636	0.950000
50, 35	2	0.416645	0.059222	0.115242	0.300570	0.532720	0.232150	0.960000
60, 35	2	0.388899	0.050428	0.104176	0.290060	0.487738	0.197678	0.940000
60, 40	2	0.341503	0.037570	0.081691	0.267867	0.415140	0.147273	0.950000
60, 45	2	0.450346	0.030832	0.095481	0.195635	0.316495	0.120860	0.950000

Table 6: *MLEs with RMSEs, RABs of* ϕ *&* *LACIL, UACIL, ACIsL, CPs of ACIs with* $(\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8)$

(n_i, m_i)	CS	MLE	RMSE	RAB	LACIL	UACIL	ACIsL	CP
20, 10	1	0.534378	0.248217	0.339334	0.047871	1.020884	0.973012	0.940000
20, 15	1	0.628356	0.226171	0.291756	0.185060	1.071652	0.886592	0.970000
30, 15	1	0.716714	0.218887	0.253214	0.287696	1.145733	0.858036	0.960000
30, 20	1	0.568052	0.204952	0.291313	0.166346	0.969758	0.803412	0.950000
30, 25	1	0.788660	0.188148	0.175430	0.419890	1.157431	0.737541	0.940000
40, 20	1	0.567105	0.180587	0.266770	0.213155	0.921055	0.707900	0.970000
40, 25	1	0.763205	0.171400	0.166184	0.427261	1.099149	0.671888	0.950000
40, 30	1	0.677318	0.165821	0.198586	0.352309	1.002327	0.650018	0.950000
50, 25	1	0.552460	0.158731	0.226367	0.241347	0.863573	0.622227	0.940000
50, 30	1	0.541943	0.131083	0.189980	0.285021	0.798864	0.513844	0.970000
50, 35	1	0.622779	0.130376	0.170692	0.367242	0.878316	0.511074	0.960000
60, 35	1	0.603203	0.126617	0.150484	0.355034	0.851372	0.496338	0.970000
60, 40	1	0.628403	0.101000	0.126772	0.430444	0.826363	0.395920	0.950000
60, 45	1	0.578427	0.095940	0.128139	0.390385	0.766470	0.376086	0.960000
20, 10	2	0.562503	0.261281	0.357193	0.050391	1.074614	1.024223	0.940000
20, 15	2	0.661427	0.238075	0.307111	0.194800	1.128055	0.933255	0.950000
30, 15	2	0.754436	0.230407	0.266541	0.302838	1.206034	0.903196	0.940000
30, 20	2	0.597950	0.215739	0.306645	0.175101	1.020798	0.845697	0.960000
30, 25	2	0.830169	0.198051	0.184663	0.441990	1.218348	0.776359	0.950000
40, 20	2	0.596952	0.190091	0.280810	0.224373	0.969531	0.745158	0.940000
40, 25	2	0.803374	0.180421	0.174930	0.449749	1.156999	0.707250	0.970000
40, 30	2	0.712967	0.174548	0.209037	0.370852	1.055081	0.684230	0.970000
50, 25	2	0.581537	0.167086	0.238281	0.254049	0.909024	0.654975	0.950000
50, 30	2	0.519055	0.156485	0.257407	0.212344	0.825767	0.613423	0.940000
50, 35	2	0.570466	0.137982	0.199979	0.300022	0.840910	0.540888	0.940000
60, 35	2	0.655557	0.137238	0.179676	0.386571	0.924543	0.537972	0.960000
60, 40	2	0.634950	0.133281	0.158404	0.373720	0.896181	0.522461	0.940000
60, 45	2	0.661477	0.106316	0.133444	0.453098	0.869856	0.416758	0.940000

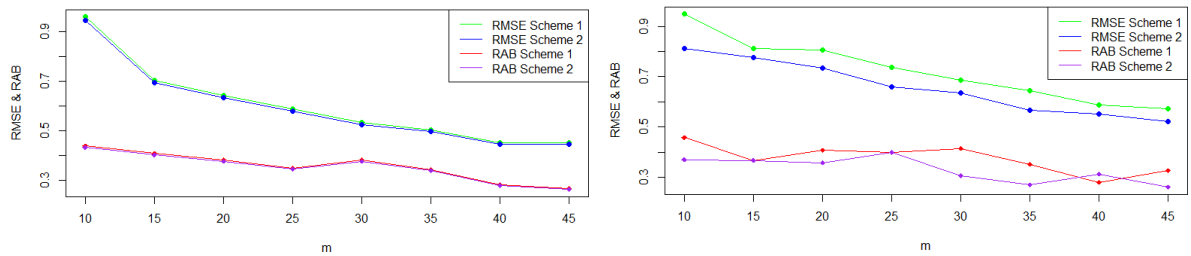


Figure 1: *RMSEs and RABs of the estimates of* ψ_0 *with* $(\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5)$ & $(\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8)$

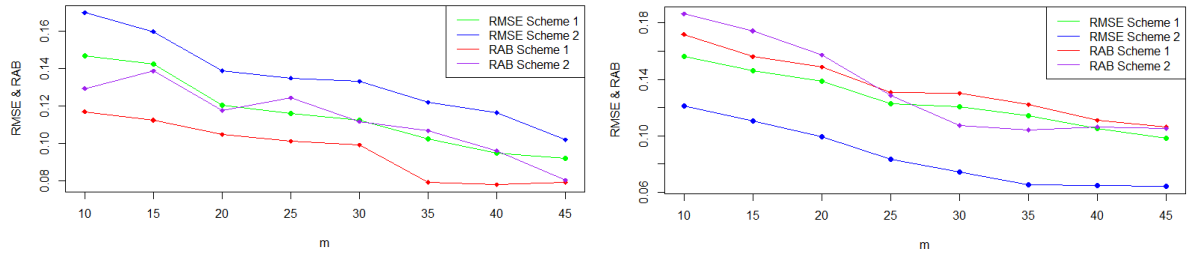


Figure 2: *RMSEs and RABs of the estimates of* ϑ *with* $(\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5)$ & $(\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8)$

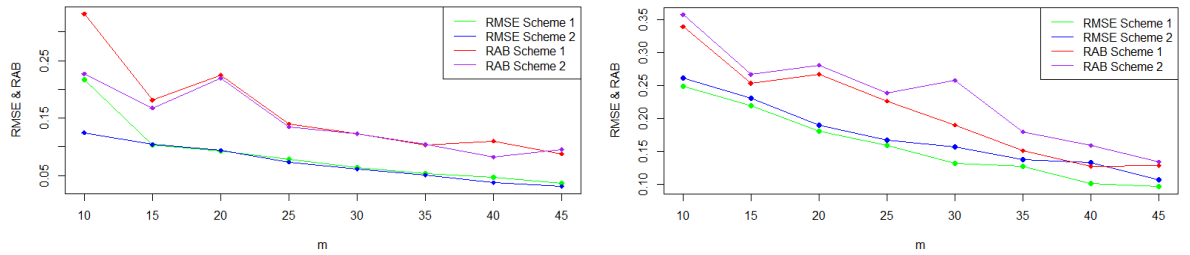


Figure 3: *RMSEs and RABs of the estimates of* ϕ *with* $(\psi_0 = 1.2, \vartheta = 0.75, \phi = 0.5)$ & $(\psi_0 = 1.5, \vartheta = 0.5, \phi = 0.8)$

It is evident from the results in tables 1-6 and Figures 1-3 that the results are consistent and that the estimates are quite closer to the real values of the parameters. The following observations can be made in general:

1. In all situations, the RMSEs and RABs decrease as the values of n_i and m_i increase, which is to be expected because greater samples produce more accurate results.
2. The lengths and CPs of 95% of ACIs are relatively precise in all situations, as shown in table 1-6. However, the ACIs are narrower for parameter setting 1 and removal scheme 2.
3. With increasing values of n_i and m_i, it can also be seen that the lengths of 95% percent ACIs are getting smaller and the CPs are getting larger.

As an outcome of the above observations, it is reasonable to infer that the proposed model and estimation method in this paper performed well, and that all statistical assumptions for fitting the model and estimation are suitable.

V. Conclusions

In this paper, the CSALT model has been considered with k levels of constant stress. The observed TIIPC failure data was assumed to come from a GnP distribution. The distribution's shape parameter has been assumed to be independent of the stress, whereas the scale parameter was assumed to have a log linear relationship with the stress variable. Model parameters are estimated using the MLE approach, and their performance is evaluated using the corresponding RABs and MSEs. The performance of MLEs has been found to be satisfactory, as the estimated values approaching real values as the sample size increases. The ACIs have also been constructed

based on the asymptotic properties of the MLEs. The performance of ACIs was evaluated in terms of their corresponding CPs and ACIIL. An alternative lifetime distribution can be considered in the future research, and the corresponding inferences under various censoring methods can be developed using the BE technique.

References

[1] Nelson, W. (1990). Accelerated Testing: Statistical Models, Test Plans and Data Analysis. John Wiley & Sons, New York.

[2] Yang, G. B. (1994). Optimum constant-stress accelerated life-test plans. *IEEE transactions on reliability*, *43*(4), 575-581.

[3] Watkins, A. J., & 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), 768-786.

[4] Zarrin, S., Kamal, M., & Saxena, S. (2012). Estimation in constant stress partially accelerated life tests for Rayleigh distribution using type-I censoring. *Reliability: Theory & Applications, 7(4)*, 41-52.

[5] Kamal, M. (2013). Application of geometric process in accelerated life testing analysis with type-I censored Weibull failure data. *Reliability: Theory & Applications, 8(3)*, 87-96.

[6] Ma, H., & Meeker, W. Q. (2010). Strategy for planning accelerated life tests with small sample sizes. *IEEE Transactions on Reliability*, *59*(4), 610-619.

[7] Kamal, M., Zarrin, S., & Islam, A. (2014). Design of accelerated life testing using geometric process for type-II censored Pareto failure data., *International Journal of Mathematical Modelling & Computations, 4(2)*, 125-134.

[8] Gao, L., Chen, W., Qian, P., Pan, J., & He, Q. (2016). Optimal time-censored constant-stress ALT plan based on chord of nonlinear stress-life relationship. *IEEE Transactions on Reliability*, *65*(3), 1496-1508.

[9] Han, D., & Bai, T. (2019). On the maximum likelihood estimation for progressively censored lifetimes from constant-stress and step-stress accelerated tests. *Electronic Journal of Applied Statistical Analysis*, *12*(2), 392-404.

[10] Miller, R., & Nelson, W. (1983). Optimum simple step-stress plans for accelerated life testing. *IEEE Transactions on Reliability*, *32*(1), 59-65.

[11] Bai, D. S., Kim, M. S., & Lee, S. H. (1989). Optimum simple step-stress accelerated life tests with censoring. *IEEE transactions on reliability*, *38*(5), 528-532.

[12] Saxena, S., Zarrin, S., Kamal, M., & & Islam, A. U. (2012). Optimum Step Stress Accelerated Life Testing for Rayleigh Distribution. *International journal of statistics and applications, 2(6)*, 120-125.

[13] Kamal, M., Zarrin, S. & Islam, A. (2013). Step stress accelerated life testing plan for two parameter Pareto distribution. *Reliability: Theory & Applications, 8(1)*, 30-40.

[14] Han, D., & Bai, T. (2020). Design optimization of a simple step-stress accelerated life test–Contrast between continuous and interval inspections with non-uniform step durations. *Reliability Engineering & System Safety*, *199*, 106875.

[15] Bai, X., Shi, Y., & Ng, H. K. T. (2020). Statistical inference of Type-I progressively censored step-stress accelerated life test with dependent competing risks. *Communications in Statistics-Theory and Methods*, 1-27.

[16] Kamal, M., Rahman, A., Ansari, S. I., & Zarrin, S. (2020). Statistical Analysis and Optimum Step Stress Accelerated Life Test Design for Nadarajah Haghighi Distribution. *Reliability: Theory & Applications, 15(4)*, 1-9.

[17] Hakamipour, N. (2021). Comparison between constant-stress and step-stress accelerated life tests under a cost constraint for progressive type I censoring. *Sequential Analysis*, *40*(1), 17-31.

[18] Nassar, M., Okasha, H., & Albassam, M. (2021). E-Bayesian estimation and associated properties of simple step–stress model for exponential distribution based on type-II censoring. *Quality and Reliability Engineering International*, *37*(3), 997-1016.

[19] Klemenc, J., & Nagode, M. (2021). Design of step-stress accelerated life tests for estimating the fatigue reliability of structural components based on a finite-element approach. *Fatigue & Fracture of Engineering Materials & Structures*, *44*(6), 1562-1582.

[20] Yin, X. K., & Sheng, B. Z. (1987). Some aspects of accelerated life testing by progressive stress. *IEEE transactions on reliability*, *36*(1), 150-155.

[21] Srivastava, P. W., & Mittal, N. (2012). Optimum multi-objective ramp-stress accelerated life test with stress upper bound for Burr type-XII distribution. *IEEE Transactions on Reliability*, *61*(4), 1030-1038.

[22] Abdel-Hamid, A. H., & AL-Hussaini, E. K. (2014). Bayesian prediction for type-II progressive-censored data from the Rayleigh distribution under progressive-stress model. *Journal of Statistical Computation and Simulation*, *84*(6), 1297-1312.

[23] Chen, Y., Sun, W., & Xu, D. (2017). Multi-stress equivalent optimum design for ramp-stress accelerated life test plans based on D-efficiency. *IEEE Access*, *5*, 25854-25862.

[24] El-Din, M. M., Amein, M. M., Abd El-Raheem, A. M., Hafez, E. H., & Riad, F. H. (2020). Bayesian inference on progressive-stress accelerated life testing for the exponentiated Weibull distribution under progressive type-II censoring. *J. Stat. Appl. Pro. Lett*, *7*, 109-126.

[25] Ma, Z., Liao, H., Ji, H., Wang, S., Yin, F., & Nie, S. (2021). Optimal design of hybrid accelerated test based on the Inverse Gaussian process model. *Reliability Engineering & System Safety*, *210*, 107509.

[26] Fan, T. H., & Hsu, T. M. (2014). Constant stress accelerated life test on a multiple-component series system under Weibull lifetime distributions. *Communications in Statistics-Theory and Methods*, *43*(10-12), 2370-2383.

[27] El-Din, M. M., Abu-Youssef, S. E., Ali, N. S., & Abd El-Raheem, A. M. (2016). Estimation in constant-stress accelerated life tests for extension of the exponential distribution under progressive censoring. *Metron*, *74*(2), 253-273.

[28] Abdel Ghaly, A. A., Aly, H. M., & Salah, R. N. (2016). Different estimation methods for constant stress accelerated life test under the family of the exponentiated distributions. *Quality and Reliability Engineering International*, *32*(3), 1095-1108.

[29] Wang, L. (2017). Inference of constant-stress accelerated life test for a truncated distribution under progressive censoring. *Applied Mathematical Modelling*, *44*, 743-757.

[30] Lin, C. T., Hsu, Y. Y., Lee, S. Y., & Balakrishnan, N. (2019). Inference on constant stress accelerated life tests for log-location-scale lifetime distributions with type-I hybrid censoring. *Journal of Statistical Computation and Simulation*, *89*(4), 720-749.

[31] Cui, W., Yan, Z., & Peng, X. (2019). Statistical analysis for constant-stress accelerated life test with Weibull distribution under adaptive Type-II hybrid censored data. *IEEE Access*, *7*, 165336-165344.

[32] Rahman, A., Sindhu, T. N., Lone, S. A., & 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, 16(3),* 577-586.

[33] Xin, H., Liu, Z., Lio, Y., & Tsai, T. R. (2020). Accelerated Life Test Method for the Doubly Truncated Burr Type XII Distribution. *Mathematics*, *8*(2), 162.

[34] Abd-El-Raheem, A. M., Almetwally, E. M., Mohamed, M., & Hafez, E. H. (2021). Accelerated life tests for modified Kies exponential lifetime distribution: binomial removal, transformers turn insulation application and numerical results. *AIMS Mathematics*, *6*(5), 5222-5255.

[35] Balakrishnan, N., & Sandhu, R. A. (1995). A simple simulation algorithm for generating progressive Type-II censored samples. *The American Statistician, 49(2),* 229-230.

Validation Of DNAFIDs Model Through Finite State Machine

Yogesh Pal
•
Research Scholar, Dept. of Computer Science & Engineering
Maharishi University of Information Technology, Lucknow, INDIA
Er.yogeshpal15@gmail.com

Santosh Kumar
•
Associate Professor, Dept. of Computer Applications
Shri Ramswaroop Memorial University Lucknow, Deva Road, Barabanki, INDIA
Sant7783@hotmail.com

Madhulika Singh
•
Professor, School of Science
Maharishi University of Information Technology, Lucknow, INDIA
Madhulika.anil@gmail.com

Shweta Dwivedi
•
Assistant Professor, Dept. of Computer Science & Engineering
Maharishi University of Information Technology, Lucknow, INDIA
Muit23412@gmail.com

Abstract

The assurance of quality and reliability of process models and workflows is essential for model driven software development. There are numerous ways to achieve these objectives. One is model checking through which it can be verified that a model satisfies specific logical rules. The model to be checked is usually given as finite state machine. Rules have to be specified at the level required by the model checker. In this work, we develop a model for validating the DNA profiling through finite state.. This enables the research/business process professionals to use model checking techniques and to produce higher quality research/business models for subsequent software development. The approach is demonstrated by validating event-driven process chains.

Keywords: DNAFIDs, FSM, Class Diagram, State Diagram, Transition Table

I. Introduction

Model checking permits confirming the grouping of dynamic communications in a model naturally. While the utilization of model checking in equipment related areas is wide-spread and has effectively modern importance, the use of this proper technique in the space of programming items, in any case, is as yet in its beginnings. Albeit model checking is an extremely encouraging method, it has three general issues:

Initially, model checking isn't valuable for the confirmation of a wide range of programming models and code. Particularly at source code level the model development, including information in the state depictions of the state-change framework, prompts the state blast issue. At the plan stage it must be applied for explicit confirmation errands again because of the state blast issue. Furthermore, to utilize model checking, we need to develop the issue to be approved (the conceivable conduct of the framework) in a proper model. This model development issue is considerably more hard to manage programming when contrasted with equipment frameworks. Thirdly, to acquire the advantages of model checking e. g. with regards to business measure displaying, the approval rules must be reasonable for business measure engineers. Right now the fleeting rationale rules must be indicated as text on the low level of the model checker model (rule determination issue).

Numerous ebb and flow research action centres around the decrease of states to keep away from the state blast issue. A few methodologies manage the model development issue, however there has been almost nothing done to tackle the third issue. A first methodology was the meaning of frequently utilized details in property designs. Nonetheless, these examples are as yet text based and fair and square of the model checker model.

A Unified Modeling Language is utilized here for planning the state chart of the proposed DNAFIDs model and this model is approved by carrying out it in Finite State Machine. Consequently UML is a notable displaying language which gives a ton of demonstrating devices and graphical documentations for taking care of complex the item arranged issues in the field of programming. It additionally gives normalization in indicating, recording, composing outline and imagining the ancient rarities of programming escalated framework a work in progress. UML gives a bunch of documentations to portraying the condition of any item through the state talk charts which is perhaps the most flexible apparatuses for depicting the existence pattern of an article from its instatement to end. State outline charts address the powerful conduct of any product framework in graphical structure, which shows every one of the ways through which an item changes its state during as long as its can remember and these ways further graphically addressed by the utilization of the idea of Finite State Machine (FSM).

FSM gives a computational model for dynamic as well as static behaviour of any software system. It is an abstract machine that produces a finite number of states and it produces one state at a time by reading input symbols. The working of FSM is started from the initial state and end on the final state and it can accept any length of string; if an automaton reaches its final state by reading input symbols one by one otherwise it rejects the string. The input is a finite set of alphabets. The finite-state automata can accept or reject an input string.

II. Background

Singh et al [1] have broken down DNA fingerprinting based recognizable proof and planned a DNA fingerprinting based ID model alongside DNA information base administration framework for 360 degree interlinking for example all administrations and advances will be advanced by DNAFIDs and data set. Saxena and Kumar [2] have introduced a way to deal with approve the UML class model through FSM is portrayed with a production of the progress table. Rafi et al [3] have reviewed for Interlinking of DNA Models with Aadhaar Real-Time Records for Enhanced Authentication. Chaturvedi [4] has examined carries out and future works in bioinformatics with Hadoop and furthermore considered the MapReduce calculation from calculation lay by point and exhibit the appropriates of our methodology by following and breaking down productive MapReduce calculations for arranging and recreation issue of equal calculations indicated in the assistance of categorize rule. Singh and Sharma [5] have explored DNA based cryptography for information covering up. Mishra [6] has presented an AAdhar based smartcard framework which will help the South Asian nations in emerging from defilements and working on their economies. O'Keefe et al [7] have introduced a microfluidic stage for atom by-particle recognition of heterogeneous epigenetic examples of uncommon tumor-inferred DNA by exceptionally parallelized computerized liquefy appraisal. Baans and Jambek [8] possess investigated the computational energy for this microarray picture handling. The outcomes show that the force extraction burns-through larger part of the generally computational time. Padmavathi et al [9] have proposed robotization in apportion appropriation utilizing brilliant card dependent on Aadhar card innovation. In this framework, they utilized a model dependent on ATM machine. Mhamane and Shriram [10] have proposed ticket checking is managed without human mediation. Prakasha et al [11] have meant to conquer this downside of manual distinguishing proof and verification of client and accomplish client ID and confirmation through a robotized technique utilizing the Aadhar card. Aadhar project is created by the Unique Identification Authority of India by consolidating biometrics and digitization. Vishal et al [12] have managed the web based democratic framework that will make the democratic framework keen. OVS(online casting a ballot system)is got and it have straightforward plan.

III. Methodology

I. UML Class Diagram and Sequence Diagram For DNAFIDs

a) UML Class Diagram

The UML Class diagram for DNA Profiling or DNAFIDs is presented here. There are ten major classes with their attributes are represented in figure 1.1which is developed by Singh et al [1]. The model shows the complete process of DNA profiling. The Information_Cofigurationhas multiple associations with the Sample_Collection class because it collects the many DNA samples from the information_configuration class whilesample_collection class has single associate with the Audit_database class that has two types of databases like sample fingerprinting database and local fingerprinting database, both these database and the Audit_Database class is directly connected the main DNA Fingerprinting database to access all the audit reports regarding both databases. The SSR_Analysis class is directly connected to the DNA Fingerprinting Database to store and fetch the analysis data for Fingerprinting _Web Services. The Gene_Mapper class, Gene_Excel and Gene_Zip_Packageare the main classes that involved in experiment of DNA fingerprinting are directly connected to the DNA Fingerprinting Database and Experimental Fingerprinting Database because the result of experimental DNA fingerprinting is stored in it after developing the DNA fingerprinting that is accessed by the main DNA fingerprinting database.

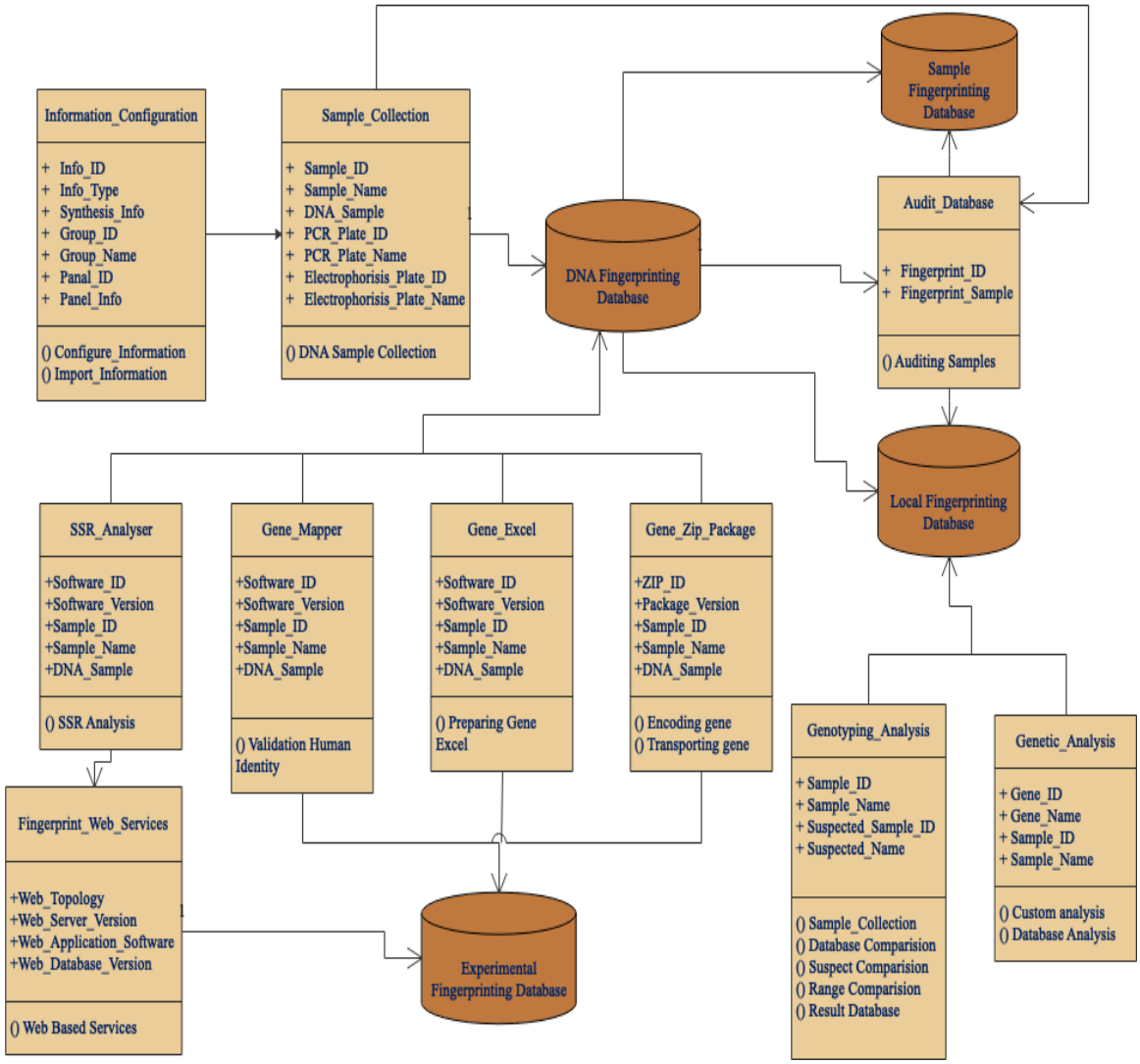


Figure 1.1: UML *Class Diagram for DNA Profiling/DNAFIDs*

b) Sequence Diagram

A sequence diagram is designed here to represent the dynamic behavior of the DNA profiling or DNA fingerprinting from the figure 1.2 it shows the complete process of DNA profiling where six major objects like Sample_Collection, DNA_Isolation, PCR, Gel_Electrophoresis, Blotting and Sanger_Sequencing. The objects are communicated between each other through the massages that is shown along with the solid arrows while the reply message shown by the dotted arrow and the life line of the object is shown by the vertical dotted lines. Therefore itis shown in the sequence diagram that the DNA sample is collected in DNA_Sample collection center and then it is send to the DNA_Isolation point where the DNA is extracted by including some solution in it. After extracting the DNA it is send to the PCR point where the DNA is fragmented and transferred to Electrophoresis station where DNA is separated through the agarose gel during the electrophoresis technique.As the DNA fragment is separated the based pattern is transformed to nylon membrane through southern blotting technique. The radioactive DNA probes binds to specific DNA sequences on the membrane for DNA_Sequencing. As the DNA sequencing is prepared an x-ray film is developed to make visible the DNA pattern after detecting the radioactive pattern, this is known as DNA profiling or DNA fingerprinting.

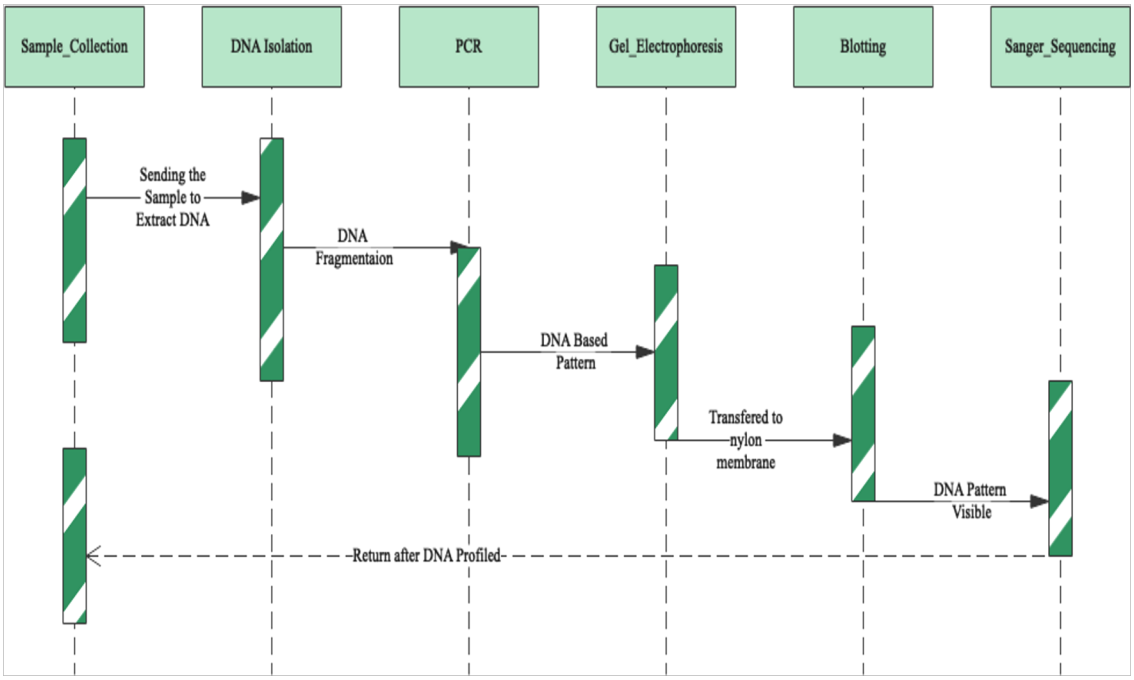


Figure 1.2: UML *Sequence Diagram for DNA Profiling/DNAFIDs*

II. State Transition Diagram For DNAFIDs and Convection Into Finite State Machine

The state transition diagram is illustrated for DNA profiling which is shown below in the figure 2.1:

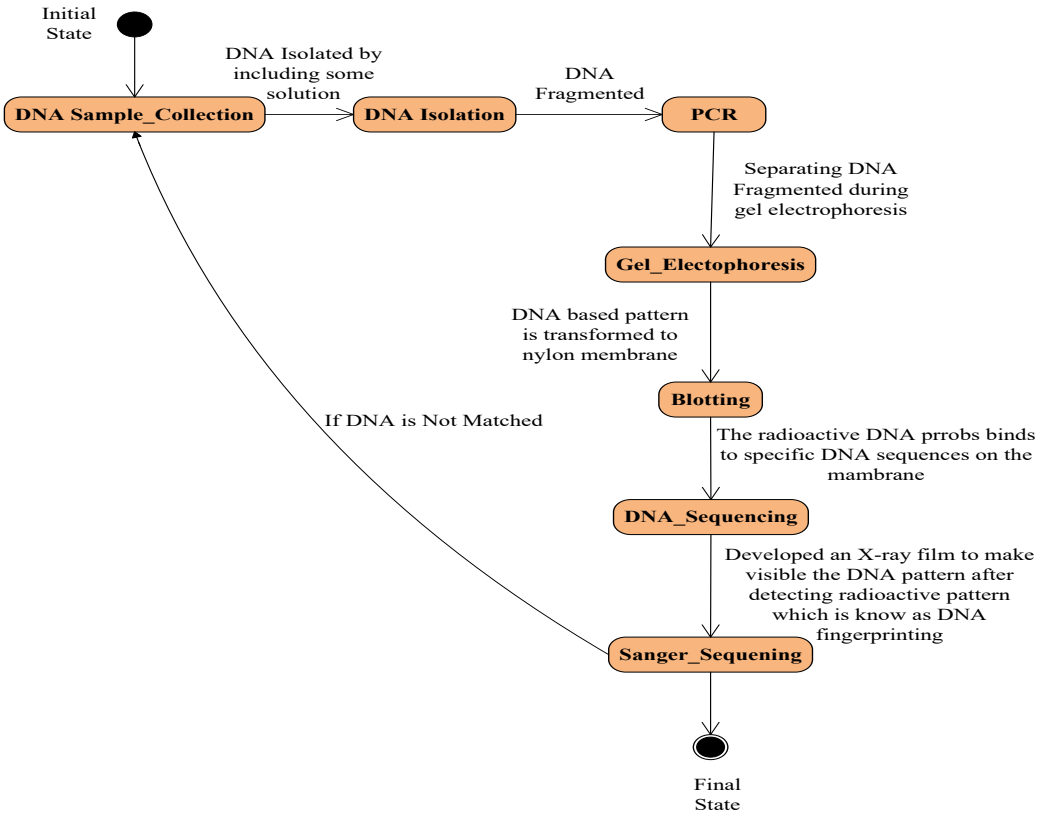


Figure 2.1: State Transition Diagram for DNA Profiling

From the above state transition diagram, it is assumed that the DNA sample is collected at the state of DNA Sample_Colletion it is the initial state which is equivalent to "q_0" and the sample send to the DNA_Isolation state it is equivalent to "q_1" where the DNA is separated by including some separation gel say "a". At the PCR state which is equivalent to "$q_{2''}$ the DNA is Fragmented say "b" and this fragmented DNA is send to the Gel_Eletrophoresis state where DNA is Separated from the fragment though the electrophoresis gel say "c" this state is equivalent to "q_3". After separating the DNA from the fragments a new state is appear which is equivalent to "q_4" where a DNA based pattern is transformed to the nylon membrane say "d" named as Blotting. At the DNA-Sequencing state ("q_5") the radioactive DNA probs binds to specific DNA sequences on the membrane say "e" and a DNA sequence is found. An x-ray film is developed to make visible the DNA pattern which is known a DNA fingerprinting say "f" at the final state Sanger_Sequening it is equivalent to "q_6". If the DNA matches the process is in final state i.e. "q_6" and if the DNA is not matches then the process went to initial state i.e. "q_0".

The finite state machines for DNA profiling through the set of these states equivalencies can be drawn as shown in the figure 2.1. The equivalent finite state machine of the above UML state diagram is as shown in figure 2.2:

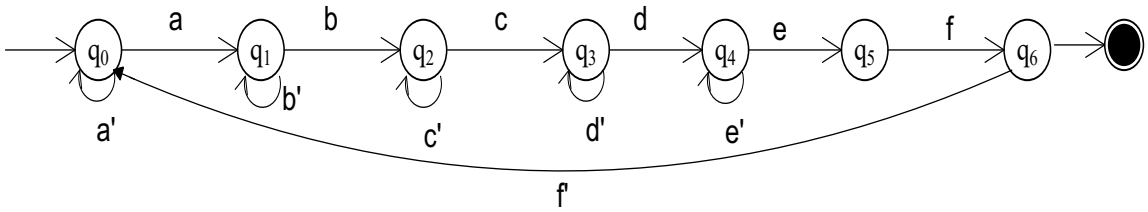


Figure 2.2: UML *Finite State Machine for DNA Profiling/DNAFIDs*

The transformation of one state to another state is done on the basis of {a, a', b, b', c, c', d, d', e, e', f, f'} inputs which is considered as terminals and the set of states {q_0, q_1, q_2, q_3, q_4, q_5, q_6 }are the non-terminals states where q_0 is the initial state and q_6 is the final state. There are several production can be induced for the above finite state machine and the corresponding transition table is as shown below in table 2.2:

Table 1: *Transition Table for DNA Profiling/DNAFIDs*

	Inputs										
States	'a'	a'	'b'	b'	'c'	c'	'd'	d'	'e'	'f'	f'
→q0	q1	q0	-	-	-	-	-	-	-	-	-
q1	-	-	q2	q1	-	-	-	-	-	-	-
q2	-	-	-	-	q3	q2	-	-	-	-	-
q3	-	-	-	-	-	-	q4	q3	-	-	-
q4	-	-	-	-	-	-	-	-	q5	q4	-
q5	-	-	-	-	-	-	-	-	-	q6	-
q6	-	-	-	-	-	-	-	-	-	-	q0

III. Test Cases For Validating DNAFIDs Model

From the above work there are some test cases are generated to validate the designed model and described below in brief:

Test case 1:

The DNA is fragmented after isolation by including some solution in collected sample.

$\rightarrow q_0 \rightarrow aq_1$

$q_1 \rightarrow bq_2$

Test case 2:

The radioactive DNA probes binds the specific DNA sequence on Membrane by DNA pattern is transformed to nylon membrane.

$q_2 \rightarrow cq_3$

$q_3 \rightarrow dq_4$

$q_4 \rightarrow eq_5$

Test case 3:

An X-ray film is made to visible the DNA pattern and matched it, if the DNA matched the final state occurs if not then initial state occurs.

$q_5 \rightarrow fq_6$

$q_6 \rightarrow f'q_0$

IV. Result & Discussion

From the above work, it is concluded that UML is a powerful modelling language for modelling the various kinds of the research problems and one can depict the static as well as the dynamic behaviour of the system. The above work is based upon the of validation technique through FSM for the designed DNAFIDs model which show the complete process of DNA profiling. The proposed model for DNA profiling/DNAFIDs is validated through various test cases drawn from the FSM.

References

[1] Singh P.Y., Kumar S. and Singh M., "Analysis and Designing A DNA Fingerprinting Based Identifications (DNAFIDs) Model and Database Management System", Reliability Theory & Applications, Vol. 16 No. 1(61), 2021.

[2] Saxena V. and Kumar S., "Validation of UML Class Model through Finite State Machine", International Journal of Computer Applications, Volume 41– No.19, March 2012.

[3] S. Mohammad Rafi, N. Phani Kumar and D. Jeevan Kumar, "Survey for Interlinking of DNA Models with Aadhaar Real-Time Records for Enhanced Authentication", 5th International Conference on Advanced Computing & Communication Systems (ICACCS), doi:10.1109/ICACCS.2019.8728512, pp. 208-211, 2019.

[4] Chaturvedi P., " A design of approximation algorithm for efficient DNA mapping using Hadoop technology", International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), doi:10.1109/ICACCCT.2016.7831726, pp. 681-684, 2016.

[5] Singh S. and Sharma Y., "A Review on DNA based Cryptography for Data hiding", International Conference on Intelligent Sustainable Systems (ICISS), doi:10.1109/ISS1.2019.8908026 , pp. 282-285, 2019.

[6] Mishra K. N., "AAdhar based smartcard system for security management in South Asia", International Conference on Control, Computing, Communication and Materials (ICCCCM), doi:10.1109/ICCCCM.2016.7918256, pp. 1-6, 2016.

[7] O'Keefe C. M., Li S., and Wang T. J., "Digital DNA Sequence Profiling Of Rare Epigenetic Cancer Biomarkers In A Highly Parallelized Microfluidic Platform", 20th International Conference on Solid-State Sensors, Actuators and Microsystems & Eurosensors XXXIII (TRANSDUCERS & EUROSENSORS XXXIII), 10.1109/TRANSDUCERS.2019.8808784, pp. 2-5, 2019.

[8] Baans S. O. and Jambek A. B., "Software Profiling Analysis for DNA Microarray Image Processing Algorithm", IEEE International Conference on Signal and Image Processing Applications (IEEE ICSIPA 2017), doi:10.1109/ICSIPA.2017.8120592, pp. 129–132, Malaysia, September 12-14, 2017.

[9] Padmavathi. R , Azeezulla K.M M., Venkatesh P. , Mahato K. K. and Nithin G., "Digitalized Aadhar Enabled Ration Distribution Using Smart Card", IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), doi:10.1109/RTEICT.2017.8256670, pp. 615–618, May 19-20, 2017, India.

[10] Mhamane S. and Shriram P., "Railway Ticket Verification and Dynamic seat Allocation using Aadhar Card", IEEE 3rd International Conference on Inventive Computation Technologies (ICICT) - Coimbatore, India, doi:10.1109/icict43934.2018.9034437, pp.293-296, 2018.

[11] Prakasha K., Muniyal B. and Acharya V., "Automated User Authentication in Wireless Public Key Infrastructure for Mobile Devices using Aadhar card", IEEE Access, vol. 7, pp. 17981 – 18007, doi:10.1109/ACCESS.2019.2896324, 2019.

[12] Vishal, Chinmay V., Garg R. and Yadav P., "Online voting system linked with AADHAR",3rd International Conference on Computing for Sustainable Global Development (INDIACom), https://ieeexplore.ieee.org/document/7724864, 16-18 March 2016.

Design of One-sided Modified *S* Control Charts for Monitoring a Finite Horizon Process

Mei Tuan Teng[1], Sin Yin Teh[2], Khai Wah Khaw[3*], XinYing Chew[4], Wai Chung Yeong[5]

•

[1,2,3]School of Management, Universiti Sains Malaysia, 11800 Penang, Malaysia
[4]School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
[5]School of Mathematical Sciences, Sunway University, 47500 Petaling Jaya, Malaysia
[1] teng.meituan@gmail.com, [2]tehsyin@usm.my, [3]khaiwah@usm.my, [4]xinying@usm.my, [5]waichungy@sunway.edu.my
*Corresponding Author

Abstract

Control charting techniques are widely used in the manufacturing industry. One of the common charts that are used to monitor process variability is the S control chart. Finite horizon process monitoring has received great attention in the last decade. In the current literature, no attempt has been made to monitor the process variability in a finite horizon process. To fill this gap in research, this paper proposes two one-sided modified S charts for monitoring the standard deviation in a finite horizon process. The performance of the proposed charts is evaluated in terms of the truncated average run length and truncated standard deviation of the run-length criteria. The numerical performances of the proposed charts are shown with the selection of numerous process shifts. The effect of the sample sizes, the number of inspections and the process shifts are studied.

Keywords: Control charting technique, finite horizon process, process variability, statistical quality control, truncated average run length

I. Introduction

Control charting techniques are common-used techniques in the process of signal detection to improve the quality of process monitoring [1 – 2]. Generally, two types of variables control charts are commonly implemented in-process monitoring, i.e. (1) process mean and (2) process standard deviation. Shewhart $\bar{X}$ was introduced for monitoring the process mean whereas *R* and *S* charts are used to monitor the process variability of the quality characteristic. The statistical control charts are widely used in the manufacturing and service industries to monitor both products and processes. For instance, Li et al. [3] implemented the multivariate weighted Poisson chart to monitor the two-dimensional telecommunications data. Chew et al. [4] considered the multivariate variable parameter coefficient of variation chart for monitoring the spring manufacturing process. The yoghurt cup filling process was monitored by Shongwe et al. [5] using the run rules chart. Chew et al. [6] applied the run rules chart for the steel sleeve measurement while Castagliola et al. [7] utilized the variable sample size chart for monitoring the die casting hot chamber process.

Numerous charts were proposed for monitoring the process variance in the last few decades.

Page [8] proposed a one-sided cumulative sum (CUSUM) *S* chart based on the subgroup range. An exponentially weighted moving average (EWMA) *S* chart was then developed by Crowder and Hamilton [9] based on the logarithmic transformation of sample variance. Shu and Jiang [10] discussed a new EWMA chart to monitor the process standard deviation, where the latter chart's performances are better than the former chart. Klein [11] modified the conventional *S* chart by approaching the equal and unequal tail chi-square distribution probabilities. The modified *S* chart from Khoo [12] can circumvent the drawbacks of the conventional *S* chart by adjusting the type I error. Rakitzis and Antzoulakos [13] improved the *S* chart by varying the sample size and sampling interval (VSSI). This chart was then surpassed by the new proposed VSSI$_t$ chart with three sample intervals [14]. Kuo and Lee [15] designed the *S* chart with one of the best adaptive strategy, i.e. variable parameter strategy while Adeoti and Olaomi [16] investigated a moving average *S* chart. Moreover, Abujiya et al. [17] and Costa and Neto [18] suggested a new combined Shewhart-CUSUM *S* and variable charting statistics *S* charts, respectively.

Most of the traditional control charting techniques are implemented for a mass production process, which means its lot sizes are large. This process can be called an infinite horizon process. When dealing with low-volume and high-variety production, a finite horizon process arises. The finite horizon process has received great attention in the last decade as many companies have become more flexible and specialized in their products and services, based on the frequent changes in demand. A well-known example of the finite horizon process is the Just-in-Time manufacturing setting, which emphasizes the minimization of surplus inventory, time of waiting and costs of overproduction. Numerous research works on a finite horizon process were extended to a wide variety of control charts. For instance, CUSUM and variable sampling interval charts for monitoring the process mean in a finite horizon process were discussed by Nenes and Tagaras [19] and Nenes et al. [20], respectively. The Shewhart *t* and the EWMA *t* charts for a finite horizon process were suggested by Celano et al. [21] to address the problems of the estimation error of the process standard deviation due to the availability of limited reliable historical data. According to Amdouni et al. [22], their proposed variable sampling interval coefficient of variation short-run chart has better performance than the standard coefficient of variation short-run chart of Castagliola et al. [23]. More recently, Chong et al. [24] and Chew et al. [25] recommended a variable sample size Hotelling's T^2 and run rules T^2 charts for monitoring the multivariate finite horizon process.

Tuprah and Ncube [26] indicated that the *S* chart has better performance than the *R* chart, for the detection of small to moderate shifts in the standard deviation, when the sample sizes increase. Thus, monitoring the process standard deviation in a finite horizon process is considered very important in Statistical Process Control. However, none of the studies is available on the *S* chart for a finite horizon process in the existing literature. This makes it difficult for quality engineers who wish to monitor the process standard deviation in a finite horizon process. This paper aims to fill the gap in research by proposing the modified *S* charts in a finite horizon process. The numerical performance of the proposed charts will be measured in terms of the truncated average run length (TARL) and truncated standard deviation of the run length (TSDRL) criteria, for monitoring both the upward and downward shifts. The remainder of the sections are organized as follows: Section 2 shows the properties of the classical *S* chart. Section 3 discusses the design of the two one-sided modified *S* charts for monitoring a finite horizon process. The derivations of the formulae and algorithms to compute the TARL and TSDRL values are illustrated in Section 4. Statistical performances of the proposed charts are enumerated in Section 5. Lastly, the research findings and suggestions for future research are shown in the last section.

II. Methods

I. Properties of Classical *S* Chart

Montgomery [27] introduced the process variability that can be monitored with the sample standard deviation of the classical *S* chart, which is based on the ±3σ limits underlying normal distribution.

Let $\{X_{i1}, X_{i2}, \ldots, X_{in}\}$, for $i = 1, 2, \ldots, I$, be a sample of n independent random variables, having a normal $N\left(\mu, \sigma_0^2\right)$ distribution, where μ is the process mean and σ_0^2 is the nominal process variance. When the standard deviation, σ of a process, is known, the upper control limit (UCL), centerline (CL) and lower control limit (LCL) can be computed as [27]:

$$\mathrm{UCL} = c_4\sigma + 3\sigma\sqrt{1-c_4^2} \tag{1}$$

$$\mathrm{CL} = c_4\sigma \tag{2}$$

and

$$\mathrm{LCL} = c_4\sigma - 3\sigma\sqrt{1-c_4^2}, \tag{3}$$

where c_4 is a constant that depends on the sample size n, which can be obtained from Montgomery [27].

If σ is unknown, then it can be estimated through analyzing past data. Assume that there are m preliminary samples, each of size n and let S_i be the standard deviation of the i^{th} sample. The average sample standard deviation is

$$\overline{S} = \frac{1}{m}\sum_{i=1}^{m} S_i, i = 1, 2, \ldots, m \tag{4}$$

The statistics $\dfrac{\overline{S}}{c_4}$ is an unbiased estimator of σ. Thus, the UCL, CL and LCL of the S chart can be obtained as

$$\mathrm{UCL} = \overline{S} + 3\frac{\overline{S}}{c_4}\sqrt{1-c_4^2} \tag{5}$$

$$\mathrm{CL} = \overline{S} \tag{6}$$

$$\mathrm{LCL} = \overline{S} - 3\frac{\overline{S}}{c_4}\sqrt{1-c_4^2} \tag{7}$$

Subsequently, the sample standard deviation of the S chart can be denoted as

$$S_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_{ij} - \overline{X}_i\right)^2} \tag{8}$$

where $\overline{X}_i$ is the mean of sample i. An out-of-control signal is indicated when S_i is being plotted outside the UCL or LCL.

II. The One-Sided Modified *S* Charts in a Finite Horizon Process

An industrial finite horizon process is scheduled to generate a small lot of N parts with finite length H, where I refers to the number of scheduled inspections within H. Subsequently, the sampling frequency between two consecutive inspections is denoted as $h = H/(I + 1)$ hours due to absence of inspection at the end of the process. With the subgroup $\{X_{i1}, X_{i2}, \ldots, X_{in}\}$ defined in Section 2, the UCL and LCL of the one-sided modified S control charts with in-control ARL (ARL_0) equal to $\dfrac{1}{\theta}$ are given as [11]

$$\mathrm{UCL} = \sqrt{\frac{\sigma^2\left(\chi^2_{n-1,\theta}\right)}{n-1}} \tag{9}$$

and

$$\mathrm{LCL} = \sqrt{\frac{\sigma^2\left(\chi^2{}_{n-1,1-\theta}\right)}{n-1}}, \tag{10}$$

where $\chi^2_{n-1,\theta}$ and $\chi^2_{n-1,1-\theta}$ denote the $100\theta^{\text{th}}$ percentile of the chi-square distribution with n - 1 degrees of freedom and σ^2 is the process variance. Note that $0 < \theta < 1$ is selected to satisfy the desired in-control TARL (TARL_0) value. Additionally, if the computed LCL value is negative, then the value is rounded up to zero since $S_i > 0$.

III. Performance Measures of the Modified *S* Charts in a Finite Horizon Process

The statistical performance of the control chart is measured by average run length (ARL) and standard deviation of the run length (SDRL) criteria in an infinite horizon process monitoring. For monitoring a finite horizon process, the ARL criterion is replaced by TARL and TSDRL criteria. This is because the chart's performance measure must be a function of the finite number I of scheduled inspections. TARL can be denoted as the average number of plotted samples in the chart up till a signal is given or up till the process is completed, whichever occurs first. According to Nenes and Tagaras [19], the TARL value equals $I + 1$ if the production run is completed without detecting any signal in the I inspections. The TARL and TSDRL values of the modified *S* chart are given as

$$\mathrm{TARL} = (1-\beta)\sum_{l=1}^{I} l\beta^{l-1} + (I+1)\beta^I = \frac{1-\beta^{I+1}}{1-\beta} \tag{11}$$

and

$$\mathrm{TSDRL} = \frac{\sqrt{\beta\left(1-\beta^{2I+1}\right) - (1-\beta)\beta^{I+1}(1+2I)}}{1-\beta}, \tag{12}$$

respectively, where β represents the probability of Type-II error of the chart and its value can be obtained as $\beta = 1 - F_{\chi^2}(\mathrm{LCL})$ for the downward case and $\beta = F_{\chi^2}(\mathrm{UCL})$ for the upward case. Here, $F_{\chi^2}(\cdot)$ is the cumulative distribution function (cdf) of a non-central chi-square random variable.

III. Results

Tables 1 - 4 display the TARL_1 and TSDRL_1 values for the upward and downward modified *S* charts in a finite horizon process, for monitoring the upward and downward shifts, when sample size $n \in$ {5, 7, 10, 15}, $I \in$ {10, 20, 30, 40, 50}, $\delta \in$ {1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0} (for the upward case) and $\delta \in$ {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9} (for the downward case). Note that when $\delta = 1.0$ indicates the process is in-control, where the $\mathrm{TARL}_0 = I$. The results show that when the θ value, which is used to compute the UCL and LCL, is decreasing when the I value increases. For example, in Table 1, θ = 0.0193, 0.0050, 0.0022, 0.0013 and 0.0008 when I = 10, 20, 30, 40 and 50. In addition, another notable trend observed is that the larger shift δ provides smaller TARL_1 and TSDRL_1 values regardless of the I and n values. An example is shown for monitoring the upward shifts, in Tables 1 and 2, for $n \in$ {5, 7, 10, 15} and I = 10, where the $\mathrm{TARL}_1 \in$ {(8.81, 8.56, 8.23, 7.77), (3.64, 2.85, 2.19, 1.65), (1.76, 1.42, 1.19, 1.06)} and $\mathrm{TSDRL}_1 \in$ {(3.34, 3.45, 3.56, 3.68), (2.79, 2.20, 1.60, 1.03), (1.15, 0.77, 0.48, 0.25)}, when $\delta \in$ {1.1, 1.5, 2.0}. Another example is shown for monitoring the downward shifts, in Tables 3 and 4, for $I \in$ {10, 20, 30, 40, 50} and n = 10, the $\mathrm{TARL}_1 \in$ {(1.53, 2.89, 4.50, 6.30, 8.76), (5.01, 12.02, 20.21, 28.47, 37.79), (9.06, 18.85, 28.79, 38.64, 48.70)} and $\mathrm{TSDRL}_1 \in$ {(0.91, 2.25, 3.95, 5.74, 8.15), (3.46, 7.35, 10.91, 14.23, 17.11), (3.20, 5.03, 6.37, 7.67, 8,54)}, when $\delta \in$ {0.5, 0.7, 0.9}. When n and δ values are fixed, the TARL_1 and TSDRL_1 values increase consistently by increasing the I value. For

example, in Tables 1 and 2, for $n = 10$, $\delta = 1.3$, the $TARL_1 \in \{4.10, 7.57, 11.14, 14.40, 18.16\}$ and $TSDRL_1 \in \{3.05, 6.02, 9.03, 11.88, 14.98\}$, when $I \in \{10, 20, 30, 40, 50\}$.When I and δ values are fixed while n value increases, the $TARL_1$ value decreases. For example, in Table 3, for $I = 30$ and $\delta = 0.8$, $TARL_1 \in \{28.68, 27.73, 26.03, 22.70\}$ when $n \in \{5, 7, 10, 15\}$. Figures 1 and 2 present the graphical view of $TARL_1$ values for the upward and downward modified *S* charts.

IV. Conclusion

In the existing literature, no attempt has been made to monitor the process variability in a finite horizon process. This paper proposes the one-sided upward and downward modified *S* charts for monitoring a finite horizon process. The performance of the proposed charts is evaluated in terms of the $TARL_1$ and $TSDRL_1$ criteria. The formulas of control limits for the modified *S* charts, $TARL_1$ and $TSDRL_1$ are discussed. Different parameter combinations, in terms of the sample size, the number of inspections and process shifts are applied to the proposed charts, for both the upward and downward cases. The results showed that the sample size and the number of inspections affected the $TARL_1$ and $TSDRL_1$ values, for monitoring the different upward and downward process shifts in a finite horizon process. Additionally, the two one-sided modified *S* charts can provide unbiased performance, in terms of $TARL_1$ and $TSDRL_1$ criteria. In future research, the study of proposed charts can be extended in the case of estimated parameters.

Funding

This work was funded by the Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme [Grant Number 203.PMGT.6711755], for the Project entitled "A New Hybrid Model for Monitoring the Multivariate Coefficient of Variation in Healthcare Surveillance".

Declaration of Conflict Interests

The Author(s) declare(s) that there is no conflict of interest.

Table 1. TARL1 values of the upward modified *S* chart when $I \in \{10, 20, 30, 40, 50\}$, $n \in \{5, 7, 10, 15\}$ and $\delta \in \{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$

			δ										
θ	I	n	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.0193	10	5	10.00	8.81	7.31	5.82	4.58	3.64	2.98	2.51	2.18	1.94	1.76
		7	10.00	8.56	6.71	5.00	3.71	2.85	2.30	1.94	1.70	1.54	1.42
		10	10.00	8.23	5.98	4.10	2.89	2.19	1.78	1.53	1.37	1.26	1.19
		15	10.00	7.77	5.03	3.12	2.14	1.65	1.39	1.24	1.15	1.09	1.06
0.0050	20	5	19.98	18.06	15.00	11.49	8.41	6.15	4.64	3.66	3.01	2.56	2.24
		7	19.98	17.64	13.77	9.64	6.48	4.50	3.35	2.65	2.20	1.90	1.69
		10	19.98	17.08	12.19	7.57	4.70	3.20	2.39	1.93	1.64	1.46	1.33
		15	19.98	16.22	10.03	5.37	3.18	2.19	1.70	1.43	1.27	1.17	1.11
0.0022	30	5	30.00	27.58	23.10	17.44	12.25	8.50	6.11	4.63	3.68	3.05	2.61
		7	30.00	27.04	21.27	14.48	9.18	6.00	4.24	3.22	2.59	2.18	1.90
		10	30.00	26.28	18.83	11.14	6.41	4.08	2.91	2.25	1.86	1.61	1.44
		15	30.00	25.12	15.39	7.59	4.11	2.66	1.96	1.59	1.37	1.24	1.16
0.0013	40	5	39.95	37.03	31.08	23.13	15.74	10.51	7.30	5.39	4.20	3.42	2.89
		7	39.95	36.35	28.63	19.03	11.55	7.24	4.95	3.67	2.89	2.39	2.06
		10	39.95	35.41	25.32	14.40	7.85	4.79	3.31	2.50	2.02	1.72	1.52
		15	39.95	33.93	20.55	9.54	4.88	3.02	2.16	1.70	1.44	1.29	1.19
0.0008	50	5	49.99	46.79	39.66	29.47	19.68	12.75	8.61	6.21	4.75	3.81	3.17
		7	49.99	46.03	36.67	24.20	14.25	8.62	5.72	4.14	3.20	2.61	2.21
		10	49.99	44.97	32.54	18.16	9.49	5.58	3.73	2.75	2.19	1.83	1.60
		15	49.99	43.27	26.46	11.80	5.73	3.41	2.36	1.82	1.52	1.34	1.22

Table 2. TSDRL1 values of the upward modified *S* chart when $I \in \{10, 20, 30, 40, 50\}$, $n \in \{5, 7, 10, 15\}$ and $\delta \in \{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$

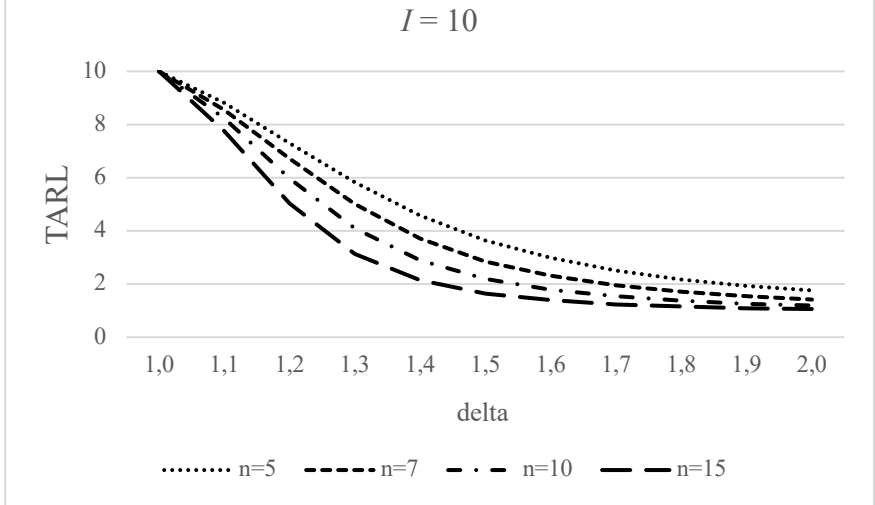
			δ										
θ	I	n	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0.0193	10	5	2.48	3.34	3.75	3.68	3.29	2.79	2.30	1.90	1.59	1.34	1.15
		7	2.48	3.45	3.77	3.46	2.83	2.20	1.71	1.35	1.09	0.91	0.77
		10	2.48	3.56	3.71	3.05	2.23	1.60	1.18	0.90	0.71	0.58	0.48
		15	2.48	3.68	3.47	2.42	1.55	1.03	0.73	0.54	0.41	0.32	0.25
0.0050	20	5	3.60	5.70	7.12	7.29	6.43	5.16	3.99	3.09	2.45	2.00	1.67
		7	3.60	6.00	7.33	6.88	5.38	3.87	2.79	2.08	1.62	1.31	1.08
		10	3.60	6.33	7.36	6.02	4.04	2.64	1.83	1.34	1.03	0.82	0.67
		15	3.60	6.73	7.00	4.59	2.62	1.62	1.09	0.79	0.59	0.45	0.36
0.0022	30	5	4.41	7.69	10.31	10.92	9.56	7.41	5.47	4.08	3.14	2.50	2.05
		7	4.41	8.17	10.77	10.36	7.88	5.38	3.70	2.67	2.03	1.61	1.31
		10	4.41	8.74	10.98	9.03	5.74	3.54	2.35	1.68	1.26	0.99	0.80
		15	4.41	9.45	10.60	6.72	3.57	2.10	1.37	0.96	0.71	0.54	0.43
0.0013	40	5	5.23	9.66	13.49	14.51	12.54	9.39	6.69	4.85	3.67	2.88	2.34
		7	5.23	10.32	14.19	13.75	10.15	6.64	4.42	3.13	2.34	1.83	1.47
		10	5.23	11.11	14.58	11.88	7.20	4.26	2.76	1.93	1.44	1.11	0.89
		15	5.23	12.12	14.13	8.64	4.35	2.47	1.58	1.09	0.80	0.61	0.47
0.0008	50	5	5.74	11.23	16.42	18.15	15.71	11.54	8.01	5.68	4.22	3.27	2.63
		7	5.74	12.06	17.44	17.31	12.63	8.02	5.20	3.61	2.66	2.05	1.64
		10	5.74	13.07	18.12	14.98	8.82	5.05	3.19	2.20	1.61	1.23	0.98
		15	5.74	14.41	17.81	10.80	5.21	2.87	1.80	1.23	0.89	0.67	0.52

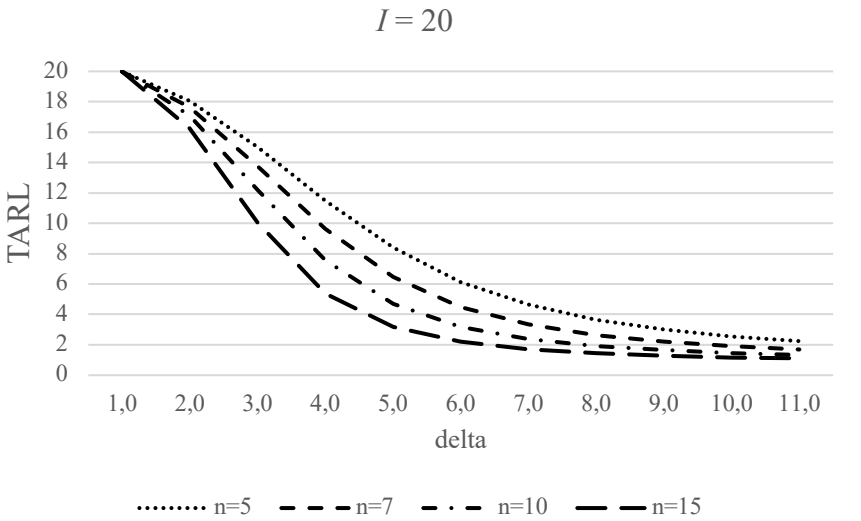
Table 3. TARL1 values of the upward modified *S* chart when $I \in \{10, 20, 30, 40, 50\}$, $n \in \{5, 7, 10, 15\}$ and $\delta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$

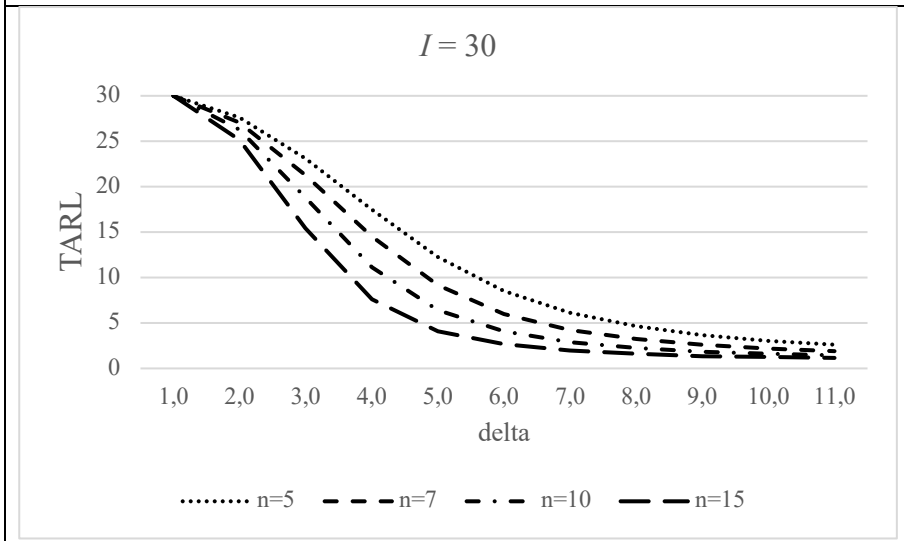
			δ									
θ	I	n	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0193	10	5	1.00	1.03	1.47	2.62	4.46	6.37	7.87	8.89	9.56	10.00
		7	1.00	1.00	1.06	1.47	2.57	4.49	6.62	8.28	9.35	10.00
		10	1.00	1.00	1.00	1.08	1.53	2.76	5.01	7.40	9.06	10.00
		15	1.00	1.00	1.00	1.00	1.10	1.64	3.23	6.07	8.61	10.00
0.0050	20	5	1.00	1.37	3.13	6.94	11.60	15.16	17.39	18.71	19.50	19.98
		7	1.00	1.01	1.38	2.83	6.27	11.23	15.34	17.86	19.24	19.98
		10	1.00	1.00	1.02	1.40	2.80	6.44	12.02	16.44	18.85	19.98
		15	1.00	1.00	1.00	1.03	1.42	2.95	7.38	13.94	18.20	19.98
0.0022	30	5	1.01	1.98	5.71	13.03	20.16	24.70	27.26	28.68	29.51	30.00
		7	1.00	1.05	1.89	4.77	11.39	19.35	24.79	27.73	29.23	30.00
		10	1.00	1.00	1.08	1.85	4.50	11.33	20.21	26.03	28.79	30.00
		15	1.00	1.00	1.00	1.09	1.82	4.59	12.66	22.70	28.04	30.00
0.0013	40	5	1.04	2.69	8.70	19.57	28.78	34.13	36.99	38.55	39.43	39.95
		7	1.00	1.12	2.44	6.92	16.92	27.57	34.15	37.48	39.12	39.95
		10	1.00	1.00	1.16	2.31	6.30	16.51	28.47	35.51	38.64	39.95
		15	1.00	1.00	1.00	1.15	2.21	6.28	18.14	31.44	37.78	39.95
0.0008	50	5	1.10	3.71	12.99	27.78	38.48	44.17	47.08	48.62	49.49	49.99
		7	1.00	1.23	3.20	9.96	24.04	36.95	44.11	47.55	49.18	49.99
		10	1.00	1.00	1.26	2.91	8.76	23.14	37.79	45.48	48.70	49.99
		15	1.00	1.00	1.00	1.24	2.71	8.52	24.96	41.00	47.81	49.99

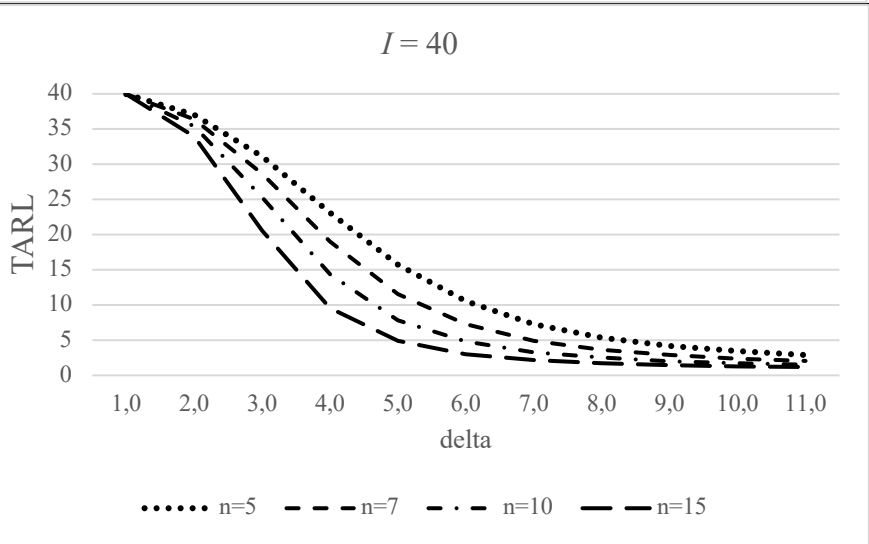
Table 4. TSDRL1 values of the upward modified *S* chart when $I \in \{10, 20, 30, 40, 50\}$, $n \in \{5, 7, 10, 15\}$ and $\delta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$

			δ									
θ	I	n	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0193	10	5	0.00	0.19	0.84	2.01	3.24	3.76	3.66	3.30	2.87	2.48
		7	0.00	0.01	0.24	0.84	1.96	3.25	3.77	3.55	3.02	2.48
		10	0.00	0.00	0.03	0.29	0.91	2.13	3.46	3.74	3.20	2.48
		15	0.00	0.00	0.00	0.05	0.34	1.03	2.50	3.72	3.42	2.48
0.0050	20	5	0.02	0.71	2.58	5.67	7.31	7.08	6.16	5.17	4.31	3.60
		7	0.00	0.10	0.73	2.27	5.24	7.26	7.04	5.85	4.62	3.60
		10	0.00	0.00	0.16	0.75	2.25	5.35	7.35	6.64	5.03	3.60
		15	0.00	0.00	0.01	0.18	0.77	2.39	5.92	7.31	5.60	3.60
0.0022	30	5	0.09	1.39	5.11	9.88	10.92	9.66	7.98	6.51	5.33	4.41
		7	0.00	0.24	1.30	4.22	9.16	10.97	9.62	7.56	5.77	4.41
		10	0.00	0.01	0.30	1.25	3.95	9.13	10.91	8.91	6.37	4.41
		15	0.00	0.00	0.02	0.31	1.22	4.05	9.73	10.43	7.24	4.41
0.0013	40	5	0.19	2.13	7.94	13.90	14.16	12.00	9.70	7.80	6.34	5.23
		7	0.00	0.37	1.88	6.34	13.04	14.38	11.99	9.17	6.90	5.23
		10	0.00	0.02	0.42	1.73	5.74	12.87	14.23	11.03	7.67	5.23
		15	0.00	0.00	0.04	0.42	1.63	5.72	13.48	13.36	8.81	5.23
0.0008	50	5	0.32	3.18	11.73	18.00	16.88	13.74	10.88	8.66	7.00	5.74
		7	0.00	0.53	2.66	9.25	17.27	17.37	13.79	10.28	7.63	5.74
		10	0.00	0.04	0.57	2.36	8.15	17.01	17.11	12.60	8.54	5.74
		15	0.00	0.00	0.06	0.55	2.15	7.92	17.50	15.79	9.92	5.74


I = 10
TARL
delta
n=5
n=7
n=10
n=15


I = 20
TARL
delta
n=5
n=7
n=10
n=15


I = 30
TARL
delta
n=5
n=7
n=10
n=15


I = 40
TARL
delta
n=5
n=7
n=10
n=15

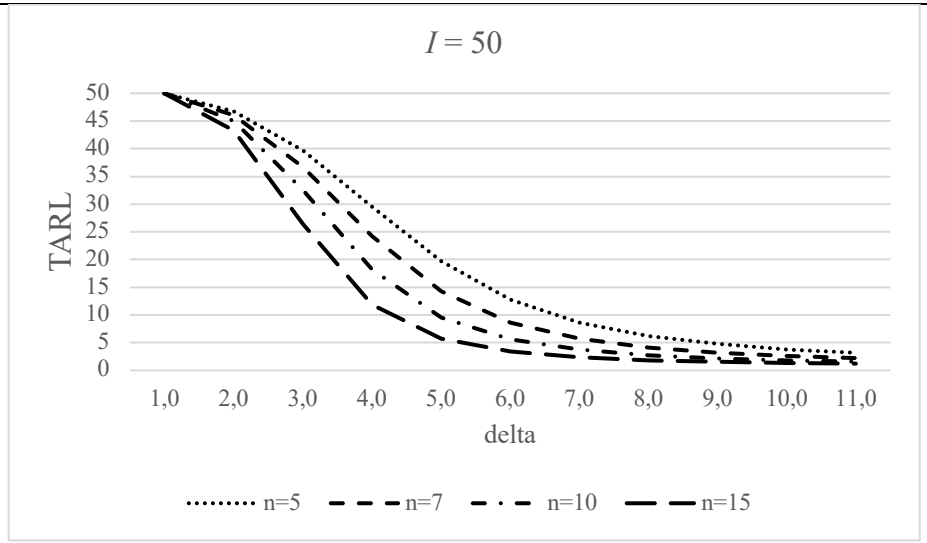
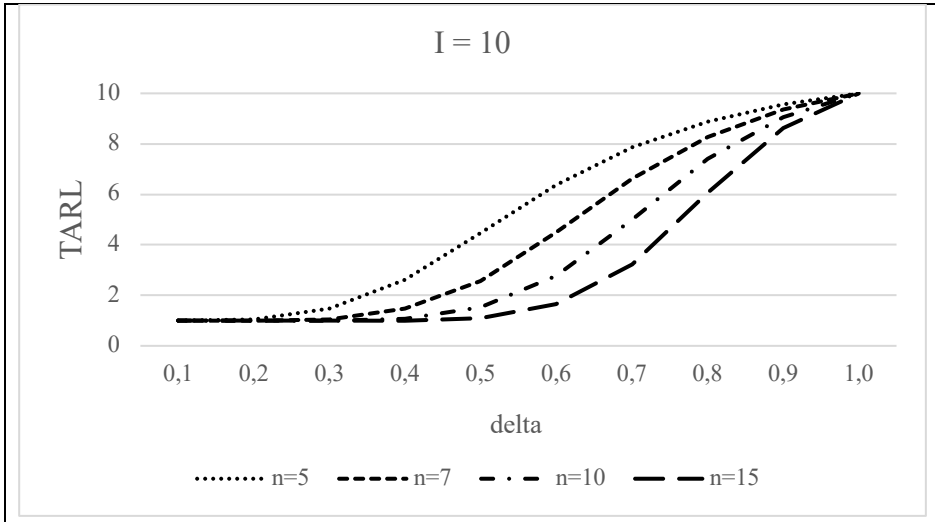
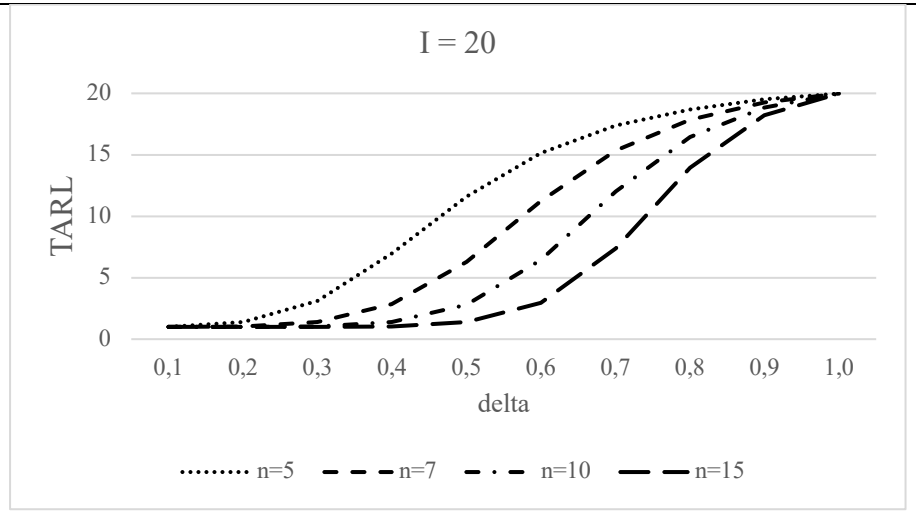
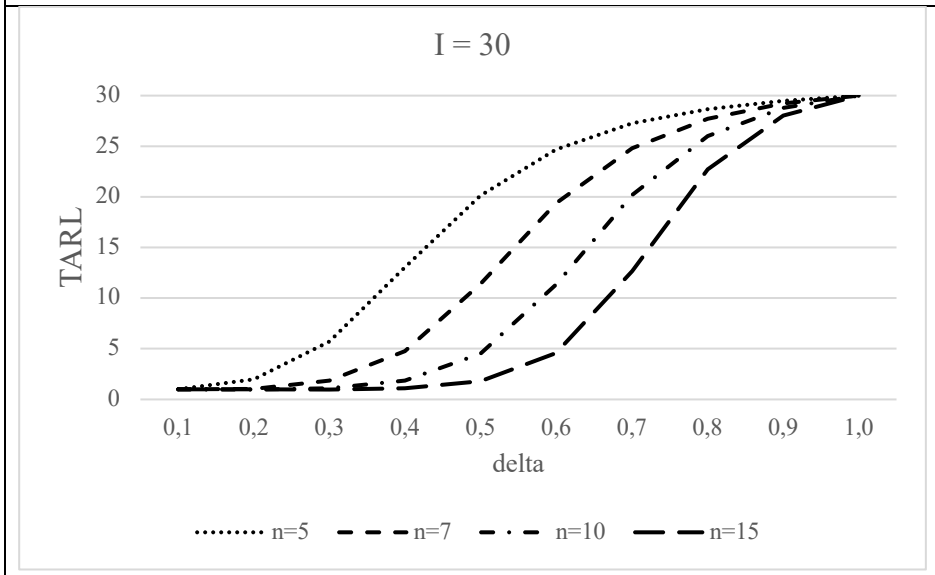
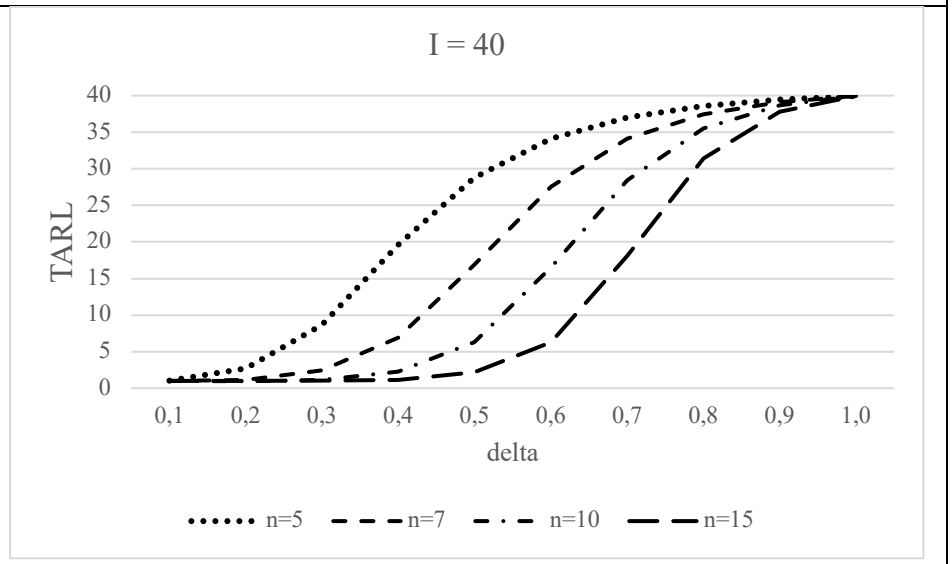


Figure 1. Comparison of $TARL_1$ values for the upward modified *S* chart, when $I \in \{10, 20, 30, 40, 50\}$, $n \in \{5, 7, 10, 15\}$ and $\delta \in \{1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0\}$


I = 10
TARL
delta
n=5
n=7
n=10
n=15


I = 20
TARL
delta
n=5
n=7
n=10
n=15


I = 30
TARL
delta
n=5
n=7
n=10
n=15


I = 40
TARL
delta
n=5
n=7
n=10
n=15

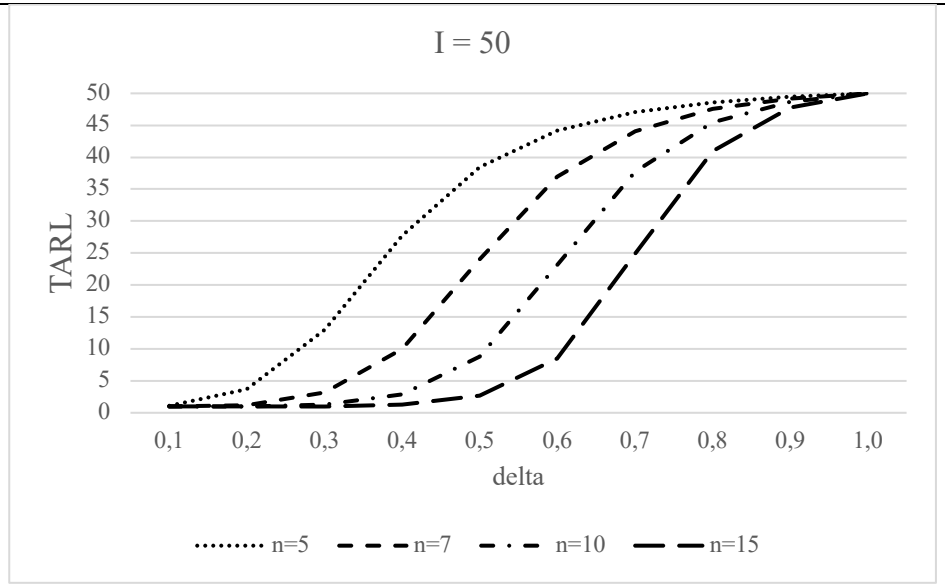


Figure 2. Comparison of $TARL_1$ values for the downward modified *S* chart, when $I \in \{10, 20, 30, 40, 50\}$, $n \in \{5, 7, 10, 15\}$ and $\delta \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$

References

[1] Noorian, S., Ahmadabadi, M. N. (2018). The use of the extended generalized lambda distribution for controlling the statistical process in individual measurements. *Statistics, Optimization & Information Computing*, 6:536-546.

[2] Khaw, K. W. and Chew, X. Y. (2019). A re-evaluation of the run rules control charts for monitoring the coefficient of variation. *Statistics, Optimization & Information Computing*, 7: 716-730.

[3] Li, C. I., Pan, J. N. and Huang, M. H. (2019) A new demerit control chart for monitoring the quality of multivariate Poisson processes. *Journal of Applied Statistics*, 46: 680-699.

[4] Chew, X. Y., Khoo, M. B. C., Khaw, K. W., Yeong, W. C. and Chong, Z. L. (2019). A proposed variable parameter control chart for monitoring the multivariate coefficient of variation. *Quality and Reliability Engineering International*, 35: 2442-2461.

[5] Shongwe, S. C., Malela-Majika, J.-C., and Castagliola, P. (2020). On monitoring the process mean of autocorrelated observations with measurement errors using w-of-w runs-rules scheme. *Quality and Reliability Engineering International*, 36: 1144-1160.

[6] Chew, X. Y., Khaw, K. W. and Yeong, W. C. (2020). The efficiency of run rules schemes for the multivariate coefficient of variation: A Markov chain approach. *Journal of Applied Statistics*, 47: 460-480.

[7] Castagliola, P., Achouri, A., Taleb. H., Celano, G. and Psarakis, P. (2015). Monitoring the coefficient of variation using a variable sample size control chart. *The International Journal of Advanced Manufacturing Technology*, 80: 1561-1576.

[8] Page, E. S. (1963). Controlling the standard deviation by CUSUMs and warning lines. *Technometrics,* 5: 307-315.

[9] Crowder, S. V., and Hamilton, M. D. (1992). An EWMA for monitoring a process standard deviation. *Journal of Quality Technology,* 24: 12-21.

[10] Su, L. and Jiang, W. (2008). A new EWMA chart for monitoring process dispersion. *Journal of Quality Technology*, 40: 319-331.

[11] Klein, M. (2000). Modified S-charts for controlling process variability. *Communications in Statistics – Simulation and Computation*, 29: 919-940.

[12] Khoo, M. B. C. (2005). A modified *S* Chart for the process variance. *Quality Engineering,* 17: 567-577.

[13] Rakitzis, A. C. and Antzoulakos, D. L. (2011). On the improvement of one-sided *S* control charts. *Journal of Applied Statistics*, 38: 2839-2858.

[14] Khoo, M. B. C., See, M. Y., Chong, N. L. and Teoh, W. L. (2019). An improved variable sample size and sampling interval *S* control chart. *Quality and Reliability Engineering International*, 35: 392-404.

[15] Kuo, T. and Lee, P. (2013). Design of adaptive *S* control charts. *Journal of Statistical Computation and Simulation*, 83: 2002-2014.

[16] Adeoti, O. A. and Olaomi, J. O. (2016). A moving average *S* control chart for monitoring process variability. *Quality Engineering*, 28: 212-219.

[17] Abujiya, M. R., Riaz, M. and Lee, M. H. (2016). A new combined Shewhart-Cumulative sum *S* chart for monitoring process standard deviation. *Quality and Reliability Engineering International*, 32: 1149-1165.

[18] Costa, A. F. B. and Neto, A. F. (2017). The *S* chart with variable charting statistic to control bi and trivariate processes. *Computers & Industrial Engineering*, 113: 27-34.

[19] Nenes, G. and Tagaras, G. (2010). Evaluation of CUSUM charts for finite-Horizon processes. *Communications in Statistics - Simulation and Computation,* 39: 578-597.

[20] Nenes, G., Castagliola, P., Celano, G. and Panagiotidou, S. (2014). The variable sampling interval control chart for finite-horizon processes. *IIE Transactions*, 46: 1050-1065.

[21] Celano, G., Castagliolab, P., Trovatoc, E. and Fichera, S. (2011). Shewhart and EWMA t control charts for short production runs. *Quality and Reliability Engineering International*, 27: 313-326.

[22] Amdouni, A., Castagliola, P., Taleb, H. and Celano, G. (2017). A variable sampling interval Shewhart control chart for monitoring the coefficient of variation in short production runs. *International Journal of Production Research*, 55: 5521-5536.

[23] Castagliola, P., Amdouni, A., Taleb, H. and Celano, G. (2015). One-sided Shewhart-type charts for monitoring the coefficient of variation in short production runs. *Quality Technology & Quantitative Management,* 12: 53-67.

[24] Chong, N. L., Khoo, M. B. C., Haq, A. and Castagliola, P. (2019). Hotelling's T^2 charts with fixed and variable sample sizes for monitoring short production runs. *Quality and Reliability Engineering International*, 35: 14-29.

[25] Chew, X. Y., Khoo, M. B. C., Khaw, K. W. and Lee, M. H. (2021). An improved Hotelling's T^2 chart for monitoring a finite horizon process based on run rules schemes: A Markov-chain approach. *Applied Stochastic Models in Business and Industry*, 37: 577-591.

[26] Tuprah, K. and Ncube, M. (1987). A Comparison of dispersion quality control charts. *Sequential Analysis,* 6: 155-163.

[27] Montgomery, D. C. (2013). *Statistical Quality Control: A Modern Introduction*. John Wiley & Sons, Inc.

Availability And Performance Analysis Of Computer Network With Dual-Server Using Gumbel-Hougaard Family Copula Distribution

Ismail Tukur
•
Department of Art and Humanities, Kano State Polytechnic, Kano, Nigeria.
ismailatukur35@gmail.com

Kabiru H. Ibrahim
•
Department of Mathematics, Aminu Kano College of Islamic and Legal Studies Kano, Nigeria
kbrhamis032@gmial.com

Muhammad Salihu Isa
•
Department of Mathematics, Federal University Dutse, Jigawa, Nigeria
salihu.muhd.isa@gmail.com

Ibrahim Yusuf
•
Department of Mathematical Sciences, Bayero University, Kano, Nigeria
iyusuf.mth@buk.edu.ng

Abstract

The determination of this paper is to study reliability measures and routine analysis of computer network, which is a combination of four subsystems A, B, C and D and all the subsystem connected in series parallel pattern, the subsystem A is client, the subsystem B is load balancer and subsystem C is servers which is divided in to two subsystem (i.e. subsystem C1 and subsystem C2) and C1andC2 served as computer servers together with two unit each and working 1-out-of-2: G policy, and subsystem D is centralized server. The system has two types of failure, degraded and complete failure. The system can completely fail due to failure of one of the following subsystems A, B, C and D. The system is at partial failed state if at least one unit is working in either subsystem C1 or subsystem C2.The system is examined using supplementary variables techniques and Laplace transform. General distribution and copula family are employed to restore degraded and complete failed state respectively. Calculated results have been highlighted by the means of tables and graphs.

Keywords: Reliability, Availability, Sensitivity, (MTTF) mean time to system failure, Gumbel − Hougaard, family, cost Analysis and supplementary variable techniques.

I. Introduction

The study of reliability modeling was started During the World War II in 1939, later then, substantial efforts have been made in this direction to create the comprehensive theoretical background for reliability modeling. The discipline is mainly anxious with requirement and valuations of the probability of device performing its purpose sufficiently for the period intended under the encountered functioning condition. Reliability has introduced new dimensions in recent years because of the difficulty of larger systems and suggestions of their failure. Undependability in the modern age of technology causes inefficiency of a system, overgenerous maintenance and can also risk human life. In today's technological world, nearly everyone be subject to upon the continued function of a wide array of machinery and equipment for our safety, security, mobility and financial welfare. We receive our electronic agreements from illuminations, hospital monitoring control, next generation aircraft, nuclear power plants data exchange system and aerospace application, to function whenever we need them when they fail, the results can be a catastrophic, injurious or eventual loss of life. The theory of reliability is the scientific discipline that studies the general regularity that must be maintained under design, research, manufacture, acceptance, and use of units/ components to acquire the highest effectiveness of their use.

The most common vital mode of redundancy is *k*-out-of-*n* redundancy. This type of redundancy is more branded into two classifications, these are *k*-out-of-*n*: G and *k*-out-of-*n*: F. A *k*-out-of-*n*: G redundancy indicates that for successful operation of the system at least *k* units out of *n* units are essential to be good (i.e. essential to work properly). If less than *k* units are good then the system fails. A *k*-out-of-*n*: F redundancy indicates that if *k* units out of *n* units have failed then the system has failed. The system reliability has been widely studied and used by numerous authors like, Yusuf et al [1] studied performance analysis of multi computer system consisting of active parallel homogeneous, Abubakar and Singh [2] have studied assessment and performance of industrial system using Gumbel Hougaard copula approach, Singh et al. [3] have studied the reliability characteristic for Internet data center with a redundant server including a main mail sever, M. Ram et.al. [4] Have discussed the reliability of a system with different failure and common cause failure under the preemptive resume policy using Gumbel-Hougaard family copula distribution. Minjae Park [5] analyzed the multi-component system with imperfect repair during warranty using renewal process. Zhang [6] analyzed on computer network reliability analysis based on intelligent cloud computing method. Kudeep et al [7] studied tree topology network environment analysis under reliability approach, nonlinear. Muhammad et al [8] studied cost benefit analysis of tree different series parallel dynamo configurations. Dillon et al. [9] have discussed the common cause's failure analysis of k-out-of-n: G systems which consisting repairable units. Ibrahim Yusuf [10] evaluates the performance of a repairable system with the concept of minor deterioration under imperfect repair. Tseng- Chang Yen et al. [11] studied Reliability and sensitivity analysis of the controllable repair system with warm standbys and working breakdown. v. v. Singh et al [12] evaluates the performance and cost assessment of repairable complex system with two subsystems connected in series configuration. v. v. Singh and Jyoti Gulati. [13] Studied performance assessment of computer Centre at Yobe state university Nigerian under different policies using copula. Kabiru H. Ibrahim et al. [14] have studied the availability and cost analysis of complex tree topology of computer network with multi-server using Gumbel-Hougaard family approach. Geon Yoon, Dae Hyun et al [15] focused on ring topology-based redundancy Ethernet for industrial network. Pratap et al [16] have examined on the assessment of complex system with two subsystems and multi types failure and repair. Rawal et al. [17] have discussed the reliability of Internet Data Centre having one mail server and one redundant server especially for the use of Interne.

Ibrahim Yusuf and Hussaini [18] examine a three-unit redundant system with three types of failure and general repair. Negi and Singh [19] studied reliability characteristics of a non-repairable complex system connected in series. C. K. Goel.et al. [20] performance assessment of repairable system in series configuration under different types of failure and repair policies using copula linguistics. However, researchers have considered different models and scrutinized the performances and availability of a complex system to forecast better performance. Most of the authors studied the complex repairable systems that are treated as single repair between two contiguous transition states. In the present paper, several reliability measures of a complex repairable system consist of four subsystems together with k-out-of-n; G configuration using two types of repair have been studied. The designed structure of the system consist of four subsystems A, B, C, and D. where C consist of C1 and C2 served as computer servers together with two unit each and working 1-out-of-2: G policy respectively. The system both are working in a series and parallel arrangement. Gumbel Hougaard family copula distribution employed for calculation and illustration.

Lastly, {S1, S3, S5, S6} represents the states of operation in degraded/partial failure while {S2, S4, S7, S_8 and S9} are entirely failed states and S_0 is at flawless operational state. The degraded points have repaired by general repair and completely failed states have repaired under Gumbel Hougaard family copula. the supplementary variables and Laplace transformation use analyze the system. Measures in reliability among availability, reliability, and MTTF and cost analysis are all treated by the means of tables and graphs.

II. State Description, Assumption and Notations

Table1: State description

State	State Description
S_0	The state S_0 indicates a perfect state in which both the subsystems are in outstanding working condition.
S_1	The state S_1 reveals a degraded state with partial failure in subsystem C, due to the failure of first server in subsystem C_1.
S_2	In this state S_2 indicate a complete failure state, due to the failure of first and second servers in subsystem C_1. The system is under repair using copula.
S_3	The state S_3 account a degraded state with partial failure in subsystem C, due to the failure of third server in subsystem C_2.
S_4	In this state S_4 indicate a complete failure state, due to the failure of third and fourth servers in subsystem C_2. The system is under repair using copula.
S_5	The state S_5 account a degraded state with partial failure in subsystem C, due to the failure of third server in subsystem C_2.
S_6	The state S_6 represents a degraded state with partial failure in subsystem C, due to the failure of first server in subsystem C_1.
S_7	The state S_7 represents a complete failed state, due to the failure of subsystem D.The system is under repair using copula.
S_8	The state S_8 represents a complete failed state, due to the failure of subsystem A. The system is under repair using copula.
S_9	The state S_9 represents a complete failed state, due to the failure of subsystem B.The system is under repair using copula.

The state description indicates, that S_0 is a perfect state where both subsystems are in good working condition. The states S_1, S_3, S_4, and S_6 are the operational states. The states S_2, S_4, S_8, and S_9 of this modeling are a complete failed state in which the system is inoperative mode.

I. Assumption

The conventions of the model is discussion below:

- At initial state, all subsystems are in good working condition.
- One unit of each subsystem C1 and subsystem C2 is necessary for operational mode.
- The system will be at complete fail state if both subsystems and C1 and C2 fail.
- Failure of all if subsystems A, B, C and D will damage the entire system.
- Failed unit of the system can be repaired when it is in operative or failed state.
- All failure rates are constant and assumed to follow exponential distribution
- The repairs follow a general distribution.
- It is assumed that a repaired system works like a new system and no damage appears during repair.
- As soon as the failed unit gets repaired, it is ready to perform the task.

Table2: Notations

t:	Time variable on time scale.
S:	A variable for Laplace transform for all expressions.
λ_{s_1} / λ_{s_2}:	Failure rates of servers of subsystems C_1
λ_{s_3} / λ_{s_4}:	Failure rates of servers of subsystems C_2
λ_B / λ_C / λ_{Cs}:	Failure rates of subsystems B, A, and D
$\phi(x)$: $\mu(z)$	Repair rates for all subsystems i.e. A, B, C and D
$u_0(x)$, $\mu(y)$, $\mu(\alpha)$:	Repair rates for complete failed states.
$p_i(x,t)$:	The probability that the system is in Si state at instant's' for I =0 to 9.
$p_i(s)$	Laplace transformation of state transition probability P (t).
$E_P(t)$	Expected profit during the time interval [0, t).
K1, K2:	Revenue and service cost per unit time in the interval [0, t) respectively.
$s_\phi(x)$	Standard repair distribution function $s_\phi(x) = \phi(x)\,\ell^{-\int_0^\infty \phi(x)\partial(x)}$
$u_0(x) = C_\theta(u_1(x), u_2(x))$	The expression of joint probability (failed state Si to good state S0) according to Gumbel Hougaard family copula is given as $C_\theta(u_1(x), u_2(x)) = \exp[x^\theta + \{\log\phi(x)\}]^{\frac{1}{\theta}}$ where, u1 = ϕ(x), and u2 = ex, where θ as a parameter, 1≤ θ≤ ∞.

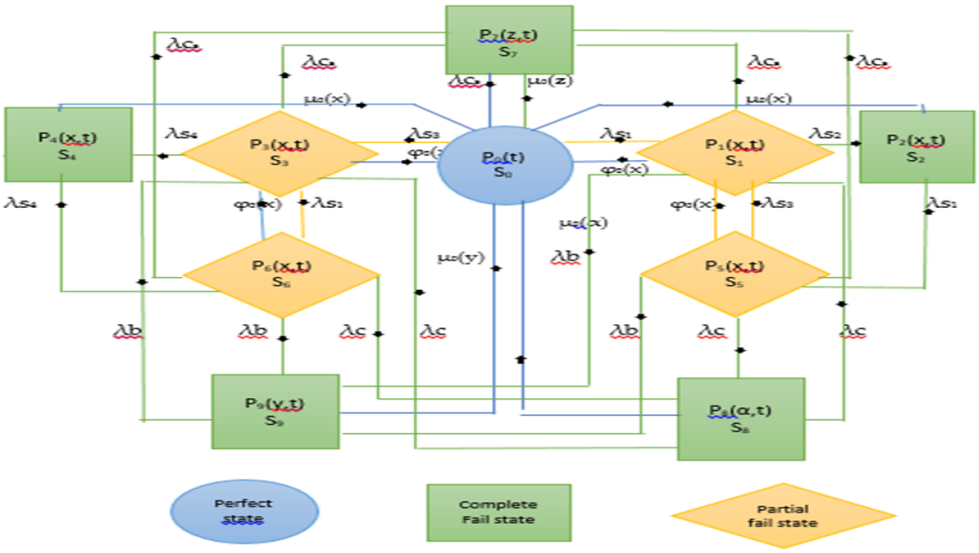


Figure 1: *Transition Diagram*

III. Formulation and Solution of Mathematical Model

By the related literature review, the following differential equation were generated for the mathematical classical.

$$\left(\frac{\partial}{\partial t}+\lambda_{s1}+\lambda_{s3}+\lambda_{Cs}\right)P_0(t) = \int_0^\infty \Phi_0(x)P_1(x,t)dx+\int_0^\infty \Phi_0(x)P_3(x,t)dx+\int_0^\infty \mu_0 P_4(x,t)dx + \int_0^\infty \mu_0 P_2(x,t)dx+\int_0^\infty \mu_0 P_7(z,t)dz+\int_0^\infty \mu_0 P_9(y,t)dy + \int_0^\infty \mu_0 P_8(\alpha,t)d\alpha \quad (1)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_c+\lambda_{cs}+\Phi(x)\right)P_1(x,t)=0 \quad (2)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\mu_0(x)\right)P_2(x,t)=0 \quad (3)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_c+\lambda_{cs}+\Phi(x)\right)P_3(x,t)=0 \quad (4)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\mu_0(x)\right)P_4(x,t)=0 \quad (5)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_{s1}+\lambda_B+\lambda_c+\lambda_{cs}+\Phi(x)\right)P_5(x,t)=0 \quad (6)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial x}+\lambda_{s4}+\lambda_B+\lambda_c+\lambda_{cs}+\Phi(x)\right)P_6(x,t)=0 \quad (7)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial z}+\mu_0(z)\right)P_7(z,t)=0 \quad (8)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial \alpha}+\mu_0(\alpha)\right)P_8(\alpha,t)=0 \quad (9)$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial y}+\ \mu_0\left(y\right)\right)P_9(y,t)=0 \quad (10)$$

Boundary conditions

$$P_1(0,t)=\lambda_{s1}P_0(t) \quad (11)$$
$$P_2(0,t)=\lambda_{21}\lambda_{s1}P_0(t) \quad (12)$$
$$P_3(0,t)=\lambda_{s3}P_0(t) \quad (13)$$
$$P_4(0,t)=\lambda_{s4}\lambda_{s3}P_0(t) \quad (14)$$
$$P_5(0,t)=\lambda_{s3}\lambda_{s1}P_0(t) \quad (15)$$
$$P_6(0,t)=\lambda_{s1}\lambda_{s3}P_0(t) \quad (16)$$
$$P_7(0,t)=(1+\ \lambda_{s1}+\lambda_{s3}+\ 2\lambda_{s1}\lambda_{s3})\lambda_{cs}P_0(t) \quad (17)$$
$$P_8(0,t)=(\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\lambda_cP_0(t) \quad (18)$$
$$P_9(0,t)=(\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\lambda_BP_0(t) \quad (19)$$

I. Solution of the Model

Equation 20 to 38 below is obtained by taking Laplace transformation of equations 1 to 19 with the help of initial condition$P_0(0)=1$,

$$\begin{aligned}(s+\lambda_A+\lambda_B+2\lambda_C+2\lambda_D)\bar{P}_0(s)&=1\\&+\int_0^\infty\Phi_0(x)P_1(x,s)dx+\int_0^\infty\Phi_0(x)P_3(x,s)dx+\int_0^\infty\mu_0(x)P_2(x,s)dx\\&+\int_0^\infty\mu_0(x)P_4(x,s)dx+\int_0^\infty\mu_0(z)P_7(z,s)dz+\int_0^\infty\mu_0(\alpha)P_8(\alpha,s)d\alpha\\&+\int_0^\infty\mu_0(y)P_9(y,s)dy\end{aligned} \quad (20)$$

$$\left(s+\frac{\partial}{\partial x}+\lambda_{s2}+\lambda_{s3}+\lambda_c+\ \lambda_{cs}+\lambda_B+\Phi_0(x)\right)\bar{P}_1(x,s)=0 \quad (21)$$
$$\left(s+\frac{\partial}{\partial x}+\mu_0(x)\right)\bar{P}_2(x,s)=0 \quad (22)$$
$$(s+\frac{\partial}{\partial x}+\lambda_{s4}+\lambda_{s1}+\lambda_c+\ \lambda_{cs}+\lambda_B+\Phi_0(x))\bar{P}_3(x,s)=0 \quad (23)$$
$$\left(s+\frac{\partial}{\partial x}+\mu_0(x)\right)\bar{P}_4(x,s)=0 \quad (24)$$
$$\left(s+\frac{\partial}{\partial x}++\lambda_{s1}+\lambda_c+\ \lambda_{cs}+\lambda_B+\Phi_0(x)\right)\bar{P}_5(x,s)=0 \quad (25)$$
$$\left(s+\frac{\partial}{\partial x}++\lambda_{s4}+\lambda_c+\ \lambda_{cs}+\lambda_B+\Phi_0(x)\right)\bar{P}_6(x,s)=0 \quad (26)$$
$$\left(s+\frac{\partial}{\partial x}+\mu_0(z)\right)\bar{P}_7(z,s)=0 \quad * \quad (27)$$
$$\left(s+\frac{\partial}{\partial x}+\mu_0(\alpha)\right)\bar{P}_8(\alpha,s)=0 \quad (28)$$
$$\left(s+\frac{\partial}{\partial x}+\mu_0(y)\right)\bar{P}_9(y,s)=0 \quad (29)$$

The Laplace transformations of the boundary conditions are:

$$\bar{P}_1(0,s)=\lambda_{s1}\overline{P_0}(s) \quad (30)$$
$$\bar{P}_2(0,s)=\lambda_{s2}\lambda_{s1}\overline{P_0}(s) \quad (31)$$
$$\bar{P}_3(0,s)=\lambda_{s3}\overline{P_0}(s) \quad (32)$$
$$\bar{P}_4(0,s)=\lambda_{s4}\lambda_{s3}\overline{P_0}(s) \quad (33)$$
$$\bar{P}_5(0,s)=\lambda_{s3}\lambda_{s1}\overline{P_0}(s) \quad (34)$$
$$\bar{P}_6(0,s)=\lambda_{s1}\lambda_{s3}\bar{P}_0(s) \quad (35)$$
$$\bar{P}_7(0,s)=\ (\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\ \lambda_{cs}\ \bar{P}_0(s) \quad (36)$$
$$\bar{P}_8(0,s)=(\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\ \lambda_c\bar{P}_0(s) \quad (37)$$
$$\bar{P}_9(0,s)=(\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\ \lambda_B\ \overline{P}_0(s) \quad (38)$$

Now solving equations (20) to (38) with the help of equations (11) to (19), yields,

$$\bar{P}_0(s)=\frac{1}{L(s)} \quad (39)$$
$$\bar{P}_1(s)=\frac{\lambda_{s1}}{L(s)}\left\{\frac{1-S_\phi(S+\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_C+\lambda_{CS})}{S+\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_C+\lambda_{CS}}\right\} \quad (40)$$

$$\bar{P}_2(s) = \frac{\lambda_{s1}\lambda_{s2}}{L(s)}\left\{\frac{1-S\mu_0(x)}{S}\right\} \quad (41)$$

$$\bar{P}_3(s) = \frac{\lambda_{s1}}{L(s)}\left\{\frac{1-S_\phi(S+\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS})}{S+\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}}\right\} \quad (42)$$

$$\bar{P}_4(s) = \frac{\lambda_{s4}\lambda_{s3}}{L(s)}\left\{\frac{1-S\mu_0(x)}{S}\right\} \quad (43)$$

$$\bar{P}_5(s) = \frac{\lambda_{s1}\lambda_{s3}}{L(s)}\left\{\frac{1-S_\phi(S+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS})}{S+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}}\right\} \quad (44)$$

$$\bar{P}_6(s) = \frac{\lambda_{s1}\lambda_{s3}}{L(s)}\left\{\frac{1-S_\phi(S+\lambda_{s4}+\lambda_B+\lambda_C+\lambda_{CS})}{S+\lambda_{s4}+\lambda_B+\lambda_C+\lambda_{CS}}\right\} \quad (45)$$

$$\bar{P}_7(s) = \frac{\lambda_{cs}}{L(s)}\ (1+\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\left\{\frac{1-S\mu_0(x)}{S}\right\} \quad (46)$$

$$\bar{P}_8(s) = \frac{\lambda_c}{L(s)}\ (\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\left\{\frac{1-S\mu_0(x)}{S}\right\} \quad (47)$$

$$\bar{P}_9(s) = \frac{\lambda_B}{L(s)}\ (\lambda_{s1}+\lambda_{s3}+2\lambda_{s1}\lambda_{s3})\left\{\frac{1-S\mu_0(x)}{S}\right\} \quad (48)$$

$$\bar{P}_{up}(s) = \bar{P}_0(s) + \bar{P}_1(s) + \bar{P}_2(s) + \bar{P}_3(s) + \bar{P}_5(s) + \bar{P}_6(s) \quad (49)$$

$$\bar{P}_{down}(s) = 1 - \bar{P}_{up}(s) \quad (50)$$

Where,

$$L(s) = \left\{\left(s+\lambda_{S1}+\lambda_{S3}+\lambda_{CS}+\right) - \left(\begin{array}{l}\lambda_{S1}\left(s+\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_C+\lambda_{CS}\right)+\lambda_{S3}\left(s+\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}\right)\\ +\lambda_{S2}\lambda_{S1}S_{\mu 0}(s)+\lambda_{S3}\lambda_{S1}S_{\mu 0}(s)+\left(1+\lambda_{S1}+\lambda_{S3}+2\lambda_{S2}\lambda_{S1}\right)\lambda_{cs}S_{\mu 0}(s)+\\ \left(1+\lambda_{S1}+\lambda_{S3}+2\lambda_{S1}\lambda_{S3}\right)\lambda_c S_{\mu 0}(s)+\left(1+\lambda_{S1}+\lambda_{S3}+2\lambda_{S1}\lambda_{S3}\right)\lambda_B S_{\mu 0}(s)\end{array}\right)+\right\} \quad (51)$$

The $\bar{P}_{up}(s)$ and $\bar{P}_{dawn}(s)$ are the system Laplace transform of the state probabilities in operative and failed state. Then,

$$\overline{P_{up}}(s) = \sum_{i=0}^{12}\overline{P_i}(s) \quad and \quad \overline{P}_{dawn}(s) = 1-\overline{P}_{up}(s)$$

$$P_{up}(t) = \left\{\begin{array}{l}1+\lambda_{S1}\left(\frac{1-S_\phi\left(s+\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_C+\lambda_{CS}\right)}{\left(s+\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_C+\lambda_{CS}\right)}\right)+\lambda_{S3}\left(\frac{1-S_\phi\left(s+\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}\right)}{\left(s+\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}\right)}\right)\\ \lambda_{S3}\lambda_{S1}\left(\frac{1-S_\phi\left(s+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}\right)}{\left(s+\lambda_{s1}+\lambda_B+\lambda_C+\lambda_{CS}\right)}\right)+\lambda_{S3}\lambda_{S1}\left(\frac{1-S_\phi\left(s+\lambda_{s4}+\lambda_B+\lambda_C+\lambda_{CS}\right)}{\left(s+\lambda_{s4}+\lambda_B+\lambda_C+\lambda_{CS}\right)}\right)\end{array}\right\} \quad (52)$$

III. Numerical Study of the Model

I. Availability Analysis

By Setting $S_{\mu_0}(s) = \bar{S}_{\exp[x^\theta+\{\log\varphi(x)\}^\theta]^{1/\theta}}(s) = \dfrac{\exp[x^\theta+\{\log\varphi(x)\}^\theta]^{1/\theta}}{s+\exp[x^\theta+\{\log\varphi(x)\}^\theta]^{1/\theta}}$, $\bar{S}_{\varphi_S}(s) = \dfrac{\varphi_S}{s+\varphi_S}$,

The expression of availability is obtained by taking the inverse Laplace transform of equation 52 together with the values of failure rates, λ_{s1}=0.0001, λ_{s2}=0.0002, , λ_{s3}=0.0003, , λ_{s4}=0.0004,λ_B=0.0005, λ_c=0.0006, λc_s = 0.0007 at $\phi(x) = \theta = x = 1$ and $\mu_0(x) = \mu_0(y) = 2.781$

$$A(x) = \left\{\begin{array}{l}-5.99867865610^{-9}e^{-1.002200000t}+1.000000381e^{-0.000699997787364t}+3.16650339610^{-7}e^{-2.718300861t}\\ -6.76892596410^{-7}e^{-1.002699361t}-1.49939921910^{-8}e^{-1.001900000t}\end{array}\right\} \quad (53)$$

The values of P_{up}(t) through variation of time t= 0, 10, 20, 30, 40, 50, 60, 70, 80, 90. as shown in Table 1 and figure 2.

Table 3:*Variation of Availability with respect to time (t)*

Time(t)	Availability
0	1
10	0.993
20	0.986
30	0.979
40	0.972
50	0.965
60	0.958
70	0.952
80	0.945
90	0.938
100	0.932

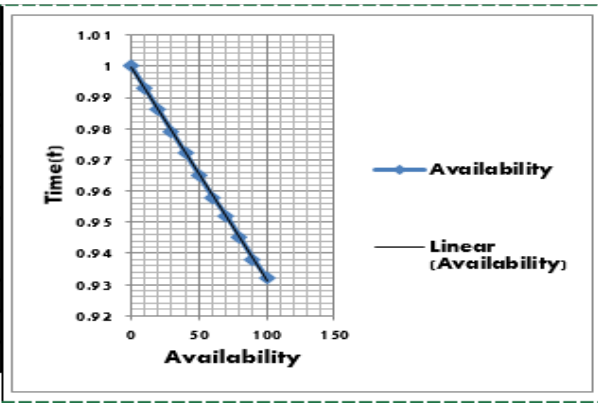


Figure 2:*Variation of Availability with respect to time (t)*

II. Reliability Analysis

Compelling all repair rates to zero with the same value of failure and repair rates in equation 53, i.e $\phi(x)$ and $\mu_0(x)$ and λ_{s1}=0.0001, λ_{s2}=0.0002, λ_{s3}=0.0003, λ_{s1}=0.0004 λ_B=0.0005, λ_C=0.0006, λ_s=0.0007, and then taking inverse Laplace transform, we obtained the expression of reliability.

$$R(t)=\begin{Bmatrix}-0.0000375000000e^{-0.00190000000t}-0.3333333333e^{-0.00230000000t}-0.00002727272727e^{-0.002200000000t}\\+1.333398106e^{-0.001100000000t}\end{Bmatrix} \quad (54)$$

Aimed at different values of time *t*= 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 units of time, we get different values of Reliability as shown in Table 2. and Figure. 3.

Table 4*: Variation of reliability with respect to time (t)*

Time(t)	Reliability
0	1
10	0.993
20	0.986
30	0.979
40	0.972
50	0.965
60	0.958
70	0.951
80	0.944
90	0.937
100	0.93

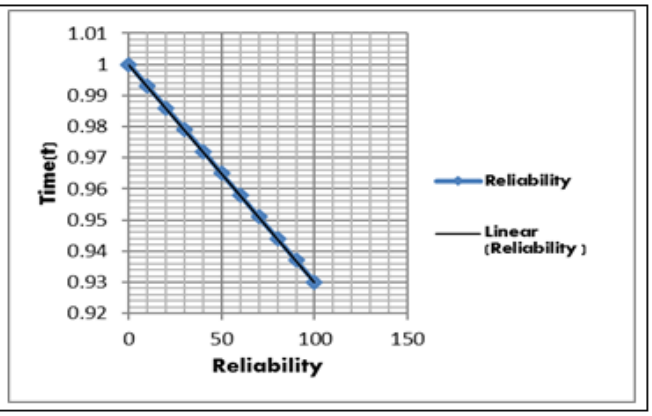


Figure 3: *Variation of reliability with respect to time (t)*

III. Mean Time To Failure (MTTF) Analysis

The expression for MTTF is found by pleasing all repairs zero in equation (53), and set the limit of *s* tends to zero:

$$M.T.T.F.=\lim_{s\to 0}\bar{P}_{up}(s)=\frac{1}{\left\{\left(\lambda_{s1}+\lambda_{s3}+\lambda_{C3}\right)\right\}}\left\{1+\frac{\lambda_{s1}}{\lambda_{s2}+\lambda_{s3}+\lambda_B+\lambda_c+\lambda_{cs}}+\frac{\lambda_{s3}}{\lambda_{s4}+\lambda_{s1}+\lambda_B+\lambda_c+\lambda_{cs}}+\frac{\lambda_{s1}\lambda_{s3}}{\lambda_{s1}+\lambda_B+\lambda_c+\lambda_{cs}}+\frac{\lambda_{s1}\lambda_{s3}}{\lambda_{s4}+\lambda_B+\lambda_c+\lambda_{cs}}\right\}$$

Location λ_{s1}=0.0001, λ_{s2}=0.0002, λ_{s3}=0.0003, λ_{s4}= 0.0004, λ_B=0.0005, λ_c=0.0006, λ_{cs}=0.0007 and changing

λ_A, λ_B, λ_C and λ_D respectively as, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, in (53), the variation of M.T.T.F. is found with respect to failure rates as shown in Table.3 and corresponding Figure.4.

Table 5*: Variation of MTTF with respect to failure rate*

Failure Rate	MTTF λ_{s1}	MTTF λ_{s2}	MTTF λ_{s3}	MTTF λ_{s4}	MTTF λ_b	MTTF λ_c	MTTF λ_{cs}
0.01	488.6	1035.2	496	971.6	939.9	940.2	99.47
0.02	462.5	1031.8	466.5	961.1	925.8	925.9	49.93
0.03	453.5	1030.5	456.2	957.2	920.5	920.6	33.31
0.04	448.9	1029.9	450.9	955.1	917.8	917.8	24.99
0.05	446.1	1029.4	447.8	953.9	916.1	916.1	19.91
0.06	444.3	1029.2	445.6	953	915	915	16.66
0.07	442.9	1028.1	444.1	952	914.1	914.2	14.28
0.08	441.9	1028.8	443	952	913.5	913.5	12.41
0.09	441.2	1028.7	442.1	951.6	913.1	913.1	11.11

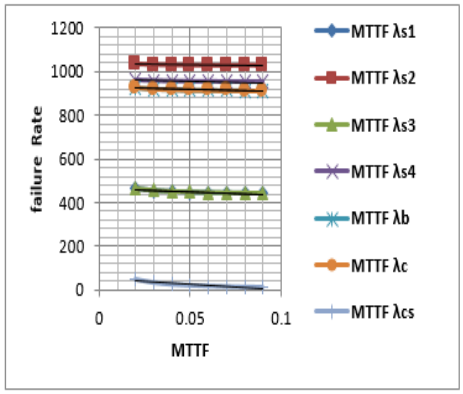


Figure 4: *Variation of MTTF with respect to failure rate*

IV. Sensitivity Analysis of MTTF

The calculation of sensitivity MTTF is studied through the partial differentiation of MTTF with respect to the failure rates of the system, by introducing the set of parametric variation of the failure rates λ_{s1} = 0.0001 λ_{s2}=0.0002, λ_{s3}=0.03, λ_{s1}=0.0004,, λ_B=0.0005, λ_c=0.0006, λ_{cs}=0.0001, from the resulting expression, the MTTF sensitivity is calculated as shown in Table 4 and the corresponding value in Figure.5 However the failure rate with respect to λ_B=0.0005 and λ_{s3}=0.03 that is the load balancer and a server from subsystem two must be given a special consideration.

Table 6*: Sensitivity of MTTF as a function of failure rates*

Failure Rate	MTTF λ_{s1}	MTTF λ_{s2}	MTTF λ_{s3}	MTTF λ_{s4}	MTTF λ_b	MTTF λ_c	MTTF λ_{cs}
0.01	-5058	-621	-5726	-1926	-2611	-2657	-9850
0.02	-1342	-186	-1528	-569	-765	-772	-2490
0.03	-607	-88	-694	-268	-360	-362	-1109
0.04	-345	-51	-395	-155	-208	-209	-624
0.05	-222	-33	-254	-101	-136	-136	-400
0.06	-155	-24	-177	-71.2	-95.2	-95.5	-278
0.07	-114	-17	-131	-52.8	-70.4	-70.7	-204
0.08	-87.3	-13	-100	-40.7	-54.4	-54.5	-156
0.09	-69.1	-11	-79.4	-32.3	-43.2	-43.3	-123

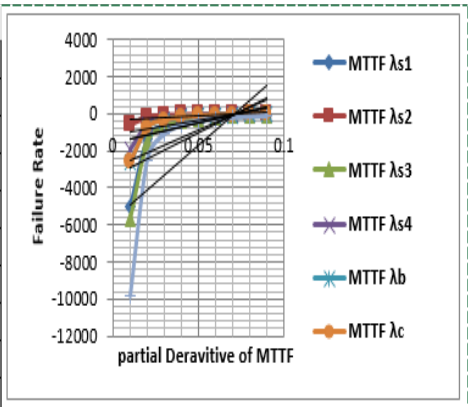


Figure 5: *Sensitivity of MTTF as a function of failure rates*

V. Cost Analysis

The predictable profit over the time interval [0, t), can be estimate by the following relation

$E_p(t) = K_1 \int_0^t P_{up}(t)dt - K_2 t$. If the service facility of the system is always available, where k_1 is revenue generated and k_2 service cost per unit time. For the same set of the parameter of failure and repair rates in (53), the expression of cost benefit analysis is attained.

$$E_p(t) = \begin{Bmatrix} 5.98551053310^{-9}\, e^{-1.0022000000t} - 1429.023674\, e^{-0.0006997787364t} + 6.75070337910^{-7}\, e^{-1.002699361t} \\ -1.16488334410^{-7}\, e^{-2.7183008861t} + 1.49655576310^{-8}\, e^{-1.0019000000t} + 1429.023673 \end{Bmatrix} \quad (55)$$

By fixing the revenue K_1= 1 and taking the values K_2= 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 respectively together with the variation of t =0, 1, 2, 3, 4, 5, 6, 7, 8, 9, Units of time, we obtained the results for expected profit as shown in Table 5 and Figure.6

Table 7: *Expected Profit in [0,t) t=0,1 ,2, 3, 4, 5,6.7,8,9 & 10*

Time (t)	$E_P(t)$: K_2=0.6	$E_P(t)$: K_2=0.5	$E_P(t)$: K_2=0.4	$E_P(t)$: K_2=0.3	$E_P(t)$: K_2=0.2	$E_P(t)$: K_2=0.1
0	0	0	0	0	0	0
1	0.4	0.5	0.6	0.7	0.8	0.9
2	0.8	1	1.2	1.4	1.6	1.8
3	1.2	1.5	1.8	2.1	2.4	2.7
4	1.59	1.99	2.39	2.79	3.19	3.59
5	1.99	2.49	2.99	3.49	3.99	4.49
6	2.39	2.99	3.59	4.19	4.79	5.39
7	2.78	3.48	4.18	4.88	5.58	6.28
8	3.18	3.98	4.78	5.58	6.38	7.18
9	3.57	4.47	5.37	6.27	7.17	8.07
10	3.97	4.97	5.97	6.97	7.97	8.97

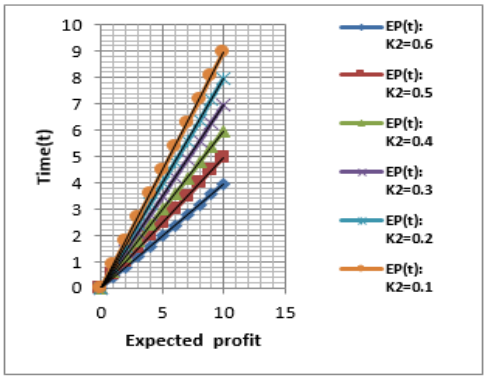


Figure 6: *Expected Profit in [0,t) t=0,1 ,2, 3, 4, 5, 6, 8,9 & 10*

IV. Discussion

The presentation of the system under the valuation of reliability measures for different values of failure and repair rates. Table.1 and figure 2 shows the evidence of availability of the complex with respect to time when the failure rates are fixed at different values mainly, λ_{s1}=0.0001, λ_{s2}=0.0002, , λ_{s3}=0.0003, , λ_{s4}=0.0004,λ_B=0.0005, λ_c=0.0006, λc_s = 0.0007. The availability of the system decreases slowly, as the probability of failure increases, the system availability will tend to zero at t very large. However, one can simply predict the future behavior of the complex system at any stage for any given set of parametric values.

Table.2 and figure.3 figure out the reliability of the system if there is no repair. The figure shown clearly that the reliability of the system is decreasing faster compare to availability, which evidently proved that when the repairs provided the performance of the system is quite better.

Table.3 and figure.4 assess the information of mean time to failure of the system (MTTF) with respect to variation of failure rates. The value change of MTTF is directly propotional to the system reliability. The computations MTTF for different values of failure rates, λ_{s1}, λ_{s2}, , λ_{s3}, λ_{s4},λ, λ_c, and λc_s from the figure the variation in MTTF corresponding to failure rates λ_{s1} is high compared to other failure which indicates that the system will not be affected with higher variations in values.

Table.4 and figure.5 demonstrations of sensitivity MTTF with respect to the values of parameters. which obtained from partial derivative of MTTF with respect to the corresponding failure rate, Moreover the variation of sensitivity MTTF corresponding to failure rates λ_B=0.0005 and λ_{s3}=0.03 that is the load balancer and a server from subsystem two, must be given a special consideration.

Table.5 and figure.6 afford the information on how the profit has been generated, by fixing revenue cost per unit time K_1= 1, and varies the service costs K_2 = 0.6, 05. 0.4, 0.3, 0.2 and 0.1, if we examine critically from Figure.6 we can reveals that the expected profit increases for low service cost. Which finally shows the Networking system of tree topology system is reliable.

Reference

[1] Yusuf Ibrahim, Sunusi Abdullahi, Ismail Abdulkareem Lado, muhaammmad Salihu Isa, K. Suleiman, Bala Shehu and Ali U.a (2020). Performance analysis of multi computer system consisting of active parallel homogeneous. *Annals of optimization Theory and Practice, volume1,no1.pp1-8. DOI:10.22121/aotp.2020.239383.1032*

[2] M.I. Abubakar and V. V. Singh (2019). Performance Assessment of an Industrial System (African Textile Manufacturers, LTD). Through Copula Approach. *Journal of Operations Research and Decisions. No. 4. Doi: 10.37190/ord190401*

[3] V. V. Singh, S. B. Singh, C. K. Goel, Stochastic analysis of internet data centre, International J. of Math. Sci& Engg. Appls. IJMESA, Vol. 3(4), (2009) 231-244.

[4] Lirong Cui, Haijun Li, (2007), Analytical method for reliability and MTTF assessment of coherent systems with dependent components, Reliability Engineering & System Safety 300-307.

[5] Minjae Park, Warranty cost analysis for multi- component systems with imperfect repair, *International Journal of Reliability and Applications,* **15**(1), (2014), 51-64.

[6] Zhang, F. (2019). Research on reliability Analysis of Computer Network Based on Intelligent Cloud Computing Method, *International Journal of Computers and Applications, 41:4, 283-288.*

[7] Kudeep Nagiya,Mangey Ram and Ayush Kumar Dua.(2017). A Tree Topology Network Environment Analysis Under reliability Approach, *Nonlinear Studies, 24(1), 199-202.*

[8] Muhammad Salihu Isa, U.A.Ali, Bashir Yusuf and Ibrahim Yusuf (2020) cost benefit analysis of tree different series parallel dynamo configurations. Life Cycle Reliability and Safety Engineering (2020) 9:413-423.*https://doi.org/10.1007/s41872-020-00141-0*

[9] B.S. Dhillon and O.C. Anude, Common cause failure analysis of a K-out-of-N: G system with repairable unit, *Microelectronics Reliability*, **34**(3), (1994), 429-442.

[10] Ibrahim Yusuf, Availability modeling and evaluation of a repairable system subject to minor Deterioration under imperfect repairs, *International journal Mathematics in Operational Research*, **7**(1), (2015), 45-51.

[11] Tseng- Chang Yen, Wu- Lin Chen and Jia-Yu Chen, Reliability and sensitivity analysis of the controllable repair system with warm standby and working breakdown, *Computers and Industrial Engineering*, **97**(3), (2016), 84-97 .

[12] Abdulkareem Lado, V. V. Singh, Kabiru, H, Ibrahim, Yusuf Ibrahim, Performance and Cost assessment of repairable complex system with two subsystems connected in series configuration, International journal of reliability and Applications. Vol.19, No1, pp.22-42.2018.

[13] V. V. Singh, and Jyoti Gulati (2015), performance assessment of computer centre at yobe state University Nigeria under Different repair policies using copula, International conference on energy, environment and chemical engineering (ICEECE2015) ISBN: 978-1-60595-257-4

[14] Kabiru, H. Ibrahim, Muhammad Salihu Isa, M. I. Ibrahim Yusuf, Ismail Tukur, Availability and cost Analysis of complex Tree Topology of computer network with Multi-server using Gumbel-hougaard family copula approach. R T AND A, NO1 (61) VOLUME 16, MARCH 2021.

[15] Geon Yoon, Dae Hyun Kwan Soon Chang Kwon, Yong Oon Park,Young Joon Lee (2006) Ring Topology-based Redundancy Ethernet for Industrial Network. *Siceicase International Joint Conference, pp.1404-1407, 18-21.*

[16] Pratap Kumar, Kabiru H. Ibrahim, M.I. Abubakar and V.V. Singh (2020) Probabilistic Assessment of Complex System with Two Subsystems in Series Arrangement With MultiTypes Failure and Two Types of Repair Using Copula. *Strategic System Assurance and Business Analytics, tttps://doi.org/10.1007/978-981-15-3647-2-2.*

[17] D. R. Cox, The analysis of non-Markov stochastic processes by the inclusion of supplementary variables, Proc. Comb. Phil. Soc. (Math. Phys. Sci.), Vol. 51, (1995) 433-44.

[18] Yusuf, I. and Hussaini, N. (2014) 'A comparative analysis of three unit redundant systems with three types of failures', *Arabian Journal for Science and Engineering*, Vol. 39, No. 4, pp.3337–3349.

[19] Negi, S. and Singh, S.B. (2015) 'Reliability analysis of non-repairable complex system with weighted subsystems connected in series', *Applied Mathematics and Computations*, Vol. 262, pp.79–89.

[20] Vijay Vir Singh, Hamisu Ismail Ayagi, C. K. Goel, Performance assessment of repairable system in series configuration under different types of failure and repair policies using copula linguistics *Int. J. Reliability and Safety, Vol. 12, No. 4, 2018*

Reliability Optimization Using Heuristic Algorithm In Pharmaceutical Plant

Tripti Dahiya[1], Deepika Garg[2] and Sarita Devi[4]

•

[1,2,4]G D Goenka University, Gurgaon – 122103, Haryana, India
Tripti.dahiya@gdgu.org, deepika.garg@gdgoenka.ac.in,Sarita.fame@gmail.com

Rakesh Kumar[3]

•

[3]Department of Mathematics and Statistics, Namibia University of Science and Technology, Windhoek
rkumar@nust.na

Abstract

In this paper, reliability of the liquid medicine manufacturing system of pharmaceutical plant named as Yaris Pharmaceuticals is enhanced on solving a redundancy allocation problem with the help of three algorithms HA_{SL1} (Heuristic algorithm with selection factor 1), HA_{SL2} (Heuristic Algorithm with Selection factor 2) and HA_{SL3} (Heuristic Algorithm with Selection factor 3). It is ensured that redundancy is allocated within given cost constraints to maximize system reliability. Post allocation of redundancy the results of these algorithms are analyzed with the help of graphs, it has been found that the reliability of the system is optimized.

Keywords: redundancy, optimization, heuristic method, system reliability

I. Introduction

In the present scenario, there is a massive growth in the industrialization due to advancement in the technology. In order to achieve the desired output, production process and machinery system has become bit complex. It is very essential that during their expected life span machines work in an effective and efficient manner. That is why concept of reliability comes into play. With the advancement of technology and growing complexity of an industrial system, it has become imperative for all production systems to perform satisfactorily during their expected life span. However, it must also be understood that failure is an inherent phenomenon that is likely to occur in most of the machines. Concept of reliability engineering aims to achieve the required and desired levels of reliability in the complex system during planning, testing and designing procedures. Reliability Redundancy Allocation Problem (RRAP) aims to maximize the system reliability under the several constraints like cost, weight and volume. A system structure achieves higher levels of reliability using minimum cost either by exchanging the existing components with more reliable components or by using redundant components in parallel. But exchanging the existing component stops the regular functioning of the system that affects the man power working on the system, their health, safety and also production & delivery services. That is why

solving cost constraint redundancy allocation in complex system has become quite popular to maximize its reliability and enhance the quality after imposition of cost related constraints. Various objective optimization techniques like heuristic algorithm, Meta-heuristic algorithms, nonlinear programming, fuzzy method etc. are used while solving a large variety of problems.

Shi method [14] used to solve constrained RAP in complex systems and the performance of this method was also compared with the existing heuristic methods. Hwan and Bong [9] proposed a heuristic method to solve constrained RAP in complex systems that allows digression over bounded infeasible region and reduces the risk of being stuck at local optimum. Garg et al. [7] introduced three heuristic algorithms for the optimization of constrained redundancy allocation problem in pharmaceutical plant and compared them to find best optimal solution. A-G method [1] formed on penalty function used to solve constrained RAP was identified and explored its feasible and infeasible regions to find optimal solution. Kumar et al. [10] analysed profit of an edible oil refinery containing four independent subsystems with the help of fuzzy logic in order to deduce that output capacity of a unit is dependent on failure and repair rates of selected independent subsystems. Kumari and Kumar [11] studied various categories of conjunctions and visual graphs to investigate conjunctions uses and effects in Mathematics writings. Kumar and Singh [12] developed a method to analyse fuzzy reliability and handle the vagueness & incompleteness of defined rough intuitionistic fuzzy set. Garg and Kumar [4] developed a mathematical model and solved differential equations with the help of Markov birth- death process and matrix method respectively, also developed C-program in order to study the variation of reliability with respect to time. Garg et al. [5] developed a mathematical model with the help of Markov birth-death process and solved differential equations based on probabilistic approach with the help of transition diagram for steady state in tab manufacturing plant. Devi and Garg [2] introduced a heuristic algorithm and constrained optimization genetic algorithm in cost constrained RAP to find best optimal solutions and compared their results in order to know which one is better algorithm. Kumar et al.[13] used Regenerative Point Graphical Technique in order to study behavioural changes happening in washing unit of a paper factory. Hsieh [8] solved nonlinear redundancy allocation problems that are multiple linear constrained with the help of introduced a two-phase linear programming approach. Tyagi et al. [15] evaluated a reliability function of renewable energy-based system by using Universal Generating Function technique and tail signature, signature, expected cost rate, expected life time and Barlow-Proschan index were found. Garg and Sharma [6] introduced a particle swarm optimization algorithm to solve a non linear constrained redundancy allocation problem in order to get best optimal solution. Devi and Garg [3] solved a constrained redundancy allocation problem with the help of fault tree analysis in manufacturing plant in order to get best optimal solution.

Further in section 2, proposed problem of the system for the manufacturing of liquid medicines in Yaris Pharmaceutical plant is given. In section 3, methodology to enhance system's reliability is described. In section 4, results of redundancy allocation of the proposed problem to increase system reliability is discussed.

II. Problem Formulation

I. System Description

The Yaris Pharmaceuticals consists of various units viz. Stirrer, colloid mill and twin head volumetric filling machine, that are placed in series for the manufacturing of liquid medicines. In the manufacturing process, there exist certain inherent factors (such as temperature, mixing time, shear rate, speed etc.) as shown in figure 1 which affect the functioning of the units of a system. Manufacturing process of liquid medicines undergoes the following procedure:

Initially required composition of raw materials is placed into a tank with stirrer machine connected to it that makes liquid solution by mixing the raw materials at required temperature and

time. Liquid solution is then transferred to colloid mill machine in order to reduce the size of particles at required shear rate and size of droplet in the liquid suspension.

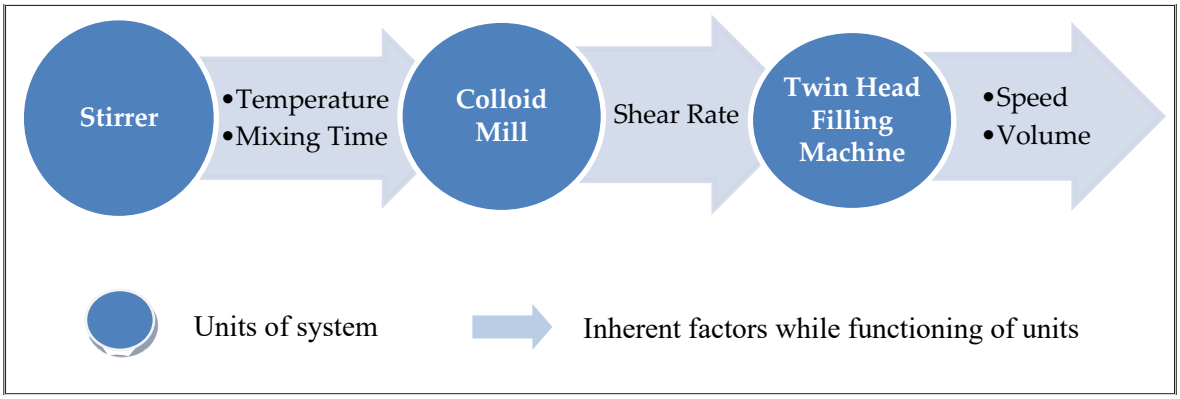


Figure 1: *Units & Respective inherent Factors*

Finally, twin head volumetric filling machine/ sealing machine is used to transfer prepared liquid for further packaging with specific speed and volume. Units of liquid medicines manufacturing system are shown in figure 2 that are connected in series.

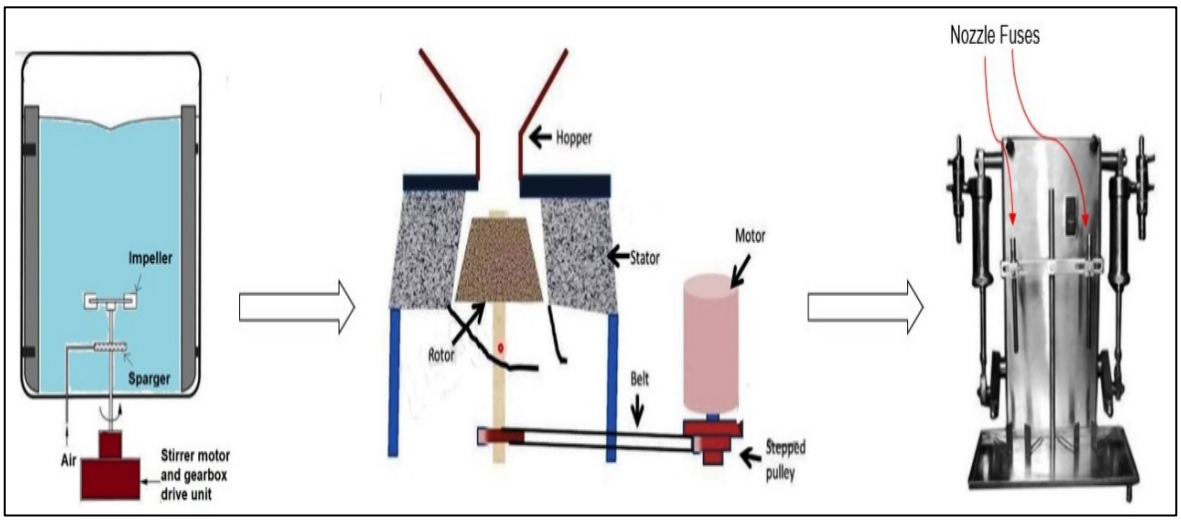


Figure 2: *Subsystems of Liquid Medicine Manufacturing Plant*

II. Assumptions

Assumptions made in this paper are as follows:

1. Coherence: A property of a system or components, as defined by Kim, Yum [9].
2. Path set: A set of components such that, if all components in the set operate, the system is guaranteed to operate.
3. Structure of total number of subsystems in system that are other than coherence is not restricted.

4. Minimal path set: A path set such that, if any subsystem is removed from the set, remaining components no longer form a path set.
5. Redundant components are not crossing the boundaries of unit in system.

III. Notations

t_i	i^{th} Subsystem/Component of system
$R_i(t_i)$, $Q_i(t_i)$	Reliability, unreliability of subsystem- t_i
$R_s(t)$	System reliability
α_i	Number of i^{th} subsystems
t	$(t_1,.....,t_\alpha)$
t^*	$(\alpha_1, \alpha_2, \alpha_3)$ is optimal solution
ΔR_i	Difference in reliability of i^{th} subsystem by adding one more redundant subsystem
$f^j(t_i)$	j^{th} resource- consumed by i^{th} subsystem
M_j	Maximum of resource-j
N_k	Minimal path set of the system
α	Number of subsystems
β	Number of constraints
γ	Number of minimal path sets
x(.)	A function that yields the system reliability, based on unique subsystems, and which depends on the configuration of the subsystems
H_i	Selection factor

IV. Redundancy Allocation Problem in Yaris Pharmaceuticals

In Yaris Pharmaceuticals, problem is to maximize the reliability with cost constraint.
Considering cost constraint M_1=Rs. 335000 and $R_s(t)$= 0.7734 (here number of constraints is one that is β=1, hence j=1)
Problem is to maximize

$$R_s(t) = g(R_1(t_1), R_2(t_2), \dots\dots\dots\dots\dots\dots\dots, R_\alpha(t_\alpha)) = \prod_{i=1}^{\alpha} R_i(t_i) \quad (1)$$

Subject to:

$$\sum_{i=1}^{\alpha} f^1(t_i) * \alpha_i \leq 335000 \quad (2)$$

Where $f^1(t_i)$ is cost of i^{th} subsystem and α_i is total number of subsystems of i^{th} subsystem. Reliability and cost of each subsystem is given by Yaris pharmaceuticals management are:

Table 1: *Reliability and cost of each component of Yaris pharmaceuticals*

Subsystem	t_1	t_2	t_3
Reliability of subsystem $R_i(t_i)$	0.9652	0.8261	0.97
Cost of subsystem $f^1(t_i)$	22444	85000	118000

III. Methods

Heuristic algorithm is one of the faster, accurate and feasible approach to solve a problem as compared to traditional methods. Here, three different heuristic algorithms have been designed to solve a problem effectively in order to get accurate and feasible solution. Also, comparison of these algorithms suggests the best one among them.

I. HA_{SL1}(Heuristic Algorithm with Selection factor 1)

Here selection factor (H_i) is defined as minimum stage reliability. Taking into account this selection factor, redundancy allocation problem is solved as follows.

Algorithm 1

Step 1
Initialize XX=0, X=0, E(0)=0
Step 2
Initialize α_i=1 for 1≤i≤α
Step 3
Find reliability of each component

$$\left[R_i(t_i) = \left(1 - \left(Q_i(t_i)\right)^{\alpha_i}\right)\right]$$

Step 4
Identify the subsystem with minimum reliability

- Find j such that

$$R_j(t_j) = min\{(R_i(t_i))\ 1 \le i \le \alpha, i \ne E(XX) forall\ 0 \le XX \le X\}$$

- If $j_1$and j_2 are such that

$$R_{j_1}(t_{j_1}) = R_{j_2}(t_{j_2}) = min\{(R_i(t_i))\ 1 \le i \le \alpha, i \ne E(XX) forall\ 0 \le XX \le X\}$$

Then check
If $f^1(t_{j_1}) < f^1(t_{j_2})$
Then take $j = j_1$ otherwise $j = j_2$
Step 5
Check for violation of the constrained on addition of additional redundant component in j[th] component

$$\sum_{\substack{i=1 \\ i \ne j}}^{\alpha} \alpha_i f^1(t_i) + (\alpha_j + 1) f^1(t_j) > M_1$$

- If violation of constrained occurred, remove this subsystem from selection list $[X = X + 1, E[X] = J]$ and go to step 6
- If violation of constrained is not occurred then check whether on addition of this subsystem reliability is improving up to desired limit i.e. $\Delta R_j > 0.0001$
- If on addition of this subsystem reliability improved up to desired limit i.e. $\Delta R_j > 0.0001$, increase that number of subsystem by one i.e. $\alpha_j = \alpha_j + 1$ and go to step 3
- If on addition, this subsystem reliability does not improve to desired limit i.e. $\Delta R_j \le 0.0001$. Remove this subsystem from further selection process i.e. X=X+1, E[X]=J and go to step 6

Step 6

- If all components are removed from further selection then $t^* = (\alpha_1, \alpha_2, \alpha_3)$ is the optimal solution
- Else go to step 3

Step 7

At the end calculate system reliability i.e. $R_s(t^*)$ using equation (1).

II. HA_{SL2}(Heuristic Algorithm with Selection factor 2)

In this algorithm selection factor (H_i) is defined as given below

$$H_i = h_i(t_i) = \frac{subsystemreliability}{percentageofresourcesconsumed} = \frac{R_i(t_i)}{\left(f^1(t_i)/M_1\right)}$$

Applying this selection factor, algorithm is as follows

Algorithm 2

Step 1

Initialize XX=0, X=0, E(0)=0

Step 2

Initialize α_i=1 for 1≤i≤α

Step 3

- Find reliability of each component

$$\left[R_i(t_i) = \left(1 - \left(Q_i(t_i)\right)^{\alpha_i}\right)\right]$$

- Calculate selection factor of each component

$$H_i = h_i(t_i) = \frac{R_i(t_i)}{\left(f^1(t_i)/M_1\right)}$$

Step 4

Identify the subsystem with minimum reliability

- Find j such that

$$h_j(t_j) = max\{(h_i(t_i))\ 1 \le i \le \alpha, \quad i \ne E(XX) forall\ 0 \le XX \le X\}$$

- If $j_1$and j_2 are such that

$$h_{j_1}(t_{j_1}) = h_{j_2}(t_{j_2}) = min\{(h_i(t_i))\ 1 \le i \le \alpha, i \ne E(XX) forall\ 0 \le XX \le X\}$$

Then check

If $f^1(t_{j_1}) < f^1(t_{j_2})$

Then take $j = j_1$ otherwise $j = j_2$

Step 5

Check for violation of the constrained on addition of additional redundant component in j^{th} component

$$\sum_{\substack{i=1 \\ i \ne j}}^{\alpha} \alpha_i f^1(t_i) + (\alpha_j + 1) f^1(t_j) > M_1$$

- If violation of constrained occurred, remove this subsystem from selection list $[X = X + 1, E[X] = J]$ and go to step 6
- If violation of constrained is not occurred then check whether on addition of this subsystem reliability is improving up to desired limit i.e. $\Delta R_j > 0.0001$
- If on addition of this subsystem reliability improved up to desired limit i.e. $\Delta R_j > 0.0001$, increase that number of subsystem by one i.e. $\alpha_j = \alpha_j + 1$ and go to step 3
- If on addition, this subsystem reliability does not improve to desired limit i.e. $\Delta R_j \le 0.0001$. Remove this subsystem from further selection process i.e. X=X+1, E[X]=J and go to step 6

Step 6

- If all components are removed from further selection then $t^* = (\alpha_1, \alpha_2, \alpha_3)$ is the optimal solution.

- Else go to step 3

Step 7

At the end calculate system reliability i.e. $R_s(t^*)$ using equation (1).

III. HA$_{SL3}$ (Heuristic Algorithm with selection factor 3)

In this algorithm selection factor (H_i) is defined as given below

$$H_i = h_i(t_i) = \frac{Changeinsubsystemreliability}{percentageofresourcesconsumed} = \frac{\Delta R_i}{\left(f^1(t_i)/M_1\right)}$$

Where $\Delta R_i = (1 - Q_i(t_i)^{\alpha_i+1}) - R_i(t_i)$

Algorithm 3

Step 1

Initialize XX=0, X=0, E(0)=0

Step 2

Initialize α_i=1 for 1≤i≤α

Step 3

- Find reliability of each component

$$\left[R_i(t_i) = \left(1 - \left(Q_i(t_i)\right)^{\alpha_i}\right)\right]$$

- Calculate selection factor of each component

$$H_i = h_i(t_i) = \frac{\Delta R_i}{\left(f^1(t_i)/M_1\right)}$$

Step 4

Identify the subsystem with minimum reliability

- Find j such that

$$h_j(t_j) = max\{(h_i(t_i))\ 1 \le i \le \alpha, \quad i \ne E(XX) forall\ 0 \le XX \le X\}$$

- If $j_1$and j_2 are such that

$$h_{j_1}(t_{j_1}) = h_{j_2}(t_{j_2}) = min\{(h_i(t_i))\ 1 \le i \le \alpha, i \ne E(XX) forall\ 0 \le XX \le X\}$$

Then check

If $f^1(t_{j_1}) < f^1(t_{j_2})$

Then take $j = j_1$ otherwise $j = j_2$

Step 5

Check for violation of the constrained on addition of additional redundant component in jth component

$$\sum_{\substack{i=1 \\ i \ne j}}^{\alpha} \alpha_i f^1(t_i) + (\alpha_j + 1) f^1(t_j) > M_1$$

- If violation of constrained occurred, remove this subsystem from selection list $[X = X + 1, E[X] = J]$ and go to step 6
- If violation of constrained is not occurred then check whether on addition of this subsystem reliability is improving up to desired limit i.e. $\Delta R_j > 0.0001$
- If on addition of this subsystem reliability improved up to desired limit i.e. $\Delta R_j > 0.0001$, increase that number of subsystem by one i.e. $\alpha_j = \alpha_j + 1$ and go to step 3
- If on addition, this subsystem reliability does not improves to desired limit i.e. $\Delta R_j \le 0.0001$. Remove this subsystem from further selection process i.e. X=X+1, E[X]=J and go to step 6

Step 6

- If all components are removed from further selection then $t^* = (\alpha_1, \alpha_2, \alpha_3)$ is the optimal solution
- Else go to step 3

Step 7

At the end calculate system reliability i.e. $R_s(t^*)$ using equation (1).

IV. Results & Discussions

In this section, application of above mentioned methodologies are discussed on the data collected from Yaris Pharmaceuticals.

I. Application of HA_{SL1}

After applying HA_{SL1} on proposed problem, following result is obtained.

Table 2: *Result of HA_{SL1}*

S. No.	Number of components in each subsystem			Consumed Resources	Subsystem Selection factor		
	α_1	α_2	α_3	$\sum_{i=1}^{\alpha} f^1(t_i) * \alpha_i$	H_1	H_2	H_3
1.	1	1	1	225444	0.9652	0.8261	0.97
2.	1	2	1	310444	0.9652	0.9697	0.97
3.	2	2	1	332888	0.9987	0.9697	0.97
4.	2	3	1	**417888?**	0.9987	!	0.97
5.	2	2	2	**450888?**	0.9987	!	!
6.	3	2	1	**355332?**	!	!	!
7.	2	2	1	332888	Algorithm stops here		

! denote 'this subsystem is removed from further selection process due to cost constraint violation'.
? denote 'the cost constraint is violated'.
Here, cost constraint is violated at 4th iteration and the optimal solution for proposed problem from table 2 is $t^* = (2,2,1)$ and system reliability is $R_s(t^*) = 0.9405$

II. Application of HA_{SL2}

After applying HA_{SL2} on proposed problem, following result is obtained

Table 3: *Result of HA$_{SL2}$*

S. No.	Number of components in each subsystem			Consumed Resources	Subsystem Selection factor		
	$\alpha_1\alpha_2\alpha_3$			$\sum_{i=1}^{\alpha} f^1(t_i) * \alpha_i$	H_1	H_2	H_3
1.	1	1	1	225444	14.4071	3.2558	2.7538
2.	2	1	1	247888	7.4533	3.2558	2.7538
3.	3	1	1	**270332#**	#	3.2558	2.7538
4.	2	2	1	332888	#	1.9108	2.7538
5.	2	2	2	**450888?**	#	1.9108	**!**
6.	2	3	1	**417888?**	#	**!**	**!**
7.	2	2	1	332888	Algorithm stops here		

! denote 'this subsystem is removed from further selection process due to cost constraint violation'.
? denote 'the cost constraint is violated'
denote 'this subsystem is removed from further selection process due to violation of ΔR_i desired limit'.
Here, cost constraint is violated at 5th iteration and the optimal solution for proposed problem from table 3 is $t^* = (2,2,1)$ and system reliability is $R_s(t^*) = 0.9394$.

III. Application of HA$_{SL3}$

After applying HA$_{SL3}$ on proposed problem, following result is obtained

Table 4: *Result of HA$_{SL3}$*

S. No.	Number of components in each subsystem			Consumed Resources	Subsystem Selection factor		
	$\alpha_1\alpha_2\alpha_3$			$\sum_{i=1}^{\alpha} f^1(t_i) * \alpha_i$	H_1	H_2	H_3
1.	1	1	1	225444	0.5051	0.5563	0.0826
2.	1	2	1	310444	0.5051	0.6646	0.0826
3.	1	3	1	**395444?**	0.5051	**!**	0.0826
4.	2	2	1	332888	0.5188	**!**	0.0826

5.	3	2	1	**355332?**		!	!	0.0826
6.	2	2	2	**450888?**	!	!	!	
7.	2	2	1	332888		Algorithm stops here		

Here, cost constraint is violated at 3rd iteration and the optimal solution for proposed problem from table 4 is $t^* = (2,2,1)$ and system reliability is $R_s(t^*) = 0.9676$.

Discussion

On applying HASL1, HASL2& HASL3 on proposed problem, redundancy allocation is done in order to improve system's reliability under given cost constrained. As a result, redundancy obtained is same i.e. (2, 2, 1) for subsystems in all the three cases that are shown in figure 3 as a common illustration for them.

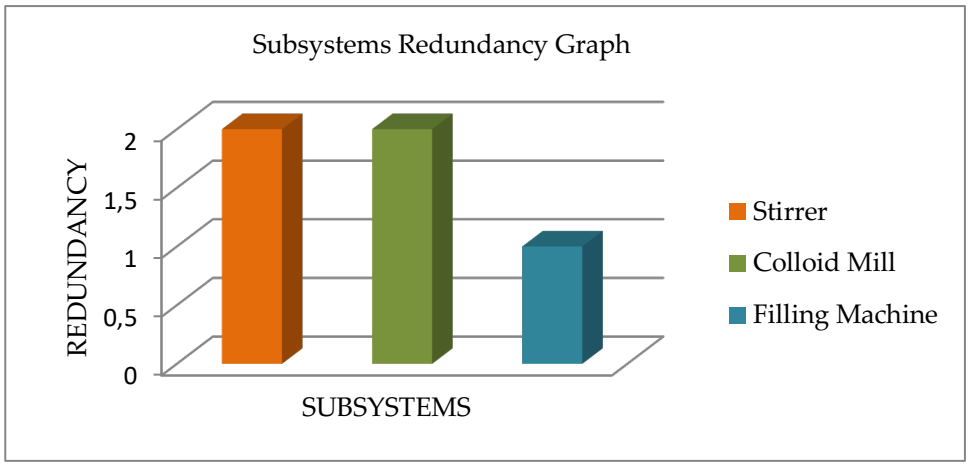


Figure 3: *Redundancy allocated to subsystems by HASL1, HASL2& HASL3*

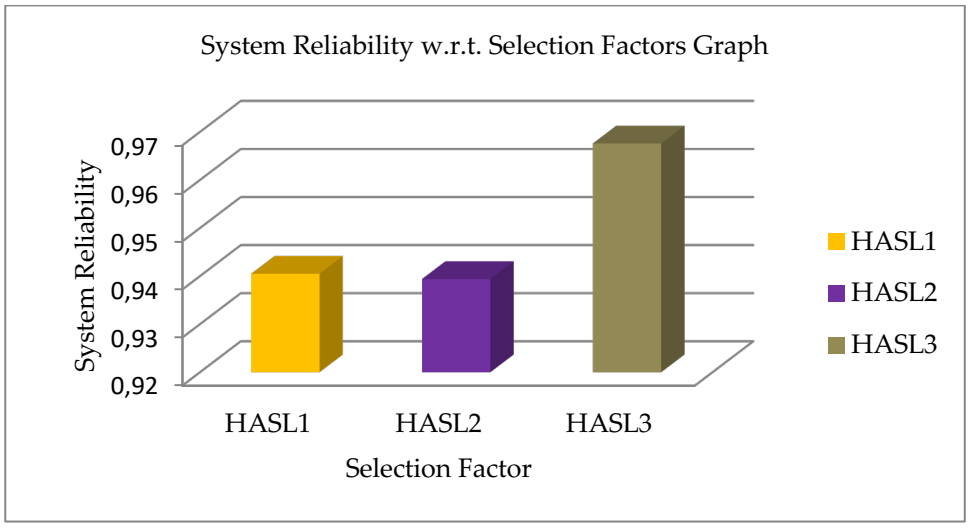


Figure 4: *Systems reliability w.r.t. selection factors of HASL1, HASL2& HASL3*

Even after getting same redundancy in case of HASL1, HASL2& HASL3, system's reliability varies in

all the three cases as shown in figure 4 that is due the selection factors of HA_{SL1}, HA_{SL2} and HA_{SL3} in this paper. It clearly states that not only redundancy allocation but selection factors also plays important role in the variation of reliability of a system.

V. Conclusion

Solving a redundancy allocation problem of pharmaceutical plant (Yaris Pharmaceuticals) using described three algorithms, reliability of the liquid medicine manufacturing system is enhanced under the given cost constrained. Before applying algorithms system reliability was 0.7734, after the application of HA_{SL1}, HA_{SL2} and HA_{SL3} reliability of liquid medicine manufacturing system is improved by 20.06%.

References

[1] Agarwal, M. and Gupta, R. (2005). Penalty function approach in heuristic algorithms for constrained redundancy reliability optimization.*Institute of Electrical and Electronics Engineers (IEEE) Trans. Reliab.*, 54(3):549-558.

[2] Devi, S. and Garg, D. (2017). Redundancy-Allocation in Neel Metal Products Limited. *Indian Journal of Science and Technology*, 10(30).

[3] Devi, S. and Garg, D. (2019). Reliability Analysis of Manufacturing Plant via Fault Tree Analysis.*Journalof Advances and Scholarly Researches in Allied Education (JASRAE)*, 16(5):256-259.

[4] Garg, D. and Kumar, K. (2009). Matrix- Based System Reliability Method and its Application to Rice plant.*The IUP Journal of Computational Mathematics*, 2(4):17-29.

[5] Garg, D., Kumar, K. and Sigh, J. (2010). Decision support system of a tab manufacturing plant. *Journal of Mechanical Engineering*, 41(1):71-79.

[6] Garg, H., Sharma, S. P. (2013). Reliability–redundancy allocation problem of pharmaceutical plant. *Journal of Engineering Science and Technology*, 8(2):190 – 198

[7] Garg, D., Kumar, K., Pahuja, G. L. (2010). Redundancy-allocation in pharmaceutical plant. *International Journal of Engineering Science and Technology*, 2(5):1088–97.

[8] Hsieh, Y.-C. (2002). A two-phase linear programming approach for redundancy allocation problems. *Yugoslav journal of Operation Research*, 12(2): 227-236.

[9] Hwan, J. K. and Bong, J. Y. (1993). A heuristic method for solving redundancy-optimization problems in complex systems. *Institute of Electrical and Electronics Engineers (IEEE) transaction on reliability*, 42(4):572-578.

[10] Kumar, A., Garg, D. and Goel, P. (2017). Mathematical Modeling and Profit Analysis of An Edible Oil Refinery Industry. *Airo International Research journal*, (13) ISSN 2320-3714: 1-14.

[11] Kumari, A. and Kumar, D. (2019). Conjunction as cohesive devices in the writing of Mathematicians. *The International Organization of Scientific Research Journal of Mathematics (IOSR-JM)*, 15(5):17-22.

[12] Kumar, D. and Singh, S. B. (2014). Evaluating fuzzy reliability using rough intuitionistic fuzzy set. *Innovative Application of Computational Intelligence on Power, Energy and controls with their impact on humanity , Institute of Electrical and Electronics Engineers Xplore*, 138-142.

[13] Kumar, A., Garg, D. and Goel, P. (2019). Mathematical modeling and behavioral analysis of a washing unit in paper mill. *International Journal of System Assurance Engineering and Management*, 10:1639-1645.

[14] Shi, D. H. (1987). A New heuristic Algorithm for constrained redundancy-optimization in complex systems. *Institute of Electrical and Electronics Engineers (IEEE) Transaction on reliability*, 36(5):621-623.

[15] Tyagi, S., Kumar, A., Bhandari, A. S. and Ram, M. (2020). Signature reliability evaluation of renewable energy system. *Yugoslav journal of operation research*, 20(20).

On Estimating Standby Redundancy System in a MSS Model with GLFRD Based on Progressive Type II Censoring Data

Marwa KH. Hassan
•
Department of Mathematics, Faculty of Education,
Ain Shams University, Cairo, Egypt.
E-mail: marwa_khalil2006@hotmail.com

Abstract

Redundancy is an approach to improve the reliability system. There are three main models of redundancy. In a system with standby redundancy, there are number of components only one of which works at a time and the other remain as standbys. When an impact of stress exceeds the strength of the active component, for the first time, it fails and another from standbys, if there is any, is activated and faces the impact of stresses, not necessarily identical as faced by the preceding component and the system fails when all the components have failed. In This paper, we consider the problem of estimation the reliability of a multicomponent stress- strength system called N-M- cold - standby redundancy. This system includes N- subsystem consisting of M- independent distributed strength components only one of which works under the impact of stress. The system fails when all the components have failed. Assuming the stress and strength random variables have the generalized linear failure rate distribution with common scale parameters and different shape parameter. The reliability estimated based on progressive type II data. Simulation study is used to compare the performance of the estimators. Finally, real data set is used the proposed model in practice.

Keywords: N-M Standby Redundancy System; Progressive Type II Censoring; Generalized Linear Failure Rate Distribution (GLFRD); Multicomponent stress- strength (MSS); Simulation study; Bootstrap; Bayes estimator; Maximum Likelihood Method.

I. Introduction

Reliability is defined as the probability of not failing in an environment for a mission time. Reliability is a statistical probability and are no absolutes or guarantees. For stress and strength models, both the strength X of the system and stress Y are random variables. The stress-strength model describes the life of a component which has a random strength X and is subjected to random stress Y. The component fails at the instant that the stress applied to it exceeds the strength and the component will function satisfactory whenever $Y < X$. Thus $R = P[Y < X]$ is a measure of component reliability. The idea of stress-strength model was presented by Birnbaum [1], for more reference see Kotz et al [2]. Recently it studied by Basirat et al [2], Asgharzadeh et al [4] and Hassan [5-6]. The stress- strength model is applied in many fields such as quality control, engineering, medicine, biostatistics and economics. Also, the reliability is considered in a multicomponent stress-strength model, which is introduced by Bhattacharyya and Johnson [7] and recently see Hassan and Alohali [8] Sriwastav and Kakati [9] are considered the stress-strength reliability of standby redundancy, that is, there are number of components only one of which works at a time and the other remains as standby.

When an impact of stress exceeds the strength of the active component, for the first time, it fails and another component from standby. The system is fails when all the components has failed. The standby redundancy systems have many applications such as military satellite, standby redundancy system can improve the lifetime of satellite. For stress-strength of standby redundancy system see Khan and Jan [10] ke et al [11], Gogoi et al [12 and Liu et al [13]. In This paper we consider the estimation of N-M- cold- standby redundancy system in a multicomponent stress-strength model based on generalized failure rate distribution (GLFRD). Sarhan and Kundu [14] introduced this distribution has the following pdf and cdf are given by

$$f(x; a, b, \alpha) = \alpha\,(a + b\,x)\,\mathrm{Exp}\left[-\left(ax + \frac{b}{2}\,x^2\right)\right]\,\left(1 - \mathrm{Exp}\left[-\left(ax + \frac{b}{2}\,x^2\right)\right]\right)^{\alpha-1}$$

and

$$F(x;; a, b, \alpha) = \left(1 - \mathrm{Exp}\left[-\left(ax + \frac{b}{2}\,x^2\right)\right]\right)^{\alpha}$$

Where $x > 0$, $a, b > 0$ are the scale parameters and $\alpha > 0$ is the shape parameter.

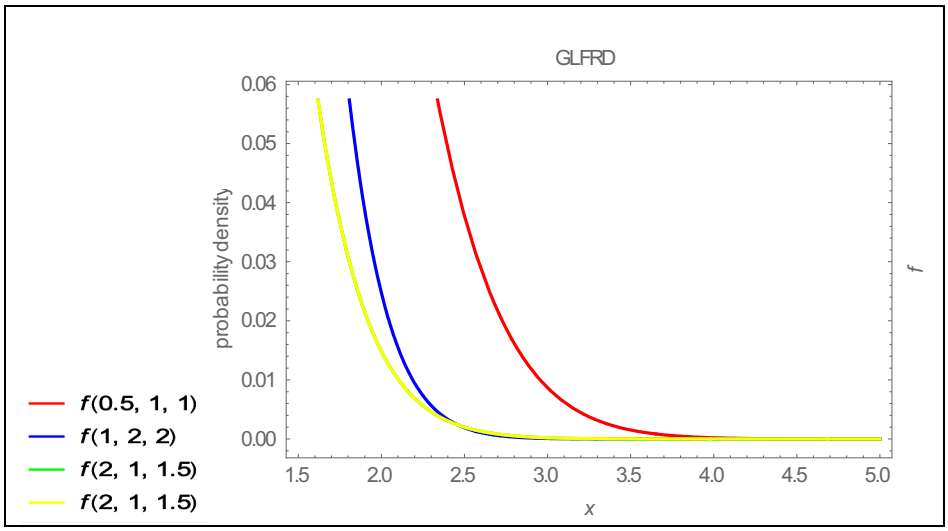


Figure (1): *Different shapes of pdf of GLFRD*

This distribution used as a lifetime model because it has increasing, decreasing or bathtub shaped hazard rate function. Figure (1) show the pdf of GLFRD which may have no mode at all.

II. METHODOLOGY

Reliability is an important concept at the planning, designing, manufacturing, and operating stages operating stages of systems ranging from simple to complex. For multicomponent system to make it more reliable use redundant parts. Redundancy plays an important role in enhancing system reliability. One of the commonly used forms of the redundancy is standby redundancy system. We consider the standby redundancy system which consists of certain number of same subsystems with series structure. Suppose $X_{i1} \ldots \ldots X_{iM}$, $i = 1 \ldots \ldots N$ are M independent strength random variables follow have $GLFRD(a, b, \alpha)$ in i^{th} subsystem. Let $Z_i = \min(X_{i1} \ldots \ldots X_{iM})$, $i = 1 \ldots \ldots N$, then $Z_1 \ldots \ldots Z_N$ is the set of N independent strength random variables. Let $Y_1 \ldots \ldots Y_N$ be independent stress random variables follow have $GLFRD(a, b, \beta)$. Then the reliability of the system is given by:

$$R = R(1) + \cdots + R(N). \qquad (1)$$

The marginal reliability R(i) of i^{th} subsystem *is*

$$R(i) = P[Z_1 < Y_1, \dots\dots, Z_{i-1} < Y_{i-1}, Z_i > Y_i], \qquad i = 1 \dots\dots N.$$

$$= \int_0^\infty F_i(y_1)h_1(y_1)dy_1 \dots\dots \int_0^\infty F_{i-1}(y_{i-1})h_{i-1}(y_{i-1})dy_{i-1} \int_0^\infty (1 - F_i(y_i))h_i(y_i)dy_i. \quad (2)$$

Since $Z_i's \approx GLFRD(a, b, \alpha)$, *then*

$$f_i(z; a, b, \alpha_i) = \alpha_i\,(a + b\,z)\,Exp\left[-\left(az + \frac{b}{2}\,z^2\right)\right]\,(1 - Exp\left[-\left(az + \frac{b}{2}\,z^2\right)\right])^{\alpha_i - 1}$$

and

$$F_i(z;; a, b, \alpha_i) = \,(1 - Exp\left[-\left(az + \tfrac{b}{2}\,z^2\right)\right])^{\alpha_i}, i = 1 \dots\dots N.$$

Since $Y_i's \approx GLFRD(a, b, \beta)$, *then*

$$h_i(y; a, b, \beta_i) = \beta_i\,(a + b\,y)\,Exp\left[-\left(ay + \frac{b}{2}\,y^2\right)\right]\,(1 - Exp\left[-\left(ay + \frac{b}{2}\,y^2\right)\right])^{\beta_i - 1}$$

and

$$H_i(y; a, b, \beta_i) = \,(1 - Exp\left[-\left(ay + \tfrac{b}{2}\,y^2\right)\right])^{\beta_i}, i = 1 \dots\dots N.$$

Hence,

$$R(i) = \frac{\beta_i}{\beta_i + \alpha_i} \prod_{j=1}^{i-1} \frac{\alpha_j}{\alpha_j + \beta_j} \quad (3)$$

And the reliability system is

$$R = \frac{\beta_1}{\beta_1 + \alpha_1} + \sum_{i=2}^{N} \frac{\beta_i}{\beta_i + \alpha_i} \prod_{j=1}^{i-1} \frac{\alpha_j}{\alpha_j + \beta_j} \quad (4)$$

I. Point Estimators of Standby Redundancy System in a MSS Model with GLFRD Based on Progressive Type II Censoring Sample:

we will derive the different point estimator for R based on progressive type II censored sample. First, we will introduce a brief description for this date type. It is very useful in lifetime studies. It saves cost and time, because it allows to cancel surviving units during the experiment time. In this progressive censoring scheme, let we want to study n units but m failures are completely observed. At the first failure time $Z_{1:m:n}$, S_1 of surviving units are randomly selected and removed from remaining $n - 1$ units, in the second failure time $Z_{2:m:n}$ observed, S_2 of the surviving units are randomly selected and removed from $n - 2 - S_1$, finally at the m^{th} failure time, when m^{th} failure $Z_{m:m:n}$ observed all the S_m surviving units are removed. This approach getting the censored sample of size m is called progressive type II censored sample with censoring scheme $S_1, \dots, S_m$ for more details see Balakrishnan and Aggarwala [15], Balakrishnan [16], and Krishna and Kumar [17].

1. Maximum Likelihood Estimator of R

To get the maximum likelihood estimator of R, Let the progressive type II censored sample $Z_{1:n_i:N_i}. \dots Z_{n_i:n_i:N_i}$ of Z_i random variables with censoring scheme $\{n_i, N_i, r_{i1} \dots r_{in_i}\}$ and similar $Y_{1:m_i:M_i}. \dots Y_{m_i:m_i:M_i}$ of Y_i random variables with censoring scheme $\{m_i, M_i, \grave{r}_{i1} \dots \grave{r}_{im_i}\}$. To simplify our symbols, we replace $Z_{1:n_i:N_i}. \dots Z_{n_i:n_i:N_i}$ by $Z_{1:n_i}. \dots Z_{n_i:n_i}$ and replace $Y_{1:m_i:M_i}. \dots Y_{m_i:m_i:M_i}$ by $Y_{1:m_i}. \dots Y_{m_i:m_i}$.

Let Z_{ij} and Y_{il}, $i = 1 \dots N$, $j = 1 \dots n_i$ and $l = 1 \dots m_i$ are independent random variables having $GLFRD(a, b, \alpha_i)$ and $GLFRD(a, b, \beta_i)$ respectively.

The Likelihood function of Z_{ij} and Y_{il} is

$$L(\alpha_i, \beta_i) = \prod_{i=1}^{N} f(z_{i1} \dots z_{in_i})\, h(y_{i1} \dots y_{im_i})$$

Where,

$$f(z_{i1} \dots z_{in_i}) = C_i\, \alpha_i^{n_i}\ Exp(-T_i(a,b) + S_i(a,b) + (\alpha_i - 1)\, Q_i(a,b) + \textstyle\sum_{j=1}^{n_i} r_{ij} Ln(1 - A_{ij}^{\alpha_i}(a,b)))\,,$$

$$h(y_{i1} \dots y_{im_i}) = \grave{C}_i\, \beta_i^{m_i}\ Exp(-\grave{T}_i(a,b) + \acute{S}_i(a,b) + (\beta_i - 1)\, \grave{Q}_i(a,b) + \sum_{l=1}^{m_i} \grave{r}_{il} Ln(1 - \grave{A}_{il}^{\beta}(a,b)))$$

where,

$$C_i = N_i\,(N_i - 1 - r_{i1}) \dots .(N_i - n_i + 1 - \textstyle\sum_{j=1}^{n_i - 1} r_{ij}),$$

$$\grave{C}_i = M_i\,(M_i - 1 - \grave{r}_{i1}) \dots .(M_i - m_i + 1 - \textstyle\sum_{l=1}^{m_i - 1} \grave{r}_{il}),$$

$$T_i(a,b) = \textstyle\sum_{j=1}^{n_i}(a\, z_{ij} + b\, z_{ij}^2),\ \ \grave{T}_i(a,b) = \textstyle\sum_{l=1}^{m_i}(a\, y_{il} + b\, y_{il}^2)$$

$$S_i(a,b) = \textstyle\sum_{j=1}^{n_i} Ln(a + b\, z_{ij}),\ \grave{S}_i(a,b) = \textstyle\sum_{l=1}^{m_i} Ln(a + b\, y_{il})$$

$$Q_i(a,b) = \textstyle\sum_{j=1}^{n_i} Ln(A_{ij}(a,b))\,,\ \grave{Q}_i(a,b) = \textstyle\sum_{l=1}^{m_i} Ln(\grave{A}_{il}(a,b))$$

$$A_{ij}(a,b) = 1 - exp\,(-\left(a\, z_{ij} + \tfrac{b}{2}\, z_{ij}^2\right)),\ and\ \grave{A}_{il}(a,b) = 1 - exp\,(-\left(a\, y_{il} + \tfrac{b}{2}\, y_{il}^2\right)).$$

Then, the log-likelihood function is given by

$$\begin{aligned} Ln\left(L(\alpha_i, \beta_i)\right) = Ln(k) &+ \sum_{i=1}^{N} n_i \ln(\alpha_i) - \sum_{i=1}^{N} T_i(a,b) + \sum_{i=1}^{N} S_i(a,b) \\ &+ \textstyle\sum_{i=1}^{N}(\alpha_i - 1)\, Q_i(a,b) + \sum_{i=1}^{N}\sum_{j=1}^{n_i} r_{ij} Ln(1 - A_{ij}^{\alpha_i}(a,b)) + \\ &+ \sum_{i=1}^{N} m_i \ln(\beta_i) - \sum_{i=1}^{N} \grave{T}_i(a,b) + \sum_{i=1}^{N} \grave{S}_i(a,b) + \\ \textstyle\sum_{i=1}^{N}(\beta_i - 1)\ \grave{Q}_i(a,b) &+ \textstyle\sum_{i=1}^{N}\sum_{l=1}^{m_i} \grave{r}_{il} Ln(1 - \grave{A}_{il}^{\beta_i}(a,b)) \end{aligned}$$

If shape parameters a and b are known, then the maximum likelihood estimators of α_i and β_i are the solution of the following nonlinear equations

$$\frac{n_i}{\alpha_i} + Q_i(a,b) - \textstyle\sum_{j=1}^{n_i} r_{ij}\left(\frac{A_{ij}^{\alpha_i}(a,b)\, Ln\left(A_{ij}(a,b)\right)}{1 - A_{ij}^{\alpha_i}(a,b)}\right) = 0 \qquad (5)$$

$$\frac{m_i}{\beta_i} + \grave{Q}_i(a,b) - \textstyle\sum_{l=1}^{m_i} \grave{r}_{il}\left(\frac{\grave{A}_{il}^{\beta_i}(a,b)\, Ln\left(\grave{A}_{il}(a,b)\right)}{1 - \grave{A}_{il}^{\beta_i}(a,b)}\right) = 0 \qquad (6)$$

using numerical nonlinear maximization techniques are $\hat{\alpha}_i$ and $\hat{\beta}_i$ then the maximum likelihood estimator of R is getting using the invariance property as

$$R_{MLE} = \frac{\hat{\beta}_1}{\hat{\beta}_1 + \hat{\alpha}_1} + \textstyle\sum_{i=2}^{N} \frac{\hat{\beta}_i}{\hat{\beta}_i + \hat{\alpha}_i} \prod_{j=1}^{i-1} \frac{\hat{\alpha}_j}{\hat{\alpha}_j + \hat{\beta}_j} \qquad (7)$$

2. Bayes estimator of R

To get the Bayes estimator of R, suppose that the prior distributions for

$\alpha_i \approx$ Gamma (t_{i1}, s_{i1}) and $\beta_i \approx$ Gamma (t_{i2}, s_{i2}), then the Bayes estimator of R under squared error loss function is defined by

$$R_{Bayes} = E[R(\alpha_i, \beta_i|data)]$$

$$= \int_0^\infty \int_0^\infty R(\alpha_i, \beta_i)\pi(\alpha_i, \beta_i|data)\ d\alpha_i\ d\ \beta_i$$

Where,

$$\pi(\alpha_i, \beta_i|data) = W\ L(\alpha_i, \beta_i|data)\ \pi(\alpha_i)\ \pi(\beta_i)\ ;$$

$$W^{-1} = \int_0^\infty \int_0^\infty L(\alpha_i, \beta_i|data)\ \pi(\alpha_i)\ \pi(\beta_i)\ d\alpha_i\ d\ \beta_i;$$

$$\pi(\alpha_i) = \frac{s_{i1}^{t_{i1}}}{\Gamma(t_{i1})}\ \alpha_i^{t_{i1}-1} e^{-\ s_{i1}\ \alpha_i},\ \pi(\beta_i) = \frac{s_{i2}^{t_{i2}}}{\Gamma(t_{i2})}\ \alpha_i^{t_{i2}-1} e^{-\ s_{i2}\ \beta_i}.$$

But this integral is difficult to calculate, we will use the Lindley approximation which introduced by Lindley [18] more details see Ahmed et al [19]. Then the Bayes estimator of R using the Lindley approximation is defined as

$$R_{Bayes} = R + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}(R_{ij} + 2\ R_i\ \rho_j)\sigma_{ij} + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} L_{ijk} R_i\ \sigma_{ij}\ \sigma_{kl} \qquad (8)$$

Where all calculations are computed at by $\hat{\alpha}_i$ and $\hat{\beta}_i$. Let $\Theta = (\theta_1, \theta_2) = (\alpha_i, \beta_i)$, $\rho(\Theta)$ is the log of joint prior of Θ, then

$$R_i = \frac{\partial R}{\partial \theta_i},\ \rho_i = \frac{\partial \rho}{\partial \theta_i}\ ,\ i = 1,2.$$

$$R_{ij} = \frac{\partial^2 R}{\partial \theta_i\, \partial \theta_j}, i, j = 1,2.$$

$$L_{ijk} = \frac{\partial^3 Log(L)}{\partial \theta_i\ \partial \theta_j\ \partial \theta_k}, i, j, k = 1,2$$

and σ_{ij} are the elements of the Fisher information matrix of Θ.

II. Interval Estimation of Standby Redundancy System in a MSS Model with GLFRD Based on Progressive Type II Censoring Sample:

we consider the interval estimation of R. Construct the asymptotic confidence interval (ACI) and bootstrap confidence interval (BCI) of R.

1. Bootstrap confidence interval of *R* (Boot-P Method).

Eforn [20] suggested confidence interval based on nonparametric bootstrap method called Boot-P. as follows:

1. Generate the progressive type II censored data $Z_{1:n_i}. \ldots Z_{n_i:n_i}$ for given parameters a, b, α_i of GLFRD.
2. Generate the progressive type II censored data $Y_{1:m_i}. \ldots Y_{m_i:m_i}$ for given parameters a, b, β_i of GLFRD.

3. Compute the maximum likelihood estimators of α_i, β_i and R according to equation (5).
4. Based on pre-specified progressive censoring schemes $\{r_{i1} \dots r_{in_i}\}$ and $\{\grave{r}_{i1} \dots \grave{r}_{im_i}\}$ to generate type II progressive censoring samples $Z^*_{1:n_i} \dots Z^*_{n_i:n_i}$ and $Y^*_{1:m_i} \dots Y^*_{m_i:m_i}$ from $GLFRD(a, b, \alpha_i^*)$ and $GLFRD(a, b, \beta_i^*)$ respectively.
5. Find the maximum likelihood estimators of α_i^*, β_i^* and R^*.
6. Repeat step 3 and 4 B-times.
7. Sort R_i^* , $i = 1 \dots B$ in ascending order $R^*_{(1)} \dots R^*_{(B)}$.
8. Compute the approximate $(1-\alpha)$% Boot-P confidence interval of R as $(\widehat{R}_{Boot-p(\alpha/2)}, \widehat{R}_{Boot-p(1-\alpha/2)})$ where $\alpha = 0.05$ and $\widehat{R}_{Boot-p}$ is the cumulative distribution of $R^*_{(i)}$ $i = 1 \dots B$.

2. Asymptotic confidence interval (ACI) of R

To obtain the asymptotic confidence interval of R, we must get the asymptotic distribution of R because the exact distribution of R does not exist. First, derive the asymptotic distribution of α_i, β_i, $i = 1 \dots N$. Then obtain the asymptotic distribution of R. To derive the asymptotic distribution, compute the Fisher information matrix of $(\alpha_1 \dots \alpha_N, \beta_1 \dots \beta_N)$ as

$$I(\alpha_1 \dots \alpha_N, \beta_1 \dots \beta_N) = \begin{bmatrix} I_{1,1} & \cdots & I_{1,2N} \\ \vdots & \ddots & \vdots \\ I_{2N,1} & \cdots & I_{2N,2N} \end{bmatrix}$$

$$= -\begin{bmatrix} E(\frac{\partial^2 Log(L)}{\partial \alpha_1^2}) & \cdots & E(\frac{\partial^2 Log(L)}{\partial \alpha_1 \partial \beta_N}) \\ \vdots & \ddots & \vdots \\ E(\frac{\partial^2 Log(L)}{\partial \alpha_N \partial \beta_1}) & \cdots & E(\frac{\partial^2 Log(L)}{\partial \beta_N^2}) \end{bmatrix}$$

Where,

$$E\left(\frac{\partial^2 Log(L)}{\partial \alpha_i^2}\right) = -\frac{n_i}{\alpha_i^2} + P_i, \quad P_i = NE\left(\sum_{j=1}^{n_i} r_{ij} \left(\frac{A_{ij}^{\alpha_i}(a,b) Ln\left(A_{ij}(a,b)\right)^2}{(1-A_{ij}^{\alpha_i}(a,b))^2}\right)\right), \ i = 1 \dots N.$$

$$E\left(\frac{\partial^2 Log(L)}{\partial \beta_i^2}\right) = -\frac{m_i}{\beta_i^2} + P'_i; \ P'_i = NE(\alpha \sum_{l=1}^{m_i} \grave{r}_{il} \left(\frac{A_{il}^{\beta_i}(a,b) Ln\left(\grave{A}_{il}(a,b)\right)^2}{(1-\grave{A}_{il}^{\beta_i}(a,b))^2}\right)), \ i = 1 \dots N.$$

$$E\left(\frac{\partial^2 Log(L)}{\partial \alpha_i \partial \beta_j}\right) = E\left(\frac{\partial^2 Log(L)}{\partial \alpha_i \partial \alpha_j}\right) = E\left(\frac{\partial^2 Log(L)}{\partial \beta_i \partial \beta_j}\right) = 0, i \neq j, i, j = 1 \dots N.$$

Theorem: $n_i \to \infty$ and $m_i \to \infty$, $i = 1 \dots N$, then $(\widehat{\Theta} - \Theta) \approx N(0, I^{-1})$. Where, $\widehat{\Theta} = (\widehat{\alpha}_1 \dots \widehat{\alpha}_N, \widehat{\beta}_1 \dots \widehat{\beta}_N)$ and $\Theta = (\alpha_1 \dots \alpha_N, \beta_1 \dots \beta_N)$.

Proof: See, Ferguson[21].

The asymptotic distribution of R_{MLE} according to Delta method see Rao [22] and Wasserman [23] is $(R_{MLE} - R) \sim N(0, H^T I^{-1} H)$,

where

$$H^T = \left(\frac{\partial R}{\partial \alpha_1} \dots \frac{\partial R}{\partial \alpha_N}, \frac{\partial R}{\partial \beta_1} \dots \frac{\partial R}{\partial \beta_N}\right)$$

and,

$$\frac{\partial R}{\partial \alpha_i} = \frac{-\beta_1}{(\alpha_1+\beta_1)^2} + \sum_{i=2}^{N}\left(\left(\frac{-\beta_i}{(\alpha_i+\beta_i)^2}\prod_{j=1}^{i-1}\frac{\alpha_j}{(\alpha_j+\beta_j)}\right) + \left(\frac{\beta_i}{(\alpha_i+\beta_i)}\prod_{j=1}^{i-1}\frac{\beta_j}{(\alpha_j+\beta_j)^2}\right)\right)$$

$$\frac{\partial R}{\partial \beta_i} = \frac{\alpha_1}{(\alpha_1+\beta_1)^2} + \sum_{i=2}^{N}\left(\left(\frac{\alpha_i}{(\alpha_i+\beta_i)^2}\prod_{j=1}^{i-1}\frac{\alpha_j}{(\alpha_j+\beta_j)}\right) + \left(\frac{\beta_i}{(\alpha_i+\beta_i)}\prod_{j=1}^{i-1}\frac{-\alpha_j}{(\alpha_j+\beta_j)^2}\right)\right)$$

Hence, an asymptotic 100(1-α)% confidence interval of R

$$\left(R_{MLE} - Z_{\alpha/2}\sqrt{Var(R_{MLE})}, R_{MLE} + Z_{\alpha/2}\sqrt{Var(R_{MLE})}\right)$$

Where $Z_{\alpha/2}$ the upper $\frac{\alpha}{2}$th quantile of standard normal distribution and $Var(R_{MLE})$ is the variance at the maximum likelihood estimator.

III. Results & Discussion

Monte-Carlo simulation study is present for 1-2- cold standby redundancy system to compare the performance of different point estimators using biases, mean square error and relative efficiency. Also, the comparison of different confidence intervals is made using the average length (ACL) and converge probability (CP). We use different parameters and different censoring schemes. For Bayes estimator three prior as follows

Prior 1	$s_i = 0$	$t_i = 0$	$i = 1,2$	(non- informative gamma prior)
Prior 2	$s_i = 1$	$t_i = 2$	$i = 1,2$	(informative gamma prior)
Prior 3	$s_i = 2$	$t_i = 3$	$i = 1,2$	(informative gamma prior)

We use three censoring schemes as

r_1	$(0,0,...,M-m)$
r_2	$(M-m,0,...,0)$
r_3	All take the same number

For the population parameters, we assume $\alpha_1 = \alpha_2 = \alpha$, $\beta_1 = \beta_2 = \beta$. Also, the censoring schemes of Z_1, Z_2, Y_1, Y_2 are the same i.e. $n_1 = n_2 = m_1 = m_2 = m = 10$, $N_1 = N_2 = M_1 = M_2 = M = 30$. Biases, MSE's and relative efficiency of two-point estimators are computed. For interval estimation, average converge length (A.C.L) and converge probability (Cp) are computed. To perform the simulation study, we use the following algorithm:

1. Use different value $(a, b, \alpha_1, \alpha_2, \beta_1, \beta_2)$= (0.5,1,1,1,1,1), (1,2,1,2,2,3) and (2,1,3,2,1.5,2.5).
2. Use the following three censoring schemes (CS)

$$r_1 = (0,0,0,020,0,0,0,0,0,)$$

$$r_2 = (20,0,0,0,0,0,0,0,0,0)$$

$$r_3 = (2,2,2,2,2,2,2,2,2,2)$$

3. For a, b use the numerical method to solve equation 5 and 6 to get the maximum likelihood estimators for $(\alpha_1, \alpha_2, \beta_1, \beta_2)$.
4. Use equation 7 to calculate the maximum likelihood estimator of R.

5. Use equation 8 to calculate the Bayes estimator of R as

$$R_{Bayes} = R_{MLE} + \frac{1}{2}[(R_{11} + 2\,R_1\,\rho_1)\sigma_{11} + (R_{22} + 2\,R_2\,\rho_2)\sigma_{22} + (R_{33} + 2\,R_3\,\rho_3)\sigma_{33} +$$

$$(R_{44} + 2\,R_4\,\rho_4)\sigma_{44}] + \frac{1}{2}[L_{111}\,R_1\sigma_{11}\sigma_{11} + L_{222}\,R_2\sigma_{22}\sigma_{22} + L_{333}\,R_3\sigma_{33}\sigma_{33} +$$

$$L_{444}\,R_4\sigma_{44}\sigma_{44}]$$

Where,

$R_1 = \frac{\partial R}{\partial \alpha_1} = \frac{-\beta_1}{(\alpha_1+\beta_1)^2}$, $R_2 = \frac{-\beta_1}{(\alpha_1+\beta_1)^2} - \frac{\beta_2\,\alpha_1}{(\alpha_1+\beta_1)(\alpha_2+\beta_2)^2} + \frac{\beta_1\,\beta_2}{(\alpha_2+\beta_2)(\alpha_1+\beta_1)^2}$, $R_3 = \frac{\alpha_1}{(\alpha_1+\beta_1)^2}$,

$R_4 = \frac{\alpha_1}{(\alpha_1+\beta_1)^2} + \frac{\alpha_2\,\alpha_1}{(\alpha_1+\beta_1)(\alpha_2+\beta_2)^2} - \frac{\alpha_1\,\beta_2}{(\alpha_2+\beta_2)(\alpha_1+\beta_1)^2}$, $R_{11} = \frac{\partial^2 R}{\partial \alpha_1^2} = \frac{2\,\beta_1}{(\alpha_1+\beta_1)^3}$,

$R_{22} = \frac{\partial^2 R}{\partial \alpha_2^2} = \frac{-\beta_2(-2\,\alpha_1^2 - 2\,\alpha_1\,\beta_1 + \beta_1(\alpha_2+\beta_2))}{(\alpha_1+\beta_1)^2(\alpha_2+\beta_2)^3}$, $R_{33} = \frac{\partial^2 R}{\partial \beta_1^2} = \frac{-2\,\alpha_1}{(\alpha_1+\beta_1)^3}$, $R_{44} = \frac{\partial^2 R}{\partial \beta_2^2} = \frac{-\alpha_1\alpha_2\,(2\,\alpha_1+\alpha_2+2\,\beta_1+\beta_2)}{(\alpha_1+\beta_1)^2(\alpha_2+\beta_2)^3}$,

$\rho_1 = \frac{\partial \rho}{\partial \alpha_1} = \frac{t_{11}-1}{\alpha_1} - s_{11}$, $\rho_2 = \frac{\partial \rho}{\partial \alpha_2} = \frac{t_{21}-1}{\alpha_2} - s_{21}$, $\rho_3 = \frac{\partial \rho}{\partial \beta_1} = \frac{t_{12}-1}{\beta_1} - s_{12}$, $\rho_4 = \frac{\partial \rho}{\partial \beta_2} = \frac{t_{22}-1}{\beta_2} - s_{22}$,

$$L_{111} = \frac{\partial^3 \text{Log}\,L}{\partial \alpha_1^3} = \frac{2\,n_1}{\alpha_1^3} + \sum_{j=1}^{n_1} \frac{A_{ij}^{\alpha_1}(a,b)\left(1+A_{ij}^{\alpha_1}(a,b)\right)\text{Log}[A_{ij}(a,b)]^3}{(A_{ij}^{\alpha_1}(a,b)-1)^3},$$

$$L_{222} = \frac{\partial^3 \text{Log}\,L}{\partial \alpha_2^3} = \frac{2\,n_2}{\alpha_2^3} + \sum_{j=1}^{n_2} \frac{A_{ij}^{\alpha_2}(a,b)\left(1+A_{ij}^{\alpha_2}(a,b)\right)\text{Log}[A_{ij}(a,b)]^3}{(A_{ij}^{\alpha_2}(a,b)-1)^3},$$

$$L_{333} = \frac{\partial^3 \text{Log}\,L}{\partial \beta_1^3} = \frac{2\,m_1}{\beta_1^3} + \sum_{l=1}^{m_1} \frac{\grave{A}_{il}^{\beta_1}(a,b)\left(1+\grave{A}_{il}^{\beta_1}(a,b)\right)\text{Log}[\grave{A}_{il}(a,b)]^3}{(\grave{A}_{il}^{\beta_1}(a,b)-1)^3},$$

$$L_{444} = \frac{\partial^3 \text{Log}\,L}{\partial \beta_2^3} = \frac{2\,m_2}{\beta_2^3} + \sum_{l=1}^{m_2} \frac{\grave{A}_{il}^{\beta_2}(a,b)\left(1+\grave{A}_{il}^{\beta_2}(a,b)\right)\text{Log}[\grave{A}_{il}(a,b)]^3}{(\grave{A}_{il}^{\beta_2}(a,b)-1)^3},$$

6. Calculate bias, mean square error and relative efficiency as $B = E[\hat{R} - R]$, $MSE = E[\hat{R} - R]^2$ and $RE = \frac{MSE(R_{Bayes})}{MSE(R_{MLE})}$. If $RE > 1$, then the maximum likelihood estimator of R is more efficient than the Bayes estimator and if $RE < 1$, then the Bayes estimator is more efficient than the maximum likelihood estimator.

7. Calculate the BCI and ACI use section 4.1 and section 4.2 respectively. Also calculate ACL and CP to compare between two interval estimations.

Note that the results of simulation study based on 1000 replications. The results of point estimators show in Table (1), we observe the biases of Bayes estimator for three priors are less than the biases of maximum likelihood estimator but for the biases of three priors, we get the biases of the third prior is smallest. Table (2) shows mean square errors and relative efficiency, The Bayes estimator may be the best for some cases and the maximum likelihood estimator may be the best for another cases. Table (3) shows interval estimation, we get ACL for asymptotic confidence interval is less than its counterpart for bootstrap confidence interval and for CP, we get it is for ACI are closer to nominal value for BCI counterparts. So, we can decide the confidence interval is more efficient than the bootstrap confidence interval.

IV. Application

In this section, we show the implementation of point and interval estimation procedure proposed in this paper. We use a real data sets from Lawless [24]. The data sets represent failure time in minutes. For two types of electrical insulation in an experiment in which insulation is subjected to a continuously increasing voltage stress. Twelve electrical insulations of each type are tested and recorded. The failure time of the first type (z) are 21.8-70.7-24.4-138.6-151.9-75.3-12.3-95.5-98.1-43.2-28.6-16.9 and the failure times of the second type (Y) are 219.3-79.4-86.0-150.2-21.7-18.5-121.9-40.5-147.1-35.1-42.3-48.7. First, we must check wither the GLFRD fit to the data (Z, Y) or not. For this check, we use the Kolmogorov–Smirnov test (K-S).

The results show in Table (4) as. So, at significant level 0.05 cannot reject that the hypothesis that the data are coming from GLFRD. We check graphically the adequacy of the GLFRD to the real data. The probability plot in Figure 2 and Figure 3 shows an excellent goodness of fit of GLFRD. Also, Now, consider 1-2-cold-standby redundancy system consisting of the first electrical insulation and system consisting of a single second type electrical insulation. The maximum likelihood estimator, Bayes estimator and asymptotic confidence interval are computed for the probability R of 1-2-cold-stand by redundancy system consisting of the first type electrical insulation with longer life for different censored schemes and the results show in Table (5).

V. Conclusions

In this paper, we consider the multicomponent system of reliability called N-M-cold- standby redundancy system. Where the distributions for stress and strength variables are GLFRD with different shape parameter, Sarhan et al [25] studied the statistical properties of this distribution and find many physical interpretations. The reliability system is estimated by maximum likelihood method, Bayes estimator and the ACI and BIC are computed.

All estimators calculated under progressive type-II censoring data. Simulation study is performed. It is showed that the bias of Bayes estimator is less than the bias of maximum likelihood estimator. The MSE and RE showed that the MLE may be more efficient for some cases and Bayes estimator is more efficient for another cases. In the context of interval estimation, the comparison between ACI and BCI is made. Finally, we discuss the real data set represents failure time in minutes to illustrate the implementation of point and interval estimation procedures which proposed in this paper.

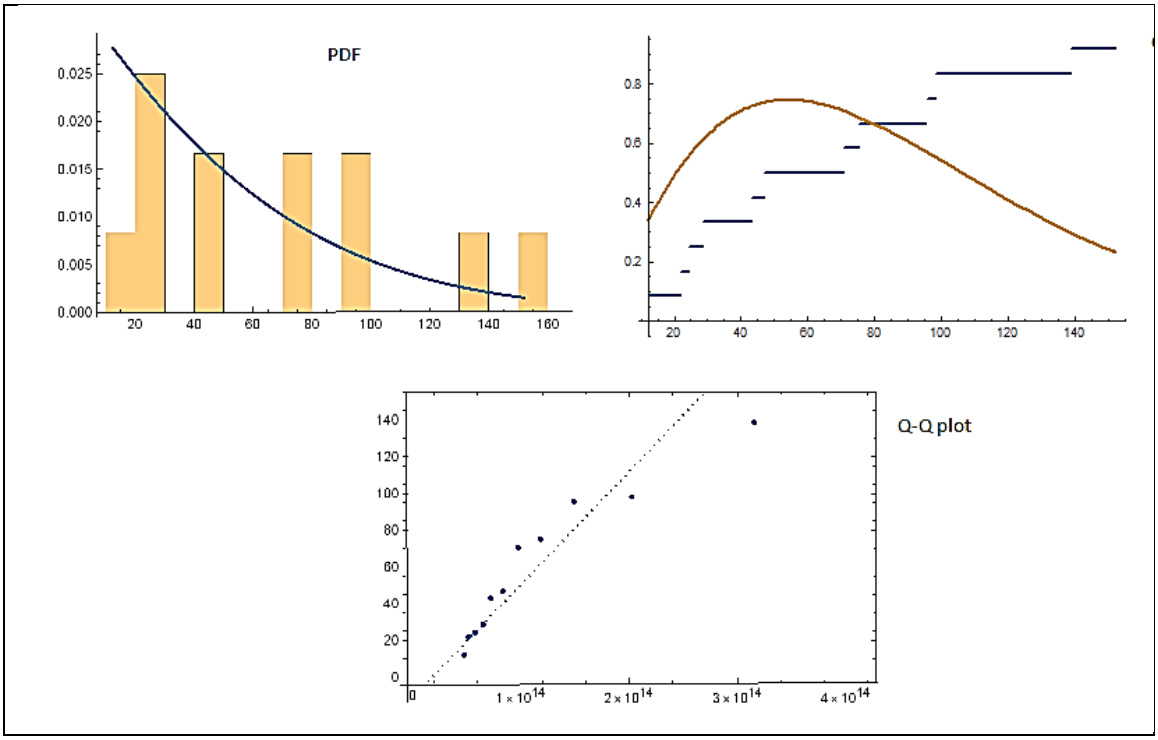


Figure (2): *Graphical goodness-of-fit on the failure time of the first type.*

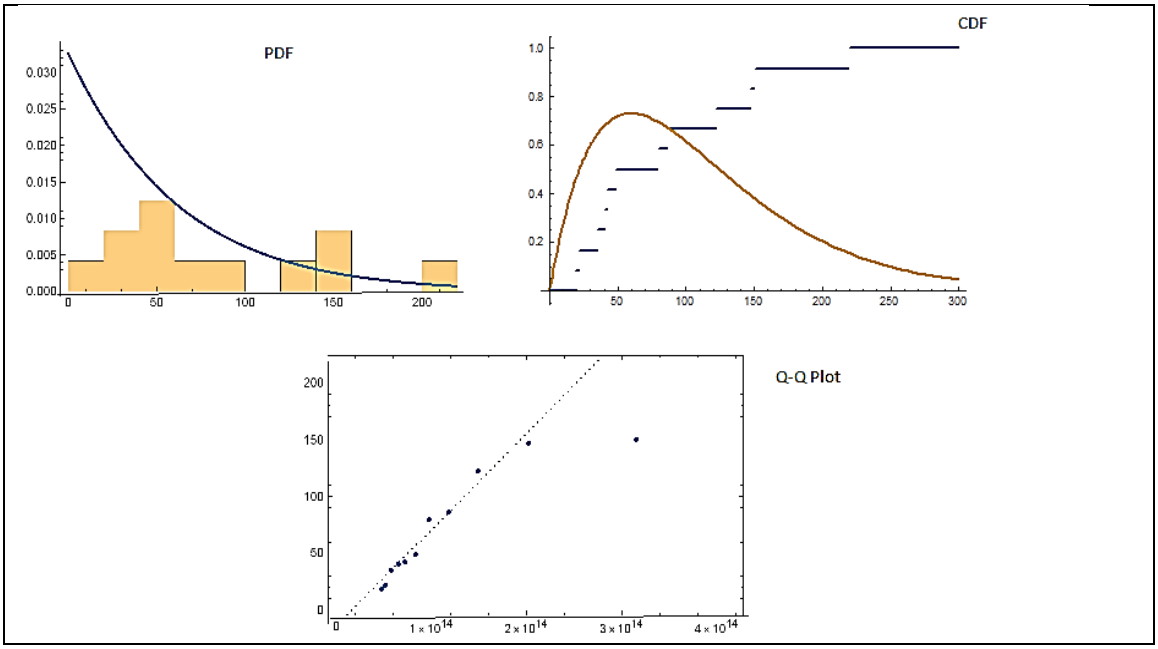


Figure (3): *Graphical goodness-of-fit on the failure time of the Second type.*

Table 1: *The Maximum likelihood and Bayes estimators of R and its biases*

		$a = 0.5,\quad b = 1,\ \alpha_1 = 1, \alpha_2 = 1,\ \beta_1 = 1,\ \beta_2 = 1$								
CS	$(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)$		**MLE**		**Prior (1)**		**Prior (2)**		**Prior (3)**	
		R_{true}	R_{MLE}	**Bias**	R_{Bayes}	**Bias**	R_{Bayes}	**Bias**	R_{Bayes}	**Bias**
(r_1, r_1)	(0.4,0.4,0.5,0.5)	0.75	0.8024	0.0262	0.7933	0.0216	0.7643	0.0071	0.7029	-0.0235
(r_1, r_2)	(0.44,0.44,0.36,0.36)		0.6975	-0.0262	0.7494	-0.0002	0.7205	-0.0147	0.6590	-0.0454
(r_1, r_3)	(0.38,0.38,0.41,0.41)		0.7686	0.0093	0.7085	-0.0207	0.6870	-0.0314	0.6486	-0.0506
(r_2, r_2)	(0.5,0.5,0.55,0.55)		0.7732	0.0116	0.7853	0.0167	0.7803	0.0151	0.7685	0.0092
(r_2, r_3)	(0.37,0.37,0.47,0.47)		0.8059	0.0279	0.7248	-0.0125	0.6759	-0.0370	0.5836	-0.0831
(r_3, r_3)	(0.41,0.41,0.42,0.42)		0.7559	0.0029	0.7354	-0.0072	0.7303	-0.0098	0.7206	-0.0145
		$a = 1,\quad b = 2,\ \alpha_1 = 1,\ \alpha_2 = 2, \beta_1 = 2,\ \beta_2 = 3$								
(r_1, r_1)	(0.35,0.33,0.37,0.54)	0.817	0.8445	0.0137	0.6104	-0.1032	0.5452	-0.1358	0.4328	-0.1920
(r_1, r_2)	(0.36,0.39,0.39,0.46)		0.8136	-0.0016	0.7374	-0.0397	0.7065	-0.0552	0.6509	-0.0830
(r_1, r_3)	(0.359,0.33,0.35,0.44)		0.8212	0.0021	0.6607	-0.0781	0.6378	-0.0895	0.6025	-0.1072
(r_2, r_2)	(0.35,0.38,0.35,0.48)		0.8503	0.0166	0.6322	-0.0923	0.5784	-0.1192	0.4884	-0.1642
(r_2, r_3)	(0.384,0.326,0.31,0.47)		0.7983	-0.0093	0.9302	0.0566	0.9420	0.0625	0.9563	0.0696
(r_3, r_3)	(0.385,0.371,0.383,0.492)		0.8200	0.0015	0.7695	-0.0237	0.7455	-0.0357	0.7007	-0.0581
		$a = 2,\quad b = 1,\ \alpha_1 = 3,\quad \alpha_2 = 2,\ \beta_1 = 1.5,\ \beta_2 = 2.5$								
(r_1, r_1)	(0.36,0.387,0.376,0.319)	0.809	0.6467	-0.0811	0.8562	0.0236	0.8972	0.0441	0.9581	0.0746
(r_1, r_2)	(0.309,0.338,0.339,0.355)		0.7085	-0.0502	02894	-0.2598	0.2239	-0.2925	0.1245	-0.3422
(r_1, r_3)	(0.356,0.394,0.363,0.375)		0.8094	0.0002	0.7853	-0.0118	0.7845	-0.0122	0.7830	-0.0129
(r_2, r_2)	(0.332,0.385,0.335,0.363)		0.8981	0.0445	0.9229	0.0569	0.9527	0.0718	0.9993	0.0951
(r_2, r_3)	(0.330,0.393,0.391,0.332)		0.8357	0.0133	0.7894	-0.0097	0.7812	-0.0138	0.7643	-0.0223
(r_3, r_3)	(0.320,0.356,0.362,0.336)		0.7033	0.0528	0.5245	-0.1422	0.4965	-0.1562	0.4572	-0.1788

Table 2: *MSE for The Maximum likelihood and Bayes estimators of R and RE*

CS	MSE				RE		
	MLE	Bayes					
		Prior (1)	Prior (2)	Prior (3)	RE1	RE2	RE3
$a = 0.5,$ $b = 1,\ \alpha_1 = 1, \alpha_2 = 1,\ \beta_1 = 1,\ \beta_2 = 1$							
(r_1, r_1)	0.0006	0.0004	0.0001	0.0005	0.6666	0.1666	0.8333
(r_1, r_2)	0.0006	0.0005	0.0002	0.0020	0.8333	0.3333	3.333
(r_1, r_3)	0.0001	0.0004	0.0009	0.0025	4	9	25
(r_2, r_2)	0.0001	0.0002	0.0002	0.0001	2	2	1
(r_2, r_3)	0.0007	0.0001	0.0013	0.0069	0.1428	1.85	9.8
(r_3, r_3)	0.0001	0.00005	0.00009	0.0002	5	9	2
$a = 1,$ $b = 2,\ \alpha_1 = 1,\ \alpha_2 = 2, \beta_1 = 2,\ \beta_2 = 3$							
(r_1, r_1)	0.0001	0.0106	0.0184	0.0368	106	184	368
(r_1, r_2)	0.0002	0.0015	0.0030	0.0068	7.5	15	34
(r_1, r_3)	0.0004	0.0061	0.0080	0.0115	15.25	20	28.75
(r_2, r_2)	0.0002	0.0085	0.0142	0.0269	42.5	71	14.5
(r_2, r_3)	0.0001	0.0032	0.0039	0.0048	32	39	40
(r_3, r_3)	0.0003	0.0005	0.0012	0.0033	1.6	4	71
$a = 2,$ $b = 1,\ \alpha_1 = 3,$ $\alpha_2 = 2,\ \beta_1 = 1.5,\ \beta_2 = 2.5$							
(r_1, r_1)	0.0065	0.0005	0.0019	0.0055	0.07	0.29	0.84
(r_1, r_2)	0.0025	0.0674	0.0855	0.1171	26.9	34.2	46.84
(r_1, r_3)	0.0001	0.0001	0.0001	0.0001	1	1	1
(r_2, r_2)	0.0019	0.0032	0.0051	0.0090	1.6	2.6	4.7
(r_2, r_3)	0.0001	0.0001	0.0002	0.0004	1	2	4
(r_3, r_3)	0.0027	0.0202	0.0244	0.0319	7.4	8.8	11.8

Table 3: *Results of interval estimation*

CS	ACI		BCI	
	ACL	CP	ACL	CP
$a = 0.5,$ $b = 1,\ \alpha_1 = 1, \alpha_2 = 1,\ \beta_1 = 1,\ \beta_2 = 1$				
(r_1, r_1)	0.1170	0.9661	0.3344	0.9052
(r_1, r_2)	0.1344	0.9579	0.4498	0.8640
(r_1, r_3)	0.1178	0.9650	0.3581	0.8965
(r_2, r_2)	0.1449	0.9571	0.3840	0.8876
(r_2, r_3)	0.1104	0.9681	0.3246	0.9076
(r_3, r_3)	0.1244	0.9626	0.3820	0.8903
$a = 1,$ $b = 2,\ \alpha_1 = 1,\ \alpha_2 = 2, \beta_1 = 2,\ \beta_2 = 3$				
(r_1, r_1)	0.0838	0.9765	0.3376	0.9047
(r_1, r_2)	0.0980	0.9719	0.3536	0.8991
(r_1, r_3)	0.0921	0.9737	0.3593	0.8977
(r_2, r_2)	0.0756	0.9789	0.3739	0.8899
(r_2, r_3)	0.1077	09687	0.4609	0.8637
(r_3, r_3)	0.1007	0.9712	0.3835	0.8898
$a = 2,$ $b = 1,\ \alpha_1 = 3,$ $\alpha_2 = 2,\ \beta_1 = 1.5,\ \beta_2 = 2.5$				
(r_1, r_1)	0.1290	0.9582	0.3500	0.8931
(r_1, r_2)	0.1216	0.9622	0.3185	0.9024
(r_1, r_3)	0.1103	0.9682	0.3687	0.8695
(r_2, r_2)	0.1051	0.9719	0.4323	0.8695
(r_2, r_3)	0.1007	0.9716	0.3261	0.9022
(r_3, r_3)	0.1186	0.9630	0.2747	0.9175

Table 4: ***(K-S) Test***

Data Set	Test Statistic	P-Value
Z	1	0.607
Y	5	0.082

Table 5: *Results of application*

CS	MLE	Bayes Estimator			ACI
		Prior (1)	Prior (2)	Prior (3)	
(r_1, r_1)	0.7420	0.8605	0.9394	0.9999	(0.7416,0.7709)
(r_1, r_2)	0.9666	0.1302	0.2895	0.3820	(0.8516,1)
(r_1, r_3)	0.9617	0.3648	0.1186	0.0523	(0.9579,0.9654)
(r_2, r_2)	0.9019	0.9922	0.7092	0.6240	(0.8935,0.9103)
(r_2, r_3)	0.9707	0.8009	0.8127	0.8245	(0.7511,1)
(r_3, r_3)	0.7510	0.7804	0.7366	0.7178	(0.7494,0.7539)

References

[1] Birnbaum, Z. W. (1956). On a use of Mann-Whitney statistics. in Proc. 3rd Berkeley Symposium on Mathematical Statistics and Probability,**1:** 13-17. https://projecteuclid.org/euclid.bsmsp/1200501643.

[2] Kotz, S., Lumelskii, I. , Pensky, M.(2003). The Stress- Strength Model and its Generalizations: Theory and Applications. Singapore: World Scientific. https://doi.org/10.1142/5015.

[3] Basirat, M., Baratpour, S. , Ahmadi, J. (2016). On estimation of stress-strength parameter using record values from proportional hazard rate models. Communication in Statistics- Theory and Methods, **45(19):** 5787-5801. https://doi.org/10.1080/03610926.2014.948727.

[4] Asgharzadeh, A., Kazemi, M. , Kundu D. (2017). Estimation of P (X > Y) for Weibull distribution based on hybrid censored samples. *International Journal* of *System Assurance Engineering* and *Management*, **8(1):** 489-498. https://doi:10.1007/s13198-015-0390-2.

[5] Hassan, M. (2017). Comparison of Different Estimators of P (Y < X) for Two Parameter Lindley Distribution. Journal of Reliability and Applications, 18(2): 83-98.

[6] Hassan, M. (2018). A New Application of Beta Gompertz Distribution in Reliability. Journal of Testing and Evaluation, 46(2): 736-744. https://doi.org/10.1520/JTE20160440.

[7] Bhattacharyya, G. k., Johnson, R. A.(1974). Estimation of reliability in a multicomponent stress-strength model. *Journal* of the *American* Statistical *Association,* **69(348):** 966–970.. https://www.jstor.org/stable/2286173.

[8] Hassan, M. , Alohali, M. (2018). Estimation of Reliability in a Multicomponent Stress-Strength Model Based on Generalized Linear Failure Rate Distribution. Hacettepe *Journal* of *Mathematics* and *Statistics,* **47 (6):** 1634-1651. .DOI: 10.15672/HJMS.2016.390.

[9] Sriwastav, G. , Kakati, M. (1981).A stress-strength model with redundancy, Indian Association of Productivity Quality & Reliability *Transactions*, **6(1):** 21–27

[10] Khan, A. G , Jan, T. (2014). Estimation of multi component systems reliability in stress-strength modelsThe Journal of Modern Applied Statistical Methods, **13(2):** 389–398.. DOI:10.22237/jmasm/1414815600.

[11] Ke, J. B., Lee, W. C. , Wang, K. H. (2007). Reliability and sensitivity analysis of a system with multiple unreliable service stations and standby switching failures, Physica A, 380(1): 455–469.. https://doi.org/10.1016/j.physa.2007.02.095.

[12] Gogoi, J., Borah, M. , Sriwastav, G. (2010).An interference model with number of stresses a Poisson process, Indian Association of Productivity Quality & Reliability *Transactions*, **34 (2):** 139–152.
[13] Liu, Y., Yimin, S., Xuchao, B. , Pei, Z. (2018). Reliability estimation of a N-M-cold-standby redundancy system in a multicomponent stress–strength model with generalized half-logistic distribution. Physica A., **490:** 231- 249...https://doi.org/10.1016/j.physa.2017.08.028.
[14] Sarhan, A.,,Kundu, D. (2009). Generalized linear failure rate distribution, Communication in Statistics- Theory and Methods, **38:** 642 – 660.. https://doi.org/10.1080/03610920802272414.
[15] Balakrishnan, N., Aggarwala, R. (2000). Progressive Censoring: Theory, Methods, and Applications, Springer Science & Business Media. DOI.10.1007/978-1-4612-1334-5.
[16] Balakrishnan, N. (2007). Progressive censoring methodology: An appraisal, Test, **16:** 211–259.. https://doi.org/10.1007/s11749-007-0061-y.
[17] Krishna, H. , Kumar, K.(2011). Reliability estimation in Lindley distribution with progressively type II right censored sample. Mathematics and Computers in Simulation ,**82(2):** 281–294..https://doi.org/10.1016/j.matcom.2011.07.005
[18] Lindley, D. V. (1980). Approximate Bayes method. Trabajos de Estadistica, (3): 281-288.. https://doi.org/10.1007/BF02888353.
[19] Ahmad, K. E., Fakhry, M. E., Jaheen, Z. F. (1997). Empirical Bayes estimation of P(Y <X) and characterizations of the Burr-typeX model. Journal of *Statistical Planning* and *Inference*, **64***:* 297-308.. https://doi.org/10.1016/S0378-3758(97)00038-4.
[20] Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. In:CBMS-NSF Regional Conference Series in Applied Mathematics,38, SIAM, Phiadelphia, PA.. http://dx.doi.org/10.1137/1.9781611970319.
[21] Ferguson T. (1967). Mathematical statistics: A decision theoretic approach, academic press, New Yourk.. https://doi.org/10.1002/bimj.19700120322.
[22] Rao, C. R. (2009). Linear Statistical Inference and its Applications, John Wiley and Sons
[23] Wasserman L. (2004). All of Statistics, Springer, New York. DOI.10.1007/978-0-387-21736-9.
[24] Lawless, J. F.(2011). Statistical Models and Methods for Lifetime Data, John Wiley and Sons.
[25] Sarhan, A. M., Tadj, L., Al-Malki, S. (2008). Estimation of the Parameters of the Generalized Linear Failure Rate Distribution. Bulletin of Statistics and Economics. **S08:** 52-63.

Critical Review Of Rams Tools And Techinques For The Analysis Of Multi Component Complex Systems

Shanti Parkash*

•

Ph.D. Scholar, Department of Production and Industrial Engineering, National Institute of Technology, Kurukshetra-136119, INDIA, Shanti_62000012@nitkkr.ac.in.

P.C. Tewari

•

Professor, Department of Mechanical Engineering, National Institute of Technology, Kurukshetra-136119, INDIA, pctewari1@gmail.com.

Abstract

This work provides the critical review of usefulness of Reliability, Availability, Maintainability and Safety (RAMS) approaches in complex mechanical systems. A broad range of research works available such as articles, conference proceedings and books covering RAMS approaches in industries as well as in the field of research is critically reviewed. These include different tools, techniques and methods which may be helpful in qualitative as well as in quantitative analysis. It provides the informations about the past and current scenario of RAMS practices in industries as well as in research. In this work the authors look for certain articles which included two or more aspects of RAMS. Limited work is reported in the field of safety.

Keywords- Reliability, Availability, Maintainability, Safety, Decision Support System, Markov, Petri Nets.

1. Introduction

Modern day industrial scenario has improved the process of designing and manufacturing of the systems which are more complex, high capacity and cost. It requires the high availability at reasonable cost. The consequences of low availability of these systems led to desire for high reliability and Maintainability (Saraswat et al., 2008). Reliability, Availability, Maintainability and Safety (RAMS) are four system dimensions that are of great interest to system developers, engineers, logisticians, and users. It affects the utility and life-cycle costs of system collectively. RAMS analysis has becomes a dynamic field of research to measure the performance of any operational system as per its required features. However the accuracy of analysis largely depends upon how the issues and challenges related to RAMS are addressed while planning the operational strategies which enhance the system performability (Hameeda et al., 2012).This paper critically reviews the literature on RAMS tools and techniques which increases the system availability and reduces the running cost. **Reliability** may be defined as the probability of success of a system that will perform its intended function adequately for a specified period of time (ASQ, 2011).**Maintainability** is the ability of a system to restored its functional state using prescribed maintenance procedures (BS4778, 1991).**Availability** generally known as operational reliability, is the ability of an system to perform its adequate function over a stated period of time in the given environment (BS4778, 1991).

Plant availability is mainly depends upon the reliability and maintainability of the system during design as well as operational stage i.e. A= f(R, M). **Safety** is a dimension which is essential for product cycle. It may be defined as the state of being free from any harm or danger. Principles of safety management can be suitably applied to various industries such as automotive, aviation, refineries, healthcare, workplace and food quality (ASQ, 2011).RAMS assists in prioritizing system maintenance. These priorities will be given to those subsystems which has a high failure rate. The performance of system can be enhanced with RAMS analysis which utilized the best combination of failure and repair rate. If the system is unreliable and not available for long time then it reduces the efficiency of the plant. It leads to the failure of its production unit. RAMS plays an important role to reduce the cost of the plant which helps to achieve the break even point rapidly. Break-Even Analysis is best used as a preliminary planning tool. (Eti et al., 2006) minimized the system maintenance cost using the combination of RAMS. The break even analysis curve draws between cost and volume of product, when the plant is available for longer period of time than large production is done due to which time period is reduces to achieve the break-even point, similarly (R.C. Mishra et al., 2003) explains that maintenance in time can yield better cost control as compare to time- availability maintenance.

2. Literature Review

Among various tools and techniques for system performance modeling like FMEA, RAM, RCA, Quality Control Tools and RAMS etc. RAMS is considered as a tool which covers both engineering and management aspect of the plant. RAMS is a tool which evaluates the performance of system in design as well as in production stage. In this literature review, papers of last two decades are considered those include RAMS analysis using different tools and techniques.

Tsarouhas [2020] provided findings of RAM analysis of an ice cream industry. Mean Time Between Failure and Mean Time To Repair data of every system's equipment was obtained using Statistical and Pareto analysis. These data were collected from the maintenance log book of the plant which helps to further improve the performance of the plant. Moreover, this approach evaluates the next move of the engineers and managers with respect to the availability of the system.

Kumar et al. [2020] analysed the performance behavior of the Veneer layup system of a plywood manufacturing plant in terms of its operational availability. Petri Nets technique was used for modeling. The impact of the failure and repair rates of different subsystem on the availability of the system was investigated. Based on the results obtained they identified most critical subsystems which needs to be put on priority. The outcomes provide appropriate planning strategies to the maintenance engineers.

Swiderski et al. [2020] discussed semi-Markov and Markov models as one of the most popular tools to determine the system reliability. This works evaluate the individual effect of each element on the entire system which helps to calculate the total reliability of the system. In this comparison exponential form was assumed. The main focus was occurring on the inconsistencies obtained while preliminary assessment of the data is done and also on the diagnostics of the machine readiness.

Kumar et al. [2019] described an effort that has made to reduce the uncertainties and incidental shutdowns of the plant using RAM approach. In this work, Petri nets approach is used for performance modeling in a milk processing plant. They identified that centrifugal pump was the most critical subsystem. This paper also highlight that software used for modeling reduces the effort as compare to markov modeling. An attempt has been made to reduce the operation costs,

maintenance costs and enhance the production volume by providing proper maintenance strategies.

Berrouane et al. [2019] performed RAMS analysis using stochastic Petri nets modeling of oil and gas processing facilities. Small-sized P-N blocks are used to represent each component of the system. These blocks are communicated through Boolean algebra due to which structure become easily trackable and reduce the structural complexity of the system. This RAMS model is built with three solid features like time dependency, durability, and clear graphical structure.

Soltanali et al. [2018] discussed RAM approach in the automotive manufacturing industries. In this work, RAM analysis was occurred on the automotive assembly line which shows that forklift and loading equipments having the main bottlenecks on this line. This analysis helps to find the maintenance schedules which improve the availability of the plant.

Wu et al. [2018] introduced an approach using Safety Instrumented System (SIS) with time dependent failure rates for reliability assessment. Weibull distributions were adopted for modeling and failure rate evaluation done by Approximation formulas. Validation of modeling was done using Petri Nets.

Corvaro et al. [2017] provided a study on the reciprocating compressor systems which were used in the gas and oil processing plants. The study was based on the RAM analysis of each part of reciprocating compressor which affects the availability of the plant. The main aim of the study to provided the better maintenance planning as compare to previous ones, so that the availability of the plant is increases.

Bosseet al. [2016] described the multi-objective redundancy allocation problem for IT services where high availability was required by the customer. This deals with cost and availability using Monte Carlo and Petri Nets for solving these multi objective problems. The approach was used by IT services for their feasibility and suitability of design.

Zhang et al. [2015] considered reliability and cost simultaneously for the individual component under practical condition. Particle Swarm Optimization was used to solve this multi objective problem. A relation for interval-valued function was formed using interval-valued numbers. This approach presented in the case study of water management system by Super Supervisory Control and Data Acquisition (SCADA).

Mortezaet al. [2014] described the new methodology for complex system for reliability design. The simulation and modeling of the system was accomplished using Monte Carlo and Reliability Block Diagram (RBD) method, this methodology evaluates availability and reliability of the system in the design stage. Drilling equipment was used for testing the proposed methodology which verifies the failure and repair data for evaluating the performance of the systems.

Ravinder et al. [2014] presented the performance of a 210 MW thermal power plant. Both economic and thermodynamics analysis were carried out to predicted mass flow of steam, equipment cost, fuel cost, consumption rate of coal and overall thermal efficiency of the plant. This works found that the effect of condensate extraction pump was more sensitive than boiler feed pump on net present value.

Liu et al. [2013] discussed double 2-out-of-2 system which obtained time dependent safety and reliability of the system. Markov modeling was used to evaluate the performance analysis of the system. This work concluded that safety and reliability affects by weak diagnostic as compare to common cause of failure

Wolde et al. [2013] stated the inspections and maintenance problem of railway carriers. This research relates failure and repair rate with the performance of a system using mathematical modeling. This modeling was applied to any system to evaluate inspection plans which further optimized its cost.

Kumar et al. [2012] evaluated the availability of a thermal power plant using markovian approach. This simulation model was used to predict the performance of the system using failure and repair rates of their subsystems. The availability of the power plant may be further increased with providing preventive maintenance for critical subsystems.

Vora et al. [2011] described stochastic and performance analysis of condensate system of a thermal plant. Markov model has been developed for six subsystems using transition diagram. This modeling evaluates the Performance level of each subsystem and finds the critical subsystem. These results help the management to take future decisions.

Andre and Vitali [2010] evaluated the availability of critical systems using Stochastic Petri Nets (SPN) model because Traditional methods to evaluate system reliability such as markov chain were not suited to the non-Poisson failures. This modeling was applied on the case study of electronics airbag controller which calculates the availability on demand.

Sachdeva et al. [2009] applied Petri nets (PN) and markovian approach for modeling and performance analysis for a feeding system of a paper industry respectively. PN used for modeling the active or standby equipments in the system. Methodology of this work provides a better maintenance planning to the management which reduced the operating and maintenance cost.

Gupta et al. [2008] analyzed the performance and provided a Decision Support System (DSS) for the feed water system of a thermal power plant. The DSS for this system were developed with the help of markovian approach. Such decision matrices can helps the maintenance department to evaluate the maintenance intervals.

Javad et al. [2007] identified the chokepoints in the system and find the critical subsystems in the crushing plant at a Bauxite Mine. This paper presents a case study of system comprises of six subsystems in which parameters such as Exponential, Lognormal and Weibull distributions had been considered using Weibull++6 version software. The results also show the reliability and availability analysis of the findings were very useful for maintenance decisions making.

Marseguerra et al. [2006] presented the methodology of RAMS analysis and performed reliability redundancy allocation and maintenance evaluation using Genetic Algorithm (GA). These multi-objective optimization problems were formulated for further analysis and evaluation.

Bertolini et al. [2006] described a technique which used Failure Mode, Effects & Criticality Analysis methodology. Reliability and maintenance policies were evaluated by Petri Nets simulation. This combination of FMECA and Petri Nets simulation for reliability modeling had been applied to an Italian oil refinery. This technique proving to be a very useful which further validated by maintenance experts.

Samrout et al. [2005] suggested to minimize the Preventive Maintenance (PM) cost using Ant Colony Optimization (ACO) of series–parallel systems. This works makes a comparison of an Ant Colony Algorithms and Classic Genetic Algorithm in detail. A hybrid algorithm was developed by the ACO and CGA combination which further evaluate the PM cost.

Rauzy [2004] reported the six methods which compute the time dependent probabilities of Markov models. The methods like full matrix exponentiation, Euler method, Runge-Kutta method and Adams-Bash forth multi-steps methods of order 2 and 4 were discussed in detail, which concluded the Markov graphs with up to millions of transitions can be handled on computers in today's world.

Elegbede and Adjallah [2003] discussed the multi-objective methodology based on GA based approach. An experimental plan was formulated which maximizing the availability and minimizing the cost of repairable simultaneously. The methodology has the following two main steps: (1) Working out a plan for calibrating the parameter (2) Selection and implementation of the GA parameters.

Avontur et al. [2002] presented reliability analysis of hydraulic and mechanical systems using finite element approach. A complete description of finite element approach was explained for these systems which show the solid and fluidic components with a single set of equations. These equations were also capable to describe non-linear behavior of the systems like non-return valves theory which produces comprehensive results for the designer.

Tang [2001] described a method which based on the concept of graph theory and Boolean algebra to determine the reliability in a process industry. The former approach used for derive the formula where as latter one was used for finding the failure interaction of two elements of the systems. This combined approach assessing the reliability of a complex system.

Cochran et al. [2001] discussed the decision making approach in the chemical plant availability during operation. This approach includes inspection routines and maintenance strategies. It uses generic Markov model which produce accurate results in exponential failure and repair rates.

Borgonovo et al. [2000] proposed the Monte Carlo modeling that suggested the flexible tool for the management and operation of plants such as system repair, renovation, aging, obsolescence etc. Economic constraints arising from safety and reliability requirements. This analysis evaluates the maintenance and operating procedures.

3. Research Gaps

In this section brief findings of literature are going to discuss using RAMS approach in the last two decades are:

1. Literature survey revealed that many researchers have done work on RAM approaches however the limited work is reported on RAMS. They generally neglect the important aspects of the plant i.e. safety. Safety is a paramount importance in any field as a harmless work environment, enhances the morale of team members working in any hazardous environment. To alleviate these conditions a separate Safety Instrumented System (SIS) has been installed in the plant. The SIS contains a set of hardware and software controls that are widely used in critical processes. SIS usually consists of logic solvers (s), sensors, final elements etc.
2. Many researchers reported their work to increase plant availability by adopting proper maintenance procedure. But few are reported the relation between the availability and break even analysis where as there is direct relationship between availability and break even analysis, when the plant is available for long time than break even point will achieve in short span of time.

3. From studying the literature survey it is found that several methods are used for quantitative and qualitative analysis of plant like reliability block diagram (RBD), fault tree analysis, Markov model, Petri nets etc. These methods have their own advantages and disadvantages like the **Markov chain** is a powerful modeling and analysis tool with robust applications in stochastic reliability and availability analysis. The major difficulty with this modeling approach is the explosion of a number of states even though it works with very small systems. **Petri Nets (PN)** contains Places, Transitions, Arcs, and Tokens. Tokens are stored locally on their places and these are travel from one place to another via arcs through transition. Petri nets have increased attention because of their simplicity and make a balance of modeling and decision-making power but increase the complexity of the system.
4. Generally, the RAMS approach used by the researchers in the operational stages of plant, very few reported in the design stages of the plant if this approach is used in design stages than availability of the plant is increases.

4. Conclusions And Future Scope

The critical overview of available literature show the usefulness of various RAMS tools and techniques applied in different plants, so as the cost of non availability of plant could be minimized. Every plant is divided into several systems or subsystems for appropriate maintenance planning which keeps the systems to remain available for long duration. The decision support framework under statistical analysis has been developed using various tools and techniques such as Reliability Hazard Analysis, Failure Mode and Effects Analysis (FMEA), Reliability Block Diagram, Fault Tree Analysis, Reliability Growth Analysis, Root Cause Analysis, Finite Element Analysis, Markov Analysis, Petri Nets etc. These techniques have their own advantages and disadvantages which has been discussion in the paper.

In future, time dependent failure and repair rates are consider which looks more appropriate to the plant because there is a continuous wear and tear occurs. Such problems are not sorted out by ordinary modeling and optimization techniques. Researchers can use such techniques like Genetic Algorithms (GA), Artificial Neutral Network (ANN), Particle Swam Optimization (PSO) and Fuzzy etc. to solve these multi objective optimization problems. These techniques are also merged with conventional modeling techniques to obtain multi objective optimization techniques like Markovian GA, Fuzzy Petri Nets approach etc. These techniques can be further programmed using MATLAB software. These newly developed modeling techniques can be applied in both design as well as operational stage to further increase the performability of the plant.

Acknowledgements: This paper would not be possible without the exceptional support of Director, National Institute of Technology, Kurukshetra. The authors would like to extend their deep appreciation of the suggestions and comments of the learned referees that have enabled them to improve the quality of the manuscript. I want to thank my parents, my wife and my son for their love and support through the time I writing this manuscript.

References

[1] Avontuur G. C. and Werff K. V., "Systems reliability analysis of mechanical and hydraulic drive systems", Reliability Engineering and System Safety, 2002, pp. 121–130.
[2] Barabady J., Kumar U., "Reliability analysis of mining equipment: A case study of a crushing plant at Jajarm Bauxite Mine in Iran", Reliability Engineering and System Safety, 2007, pp. 647–653.

[3] Berrouane M., Khan F.,Kamil M., "Dynamic RAMS Analysis Using Advanced Probabilistic Approach", Chemical Engineering Transactions, Vol. 77, 2019, pp.2241-2246.

[4] Bertolini M. and Bevilacqua M., "Reliability design of industrial plants using Petri nets", Journal of Quality in Maintenance Engineering, Vol. 12, 2006, pp. 397-411.

[5] Borgonovo E., Marseguerra M., Zio E., "A Monte Carlo Methodological Approach to plant Availability Modeling with Maintenance, Aging and Obsolescence". Reliability Engineering and System Safety, Vol. 67, 2000, pp. 61–73.

[6] Bosse S., Splieth M. and Turowski K., "Multi-objective optimization of IT service availability and costs" Reliability Engineering and System, 2016, pp.142–155.

[7] Cochran K.J., Murugan A. and Krishnamurthy," Generic markov models for availability, estimation and failure characterization in petroleum refineries". Computers and Operations Research, 2001, pp. 1–12.

[8] Corvaro F., Giacchetta G.,Marchetti B. and Recanati M., "Reliability, Availability, Maintainability (RAM) study, on reciprocating compressors API 618", ADVS, Vol.3, 2017, pp.266-272.

[9] Ebrahimi N.B., "Indirect assessment of System Reliability", IEEE Transactions on Reliability, Vol. 52, 2003.

[10] Gupta S. and Tewari P. C., "Simulation model for stochastic analysis and performance evaluation of steam generator system of a thermal power plant", International Journal of Engineering Science and Technology, Vol. 3, 2011, pp. 5141-5149.

[11] Gupta S., Tewari P.C. and Sharma A.K., "A Performance Modeling and Decision Support System for a pp. Feed Water Unit of a Thermal Power Plant", South African Journal of Industrial Engineering, Vol. 19(2), 2008, 125-134.

[12] Hameeda Z., Vatna J., "Important challenges for 10 MW reference wind turbine from RAMS perspective", Energy Procedia, 2012, pp. 263 – 270.

[13] Kleyner A., Volovoi V.,"Application of Petri nets to reliability prediction of occupant safety systems with partial detection and repair", Reliability Engineering and System Safety,2010.

[14] Kumar N., Tewari P.C. and Sachdeva A., "Petri Nets Modeling and Analysis of the Veneer Layup System of Plywood Manufacturing Plant" Engineering modeling 2020, pp. 95-107.

[15] Kumar N., Tewari P.C. and Sachdeva A., "Performance Modeling and Analysis of Refrigeration System of a Milk Processing Plant using Petri Nets", IJPE, 2019,Vol. 15, pp. 1751-1759.

[16] Kumar R., Sharma A. and Tewari P. C., "Thermal Performance and Economic Analysis of 210MW Coal-Fired Power Plant", Journal of Thermodynamics, 2014, pp.1-10.

[17] Kumar R., Sharma A. and Tewari P. C., "Markov approach to evaluate the availability simulation model for power generation system in a thermal power plant", International Journal of Industrial Engineering Computations, Vol. 3, 2012, pp.743–750.

[18] Laibin S., Mary Z., Yiliu A. and Zheng L. W., "Reliability assessment for final elements of SISs with time dependent failures" Journal of Loss Prevention in the Process Industries, Vol. 51, 2018, pp. 186-199.

[19] Liu Z., Liu Y., Cai B., Li J. and Tian X., "Markov Modeling of Double 2-out-of-2 System with Imperfect Detection and Common Cause Failures", International Conference on Computer, Networks and Communication Engineering, 2013.

[20] Marseguerraa M., Zioa E. and Martorell S., "Basics of genetic algorithms optimization for RAMS applications", Reliability Engineering and System Safety, Vol. 91, 2006, pp. 977–991.

[21] Rauzy A., "An experimental study on iterative methods to compute transient solutions of large Markov models", Reliability Engineering and System Safety, Vol.86, 2004, pp.105–115.

[22] Sachdeva A., Kumar P. and Kumar D., "Behavioral and performance analysis of feeding system using stochastic reward petri nets", IJAT, 2009, pp.156-169.

[23] Samrouta M., Yalaouia F., Chateleta E. and Chebbo N., "New methods to minimize the preventive maintenance cost of series–parallel systems using ant colony optimization", Reliability Engineering and System Safety, 2005, pp. 346–354.

[24] Saraswat S. and Yadava G.S., "An overview on reliability, availability, maintainability and supportability (RAMS) engineering", International Journal of Quality &Reliability Management, Vol. 25 No. 3, 2008, pp. 330-344.

[25] Soleimani M., Mohammad P., Rostami A., and Ghanbari A., "Design for Reliability of Complex System: Case Study of Horizontal Drilling Equipment with Limited Failure Data", Journal of Quality and Reliability Engineering, 2014.

[26] Soltanali H., Garmabaki A., Thaduri A., Parida A.,Kumar U. and Rohani A.,"Sustainable production process: An application of reliability, availability, and maintainability methodologies in automotive manufacturing", Journal of Risk and Reliability, 2018, pp.1-16.

[27] Swiderski A., Borucka A., Grzelak M. and Gil L., "Evaluation of Machinery Readiness Using Semi-Markov Processes", Appl. Sci. 2020, Vol.10, pp.1-15.

[28] Tang J., "Mechanical system reliability analysis using a combination of graph theory and Boolean function", Reliability Engineering and System Safety, Vol. 72, 2001, pp. 21-30.

[29] Tsarouhas P., "Reliability, Availability, and Maintainability (RAM) Study of an Ice Cream Industry", Appl. Sci. 2020, Vol.10, pp.1-20.

[30] Wolde M., Ghobbar A., "Optimizing inspection intervals Reliability and availability in terms of a cost model: A case study on railway carriers", Reliability Engineering and System Safety, 2013, pp. 137–147.

[31] Zhang E., Chen Q., "Multi-objective reliability redundancy allocation in an interval Environment using particle swarm optimization", Reliability Engineering and System Safety, 2015, pp. 83–92.

Effect Of Activation Function In Speech Emotion Recognition On The Ravdess Dataset

Komal D. Anadkat[1], Dr.Hiteishi M. Diwanji[2]

Information Technology department[12]
Government Engineering College[1], L.D.College of Engineering[2],
Gandhinagar, India[1], Ahmedabad, India[2]
komalanadkat@gecg28.ac.in[1], hiteishi.diwanji@gmail.com[2]

Abstract

Since last decade, Speech Emotion recognition has attracted extensive research attention to identify emotions by user's pitch and voice. Many research has been done in this field to recognize emotions using different machine learning as well as deep learning approaches. In this paper ,we tried three different machine learning algorithms named SVM, Logistic regression and Random forest which take four different features named MFCC,Chroma,Mel-scale spectrogram and tonnetz as an input on RAVDESS dataset where SVM is more accurate than others. As deep learning approaches are more capable to identify hidden patterns and classify the data more accurately, we tried popular algorithm like MLP,CNN and LSTM. In deep learning approach, activation function is one of the most dominant parameters which a designer can choose to make classification more accurate. In this paper, we tried to show the effect of different activation functions on the overall accuracy of the model and analyzed the results.

Keywords: CNN; MLP; activation function; neural network; deep learning; SER;

I. Introduction

Through all the inherently available senses, people can identify the emotional condition of their correspondence accomplice. This emotion detection is normal for people, yet it is troublesome undertaking for computers; Emotion is a way to express how an individual feels. Emotion recognition play a significant factor in many important areas, like health-care, education, and human resources. Emotion recognition is an intense errand as each individual has an alternate tone and pitch of voice [1].SER is the demonstration of endeavoring to perceive human feeling and emotional states from speech. This is gaining by the way that voice frequently reflects fundamental feeling through tone and pitch. Speech Emotion recognition (SER) is the common and quickest method of correspondence among people and PCs and assumes a significant part progressively uses of human-machine association. To recognize the emotion of the individual from voice is one of the interesting research area for various researcher. In SER, the robust and discriminative features selection and extraction is a most difficult task [2]. In this paper, we compare different machine learning algorithms like Linear SVM, Logistic regression and Random forest, with different neural network algorithms for classifying emotions from speech.

Along with all major problems in machine learning, SER has started to gain an advantage from the tools made available by deep learning. Feed-forward Neural Network takes input, processes input with 1 or more hidden layers and then outputs with the help of output layer at the end. In between each layer, no matter which type of layer it be, there is always an activation function which

decides what amount of input should be carried forward. Different neural networks serve different and specific purposes. For a case to consider, for image classification task, convolutional neural network (CNN) has been successful in achieving great results in various applications in machine learning domain in recent years [4] [5]. Whereas an artificial neural network with the capability of back-propagation learning helps in approximating multi-variate and non-linear kind of relationship between input(s) and output(s)[6].

One of the most important part of any neural network is the activation function [7] that helps decide if the current input and its setting allow what amount of input is to be carried forward in the neural network setting. Deep learning algorithms work just like our brains work and the activation functions act as if what is the limit to which a neuron can allow certain inputs to be tolerated, which is called threshold [8]. Activation function helps classifying input data into binary or multi-class based on certain threshold value, which is the basic use of an activation function of any neuron [9]. There are various types of activation functions, some of which are depicted below. It should be noted that this list and its details are not exhaustive.

II. Related Work

Many Researchers have previously done work on speech emotion recognition using different datasets and algorithms. Main two tasks of Speech emotion recognition are :- (i) Extracting and selecting the important and salient features from dataset and (ii) The selection of appropriate classifier which can classify the accurate emotions from the given speech. Many authors have used handcrafted features like MFCC, Contrast, Mels Spectrograph Frequency, Chroma and Tonnetz, while some authors have used deep learning approaches to improve the recognition accuracy.

I. Hand-Crafted Feature-Based Speech Emotion Recognition (SER)

Many researchers tried to efficiently recognize emotions from person's voice using handcrafted features. Audio signal have many features and so that it is major concern of researcher to select the appropriate features in the SER task. One researcher , Dave et al. [10] analyzed various speech emotions features and conclude that Mel frequency cepstral coefficient (MFCC) [11] features are better to use for SER than other low-level features like linear productivity code (LPC) [12], like formant, loudness,etc. On the other side Liu et al. [13] analyzed that to increase the Unweighted accuracy upto 3.6 %, it is better to use gamma tone frequency cepstral coefficient (GFCC) features using additional voice features like jitter and shimmer for SER. Liu et al. [14] proposed a novel method to extract the features from Chinese speech dataset (CASIA), using correlation and an extreme learning machine (ELM)-based decision tree for classification. Fahad et al. [15] proposed a technique which select glottal and MFCC features. These features fed to train a model based on deep neural network for SER. Wei and Zhao [16] proposed a novel method which used sparse classifier and auto encoders on a Chinese speech emotion dataset. To train support vector machine (SVM) classifier to recognize the emotion of a person from speech, author used auto encoder which extracts the large dimensions features and sparse network which extract small dimensions sparse features.

II. Deep Learning

Many researchers have worked to improve the efficiency of Speech emotion recognition using many different methods and using various datasets. Many of them used different neural network architecture and different pre trained networks to classify the emotions. Navya Damodar[17] have proposed novel approach to classify the emotions using CNN and decision tree. They used MFCC to extract the useful features from the Pre-processed audio files and applied these features with Decision tree as well as CNN both to compare the accuracy .They concluded that CNN outperforms Decision tree with 72% accuracy ,where the other achieved 63%. Jianfeng Zhao[18] have proposed

an approach which used Merged Deep CNN to learn deep features for speech emotion recognition. The merged deep CNN is the combination of 1D and 2D CNN models. These two models designed and evaluated individually and them merged. 1 D and 2D CNNs are designed to learn deep features from audio clips and from log-mel spectrograms respectively. Authors have used transfer learning to train the model which makes training faster. Fei and Liu [19] have proposed a method which uses advance long short-term memory (A-LSTM) to learn the sequences using pooling recurrent neural network (RNN) scheme in SER. Then they compare the results of A-LSTM with simple LSTM and found that former is better. Saurabh et al. [20] used IEMOCAP dataset and autoencoder approach for Speech emotion recognition. Hajar and Hasan [21] proposed a novel technique in which they split the audio signal into frames and extract MFCCs features. They used K-means clustering algorithm to select key spectrograms and used 3D CNN to predict speech emotions.

III. Proposed Methodology

Every individual express his/her emotion by face, speech, and text etc. The Speech emotion can be captured by tone and pitch of a person. In this paper we tried to recognize the different types of emotions of a person like sad, happy, angry , neutral ,disgust and surprised. In this paper the emotions in the speech are predicted using different machine learning algorithm along with neural networks. We have also enlighten the effect of different activation functions on accuracy. For analysis purpose we have used RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song dataset) dataset.

Fig.1 shows the work flow of speech emotion recognition. First, we load and read the RavDess dataset. In the next step, important features like MFCC, Mel, Chroma and Tonnetz are extracted from the data. In the third step, these extracted features become inputs for the training algorithm. Here we have used three different machine learning algorithms, Multilayer Perceptron, CNN and CNN-LSTM for the analysis purpose. When we used neural network algorithms, we tried to show the effect of different activation functions on the classification accuracy. Hence we compared each neural network on the bases of accuracy they achieved with different activation functions.

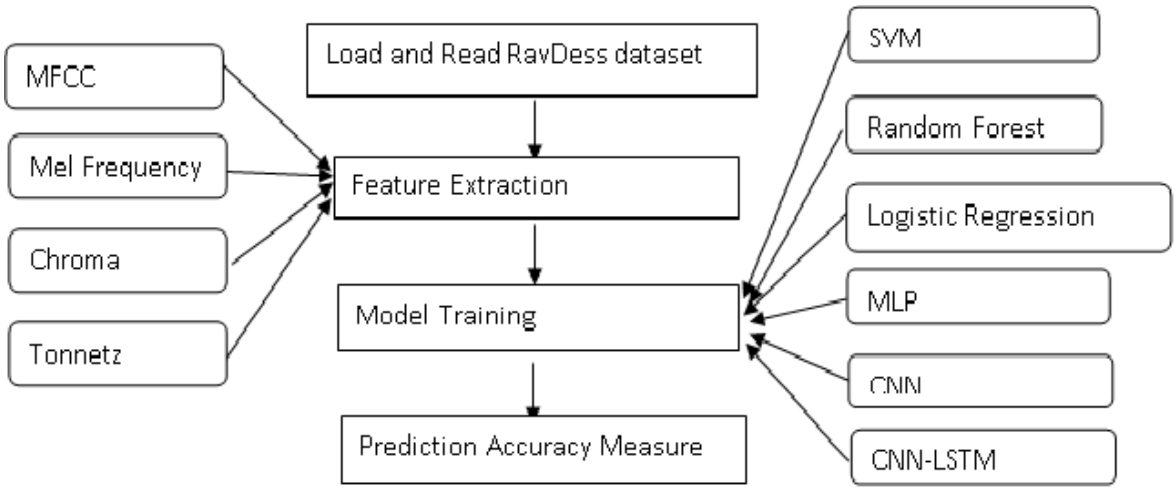


Figure 1: *Work Flow of Speech emotion recognition*

I. Traditional Methods

Before the era of deep neural networks, emotion recognition from speech has been mostly done

using machine learning and signal processing techniques. Machine learning algorithms like Decision trees, Random forest, SVM and logistic regression were used to classify the emotions automatically from speech.

Machine learning algorithm based solutions of SER problems need deep understanding of feature extraction and selection methods. Speech signals have many features available like Mel,Chroma, mel-frequency cepstrum coefficients (MFCC), modulation spectral features ,pitch,, energy, linear prediction coefficients (LPC) and Tonnetz .The main challenge here is to select the relevant and useful features for the classification task. After feature extraction, the next stage is classification. In the classification step, emotions will be classify in different class. Many classification algorithms are available and they can't be compare with each other as each of them has its own pros and cons. In this paper, we chose Linear SVM, logistic regression and Random Forest to compare the efficiency of three classifiers.

II. Multi-Layer Perceptron Classifier

Multi-layer Perceptron Classifier (MLP Classifier) is fall in the category of neural network used for the classification task. It uses MLP algorithm and trains model which uses Back-propagation to update the weights of neurons. Below are the steps to build MLP classifier:-

- In the first step, We need to initialize the parameters and define the classifier. Here we experimented different activation functions to find the effect on accuracy of the model.
- Then in the next step data is feed to neural network to train the model
- Now model trained in the step -2 is used to predict the output of the test data.
- Now Measure the accuracy of the prediction.

III. Convolutional Neural Networks

Convolutional neural networks (CNNs) are one of the most popular deep learning models that have manifested remarkable success in the research areas such as 14 object recognition , face recognition , handwriting recognition , speech recognition , and natural language processing . The term convolution comes from the fact that convolution—the mathematical operation—is employed in these networks. Generally, CNNs have three fundamental building blocks: the convolutional layer, the pooling layer, and the fully connected layer.

IV. Long Short Term Memory Networks

LSTM Networks are Recurrent neural network, which can learn long-term dependencies. These networks are used to solve many complex real world problems like speech recognition, object detection, facial expression recognition etc. the architecture of LSTM contains one input layer, one hidden layer and one output layer. The hidden layers have memory cells with gate units named the input, output and forget gates.

IV. Results and Discussion

I. Traditional methods Experiments

As A traditional methods, we have used SVM, Logistic regression and Random Forest. Table:-1 shows the comparison of three machine learning algorithm and it shows that SVM has achieved highest accuracy of 69.17% with very less training time. Other two algorithms have nearly performed well in terms of accuracy but random forest take more time in training and testing phase.

Table 1: *Performance Comparison of Tradition methodS*

Algorithm	Accuracy	Training Time(s)	Testing Time(s)
SVM	69.17%	0.024	0.001
Logistic Regression	67.10%	0.549	0.001
Random Forest	67.19%	3.413	0.155

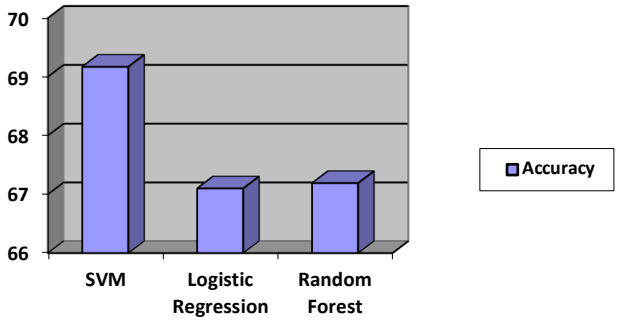


Figure 2: *Accuracy analysis of Tradition method*

II. MLP classifier Experiments

For MLP classifier, we use categorical cross entropy as our loss function. We use the Adam optimizer for optimizing our model. It combines various improvements to traditional stochastic gradient descent . Here highest accuracy achieved by Tanh activation function which was 81.77%, where Logistic also performs almost similar with 81.27% but it takes much time to train the model.

Table 2: *Effect Of Activation Functions On Accuracy Of MLP Classifier*

Activation Function	Accuracy	Training Time(s)	Testing Time(s)
Relu	71.35%	2.408	0.002
Tanh	81.77%	9.681	0.002
Logistic	81.27%	16.41	0.003
Identity	52.08%	1.172	0.002

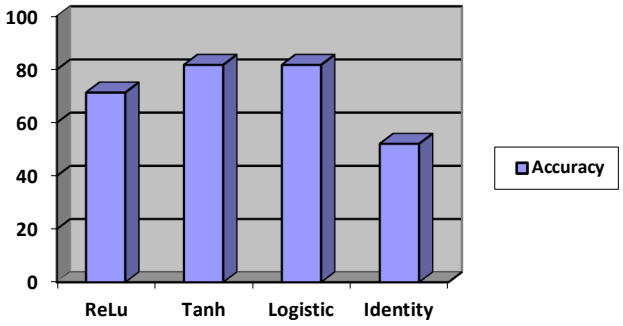


Figure 3: *Accuracy analysis of MLP classifier with different activation function*

III. Convolutional Neural Networks Experiments

The architecture followed here is 3 convolution layers followed by Max pooling layer, a fully connected layer and softmax layer respectively with 75 epochs. Multiple filters are used at each convolution layer, for different types of feature extraction. Here highest accuracy achieved by ReLu activation function which was 87.46%, where Tanh also performs almost similar with 85.89%.

Table 3: *Effect Of Activation Functions On Accuracy Of CNN Classifier*

Activation Function	Testing Accuracy	Training Accuracy
Relu	87.46%	96.50%
Tanh	85.89%	98.11%
Linear	75.61%	81.11%
Sigmoid	70.19%	75.40%

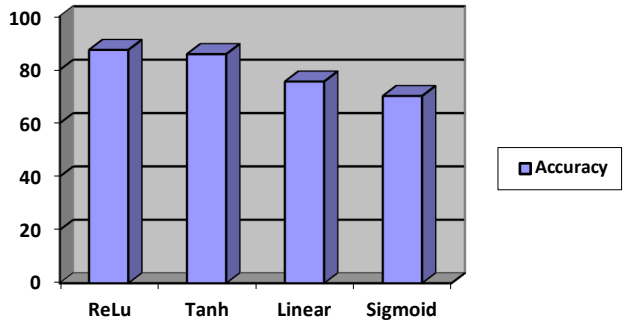


Figure 4: *Accuracy analysis of CNN classifier with different activation function*

IV. CNN-LSTM Experiments

Here we have created a 2D convolutional network as comprised of Conv2D and MaxPooling2D layers ordered into a stack of the required depth.We have defined a CNN LSTM model in Keras by first defining the CNN layer or layers, wrapping them in a TimeDistributed layer and then defining the LSTM and output layers. We have defined the CNN model first, then add it to the LSTM model by wrapping the entire sequence of CNN layers in a TimeDistributed layer. Here highest accuracy was achieved by ReLu activation function which was 84.97%.

Table 3: *Effect Of Activation Functions On Accuracy Of LSTM Classifier*

Activation Function	Testing Accuracy	Training Accuracy
elu	74.31%	90.81%
Tanh	54.67%	69.99%
Relu	84.97%	88.05%
Sigmoid	40.14%	50.95%

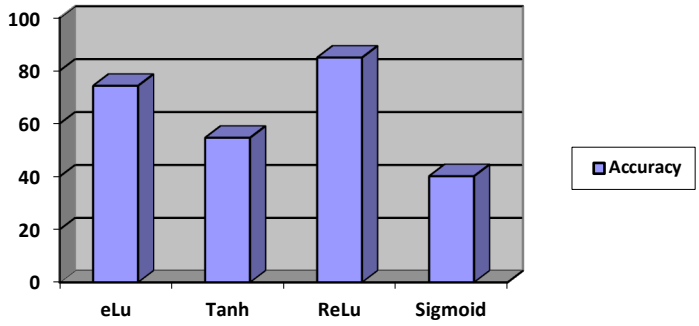


Figure 5: *Accuracy analysis of LSTM classifier with different activation function*

V. Conclusion & Future Work

Even though many works have been done for the speech emotion recognition area, still it faces many challenges. We applied three different machine learning algorithm named SVM, Logistic regression and Random forest on RevDess dataset where SVM outperforms with 69.17% accuracy and it takes less time to train the model. Then we applied ReLu, Tanh, Identity Logistic activation functions on MLP classifier where Tanh outperform with 81.77% of accuracy. After having ReLu, Tanh, Linear and sigmoid on CNN classifier, ReLu outperform with 87.46% accuracy. At Last we applied eLu, Tanh, Relu and sigmoid activation functions on CNN-LSTM model where ReLu outperform with 84.97% accuracy. Deep learning algorithms give better performance than traditional machine learning algorithms. These results of machine learning algorithm, MLP, CNN, LSTM are for Speech emotion recognition task on Revdess dataset. In future these results can be cross verified on other wide and real datasets of speech emotion recognition. Speech signals have

many features, here we have selected Mel, MFCC, Chroma and Tonnetz. In future, other features can also be used as a feature extraction methods and check whether it can improve the accuracy.

References

[1] Jerry Joy1, Aparna Kannan2, Shreya Ram3, S. Rama4," Speech Emotion Recognition using Neural Network and MLPClassifier ", IJESC,April-2020.

[2] Mustaqeem and Soonil Kwon * ,"A CNN-Assisted Enhanced Audio Signal Processing for Speech Emotion Recognition", Sensors 2020, 20, 183; doi:10.3390/s20010183.

[3] Dave, N. Feature extraction methods LPC, PLP and MFCC in speech recognition. Int. J. Adv. Res. Eng.Technol. 2013, 1, 1–4.

[4] Y. Yin, L. Want, E. Gelenbe, "Multi-Layer Neural Networks for Quality of Service oriented Server-State Classification in Cloud Servers", 2017, ISBN No. 978-1-5090-6182-2/17, pp. 1623-1627

[5] Y. Guo, L. Sun, Z. Zhang, H. He, "Algorithm Research on Improving Activation Function of Convolutional Neural Networks", 2019, ISBN No. 978-1-7281-0106-4/19, pp. 3582-3586

[6] Y. Zhang, Q. Hua, D. Xu, H. Li, Y. Bu, P. Zhao, "A complex valued CNN for different activation functions in polarsar image classification", 2019, IGARSS 2019, pp. 10023-100

[7] J. Demby, Y. Gao, G. N. DeSouze, "A study on Solving the Inverse Kinematics of Serial Robots using Aritifical Neural Network and Fuzzy Neural Network", 2019, ISBN No. 978-1-5386-1728-1/19

[8] S. Baraha, P. K. Biswal, "Implementation of Activation Functions for ELM based Classifiers", 2017, WiSPNET 2017, pp. 1038-1042

[9] S. Jeyanthi, M. Subadra, "Implementation of Single Neuron Using Various Activation Functions with FPGA", 2014, "International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), pp. 1126-1131

[10] Luque Sendra, A.; Gómez-Bellido, J.; Carrasco Muñoz, A.; Barbancho Concejero, J. Optimal Representation of Anuran Call Spectrum in Environmental Monitoring Systems UsingWireless Sensor Networks. Sensors 2018, 18, 1803. [CrossRef]

[11] Erol, B.; Seyfioglu, M.S.; Gurbuz, S.Z.; Amin, M. Data-driven cepstral and neural learning of features for robust micro-Doppler classification. In Proceedings of the Radar Sensor Technology XXII, Orlando, FL, USA, 16–18 April 2018; p. 106330J.

[12] Liu, G.K. Evaluating Gammatone Frequency Cepstral Coe_cients with Neural Networks for Emotion Recognition from Speech. arXiv 2018, arXiv:1806.09010.

[13] Liu, C.-L.; Yin, F.; Wang, D.-H.; Wang, Q.-F. CASIA online and o_ine Chinese handwriting databases.In Proceedings of the 2011 International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011; pp. 37–41.

[14] Liu, Z.-T.; Wu, M.; Cao, W.-H.; Mao, J.-W.; Xu, J.-P.; Tan, G.-Z. Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neurocomputing 2018, 273, 271–280. [CrossRef]

[15] Fahad, M.; Yadav, J.; Pradhan, G.; Deepak, A. DNN-HMM based Speaker Adaptive Emotion Recognition using Proposed Epoch and MFCC Features. arXiv 2018, arXiv:1806.00984.

[16] Zhu, L.; Chen, L.; Zhao, D.; Zhou, J.; Zhang, W. Emotion recognition from Chinese speech for smart services using a combination of SVM and DBN. Sensors. 2017, 17, 1694.

[17] Navya Damodar, Vani H Y, Anusuya M A. Voice Emotion Recognition using CNN and Decision Tree. International Journal of Innovative Technology and Exploring Engineering (IJITEE), October 2019.

[18] Jianfeng Zhao, Xia Mao, Lijiang Chen. Learning Deep features to Recognise Speech Emotion using Merged Deep CNN. IET Signal Process., 2018

[19] Tao, F.; Liu, G. Advanced LSTM: A study about better time dependency modeling in emotion recognition.In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2906–2910.

[20] Sahu, S.; Gupta, R.; Sivaraman, G.; AbdAlmageed,W.; Espy-Wilson, C. Adversarial auto-encoders for speech based emotion recognition. arXiv 2018, arXiv:1806.02146.

[21] Hajarolasvadi, N.; Demirel, H. 3D CNN-Based Speech Emotion Recognition Using K-Means Clustering and Spectrograms. Entropy 2019, 21, 479.

RT&A, No 3 (63)
Volume 16, September 2021

On the Use of Entropy as a Measure of Dependence of Two Events

Valentin Vankov Iliev

•

Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria
viliev@math.bas.bg

Abstract

We define degree of dependence of two events A and B in a probability space by using Boltzmann-Shannon entropy function of an appropriate probability distribution produced by these events and depending on one parameter (the probability of intersection of A and B) varying within a closed interval I.

The entropy function attains its global maximum when the events A and B are independent. The important particular case of discrete uniform probability space motivates this definition in the following way. The entropy function has a minimum at the left endpoint of I exactly when one of the events and the complement of the other are connected with the relation of inclusion (maximal negative dependence). It has a minimum at the right endpoint of I exactly when one of these events is included in the other (maximal positive dependence).

Moreover, the deviation of the entropy from its maximum is equal to average information that carries one of the binary trials $A \cup A^c$ and $B \cup B^c$ with respect to the other. As a consequence, the degree of dependence of A and B can be expressed in terms of information theory and is invariant with respect to the choice of unit of information.

Using this formalism, we describe completely the screening tests and their reliability, measure efficacy of a vaccination, the impact of some events from the financial markets to other events, etc.

A link is available for downloading an Excel program which calculates the degree of dependence of two events in a sample space with equally likely outcomes.

Keywords: entropy; average information; degree of dependence; probability space; probability distribution; experiment in a sample space; linear system; affine isomorphism; classification space.

1. Introduction

In this paper we study the set of ordered pairs (A, B) of events in a probability space in order to define a measure of dependence (the power of relations) between A and B. This is done by means of Boltzmann-Shannon entropy of a variable probability distribution that arises naturally out of the pair (A, B). This approach is radically different from the standard ones where most of the measures of dependence are linear functions in the probability of intersection $A \cap B$, see Section 4. For a detailed study of these measures refer to [1].

The ordered pairs (A, B) in the form of two-attributed tables are used meticulously by G. Udny Yule in his memoir [9] in order to "...classify the objects or individuals observed into two classes only". He presents various examples and defines different indices to study the degree of association (dependence) of the corresponding two events. Any such ordered pair of events is said to be, as we termed it, an *Yule's pair of events*. G. Udny Yule himself noted that W. R. Macdonell in [4] used a two-attributed table as a tool to study "...the degree of effectiveness of vaccination in small-pox".

The paper is organized as follows. In Section 2 we parameterize the members of an equivalence class consisting of Yule's pairs with fixed probabilities α and β (that is, Yule's pairs of type (α, β)). We use the fact that the probability distribution produced by the probabilities of results of the experiment corresponding to a Yule's pair (cf. [3, I,§5]) is solution of a linear system with one free variable θ (see (3)). The system of inequalities that restrict the components of this solution is equivalent to the restriction of the variation of θ within a closed interval $I(\alpha, \beta) \subset [0, 1]$. Thus, we naturally introduce (α, β, θ)-equivalence classes of Yule's pairs, whose members are said to be Yule's pairs of type (α, β, θ). Note that for any such pair, θ is the probability of the intersection of its components.

When we vary $(\alpha, \beta) \in [0, 1]^2$, the segment $\{\alpha\} \times \{\beta\} \times I(\alpha, \beta)$ sweeps a tetrahedron T_3 in $\mathbb{R}^3$, so the (α, β, θ)-equivalence classes of Yule's pairs are represented by some points in T_3, which, in turn, form so called dotted tetrahedron $T_3^{(\cdot)}$.

On the other hand, the affine isomorphism (4) which transforms $\mathbb{R}^3$ onto the hyperplane H in $\mathbb{R}^4$ that contains the solutions of the linear system (3), transforms the tetrahedron T_3 onto the 3-simplex $\Delta_3 \subset H$. Moreover, the dotted tetrahedron $T_3^{(\cdot)}$ is mapped on the dotted 3-simplex $\Delta_3^{(\cdot)} \subset H$, the latter classifying the probability distributions produced by all Yule's pairs. For the precise statements see Theorem 1 and Figure 1.

Given a Yule's pair of type (α, β, θ), Boltzmann-Shannon entropy $E_{\alpha,\beta}(\theta)$ of its distribution is a continuous function in $\theta \in I(\alpha, \beta)$ and its behaviour is described in Theorem 2 from Section 3. In particular, we show that $E_{\alpha,\beta}(\theta)$ attains its global maximum at the only point $\theta_0 = \alpha\beta$ for which the components of all Yule's pairs of type $(\alpha, \beta, \theta_0)$ (if any) are independent. The special case of a sample space with equally likely outcomes illustrates the fact that the maximum of dependence occurs at the endpoints of the interval $I(\alpha, \beta)$. More precisely, at the left endpoint we have $A \subset B^c$ or $B^c \subset A$ and at the right endpoint — $A \subset B$ or $B \subset A$.

Finally, $E_{\alpha,\beta}(\theta) = E_{\beta,\alpha}(\theta)$ and this common entropy function strictly increases to the left of θ_0 an strictly decreases to the right.

All of this motivates the use of entropy function $E_{\alpha,\beta}(\theta)$ as a measure of dependence of two events A and B with $\Pr(A) = \alpha$ and $\Pr(B) = \beta$: Negative dependence to the left of θ_0 and positive dependence to the right. By modifying appropriately $E_{\alpha,\beta}(\theta)$ by linear functions, we obtain a strictly increasing continuous function $e_{\alpha,\beta}$ which maps the range of θ onto the interval $[-1, 1]$ and serves (and is termed) as degree of dependence of the events A and B.

From the link provided in Remark 1 one can download a simple Excel program which calculates the degree of dependence of two events in a discrete uniform probability space with given cardinalities and given cardinality of their intersection.

It turns out that the expression for entropy function $E_{\alpha,\beta}$ is a particular case of what Shannon called in [6, Part I, Sction 6] the entropy of the joint event. More precisely, this is the complete amount of information which contains in the results of the experiment $\mathfrak{J}$ from (1). On the other hand, $\mathfrak{J}$ is the joint experiment of two binary trials: $\mathfrak{A} = A \cup A^c$ and $\mathfrak{B} = B \cup B^c$. Theorem 3 shows that the mutual information $I(\mathfrak{A}, \mathfrak{B})$ of the experiments $\mathfrak{A}$ and $\mathfrak{B}$ is equal to the deviation of the entropy $E_{\alpha,\beta}(\theta)$ from its maximum $E_{\alpha,\beta}(\alpha\beta)$. In accord with the expression (7) which represents the function $e_{\alpha,\beta}(\theta)$ as a fraction of amounts of information, the degree of dependence of two events is invariant with respect to change of unit of information (bits, nats, etc.).

In case Yule's pairs are models of a screening tests, the probability $F_-(\theta)$ of false negative and the probability $F_+(\theta)$ of false positive test are tending from statistically insignificant nearby the left endpoint of the range of θ to statistical significance in a neighbourhood of the right endpoint. Moreover, on the complement of any such neighbourhood the product $F_-(\theta)F_+(\theta)$ is bound below by a positive constant. In other words, a kind of uncertainty principle holds — see Subsection 5.4.

In Subsection 5.5 we show that the degree of dependence of pairs of events can be used as a measure of effectiveness of vaccine for a particular decease. As an example we estimate the efficacy of vaccine for small-pox tested via the epidemic at Sheffield in 1887-88, the statistical data taken from [9, I] .

In Section 4 we give several examples of other measures of dependence which are evaluated

by using Sheffield's sample.

2. Definitions and Notation

Let $(\Omega, \mathcal{A}, \Pr)$ be a probability space with set of outcomes Ω, σ-algebra $\mathcal{A}$, and probability function Pr. In this paper we are using only the structure of Boolean algebra on $\mathcal{A}$.

We introduce the following notation:

R is the range of the probability function $\Pr\colon \mathcal{A} \to \mathbb{R}$; $[(\alpha, \beta)]$ is the fiber of the surjective map $\mathcal{A}^2 \to R^2, (A, B) \mapsto (\Pr(A), \Pr(B))$, over $(\alpha, \beta) \in R^2$; $\theta^{(A,B)} = \Pr(A \cap B)$, $(A, B) \in \mathcal{A}^2$; $[(\alpha, \beta, \theta)]$ is the fiber of the map $[(\alpha, \beta)] \to R$, $(A, B) \mapsto \theta^{(A,B)}$, with image $R^{(\alpha,\beta)} \subset R$, over any $\theta \in R^{(\alpha,\beta)}$.

We note that the fibers $[(\alpha, \beta)]$ for $(\alpha, \beta) \in R^2$ form a partition of $\mathcal{A}^2$ and the fibers $[(\alpha, \beta, \theta)]$ for $\theta \in R^{(\alpha,\beta)}$ form a partition of $[(\alpha, \beta)]$.

As usual, the events $\emptyset$ and Ω are called *trivial*. The members of the equivalence class $[(\alpha, \beta)]$ (resp., the equivalence class $[(\alpha, \beta, \theta)]$) are said to be *Yule's pairs of type* (α, β) (resp., *Yule's pairs of type* (α, β, θ)).

3. Methods

In this paper we are using fundamentals of:

- Linear algebra,
- Affine geometry,
- Information Theory.

4. Classification of Yule's Pairs

4.1. The Probability Distribution of a Yule's Pair

Any ordered pair $(A, B) \in \mathcal{A}^2$ produces an experiment

$$\mathfrak{J} = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \cup (A^c \cap B^c) \tag{1}$$

(cf. [3, I,§5]) and the probabilities of its results:

$$\xi_1^{(A,B)} = \Pr(A \cap B), \xi_2^{(A,B)} = \Pr(A \cap B^c),$$

$$\xi_3^{(A,B)} = \Pr(A^c \cap B), \xi_4^{(A,B)} = \Pr(A^c \cap B^c).$$

For any $(A, B) \in [(\alpha, \beta)]$, the probability distribution

$$(\xi_1, \xi_2, \xi_3, \xi_4) = (\xi_1^{(A,B)}, \xi_2^{(A,B)}, \xi_3^{(A,B)}, \xi_4^{(A,B)}) \tag{2}$$

satisfies the linear system

$$\left|\begin{array}{ccccccccc} \xi_1 & + & \xi_2 & & & & & = & \alpha \\ & & & & \xi_3 & + & \xi_4 & = & 1-\alpha \\ \xi_1 & & & + & \xi_3 & & & = & \beta \\ & & \xi_2 & & & + & \xi_4 & = & 1-\beta. \end{array}\right. \tag{3}$$

Let H be the affine hyperplane in $\mathbb{R}^4$ with equation $\xi_1 + \xi_2 + \xi_3 + \xi_4 = 1$. The solutions of (3) depend on one parameter, say $\theta = \xi_1$, and form a straight line $\ell_{\alpha,\beta}$ in H with parametric representation

$$\ell_{\alpha,\beta}\colon \xi_1 = \theta, \xi_2 = \alpha - \theta, \xi_3 = \beta - \theta, \xi_4 = 1 - \alpha - \beta + \theta.$$

The map

$$\iota\colon \mathbb{R}^3 \to H, (\alpha, \beta; \theta) \mapsto (\theta, \alpha - \theta, \beta - \theta, 1 - \alpha - \beta + \theta) \tag{4}$$

is an affine isomorphism with inverse affine isomorphism

$$\chi\colon H \to \mathbb{R}^3, \xi \mapsto (\xi_1 + \xi_2, \xi_1 + \xi_3, \xi_1).$$

The trace of the 4-dimensional cube $\{\xi \in \mathbb{R}^4 | 0 \leq \xi_k \leq 1, k = 1,2,3,4\}$ onto the hyperplane H is the 3-dimensional simplex, that is, the tetrahedron, Δ_3 defined in H by the inequalities

$$\xi_1 \geq 0, \xi_2 \geq 0, \xi_3 \geq 0, \xi_1 + \xi_2 + \xi_3 \leq 1.$$

The inverse image $T_3 = \iota^{-1}(\Delta_3)$ via the affine isomorphism ι is the tetrahedron in $\mathbb{R}^3$ defined by the system of inequalities

$$\theta \leq \alpha, \theta \leq \beta, \theta \geq \alpha + \beta - 1, \theta \geq 0. \tag{5}$$

In other words, this is the tetrahedron with vertices $O(0,0,0)$, $M(1,0,0)$, $N(0,1,0)$, $P(1,1,1)$ — see Figure 1.

For any fixed $(\alpha, \beta) \in \mathbb{R}^2$ we set $\lambda_{\alpha,\beta} = \{\alpha\} \times \{\beta\} \times \mathbb{R}$, $C(\alpha,\beta) = \lambda_{\alpha,\beta} \cap T_3$, so $C(\alpha,\beta) = \{\alpha\} \times \{\beta\} \times I(\alpha,\beta)$, $I(\alpha,\beta) \subset \mathbb{R}$. The system (5) yields that $I(\alpha,\beta)$ equals the closed interval $[\ell(\alpha,\beta), r(\alpha,\beta)]$, $\ell(\alpha,\beta) = \max(0, \alpha+\beta-1), r(\alpha,\beta) = \min(\alpha,\beta)$. We have $\alpha\beta \in I(\alpha,\beta)$ and denote by $\mathring{I}(\alpha,\beta)$ the interior of the interval $I(\alpha,\beta)$. We obtain immediately:

Lemma 1. Let $(\alpha,\beta) \in [0,1]^2$. The next three statements are equivalent:
(i) One has $(\alpha,\beta) \in (0,1)^2$.
(ii) One has $\alpha\beta \in \mathring{I}(\alpha,\beta)$.
(iii) One has $\mathring{I}(\alpha,\beta) \neq \emptyset$.
(iv) Under the above conditions, one has $\xi_k(\theta) > 0$ for all $\theta \in \mathring{I}(\alpha,\beta)$ and all $k = 1,2,3,4$.
(v) Conversely, if there exists $\theta \in I(\alpha,\beta)$ such that $\xi_k(\theta) > 0$ for all $k = 1,2,3,4$, then (i) — (iii) hold.

We have $R^{(\alpha,\beta)} \subset I(\alpha,\beta)$ and define the *dotted interval* $I^{(\cdot)}(\alpha,\beta) = R^{(\alpha,\beta)}$. The *dotted segment* $C^{(\cdot)}(\alpha,\beta) = \{\alpha\} \times \{\beta\} \times I^{(\cdot)}(\alpha,\beta)$, $(\alpha,\beta) \in R^2$, is the locus of all triples of probabilities $(\alpha,\beta,\theta^{(A,B)})$, where $(A,B) \in [(\alpha,\beta)]$.

For any $(\alpha,\beta) \in \mathbb{R}^2$ we set $D(\alpha,\beta) = \iota(C(\alpha,\beta))$. Since $\iota(\lambda_{\alpha,\beta}) = \ell_{\alpha,\beta}$, we obtain that $D(\alpha,\beta) = \ell_{\alpha,\beta} \cap \Delta_3$.

Let $I_k(\alpha,\beta) = [\ell_k(\alpha,\beta), r_k(\alpha,\beta)]$ be the corresponding range of the real variable $\xi_k(\theta)$ for $k = 1,2,3,4$, with $I_1(\alpha,\beta) = I(\alpha,\beta)$. The line segment $D(\alpha,\beta)$ in $\ell_{\alpha,\beta}$ has endpoints

$$(\ell_1(\alpha,\beta), \ell_2(\alpha,\beta), \ell_3(\alpha,\beta), \ell_4(\alpha,\beta)), (r_1(\alpha,\beta), r_2(\alpha,\beta), r_3(\alpha,\beta), r_4(\alpha,\beta)).$$

Since ι is also a homeomorphism, we have $\mathring{D}(\alpha,\beta) = \ell_{\alpha,\beta} \cap \mathring{\Delta}_3$.

The line segment $D(\alpha,\beta)$ contains the *dotted segment* $D^{(\cdot)}(\alpha,\beta)$ which is the locus of all probability distributions (2) for which $(A,B) \in [(\alpha,\beta)]$.

Finally, we note that $T_3 = \cup_{(\alpha,\beta)\in[0,1]^2} C(\alpha,\beta)$, $\Delta_3 = \cup_{(\alpha,\beta)\in[0,1]^2} D(\alpha,\beta)$, and the unions $T_3^{(\cdot)} = \cup_{(\alpha,\beta)\in R^2} C^{(\cdot)}(\alpha,\beta)$, $\Delta_3^{(\cdot)} = \cup_{(\alpha,\beta)\in R^2} D^{(\cdot)}(\alpha,\beta)$ are the corresponding *dotted tetrahedrons*.

The above considerations and Figure 1 yield the following theorem and its corollary.

Theorem 1. (i) The affine isomorphism $\iota\colon \mathbb{R}^3 \to H$ from (4) is a strictly increasing transformation of any line segment $C(\alpha,\beta)$ (resp., dotted line segment $C^{(\cdot)}(\alpha,\beta)$) onto the line segment $D(\alpha,\beta)$ (resp., onto the dotted line segment $D^{(\cdot)}(\alpha,\beta)$).

(ii) ι maps the tetrahedron T_3 (resp., dotted tetrahedron $T_3^{(\cdot)}$) onto the tetrahedron Δ_3 (resp., onto the dotted tetrahedron $\Delta_3^{(\cdot)}$).

(iii) The dotted tetrahedron $T_3^{(\cdot)}$ is the classification space of all equivalence classes $[(\alpha,\beta,\theta)]$ of Yule's pairs.

(iv) The dotted tetrahedron $\Delta_3^{(\cdot)}$ is the classification space of all probability distributions (2) produced by Yule's pairs.

Corollary 1. (i) One has $\xi_1(\theta) = 0$ if and only if $(\alpha,\beta,\theta) \in MON$.
(ii) One has $\xi_2(\theta) = 0$ if and only if $(\alpha,\beta,\theta) \in NOP$.
(iii) One has $\xi_3(\theta) = 0$ if and only if $(\alpha,\beta,\theta) \in MOP$.
(iv) One has $\xi_4(\theta) = 0$ if and only if $(\alpha,\beta,\theta) \in MNP$.

5. Entropy and Dependence of Yule's Pairs

5.1. Entropy of a Yule's Pair

Let us suppose that $(\alpha,\beta)\in(0,1)^2$. Then Lemma 1 implies $\mathring{I}(\alpha,\beta)\neq\emptyset$ and $\xi_k(\theta)>0$ for $\theta\in\mathring{I}(\alpha,\beta)$ and for all $k=1,2,3,4$. Therefore Boltzmann-Shannon entropy of the probability distribution $(\xi_1(\theta),\xi_2(\theta),\xi_3(\theta),\xi_4(\theta))$ is defined (cf. [6], [7]):

$$E_{\alpha,\beta}(\theta)=-\sum_{k=1}^{4}\xi_k(\theta)\ln(\xi_k(\theta)),\theta\in\mathring{I}(\alpha,\beta).$$

Theorem 2. Let $(\alpha,\beta)\in(0,1)^2$.

(i) For any $\theta\in\mathring{I}(\alpha,\beta)$ one has

$$E'_{\alpha,\beta}(\theta)=\ln\frac{\xi_2(\theta)\xi_3(\theta)}{\xi_1(\theta)\xi_4(\theta)}.$$

(ii) The function $E_{\alpha,\beta}(\theta)$ in θ strictly increases on the interval $(\ell(\alpha,\beta),\alpha\beta]$ and strictly decreases on the interval $[\alpha\beta,r(\alpha,\beta))$, having a global maximum at $\theta=\alpha\beta$.

(iii) The function $E_{\alpha,\beta}(\theta)$ can be extended uniquely as a continuous function on the closed interval $I(\alpha,\beta)$, which strictly increases on the interval $[\ell(\alpha,\beta),\alpha\beta]$ and strictly decreases on the interval $[\alpha\beta,r(\alpha,\beta)]$.

(iv) One has $I(\alpha,\beta)=I(\beta,\alpha)$ and $E_{\alpha,\beta}=E_{\beta,\alpha}$.

Proof. (i) We have

$$E'_{\alpha,\beta}(\theta)=-\sum_{k=1}^{4}\xi'_k(\theta)\ln(\xi_k(\theta))-\sum_{k=1}^{4}\xi_k(\theta)\frac{\xi'_k(\theta)}{\xi_k(\theta)}=\ln\frac{\xi_2(\theta)\xi_3(\theta)}{\xi_1(\theta)\xi_4(\theta)}.$$

(ii) The equation $E'_{\alpha,\beta}(\theta)=0$ (resp., the inequality $E'_{\alpha,\beta}(\theta)>0$) is equivalent to $\theta=\alpha\beta$ (resp., $\theta<\alpha\beta$).

(iii) According to Corollary 1, one or two functions $\xi_k(\theta)$ are zero at any endpoint of each interval $I(\alpha,\beta)$ and all functions $\xi_k(\theta)$ are strictly positive on the interior $\mathring{I}(\alpha,\beta)$. For a fixed interval $I(\alpha,\beta)$ and a fixed endpoint a of $I(\alpha,\beta)$ the limit $\lim_{\theta\to a}E_{\alpha,\beta}(\theta)$ exists and we extend $E_{\alpha,\beta}(\theta)$ as continuous at the point a.

(iv) We have $I(\alpha,\beta)=I(\beta,\alpha)$ and the transposition of events A and B yields transposition of the functions $\xi_2(\theta)$ and $\xi_3(\theta)$.

■

The continuous function $E_{\alpha,\beta}(\theta)$ in $\theta\in I(\alpha,\beta)$ is said to be the *entropy function of Yule's pairs of type* (α,β) and its value at $\theta=\theta^{(A,B)}$ is called *entropy of Yule's pair* (A,B) *of type* (α,β).

Theorem 2 implies immediately

Corollary 2. Let $(\alpha,\beta)\in(0,1)^2$. The following three statements are equivalent:

(i) One has $\theta_0=\alpha\beta$.

(ii) If Yule's pair (A,B) of type (α,β) satisfies the equality $\xi_1^{(A,B)}=\theta_0$, then the events A and B are independent.

(iii) The entropy function $E_{\alpha,\beta}(\theta)$ of Yule's pairs of type (α,β) attains its global maximum at the point θ_0.

5.2. Degree of Dependence of Pairs of Events

We "normalize" the entropy function by composing the functions $E_{\alpha,\beta}(\theta)$ and $2E_{\alpha,\beta}(\alpha\beta)-E_{\alpha,\beta}(\theta)$ on the intervals of their increase by the appropriate linear functions and obtain for any pair $(\alpha,\beta)\in(0,1)^2$ a continuous function $e_{\alpha,\beta}\colon I(\alpha,\beta)\to[-1,1]$. In accord with Theorem 2, (iv), we

have $e_{\alpha,\beta} = e_{\beta,\alpha}$. The value of the function $e_{\alpha,\beta}$ at $\theta \in I(\alpha,\beta)$, $\theta = \theta^{(A,B)}$, is said to be *degree of dependence of the events A and B with $\alpha = \Pr(A)$, $\beta = \Pr(B)$*.

The function $e_{\alpha,\beta}$ strictly increases on the interval $I(\alpha,\beta)$ from -1 to 1 and attains value 0 at the point $\alpha\beta$. The events A and B are said to be *negatively dependent* if $\theta^{(A,B)} < \alpha\beta$ and *positively dependent* if $\theta^{(A,B)} > \alpha\beta$. When $\theta^{(A,B)} = \alpha\beta$ the events A and B are independent (the entropy is maximal). In a small neighbourhood of the left endpoint $\ell(\alpha,\beta)$ of the interval $I(\alpha,\beta)$ the dependence is *negatively strong*, with "maximum" $1 = |-1|$ at $\theta = \ell(\alpha,\beta)$ (if attainable). In a small neighbourhood of the right endpoint $r(\alpha,\beta)$ of the interval $I(\alpha,\beta)$ the dependence is *positively strong*, with maximum 1 at $\theta = r(\alpha,\beta)$ (if attainable). In both cases, the entropy is minimal at the endpoints $\ell(\alpha,\beta)$ and $r(\alpha,\beta)$ of the corresponding semi-intervals. Note that in a small neighbourhood of the point $\theta = \alpha\beta$ the events A and B are "almost independent" (the entropy is close to its maximum).

Remark 1. One can find below the link to a simple Excel program which calculates the degree of dependence of two events in a sample space with equally likely outcomes:
`http://www.math.bas.bg/algebra/valentiniliev/`

5.3. A Glance at the Information Theory

The experiment $\mathfrak{J}$ from (1) is the joint experiment (see [6, Part I, Section 6]) of two simpler binary trials: $\mathfrak{A} = A \cup A^c$ and $\mathfrak{B} = B \cup B^c$ with $\Pr(A) = \alpha$, $\Pr(B) = \beta$. The *average quantity of information of one of the experiments $\mathfrak{A}$ and $\mathfrak{B}$, relative to the other*, (see [2, §1]), is defined in this particular case by the formula

$$I(\mathfrak{A},\mathfrak{B}) = \xi_1 \ln\frac{\xi_1}{\alpha\beta} + \xi_2 \ln\frac{\xi_2}{\alpha(1-\beta)} + \xi_3 \ln\frac{\xi_3}{(1-\alpha)\beta} + \xi_4 \ln\frac{\xi_4}{(1-\alpha)(1-\beta)}. \tag{6}$$

The definition of the degree function $e_{\alpha,\beta}(\theta)$ and (6) yield immediately the following:

Theorem 3. (i) One has

$$I(\mathfrak{A},\mathfrak{B})(\theta) = E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)$$

for all $\theta \in I(\alpha,\beta)$.

(ii) One has

$$e_{\alpha,\beta}(\theta) = \begin{cases} -\frac{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)}{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\ell(\alpha,\beta))} & \text{if } \ell(\alpha,\beta) \le \theta \le \alpha\beta \\ \frac{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\theta)}{E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(r(\alpha,\beta))} & \text{if } \alpha\beta \le \theta \le r(\alpha,\beta). \end{cases} \tag{7}$$

Remark 2. Since

$$E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(\ell(\alpha,\beta)) = \max_{\ell(\alpha,\beta)\le\tau\le\alpha\beta} I(\mathfrak{A},\mathfrak{B})(\tau),$$

$$E_{\alpha,\beta}(\alpha\beta) - E_{\alpha,\beta}(r(\alpha,\beta)) = \max_{\alpha\beta\le\tau\le r(\alpha,\beta)} I(\mathfrak{A},\mathfrak{B})(\tau),$$

we can write down the equality (7) in the form

$$e_{\alpha,\beta}(\theta) = \begin{cases} -\frac{I(\mathfrak{A},\mathfrak{B})(\theta)}{\max_{\ell(\alpha,\beta)\le\tau\le\alpha\beta} I(\mathfrak{A},\mathfrak{B})(\tau)} & \text{if } \ell(\alpha,\beta) \le \theta \le \alpha\beta \\ \frac{I(\mathfrak{A},\mathfrak{B})(\theta)}{\max_{\alpha\beta\le\tau\le r(\alpha,\beta)} I(\mathfrak{A},\mathfrak{B})(\tau)} & \text{if } \alpha\beta \le \theta \le r(\alpha,\beta). \end{cases}$$

Part (ii) of the above theorem implies

Corollary 3. The degree of dependence of two events does not depend on the choice of unit of information.

The graphs of $E_{\alpha,\beta}$ and $e_{\alpha,\beta}$ for some particular $(\alpha,\beta) \in (0,1)^2$ are presented in Figures 2 and 3.

5.4. Application: Description of a Screening Test

According to Merriam-Webster Dictionary, a screening test is "...a preliminary or abridged test intended to eliminate the less probable members of an experimental series". In other words, some of the equally likely outcomes of a sample space Ω possess a property and, in this way, form an event A. On the other hand, there exists an event B consisting of all outcomes which as if have this property after conducting the test. Thus, we obtain a Yule's pair of events in a sample space Ω. The test does not always work perfectly — sometimes it is negative under the condition that the property is present (that is, *false negative*), and sometimes it is positive under the condition that the property is absent (that is, *false positive*). Let us suppose that all members of the population are tested and that $\Pr(A) = \alpha$, $\Pr(B) = \beta$, where $(\alpha, \beta) \in (0,1)^2$. Yule's pair (A, B) produces the experiment (1) and in turn, the probability distribution (2) consisting of its results. In the notation introduced in Subsection 4.1, the probability F_- of false negative result is $\frac{\xi_2^{(A,B)}}{\alpha} = \frac{\alpha-\theta}{\alpha}$ and the probability F_+ of false positive result is $\frac{\xi_3^{(A,B)}}{1-\alpha} = \frac{\beta-\theta}{1-\alpha}$, where $\theta = \theta^{(A,B)}$. The product $F_-(\theta)F_+(\theta)$ is a quadratic function in θ which strictly decreases on the interval $I(\alpha, \beta)$ and assumes value 0 at its right endpoint $r(\alpha, \beta)$. In particular, for the complement of any open neighbourhood of $r(\alpha, \beta)$ in the interval $I(\alpha, \beta)$, there exists a positive constant K such that $F_-(\theta)F_+(\theta) \geq K$ for any point θ from this complement. In other words, both $F_-(\theta)$ and $F_+(\theta)$ can not be simultaneously as small as we want (a kind of uncertainty principle). The conditional probabilities $F_-(\theta)$ and $F_+(\theta)$ are statistically acceptable in a small neighbourhood of the point $r(\alpha, \beta)$, at least one being 0 at this point. The reliability of $F_-(\theta)$ and $F_+(\theta)$ decreases when θ approaches the left endpoint $\ell(\alpha, \beta)$ of $I(\alpha, \beta)$. When $\theta = \ell(\alpha, \beta)$, at least one of $F_-(\theta)$ and $F_+(\theta)$ is equal to 1. In terms of the degree of dependence $e_{\alpha,\beta}(\theta)$, $\theta = \theta^{(A,B)}$, of the events A and B, this behaviour can be described in the following way: When $e_{\alpha,\beta}(\theta)$ is close to -1, then the test is not reliable but its effectiveness increases when $e_{\alpha,\beta}(\theta)$ approaches 1. In a small neighbourhood of 1 the test is statistically acceptable.

5.5. Application: Effectiveness of Vaccination

Let us consider a population whose members have a particular disease for which a vaccine is developed. Let A be the set of all those who have recovered and let B be the set of vaccinated members. Then A^c is the set of all fatal endings and B^c is the set of unvaccinated. If $\alpha = \Pr(A)$, $\beta = \Pr(B)$, then the degree of dependence $e_{\alpha,\beta}(\theta)$, $\theta = \theta^{(A,B)}$, of the events A and B measures the effectiveness of the corresponding vaccine. More precisely, when $e_{\alpha,\beta}(\theta)$ is close to -1, then the vaccine is counterproductive and its effectiveness increases being negative when $e_{\alpha,\beta}(\theta) < 0$ and positive when $e_{\alpha,\beta}(\theta) > 0$. In case $e_{\alpha,\beta}(\theta) = 0$ the vaccination does not influence the recovery and it is very positively effective when $e_{\alpha,\beta}(\theta)$ is close to 1.

In his memoir [9, Section I, Table I], G. Udny Yule presents a table used by W. R. Macdonell in [4] in order to show "...the recoveries and deaths amongst vaccinated and unvaccinated patients during the small-pox epidemic at Sheffield in 1887-88", see Table 1.

We have $\alpha = 0.88262811$, $\beta = 0.899213268$, $\theta = 0.840102063$, and $e_{\alpha,\beta}(\theta) = 0.268810618$. Therefore the results of this vaccination are faintly positive (the recovery is not only due to vaccination!).

Yule's pair (A, B) considered as a screening test has statistically acceptable false negative probability $\Pr(B^c|A) \approx 0.0482$ (the probability that a member is unvaccinated under the condition that he/she is recovered). On the other hand, this test has not statistically significant false positive probability $\Pr(B|A^c) \approx 0.4964$ (the probability that a member of the population was vaccinated under the condition that he/she is dead). Equivalently: the matter of life and death depended of the result of tossing an almost fair coin!

6. Other Measures of Dependence

In this section we assume that Ω is a sample space with equally likely outcomes.

6.1. Yule's Q

The difference $\delta = \Pr(A \cap B) - \Pr(A)\Pr(B) = \theta - \alpha\beta$ (the deviation from independence) is called *copula between A and B*. It is cited by G. Udny Yule in [9, Section I, n_o 5]. He notes there that the relation $\delta = \xi_1(\theta)\xi_4(\theta) - \xi_2(\theta)\xi_3(\theta)$ is due to Karl Pearson (one of his teachers).

In [9, Section I, n_o 9], G. Udny Yule introduces his measure of association first given in [8, Section I, n_o 9]:

$$Q = \frac{\xi_1(\theta)\xi_4(\theta) - \xi_2(\theta)\xi_3(\theta)}{\xi_1(\theta)\xi_4(\theta) + \xi_2(\theta)\xi_3(\theta)} = \frac{\theta - \alpha\beta}{2\theta^2 - (2\alpha + 2\beta - 1)\theta + \alpha\beta}.$$

It has the necessary properties: (a) $Q = 0$ if and only if A and B are independent; (b) $Q = 1$ if and only if $A \subset B$ or $B \subset A$; (c) $Q = -1$ if and only if $A \subset B^c$ or $B^c \subset A$. Finally, $-1 \leq Q \leq 1$. In the case of Sheffield's epidemic, we have $Q = 0.902299648$.

We define the function

$$Q_{\alpha,\beta}(\theta) = \frac{\theta - \alpha\beta}{2\theta^2 - (2\alpha + 2\beta - 1)\theta + \alpha\beta}, \theta \in I(\alpha, \beta),$$

which produces Yule's Q (see Figure 7).

Remark 3. There are infinitely many functions of the form $h(\theta) = \frac{f(\theta) - g(\theta)}{f(\theta) + g(\theta)}$ with the properties (a), (b), (c), and $-1 \leq h(\theta) \leq 1$, defined on the interval $I(\alpha, \beta)$, $(\alpha, \beta) \in (0, 1)^2$. For example, there exist infinitely many pairs of cubic polynomials $f(\theta)$ and $g(\theta)$ which work.

6.2. Obreshkoff's Measures of Dependence

The properties of the copula δ are also discussed by N. Obreshkoff in his textbook [5, Chapter 3,§6] and in [3]. In particular, the relation $\Pr(B|A) = \Pr(B) + \frac{\delta}{\Pr(A)}$ shows that "... the probability of one of these events increases under the condition that the other comes true in case $\delta > 0$ and decreases in case $\delta < 0$". Moreover, $-\delta = \Pr(A \cap B^c) - \Pr(A)\Pr(B^c)$.

The number

$$\rho(B; A) = \Pr(B|A) - \Pr(B|A^c) = \frac{\theta - \alpha\beta}{\alpha(1 - \alpha)}$$

is called *coefficient of regression of B with respect to A*. It measures the influence of A on B. We have $-1 \leq \rho(B; A) \leq 1$.

It has the following properties: (a) $\rho(B; A) = 0$ if and only if A and B are independent, (b) $\rho(B; A) = 1$ if and only if $A = B$, (c) $\rho(B; A) = -1$ if and only if $A^c = B$.

In the above example of small-pox epidemic at Sheffield we have $\rho(B; A) = 0.44819565$.

We define the functions

$$\rho_{\alpha;\beta}(\theta) = \frac{\theta - \alpha\beta}{\alpha(1 - \alpha)}, \rho_{\beta;\alpha}(\theta) = \frac{\theta - \alpha\beta}{\beta(1 - \beta)}, \theta \in I(\alpha, \beta),$$

which produce the corresponding coefficients of regression (see Figures 4 and 5).

The numbers $\rho(B; A)$ and $\rho(A; B)$ have the same sign and, in general, are not equal. Their geometric mean

$$R(A, B) = \pm\sqrt{\rho(A; B)\rho(B; A)} = \frac{\theta - \alpha\beta}{\sqrt{\alpha\beta(1 - \alpha)(1 - \beta)}},$$

where $\pm$ is chosen to be the common sign of $\rho(B; A)$ and $\rho(A; B)$, is said to be *coefficient of correlation between A and B*. This coefficient has the above properties (a) — (c). In the case of Sheffield's epidemic we have $R(A, B) = 0.4791876$.

We define the function

$$R_{\alpha;\beta}(\theta) = \frac{\theta - \alpha\beta}{\sqrt{\alpha\beta(1 - \alpha)(1 - \beta)}}, \theta \in I(\alpha, \beta),$$

which produces the corresponding coefficient of correlation (see Figure 6).

7. Conclusions

This paper presents an original approach to the problem of measuring the degree of dependence of two events A and B in a probability space. It uses the only reliable way of evaluation of the power of relations between these events, borrowed from statistical physics and information theory: this is the utilization of Boltzmann-Shannon entropy. More precisely, we start with the joint experiment assembled by the two binary trials $A \cup A^c$ and $B \cup B^c$. The four probabilities of results of this experiment constitute a variable probability distribution and satisfy a simple linear system whose general solution $(\xi_1, \xi_2, \xi_3, \xi_4)$ depends on one parameter θ (the probability of intersection $A \cap B$). Note that due to the natural constraints on ξ_k, $k = 1, 2, 3, 4$, θ varies throughout a closed interval $I(\alpha, \beta)$, where $\alpha = \Pr(A)$ and $\beta = \Pr(B)$. We modify naturally the entropy function of the distribution $(\xi_1(\theta), \xi_2(\theta), \xi_3(\theta), \xi_4(\theta))$ and obtain the degree of dependence function $e_{\alpha,\beta}(\theta)\colon I(\alpha, \beta) \to [-1, 1]$. By definition, if $\theta = \Pr(A \cap B)$, then $e_{\alpha,\beta}(\theta)$ measures the intensity of relations between A and B.

Our degree of dependence is still within the probation period. In its defence it can be said that evaluates the mutual information which is exchanged between the random objects A and B and, moreover, does not depend on the choice of unit of information. It also reflects plausibly the behaviour of a screening test or impact of a vaccination on the survival of a person. The function $e_{\alpha,\beta}(\theta)$ can also be used for measuring the effectiveness of a drug or medical treatment, the association of adverse events with use of some particular drug, the association of certain events with the stock market prices, etc.

8. Figures and Tables

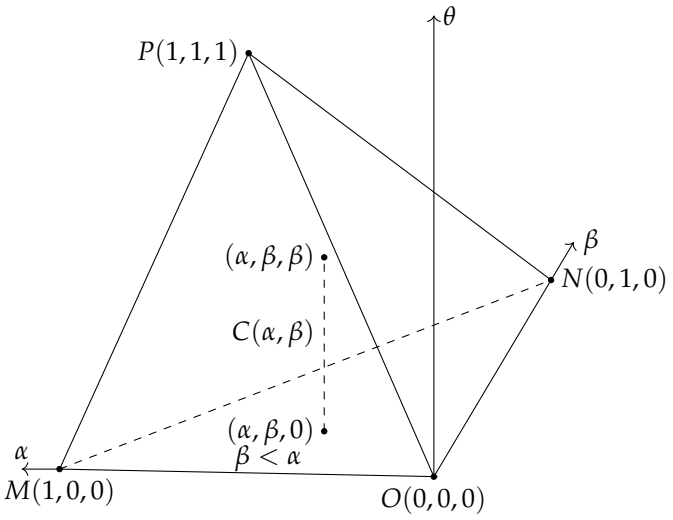


Figure 1

	B	B^c	Total
A	3951	200	4151
A^c	278	274	552
Total	4229	474	4703

Table 1

In all graphs below we use Sheffield's sample data.

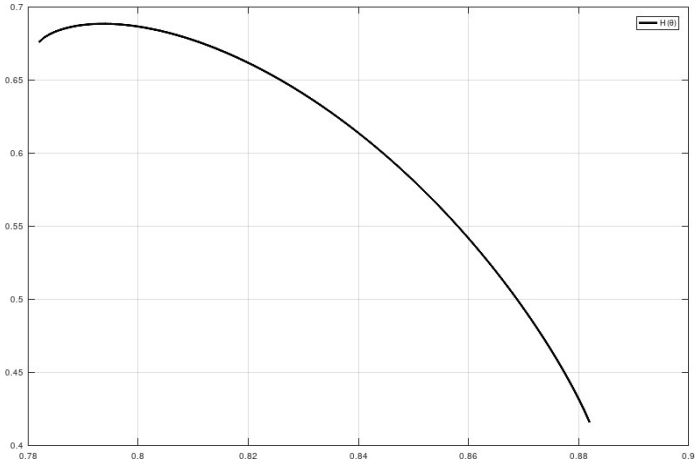


Figure 2

Graph of the entropy function $E_{\alpha,\beta}(\theta)$

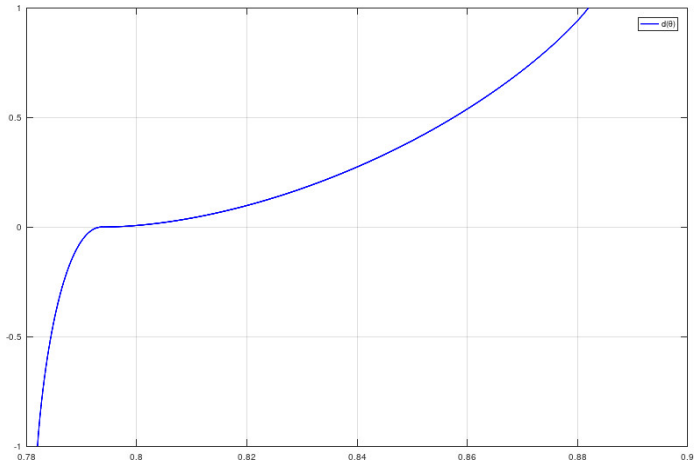


Figure 3
Graph of the degree function $e_{\alpha,\beta}(\theta)$

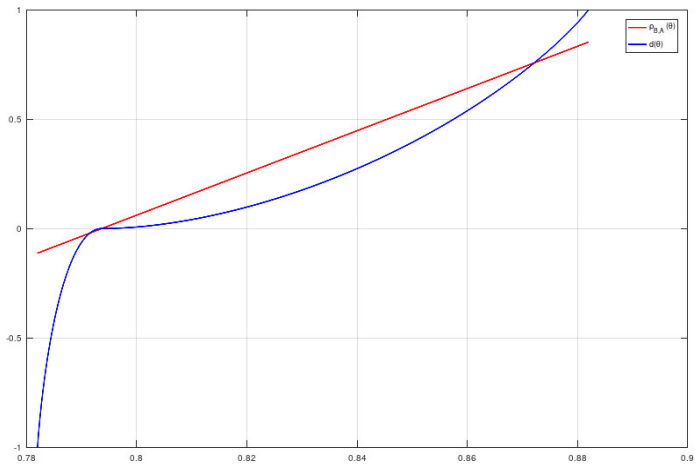


Figure 4
Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $\rho_{\beta;\alpha}(\theta)$

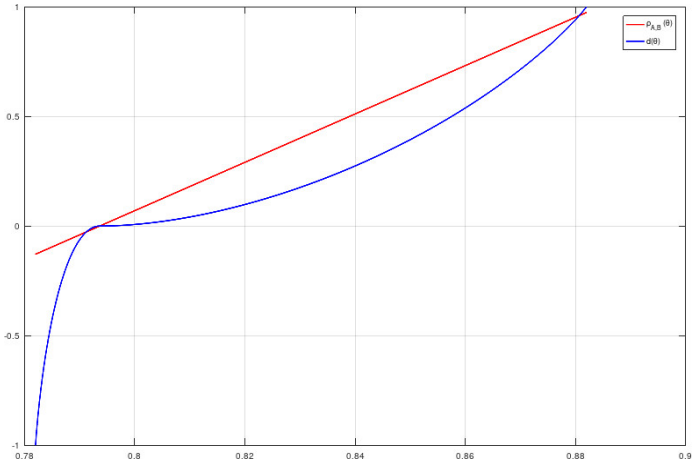


Figure 5
Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $\rho_{\alpha;\beta}(\theta)$

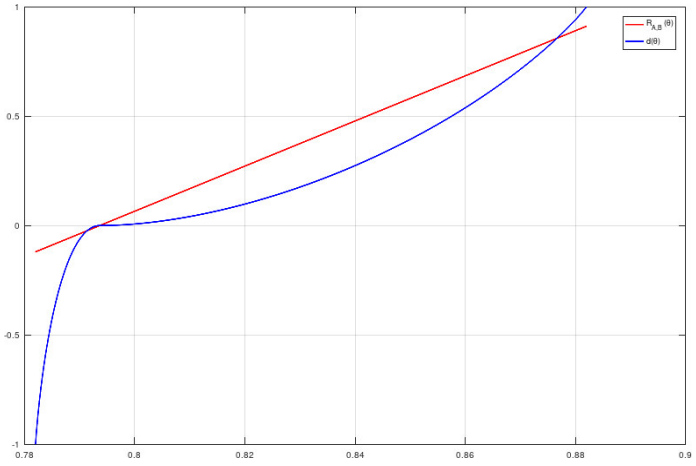


Figure 6
Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $R_{\alpha;\beta}(\theta)$

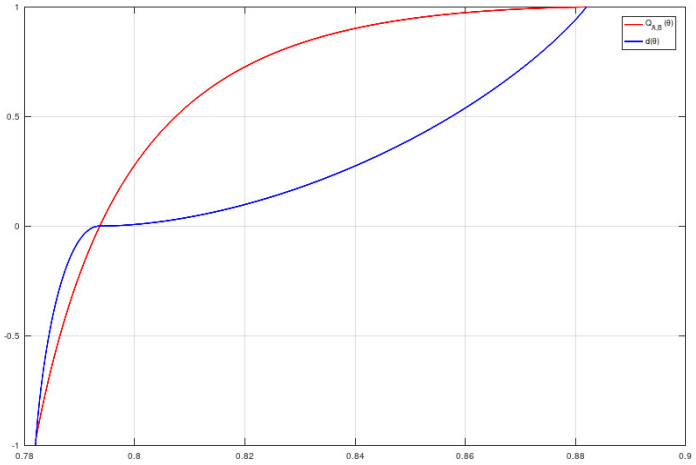


Figure 7
Comparison of the graphs of $e_{\alpha,\beta}(\theta)$ and $Q_{\alpha,\beta}(\theta)$

Acknowledgements

It is a pleasure for me to cordially thank Boyan Dimitrov, Kettering University, MI, USA, whose notes and suggestions are invaluable for making this paper more readable.

I would also like to express my sincere thanks to Dimitar Guelev for sending me references to the works of A. N. Kolmogorov on information theory, to Hristo Iliev for his mastery of drawing the graphs in this paper, and to the administration of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences for creating perfect and safe conditions of work.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

9. Declaration of Conflicting Interests

The Author declares that there is no conflict of interest.

References

[1] Dimitrov B. (2010) Some Obreshkov Measures of Dependence and Their Use. *Comptes rendus de l'Academie bulgare des Sciences* 63:5–18.

[2] Gelfand I. M., Kolmogorov A. N., Yaglom A. M. (1993), Amount of Information and Entropy for Continuous Distributions. Mathematics and Its Applications, Selected Works of A. N. Kolmogorov, III: Information Theory and the Theory of Algorithms, 33–56, Springer Science+Business Media Dordrecht 1993.

[3] Kolmogorov A. N. (1956). Foundations of the Theory of Probability, Chelsea Publishing Company, New Yourk 1956.

[4] Macdonell W. R. (1902) On the influence of previous vaccination in cases of small-pox. *Biometrika*, 1:375–383.

[5] Obreshkoff N. (1963). Theory of Probability, Nauka i Izkustvo, Sofia 1963 (In Bulgarian)

[6] Shannon C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379-423.

[7] Shannon C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:523–656.

[8] Yule G. Udny. (1900) On the Association of the Attributes in Stattistics. *Phil. Trans. Roy. Soc., A* 194: 257-319.

[9] Yule G. Udny. (1912) On the Methods of Measuring Association Between Two Attributes. *Journal of the Royal Statistical Society* 75: 579–652.

Statistical Analysis of Marshall-Olkin inverse Maxwell Distribution: Estimation and Application to Real Data

C. P. Yadav*, Jitendra Kumar**, M. S. Panwar*

•

*Department of Statistics, Banaras Hindu University, Varanasi-221005, India.
**Directorate of Economics & Statistics, Planning Department, Delhi-110054, India.
cahndrap.yadav4@bhu.ac.in, statsjkumar@gmail.com,vir.panwar@gmail.com

Abstract

In this paper, Marshall-Olkin inverse Maxwell distribution is proposed by generalizing the inverse Maxwell distribution under the Marshall-Olkin family of distribution that leads to greater flexibility in modeling various new data types. The basic statistical properties for the proposed distribution including moments, quantile function, median, skewness, kurtosis, and stochastic ordering are derived. Point estimates for the parameters are obtained by using two well known methods maximum likelihood and maximum spacing methods. The confidence intervals are used by using asymptotic properties of maximum likelihood estimators and boot-p methods. We have applied the proposed distribution under different real-life scenarios such as record value problem, system lifetime distributions, stress-strength reliability and random censored problems. For illustration purposes, simulation and real data results are established.

Keywords: Marshall-Olkin Family, Tilt Parameter, Inverse Maxwell Distribution, Stochastic Ordering, Maximum Likelihood Estimate, Maximum Spacing, Record Value, Stress-Strength Reliability, Random Censoring

1. Introduction

1.1. Literature

In lifetime experiments, the problem of finding the appropriate lifetime distributions is a concern for a long time. In literature, we have a lot of distributions that study different types of hazard nature. There are many scenarios in real life where some standard distributions are not applied or less suitable to fit actual data. Another application of generalizing distributions is to gain flexibility and better fit to real-life data. Since the several lifetime distributions had been proposed consisting of increasing, decreasing and bathtub hazard rates. Some well-known distributions consisting of these hazard natures are Exponential, Weibull, Gamma, Normal, etc. but these standard distributions are not always fulfilling our purposes. A situation in which the hazard rate initially increases attains a maximum point and then again starts decreasing generates an upside-down bathtub (UBT) shaped hazard rate. There are some distributions in literature containing UBT shapes as, inverse Gamma distribution, inverse Gaussian distribution, Log-Normal distribution, Log-Logistic distribution, Birnbaum-Saunders distribution, inverse Weibull distribution and inverse Maxwell distribution (InvMWD). For example, the lifetime models that present upside-down bathtub failure rates curves can be observed in the course of a disease whose mortality reaches a peak after some finite period and then declines gradually. Also, it is observed that the risks of dying a patient just after an operation increase due to infection and then decrease with recovery.

Many authors have been discussed the situations where the data shows decreasing, increasing, bathtub and UBT shape hazard rates. Proschan (1963) discussed the air-conditioning systems of

planes follow decreasing failure rate. They obtained the reliability characteristics of an aircraft air-conditioning system of the plane in an airline-use environment. In the paper, Kuş (2007) analyzed the earthquake data in the last century in the North Anatolian fault zone and found that the fitting the decreasing. Folks and Chhikara (1978) showed the behavior of the inverse Gaussian distribution as an upside-down bathtub model and reviewed its important statistical properties. Langlands et al. (1979) studied the pattern of mortality for the breast carcinoma data and concluded that it increases initially and after a certain time it started declining. Bennett (1983) studied the failure rate of lung cancer data applied to uni-model Log-Logistic distribution. In article Efron (1988), authors analyzed the data set in the context of head and neck cancer, in which the hazard rate initially increased, attained a maximum and then decreased before it stabilized owing to a therapy. Sharma et al. (2015) discussed the inverted versions of usual distributions are capable of modeling the data with UBT shaped failure rate. In a recent study in article Tomer and Panwar (2020), authors had obtained the estimates of the InvMWD under classical and Bayesian paradigm along with basic statistical properties and applications in different scenarios. In this article, we propose a new lifetime distribution utilizing the well known M-O family proposed by Marshall and Olkin (1997) taking InvMWD as the baseline distribution. The proposed distribution is named as Marshall-Olkin inverse Maxwell distribution (M-O InvMWD). We discuss the statistical properties of the proposed distribution and show the application for lifetime experiment.

1.2. Inverse Maxwell Distribution

Singh and Srivastava (2014) proposed the InvMWD and discusses the basic properties. Recently, Tomer and Panwar (2020) reviewed the InvMWD and established its important statistical properties with its applications in many fields. A real valued random variable Y following InvMWD has probability density function (pdf) as

$$f(y;\theta) = \frac{4}{\sqrt{\pi} y^4 \theta^{\frac{3}{2}}} \exp\left\{-\frac{1}{\theta y^2}\right\}; \quad y > 0, \theta > 0. \tag{1}$$

The survival function of Y is given by

$$S(y;\theta) = \frac{2}{\sqrt{\pi}} \gamma\left(\frac{3}{2}, \frac{1}{\theta y^2}\right) = 1 - \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}, \frac{1}{\theta y^2}\right), \tag{2}$$

where, $\gamma(a,z) = \int_0^z u^{a-1} e^{-u}\, du$ and $\Gamma(a,z) = \int_z^\infty u^{a-1} e^{-u}\, du$ are lower and upper incomplete gamma functions, respectively. The hazard function of random variable Y is define as

$$h(y;\theta) = \frac{2\theta^{-\frac{3}{2}}}{y^4 \gamma\left(\frac{3}{2}, \frac{1}{\theta y^2}\right)} \exp\left\{-\frac{1}{\theta y^2}\right\}. \tag{3}$$

The hazard function of InvMWD is a UBT in nature, i.e. it increases sharply in the initial phase of time passes, and then after reaching a peak point it deepens gradually and tends to zero. This means InvMWD represents the lifetime of such individuals who have an increased chance of failing in the early age of life span after survival up to a speci?c age, the rate of failure starts decreasing as age increases.

1.3. Marshall-Olkin Family

Since Marshall and Olkin (1997) proposed a modern methodology to develop a model which is more compatible with real-life experiments than existing lifetime models. In this methodology, a real-valued parameter is added with the existing baseline distribution and constitutes a modern family distribution. This family of distribution is called the Marshall-Olkin (M-O) extended

family of distribution. If X is a random variable (rv) having the cdf $F(X,\theta)$ and it is considered the baseline distribution function then the corresponding cdf of the M-O family can be defined as

$$F_{MO}(x;\alpha,\theta) = \frac{F(x;\theta)}{1-(1-\alpha)S(x;\theta)}; \quad -\infty < y < \infty, \alpha > 0, \tag{4}$$

where, α is called a tilt parameter. Here it is to be noticeable that for $\alpha = 1$, we have $F_{MO}(x,\alpha,\theta) = F(x,\theta)$, i.e. the M-O distribution will reduce to the baseline distribution.

In the last decade, many authors studied the different distributions utilizing the M-O family. Ghitany et al. (2012) proposed a two-parameter M-O extended Lindley distribution and derived some basic statistical properties. They obtained the parameter estimate of the proposed distributions and standard error by utilizing the limiting distribution of maximum likelihood (ML) estimate under randomly censored data. A new distribution, namely M-O Frechet distribution was proposed by Krishna et al. (2013) and discussed the basic statistical properties such as moments, quantiles, Renyi entropy and order statistics. They obtained the point estimates by utilizing three different iterative procedures. Finally, the M-O Frechet distribution is applied to the survival time data. Mansour et al. (2017) proposed Kumaraswamy M-O Lindley distribution having four parameters and derived the statistical properties. They also obtained the parameter estimates and two real data sets analyzed for illustration purposes. Benkhelifa (2017) proposed a new three-parameter model called the M-O extended generalized Lindley distribution. They derived various structural properties of the proposed model including expansions for the density function, ordinary moments, moment generating function, quantile function, mean deviations, Bonferroni and Lorenz curves, order statistics and their moments, Renyi entropy and reliability function. The parameter estimates of the given distribution have been obtained and for illustration purposes, the simulation study and two real data sets have been discussed. Pakungwati et al. (2018) studied the M-O extended Inverse Weibull distribution and applied it to wind speed data. The basic properties and ML estimate derived for M-O length-biased exponential distribution discussed by UL Haq et al. (2019). The distribution applied to the tensile strength of 100 carbon fibers data. Maxwell et al. (2019) proposed a new distribution M-O inverse Lomax distribution by adding a new parameter to the existing inverse Lomax distribution which facilitates modeling of various kinds of data sets. Raffiq et al. (2020) derived a new distribution from the M-O family called M-O inverted Nadarajah-Haghighi distribution. The distribution appeared to give flexible shapes of hazard and pdf that existing model. A three-parameter flexible model named M-O extended inverted Kumaraswamy is derived by Usman and UL Haq (2020). They also obtained the point estimates for model parameters. Finally, simulation and real data studies were done for illustration purposes.

In this paper, we used the M-O family and proposed a new distribution named M-O InvMWD, where InvMWD is considered as a baseline distribution. The pdf and cdf of M-O InvMWD have been established with the discussion of the nature of survival and hazard function in Section 2. In Section 3, some important statistical properties of M-O InvMWD have been derived. In Section 4, the point estimation procedure for the parameters has been discussed by using ML and maximum spacing (MS) methods. We also discuss the asymptotic confidence and boot-p method to calculate the exact confidence intervals for the parameters. In Section 5, we show the applicability of the proposed distribution for several statistical problems. The mathematical expression for record data-based problem, series, parallel and k-out-of-n systems, coherent systems, stress-strength reliability and random censoring, the expressions have been derived. Section 6 deals with the simulation study for the proposed models. The flexibility of the proposed distribution is judged based on the likelihood function, AIC and BIC criteria. The real data analysis is done to support the proposed model setup in Section 7.

2. Marshall-Olkin inverse Maxwell Distribution

Now we use the new methodology proposed by Marshall and Olkin (1997) for the construction of a flexible model. So the given section is completely dedicated to establishing the M-O InvMWD.

For this purpose, the InvMWD is considered the baseline distribution for the M-O family. A rv Y is said to follow the M-O InvMWD with cdf $F_{MO}(y;\alpha,\theta),\ y \in \mathcal{R}^+$, if it is defined as follows

$$F_{MO}(y;\alpha,\theta) = \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]};\quad \alpha,\theta>0. \tag{5}$$

For the incomplete gamma function, we know that $\Gamma(a,0)=\Gamma(a)$ and $\gamma(a,0)=0$ and one can directly write that $\lim_{y\to\infty} F_{MO}(y;\alpha,\theta)=1$. Henceforth, the cdf expressions of $M-O$ InvMWD in (5) represents a proper density.

As it is defined that two parameter points (α_1,θ_1) and (α_2,θ_2) are said to be observationally equivalent if $F(y;\alpha_1,\theta_1)=F(y;\alpha_2,\theta_2)\ \forall\ y\in\mathcal{R}$. Additionally, a parameter point (α^0,θ^0) in $\omega\subset\mathcal{R}^2$ is said to be identifiable if there is no other point (α,θ) in ω which is observationally equivalent.

Lemma 1. The cdf of M-O InvMWD represents a proper density i.e.

$$\lim_{y\to\infty} F_{MO}(y;\alpha,\theta)=1.$$

Proof.

$$\begin{aligned}
\lim_{y\to\infty} F_{MO}(y;\theta,\alpha) &= \lim_{y\to\infty}\left[\frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]}\right]\\
&= \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta\infty^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta\infty^2}\right)\right]}\\
&= \frac{\frac{2}{\sqrt{(\pi)}}\Gamma(\frac{3}{2},0)}{1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma(\frac{3}{2},0)}\\
&= \frac{\frac{2}{\sqrt{\pi}}\times\frac{1}{2}\times\Gamma(\frac{1}{2})}{1-0}\\
&= 1,
\end{aligned}$$

From here we can see that MO-InvMWD has a proper pdf.

Lemma 2. The M-O InvMWD is identifiable i.e.

$$F_{MO}(y;\alpha_1,\theta_1)=F_{MO}(y;\alpha_2,\theta_2)\ \forall\ y\in\mathcal{R}^+ \quad \text{iff} \quad (\alpha_1,\theta_1)=(\alpha_2,\theta_2),$$

where (α,θ) in $\omega\subset\mathcal{R}^2$.

Proof. We known that InvMWD is identifiable, see Tomer and Panwar (2020), i.e. $F(y;\theta_1)=F(y;\theta_2),\ y\in\mathcal{R}^+$ iff $\theta_1=\theta_2$. Let first assume that the two real valued parameter points (α_1,θ_1) and (α_2,θ_2) are such as $(\alpha_1,\theta_1)=(\alpha_2,\theta_2)$. So

$$\begin{aligned}
F_{MO}(y;\alpha_1,\theta_1) &= \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta_1 y^2}\right)}{\left[1-\frac{2(1-\alpha_1)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y^2}\right)\right]}\\
&= \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta_2 y^2}\right)}{\left[1-\frac{2(1-\alpha_2)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y^2}\right)\right]} \qquad \because F(y;\theta_1)=F(y;\theta_2)\\
&= F_{MO}(y;\alpha_2,\theta_2).
\end{aligned}$$

Similarly it can be shown that if $F_{MO}(y;\alpha_1,\theta_1)=F_{MO}(y;\alpha_2,\theta_2)$ then $(\alpha_1,\theta_1)=(\alpha_2,\theta_2)$. Hence the M-O InvMWD is identifiable.

The pdf, $f_{MO}(y;\alpha,\theta)$, of M-O InvMWD can be obtained by using cdf given in (5) such as $f_{MO}(y;\alpha,\theta) = \frac{d}{dy}F_{MO}(y;\alpha,\theta)$. So

$$f_{MO}(y;\alpha,\theta) = \frac{\frac{4\alpha}{\sqrt{\pi}}\frac{1}{y^4\theta^{\frac{3}{2}}}exp\left(-\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^2}; \quad y>0,\alpha,\theta>0. \tag{6}$$

For the arbitrary value of $(\alpha,\theta) = \{(0.5,1.5),(1.0,1.5),(1.8,1.5),(1.8,0.8),(0.8,0.2)\}$, the different patterns of pdf have been drawn in Figure 1. It can be seen from the pdf plot that the M-O InvMWD is a uni-modal and positively skewed distribution.

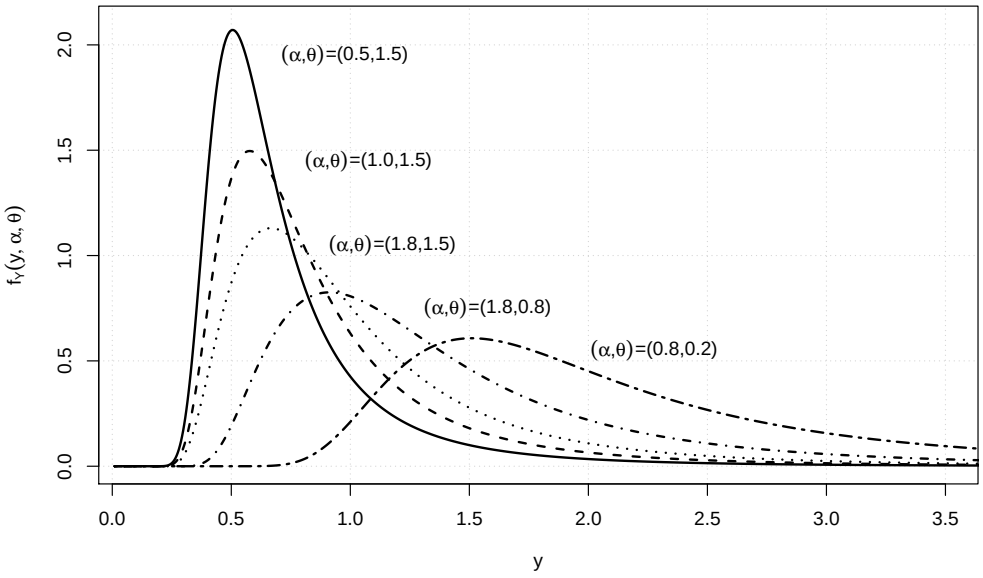


Figure 1: *The probability density function of M-O InvMWD*(α,θ)

The survival function of the M-O InvMWD is given by

$$S_{MO}(y;\alpha,\theta) = 1 - F_{MO}(y;\alpha,\theta) = \frac{\frac{2\alpha}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]} \tag{7}$$

Similarly, the hazard rate function, say $h_{MO}(y;\theta,\alpha)$, of M-O InvMWD can be defined as

$$h_{MO}(y;\alpha,\theta) = \frac{2}{\theta^{\frac{3}{2}}y^4}\frac{\left[\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^{-1}}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]}\exp\left\{-\frac{1}{\theta y^2}\right\} \tag{8}$$

For the arbitrary pair values of parameter (α,θ), the hazard rate functions have been sketched in Figure 2. It can be seen that the hazard rate of M-O InvMWD is upside down bathtub shape and quite flexible with respect to parameters of the distribution. The turning point of hazard rate can be obtained easily by solving the following equation

$$\frac{d}{dy}\ln h_{MO}(y;\alpha,\theta) = 0$$

$$\text{or} \quad 2\sqrt{\pi}\left[\exp\left(-\frac{1}{\theta y^2}\right) - (2\theta-1)\theta^{\frac{1}{2}}y^3\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right] + 4(1-\alpha)\theta^{\frac{1}{2}}(2\theta-1)y^3\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right\}^2 = 0.$$

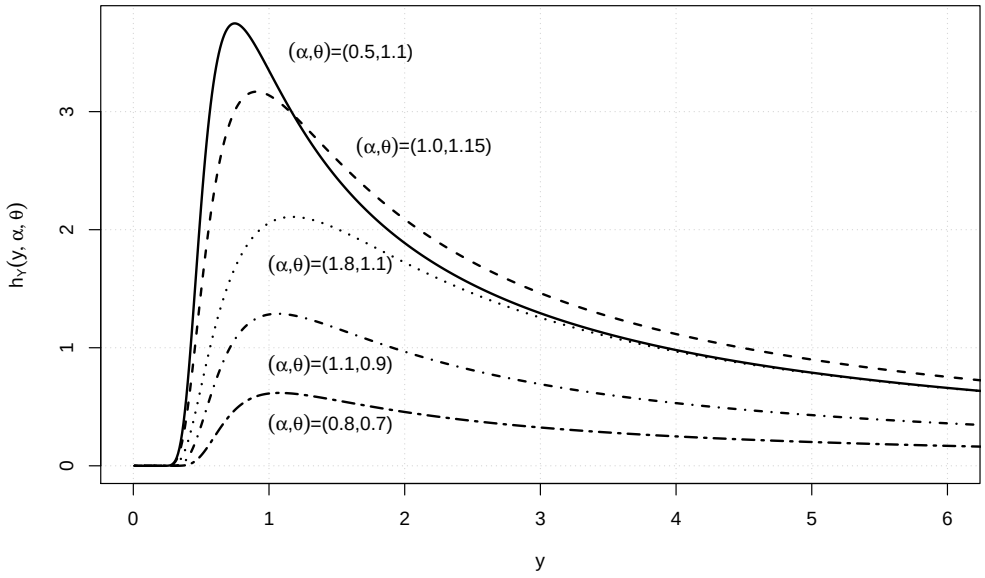


Figure 2: *The hazard rate function of M-O InvMWD*(α, θ).

By using optimization techniques, one can obtain the value of y corresponding to the turning point of the hazard rate function. For the considered choices of $(\theta, \alpha) = \{(0.5, 1.1), (1.0, 1.15), (1.8, 1.1), (1.1, 0.9), (0.8, 0.7)\}$, in Figure 2, the turning points are $0.7473, 0.9085, 1.1730, 1.0642$ and 1.0770, respectively.

3. Statistical Properties

In everyday scenes, lifetime rarely appears in unique applications. Likely researchers are discussing the statistical descriptors including the mean, the range and the variance to understand how these statistics are extracted is one goal for the study of perception. Now we discuss the statistical properties of M-O InvMWD.

3.1. Moments

Many distributions have parameters that control their respective attribute distribution. Central moments are useful because they allow us to quantify properties of distributions in ways that are location-invariant. The moment is the most important characteristic of a distribution function and it can be derived from the functional form of a distribution function. Thus, moments have a great role in defining a distribution theory. If Y follows M-O InvMWD(α, θ), then r^{th} moment about the origin of M-O InvMWD is given by

$$\begin{aligned} E(Y^r) &= \int_{-\infty}^{\infty} y^r f_{MO}(y;\alpha,\theta)dy \\ &= \frac{4\alpha}{\sqrt{\pi}\theta^{\frac{3}{2}}} \int_0^{\infty} \frac{y^{r-4} exp\left(-\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^2} dy \\ &= \frac{4\alpha}{\sqrt{\pi}\theta^{\frac{3}{2}}} \mathcal{G}(r;\alpha,\theta); \qquad \text{for } r \leq 2 \end{aligned} \tag{9}$$

where, $\mathcal{G}(y;\alpha,\theta) = \int_0^\infty y^{r-4}\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^{-2} exp\left(-\frac{1}{\theta y^2}\right)dy$ and this integral can be computed numerically.

3.2. Quantile Function

The quantile function is one way of prescribing a probability density distribution and based on the inverse distribution function. For statistical applications, researchers are required to know key percentage points of a given distribution. The q^{th} quantile, y_q, of the M-O InvMWD can be derived as follows

$$q = F_{MO}(y_q;\alpha,\theta) = \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta y_q^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y_q^2}\right)\right]}.$$

Finally, after some calculations, we get

$$y_q(\alpha,\theta) = \left[\theta\Gamma_{3/2}^{-1}\left(\frac{\sqrt{\pi}}{2}\frac{q\alpha}{1-q(1-\alpha)}\right)\right]^{-\frac{1}{2}}. \tag{10}$$

For a particular choice of $q,\ 0<q<1$, the corresponding quantile value of the distribution can be obtained. If we put $q = 0.5$ in (10), the median, $\bar{X}_{md}$, of the M-O InvMWD can be obtained such as

$$\bar{X}_{md} = \left[\theta\Gamma_{3/2}^{-1}\left(\frac{\sqrt{\pi}}{2}\frac{0.5\alpha}{1-0.5(1-\alpha)}\right)\right]^{-\frac{1}{2}} \tag{11}$$

The quantile function can also be utilized to generate random numbers from M-O InvMWD. If u is a random number from uniform distribution i.e. $u \sim \mathcal{U}(0,1)$ then one can obtain a random number y from M-O InvMWD as follows

$$y = \left[\theta\Gamma_{3/2}^{-1}\left(\frac{\sqrt{\pi}}{2}\frac{u\alpha}{1-u(1-\alpha)}\right)\right]^{-\frac{1}{2}}. \tag{12}$$

3.3. Measures of Skewness and Kurtosis

Since, the Skewness and Kurtosis are two important characteristics of a distribution function. But in this case, it is noticeable that the higher-order moment of the M-O InvMWD does not exist. For this purpose, to calculate the coefficient of skewness and kurtosis, we are using the approach of quantile functions. By using the expression of quantile function given in (10), we used the formula of Galton's measure of skewness and Moor's measure of kurtosis, Gilchrist (2000) are as follows

$$\mathcal{S}(\alpha,\theta) = \frac{y_{0.75}-2y_{0.5}+y_{0.25}}{y_{0.75}-y_{0.25}} \quad \text{and} \quad \mathcal{K}(\alpha,\theta) = \frac{y_{0.875}-y_{0.625}-y_{0.325}+y_{0.125}}{y_{0.75}-y_{0.25}} \tag{13}$$

The range of Galton's measure of skewness $S(.)$ is $(-1,1)$ and a perfectly symmetrical distribution at $S(.) = 0$. A large and positive value of $S(.)$ indicates a long tail to the right, i.e. it indicates that the pdf has a positively skewed distribution and vice versa. From Table 1, it is clear that M-O InvMWD is positively skewed. Galton's skewness is varying significantly with different values of α and θ. Also, Moor's measure of kurtosis is not influenced by the (extreme) tails of the distribution.

3.4. Stochastic Ordering

In probability theory, the stochastic orderings are relating to inequalities between expectations of functions concerning the corresponding distribution. Stochastic ordering of positive continuous

Table 1: *The mean($\bar{x}$), variance($var(x)$), median($\bar{x}_{md}$), skewness and kurtosis for different sets of parameters (α, θ).*

α	θ	$\bar{x}$	$var(x)$	$\bar{x}_{md}$	$\mathcal{S}(\alpha,\theta)$	$\mathcal{K}(\alpha,\theta)$
	0.5	1.31	0.87	1.08	0.48	0.81
0.5	1.0	0.92	0.43	0.76	0.51	0.79
	1.5	0.75	0.29	0.62	0.55	0.78
	0.5	1.59	1.45	1.30	0.51	0.82
1.0	1.0	1.12	0.72	0.91	0.59	0.81
	1.5	0.92	0.48	0.75	0.65	0.80
	0.5	1.80	1.94	1.46	0.54	0.83
1.5	1.0	1.27	0.97	1.03	0.64	0.82
	1.5	1.04	0.64	0.84	0.70	0.81

random variables is used to study the comparative behavior. A random variable X is said to be smaller than a random variable Y in the

1. stochastic order $(X \leq_{st} Y)$ if $F_X(x) \geq F_Y(x)$ for all x;
2. hazard rate order $(X \leq_{hr} Y)$ if $h_X(x) \geq h_Y(x)$ for all x;
3. mean residual order $(X \leq_{mrl} Y)$ if $m_X(x) \leq m_Y(x)$ for all x;
4. likelihood ratio order $(X \leq_{lr} Y)$ if $\left(\frac{f_X(x)}{f_Y(x)}\right)$ decreases in x.

The following result is a well known result and are given below:

$$X \leq_{lr} Y \implies X \leq_{hr} Y \implies X \leq_{mrl} Y \implies X \leq_{st} Y$$

That is, we can see that the likelihood ratio ordering implies the rest of all orderings.

Theorem 1. Let X $\sim$ M-O InvMWD (α_1,θ_1) and Y $\sim$ M-O InvMWD (α_2,θ_2). If $\alpha_1 = \alpha_2 = \alpha$ and $\theta_1 \geq \theta_2$, then (Y$\leq_{lr}$X).

Proof. The likelihood ratio is given by

$$\frac{f_X(x)}{f_Y(x)} = \left(\frac{\theta_2}{\theta_1}\right)^{(3/2)} \exp\left(-\frac{\theta_2-\theta_1}{\theta_1\theta_2 x^2}\right) \left[\frac{1-\frac{2(1-\alpha_2)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 x^2}\right)}{1-\frac{2(1-\alpha_1)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 x^2}\right)}\right]$$

If $\alpha_1 = \alpha_2 = \alpha$ and $\theta_1 > \theta_2$, then $\frac{d}{dx}\frac{f_X(x)}{f_Y(x)} \leq 0$, which implies that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$.

4. Parameter Estimation

4.1. Maximum Likelihood Estimation

The Maximum likelihood procedure is to determine the values for the unknown parameters of a model. The obtained parameter values are such that they maximize the likelihood function that the process described by the model produced the data that is actually observed. In this section, the unknown parameters of M-O InvMWD are to be estimated by using maximum likelihood estimation techniques. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the M-O InvMWD

population, and the likelihood function is given by

$$L(\alpha,\theta|y) = \prod_{i=1}^{n} f(y_i;\alpha,\theta)$$
$$= \prod_{i=1}^{n} \frac{\frac{4\alpha}{\sqrt{\pi}} \frac{1}{y_i^4 \theta^{\frac{3}{2}}} exp\left(-\frac{1}{\theta y_i^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}} \gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right]^2}$$

On taking the natural logarithm of $L(\alpha,\theta|y)$ both side, the log-likelihood function is given by

$$l(\alpha,\theta|y) \propto n\ln(\alpha) - \frac{3n}{2}\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n}\frac{1}{y_i^2} - 2\sum_{i=1}^{n}\ln\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right]$$

We can use the following normal equations for obtaining the maximum likelihood estimates of θ and α. So, we have

$$-\frac{3n}{2\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}\frac{1}{y_i^2} + \frac{4(1-\alpha)}{\sqrt{\pi}\theta^{5/2}}\sum_{i=1}^{n}\frac{\frac{1}{y_i^3}\exp\left(-\frac{1}{\theta y_i^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right]} = 0 \tag{14}$$

and

$$\frac{n}{\alpha} - \frac{4}{\sqrt{\pi}}\sum_{i=1}^{n}\frac{\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right]} = 0 \tag{15}$$

The above equations give the roots for unknown parameters of M-O InvMWD. The above nonlinear equations can be solved by using any iterative method. Let $\hat{\theta}$ and $\hat{\alpha}$ be the ML estimate of θ and α, respectively, after solving the above equations.

4.2. Maximum Spacing Estimation

The method of "maximum product of spacings" (MPS) was proposed by Cheng and Amin (1979). The method is based on the maximization of the geometric mean of spacings in the data, which are the differences between the values of the cumulative distribution function at neighboring data points. Here our main aim is to estimate the unknown parameters θ and α of the distribution function. The idea of MPS is to make the observed data as uniform as possible, based on a quantitative measure of uniformity. Let $y_1 < y_2, \ldots, < y_n$ be the complete ordered sample. Also, let us define some quantities

$$D_1 = F(y_1;\alpha,\theta) \tag{16}$$
$$D_{n+1} = 1 - F(y_n;\alpha,\theta) \tag{17}$$

And the general term we can write for the spacings given by,

$$D_i = F(y_{i:n};\alpha,\theta) - F(y_{(i-1):n};\alpha,\theta) = \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \frac{1}{\theta y_{i:n}^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_{i:n}^2}\right)\right]} - \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \frac{1}{\theta y_{(i-1):n}^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_{(i-1):n}^2}\right)\right]} \tag{18}$$

such that $\sum D_i = 1$. The MPS method choose θ which maximizes the product of spacing or other words it maximizes the geometric mean of the spacings, i.e.

$$G = \left(\prod_{i=1}^{n+1} D_i\right)^{\frac{1}{n+1}}$$

Taking the logarithm of above equation, we get

$$S = \frac{1}{n+1}\sum_{i=1}^{n+1} \ln D_i$$

Also we can write S as

$$\begin{aligned}
S =& \frac{1}{n+1}\left\{\ln D_1 + \sum_{i=2}^{n} \ln D_i + \ln D_{n+1}\right\} \\
=& \frac{1}{n+1}\ln\left(\frac{2}{\sqrt{\pi}}\right) + \frac{1}{n+1}\ln\left(\Gamma\left(\frac{3}{2}, \frac{1}{\theta y_1^2}\right)\right) - \frac{1}{n+1}\ln\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_1^2}\right)\right] \\
&+ \frac{1}{n+1}\sum_{i=2}^{n}\ln\left[\frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)}{\left\{1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right\}} - \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \frac{1}{\theta y_{i-1}^2}\right)}{\left\{1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_{i-1}^2}\right)\right\}}\right] \\
&+ \frac{1}{n+1}\ln\left[1 - \frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \frac{1}{\theta y_n^2}\right)}{\left\{1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_n^2}\right)\right\}}\right]
\end{aligned}$$

After differentiating the above equation with respect to parameters and equating them to zero, we get the normal equations from which we can obtain the required estimates.

4.3. Variance-Covariance Matrix

We have obtained the variance-covariance matrix to find out the asymptotic confidence intervals. The observed information matrix for the parameter is given by inverting the second derivative matrix with respect to the given parameters. Therefore, we get the observed approximate Fisher's Information matrix which is given by

$$I(\hat{\underline{\zeta}}) = \begin{bmatrix} I_{\alpha\alpha} & I_{\alpha\theta} \\ I_{\theta\alpha} & I_{\theta\theta} \end{bmatrix}\Bigg|_{(\alpha,\theta)=(\hat{\alpha},\hat{\theta})},$$

where, $\underline{\zeta} = (\alpha, \theta)$ is the parameter vector, and

$$I_{\theta\theta} = -\frac{\partial^2 l}{\partial\theta^2} = -\frac{3n}{2\theta^2} + \frac{2}{\theta^3}\sum_{i=1}^{n}\frac{1}{y_i^2} - \frac{4(1-\alpha)}{\sqrt{\pi}}\sum_{i=1}^{n}\psi(y_i;\alpha,\theta),$$

$$I_{\theta\alpha} = -\frac{\partial^2 l}{\partial\theta\partial\alpha} = \frac{4}{\sqrt{\pi}\theta^{\frac{3}{2}}}\sum_{i=1}^{n}\frac{\left[\frac{2(1-\alpha)}{\sqrt{\pi}}\exp\left(-\frac{1}{\theta y_i^2}\right)\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right) - \frac{1}{y_i^3}\exp\left(-\frac{1}{\theta y_i^2}\right)\left\{1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right\}\right]}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right]^2},$$

$$I_{\alpha\theta} = I_{\theta\alpha},$$

$$I_{\alpha\alpha} = -\frac{\partial^2 l}{\partial\alpha^2} = \frac{n}{\alpha^2} + \frac{8}{\pi}\sum_{i=1}^{n}\frac{\left\{\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right\}^2}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right]^2},$$

where, $\psi(y_i;\alpha,\theta)$ is given by

$$\begin{aligned}
\psi(y_i;\alpha,\theta) =& \frac{2(1-\alpha)}{\sqrt{\pi}}\xi(y_i,\theta)\phi(y_i;\alpha,\theta)\left[-\frac{1}{\theta^2 y_i^2} - \frac{5}{2\theta} - \xi(y_i,\theta)\right] \\
&+ \left\{\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right\}^{-1}\xi(y_i,\theta)\phi(y_i;\alpha,\theta)\left(\frac{5}{2\theta} + \frac{1}{\theta y_i^2}\right)
\end{aligned}$$

and $\xi(y_i,\theta) = \theta^{-5/2}y_i^{-3}\exp\left(-\frac{1}{\theta y_i^2}\right)\left\{\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right\}^{-1}$, $\phi(y_i;\alpha,\theta) = \left[\left\{\gamma\left(\frac{3}{2}, \frac{1}{\theta y_i^2}\right)\right\}^{-2} - \frac{2(1-\alpha)}{\sqrt{\pi}}\right]^{-2}$

respectively. The approximate asymptotic variance-covariance matrix for the parameters θ and α

based on MLE can be found by inverting $I(\hat{\underline{\zeta}})$ as

$$I^{-1}(\hat{\underline{\zeta}}) = \begin{bmatrix} Var(\hat{\theta}) & Cov(\hat{\theta},\hat{\alpha}) \\ Cov(\hat{\alpha},\hat{\theta}) & Var(\hat{\alpha}) \end{bmatrix}$$

Thus, using above equation, we get the $100(1-\gamma)\%$ confidence limits for $\hat{\theta}$ and $\hat{\alpha}$ given by $\hat{\theta} \pm z_{\frac{\gamma}{2}} SE(\hat{\theta})$ and $\hat{\alpha} \pm z_{\frac{\gamma}{2}} SE(\hat{\alpha})$ respectively, where $z_{(\frac{\gamma}{2})}$ is upper $100(\frac{\gamma}{2})^{th}$ percentile of standard normal variate.

4.4. Boot-p Method

The Bootstrap method is a resampling technique and used to estimate the statistic of the population by using the sampling technique with replacement. In some situations of distribution theory, the ACI does not provide the appropriate confidence interval for the parameters. This technique is used to construct confidence intervals, calculate the standard errors and perform hypothesis testing for several types of sample statistics. We have applied the boot-p method to calculate the confidence intervals for parameters. For more details, one can cite the article Tibshirani and Efron (1993). The necessary steps for applying the parametric bootstrap method are given below:

1. Based on the original sample $\underline{y} = (y_1, y_2, \ldots, y_n)$, obtain the MLE of $\hat{\zeta} = (\hat{\alpha}, \hat{\theta})$.
2. Under the same conditions to generate the sample, say $(x_1, x_2, \ldots, x_m)$, from the underlying distribution M-O InvMWD $(\hat{\zeta})$ with parameter $\hat{\zeta}$.
3. Compute the MLE of $\hat{\zeta}$ based on observed sample $(x_1, x_2, \ldots, x_m)$, say $\hat{\zeta}^*$.
4. Repeat step (2) and (3) B times and obtain $\hat{\zeta}_1^*, \hat{\zeta}_2^*, \ldots, \hat{\zeta}_B^*$.
5. Arrange $\hat{\zeta}_1^*, \hat{\zeta}_2^*, \ldots, \hat{\zeta}_B^*$ in ascending order.
6. A two-sided $100(1-\gamma)\%$ percentile bootstrap confidence interval of ζ, say $[\hat{\zeta}_L^*, \hat{\zeta}_U^*]$ is given by $[\hat{\zeta}_L^*, \hat{\zeta}_U^*] = \left[\hat{\zeta}_B^{*(\frac{\gamma}{2})}, \hat{\zeta}_B^{*(1-\frac{\gamma}{2})}\right]$

5. Statistical Application

In this section, we have presented the application of the proposed distribution in various real-life situations. Here we can show the record value estimation procedure for the M-O InvMWD. The reliability function is also derived when the system has arranged in k-out-of-n configuration (a special case of series and parallel system) when all components are *i.i.d.*. The stress-strength reliability is also discussed here. The estimation procedure for the parameter when the data is random censored. We discuss in brief the following the necessary procedures.

5.1. Order Statistics

Let $Y_1, Y_2, \ldots, Y_n$, be a random sample from the M-O InvMWD and $z_1, z_2, \ldots, z_n$ are the ascending order with their magnitude of observed sample. Then pdf of the j^{th} order statistic of M-O InvMWD is given by

$$f_{j:n}(z) = \frac{n!}{(j-1)!(n-j)!} f_{MO}(z) F_{MO}(z)^{(j-1)} \left[1 - F_{MO}(z)\right]^{(n-j)} \tag{19}$$

By putting j=1 and j=n in (19), we can obtain the distributions of minimum and maximum order statistics for M-O InvMWD respectively.

$$\begin{aligned} f_{1:n}(y) &= n f_{MO}(y)\left[1-F_{MO}(y)\right]^{(n-1)} \\ &= n\left(\frac{\frac{4\alpha}{\sqrt{\pi}}\frac{1}{y^4\theta^{\frac{3}{2}}}exp\left(-\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^2}\right)\left[\frac{\frac{2\alpha}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]}\right]^{(n-1)} \\ &= \frac{2^{n+1}\pi^{-\frac{n}{2}}n\alpha^n\theta^{-\frac{3}{2}}y^{-4}exp\left(-\frac{1}{\theta y^2}\right)\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right\}^{(n-1)}}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^{(n+1)}} \end{aligned}$$

and

$$\begin{aligned} f_{n:n}(y) &= n f_{MO}(y)F_{MO}(y)^{(n-1)} \\ &= n\left(\frac{\frac{4\alpha}{\sqrt{\pi}}\frac{1}{y^4\theta^{\frac{3}{2}}}exp\left(-\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^2}\right)\left[\frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]}\right]^{(n-1)} \\ &= \frac{2^{n+1}\pi^{-\frac{n}{2}}n\alpha\theta^{-\frac{3}{2}}y^{-4}exp\left(-\frac{1}{\theta y^2}\right)\left\{\Gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right\}^{(n-1)}}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta y^2}\right)\right]^{(n+1)}} \end{aligned}$$

5.1.1 Record Estimation

In many real-life situations instead of collecting the whole data, we collect data related to record-breaking observations which are known as "records". An observation is considered an upper (lower) record if it is greater (smaller) than all previous observations. Doostparast and Balakrishnan (2010) used the exponential record data to analyze the optimal sample size and associated optimum cost of the experiment. The record data estimation and prediction for gamma distribution derived by Sultan et al. (2008). Here, we are interested to estimate the parameters under record value setup.

Let $Y_1, Y_2, \ldots$ be a sequence of *i.i.d.* random variables having cdf and pdf given by (5) and (6). An observation Y_j is considered as an upper record value if it exceeds that of all previous observations. In other words, we say that if Y_j is an upper record value if $Y_j > Y_i$ for all $i < j$. Let $r = (r_1, r_2, \ldots, r_m)$ be the first observed m upper record values from the parent distribution with pdf given in (6). The joint pdf of given data can be constructed by method given by Arnold et al. (2011) as below

$$f(r;\theta,\alpha) = \prod_{i=1}^{m-1} h_{MO}(r_i;\alpha,\theta) f_{MO}(r_m;\alpha,\theta); \quad -\infty < r_1 < r_2 < \ldots < r_m < \infty, \tag{20}$$

where, $h_{MO}(r_i;\alpha,\theta) = \frac{f_{MO}(r_i;\alpha,\theta)}{1-F_{MO}(r_i;\alpha,\theta)}$. Thus, the likelihood function under upper record value is given by

$$\begin{aligned} L(\alpha,\theta|r) &= \bar{F}_{MO}(r_m;\alpha,\theta)\prod_{i=1}^{m}\frac{f_{MO}(r_i;\alpha,\theta)}{\bar{F}_{MO}(r_i;\alpha,\theta)} \\ &= \frac{\frac{2\alpha}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)\right]}\prod_{i=1}^{m}\frac{2\exp\left(-\frac{1}{\theta r_i^2}\right)\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)\right\}^{-1}}{\theta^{\frac{3}{2}}r_i^4\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)\right]} \end{aligned}$$

The log-likelihood function is given as

$$l(\alpha,\theta|r) \propto \ln(\alpha) + \ln\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)\right\} - \ln\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)\right] - \frac{1}{\theta}\sum_{i=1}^{m}\frac{1}{r_i^2}$$
$$-\frac{3m}{2}\ln(\theta) - \sum_{i=1}^{m}\ln\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)\right\} - \sum_{i=1}^{m}\ln\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)\right]$$

Differentiating above log-likelihood function with respect to θ and α, we get

$$\hat{\theta}^{\frac{3}{2}} = \frac{\frac{r_m^{-3}\exp\left(-\frac{1}{\theta r_m^2}\right)}{\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)} + \frac{\frac{2(1-\alpha)}{\sqrt{\pi}r_m^3}\exp\left(-\frac{1}{\theta r_m^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)\right]} - \sum_{i=1}^{m}\frac{r_i^{-3}\exp\left(-\frac{1}{\theta r_i^2}\right)}{\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)} + \sum_{i=1}^{m}\frac{\frac{2(1-\alpha)}{\sqrt{\pi}r_i^3}\exp\left(-\frac{1}{\theta r_i^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)\right]}}{\left[\frac{1}{\theta}\sum_{i=1}^{m}\frac{1}{r^2} - \frac{3m}{2}\right]} \tag{21}$$

$$\hat{\alpha}^{-1} = \frac{\frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_m^2}\right)\right]} + \sum_{i=1}^{m}\frac{\frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r_i^2}\right)\right]} \tag{22}$$

The equations (21) and (22) are not obtained in closed form, so required estimates can be obtained by using any iterative technique.

5.2. System Lifetime Distribution

5.2.1 k out of n System

The maximum likelihood estimate for reliability of k-out-of-n systems which are composed of n *iid* components having M-O InvMWD lifetimes are discussed. The system is operational if and only if at least k of out the n components are operational at the given time. In other words, as soon as $(n-k+1)$ components fail, the system fails. For example, we have three generators at an electric power plant station in which a minimum of two must be operational at all times in order to deliver the required power. Such a system is called the 2-out-of-3 system. The system reliability estimation procedure for the failure of uncensored cases (where there are n units put on test which is terminated when all the units have failed) is discussed. In particular, the system having configuration 1-out-of-n is known as parallel while n-out-of-n is known as series systems. By using this configuration, we assume that the failure time distributions of the components are independent. Then we can define the probability of exactly k out of n components functioning as:

$$P(X=k) = \binom{n}{k}\{R_{MO}(t)\}^k\{1-R_{MO}(t)\}^{(n-k)}; \quad for \quad k=0,1,2,\ldots,n,$$

where $R_{MO}(t)$ is defined as the reliability function of each component having M-O InvMWD. Also, the system failure time density is given as

$$f_s(t) = \frac{n!}{(n-k)!(k-1)!}\{R_{MO}(t)\}^k\{1-R_{MO}(t)\}^{(n-k)}f_{MO}(t); \quad for \quad k=0,1,2,...,n.$$

Thus, the reliability of k-out-of-n system, simply, is

$$R_s(t) = \sum_{i=k}^{n}\binom{n}{i}\{R_{MO}(t)\}^i\{1-R_{MO}(t)\}^{(n-i)} \tag{23}$$

5.2.2 Series and Parallel System

The series and parallel systems can be viewed as special cases of k-out-of-n systems. Suppose we have n such systems that each have k-components in series or parallel attachment. Let Y_j, $j \in \{1,2,...,k\}$, denote the sequence of failure times of all components in a system. We assume

that the sequence is composed of independent but not identical from M-O InvMWD. In this case, at a failed component, there are two observed quantities are recorded say (T,δ), where $T = min(y_1, y_2, \dots y_n)$ for the series system and for parallel system $T = max(y_1, y_2, \dots y_n)$ with $\delta = j$ if $T = Y_j$ for $j = 1, 2, \dots, k$. The δ quantity can be viewed as an indicator function of the component that caused the system failure. Consider a sample of size n be independent and identically distributed systems (either all series or all parallel systems). The observations are represented by $(T,\delta) = \{(T_i, \delta_i) : i = 1, ..., n\}$. Then, the reliability of the j^{th} component is given by $R_j(t) = P(X_j > t), j = 1, 2, ..., k$. Let us define the random variables $Y_j's$ for component's reliability with M-O InvMWD distributions parameterized by $\zeta_j = (\alpha_j, \theta_j)$, that is,

$$P(Y_j > y|\zeta_j) = R(y|\zeta_j) = \frac{\frac{2\alpha_j}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta_j y^2}\right)}{\left[1 - \frac{2(1-\alpha_j)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta_j y^2}\right)\right]} \tag{24}$$

When all components of a system are connected in series then

$$R(t) = \prod_{j=1}^{k} \frac{\frac{2\alpha_j}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta_j y^2}\right)}{\left[1 - \frac{2(1-\alpha_j)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta_j y^2}\right)\right]}$$

and system reliability when component put in parallel setup

$$R(t) = 1 - \prod_{j=1}^{k} \frac{\frac{2}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta_j y^2}\right)}{\left[1 - \frac{2(1-\alpha_j)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta_j y^2}\right)\right]}$$

By using the component reliability given in (24), we can obtain the values for the reliability of series and parallel systems.

5.3. Stress-Strength Reliability

In reliability theory, the stress-strength reliability is denoted by quantity $R = P(W > V)$, where W and V denotes the strength and stress of the system. In this regard, when the stress is greater than strength, the system will fail. The applicability of probability R is that it can be used to compare the two random variables encountered in various applied fields so the estimation of R is a great concern for a long time. Kundu and Gupta (2006) discussed the point and interval estimation procedure for stress-strength Weibull model under classical and Bayesian approaches. Chaudhary et al. (2017) analyzed the stress-strength reliability estimates when stress and strength both follow Maxwell lifetime. So for the proposed model, we have calculated expressions for stress-strength reliability estimates. Consider $W \sim$ M-O InvMWD(α,θ) and $V \sim$ M-O InvMWD(β,θ) and W and V are independently distributed. The expression for R comes out to be as:

$$\begin{aligned}
R = P(W > V) &= \int_0^\infty f_{MO}(w;\alpha,\theta) F_{MO}(w;\beta,\theta) dw \\
&= \int_0^\infty \left(\frac{\frac{4\alpha}{\sqrt{\pi}} \frac{1}{w^4 \theta^{\frac{3}{2}}} exp\left(-\frac{1}{\theta w^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta w^2}\right)\right]^2} \right) \times \left(\frac{\frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \frac{1}{\theta w^2}\right)}{\left[1 - \frac{2(1-\beta)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta w^2}\right)\right]} \right) dw \\
&= \frac{8\alpha}{\pi\theta^{\frac{3}{2}}} \int_0^\infty \frac{\Gamma\left(\frac{3}{2}, \frac{1}{\theta w^2}\right) \exp\left(-\frac{1}{\theta w^2}\right)}{\left[1 - \frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta w^2}\right)\right]^2 \left[1 - \frac{2(1-\beta)}{\sqrt{\pi}}\gamma\left(\frac{3}{2}, \frac{1}{\theta w^2}\right)\right]} dw \\
&= \frac{8\alpha}{\pi\theta^{\frac{3}{2}}} I(r;\alpha,\beta,\theta).
\end{aligned} \tag{25}$$

The quantity $I(r;\alpha,\beta,\theta) = \int_0^\infty \frac{\Gamma\left(\frac{3}{2},\frac{1}{\theta r^2}\right)\exp\left(-\frac{1}{\theta r^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r^2}\right)\right]^2\left[1-\frac{2(1-\beta)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta r^2}\right)\right]}dr$ can be calculated with help of statistical software for given values of parameters and thus numerical value of R can be also be obtained.

5.4. Random Censored Data

In real-life experiments, we are unable to observe the complete failure of the sample due to the insufficient necessary resource. Since, conducting life experiments is time taking and more expensive which demands a large amount of money, labor, and time. For reducing the cost and time of the experiments, various kinds of censoring schemes are developed in the statistical literature. In this case, when the researchers are interested in analyzing the partial part of the sample, say censored data. A special type of censoring scheme known as random censoring occurs in literature when the item is lost or removed randomly from the experiment before its failure under the study. A sample is randomly censored when the experimental unit and censoring time points are random and independent of each other outcomes. In real-life situations, especially in clinical trials, the patients do not complete the course of treatment and they leave due to several factors before the termination point of the experiment. Nandi and Dewan (2010) discussed the parameter estimation procedure for bivariate Weibull distribution under random censoring. In the article authors Kumar and Garg (2014), the authors obtained ML and Bayes estimates of parameters for generalized Inverse Rayleigh distribution under random censoring. Krishna et al. (2015) discussed the classical and Bayesian estimation procedure for the Maxwell distribution random censored sample. Kumar and Kumar (2019) obtained the ML and Bayes estimates of Inverse Weibull distribution parameters under random censoring. Here our interest lies in dealing with M-O InvMWD distribution under the random censoring setup.

Suppose n items are put on test with their lifetimes as $X_1, X_2, \ldots, X_n$ which are *iid* random variables with cdf and pdf given in (5) and (6). Also, let $T_1, T_2, \ldots, T_n$ be the random censoring times. Let pdf and cdf of $T_i's$ be $f_T(t)$ and $F_T(t)$, respectively. Further, let $X_i's$ and $T_i's$ be mutually independent. Note that, between $X_i's$ and $T_i's$, only one will actually be observed. Let the actual observation time be $Y_i = min(X_i, T_i); i = 1, 2, \ldots, n$. Also, define the indicator variable δ_i as

$$\delta_i = \begin{cases} 1, & \text{if } X_i \leq T_i \\ 0, & \text{if } X_i > T_i \end{cases}$$

Note that δ_i is a random variable with Bernoulli probability mass function given by

$$P[\delta_i = j] = p^j(1-p)^{(1-j)}; \quad j = 0, 1 \quad \& \quad p = P[X_i \leq T_i]$$

Since $X_i's$ and $T_i's$ are independent, so will be $Y_i's$ and $\delta_i's$. Now, it is simple to show that the joint probability density function of Y and Δ is

$$f_{Y,\Delta}(y,\delta) = \{f_X(y)\,(1-F_T(y))\}^{\delta}\,\{f_T(y)\,(1-F_X(y))\}^{1-\delta}; y, \lambda, \theta \geq 0, \delta = 0, 1.$$

Taking the distribution function of X & T as M-O InvMWD(α,θ_1) and M-O InvMWD(α,θ_2), respectively, and putting the expressions for their pdf & cdf in above equation, we get

$$\begin{aligned} f_{Y,\Delta}(y,\delta;\alpha,\theta_1,\theta_2) = & \frac{8\alpha^2}{\pi y^4}\theta_1^{-\frac{3\delta}{2}}\theta_2^{-\frac{3(1-\delta)}{2}}\exp\left\{-\left(\frac{\delta}{\theta_1 y^2}+\frac{(1-\delta)}{\theta_2 y^2}\right)\right\}\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y^2}\right)\right\}^{\delta} \\ & \left\{\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y^2}\right)\right\}^{(1-\delta)}\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y^2}\right)\right]^{-(\delta+1)} \\ & \left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y^2}\right)\right]^{-(\delta-2)} \end{aligned}$$

Let $(y,\delta)=\{(y_1,\delta_1),(y_2,\delta_2),\ldots,(y_n,\delta_n)\}$ be a randomly censored sample from the above model. Now, the likelihood function can be given by

$$L(\alpha,\theta_1,\theta_2|y,\delta)\propto \alpha^{2n}\theta_1^{-\frac{3m}{2}}\theta_2^{-\frac{3(n-m)}{2}}\prod_{i=1}^{n}\frac{1}{y^4}\exp\left\{-\left(\sum_{i=1}^{n}\frac{\delta_i}{\theta_1 y_i^2}+\sum_{i=1}^{n}\frac{(1-\delta_i)}{\theta_2 y_i^2}\right)\right\}$$
$$\prod_{i=1}^{n}\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)\right\}^{\delta_i}\prod_{i=1}^{n}\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)\right\}^{(1-\delta_i)}$$
$$\prod_{i=1}^{n}\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)\right]^{-(\delta_i+1)}\prod_{i=1}^{n}\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)\right]^{-(\delta_i-2)}$$

Taking natural logarithm above expression, we get the log-likelihood function as

$$l(\alpha,\theta_1,\theta_2)\propto 2n\ln(\alpha)-\frac{3m}{2}\ln(\theta_1)-\frac{3(n-m)}{2}\ln(\theta_2)-\frac{1}{\theta_1}\sum_{i=1}^{n}\frac{\delta_i}{y_i^2}-\frac{1}{\theta_2}\sum_{i=1}^{n}\frac{(1-\delta_i)}{y_i^2}$$
$$+\sum_{i=1}^{n}\delta_i\ln\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)\right\}+\sum_{i=1}^{n}(1-\delta_i)\ln\left\{\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)\right\}$$
$$-\sum_{i=1}^{n}(\delta_i+1)\ln\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)\right]-\sum_{i=1}^{n}(\delta_i-2)\ln\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)\right]$$

Here, quantity $m=\sum_{i=1}^{n}\delta_i$ denotes the number of observed failures. On differentiating the log-likelihood equation with respect to θ_1, θ_2 and α, we get the normal equations for the parameters

$$\hat{\theta}_1=\frac{\sum_{i=1}^{n}\frac{(1-\delta_i)y_i^{-3}\exp\left(-\frac{1}{\theta_1 y_i^2}\right)}{\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)}-\frac{2(1-\alpha)}{\sqrt{\pi}}\sum_{i=1}^{n}\frac{(1+\delta_i)y_i^{-3}\exp\left(-\frac{1}{\theta_1 y_i^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)\right]}}{\left[-\frac{3m}{2}+\frac{1}{\theta_1}\sum_{i=1}^{n}\frac{\delta_i}{y_i^2}\right]} \tag{26}$$

$$\hat{\theta}_2=\frac{\sum_{i=1}^{n}\frac{\delta_i y_i^{-3}\exp\left(-\frac{1}{\theta_2 y_i^2}\right)}{\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)}-\frac{2(1-\alpha)}{\sqrt{\pi}}\sum_{i=1}^{n}\frac{(\delta_i-2)y_i^{-3}\exp\left(-\frac{1}{\theta_2 y_i^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)\right]}}{\left[-\frac{3(n-m)}{2}+\frac{1}{\theta_2}\sum_{i=1}^{n}\frac{(1-\delta_i)}{y_i^2}\right]} \tag{27}$$

$$\hat{\alpha}^{-1}=\frac{1}{n\sqrt{\pi}}\left[\sum_{i=1}^{n}\frac{(1+\delta_i)\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_1 y_i^2}\right)\right]}+\sum_{i=1}^{n}\frac{(\delta_i-2)\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)}{\left[1-\frac{2(1-\alpha)}{\sqrt{\pi}}\gamma\left(\frac{3}{2},\frac{1}{\theta_2 y_i^2}\right)\right]}\right] \tag{28}$$

These equations are obtained after some simplifications, but we are not able to get closed form, so for finding roots of above equations we have to apply some iterative methods.

6. Simulation Study

In this section, we give illustrations based on the simulation study. We generate the random sample for the assumption the failure observation follows M-O InvMWD with corresponding parameters α and θ. A simulation study has been performed by setting different initial values of parameters α and θ. In this regard, we generate the random sample by using the inverse cdf method. The necessary steps for generating the random sample from M-O InvMWD are given below:

1. Set the initial values of n, α and θ.
2. Generate a standard uniform number, $u\sim U(0,1)$.

3. By using the quantile formula in (12), we obtain the value of the random variable y.
4. Repeat (2) to (3) n times to get a sample $(y_1, y_2, \ldots, y_n)$ of size n from M-O InvMWD(α,θ).

6.1. Simulation Analysis (For section - 4.1 to 4.4)

The simulated sample is generated by using the necessary steps described above and using the initial values of parameters $\alpha = 0.50, 1.20$ and $\theta = 1.50, 0.85$. The simulation study is done for different sample sizes $n = 40$, 50 and 60 for 5000 iterations. Here, we discuss two methods for obtaining the point and interval estimates for the unknown parameters. For this, two estimation procedures i.e. (i) maximum likelihood method and (ii) maximum product spacing and for interval estimation (i) asymptotic confidence interval and (ii) boot-p methods are used. We present the estimated values of unknown parameters along with their mean square error (MSE) and absolute bias (AB) in Tables 2 and 3. The MLE's are consistent as we see the value of MSE and AB decrease with increase in sample size. The interval estimates of parameters under asymptotic normality assumption and boot-p method are given in Tables 4 and 5.

Table 2: *Average values of MSE and Absolute Bias under varying sample sizes* 40, 50 *and* 60 *for simulated data under ML and MPS estimates and parameter values* $\alpha = 0.50$ *and* $\theta = 1.50$.

		ML		MPS	
n		$\hat{\theta}$	$\hat{\alpha}$	$\hat{\theta}^*$	$\hat{\alpha}^*$
40	Estimate	1.4999	0.5401	1.3293	0.4197
	MSE	0.1334	0.1215	0.1224	0.0748
	AB	0.2861	0.2423	0.2903	0.2220
50	Estimate	1.4845	0.5327	1.3832	0.4513
	MSE	0.1041	0.1012	0.1032	0.0730
	AB	0.2463	0.2244	0.2586	0.2044
60	Estimate	1.4944	0.5187	1.3841	0.4361
	MSE	0.0906	0.0699	0.0832	0.0527
	AB	0.2340	0.1944	0.2339	0.1773

Table 3: *Average values of MSE and Absolute Bias under varying sample sizes* 40, 50 *and* 60 *for simulated data under ML and MPS estimates and parameter values* $\alpha = 1.20$ *and* $\theta = 0.85$.

		ML		MPS	
n		$\hat{\theta}$	$\hat{\alpha}$	$\hat{\theta}^*$	$\hat{\alpha}^*$
40	Estimate	0.8527	1.3164	0.7207	0.9671
	MSE	0.0861	1.0600	0.0623	0.5011
	AB	0.2132	0.6542	0.2091	0.5708
50	Estimate	0.8520	1.2998	0.7407	0.9892
	MSE	0.0699	0.8172	0.0522	0.4070
	AB	0.1941	0.5824	0.1921	0.5166
60	Estimate	0.8518	1.2632	0.7660	1.0381
	MSE	0.0513	0.5179	0.0481	0.3849
	AB	0.1788	0.5164	0.1794	0.4975

It can be observed from the simulation Tables 2, 3, 4 and 5 that

1. The average values of MSE and AB decreases as the sample size increases for both parameters.

Table 4: *Interval estimates under ML and Boot-p methods along with their average lengths, shape and coverage probabilities (CP) for varying sample sizes* 40, 50 *and* 60 *and parameter values* $\alpha = 0.50$ *and* $\theta = 1.50$.

		ACIs		Boot-p	
n		$\hat{\theta}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\alpha}$
40	Length	1.5123	1.6333	1.5149	1.5356
	Shape	1.0000	1.0000	1.5180	2.6335
	CP	0.94	0.95	0.92	0.93
50	Length	1.3146	1.3633	1.3044	1.2420
	Shape	1.0000	1.0000	1.4427	2.3569
	CP	0.93	0.94	0.95	0.96
60	Length	1.1910	1.1622	1.1783	1.0975
	Shape	1.0000	1.0000	1.4033	2.2023
	CP	0.93	0.95	0.92	0.93

Table 5: *Interval estimates under ML and Boot-p methods along with their average lengths, shape and coverage probabilities (CP) for varying sample sizes* 40, 50 *and* 60 *and parameter values* $\alpha = 1.20$ *and* $\theta = 0.85$.

		ACIs		Boot-p	
n		$\hat{\theta}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\alpha}$
40	Length	1.2360	5.1358	1.1441	4.1830
	Shape	1.0000	1.0000	1.7499	3.0104
	CP	0.93	0.95	0.90	0.93
50	Length	1.0681	4.1173	1.0567	3.8256
	Shape	1.0000	1.0000	1.6795	2.7408
	CP	0.92	0.94	0.92	0.93
60	Length	0.9392	3.3435	0.9315	3.2209
	Shape	1.0000	1.0000	1.6105	2.5208
	CP	0.92	0.95	0.93	0.95

2. ML and MPS methods perform almost equally well.
3. The average values of lengths decreases as sample size increases.
4. The value of shape under the boot-p method indicates that the proposed distribution is positively skewed.
5. The ACIs and boot-p confidence intervals give almost equal performances.
6. High coverage probability values indicate the proportion of the true value to lie in interval estimate is good.

7. Real Data Study

In this section, two real data sets have been analyzed to show the applicability of the proposed model in real life situations. The first data is based on measurements of glycosaminoglycans (GAG) concentration in urine and second data is the relief time of patients receiving an analgesic.

Data 1: GAGurine Data

This GAGurine [glycosaminoglycans (GAG) concentration in urine] data discussed by Ripley (2002) and explain the concentration of GAG(in units of milligrams per millimole creatinine) in the urine of children aged up to 17 years. Analysis of such data may be helpful for pediatricians to diagnose whether the GAG concentration is normal for a child or not. Here we are considering the results for 40 children of age between 12 to 17 years given as: 5.8, 5.4, 5.7, 3.1, 6.4, 7.0, 5.7, 3.9, 9.4, 4.4, 5.0, 15.9, 3.7, 9.1, 4.7, 3.6, 3.7, 4.1, 7.9, 3.3, 6.6, 1.9, 3.0, 5.7, 3.2, 3.8, 5.3, 3.2, 4.2, 6.0, 9.7, 3.4, 3.2, 2.5, 2.0, 4.0, 4.3, 2.8, 2.2, 4.7.

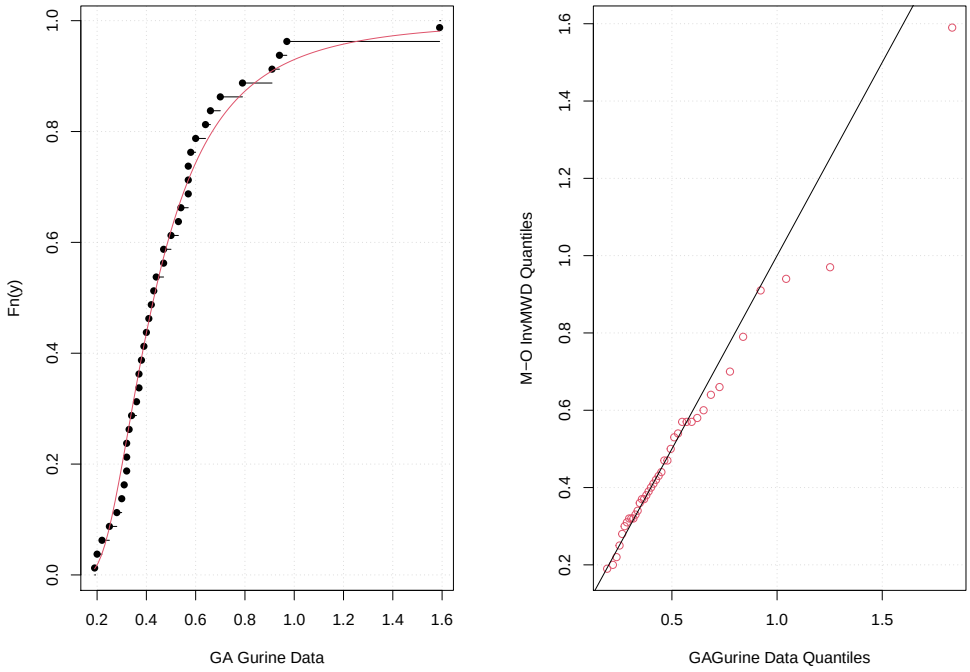
Table 6: *Descriptive Statistics for GAG Data.*

Median	Mean	Variance	Skewness	Kurtosis	Min	Max
4.25	4.99	6.90	2.07	8.74	1.90	15.90

In order, we used Anderson Darling (AD) test to show the fitting of this data to M-O InvMWD. The observed Anderson darling test statistic comes out to be 0.2458 along with high p value. For the estimated parameter, the compatibility of M-O InvMWD to GAGurine data is shown graphically by empirical cdf (ECDF) plot and Q-Q plots in Figure 3 which shows well data fitting. Here for the goodness of fit we have applied three criteria: Negative Log-Likelihood, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The distribution for which AIC and BIC values are smallest is known as the best model. The AIC and BIC are the criterion based on the likelihood and explain the information lost while fitting the model to the given data. We can see from the Table 7 that our model performs better than all other distributions.

Table 7: *Model comparison based on Negative log-likelihood (LL), AIC and BIC for GA Gurine data.*

Sr. No.	Model	Negative LL	AIC	BIC
1.	**M-O InvMWD**	**84.15**	**172.30**	**172.63**
2.	Log-Normal	84.38	172.77	176.15
3 .	Inverse Weibull	84.74	173.49	176.87
4.	Inverse Rayleigh	86.35	174.72	176.41
5.	Maxwell	90.21	182.44	184.13
6.	Inverse Lindley	103.25	208.51	210.20
7.	Inverse Exponential	103.76	209.52	211.21



(a) *ECDF plot for GA Gurine data.* **(b)** *QQ plot for GA Gurine data.*

Figure 3: *ECDF and QQ plots for GA Gurine data.*

Data 2: Relief time Data

This data set represents the relief times (in minutes) of 20 patients receiving an analgesic reported by article Gross and Clark (1975). This data-set is studied and fitted with different models by authors Fayomi (2019). The sample values are given 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.

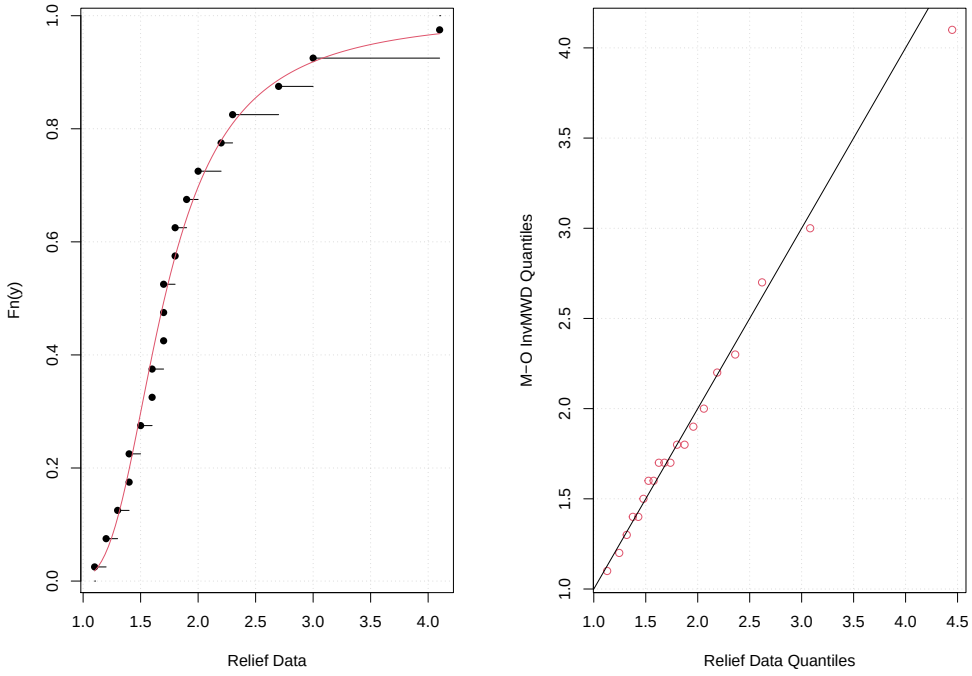
Table 8: *Descriptive Statistics for relief time Data.*

Median	Mean	Variance	Skewness	Kurtosis	Min	Max
1.70	1.90	0.50	1.72	5.92	1.10	4.10

The observed Anderson Darling test statistics comes out to be 0.1438 along with a high p value. For the estimated parameter the compatibility of M-O InvMWD to relief time data is shown graphically by QQ plot and ECDF plots in Figure 4 which shows well data fitting. We can see from the Table 9, that our model performs better than all other considered models.

Table 9: *Model comparison based on Negative log-likelihood (LL), AIC and BIC for relief data.*

Sr. No.	Model	Negative LL	AIC	BIC
1.	**M-O InvMWD**	**15.52**	**35.04**	**37.03**
2.	Log-Normal	16.77	37.54	39.53
3.	Maxwell	20.18	42.36	43.35
5.	Inverse Rayleigh	21.18	44.36	45.36
6.	Inverse Lindley	31.76	65.51	66.51
7.	Inverse Maxwell	32.35	66.69	67.68
8.	Inverse Exponential	32.67	67.34	68.33



(a) *ECDF plot for relief time data.* **(b)** *QQ plot for relief time data.*

Figure 4: *ECDF and QQ plots for patient relief data.*

Table 10: *Point and interval estimates for GA Gurine data under ML and Boot-p methods.*

	$\hat{\theta}$	$\hat{\alpha}$
ML	0.0523	1.2698
MPS	0.0422	1.2098
ACIs	(0.0378,0.0668)	(0.5725,1.9672)
Boot-p	(0.0241,0.0885)	(0.2623,3.4904)

8. Results and Findings

In the simulation study, we see that MSE and AB decrease as the sample size increases. This indicates that the parameters are consistent. Also, the estimated confidence interval obtained by

Table 11: *Point and interval estimates for relief data under ML and Boot-p methods.*

	$\hat{\theta}$	$\hat{\alpha}$
ML	0.1135	0.1232
MPS	0.0955	0.1032
ACIs	(0.0901,0.1369)	(0.0366,0.2098)
Boot-p	(0.0644,0.1739)	(0.0132,0.4447)

asymptotic confidence and boot-p methods contains the true parameters. Two real data sets were analyzed in support of the proposed model and provide a good result. We observed satisfactory results for both simulation and real data.

9. Concluding Remarks

In this paper, a new life-time distribution named M-O InvMWD is proposed by generalizing InvMWD. The nature of hazard of distribution is uni-modal which can be applicable in real life situations in which rate of failure is higher in initial phases and with the passes of time it reduces attaining a maximum point. Basic statistical characteristics for the related distribution are derived and parameters are estimated by using the maximum likelihood estimation and maximum product spacing methods. For calculation of asymptotic confidence intervals, the observed information matrix is derived. Also, the boot-p method is also discussed for obtaining interval estimates. The simulation and real data study are shown for the applicability of the proposed model. We have presented the applications of the proposed distribution under different real life situations. One can extend the work in desired directions based on their choice and availability of real data situations.

Disclosure statement

On behalf of all authors, the corresponding author declare that no potential conflict of interest was reported.

Funding

No funding was provided for the research.

References

[1] Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (2011). *Records*, volume 768. John Wiley & Sons.

[2] Benkhelifa, L. (2017). The Marshall-Olkin extended generalized Lindley distribution: Properties and applications. *Communications in Statistics-Simulation and Computation*, 46(10):8306–8330.

[3] Bennett, S. (1983). Log-logistic regression models for survival data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 32(2):165–171.

[4] Chaudhary, S., Kumar, J., and Tomer, S. K. (2017). Estimation of $P[Y < X]$ for Maxwell distribution. *Journal of Statistics and Management Systems*, 20(3):467–481.

[5] Cheng, R. and Amin, N. (1979). Maximum product-of-spacings estimation with applications to the Lognormal distribution. *Math report*, 79.

[6] Doostparast, M. and Balakrishnan, N. (2010). Optimal sample size for record data and associated cost analysis for Exponential distribution. *Journal of Statistical Computation and Simulation*, 80(12):1389–1401.

[7] Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, 104(2):293.

[8] Fayomi, A. (2019). The Odd Frechet Inverse Weibull distribution with application. *Journal of Nonlinear Sciences and Applications*, 12:165–172.

[9] Folks, J. L. and Chhikara, R. S. (1978). The inverse Gaussian distribution and its statistical application- A review. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40(3):263–275.

[10] Ghitany, M., Al-Mutairi, D., Al-Awadhi, F., and Al-Burais, M. (2012). Marshall-Olkin extended Lindley distribution and its application. *International Journal of Applied Mathematics*, 25(5):709–721.

[11] Gilchrist, W. G. (2000). *Statistical modeling with quantile functions*. Chapman and Hall/CRC.

[12] Gross, A. J. and Clark, V. A. (1975). *Survival distributions: Reliability applications in the biomedical sciences*. University microfilm international.

[13] Krishna, E., Jose, K., Alice, T., and Ristić, M. M. (2013). The marshall-olkin Fréchet distribution. *Communications in Statistics-Theory and Methods*, 42(22):4091–4107.

[14] Krishna, H., Vivekanand, and Kumar, K. (2015). Estimation in Maxwell distribution with randomly censored data. *Journal of Statistical Computation and Simulation*, 85(17):3560–3578.

[15] Kumar, K. and Garg, R. (2014). Estimation of the parameters of randomly censored generalized Inverted Rayleigh distribution. *International Journal of Agricultural and Statistical Sciences*, 10:147–155.

[16] Kumar, K. and Kumar, I. (2019). Estimation in Inverse Weibull distribution based on randomly censored data. *Statistica*, 79(1):47–74.

[17] Kundu, D. and Gupta, R. D. (2006). Estimation of $P[Y < X]$ for Weibull distributions. *IEEE Trans. Reliability*, 55(2):270–280.

[18] Kuş, C. (2007). A new lifetime distribution. *Computational Statistics & Data Analysis*, 51(9):4497–4509.

[19] Langlands, A. O., Pocock, S. J., Kerr, G. R., and Gore, S. M. (1979). Long-term survival of patients with breast cancer: A study of the curability of the disease. *Br med J*, 2(6200):1247–1251.

[20] Mansour, M. M., Nofal, Z. M., Noori, A. N., and El Gebaly, Y. M. (2017). Kumaraswamy Marshall-Olkin Lindley distribution: Properties and applications. *International Journal of Business and Statistical Analysis*, 4(01):21–28.

[21] Marshall, A. W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the Exponential and Weibull families. *Biometrika*, 84(3):641–652.

[22] Maxwell, O., Chukwu, A. U., Oyamakin, O. S., and Khaleel, M. A. (2019). The Marshall-Olkin Inverse Lomax distribution (MO-ILD) with application on cancer stem cell. *Journal of Advances in Mathematics and Computer Science*, pages 1–12.

[23] Nandi, S. and Dewan, I. (2010). An EM algorithm for estimating the parameters of bivariate Weibull distribution under random censoring. *Computational Statistics & Data Analysis*, 54(6):1559–1569.

[24] Pakungwati, R., Widyaningsih, Y., and Lestari, D. (2018). Marshall-Olkin extended inverse Weibull distribution and its application. In *Journal of Physics: Conference Series*, volume 1108, page 012114. IOP Publishing.

[25] Proschan, F. (1963). Theoretical explanation of observed decreasing failure rate. *Technometrics*, 5(3):375–383.

[26] Raffiq, G., Dar, I. S., Haq, M. A. U., Ramos, E., et al. (2020). The Marshall-Olkin Inverted Nadarajah-Haghighi distribution: Estimation and applications. *Annals of Data Science*, pages 1–16.

[27] Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.

[28] Sharma, V. K., Singh, S. K., Singh, U., and Agiwal, V. (2015). The Inverse Lindley distribution: A stress-strength reliability model with application to head and neck cancer data. *Journal of Industrial and Production Engineering*, 32(3):162–173.

[29] Singh, K. L. and Srivastava, R. (2014). Inverse Maxwell distribution as a survival model, genesis and parameter estimation. *Research Journal of Mathematical and Statistical Sciences*, 2320:6047.

[30] Sultan, K., Al-Dayian, G., and Mohammad, H. (2008). Estimation and prediction from gamma distribution based on record values. *Computational Statistics & Data Analysis*, 52(3):1430–1440.

[31] Tibshirani, R. J. and Efron, B. (1993). An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, 57:1–436.

[32] Tomer, S. K. and Panwar, M. (2020). A review on Inverse Maxwell distribution with its statistical properties and applications. *Journal of Statistical Theory and Practice*, 14:1–25.

[33] UL Haq, M. A., Usman, R. M., Hashmi, S., and Al-Omeri, A. I. (2019). The Marshall-Olkin length-biased exponential distribution and its applications. *Journal of King Saud University-Science*, 31(2):246–251.

[34] Usman, R. M. and UL Haq, M. A. (2020). The Marshall-Olkin extended inverted Kumaraswamy distribution: Theory and applications. *Journal of King Saud University-Science*, 32(1):356–365.

RT&A, No 3 (63)
Volume 16, September 2021

On Discrete Scheduled Replacement Model of a Series-Parallel System

Tijjani A. Waziri
•
School of Continuing Education, Bayero University Kano, Nigeria
tijjaniw@gmail.com
Ibrahim Yusuf
•
Department of Mathematical Sciences, Bayero University, Kano, Nigeria
iyusuf.mth@buk.edu.ng

Abstract

This paper investigated the properties of discrete scheduled replacement model of a series-parallel system, with six units. The six units of the system formed three subsystems, which are subsystems A, B and C. Subsystem A is having three parallel units, subsystem B is having a single unit and subsystem C is having two parallel units. It is assumed that, the repairable system is subjected to two categories of failures (Category I and Category II). The mathematical expressions for both reliability function and failure rates, and an elementary renewal theorem were used based on some assumptions in constructing the discrete scheduled replacement model for a series- parallel system. A simple illustrative numerical example where made available, so as to study the properties of the replacement model constructed.

Keywords: category, discrete, replacement, scheduled, time

I. Introduction

Almost all systems deteriorate owing to age and usage, and experience stochastic failures during actual operation. Deterioration raises operating costs and produces less competitive goods. Moreover, consecutive failures are dangerous to the whole system, so timely preventive maintenance is beneficial for supporting normal and continuous system operation. But sometimes an operating systems cannot be replaced at the exact optimum times due to some reasons, such as : shortage of spare units, lack of money or workers, or inconvenience of time required to complete the replacement, but can be rather replaced in idle times, e.g., weekend, month-end, or year-end.

There is an extensive literature on preventive replacement models with their modified versions. Briš *et al.* [1] introduced a new approach for optimizing a complex system's maintenance strategy that respects a given reliability constraint. Chang [2] considered a device that faces two types of failures (repairable and non-repairable) based on a random mechanism. Coria *et al*. [3] proposed a method of analytical optimization for preventive maintenance policy with historical failure time data. Enogwe *et al.* [4] used the distribution of the probability of failure times and come up with a replacement model for items that fails un-notice. Fallahnezhad and Najafian [5] investigated the

number of spare parts and installations for a unit and parallel systems, so as cut down the average cost per unit time. Jain and Gupta [6] studied optimal replacement policy for a repairable system with multiple vacation and imperfect coverage. Lim *et al.* [7] studied the characteristics of some age substitution policies. Liu *et al.* [8] developed mathematical models of uncertain reliability of some multi-component systems. Malki *et al.* [9] analyzed age replacement policies of a parallel system with stochastic dependency. Murthy and Hwang [10] discussed that, the failures can be reduced (in a probabilistic sense) through effective maintenance actions, and such maintenance actions can occur either at discrete time instants or continuously over time. Nakagawa [11] modified the continuous standard age replacement for a unit, and come up with a discrete replacement model for the unit. Nakagawa *et al.* [12] explored the advantages of some replacement policies. Safaei *et al.* [13] investigated the optimal preventive maintenance action for a system based on some conditions. Sudheesh *et al.* [14] studied age replacement policy in discrete approach. Tsoukalas and Agrafiotis [15] presented a new replacement policy warrant for a system with correlated failure and usage time. Waziri *et al.* [16] presented some discounted age replacement models with discounting factor for a serial system exposed to two forms of failures. Waziri and Yusuf [17] presented an age replacement model for a parallel-series system based on some proposed policies. Xie *et al.* [18] assessed the effects of safety barriers on the prevention of cascading failures. Yaun and Xu [19] studies a cold standby repairable system with two different components and one repairman who can take multiple vacations, where they assumed that, if there is a component which fails and the repairman is on vacation, the failed component will wait for repair until the repairman is available. Yusuf and Ali [20] considered two parallel units in which both units operate simultaneously, and the system is subjected to two types of failures. Yusuf *et al.* [21] modified the standard age replacement model by introducing random working time Y in the model, for which the system is replaced at a planned time T, at a random working time Y, or at the first non-repairable type 2 failure whichever occurs first. Zaharaddeen and Bashir [22] developed a planned time replacement model for a unit exposed to two different forms of failures. Zhao *et al.* [23] gathered some recently proposed policies on planned time replacement decision.

This paper is organized in five sections. The present section described the introductory part. Section 2 described the system and its notations. Furthermore, section 2 contained some assumptions based on which, the author will develop the proposed replacement model. Section 3 discussed the proposed replacement model. Section 4 presents some numerical results. Section 5 presents the discussion of the results obtained from both examples 1 and 2. Finally, section 6 presents the summary, conclusion and recommendations.

II. Methods

Reliability measures namely reliability function and failure rates are used to obtain the expressions of discrete scheduled replacement model involving minimal repair based on some assumptions. A numerical example was given for the purpose of investigating the characteristics of the model constructed.

III. Notations

- $r_i^*(t)$: Category I failure rate of unit A_i of subsystem A, for $i = 1, 2, 3$.
- $r_b(t)$: Category II failure rate of subsystem B.
- $r_i(t)$: Category II failure rate of unit C_i of subsystem C, for $i = 1, 2$.
- $R_i^*(t)$: reliability function of unit A_i of subsystem A, for $i = 1, 2, 3$.
- C_b : cost of minimal repair of subsystem B due to Type II failure.
- C_{1m} : cost of minimal repair of unit C_1 of subsystem C due to Category II failure.

- C_{2m} : cost of minimal repair of unit C_2 of subsystem C due to Category II failure.
- C_p : cost of scheduled replacement of the system at NT, for $N = 1, 2, 3$ ....
- C_r : cost of un-scheduled replacement of the system due to Category I failure.
- N^*: the system's optimum discrete scheduled replacement time.

IV. Description of the System

Consider a system comprising of three subsystems A, B and C in series. Subsystem A consist of three active parallel units, which are A_1, A_2 and A_3. Subsystem B consist of a single active unit. While, subsystem C consist of two active units, which are $C_1$and C_2. The three units A_1, A_2 and A_3 are all subjected to Category I failure, which is an un-repairable failure. Subsystem B is only subjected to Category II failure, which is repairable failure. Also, the two units $C_1$and C_2, are only subjected to Category II failure. The system fails due to Category I failure, if all the three units of subsystem A fails due to Category I failure, at such failure, the system is replaced completely. While the system fails due to Category II failure, if subsystem B or all the two units of subsystem C fails due to Category II failure, at such failure the system is minimally repaired. Figure 1 below is the diagram of the system.

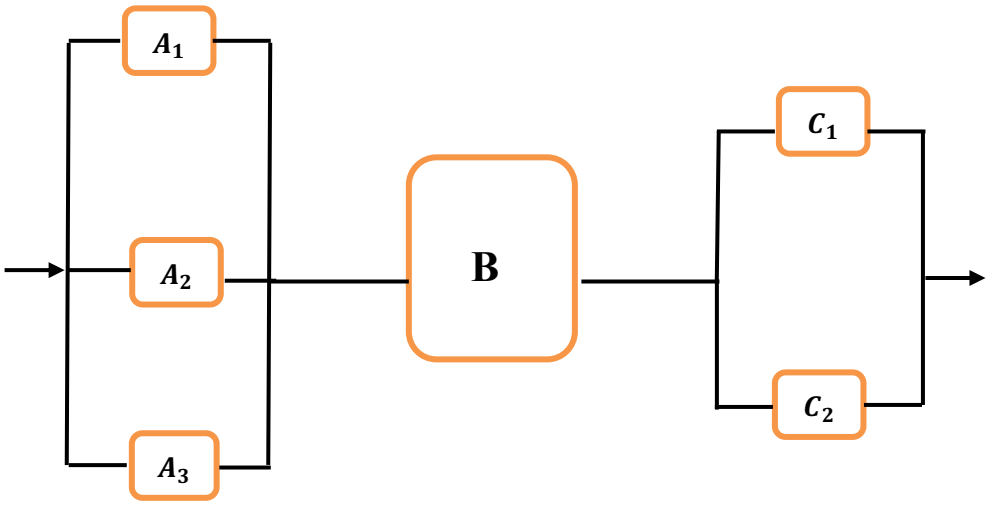


Figure 1: *Reliability block diagram of the system*

V. Discrete Scheduled Replacement Model

This section considers a fundamental discrete scheduled replacement model involving minimal repair.

Assumptions for this model:

1. If subsystem B failed due to Category II failure, then the failed subsystem undergoes minor repair, and allow the system operating from where it stopped.

2. If all the two units of subsystem C failed due to Category II failure, then the failed units will undergoes minor repair, and allow the system operating from where it stopped.
3. The system is replaced completely at scheduled time $NT(N = 1, 2, 3 \ldots)$ for a fixed T or where all the three units of subsystem A fails due to Category I failure, whichever arrives first.
4. The cost of scheduled replacement of the system is less than the cost of un-scheduled replacement.
5. Both Category I failure rate and Category II failures rate arrives according to a non-homogeneous Poisson process
6. The cost of minor repair, planned replacement and un-planned replacement are all positive numbers.

Based on the assumptions, the reliability function of component A_i with respect to Category I failure is

$$R_i^*(T) = e^{-\int_0^T r_i^*(t)dt}, \quad \text{for } i = 1,\ 2,\ 3. \tag{1}$$

Based on the assumptions, the reliability function of the system with respect to Category I failure is

$$R_A^*(NT) = 1 - \big(1 - R_1^*(NT)\big)\big(1 - R_2^*(NT)\big)\big(1 - R_3^*(NT)\big), \tag{2}$$

where $N = 1, 2, 3 \ldots$ and T is fixed.

The cost of unscheduled replacement of the system in one replacement cycle is

$$C_r\big(1 - R_A^*(NT)\big), \tag{3}$$

where $N = 1, 2, 3 \ldots$ and T is fixed.

The cost of scheduled replacement of the system in one replacement cycle is

$$C_p R_A^*(NT), \tag{4}$$

where $N = 1, 2, 3 \ldots$ and T is fixed.

The cost of minimal repair of subsystem B in one replacement cycle is

$$\int_0^{NT} C_b r_b(t) R_A^*(t)dt, \tag{5}$$

where $N = 1, 2, 3 \ldots$ and T is fixed.

The cost of minimal repair of unit C_1 of subsystem C in one replacement cycle is

$$\int_0^{NT} C_{1m} r_1(t) R_A^*(t)dt, \tag{6}$$

where $N = 1, 2, 3 \ldots$ and T is fixed.

The cost of minimal repair of unit C_2 of subsystem C in one replacement cycle is

$$\int_0^{NT} C_{2m} r_2(t) R_A^*(t) dt, \tag{7}$$

where $N = 1, 2, 3$ ... and T is fixed.

Using equations (3) to (7), the replacement cost rate of the system in one replacement cycle is

$$C(N) \frac{C_r\left(1-R_A^*(NT)\right)+C_p R_A^*(NT)+\int_0^{NT} C_b r_b(t) R_A^*(t) dt+\int_0^{NT} C_{1m} r_1(t) R_A^*(t) dt+\int_0^{NT} C_{2m} r_2(t) R_A^*(t) dt}{\int_0^{NT} R_A^*(t) dt}. \tag{8}$$

Noting the following:

1. If the value of T is taking as one (that is , T = 1), then $C(N)$ will be a continuous standard age replacement model with minimal repair.
2. $C(N)$ is adopted as an objective function of an optimization problem, and the main goal is to obtain an optimal discrete scheduled replacement time N^* that minimizes $C(N)$.

VI. Numerical Example

In this section, we will give two numerical example, so as to illustrate the characteristics of the constructed discrete scheduled replacement model.

Let the rate of Category I failure of the three units of subsystem A follows Weibull distribution:

$$r_i^*(t) = \lambda_i^* \propto_i^* t^{\propto_i^* - 1},\ t \geq 0,\ i = 1, 2, 3, \tag{9}$$

where $\propto_i^* > 1$.

Again, let the rate of Category II failure of subsystem B follows Weibull distribution:

$$r_b(t) = \lambda_b \propto_b t^{\propto_b - 1},\ t \geq 0\,, \tag{10}$$

where $\propto_b > 1$.

Also, let the rate of Category II failure of the two units of subsystem C follows Weibull distribution:

$$r_i(t) = \lambda_i \propto_i t^{\propto_i - 1},\ t \geq 0\,,\ i = 1, 2, \tag{11}$$

where $\propto_i > 1$.

Let the set of parameters and cost of repair/replacement be used throughout this particular example:

1. $\propto_b = 4, \propto_2 = 3, \propto_3 = 3.$
2. $\lambda_b = 0.03, \lambda_2 = 0.002, \lambda_3 = 0.03.$
3. $\propto_1^* = 2, \propto_2^* = 3, \propto_3^* = 4.$
4. $\lambda_1^* = 0.0014,\ \lambda_2^* = 0.003, \lambda_3^* = 0.004.$
5. $C_r = 50, C_p = 40.$
6. $C_b = 3, C_{1m} = 2\,, C_{2m} = 2.$

Consequently, by substituting the parameters in equations (9), we obtained the rate of Category I failure of three units of subsystem A as follows below:

$$r_1^*(t) = 0.0028t, \tag{12}$$

$$r_2^*(t) = 0.009t^2, \tag{13}$$

and

$$r_3^*(t) = 0.016t^3. \tag{14}$$

Also by substituting the parameters in equation (10), we obtained the rate of Category II failure of subsystem B as follows:

$$r_b(t) = 0.12t^3. \tag{15}$$

Also by substituting the parameters in equation (11), we obtained the rates of Category II failure of the two units of subsystem C as follows:

$$r_2(t) = 0.006t^2, \tag{16}$$

and

$$r_3(t) = 0.09t^2. \tag{17}$$

Table 1 below is obtained, by substituting the assumed cost of replacement/repair and failure rates of Category I, Category II and Category III failures obtained above (equations (12), (13), (14), (15), (16) and (17)) in equation (8), so as to evaluate the system's optimal discrete scheduled replacement time. When obtaining table 1, the value of $T = 1$, $T = 2$, $T = 3$, and $T = 4$ are considered so as to investigate the properties of the system's optimal discrete scheduled replacement time. Figure 2 is the graph of $C(N)$ against N, as $T = 1$. Figure 3 is the graph of $C(N)$ against N, as $T = 2$. Figure 4 is the graph of $C(N)$ against N, as $T = 3$. Figure 5 is the graph of $C(N)$ against N, as $T = 4$. Figure 6 is the graph comparing the values of $C(N)$ against N, as $T = 1, T = 2, T = 3$ and $T = 4$.

Table 1: *Values of C(N) for T = 1, T = 2, T = 3 and T = 4, versus N* $(1, 2, 3 \ldots)$.

N	C(N) as T=1	C(N) as T=2	C(N) as T=3	C(N) as T=4
1	175.78	91.22	68.69	66.79
2	91.22	66.79	112.03	223.26
3	68.69	112.03	292.11	562.90
4	66.79	223.26	562.90	1089.46
5	80.61	369.52	944.26	1709.67
6	112.03	562.90	1395.91	2289.30
7	161.45	806.97	1863.78	2703.20
8	223.26	1089.46	2289.30	2874.25
9	292.11	1395.91	2620.56	2790.45
10	369.52	1709.67	2821.38	2497.61
11	459.33	2013.13	2876.24	2075.48
12	562.90	2289.30	2790.45	1609.72
13	679.30	2523.24	2586.36	1170.14
14	806.97	2703.20	2297.18	800.04
15	944.26	2821.38	1960.01	516.19

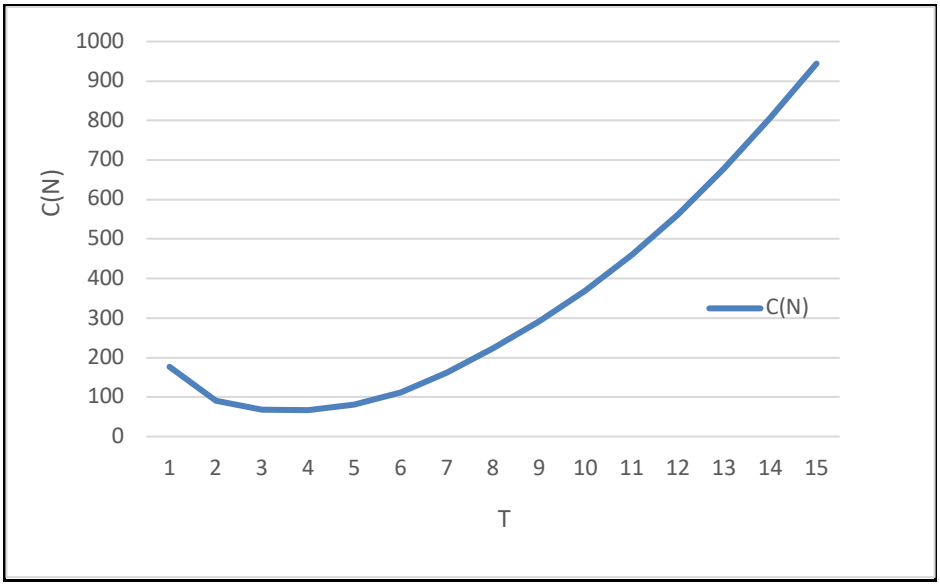


Figure 2: *C(N) against N, as T=1*

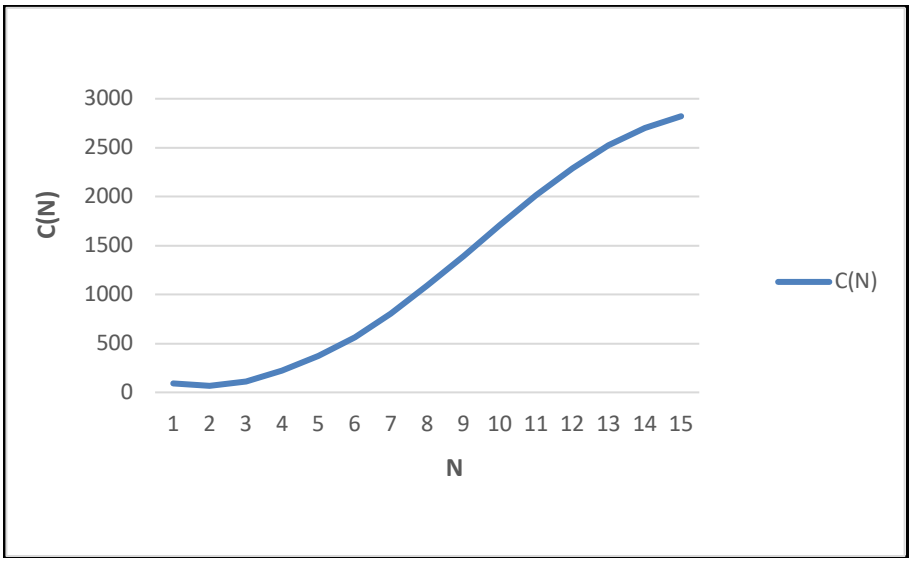


Figure 3: *C(N) against N, as T=2*

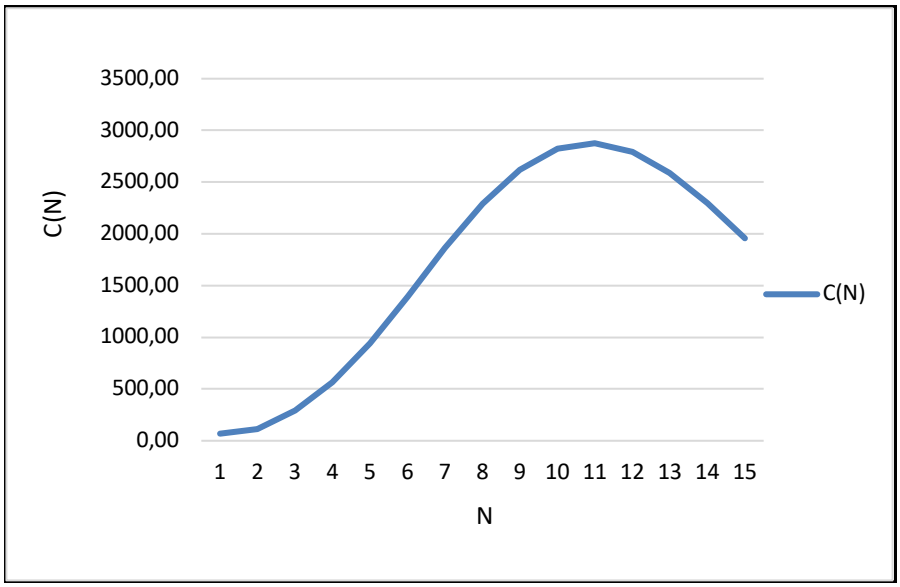


Figure 4: *C(N) against N, as T=3*

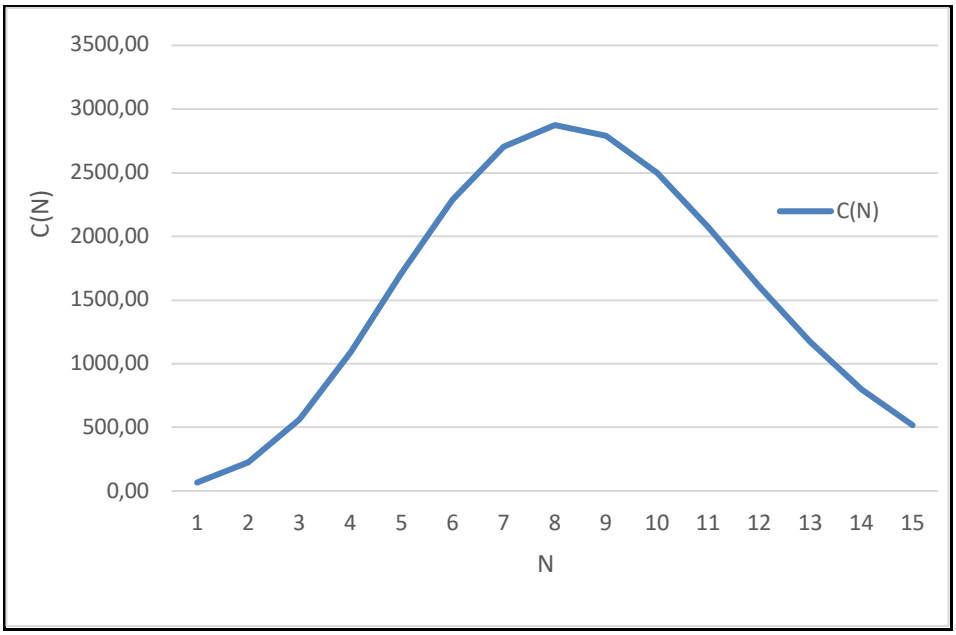


Figure 5: *C(N) against N, as T=4*

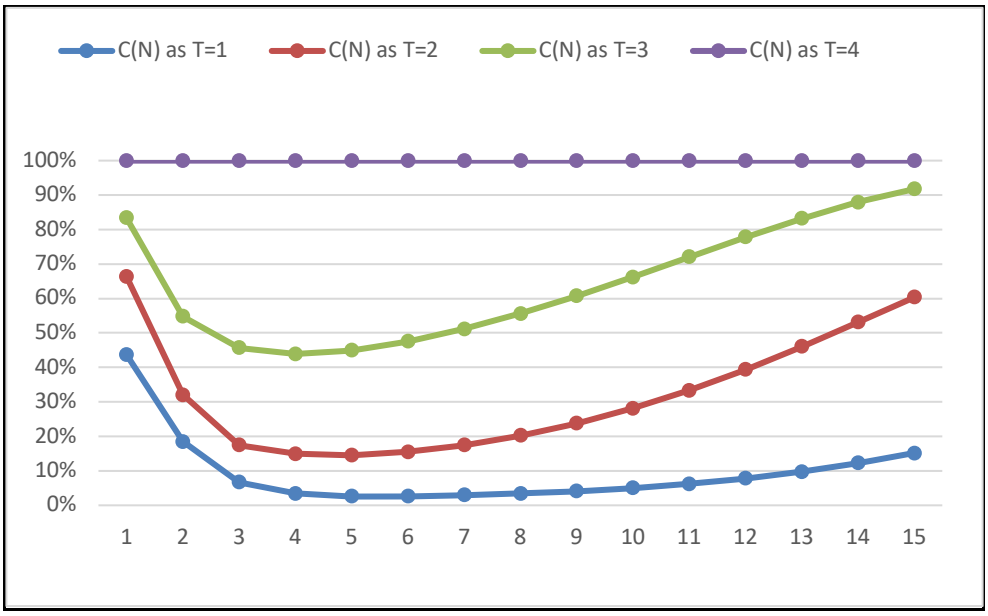


Figure 6: *Comparing C(N) as T=1, T=2, T=3 and T=4*

Some observations from the results obtained are as follows

1. Observe table 1, the optimum discrete scheduled replacement time is 4, when $T = 1$, that is, $N^* = 4$, with $C(N^* = 4) = 66.79$, when $T = 1$. See figure 2 below for the plot of $C(N)$ against N, as $T = 1$.
2. Observe table 1, the optimum discrete scheduled replacement time is 2, when $T = 2$, that is, $N^* = 2$, with $C(N^* = 3) = 66.79$, when $T = 2$. See figure 3 below for the plot of $C(N)$ against T, as $T = 2$.
3. Observe table 1, the optimum discrete scheduled replacement time is 1, when $T = 3$, that is, $N^* = 1$, with $C(N^* = 1) = 68.69$, when $T = 3$. See figure 4 below for the plot of $C(N)$ against T, as $T = 3$..
4. Observe table 1, the optimum discrete scheduled replacement time is 1, when $T = 4$, that is, $N^* = 1$, with $C(N^* = 1) = 66.79$, when $T = 4$. See figure 5 below for the plot of $C(N)$ against T, as $T = 4$.
5. Observe figure 6, $(C(N),\ T = 1) < (C(N), T = 2) < (C(N), T = 3) < (C(N), T = 4)$.
6. Observe figure 2 and figure 3, the graphs are in convex shape.
7. Observe figure 4, the graph is in s-shape.
8. Observe figure 5, the graph is in concave shape.

VII. Conclusion and Recommendations

This paper presented some properties of discrete scheduled replacement model involving minimal repairs. In trying to do that, it is assumed that a system is subjected to two categories of failures (Category I and Category II), such that, Category I is an un-repairable failure, while Category II is a repairable one. A numerical example was provided for simple illustrations. From the results obtained, it is discovered or verified that, the value of T have an effect on the discrete scheduled replacement model, because of the following reasons:

1. as the value of T decreases, the optimal discrete replacement time (N^*) increases, while as the value of T increases, the optimal discrete replacement time (N^*) decreases.
2. as the value of T increases, $C(N)$ increases, while as the value of T decreases, $C(N)$ increases decreases.

With such reasons above, it can be easily seen that, continuous scheduled replacement model (continuous age replacement model) is better than discrete scheduled replacement model (discrete age replacement model). This paper is important to engineers, maintenance managers and plant management in maintaining multi-component systems at idle times, such as weekend, month-end or year-end.

References

[1] Briš, R., Byczanski, P., Goňo, R. and Rusek, S. (2017). Discrete maintenance optimization of complex multi-component systems. *Reliability Engineering and System Safety,* 168:80–89.

[2] Chang, C. (2014). Optimum preventive maintenance policies for systems subject to random working times, replacement and minimal repair. *Computers and Industrial Engineering,* 67:185–194.

[3] Coria, V. H., Maximov, S., Rivas-Davalos, F., Melchor, C. L. and Guardado, J. L. (2015). Analytical method for optimization of maintenance policy based on available system failure data. *Reliability Engineering and System Safety,* 135:55–63.

[4] Enogwe, S. U., Oruh, B. I. and Ekpenyong, E. J. (2018). A modified replacement model for items that fail suddenly with variable replacement costs. *American Journal of Operations Research,* 8:457–473.

[5] Fallahnezhad, M. S., Najafian, E. (2017). A model of preventive maintenance for parallel, series and single item replacement systems based on statistical analysis. *Communications in Statistics-Simulation and Computation,* 46:5846–5859.

[6] Jain, M. and Gupta, R. (2013). Optimal replacement policy for a repairable system with multiple vacations and imperfect fault coverage. *Computers and Industrial Engineering,* 66: 710–719.

[7] Lim, J. H., Qu, J. and Zuo, J. M. (2016). Age replacement policy based on imperfect repair with random probability. *Reliability Engineering and System Safety*, 149: 24–33.

[8] Liu, Y., Ma, Y., Qu, Z. and Li, X. (2018). Reliability mathematical models of repairable Systems with uncertain lifetimes and repair time. *IEEE,* 6: 71285–71295.

[9] Malki, Z., Ait, D. A. and Ouali, M. S. (2015). Age replacement policies for two-component systems with stochastic dependence. *Journal of Quality in Maintenance Engineering,* 20(3):346-357.

[10] Murthy, D. N. P. and Hwang, M. C. (2007). Optimal discrete and continuous maintenance policy for a complex unreliable machine *International Journal of Systems Science,* 6(1):35–52.

[11] Nakagawa T. Maintenance Theory of Reliability: *Springer-Verlag, London Limited*, 2005.

[12] Nakagawa, T., Chen, M. and Zhao, X. (2018). Note on history of age replacement policies. *International Journal of Mathematical, Engineering and Management Sciences,* 3(2): 151–161.

[13] Safaei, F., Ahmadi, J. and Balakrishnan, N. (2018). A repair and replacement policy for systems based on probability and mean of profits. *Reliability Engineering and System Safety,* 183: 143–152.

[14] Sudheesh, K. K, Asha, G. and Krishna, K. M. J. (2019). On the mean time to failure of an age-replacement model in discrete time. *Communications in Statistics - Theory and Methods,* 50 (11): 2569–2585.

[15] Tsoukalas, M. Z. and Agrafiotis, G. K. (2013). A new replacement warranty policy indexed by the product's correlated failure and usage time.*Computers and Industrial Engineering*, 66:203–211.

[16] Waziri, T. A., Yakasai, B. M. and Yusuf, I. (2019). On some discounted replacement models of a series system. *Life Cycle Reliability and Safety Engineering*, doi.org/10.1007/s41872- 019-00101-3.

[17] Waziri, T. A. and Yusuf, I. (2020). On age replacement policy of system involving minimal repair. *Reliability Theory and Application*, 4(59): 54–62.

[18] Xie, L., Lundteigen, M. A. and Liu, Y. (2020). Reliability and barrier assessment of series-parallel systems subject to cascading failures. *Journal of Risk and Reliability*, 00 (0): 1–15.

[19] Yuan, L. and Xu, J. (2011). An optimal replacement policy for a repairable system based on its repairman having vacations. *Reliability Engineering and System Safety*, 96:868–875.

[20] Yusuf, I. and Ali, U. A. (2012). Structural dependence replacement model for parallel system of two units. *Nigerian Journal of Basic and Applied Science*, 20(4): 324–326.

[21] Yusuf, B., Yusuf, I. and Yakasai, B. M. (2015). Minimal repair - replacement model for a system with two types of failure. *Applied Mathematical Sciences*, 138(9): 6867–6875.

[22] Zaharaddeen, H. A. and Bashir, M. Y. (2014). Minimal repair and replacement policy for a complex system with different failure modes of the groups. *International Journal Applied of Mathematics*, 3(4):416–421.

[23] Zhao, X., Al-Khalifa N. K., Hamouda, A. M. and Nakagawa, T. (2017). Age replacement models: a summary of new perspectives and methods. *Reliability Engineering and System Safety*, 161:95–105.

A New Generalization of Exponential Distribution for Modelling reliability Data

V. Jilesh, Aifoona Ahammed P.M

•

Department of Statistics, Government Arts and Science College, Kozhikode. Kerala, India.
jileshstat@gmail.com
aifoonaahammed@gmail.com

Abstract

In this paper, a new generalization of the exponential distribution is proposed. Different properties, important reliability measures and special cases of this distribution are investigated. Unknown parameters are estimated using the maximum likelihood method of estimation. A simulation study is carried out to assess the accuracy of the maximum likelihood estimates. Two real data sets are successfully modelled with the proposed distribution.

Keywords: Exponential distribution, Reliability, Maximum likelihood estimation.

1. Introduction

Recently, researchers are more interested in developing new probability distributions and generalizations from the existing family of distributions. The aim of more realistic modelling of complex datasets can be attained through such generalizations. Such newly formed distributions are also showing better flexibilty and properties than the baseline distribution, becomes more suitable in reliability studies and other related fields. For instance, see Eugene et al [3], Bourguignon et al. [4], Cordeiro and Castro [3], Marshall and Olkin[9], Zografos and Balakrishnan [15], Silva et al. [14], Jayakumar and Mathew [7], Nadarajah and Kotz [10], Nadarajah and Kotz [12], Nadarajah and Gupta [11]. One important family of distribution, which is the basis of many other well studied probability distributions is the exponential distribution. One main limitation of this distribution in using the reliability modelling is it's constant hazard rate. So we can find different generalizations of the exponential distribution to overcome this problem. Some important generalizations are gamma, Weibull, Rayleigh and Generalized exponential distribution of Gupta and Kundu [5]. But, in some of the datasets, we can observe, a sudden drop in the frequency of observations after some specific data points. But the available generalizations are not appropriate to model that kind of datasets, which demands a breakage in the flexibility of the probability density function. Here we introduce a new distribution to model such data sets. The family of distrbution under study is derived using the exponential distribution and two sided power distribution of René Van Dorp and Kotz [13]. The probability density function of the two sided power distribution is given by,

$$g(x) = \begin{cases} \alpha \left(\frac{x}{\theta}\right)^{\alpha-1}, & if \quad 0 < x \leq \theta \\ \alpha \left(\frac{1-x}{1-\theta}\right)^{\alpha-1} & ,if \quad \theta \leq x < 1. \end{cases} \quad \alpha > 0, 0 < \theta < 1. \tag{1}$$

This paper is organized as follows. In section 2, we propose the new distribution and discuss it's basic properties, Reliability properties are studied in Section 3. The estimation of unknown parameters in the proposed distribution is done using maximum likelihood method and is discussed in Section 4. A simulation study to check the estimation procedure is conducted in section 5. In section 6, we successsfully modelled the fatigue failure data and aircraft failure time dataset using the proposed distribution.

2. Definition and Properties

A random variable X is said to follow a generalized exponential distribution, denoted by G(α, θ), if it's probability density function (pdf) is of the form

$$f(x) = \begin{cases} \alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1}, & if \quad 0 < x \leq -ln(1-\theta) \\ \\ \alpha e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1}, & if \quad -ln(1-\theta) \leq x < \infty, \end{cases} \tag{2}$$

where $0 < \theta < 1$ and $\alpha > 0$, α is not necessarily be an integer. When α=1, the above probability density function is the probability density function of exponential random variable with mean unity. The above probability denssity function can be derived by mixing the two sided power distribution and the exponential distribution. The shape of the probability density curves given by (2) for different values of parameters are shown in Figure (1). Note that for the values $\alpha > 1$ the density curve first increases and reach a maximum value corresponding to $x = log(\alpha)$ and then decreases, there is a cutting point at $-ln(1-\theta)$. For $0 < \alpha < 1$, the curve is always decreases with a cutting point at $-ln(1-\theta)$. The values taken by the parameter θ doesnot effect the monotone behaviour of the curve. Mode of the distribution is $log(\alpha)$. To derive the cumulative distribution function (CDF) of the proposed distribution, we have to consider two cases.

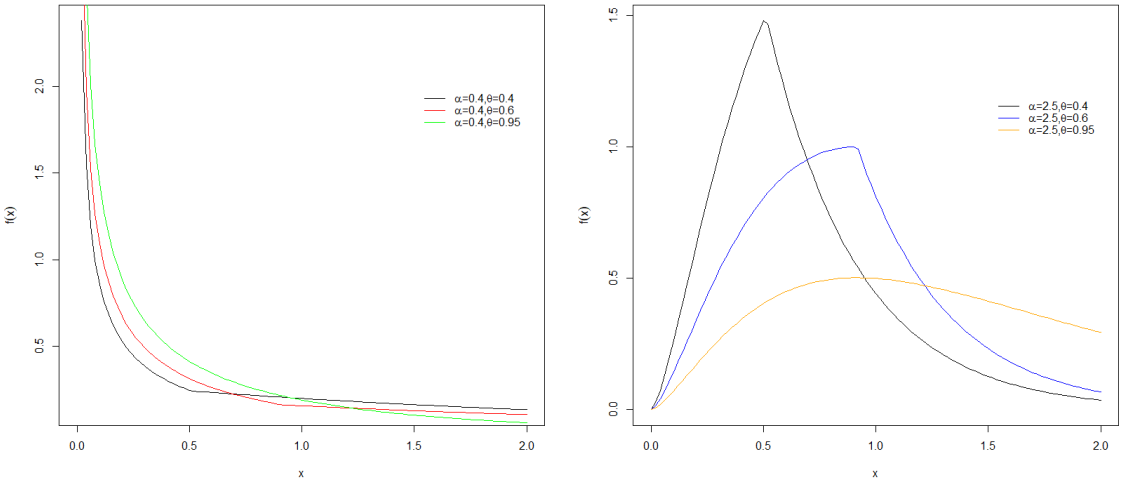


Figure 1: *Density plot of G(α,θ) for different values of θ*

Case 1: $0 < t \leq -ln(1-\theta)$,
Then

$$\begin{aligned} F(t) &= \int_0^t \alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx \\ &= \frac{\alpha}{\theta^{\alpha-1}} \int_0^t e^{-x}(1-e^{-x})^{\alpha-1} dx \end{aligned}$$

by substitution for $1-e^{-x}$ by u, we get

$$\begin{aligned} F(t) &= \frac{\alpha}{\theta^{\alpha-1}} \int_0^{1-e^{-t}} u^{\alpha-1} du \\ &= \theta\left(\frac{1-e^{-t}}{\theta}\right)^{\alpha}. \end{aligned}$$

Case 2: $-ln(1-\theta) < t < \infty$

$$\begin{aligned} F(t) &= \int_0^{-ln(1-\theta)} \alpha e^{-x} \left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx + \int_{-ln(1-\theta)}^{t} \alpha e^{-x} \left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1} dx \\ &= 1-(1-\theta)\frac{e^{-\alpha t}}{(1-\theta)^{\alpha}}. \end{aligned}$$

Thus, the cumulative distribution function is given by

$$F(x) = \begin{cases} \theta\left(\frac{1-e^{x}}{\theta}\right)^{\alpha}, & if \quad 0 < x \leq -ln(1-\theta) \\ 1-(1-\theta)\frac{e^{-\alpha x}}{(1-\theta)^{\alpha}}, & if \quad -ln(1-\theta) \leq x < \infty. \end{cases} \tag{3}$$

Similarly, the ***quantile function*** can be derived by inverting the distribution function. Thus, we obtain

$$\begin{aligned} Q(u) &= F^{-1}(u), \quad 0 < u < 1 \qquad (4) \\ &= \begin{cases} -ln\left[1-\theta\left(\frac{u}{\theta}\right)^{\frac{1}{\alpha}}\right] & 0 < u \leq \theta \\ -\frac{1}{\alpha}\left[ln(1-u)+(\alpha-1)ln(1-\theta)\right] & \theta < u < 1. \end{cases} \end{aligned}$$

The simulation of random sample having G(α,θ) distribution can be done using the above equation, where U=u is a realization from a uniform random variable, that is $U \longrightarrow U(0,1)$.
Moment generating function of a random variable X with G(α,θ) distribution, is given by,

$$\begin{aligned} M_X(t) &= \int_0^{-ln(1-\theta)} e^{tx}\alpha e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx + \int_{-ln(1-\theta)}^{\infty} e^{tx}\alpha e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1} dx \qquad (5) \\ &= \frac{\alpha}{\theta^{\alpha-1}}\int_0^{-ln(1-\theta)} e^{-x(1-t)}(1-e^{-x})^{\alpha-1}dx + \frac{\alpha}{(1-\theta)^{\alpha-1}}\int_{-ln(1-\theta)}^{\infty} e^{-x(\alpha-t)}dx \\ &= \frac{\alpha}{\theta^{\alpha-1}}B(\theta,\alpha,1-t) + \frac{\alpha}{(1-\theta)^{\alpha-1}}\frac{(1-\theta)^{\alpha-t}}{(\alpha-t)}, \end{aligned}$$

where

$$B(\theta,m,n) = \int_0^{\theta} u^{m-1}(1-u)^{n-1}du$$

is the incomplete beta function or the distribution function of beta 1^{st} kind.
The k^{th} order raw moment for G(α,θ) distribution can be written as

$$\begin{aligned} E(X^k) &= \int_0^{\infty} x^k f(x)dx \qquad (6) \\ &= \int_0^{-ln(1-\theta)} \alpha x^k e^{-x}\left(\frac{1-e^{-x}}{\theta}\right)^{\alpha-1} dx + \int_{-ln(1-\theta)}^{\infty} \alpha x^k e^{-x}\left(\frac{e^{-x}}{1-\theta}\right)^{\alpha-1} dx. \end{aligned}$$

But, it is difficult to get a good expression for the above integral. Hence, we have calculated the moments numerically. Table (1) gives the the mean and variance of G(α,θ), for different values taken by the parameters α and θ.

An ***entropy*** is a measure of variation or uncertainty. The Rényi entropy of a random variable with probability density function f(·) is defined as

$$I_R(x) = \frac{1}{1-\gamma} log \int_0^{\infty} f^{\gamma}(x)dx, \gamma > 0, \gamma \neq 1.$$

Table 1: *mean and variance of G(α,θ), for different values α and θ*

	$\theta = 0.3$		$\theta = 0.5$		$\theta = 0.8$	
α	E(X)	V(X)	E(X)	V(X)	E(X)	V(X)
0.5	1.6828	3.8627	1.4467	3.5725	1.0269	2.0617
1	1	1	1	1	1	1
1.5	0.7772	0.4552	0.8669	0.4825	1.0480	0.6123
2	0.6677	0.2606	0.8068	0.2892	1.0976	0.4406
2.5	0.6029	0.1691	0.7740	0.1946	1.1410	0.3429
3	0.5603	0.1185	0.7539	0.1408	1.1780	0.2794
3.5	0.5301	0.0879	0.7406	0.1071	1.2096	0.2346
4	0.5076	0.0678	0.7313	0.0844	1.2368	0.2012
4.5	0.4902	0.0539	0.7245	0.0685	1.2604	0.1755
5	0.4764	0.0439	0.7194	0.0567	1.2812	0.1546

The Shannon entropy of a random variable X is defined by $E\left[-logf(x)\right]$.
First derive the Rényi entropy for the corresponding to G(α,θ) distribution. We have

$$\begin{aligned}\int_0^\infty f^\gamma(x) &= \left(\frac{\alpha}{\theta^{(\alpha-1)}}\right)^\gamma \int_0^{-ln(1-\theta)} e^{-\gamma x}(1-e^{-x})^{\gamma(\alpha-1)}dx \\ &\quad + \left(\frac{\alpha}{(1-\theta)^{(\alpha-1)}}\right)^\gamma \int_{-ln(1-\theta)}^\infty e^{-\alpha\gamma x}dx \\ &= \left(\frac{\alpha}{\theta^{(\alpha-1)}}\right)^\gamma \int_0^{-ln(1-\theta)} e^{-\gamma x}(1-e^{-x})^{\gamma(\alpha-1)}dx + \frac{\alpha^{\gamma-1}(1-\theta)^\gamma}{\gamma}\end{aligned} \tag{7}$$

But, the above expresson is difficult to be express in an explicit form.

Consider the Shannon entropy, we have

$$\begin{aligned}H(X) &= E(-logf(x)) \\ &= \int_0^\infty -logf(x)f(x)dx \\ &= \int_0^{-ln(1-\theta)} -logf(x)f(x)dx + \int_{-ln(1-\theta)}^\infty -logf(x)f(x)dx \\ &= -2log(\alpha) + (\alpha+1)E(X) - (\alpha-1)\left[\theta\left(log(\theta) - \frac{1}{\alpha}\right) - 2log(\theta)\right].\end{aligned} \tag{8}$$

Order statistics refers the ranking a sample from a distribution. Let $X_1, X_2, .., X_k$ be k independent and identically distributed random variables, each with cumulative distribution function F(x) given in equation (3). We denote $X_{(r)}$ as the r^{th} order statistic, r=1,2,...,k. Then $f_r(x)$, the probability density function of $X_{(r)}$ for G(α,θ) distribution is given by

$$\begin{aligned}f_r(x) &= \frac{1}{B(r,k-r+1)}F^{r-1}(x)(1-F(x))^{k-r}f(x) \\ &= \frac{k!}{(r-1)!(k-r)!}\begin{cases}\frac{\alpha e^{-x}(1-e^{-x})^{\alpha r-1}\left[\theta^{(\alpha-1)}-(1-e^{-x})^\alpha\right]^{k-r}}{\theta^{(\alpha-1)k}}, 0 < x < -ln(1-\theta) \\ \frac{\alpha e^{-\alpha x(k-r+1)}\left[(1-\theta)^{\alpha-1}-e^{-\alpha x}\right]^{r-1}}{(1-\theta)^{(\alpha-1)k}}, -ln(1-\theta) < x < \infty.\end{cases}\end{aligned} \tag{9}$$

3. Reliability measures of the G(α, θ) distribution

From equation (3), the ***reliability function*** is

$$\begin{aligned} R(t) &= 1 - F(t) \\ &= 1 - \begin{cases} \theta\left(\frac{1-e^{x}}{\theta}\right)^{\alpha}, & if \quad 0 < x \leq -ln(1-\theta) \\ 1-(1-\theta)\frac{e^{-\alpha x}}{(1-\theta)^{\alpha}}, & if \quad -ln(1-\theta) \leq x < \infty \end{cases} \\ &= \begin{cases} 1-\frac{(1-e^{-x})^{\alpha}}{\theta^{\alpha-1}} & 0 < x \leq -ln(1-\theta) \\ \frac{e^{-\alpha x}}{(1-\theta)^{\alpha-1}} & -ln(1-\theta) \leq x < \infty. \end{cases} \end{aligned} \tag{10}$$

To identify the applicability of the introduced distribution in reliability studies, we have to familiarize with the shape characteristics of the reliability and hazard curves with the changes in parameter values. For the G(α,θ) distribution the ***hazard function*** is given by

$$\begin{aligned} h(x) &= \frac{f(x)}{R(x)} \\ &= \begin{cases} \frac{\alpha e^{-x}(1-e^{-x})^{\alpha-1}}{\theta^{\alpha-1}-(1-e^{-x})^{\alpha}} & 0 < x \leq -ln(1-\theta) \\ \alpha & -ln(1-\theta) \leq x < \infty. \end{cases} \end{aligned} \tag{11}$$

The plot of hazard function is given in Figure (2). For any fixed θ, The G(α,θ) distribution has an increasing hazard function for $\alpha > 1$ and it has decreasing hazard function for $\alpha < 1$. For α=1 the hazard function becomes 1, independent of x. These results are not very difficult to prove, it simply follows from the fact that the G(α,θ) distribution has a log-concave density for $\alpha > 1$ and it is log-convex for $\alpha \leq 1$. The hazard function of the G(α,θ) distribution behaves exactly the same way as the hazard functions of the Weibull distribution distribution.

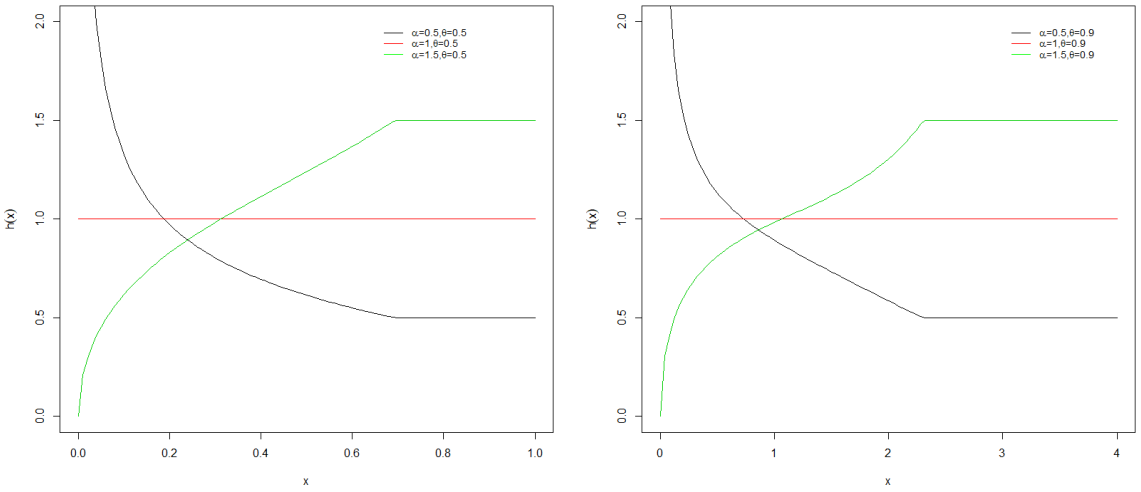


Figure 2: *Plot of hazard function for different values of parameters α and θ*

Reversed hazard function Reversed hazard has been using for the analysis of right- truncated and left-censored data and is applicable in such areas as Forensic Science. The formula of reversed hazard rate of a random life is defined as the ratio between the life probability density to its distribution function. The reversed hazard function for the G(α,θ) distribution is given by,

$$r(x,\alpha,\theta) = \begin{cases} \frac{\alpha}{e^{x}-1}, & 0 < x \leq -ln(1-\theta) \\ \frac{\alpha}{e^{\alpha x}(1-\theta)^{(\alpha-1)}-1}, & -ln(1-\theta) \leq x < \infty. \end{cases} \tag{12}$$

It has observed that, for all values of α and θ, the reversed hazard function is a decreasing function of x. Further, there is no non negative random variable have an increasing reversed hazard rate. Distributions showing the same behaviour in the reverse hazard function are weibull, lognormal.

Elasticity of a distribution express the change that, the distribution function undergoes when faced with the variation in the random variable. It is one of the most important concept in economics theory. In economics, Elasticity measures how sensitive an *output* variable is to change in an *input* variable. This classical concept of elasticity to an economic function can be extended to the cumulative distribution function of a random variable. The elasticity function e(x) is defined as

$$e(x) = \frac{dlnF(x)}{dlnx} = \frac{F'(x)/F(x)}{1/|x|} = \frac{|x|f(x)}{F(x)}, F(x) > 0. \tag{13}$$

In the case of G(α,θ) distribution, Elasticity function is given by

$$\begin{aligned} e(x) &= |x| r(x, \alpha, t) \\ &= |x| \begin{cases} \frac{\alpha}{e^x - 1}, & 0 < x \leq -ln(1-\theta) \\ \frac{\alpha}{e^{\alpha x}(1-\theta)^{(\alpha-1)} - 1}, & -ln(1-\theta) \leq x < \infty, \end{cases} \end{aligned} \tag{14}$$

which shows the close relationship that exists between the reversed hazard function and elasticity.

Mean residual life function (MRLF) is the average lifetime remaining for a component or an individual which had survived at time. For a continuous, non-negative random variable X, representing lifetime of a component or an individual is the residual life random variable at age t, denoted by $X_t = X - t|X > t$, is simply the remaining lifetime beyond that age. Then the MRLF denoted by μ(t) is defined as

$$\begin{aligned} \mu(t) &= E(X - t|X > t) \\ &= \frac{1}{\overline{F}(t)} \int_t^\infty \overline{F}(x) dx \\ &= \frac{1}{R(t)} \int_t^\infty R(x) dx \end{aligned} \tag{15}$$

Then for G(α,θ) distribution we have

$$\begin{aligned} \mu(t) = &\frac{\theta^{(\alpha-1)}}{\theta^{(\alpha-1)} - (1 - e^{-t})^\alpha} \int_t^{-ln(1-\theta)} 1 - \frac{(1 - e^{-x})^\alpha}{\theta^{(\alpha-1)}} dx \\ &+ \frac{(1-\theta)^{(\alpha-1)}}{e^{-\alpha t}} \int_{-ln(1-\theta)}^\infty \frac{e^{-\alpha t}}{(1-\theta)^{(\alpha-1)}} \end{aligned}$$

substituting for $(1 - e^{-x})$ by u in first part of the integral we get

$$\mu(t) = \frac{1}{\theta^{(\alpha-1)} - (1 - e^{-t})^\alpha} \int_{1-e^{-t}}^\theta u^\alpha (1-u)^{-1} du + \frac{1}{e^{-\alpha t}} \left[\frac{(1-\theta)^\alpha}{\alpha} \right]. \tag{16}$$

But, the above expression does not have an explicit form and thus we need to calculate it numerically.

4. Estimation of Parameters; Maximum Likelihood Estimation

The proposed derivation of the maximum likelihood estimation procedure of a G(α,θ) distribution is quite instructive. Let for a sample $\underline{X} = (X_1, X_2, .., X_s)$, the order statistics be $X_{(1)} < X_{(2)} < .. < X_{(s)}$. By definition, the likelihood function for $\underline{X}$ is

$$L(\underline{X}; \theta, \alpha) = \alpha^s \prod_{i=1}^s e^{-X_{(i)}} \left[\frac{\prod_{i=1}^r (1 - e^{-X(i)}) \prod_{i=r+1}^s e^{-X_{(i)}}}{\theta^r (1-\theta)^{s-r}} \right]^{\alpha-1}, \tag{17}$$

where, $X_{(r)} \leq -ln(1-\theta) < X_{(r+1)}$, with $X_{(0)} \equiv 0$
Then, the MLE estimators of the parameters are given by

$$\hat{\theta} = 1 - e^{-X_{(\hat{r})}} \tag{18}$$

$$\hat{\alpha} = -\frac{s}{logM(\hat{r})}, \tag{19}$$

where, $\hat{r}$ = arg $max_{r\in\{1,2,..,s\}} M(r)$ and

$$M(r) = \prod_{i=1}^{r-1} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r)}}} \prod_{i=r+1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(r)}}}. \tag{20}$$

To maximize the likelihood (17), we set

$$\max_{\alpha>0,0<\theta<1} L(\underline{X};\theta,\alpha) = \max_{\alpha>0} \left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right], \tag{21}$$

where, $\hat{M}$ is given by

$$\hat{M} = \max_{0<\theta<1} \left[\frac{\prod_{i=1}^{r}(1-e^{-X(i)}) \prod_{i=r+1}^{s} e^{-X_{(i)}}}{\theta^r(1-\theta)^{s-r}}\right], \tag{22}$$

and as above $X_{(r)} \leq -ln(1-\theta) < X_{(r+1)}$, with $X_{(0)} \equiv 0$.
Using the properties of Pitman family, $X_{(\hat{r})}$ would be the estimate of $\hat{\hat{\theta}}$= -ln(1-θ). Then by inverting we have,

$$\hat{\theta} = 1 - e^{-X_{(\hat{r})}}$$

Now,

$$log\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right] = slog(\alpha) - \sum_{i=1}^{s} X_{(i)} + (\alpha-1)log(\hat{M}) \tag{23}$$

and

$$\frac{\partial}{\partial\alpha} log\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right] = \frac{s}{\alpha} + log(\hat{M}) \tag{24}$$

equating to 0 yields,

$$\hat{\alpha} = -\frac{s}{log(\hat{M})}.$$

From equation (24), it follows that

$$\frac{\partial}{\partial\alpha} log\left[\alpha^s \prod_{i=1}^{s} e^{-X_{(i)}} \hat{M}^{\alpha-1}\right] > 0 \Leftrightarrow \hat{\alpha} < -\frac{s}{log(\hat{M})}. \tag{25}$$

Hence, $\hat{\alpha}$ corresponds to a global maximum of both (23) and (21). Note that for $i < r$, it follows that $0 < \frac{1-e^{-X_{(i)}}}{\theta} < 1$ and for $i > r$, it follows that $0 < \frac{e^{-X_{(i)}}}{1-\theta} < 1$. Hence $0 < \hat{M} < 1$ and thus $\hat{\alpha} > 0$. Using equation (22), we may write $\hat{M} = \max_{r\in\{0,..,s\}} H(r)$, where,

$$H(r) = \max_{X_{(r)}\leq\hat{\hat{\theta}}\leq X_{(r+1)}} \left[\frac{\prod_{i=1}^{r}(1-e^{-X(i)}) \prod_{i=r+1}^{s} e^{-X_{(i)}}}{\theta^r(1-\theta)^{s-r}}\right]. \tag{26}$$

We shall discuss three cases: $r \in \{1,2,...,s-1\}$, r = 0 and r = s.

Case 1: $r \in \{1,2,...,s-1\}$: Here, $X_{(r)} \leq \hat{\hat{\theta}} \leq X_{(r+1)}$. From (26),

$$H(r) = \max_{r'\in\{r,r+1\}} \prod_{i=1}^{r'-1} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r')}}} \prod_{i=r'+1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(r')}}} \tag{27}$$

Case 2: r = 0 : Here, $0 \leq \hat{\hat{\theta}} \leq X_{(1)}$. From (26) it follows that in this case

$$H(0) = \max_{0 \leq \hat{\hat{\theta}} \leq X_{(1)}} \left[\prod_{i=1}^{s} \frac{e^{-X(i)}}{1-\theta} \right] \tag{28}$$

Hence,

$$H(0) = \prod_{i=1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(1)}}} = \prod_{i=2}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(1)}}} \tag{29}$$

Case 3: r = s : Here, $X_{(s)} \leq \hat{\hat{\theta}} < \infty$. From (26) it follows that in this case

$$H(s) = \max_{X_{(s)} \leq \hat{\hat{\theta}} < \infty} \left[\frac{1-e^{X_{(i)}}}{\theta} \right] \tag{30}$$

Hence

$$H(s) = \prod_{i=1}^{s} \frac{1-e^{X_{(i)}}}{1-e^{-X_{(s)}}} = \prod_{i=1}^{s-1} \frac{1-e^{X_{(i)}}}{1-e^{-X_{(s)}}} \tag{31}$$

From (27), (29)and (31) we obtain that $\hat{M} = \max_{r \in \{1,..s\}} M(r)$, where

$$M(r) = \prod_{i=1}^{r-1} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r)}}} \prod_{i=r+1}^{s} \frac{e^{-X_{(i)}}}{e^{-X_{(r)}}} \tag{32}$$

Note that $\hat{M}$ attained at $\hat{\hat{\theta}} = X_{(\hat{r})}$ where $\hat{r} = \arg\max_{r \in \{1,2,...,s-1\}} M(r)$.

The estimates given in (18) and (19) are quite intuitive. In particular the estimator of the parameter θ is in term of a specific order statistic. Note that the approach for determining the MLE estimate $\hat{\theta}$ for the G(α,θ) distribution is similar to the approach for determining the MLE estimate $\hat{\theta}$ for a triangular distribution and STSP (θ, n) distribution (see René Van Dorp and Kotz [13]). In order to find the estimate, we use a quite different method.

Consider the matrix $\mathbf{A} = [a_{i,r}]$ where

$$a_{i,r} = \begin{cases} \frac{1-e^{-X_{(i)}}}{1-e^{-X_{(r)}}}, & \text{if} \quad i < r \\ \\ \frac{e^{-X_{(i)}}}{e^{-X_{(r)}}}, & \text{if} \quad i \geq r. \end{cases} \tag{33}$$

Then **A** will be a real matrix with unit diagonal entries. Then, we find the product of the matrix elements in the r^{th} column which are equal to the values of M(r) given by equation (32). identify the maximum value of M(r) and the corresponding r^{th} order statistic $X_{(r)}$ is taken as the estimate of $\hat{\hat{\theta}}$ and by inverting we get the maximum likelihood estimate of the parameter θ. The maximum likelihood estimate of second parameter α can be evaluated using the equation (19), where s denotes the total number of observations.

5. Simulation

To verify the estimation procedure, We have considered a simulation study. We generated samples of different sizes using the quantile function of the proposed distribution and the parameters are estimated with the above discussed procedure. It is repeated 500 times and the mean of estimates are taken as the estimate of the parameters. Table (2) gives the estimates of the parameters, the values in brackets indicates the mean squared error. Note that, as the sample size increases, the estimate becomes more close to the true value of the parameter.

Table 2: *Estimated values of the parameters and mean squared error.*

	θ			α		
Sample size	0.3	0.5	0.9	1.5	2	3
100	0.3107	0.5036	0.8985	1.5469	2.0444	3.0467
	(0.0135)	(0.0031)	(0.0003)	(0.0229)	(0.0416)	(0.0873)
150	0.3081	0.5012	0.8980	1.5354	2.0326	3.0422
	(0.0085)	(0.0022)	(0.0003)	(0.0162)	(0.0285)	(0.0636)
200	0.3072	0.5007	0.8982	1.5186	2.0161	3.0323
	(0.0056)	(0.0014)	(0.0001)	(0.0104)	(0.0196)	(0.0442)
500	0.3018	0.5004	0.8998	1.5093	2.009	3.0228
	(0.0021)	(0.0005)	(5.213×10^{-05})	(0.0048)	(0.0091)	(0.0190)

6. Data Analysis

In this section, we provide an application of the G(α,θ) distribution by modeling two real data sets.

Data set I: The first data set represents the life of fatigue fracture of Kevlar 373/epoxy that are subject to constant pressure at the 90 percentage stress level until all had failed, so we have complete data with the exact times of failure. The same datasets were used in the literature by Alizadeh et. all [1]. The estimation of the parameters is done using the maximum likelihood method as discussed in the previous section. Firstly, we identified the matrix **A** with entries defined as in (33). Then we find the product of the matrix elements in the rth column which gives the values of M(r) given by equation (32). The maximum value of M(r) is identified and the corresponding r^{th} order statistic is taken as the maximum likelihood estimate of $\hat{\hat{\theta}}$ and by inverting we obtained the estimate of the parameter θ. Similarly, the maximum likelihood estimate $\hat{\alpha}$ can be obtained using equation (19). Here, we obtained the estimate of θ as 0.9998879 and the estimate of α as 2.4003. To verify, goodness of fit of the proposed distribution, we have performed Kolmogorov-Smirnov test and the corresponding p-value is 0.07278, which indicates that the G(α,θ) is a suitable model for the data. The histogram of the data together with the fitted probability density curve is given in Figure (3).

Data set II: The second real data set represent the failure times of 84 aircraft windshield. This data is taken from Ijaz et all. [6]. The estimation was carried out as done in the previous data set. The value of M(r) and corresponding order statistic are noted and we obtain the estimte of θ and α. The parameters for the data set II are estimated as $\hat{\theta}$ = 0.9961512, $\hat{\alpha}$ = 7.827007. The p-value for Kolmogorov-Smirnov test is 0.1958. The histogram of the data set II with fitted probability curve is given in Figure (4).

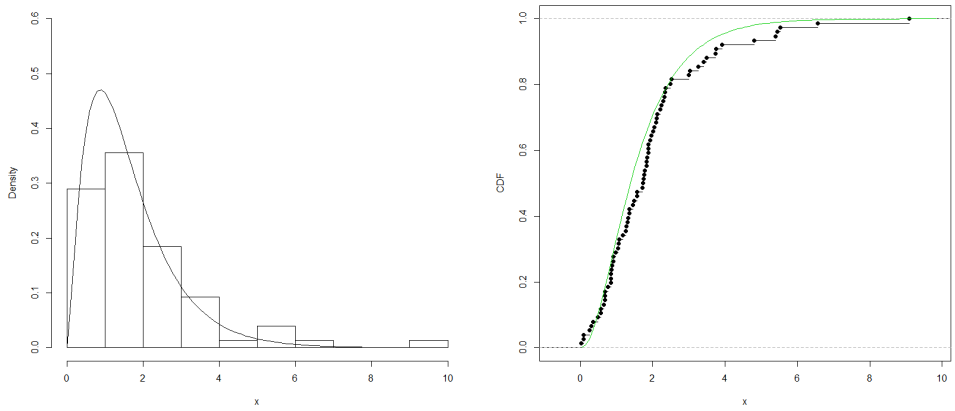


Figure 3: *(a) Histogram and fitted G(α,θ) probability density function for dataset I. (b) Theoretical and fitted distribution function for the dataset I.*

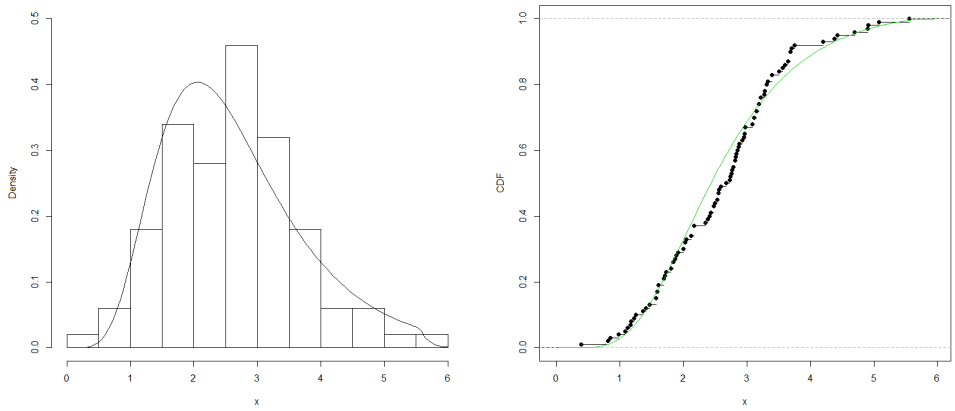


Figure 4: *(a) Histogram and fitted G(α,θ) probability density function for dataset II. (b) Theoretical and fitted distribution function for the dataset II.*

References

[1] Alizadeh, M. Merovci, F. and Hamedani, G. G. (2000). Generalized Transmuted Family of Distributions: Properties and Applications. *Hacetteppe University Bulletin of Natural science and Engineering Series B: Mathematics and Statistics*, 46: 645-668.

[2] Bourguignon, M, Silva and R B, Cordeiro, G M. (2014). The Weibull-G family of probability distributions. *Journal of Data Science*, 12: 53–68.

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

[4] Eugene, N. Lee, C and Famoye, F. (2002). Beta-normal Distribution and its applications. *Communications in Statistics – Theory and Methods*, 31: 497-512.

[5] Gupta, R. D. and Kundu, D. (1999). Generalized Exponential Distributions. *Australian and New Zealand Journal of Statistics*, 41(2): 173-188.

[6] Ijaz,M. , Asim, S. M. and Alamgir. (2019). Lomax exponential distribution with an application to real-life data, *PLOS ONE*, 14(12): 1-16.

[7] Jayakumar, K and Mathew, T. (2008). On a generalization to Marshall-Olkin scheme and its application to Burr type XII distribution. *Statistical Papers*, 49: 421–439.

[8] Lai, C.D., Zhang, L. and Xie, M. (2004). Mean Residual Life and Other Properties of Weibull Related Bathtub Shape Failure Rate Distribution. *International Journal of Reliability, Quality and Safety Engineering*, 11(2): 113-132.

[9] Marshall, A W and Olkin, I. A. (1997). new method for adding a parameter to a family of distributions with application to the exponential and weibull families. *Biometrika*, 84: 641–652.

[10] Nadarajah, S and S. Kotz. (2004). The beta Gumbel distribution. *Mathematical Problems in Engineering*, 10: 323–332.

[11] Nadarajah, S and A. K. Gupta. (2004). The beta Frechet distribution. *Far East Journal of Theoretical Statistics*, 14: 15–24.

[12] Nadarajah S and S. Kotz. (2006). The beta exponential distribution. *Reliability Engineering and System Safety* 91: 689–697.

[13] René Van Dorp, J. and Kotz, S. (2002). The Standard Two-Sided Power Distribution and it's Properties. *The American Statistician*, 56(2): 90-99.

[14] Silva, R B., Bourguignon, M., Dias, C R B, Cordeiro, G M. (2013). The compound class of extended Weibull power series distributions. *Computational Statistics and Data Analysis*, 58: 352-367.

[15] Zogarafos, K and Balakrishnan, N. (2009). On the families of beta-and generalized gamma-generated distribution and associated inference, *Statistical Methodology*, 6: 344–362.

Analysis on Dual Supply Inventory Model having Negative Arrivals and Finite Life Time Inventory

M. L. Soujanya[a,*], P. Vijaya Laxmi[b]

•

[a]Department of Mathematics,
Vignan's Institute of Information Technology (Autonomous),
Visakhapatnam, India.
drmlsoujanya@gmail.com
[b]Department of Applied Mathematics,
Andhra University, Visakhapatnam, India.
vijayalaxmiau@gmail.com

Abstract

In this paper the impact of dual supply chain on a perishable inventory model with negative arrivals is evaluated. The perishable and replenishment rates of dual suppliers are distributed exponentially. Arrival process follows Poisson distribution and the probability for an ordinary customer is p and for the negative customer is q. Limiting distribution of the assumed model is obtained. Numerical results are presented for cost function and various system performance parameters. The impact of dual suppliers on the optimal reorder points will be useful in developing strategies for handling various perishable inventory problems with replenishment rates.

Keywords: Matrix analytic method, Steady state distributions, Perishable inventory, Replenishment time, Negative arrivals, Dual supply inventory model, Cost function optimization.

I. Introduction

In an (s, S) inventory policy, an order of quantity $Q(= S - s)$ is placed if inventory drops to s, so that the maximum inventory level is S. This policy has been widely discussed for almost a century. However in inventory models with more than one supplier we can improve the quality of service, develop strong relationship with the customers, reduce loss of sales due to stock shortages, enhanced profits, etc. In dual supply (s, S) inventory policy, two orders of quantities Q_1 and Q_2 are placed whenever inventory level drops to r and s respectively. For literature on inventory models with dual supply chains one can refer [9] and [8].

A review of the literature on fixed time perishable inventory models was given by the life time of inventory items is indefinitely long in many classic inventory models, like vegetables, food items, medical products, etc., which become unusable after a certain span. That means there exists a real - life inventory system which consists of products having a finite lifetime. These types of products are called as perishable products and the corresponding inventory system can be considered as a perishable inventory system. [10] studied inventory models for perishable items with and without backlogging. A deterministic inventory model for perishable items with time dependent arrivals is developed by [5]. [1] discussed a perishable inventory model with style goals.

Finite waiting hall inventory model with negraive arrivals is introduced by [4]. For more literature on negative arrivals, one may refer [2],[3] and [6, 7].

In the present paper, an (s, S) inventory policy with dual supply chains for replenishment in which one having a shorter lead time is considered. Demands occur according to Poisson distribution. The arriving person may join the system with possibility p or remove one customer from the queue with probability q. The perishable and service rates are exponentially distributed. Limiting distributions are found. Several system performance parameters of the assumed model are presented. Also the analysis of the cost function is also carried out using direct search method.

II. Model description

In dual supply inventory model, when the inventory shrinks to a fixed level $r(> \frac{S}{2})$ an order of quantity $Q_1(= S - r)$ is placed from the first supplier and are replenished with an exponential rate η_1. If it drops to a prefixed level $s(< Q_1)$ an order of quantity $Q_2(= S - s > s + 1)$ is placed from the second supplier and is replenished with an exponential rate $\eta_2(\eta_2 > \eta_1)$. Arrival process follows Poisson distribution with rate λ. The arrived customer joins the queue with probability p and removes an existing customer from the end with probability $q(= 1 - p)$. Service time follows exponential distribution with rates μ. Let us assume that, the server stays idle at empty queue. Perishable rate follows exponential distribution with rate γ.

Let $N(t)$ be the queue length, $L(t)$ be the quantity of inventory and $\zeta(t)$ be the server state, which is defined as

$$\zeta(t) = \begin{cases} 1, & \text{server is idle at time } t; \\ 2, & \text{server is busy at time } t. \end{cases}$$

Therefore, the stochastic process $\{(\mathtt{N}(t), \zeta(t), L(t) \mid t \geq 0\}$ is a state space model with $E = \{(0,1,k) \mid 0 \leq k \leq S\} \bigcup \{(i,2,k) \mid i \geq 1,\ 1 \leq k \leq S\}\}$ and it is subdivided into level $\bar{i}$ defined as $\bar{0} = \{(0,1,0),\ (0,1,1),\ \ldots,\ (0,1,S)\}$ and $\bar{i} = \{(i,2,0),\ (i,2,1), \ldots, (i,2,S)\}$, for $i \geq 1$. Then the transition rate matrix P is

$$P = \begin{array}{c} \\ <0> \\ <1> \\ <2> \\ <3> \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} <0> & <1> & <2> & <3> & \ldots \end{array} \\ \begin{bmatrix} A_0 & C & 0 & 0 & \ldots \\ B & A & C & 0 & \ldots \\ 0 & B & A & C & \ldots \\ 0 & 0 & B & A & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}$$

where

$$[A_0] = \begin{cases} -(p\lambda + \eta_1 + \eta_2), & x = y, & y = 0; \\ -(p\lambda + \eta_1 + \eta_2 + y\gamma), & x = y, & y = 1, 2, \ldots, s; \\ -(p\lambda + \eta_1 + y\gamma), & x = y, & y = s+1, \ldots, r; \\ -(p\lambda + y\gamma), & x = y, & y = r+1, \ldots, S; \\ y\gamma, & x = y - 1, & y = 1, \ldots, S; \\ \eta_1, & x = y + Q_1, & y = 0, \ldots, r; \\ \eta_2, & x = y + Q_2, & y = 0, \ldots, r; \\ 0, & \text{otherwise.} & \end{cases},$$

$$[C] = \begin{cases} p\lambda, & x = y,\ y = 0, \ldots, S; \\ 0, & \text{otherwise.} \end{cases}, \quad [B] = \begin{cases} \mu, & x = y - 1, y = 1, \ldots, S; \\ q\lambda, & x = y,\ y = 0, \ldots, S; \\ 0, & \text{otherwise.} \end{cases},$$

$$[A] = \begin{cases} -(p\lambda + q\lambda + \eta_1 + \eta_2), & x = y, & y = 0; \\ -(p\lambda + q\lambda + \eta_1 + \eta_2 + y\gamma + \mu), & x = y, & y = 1, 2, \ldots, s; \\ -(p\lambda + +q\lambda + \eta_1 + y\gamma + \mu), & x = y, & y = s+1, \ldots, r; \\ -(p\lambda + q\lambda + y\gamma + \mu), & x = y, & y = r+1, \ldots, S; \\ y\gamma, & x = y - 1, & y = 1, \ldots, S; \\ \eta_1, & x = y + Q_1, & y = 0, \ldots, r; \\ \eta_2, & x = y + Q_2, & y = 0, \ldots, r; \\ 0, & \text{otherwise.} & \end{cases},$$

III. Analysis of the Model

Initially the stability condition of the defined model is determined and then the limiting probabilities are derived in this section.

I. Stability condition

For the stability condition, consider the matrix $G = A + B + C$ as

$$[G] = \begin{cases} -(\eta_1+\eta_2), & x=y, \quad y=0; \\ -(\eta_1+\eta_2+\mu_b+y\gamma), & x=y, \quad y=1; \\ -(\eta_1+\eta_2+y\gamma), & x=y, \quad y=2,\ldots,s; \\ -(\eta_1+y\gamma), & x=y, \quad y=s+1,\ldots,r; \\ -y\gamma, & x=y, \quad y=r+1,\ldots,S; \\ y\gamma, & x=y-1, \quad y=2,\ldots,S; \\ \eta_1, & x=y+Q_1, \ y=1,\ldots,r; \\ \eta_2, & x=y+Q_2, \ y=1,\ldots,r; \\ 0, & \text{otherwise.} \end{cases},$$

Let Π be the limiting distribution of G, i.e., $\Pi G = 0$ and $\Pi e = 1$ where $\Pi = (\Pi(2,0), \Pi(2,1), \ldots, \Pi(2,S))$. From $\Pi G = 0$, we get

$$-(\eta_1+\eta_2)\Pi(2,0) + \gamma\Pi(2,1) = 0$$
$$-(\eta_1+\eta_2+y\gamma)\Pi(2,l) + y\gamma\Pi(2,l+1) = 0, 1 \le y \le s$$
$$-(\eta_1+y\gamma)\Pi(2,l) + y\gamma\Pi(2,l+1) + \eta_2\Pi(2,l+Q_2) = 0, s+1 \le y \le r$$
$$-(\eta_1+y\gamma)\Pi(2,l) + y\gamma\Pi(2,l+1) + \eta_2\Pi(2,l+Q_1) = 0, r+1 \le y \le S$$

On solving the above two equations and by using $\Pi e = 1$, one can get the limiting probability values.

II. Computation of Steady State Vectors

The limiting distribution for the defined model is

$$\pi_i^{(j,k)} = \lim_{t\to\infty} Pr\Big[\mathtt{N}(t) = i, \zeta(t) = j, L(t) = k \mid \mathtt{N}(0), \zeta(0), L(0)\Big]$$

where $\pi_i^{(j,k)}$ is the probability of i^{th} demand at j^{th} state of the server with k inventories. These probabilities are shortly represented as π_i. The limiting distribution is given by $\pi_i = \pi_0 R^i,\ i \ge 1$.

For finding π_0 and π_1, we have from $\pi P = 0$,

$$\pi_1 = -\pi_0 C(A+RB)^{-1} = \pi_0 w,$$

where

$$w = -C(A+RB)^{-1}.$$

Further, $\pi_0 A_0 + \pi_1 B = 0$, *i.e.*, $\pi_0(A_0 + wB) = 0$.

First take π_0 as the limiting distribution of $A_0 + wB$. Then π_i, for $i \ge 1$ can be found using $\pi_1 = \pi_0 w,\ \pi_i = \pi_1 R^{i-1}, i \ge 2$. Therefore the limiting distribution of the system is obtained as follows.

$$(\pi_0 + \pi_1 + \pi_2 + \ldots)e = \pi_0(1 + w(I-R)^{-1})e.$$

IV. System performance measures

(i) **Average inventory level:** The average inventory level (E_{IL}) is defined as

$$E_{IL} = \sum_{k=0}^{S} k\pi_0^{(1,k)} + \sum_{i=1}^{\infty}\sum_{k=0}^{S} \pi_i^{(2,k)}.$$

(ii) **Average service rate:** Let E_{SR} denote the average service rate and is given by

$$E_{SR} = \sum_{i=1}^{\infty} \mu\pi_i^{(2,k)}.$$

(iii) **Average reorder rate:** Average reorder rate (E_{OR}) is given by

$$E_{OR} = \begin{cases} \sum_{k=0}^{S} \gamma\left[\pi_0^{(1,s+1)} + \pi_0^{(1,r+1)}\right] + \\ \sum_{i=1}^{\infty}\sum_{k=0}^{S}(\mu+\gamma)\left[\pi_i^{(2,s+1)} + \pi_i^{(2,r+1)}\right]. \end{cases}$$

(iv) **Average negative arrivals:** The average negative arrivals (E_{NA}) is given by

$$E_{NA} = \sum_{i=1}^{\infty}\sum_{k=1}^{S} q\lambda\pi_i^{(2,k)}.$$

(v) **Average arrival rate:** The average arrival rate (E_{AR}) is given by

$$E_{AR} = \sum_{i=1}^{\infty}\sum_{k=0}^{S} p\lambda\pi_i^{(2,k)}.$$

(vi) **Average replenishment rate from 1^{st} supplier:** The average replenishment rate from 1^{st} supplier (E_{RR1}) is

$$E_{RR1} = \sum_{k=0}^{r} \eta_1\pi_0^{(1,k)} + \sum_{i=1}^{\infty}\sum_{k=0}^{r} \eta_1\pi_i^{(2,k)}.$$

(vii) **Average replenishment rate from 2^{nd} supplier:** The average replenishment rate from 2^{nd} supplier (E_{RR2}) is

$$E_{RR2} = \sum_{k=0}^{s} \eta_2\pi_0^{(1,k)} + \sum_{i=0}^{\infty}\sum_{k=0}^{s} \eta_2\pi_i^{(2,k)}.$$

(viii) **Average lifetime:** The average lifetime (E_{FR}) is defined as

$$E_{FR} = \sum_{k=1}^{S} k\gamma\pi_0^{(1,k)} + \sum_{i=1}^{\infty}\sum_{k=1}^{S} \pi_i^{(2,k)}.$$

I. Cost analysis

Let

C_H = Carrying cost,

C_S = Set up cost,

C_{S1} = cost for the 1^{st} supplier,

C_{S2} = cost for the 2^{nd} supplier,

C_F = Failure cost,

C_N = Loss incurred due to negative customer,

C_A = Fixed cost for arrivals,

C_{ST} = service cost.

Therefore, the total average cost is defined as

$$TC(s,r) = \begin{cases} C_H\ E_{IL} + C_S\ E_{OR} + C_{S1}\ E_{RR1} + C_{S2}\ E_{RR2} + \\ C_F E_{PR} + C_N\ E_{NA} + C_A\ E_{AR} + C_{ST}\ E_{ST}. \end{cases}$$

V. Numerical analysis

For this section, let us fix the parameters as $S = 14, p = 0.6, q = 0.4, \lambda = 2.3, \mu = 1.2, \eta_1 = 2.4, \eta_2 = 4.7, \gamma = 0.2$.

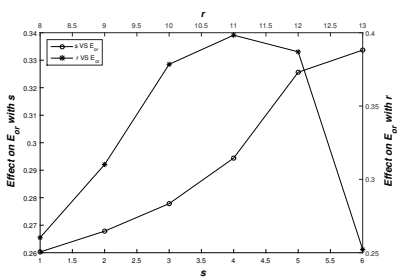


Figure 1: Impact of (s,r) on E_{OR}

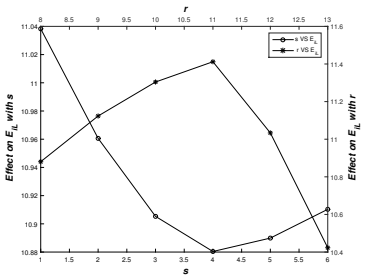


Figure 2: Impact of (s,r) on E_{IL}

Figure 1 and Figure 2 presents the influence of reordering points (s,r) on E_{OR} and E_{IL} respectively. As we know that, E_{OR} and E_{IL} increases with the increase of reorder points, which is evident from Figure 1 and Figure 2. However, one can observe that E_{OR} decreases from $r = 11$ as reordering is done very near to S. From Figure 2, one can also observe that E_{IL} is decreased till $s = 4$ since reordering is done very slowly.

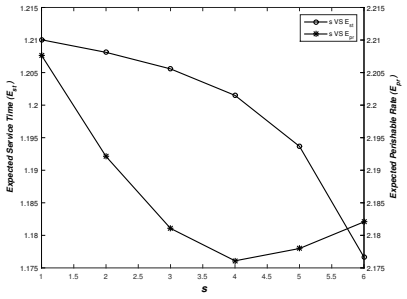


Figure 3: Impact of s on E_{PR} & E_{ST}

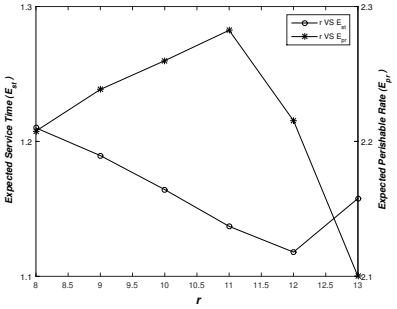


Figure 4: Impact of r on E_{PR} & E_{ST}

The influence of s and r on E_{PR} and E_{ST} are shown in Figure 3 and Figure 4, respectively. According to our assumption $s < S - r$, first order is replenished when the inventory reaches to r. Also $\eta_2 > \eta_1$, the replenishment time of the first supplier is greater than that of the second supplier, due to this E_{ST} decreases with increase in s and r, E_{PR} decreases with s and increases with r.

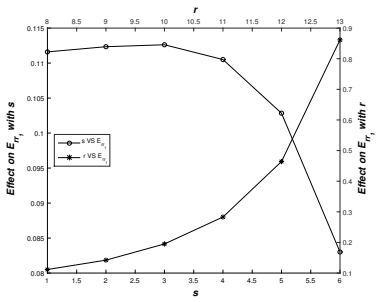


Figure 5: Impact of (s,r) on E_{RR1}

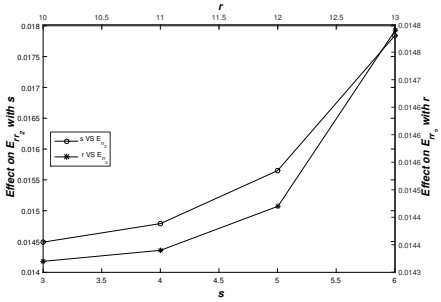


Figure 6: Impact of (s,r) on E_{RR2}

Figure 5 and Figure 6 presents the influence of s and r on E_{RR1} and E_{RR2} respectively. Since E_{RR1} is effected with r, it rises as r increases and falls as s increases as shown in Figure 5. Even though E_{RR2} is effected with s, the second replenishment order is placed only after the first replenishment order is done. Due to this E_{RR2} rises with the increase in both s and r which is clearly evident from Figure 6.

Table 1: *The values of s^*, r^* and $TC(s^*, r^*)$*

Case 1	(s^*, r^*)	(4,8)	(5,9)	(7,10)	(7,11)	(4,12)	(7,13)
	$TC(s^*, r^*)$	893.33	**891.62**	895.59	889.00	872.29	648.25
Case 2	(s^*, r^*)	(4,8)	(5,9)	(7,10)	(6,11)	(4,12)	(7,13)
	$TC(s^*, r^*)$	1437.35	**1436.10**	1462.06	1455.07	1423.96	1373.02
Case 3	(s^*, r^*)	(4,8)	(5,9)	(7,10)	(7,11)	(4,12)	(7,13)
	$TC(s^*, r^*)$	899.22	**897.79**	903.77	897.88	979.79	854.92
Case 4	(s^*, r^*)	(4,8)	(5,9)	(7,10)	(7,11)	(4,12)	(7,13)
	$TC(s^*, r^*)$	893.88	**892.27**	896.37	890.19	874.56	851.97
Case 5	(s^*, r^*)	(4,8)	(5,9)	(7,10)	(7,11)	(4,12)	(7,13)
	$TC(s^*, r^*)$	893.40	**891.69**	895.71	889.12	872.38	848.39

Table 1 gives s^* and r^* that minimize $TC(s,r)$, for different numerical examples which are defined as the following cases.

1: $C_H = 30,\ C_S = 100\ C_{S1} = 10,\ C_{S2} = 30,\ C_F = 13,\ C_N = 150,\ C_A = 250,\ C_{ST} = 20$

2: $C_H = 80,\ C_S = 100\ C_{S1} = 10,\ C_{S2} = 30,\ C_F = 13,\ C_N = 150,\ C_A = 250,\ C_{ST} = 20$

3: $C_H = 30,\ C_S = 120\ C_{S1} = 10,\ C_{S2} = 30,\ C_F = 13,\ C_N = 150,\ C_A = 250,\ C_{ST} = 20$

4: $C_H = 30,\ C_S = 100\ C_{S1} = 15,\ C_{S2} = 30,\ C_F = 13,\ C_N = 150,\ C_A = 250,\ C_{ST} = 20$

5: $C_H = 30,\ C_S = 100\ C_{S1} = 10,\ C_{S2} = 35,\ C_F = 13,\ C_N = 150,\ C_A = 250,\ C_{ST} = 20$

One can observe that the optimum reorder point in all the cases is $(s,r) = (5,9)$ as 5the optimum total cost exists. We know that the average inventory level is more when compared to other performance measures discussed in Section 4. However the total cost function increases with increase in holding cost which is evident form Case 2.

VI. Conclusion

In this paper, some investigations are done on the impact of two suppliers on an inventory model having negative arrivals and finite lifetimes. The limiting distribution of the assumed model is derived. Various system performance parameters are discussed and analyzed the assumed cost function to obtain s^* and r^*.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

References

[1] Dinesh S. and Roberts R. (2017). Inventory models for perishable items and style goods. *Problems & Solutions in Inventory Management*, 233–250.

[2] Manual P., Sivakumar B. and Arivarignan G. (2007). A perishable inventory system with service facilities, MAP arrivals and PH-service time. *Journal of Systems Science and Systems Engineering*, 16, 62–73.

[3] Manual P., Sivakumar B. and Arivarignan G.(2007). Perishable inventory systems with postponed demands and negative customers. *Journal of Applied Mathematics and Decision Science*, 1–12.

[4] Sivakumar B. and Arivarignan G.(2005). A perishable inventory system with service facilities and negative customers, *Advance Modeliling Optimization, 7, 193–210.*

[5] Sushil Kumar and Ravendra Kumar. (2015). A deterministic inventory model for perishable items with time dependent demand and shortages. International Journal of Mathematics And its Applications, 3(4-F), 105–111.

[6] Vijaya Laxmi P. and Soujanya M. L. (2017). Perishable inventory model with retrial demands, negative customers and multiple working vacations, International Journal of Mathematical Modelling and Computations, 7(4), 239–254.

[7] Vijaya Laxmi P. and Soujanya M. L. (2018). Perishable inventory model with Markovian arrival procees, retrial demands and multiple working vacations, International Journal of Inventory Research, 5(2), 79–98.

[8] Vijayashree, M. and Uthaykumar, R. (2016). Two-Echelon supply chain inventory model with controllable lead time, International Journal of System Assurance Engineering and Management, 7, 112—125.

[9] Yang M. F. and Wei-Chung, Three-Echelon inventory model with permissible delay in payments under controllable lead time and backorder consideration, Mathematical Problems in Engineering, 2014, Artical ID: 809149, 16 Pages.

[10] Yonguri Duan, Guiping Li, James M. (2012). Tien and Jiazhen Huo, Inventory models for perishable items with inventory level dependent demand rate, Applied Mathematicall Modelling, 36(10), 5015–2028.

Harris Extended Two Parameter Lindley Distribution and Applications in Reliability

Sophia P.Thomas

Lishamol Tomy

K.K.Jose

•

Department of Statistics, St.Thomas College, Pala, Kerala, India
Department of Statistics, St.Thomas College, Pala, Kerala, India
School of Mathematics, Statistics & Data Analytics, Mahatma Gandhi University, Kottayam, Kerala, India
sophiathomas92@gmail.com
lishatomy@gmail.com
kkjstc@gmail.com

Abstract

This paper introduces a new generalization of the two parameter Lindley distribution distribution namely, Harris extended two parameter Lindley distribution. Various structural properties of the new distribution are derived including moments, quantile function, Renyi entropy, and mean residual life. The model parameters are estimated by maximum likelihood method. The usefulness of the new model is illustrated by means of two real data sets on Wheaton river flood and bladder cancer. Also, we derive a reliability test plan for acceptance or rejection of a lot of products submitted for inspection with lifetimes following this distribution. The operating characteristic functions of the sampling plans are obtained. The producer's risk, minimum sample sizes and associated characteristics are computed and presented in tables. The results are illustrated using two data sets on ordered failure times of products as well as failure times of ball bearings.

Keywords: Acceptance Sampling Plan, Harris extended two parameter Lindley distribution, Lindley distribution, maximum likelihood estimation, hazard rate, extreme order statistics.

1. Introduction

The Lindley distribution was introduced by Lindley (1958) to analyze failure time data, especially in applications to modeling stress-strength reliability. The motivation of the Lindley distribution arises from its ability to model failure time data with increasing, decreasing, unimodal and bathtub shaped hazard rates. The Lindley distribution belongs to the exponential family and it can be written as a mixture of exponential and gamma distributions. The properties and inferential procedure for the Lindley distribution were studied by Ghitany et al. (2008, 2011). It is shown that the Lindley distribution is better than the exponential distribution when hazard rate is unimodal or bathtub shaped. Mazucheli and Achcar (2011) also proposed the Lindley distribution as a possible alternative to exponential and Weibull distributions.

In recent years, there have been many studies to obtain a new distribution based on modifications of the Lindley distribution for modeling data in biology, medicine, finance, and engineering. To name a few extensions, three parameter generalized Lindley suggested by Zakerzadeh and Dolati (2009), New Generalized Lindley Distribution(NGLD) proposed by Abouammoh et al. (2015), two-parameter weighted Lindley distribution developed by Ghitany et al.(2011) , power

Lindley by Ghitany et al. (2013), and Merovci (2013) proposed transmuted Lindley distribution. The Lindley distribution has been generalized by different researchers including Elbatal et al.(2013), Liyanage and Parai (2014), Nadaeajah et al. (2011), Oluyede and Yang (2014), Shanker and Fesshaye (2015), Kemaloglu and Yilmaz (2017) are some among others. Some recent works based on the Lindley distribution are wrapped Lindley distribution by Joshi and Jose (2018), three parameter generalized Lindley by Ekhosuehi and Opone (2018), Lindley Weibull distribution by Cordeiro et al. (2018), modified Lindley distribution by Chesneau et al. (2019a), wrapped modified Lindley distribution by Chesneau et al. (2019b), inverted modified Lindley distribution by Chesneau et al. (2020a), and sum and difference of two Lindley distributions by Chesneau et al. (2020b). The Lindley distribution does not provide enough flexibility for analyzing different types of lifetime data because of having only one parameter. To increase the flexibility for modelling purposes it will be useful to consider further alternatives of this distribution. Therefore, the aim of this study is to introduce a new family of distributions using the Lindley generator. In these directions, one can also study the properties of the Harris Extended generalization of two-parameter Lindley distribution. The main idea of this technique is to get more flexible structures than the base distribution.

The procedure of adding one or two parameters to a family of distributions to obtain more flexibility is a well-known technique in the existing literature. Aly and Benkherouf (2011) introduced a new family of distributions, called the Harris Extended (HE) family by adding two new parameters to a baseline distribution. The new method is based on the probability generating function (pgf) of Harris (1948) distribution. If $\bar{F}(x)$, $f(x)$, and $r_F(x)$ denote the survival function (sf), probability density function (pdf) and hazard rate function (hrf) of a parent distribution, then the sf $\bar{G}(x)$ of HE family of distribution is given by;

$$\bar{G}(x) = \left[\frac{\lambda \bar{F}(x)^k}{1-\bar{\lambda}\bar{F}(x)^k}\right]^{1/k} \quad ; \quad x>0, \quad \alpha>0, \quad \bar{\lambda}=1-\lambda, \quad k>0. \tag{1}$$

Here, the parameters α and k are additional shape parameters that aims to introduce greater flexibility. The HE density function is;

$$g(x) = \frac{\lambda^{1/k} f(x)}{[1-\bar{\lambda}\bar{F}(x)^k]^{(k+1)/k}} \quad ; \quad x>0, \quad \lambda>0, \quad \bar{\lambda}=1-\lambda, \quad k>0. \tag{2}$$

The hrf of the HE distribution is given by

$$r(x) = \frac{r_F(x)}{[1-\bar{\lambda}\bar{F}(x)^k]}, \quad x>0$$

where $r_F(x)$ denotes the failure rate function of the baseline distribution. When $k=1$, the above equations reduces to those of Marshall-Olkin family of distributions, introduced by Marshall and Olkin (1997). Hence the HE family of distributions generalizes the well-known Marshall-Olkin class of distributions. Batsidis and Lemonte (2014) considered the HE family of distributions with respect to some lifetime models. Pinho et al. (2015), introduced and studied Harris extended exponential model. More recently, Jose and Paul (2018) derived reliability test plans for percentiles based on Harris generalized linear exponential distribution. Jose et al.(2018) developed reliability test plans for Harris extended Weibull distribution.

In this article, we introduce a new variant of Harris extended model by considering the baseline distribution to be a two parameter lindley distribution suggested by Shankar et al.(2013). The article is organized as follows: The derivation of the Harris extended two-parameter Lindley distribution (HETLD), hazard rate and cumulative hazard rate are presented in Section 2. Linear representation of the density function is presented in Section 3. Statistical properties include moments, quantile function, Renyi entropy, mean residual life and maximum likelihood estimation are explored in Section 4. The new distribution is illustrated on real datasets in Section 5. In section 6, the proposed sampling plans are established for the Harris extended two parameter Lindley distribution. Finally, conclusions are given in Section 7.

2. Harris Extended Two Parameter Lindley Distribution

If $\bar{F}(x)$, $f(x)$, and $r_F(x)$ denote the survival function (sf), probability density function (pdf) and hazard rate function (hrf) of a parent distribution, then the baseline sf of two parameter Lindley distribution (TLD) is given by,

$$\bar{F}(x;\beta,\theta) = \left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)e^{-\theta x};\ \ x>0,\ \ \beta>0,\ \ \theta>0$$

and the corresponding pdf is given by

$$f(x;\beta,\theta) = \frac{\theta^2}{\theta+\beta}(1+\beta x)e^{-\theta x};\ \ x>0,\ \ \beta>0,\ \ \theta>0$$

For detailed information, see Shanker et al. (2013). Substituting the sf of TLD in (1), the distribution called (Harris extended two-parameter Lindley distribution) HETLD is obtained. It is denoted by HETLD (λ,k,β,θ) with sf

$$\bar{G}(x) = \left\{\frac{\lambda\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}}{1-\bar{\lambda}\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}}\right\}^{1/k} \tag{3}$$

where, $(\lambda,k,\beta,\theta)>0$, $\bar{\lambda}=1-\lambda$
pdf of HETLD is given by;

$$g(x) = \frac{\lambda^{1/k}\frac{\theta^2}{\theta+\beta}(1+\beta x)e^{-\theta x}}{\left\{1-\bar{\lambda}\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}\right\}^{1+\frac{1}{k}}}\ \ ;\ \ x>0,\quad (\lambda,k,\beta,\theta)>0,\quad \bar{\lambda}=1-\lambda. \tag{4}$$

Figure 1 illustrates some possible shapes of the pdf of HETLD for different values of the parameters λ, k, β, and θ.
The hrf of HETLD is given as

$$r_{HETLD}(x) = \frac{\left(\frac{\theta^2(1+\beta x)}{\theta+\beta+\beta\theta x}\right)}{\left[1-\bar{\lambda}(\frac{\theta+\beta+\beta\theta x}{\theta+\beta})^k e^{-\theta k x}\right]}$$

Plots of hrf of HETLD for different values of parameters β, θ, λ and k are given in Figure 2. It can be seen that, the hrf of HETLD is attractively flexible. Therefore, the new distribution can be used quite effectively for analyzing different types of lifetime data in practice.

The cumulative hazard rate function of HETLD is given by

$$H_{HETLD}(x) = \frac{1}{k}ln\left[1-\bar{\lambda}\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}\right] - \frac{1}{k}ln\left[\lambda\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}\right]$$

Figure 4 illustrates the behavior of the cumulative hazard rate function of the HETLD distribution at different values of the parameters.

3. Linear representation of the density function

In this section, we obtain a useful expansion for the HETLD. For $|\ z\ |<1$ and $r>0$, we have

$$(1-z)^{-r} = \sum_{i=0}^{\infty}\binom{r+i-1}{i}z^i \tag{5}$$

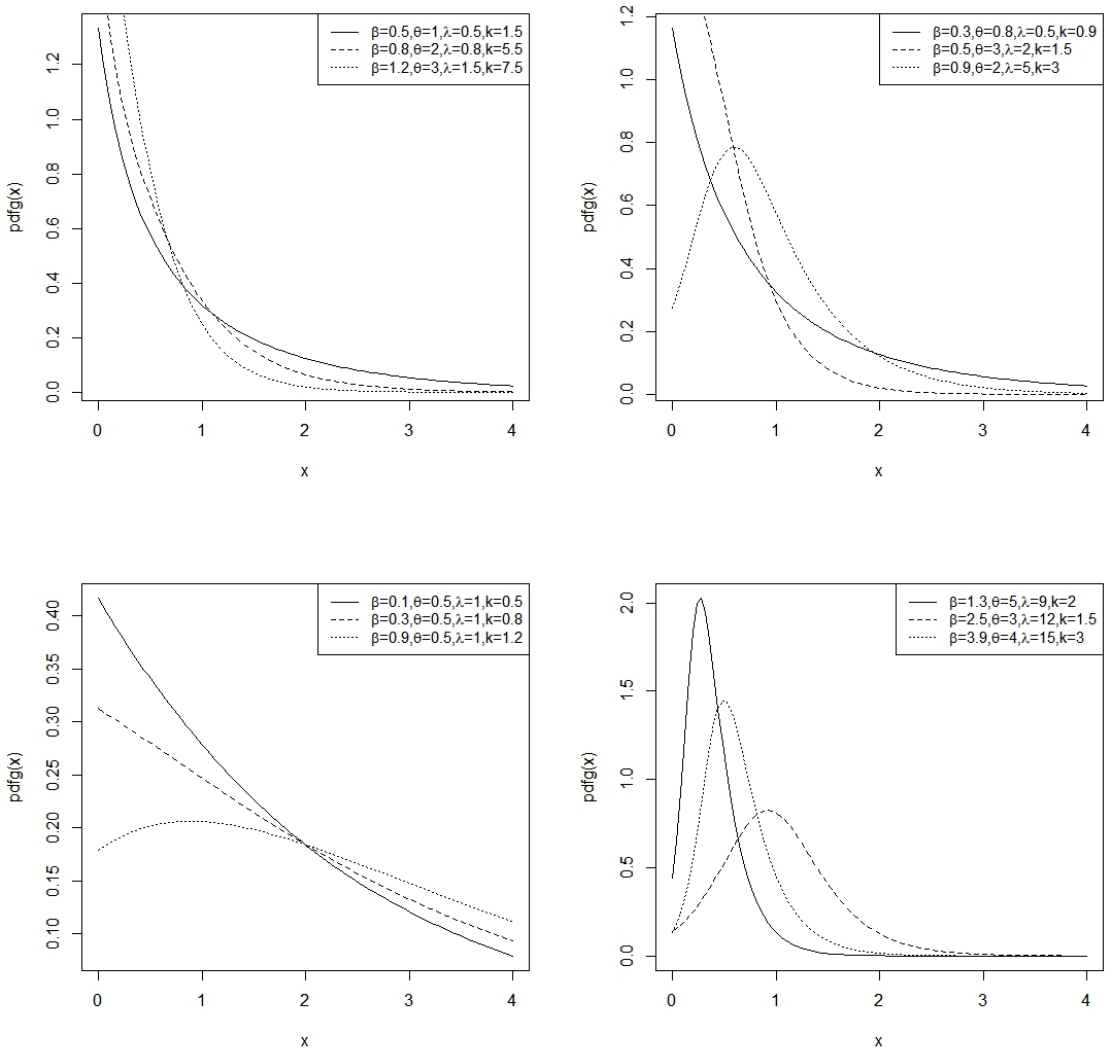


Figure 1: *pdf of HETLD distribution for various values of β, θ, λ and k.*

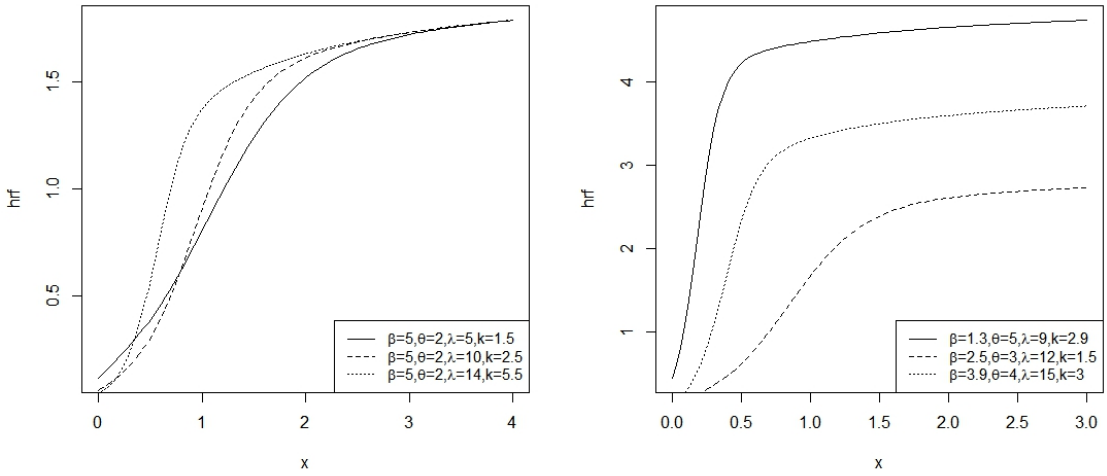


Figure 2: *hrf of HETLD for various values of β, θ, λ and k.*

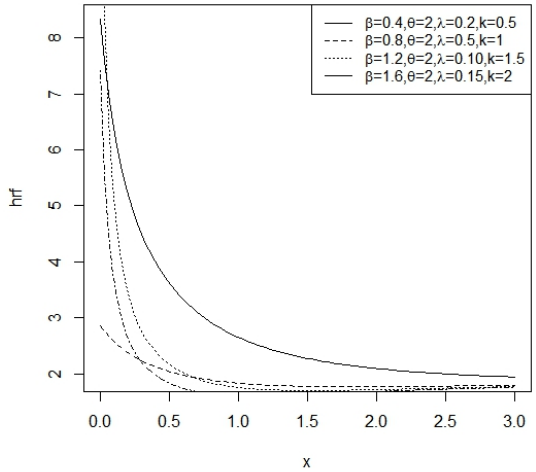


Figure 3: *HETLD for various values of β, θ, λ and k.*

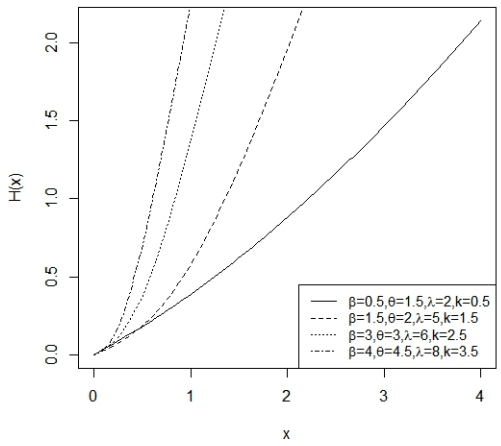


Figure 4: *The cumulative hazard rate function of the HETLD*

Applying Equation (5) in Equation (4), for $\lambda \in (0,1)$, yields

$$g(x) = \sum_{i=0}^{\infty} w_i f(x;\beta,\theta); \quad x > 0 \tag{6}$$

where,

$$w_i = \binom{i+k^{-1}}{i} \frac{\lambda^{1/k}(1-\lambda)^i}{1+ik}$$

otherwise, if $\lambda > 1$, we can obtain

$$g(x) = \sum_{i=0}^{\infty} v_i f(x;\beta,\theta); \quad x > 0 \tag{7}$$

where,

$$v_i = \frac{\frac{1}{\lambda}(-1)^i}{1+ik} \sum_{j=i}^{\infty} \binom{j+k^{-1}}{j} \binom{j}{i} (1-\frac{1}{\lambda})^j$$

and $f(x;\beta,\theta)$ denotes the Two parameter Lindley density function. Thus HETLD as an infinite linear combination of Two parameter Lindley density functions. Equations (6) and (7) have the same form except for the coefficients which are w_i's in Equation (6) and v_i's in Equation (7).

4. Statistical Properties

In this section, statistical properties of the HETLD including the moments, quantile function, Rẻnyi entropy and limiting distributions of order statistics, stochastic Orders, Lorenz and Bonferroni curves and mean residual lifetime are given. Furthermore, the estimation of the HETLD parameters based on maximum likelihood is discussed.

4.1. Quantiles

Recently, Bensid and Zeghdoudi (2017) showed that the quantile function of the two parameter Lindley distribution is given by

$$F^{-1}(v) = -\frac{\theta+\beta}{\theta\beta} - \frac{1}{\theta} W\left(-\frac{1}{\beta}(1-v)(\theta+\beta)e^{(-\frac{\theta+\beta}{\beta})}\right), \qquad 0 < v < 1,$$

where $W_{-1}(.)$ denotes the negative branch of the Lambert W function (i.e., the solution of the equation $W(z)e^{W(z)} = z$).
The quantile function, $Q(u)$, $0 < u < 1$, for the Harris Extended (HE) family of distributions is computed by using the formula of Batsidis and Lemonte (2014) as

$$G^{-1}(u) = F^{-1}\left(1 - (1-u)[\lambda + \bar{\lambda}(1-u)^k]^{-1/k}\right)$$

where $G^{-1}(.)$ and $F^{-1}(.)$ are the inverse functions of G(.) (obtained from Equation (10)) and F(.) (baseline cumulative function), respectively. The quantile function of the HETLD is given by

$$x = G^{-1}(u) = -(\frac{\theta+\beta}{\theta\beta}) - \frac{1}{\theta}W\left[-\frac{1}{\beta}\left(\frac{1-u}{[\lambda+\bar{\lambda}(1-u)^k]^{1/k}}\right)(\theta+\beta)e^{(-\frac{\theta+\beta}{\beta})}\right] \tag{8}$$

Simulation of HETL random variable follows directly from (8). The quantiles of the HETLD can be obtained by setting $u = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$.

4.2. Moments

Here, we shall derive a general expression for the moments of HETLD$(\lambda, k, \beta, \theta)$, in terms of the Probability Weighted Moments (PWMs) of the baseline distribution. From Equation (6) and for $\lambda \in (0,1)$, r^{th} moment of the HETLD can be expressed as

$$\mu'_r = \sum_{j=0}^{\infty} w_j \tau_{r,jk} \tag{9}$$

For $\lambda > 1$, Equation (7) holds replace w_j with v_j
where,

$$\begin{aligned}
\tau_{r,jk} &= \int_0^\infty x^r [\bar{F}x]^{jk} f(x)dx \\
&= \frac{\theta^2}{\beta+\theta}\int_0^\infty x^r \left[\left(1+\frac{\beta\theta x}{\theta+\beta}\right)e^{-\theta x}\right]^{jk}(1+\beta x)e^{-\theta x}dx \\
&= \sum_{i=0}^{\infty}\binom{jk}{i}\left(\frac{\beta\theta}{\theta+\beta}\right)^i \frac{\theta^2}{\beta+\theta}\left(\frac{\Gamma(r+1)}{[\theta(1+jk)]^{r+1}} + \frac{\beta\Gamma(r+2)}{[\theta(1+jk)]^{r+2}}\right)
\end{aligned}$$

Substituting $\tau_{r,jk}$ in Equation (9) μ'_r of HETLD, for $\lambda \in (0,1)$,

$$\mu'_r = E(X^r) = \frac{\lambda^{1/k}\theta^2}{(\beta+\theta)}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(1-\lambda)^j \frac{\Gamma(k^{-1}+1+j)}{\Gamma(k^{-1}+1)j!}\binom{jk}{i}\left(\frac{\beta\theta}{\theta+\beta}\right)^i \left\{\frac{\Gamma(r+1)}{[\theta(1+jk)]^{r+1}} + \frac{\beta\Gamma(r+2)}{[\theta(1+jk)]^{r+2}}\right\}$$

For $\lambda > 1$

$$\begin{aligned}
\mu'_r = E(X^r) = \frac{\lambda^{-1}\theta^2}{(\beta+\theta)}\sum_{i=0}^{\infty}\sum_{l=j}^{\infty}\binom{l}{j}(-1)^j(1-\frac{1}{\lambda})^l \frac{\Gamma(k^{-1}+1+l)}{\Gamma(k^{-1}+1)l!}\binom{jk}{i}\left(\frac{\beta\theta}{\theta+\beta}\right)^i \\
\times\left\{\frac{\Gamma(r+1)}{[\theta(1+jk)]^{r+1}} + \frac{\beta\Gamma(r+2)}{[\theta(1+jk)]^{r+2}}\right\}
\end{aligned}$$

and $\Gamma.$ is the complete gamma function.

4.3. Rényi entropy

The *Rényi* entropy of a continuous random variabe X distributed according to the HEW is derived by the following formula

$$I_R(\delta) = \frac{1}{1-\delta}log\int_0^\infty g^\delta(x)dx \ \ , \delta > 0$$

$$\int_0^\infty g^\delta(x)dx = (\lambda)^{\delta/k}\left(\frac{\theta^2}{\beta+\theta}\right)^\delta \int_0^\infty \frac{(1+\beta x)^\delta e^{-\delta\theta x}}{[1-\bar{\lambda}(1+\frac{\beta\theta x}{\theta+\beta})^k e^{-\theta kx}]^{\delta+\frac{\delta}{k}}}dx$$

Using the expansions: $(1-z)^{-\delta} = \sum_{j=0}^{\infty}\frac{\Gamma(\rho+j)}{\Gamma(\rho)j!}z^j$ to expand the expression

$$\left[1-\bar{\lambda}(1+\left(\frac{\beta\theta x}{\theta+\beta}\right)^k e^{-\theta kx}\right]^{\delta+\frac{\delta}{k}} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\Gamma(\delta+\frac{\delta}{k}+j)}{\Gamma(\delta+\frac{\delta}{k})j!}\binom{jk}{i}(\bar{\lambda})^j\left(\frac{\theta\beta}{\theta+\beta}\right)^i x^i e^{-\theta kjx}$$

then

$$\int_0^\infty g^\delta(x)dx = \left[\frac{\lambda^{1/k}\theta^2}{\beta+\theta}\right]^\delta \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\sum_{n=0}^{\infty}\frac{\Gamma(\delta+\frac{\delta}{k}+j)}{\Gamma(\delta+\frac{\delta}{k})j!}\binom{jk}{i}(\bar{\lambda})^j\left(\frac{\theta\beta}{\theta+\beta}\right)^i\binom{\delta}{n}\beta^n\int_0^\infty x^{n+i}e^{-(j+\delta)\theta x}dx$$

The integral $I = \int_0^\infty x^{n+i}e^{-(j+\delta)\theta x}dx$ can be evaluated as

$$\begin{aligned} I &= \int_0^\infty x^{n+i}e^{-(j+\delta)\theta x}dx \\ &= \frac{\Gamma(n+i+1)}{[(j+\delta)\theta]^{n+i+1}} \end{aligned}$$

Collecting all of the above evaluations, the Renyi entropy of HETLD can be written as

$$I_R(\delta) = \frac{1}{1-\delta}\left\{lnC + ln\left(\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\sum_{n=0}^{\infty}\frac{\Gamma(\delta+\frac{\delta}{k}+j)}{\Gamma(\delta+\frac{\delta}{k})j!}\binom{jk}{i}(\bar{\lambda})^j\left(\frac{\theta\beta}{\theta+\beta}\right)^i\binom{\delta}{n}\beta^n\frac{\Gamma(n+i+1)}{[(j+\delta)\theta]^{n+i+1}}\right)\right\}$$

where, $C = \left[\frac{\lambda^{1/k}\theta^2}{\beta+\theta}\right]^\delta$

4.4. The Mean Residual Lifetime

The additional lifetime given that the component has survived up to time x is called the residual life function of the component, then the expectation of the random variable X that represent the remaining lifetime is called the mean residual lifetime (MRL) and is given by

$$\begin{aligned} m(x) &= E(X-x \mid X \geq x) \\ &= \left\{\frac{1}{\bar{F}(x)}\int_x^\infty tf(t)dt\right\} - x \end{aligned}$$

The MRL function m(x) for HETLD random variable can be derived in the following steps.

$$\int_x^\infty tf(t)dt = \lambda^{1/k}\frac{\theta^2}{\beta+\theta}\int_x^\infty \frac{(t+\beta t^2)e^{-\theta t}}{[1-\bar{\lambda}(1+\frac{\beta\theta t}{\theta+\beta})^k e^{-\theta kt}]^{1+\frac{1}{k}}}dt$$

Using the expansion $(1-z)^{-\delta} = \sum_{j=0}^{\infty}\frac{\Gamma(\delta+j)}{[\Gamma(\delta)j!]}z^j, \ \mid z \mid < 1$ one has

$$[1-\bar{\lambda}(1+\frac{\beta\theta t}{\theta+\beta})^k e^{-\theta kt}]^{1+\frac{1}{k}} = \sum_{j=0}^{\infty}\frac{\Gamma(1+k^{-1}+j)}{[\Gamma(1+k^{-1})j!]}\bar{\lambda}^j(1+\frac{\beta\theta t}{\theta+\beta})^{jk}e^{-\theta kjt}$$

Similarly, using the expansion $(1+b)^n = \sum_{i=0}^{\infty}\binom{n}{i}b^{n-i}$ one can have

$$(1+\frac{\theta\beta t}{\theta+\beta})^{jk} = \sum_{i=0}^{\infty}\binom{jk}{i}(\frac{\theta\beta}{\theta+\beta})^{jk-i}t^{jk-i}$$

$$\int_x^\infty tf(t)dt = \lambda^{1/k}\frac{\theta^2}{\beta+\theta}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\Gamma(1+\frac{1}{k}+j)}{\Gamma(1+\frac{1}{k})j!}\binom{jk}{i}(\bar{\lambda})^j(\frac{\theta\beta}{\theta+\beta})^{jk-i}$$
$$\times\int_x^\infty(t^{jk-i+1}+\beta t^{jk-i+2})e^{-(1+jk)(\theta t)}dt$$

Using the substitution $u=\theta(j+1)t$

$$\int_x^\infty(t^{jk-i+1}+\beta t^{jk-i+2})e^{-(1+jk)(\theta t)}dt = \frac{\Gamma(jk-i+2,\theta(jk+1)x)}{[\theta(1+jk)]^{jk-i+1}}+\beta\frac{\Gamma(jk-i+3,\theta(jk+1)x)}{[\theta(1+jk)]^{jk-i+2}}$$

Collecting all of the above evaluations and making the necessary simplifications, the MRL can be written as

$$m(x)=\left\{\frac{1-\bar{\lambda}\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}}{\left(\frac{\theta+\beta+\beta\theta x}{\theta+\beta}\right)^k e^{-k\theta x}}\right\}^{1/k}\left(\frac{\theta^2}{\theta+\beta}\right)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{\Gamma(1+\frac{1}{k}+j)}{\Gamma(1+\frac{1}{k})j!}\binom{jk}{i}(\bar{\lambda})^j$$
$$\times(\frac{\theta\beta}{\theta+\beta})^{jk-i}\left[\frac{\Gamma(jk-i+2,\theta(jk+1)x)}{[\theta(1+jk)]^{jk-i+1}}+\beta\frac{\Gamma(jk-i+3,\theta(jk+1)x)}{[\theta(1+jk)]^{jk-i+2}}\right]-x$$

4.5. Estimation of Parameters

Let $x=(x_1,....,x_n)$ be a random sample of size n from the HETLD distribution. Then, the log-likelihood function is given by

$$L(\lambda,k,\beta,\theta)=n\left[\frac{1}{k}log\lambda+2log\theta-log(\beta+\theta)\right]+\sum_{i=1}^{n}log(1+\beta x_i)-\theta\sum_{i=1}^{n}x_i$$
$$-(1+\frac{1}{k})\sum_{i=1}^{n}logA_i(k,\lambda,\beta,\theta)$$

where,

$$A_i(k,\lambda,\beta,\theta)=1-\bar{\lambda}\left(\frac{\theta+\beta+\beta\theta x_i}{\theta+\beta}\right)^k e^{-k\theta x_i},\qquad i=1,....n$$

The MLEs $\bar{\lambda},\bar{k},\bar{\beta},\bar{\theta}$ of λ,k,β,θ are then the solutions of the following non-linear equations.

$$\frac{\partial LogL}{\partial k}=-nk^{-2}log\lambda+\frac{1}{k^2}\sum_{i=1}^{n}logA_i(k,\lambda,\beta,\theta)+(1+\frac{1}{k})\bar{\lambda}\sum_{i=1}^{n}\frac{A_{i,k}(k,\lambda,\beta,\theta)}{A_i(k,\lambda,\beta,\theta)}=0$$

$$\frac{\partial LogL}{\partial\lambda}=\frac{n}{k\lambda}-(1+\frac{1}{k})\sum_{i=1}^{n}\frac{A_{i,\lambda}(k,\lambda,\beta,\theta)}{A_i(k,\lambda,\beta,\theta)}=0$$

$$\frac{\partial LogL}{\partial\beta}=\frac{n}{\beta+\theta}+\sum_{i=1}^{n}\frac{x_i}{1+\beta x_i}+\bar{\lambda}(1+k)\sum_{i=1}^{n}\frac{A_{i,\beta}(k,\lambda,\beta,\theta)}{A_i(k,\lambda,\beta,\theta)}=0$$

$$\frac{\partial LogL}{\partial\theta}=\frac{2n}{\theta}-\frac{n}{\beta+\theta}-\sum_{i=1}^{n}x_i+\frac{\bar{\lambda}(1+k)}{(\beta+\theta)^2}\sum_{i=1}^{n}\frac{A_{i,\theta}(k,\lambda,\beta,\theta)}{A_i(k,\lambda,\beta,\theta)}=0$$

where,

$$A_{i,k}(k,\lambda,\beta,\theta)=\frac{\partial A_i(k,\lambda,\beta,\theta)}{\partial k}=\left(\frac{(\theta+\beta+\beta\theta x_i)e^{-\theta x_i}}{\theta+\beta}\right)^k log\left(\frac{(\theta+\beta+\beta\theta x_i)e^{-\theta x_i}}{\theta+\beta}\right)$$

Table 1: *Wheaton River Data.*

1.7	2.2	14.4	1.1	0.4	20.6	5.3	0.7
1.9	13.0	12.0	9.3	1.4	18.7	8.5	25.5
11.6	14.1	22.1	1.1	2.5	14.4	1.7	37.6
0.6	2.2	39.0	0.3	15.0	11.0	7.3	22.9
1.7	0.1	1.1	0.6	9.0	1.7	7.0	20.1
0.4	2.8	14.1	9.9	10.4	10.7	30.0	3.6
5.6	30.8	13.3	4.2	25.5	3.4	11.9	21.5
27.6	36.4	2.7	64.0	1.5	2.5	27.4	1.0
27.1	20.2	16.8	5.3	9.7	27.5	2.5	27.

Table 2: *Bladder Cancer Patients Data.*

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

$$A_{i,\lambda}(k,\lambda,\beta,\theta) = \frac{\partial A_i(k,\lambda,\beta,\theta)}{\partial\lambda} = \left(\frac{(\theta+\beta+\beta\theta x_i)e^{-\theta x_i}}{\theta+\beta}\right)^k$$

$$A_{i,\beta}(k,\lambda,\beta,\theta) = \frac{\partial A_i(k,\lambda,\beta,\theta)}{\partial\beta} = \left(\frac{(\theta+\beta+\beta\theta x_i)e^{-\theta x_i}}{\theta+\beta}\right)^{k-1}\frac{\theta^2 x_i e^{-\theta x_i}}{(\theta+\beta)^2}$$

$$A_{i,\theta}(k,\lambda,\beta,\theta) = \frac{\partial A_i(k,\lambda,\beta,\theta)}{\partial\theta} = \left(\frac{(\theta+\beta+\beta\theta x_i)e^{-\theta x_i}}{\theta+\beta}\right)^{k-1} x_i e^{-\theta x_i}[-\theta^2-2\beta\theta-\beta\theta^2(\beta+\theta)]$$

Here, it is not possible to find the exact solution of the estimators for $\bar{\lambda}$, $\bar{k}$, $\bar{\beta}$, and $\bar{\theta}$, so the MLEs are obtained numerically by using the appropriate optimization methods.

5. Application

In this section, we illustrate the applicability of HETLD by considering 2 real data sets. The first data set correspond to the exceedances of flood peaks (in m3=s) of the Wheaton River near Carcross in Yukon Territory, Canada. The data consist of 72 exceedances for the years 1958-1984, rounded to one decimal place. They were analysed by Choulakian and Stephens (2001) and are listed in Table 1. The second data set given by Lee and Wang (2003) which represent remission times (in months) of a random sample of 128 bladder cancer patients. Its application in survival analysis has been identified and Table2 lists the remission times of the bladder cancer.

The HETL distribution was compared to four other distributions, namely,

- Harris Extended Weibull (HEW) distribution (Batsidis and Lemonte, 2014), with cumulative distribution function (cdf) $F(x) = 1 - \left(\frac{\lambda e^{-k(\eta x)^\beta}}{1-\bar{\lambda} e^{-k(\eta x)^\beta}}\right)^{1/k}$
- exponentiated Weibull (EW) distribution (Mudholkar and Srivastava, 1993), with $F(x) = \{1 - e^{-(\eta x)^\beta}\}^\alpha$
- TLD distribution, which is the HETLD with $k = 1, \lambda = 1$
- Weibull distribution with $F(x) = 1 - e^{-(\eta x)^\beta}$

The parameters of the above distributions are estimated by the maximum-likelihood method and, then, the values of Akaike information criterion (AIC)and Bayesian information criterion (BIC) are calculated. A summary of computations are given in Table 3 and Table 4. Since the values of the AIC and BIC are smaller for the HETLD compared with those values of the other models,the new distribution seems to be a very competitive model to these data sets. For both data sets, the fitted densities and the empirical cdf plots of the HETLD model are shown in Figures 5 and 6, respectively. The figures indicate a satisfactory fit for the HETLD model too. Thus We conclude that HETLD is the best fit for bladder cancer patients data and wheaton river data.

Table 3: *MLE, maximized log-likelihood, AIC, and BIC for the Wheaton River Data.*

Distribution	MLEs of parameters	-Log L	AIC	BIC
HETLD	$(\hat{\alpha}, \hat{k}, \hat{\beta}, \hat{\theta})$=(0.0417, 7.4455,3.1565,0.1119)	247.8641	503.7282	512.8349
HEW	$(\hat{\alpha}, \hat{k}, \hat{\eta}, \hat{\beta})$ =(5.5322,0.0900,0.5750,0.7844)	250.8359	509.6719	518.7785
EW	$(\hat{\alpha}, \hat{\eta}, \hat{\beta})$=(0.5180,0.0501,1.3879)	251.0251	508.0502	514.8802
TLD	$(\hat{\beta}, \hat{\theta})$=$(7.359766e-05, 8.200532e-02)$	252.128	508.2559	512.8093
Weibull	$(\hat{\eta}, \hat{\beta})$=(0.0859,0.9010)	251.4986	506.9973	511.5506

Table 4: *MLE, maximized log-likelihood, AIC, and BIC for the Bladder Cancer Patients Data.*

Distribution	MLEs of parameters	-Log L	AIC	BIC
HETLD	$(\hat{\alpha}, \hat{k}, \hat{\beta}, \hat{\theta})$=(0.0783,0.8450,1.3110,0.0615)	409.2563	826.514	837.8932
HEW	$(\hat{\alpha}, \hat{k}, \hat{\eta}, \hat{\beta})$ =(8.8429,4.3278,0.1997,0.7319)	409.7029	827.4058	838.814
EW	$(\hat{\alpha}, \hat{\eta}, \hat{\beta})$=(2.7989,0.2993,0.6540)	410.6801	827.3602	835.9163
TLD	$(\hat{\beta}, \hat{\theta})$=(0.000014,0.1068)	414.3419	832.6838	838.3879
Weibull	$(\hat{\eta}, \hat{\beta})$=(0.1046,1.0475)	414.0869	832.1738	837.8778

6. Reliability Test Plan

As a sequel, in this section, we discuss the application of the HETLD in acceptance sampling. It is assumed that the probability distribution of life times follow the HETLD. Acceptance sampling is an inspection procedure used to determine whether to accept or reject a specific quantity of material. The decision law to accept or reject a lot according to the results of a random sample from the population is called acceptance sampling plan . The procedure is to take a random sample of size (n) and inspect each item. If the number of defectives does not exceed a specified acceptance number c, the consumer accepts the entire set of products. Any defectives found in the sample are either repaired or returned to the producer. If the number of defectives in the sample is greater than c, the consumer subjects the entire products to 100 percent inspection or

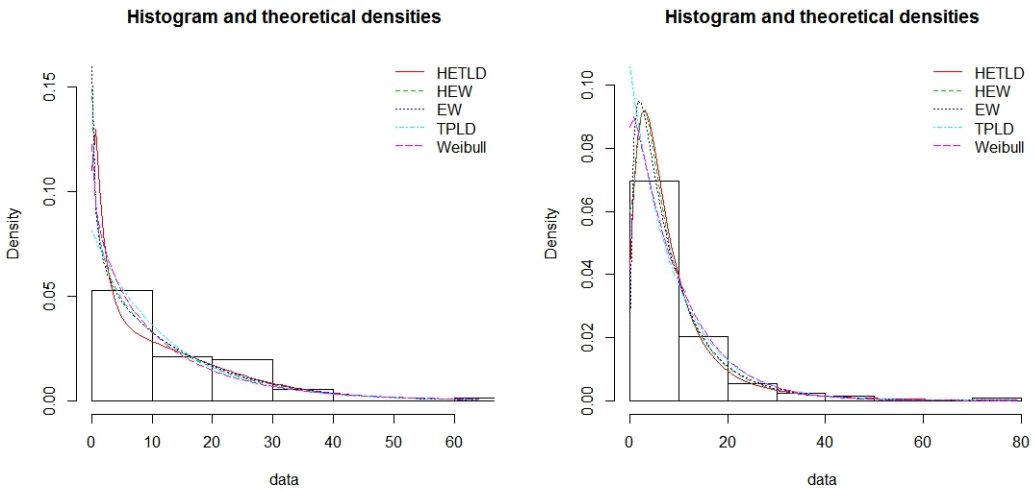


Figure 5: *Fitted densities plots for the first and the second data sets.*

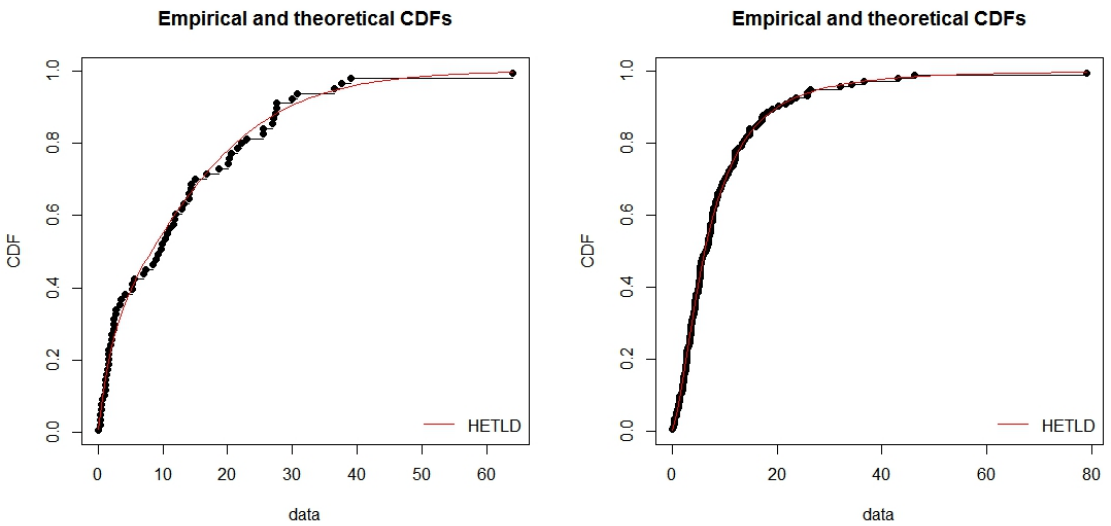


Figure 6: *Empirical cdf plots of HETLD distribution for the first and the second data sets.*

rejects the entire product and returns to the producer. There are two kinds of errors which may arise when a lot is either accepted or rejected. If a good lot is rejected it is called producer's risk and if a bad lot is accepted it is known as consumer's risk. An acceptance sampling plan should be designed in such a way that both risks are minimal. Then the procedure is termed as 'acceptance sampling based on life tests 'or 'reliability test plans'.

Gupta and Groll (1961), Good and Kao (1961), Kantam and Rosaiah (1998), Kantam et al. (2001), Rosaiah and Kantam (2005), Rosaiah et al. (2006), Lio et al. (2010), Krishna et al. (2013) etc. have discussed acceptance sampling plans for various distributions. Jose and Sivadas (2015) developed a reliability test plan for acceptance or rejection of a lot of products submitted for inspection with lifetimes governed by Negative Binomial Marshall-Olkin Rayleigh distribution. Recently, Jose and Joseph (2018) introduced reliability test plan for Gumbel - Uniform life time model. Jose and Paul (2018) derived the reliability test plans for percentiles based on Harris generalized linear exponential distribution. Jose et al. (2018) developed the reliability test plan for Harris extended Weibull distribution.

6.1. The Sampling Plans

In statistical quality control acceptance sampling plan is concerned with the inspection of a sample of products taken from a lot and the decision whether to accept or not to accept the lot based on the quality of product.

Here we discuss the reliability test plan for accepting or rejecting a lot where the life time of the product follows HETLD. In a life testing experiment the procedure is to terminate the test by a pre-determined 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 HETLD. If a scale parameter $\eta > 0$ is introduced then cumulative distribution function (cdf) of HETLD is given by,

$$\bar{G}(t) = \left\{ \frac{\lambda\left(\theta+\beta+\beta\theta(\frac{t}{\eta})\right)^k e^{-k\theta(\frac{t}{\eta})}}{(\theta+\beta)^k - \bar{\lambda}\left(\theta+\beta+\beta\theta(\frac{t}{\eta})\right)^k e^{-k\theta(\frac{t}{\eta})}} \right\}^{1/k} \tag{10}$$

where,$\bar{\lambda} = 1-\lambda$ and $(\lambda, k, \beta, \theta) > 0$are shape parameters and $\eta > 0$ is the scale parameter. The average life time depends only on η if λ, β, θ and k are known. Let η_0 be the required minimum average life time. Then

$$G(t, \lambda, k, \beta, \theta, \eta) \leq G(t, \lambda, k, \beta, \theta, \eta_0) \Leftrightarrow \eta \geq \eta_0.$$

A sampling plan is specified by 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 η_0.

The consumer's risk , i.e. the probability of accepting a bad lot should not exceed the value $1-p^*$, where p^* is a lower bound for the probability that a lot of true value η below η_0 is rejected by the sampling plan. For fixed p^* the sampling plan is characterized by $(n, c, t/\eta_0)$. 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 t/η_0 such that

$$L(p_0) = \sum_{i=0}^{c} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq 1 - p*, \tag{11}$$

where $p_0 = G(t, \lambda, k, \beta, \theta, \eta_0)$. The function $L(p)$ is called operating characteristic function of the sampling plan and it gives, i.e. the acceptance probability of the lot as a function of the failure probability $p(\eta) = G(t, \lambda, k, \beta, \beta, \eta)$. The average life time of the product is increasing with η and therefore the failure probability $p(\eta)$ decreases implying that the operating characteristic function is increasing in η. The minimum values of n satisfying (11) are obtained for $\lambda = 2$, $k = 2$, $\beta = 2$, $\theta = 2$ and $p*$=0.75, 0.90, 0.95, 0.99 and t/η_0=0.40, 0.56, 0.72, 0.88, 1, 1.5, 2 and 2.5. The results are displayed in Table 5. If $p_0 = G(t, \lambda, k, \beta, \theta, \eta_0)$ is very small and n is large, the binomial probability may be approximated by Poisson probability with parameter $\delta = np_0$ so that Equation (11) becomes

$$L_1(p_0) = \sum_{i=0}^{c} \frac{\delta^i}{i!} e^{-\delta} \leq 1 - p*. \tag{12}$$

The minimum values of n satisfying Equation (12) are obtained for the same combination of values of δ, β and t/η_0 for various values of p^* are presented in Table 6. The operating characteristic function of the sampling plan $(n, c, t/\eta_0)$gives the probability L(p) of accepting the lot with

$$L(p) = \sum_{i=0}^{c} \binom{n}{i} p_0^i (1 - p_0)^{n-i} \tag{13}$$

where $p = G(t, \eta)$ is considered as a function of η. For given p^* ,t/η_0 the choice of c and n are made on the basis of operating characteristics. Considering the fact that

$$p = G(\frac{t}{\eta_0} / \frac{\eta_0}{\eta}) \tag{14}$$

values of operating characteristic for a few sampling plans are computed and presented in Table 7.

6.2. Illustration

Assume that the life distribution is HETLD with $\lambda = 2$, $k = 2$, $\beta = 2$, $\theta = 2$. Suppose that the experimenter is interested in establishing that the true unknown average life is at least 1000 hours. Let the consumer's risk is set to be $1 - p^* = .25$. It is desired to stop the experiment at t=560 hrs. Then, for an acceptance number $c = 2$, the required n from Table 5 is 10. So the sampling plan is $(n = 10, c = 2, t/\eta_0 = .56)$. ie; if during 560 hours, not more than 2 failures out of 10 are observed then the experimenter can assert with confidence limit 0.75 that the average life is at least 1000 hours. If we use Poisson approximation to binomial the corresponding value is n=11 for the sampling plan $(n = 11, c = 2, t/\eta_0 = 0.56)$ with the consumer risk 0.25 under the HETLD, the operating characteristic values from Table 7 are,

$\frac{\eta}{\eta_0}$	2	4	6	8	10	12
L(p)	0.7790	0.9638	0.9886	0.9950	0.9974	0.9985

Application 1

Consider the following ordered failure times of the product, gathered from a software development project (Wood, 1996). The data set regarding the software reliability, presented by Wood (1996), was analyzed in the context of acceptance sampling by Rosaiah and Kantam (2005), Rao et al. (2008), Rao et al. (2009a, 2009b), Lio et al. (2010) and Rao and Kantam (2010). This data can be regarded as an ordered sample of size 12 with observations x_i, i=1,2,...12. The observations in the sample are

519, 968, 1430, 1893, 2490, 3058, 3625, 4422, 5218, 5823, 6539, 7083.

Let the specified average life be 1000 hours and the testing time is 560 hours, this leads to the ratio of $t/\eta_0 = 0.56$ with corresponding n and c as 12 and 1 from Table 7.1 for p^* =0.95. Therefore, the sampling plan for above sample data is $(n = 12, c = 1, t/\eta_0 = 0.56)$. We accept the lot only if the number of failures before 560 hours is less than or equal to 1. However, the confidence level is assured by the sampling plan only if the given lifetimes follow the HETLD. We have

Table 5: *Minimum sample size for the specified ratio t/η_0, confidence level p^*, acceptance number c, $\lambda = 2$, $k = 2$, $\beta = 2$, $\theta = 2$ using binomial approximation*

p^*	c	t/η_0							
		0.4	0.56	0.72	0.88	1	1.5	2	2.5
0.75	0	5	4	3	2	2	1	1	1
	1	11	7	5	4	4	3	2	2
	2	15	10	8	6	6	4	3	3
	3	20	14	10	8	7	5	5	4
	4	25	17	13	10	9	7	6	5
	5	29	20	15	12	11	8	7	6
	6	34	23	17	14	13	9	8	7
	7	38	26	20	16	14	10	9	8
	8	43	29	22	18	16	12	10	10
	9	47	32	24	20	18	13	11	11
	10	52	35	27	22	19	14	12	12
0.90	0	9	6	4	3	3	2	1	1
	1	15	10	7	6	5	3	3	2
	2	20	14	10	8	7	5	4	3
	3	26	17	13	10	9	6	5	5
	4	31	21	15	12	11	8	6	6
	5	36	24	18	14	13	9	7	7
	6	41	27	21	17	15	10	8	8
	7	46	31	23	19	17	12	9	9
	8	50	34	26	21	18	13	11	10
	9	55	37	28	23	20	14	12	11
	10	60	41	30	25	22	16	13	12
0.95	0	11	7	5	4	4	2	2	1
	1	18	12	9	7	6	4	3	3
	2	24	16	12	9	8	5	4	4
	3	29	20	14	11	10	7	5	5
	4	35	23	17	14	12	8	7	6
	5	40	27	20	16	14	10	8	7
	6	45	30	23	18	16	11	9	8
	7	50	34	25	20	18	12	10	9
	8	56	37	28	22	20	14	11	10
	9	61	41	30	24	21	15	13	12
	10	66	44	33	26	23	16	14	13
0.99	0	17	11	8	6	5	3	2	2
	1	24	16	12	9	8	5	4	3
	2	31	21	15	12	10	7	5	4
	3	37	25	18	14	12	8	6	6
	4	43	29	21	17	14	10	8	7
	5	49	33	24	19	16	11	9	8
	6	55	36	27	21	19	13	10	9
	7	61	40	30	24	21	14	11	10
	8	66	44	33	26	22	15	13	11
	9	71	48	35	28	24	17	14	12
	10	77	51	38	30	26	18	15	13

Table 6: *Minimum sample size for the specified ratio t/η_0, confidence level p^*, acceptance number c, $\lambda = 2$, $k = 2$, $\beta = 2$, $\theta = 2$ using Poisson approximation*

p^*	c	t/η_0							
		0.4	0.56	0.72	0.88	1	1.5	2	2.5
0.75	0	6	4	3	3	3	2	2	2
	1	11	8	6	5	5	4	3	3
	2	16	11	9	7	7	5	5	4
	3	21	15	11	10	9	6	6	6
	4	26	18	14	12	10	8	7	7
	5	31	21	16	14	12	9	9	8
	6	35	24	19	16	14	11	10	9
	7	40	28	21	18	16	12	11	10
	8	44	31	24	20	18	13	12	11
	9	49	34	26	22	19	15	13	13
	10	53	37	28	24	21	16	15	14
0.90	0	10	7	5	5	4	3	3	3
	1	16	11	9	7	7	5	5	4
	2	22	15	12	10	9	7	6	6
	3	28	19	15	12	11	9	8	7
	4	33	23	18	15	13	10	9	9
	5	38	26	20	17	15	12	11	10
	6	43	30	23	19	17	13	12	11
	7	48	33	26	21	19	15	13	12
	8	53	37	28	24	21	16	15	14
	9	58	40	31	26	23	18	16	15
	10	63	44	34	28	25	19	17	16
0.95	0	13	9	7	6	5	4	4	4
	1	20	14	11	9	8	6	6	5
	2	26	18	14	12	10	8	7	7
	3	32	22	17	14	13	10	9	8
	4	38	26	20	17	15	12	10	10
	5	43	30	23	19	17	13	12	11
	6	48	34	26	21	19	15	13	13
	7	54	37	29	24	21	16	15	14
	8	59	41	31	26	23	18	16	15
	9	64	45	34	28	25	20	18	16
	10	69	48	37	31	27	21	19	18
0.99	0	19	13	10	9	8	6	5	5
	1	27	19	15	12	11	9	8	7
	2	35	24	19	15	14	11	10	9
	3	41	29	22	18	16	13	11	11
	4	47	33	25	21	19	15	13	12
	5	54	37	29	24	21	16	15	14
	6	59	41	32	26	24	18	16	15
	7	65	45	35	29	26	20	18	17
	8	71	49	38	31	28	22	19	18
	9	78	53	41	34	30	23	21	20
	10	82	57	44	36	32	25	22	21

Table 7: *Values of the operating characteristic function of the sampling plan*$(n, c, \frac{t}{\eta_0})$

				$\frac{\eta}{\eta_0}$					
p^*	n	c	$\frac{t}{\eta_0}$	2	4	6	8	10	12
0.75	15	2	0.4	0.7613	0.9568	0.9858	0.9937	0.9967	0.9980
	10	2	0.56	0.7790	0.9638	0.9886	0.9950	0.9974	0.9985
	8	2	0.72	0.7528	0.9607	0.9879	0.9948	0.9973	0.9984
	6	2	0.88	0.7894	0.9700	0.9912	0.9963	0.9981	0.9989
	6	2	1	0.7127	0.9562	0.9870	0.9946	0.9972	0.9984
	4	2	1.5	0.7087	0.9598	0.9890	0.9956	0.9978	0.9988
	3	2	2	0.7493	0.9682	0.9918	0.9969	0.9985	0.9991
	3	2	2.5	0.5895	0.9344	0.9827	0.99346	0.9969	0.9983
0.90	20	2	0.4	0.5992	0.9111	0.9687	0.9857	0.9923	0.9954
	14	2	0.56	0.5850	0.9126	0.9702	0.9866	0.9929	0.9958
	10	2	0.72	0.6188	0.9278	0.9765	0.9897	0.9946	0.9968
	8	2	0.88	0.6159	0.9309	0.9782	0.9906	0.9951	0.9972
	7	2	1	0.6088	0.9315	0.9788	0.9910	0.9954	0.9973
	5	2	1.5	0.5269	0.9167	0.9754	0.9899	0.9949	0.9971
	4	2	2	0.4715	0.9030	0.9723	0.9890	0.9946	0.9970
	3	2	2.5	0.5895	0.9344	0.9827	0.9934	0.9969	0.9983
0.95	24	2	0.4	0.4764	0.8649	0.9499	0.9765	0.9872	0.9922
	16	2	0.56	0.4927	0.8799	0.9574	0.9805	0.9895	0.9937
	12	2	0.72	0.4908	0.8861	0.9610	0.9825	0.9907	0.9945
	9	2	0.88	0.5314	0.9058	0.9693	0.9866	0.9930	0.9959
	8	2	1	0.5094	0.9020	0.9685	0.9863	0.9929	0.9959
	5	2	1.5	0.5269	0.9167	0.9754	0.9899	0.9949	0.9971
	4	2	2	0.4715	0.9030	0.9723	0.9890	0.9946	0.9970
	4	2	2.5	0.2733	0.8172	0.9442	0.9774	0.9890	0.9938
0.99	31	2	0.4	0.3000	0.7699	0.9069	0.9543	0.9744	0.9843
	21	2	0.56	0.3013	0.7837	0.9158	0.9596	0.9777	0.9864
	15	2	0.72	0.3290	0.8114	0.9302	0.9675	0.9824	0.9894
	12	2	0.88	0.3178	0.8140	0.9330	0.9693	0.9836	0.9902
	10	2	1	0.3394	0.8315	0.9413	0.9736	0.9861	0.9918
	7	2	1.5	0.2508	0.7973	0.9315	0.9701	0.9845	0.9910
	5	2	2	0.2661	0.8140	0.9411	0.9754	0.9877	0.9930
	4	2	2.5	0.2733	0.8172	0.9442	0.9774	0.9890	0.9938

correlated the sample quantiles and the corresponding population quantiles to confirm that the given sample is generated by lifetimes following the HETLD and found a satisfactory agreement. Thus, the adoption of the decision rule of the sampling plan seems to be justified.
In the above sample there is only one failure at 519 hours before termination t=560 hours. Hence we accept the product. Here, we may note that termination time t is smaller than that of the sampling plan suggested by Krishna et al. (2013) and Rosaiah and Kantam (2005). Hence, the cost and the experimental time can be saved considerably by using present sampling plans.
Application 2
The second data set is obtained from tests on endurance of deep groove ball bearings (Lawless, 1982). The data are the number of million revolutions before failure for each of the 11 ball bearings in life test and they are,
51.84, 51.96, 54.12, 55.56, 67.80, 68.44, 68.64, 68.88, 84.12, 93.12.
Let the specified average life be 93 hours and the testing time is 53 hours, this leads to the ratio of $t/\eta_0 = 0.56$ with corresponding n and c as 10, 2 from Table 7.1 for p^* =0.75. Therefore, the sampling plan for above sample data is $(n = 10, c = 2, t/\eta_0 = 0.56)$. We accept the lot only if the number of failures before 53 hours is less than or equal to 2. Thus in the above sample of 10 failures there are 2 failures at 51.84, 51.96 before 53 hours, therefore we accept the product.

7. Conclusion

In this article, we consider the HETLD from the Harris extended family and discussed its properties. We derived expansions for the moments, hazard rate function, reversed hazard rate function, cumulative hazard rate function, mean residual lifetime distribution, quantiles and Renyi entropy. The estimation of parameters is approached by the method of maximum likelihood. The applicability and usefulness of the proposed model are presented by using the real data sets. Also a reliability test plan is developed when the lifetimes of the items follow the HETLD. The results are illustrated using two data sets on ordered failure times of products as well as failure times of ball bearings.

References

[1] Abouammoh A.M., Alshangiti A.M., Ragab I.E.(2015), A new generalized Lindley distribution. *Journal of Statistical computation and simulation* **85**, 3662-3678.
[2] Aly E.A.A., Benkherouf L. (2011), A new family of distributions based on probability generating functions, *Sankhya B*, **73**, 70-80.
[3] Batsidis A., Lemonte A.J., (2014), On the Harris extended family of distributions, *Statistics: A Journal of Theoretical and Applied Statistics*, **49**(6), 1400-1421.
[4] Barreto-Souza W., Lemonte A.J. and Cordeiro G.M.(2013), General results for the Marshall-Olkin's family of distributions, *Ann Braz Acad Science*, **85**, 3-21.
[5] Bensid A.E., Zeghdoudi H. (2017),BSG Proceedings, **24**, 1-18.
[6] Choulakian V. and Stephens M.A. (2001), Goodness-of-fit for the generalized Pareto distribution. *Technometrics*, **43**,478-484.
[7] Chesneau C., Tomy L., Gillariose J. (2019a), A new modified Lindley distribution with properties and applications, preprint.
[8] Chesneau C., Tomy L., Jose M. (2019b), Wrapped modified Lindley distribution, Journal of Statistics and Management Systems, Accepted.
[9] Chesneau C., Tomy L., Gillariose J. (2020a), The Inverted modified Lindley distributions, J Stat Theory Pract, 14 (46).
[10] Chesneau C., Tomy L., Gillariose J. (2020b), On a sum and difference of two Lindley distributions: theory and applications, REVSTAT, Accepted.
[11] Cordeiro G.M., Alfy A.Z., Yousof H.M., Cakmakyapan S., Ozel G. (2018), The Lindley Weibull distribution: properties and applications, *Anais da Academia Brasileira de Ciencias*, **90 (3)**, 2579-2598.

[12] Elbatal I., Merovi F., Elgarhy M.(2013), A New generalized Lindley distribution, *Mathematical Theory and Modelling* **3(13)**,30-37.
[13] Ekhosuehi N., Opone F. (2018), A three-parameter generalized Lindley distribution: properties and application, *Statistica*, **78(3)**, 233-249.
[14] Ghitany M., Al-Mutairi D., Balakrishnan N., Al-Enezi I.(2013), Power lindley distribution and associated inference, *computational statistics and data analysis* **6**, 20-33.
[15] Ghitany M.E., Atieh B. and Nadarajah S.(2008) Lindley distribution and its application, *Math Comput Simulat*, **78**, 493-506.
[16] Ghitany M.E., Alqallaf F., Al-Mutairi D.K. and Husain H.A.(2011), A two-parameter Lindley distribution and its applications to survival data, *Math Comput Simulat*, **81**, 1190-1201.
[17] Goode H.P., 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.
[18] Gupta S.S., Groll P.A. (1961), Gamma distribution in acceptance sampling based on life tests, *Journal of the American Statistical Association*, **56**, 942-970.
[19] Harris T.E. (1948), Branching processes, *Annals of Mathematical Statistics*, **19**, 474-494.
[20] Jose K.K., Paul A.(2018), Reliability test plans for percentiles based on Harris generalized linear exponential distribution, *Stochastics and Quality Control*, https://doi.org/10.1515/eqc-2017-0025.
[21] Jose K.K., Tomy L., Thomas S.P. (2018), On a Generalization of the Weibull Distribution and Its Application in Quality Control, *Stochastics and Quality Control*, **33(2)** , 113-124 .
[22] Joshi S., Jose K.K. (2018) Wrapped Lindley distribution, *Communications in Statistics-Theory and Methods*, **47**, 1013-1021.
[23] Jose K.K, Krishna E., (2011), Marshall-Olkin extended uniform distribution, *ProbStat Forum*, **4**, 78-88.
[24] Jose K.K., Sivadas R. (2015), Negative binomial Marshall-Olkin Rayleigh distribution and its applications, *Economic Quality Control*, **30(2)**,89-98.
[25] Jose K.K. and Joseph J. (2018), Reliability test plan for the Gumbel-Uniform distribution, *Stochastics and Quality Control*, https://doi.org/10.1515/eqc-2017-0011.
[26] Jose K.K., Paul A., (2018), Reliability test plans for percentiles based on Harris generalized linear exponential distribution, *Stochastics and Quality Control*, **33**(1), 61-70.
[27] Kantam R.R.L., Rosaiah K., Rao S.G., (2001), Acceptance sampling based on life tests: log-logistic model, *Journal of Applied Statistics*, **28**, 121-128.
[28] Kantam R.R.L., Rosaiah K., (1998), Half logistic distribution in acceptance sampling based on life tests, *IAPQR Transactions*, **23**(2), 117-125.
[29] Krishna E., Jose K.K, Ristic M.M., (2013), Applications of Marshall-Olkin Fréchet distribution, *Communications in Statistics-Simulation and Computation*, **42**, 76-89.
[30] Kemaloglu S.A. and Yilmaz M. (2017), Transmuted two parameter Lindley distribution, *Communications in Statistics - Theory and Methods*, **46(23)**, 11866-11879.
[31] Lawless J.F. (2003), Statistical Models and Methods for Lifetime Data, 2nd ed., JohnWiley and Sons, NewYork.
[32] Lindley D.V. (1958) Fiducial distributions and Bayes' theorem, *J Roy Stat Soc B*, **20**, 102-107.
[33] Lee E. T., 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.
[34] Liyanage G., Parai M. (2014), The generalized power Lindley distribution with its applications, *Asian journal of mathematics and applications* **73**, 331-359.
[35] Lio Y.L., Tsai T.R., Wu S.J., (2010), Acceptance sampling plans from truncated life tests based on the burr type XII percentiles, *Journal of the Chinese Institute of Industrial Engineers*, **27**(4), 270-280.
[36] Marshall A.W., Olkin I. (1997), A new method for adding a parameter to a family of distributions with Application to the Exponential and Weibull Families, *Biometrica*, **84**, 641-652.

[37] Mazucheli J., Achcar J.A. (2011), The Lindley distribution applied to competing risks lifetime data, *Comput Meth Prog Bio*, **104**, 188-192.
[38] Merovci F.(2013), Transmuted Lindly Distribution, *International Journal of Open Problems in Computer Science and Mathematics* **6**, 63-72.
[39] Mazucheli J., Achcar J.A., (2011), The Lindley distribution applied to competing risks lifetime data, *Computer Methods and Programs in Biomedicine*, **104**, 188-192.
[40] Mudholkar G.S., Srivastava D.K., (1993), Exponentiated Weibull family for analyzing bathtub failure rate data, IEEE Transactions on Reliability, **42**, 299-302.
[41] Nadarajah S., Bakouch H. S., Tahmasbi R. (2011), A generalized Lindley distribution, *Sankhya B*, **73(2)**, 331-359.
[42] Oluyede B.O., Yang T.(2014), A New class of generalized Lindley distribution with its applications, *Journal of Statistical computation and simulation* **85(10)**, 2072-2100.
[43] Pinho L.G.B., Cordeiro G.M., Nobre J.S. (2015), The Harris extended exponential distribution, *Communications in Statistics-Theory and methods*, **44**, 3486-3502.
[44] Rao G.S., Ghitany M.E., Kantam R.R.L., (2008), Acceptance sampling plans for Marshall-Olkin extended Lomax distribution, *International Journal of Applied Mathematics*, **21**, 315-325.
[45] Rao G.S., Ghitany M.E., Kantam R.R.L., (2009a), Marshall Olkin extended Lomax distribution: An economic reliability test plan, *International Journal of Applied Mathematics*, **22**, 139-148.
[46] Rao G.S., Ghitany M.E., Kantam R.R.L., (2009b), Reliability test plans for Marshall Olkin extended exponential distribution, *Applied Mathematical Sciences*, **3**, 2745-2755.
[47] Rao G.S., Kantam R.R.L., (2010), Acceptance sampling plans from truncated life tests based on the log-logistic distribution for percentiles, *Economic Quality Control*, **25**, 153-167.
[48] Rosaiah K., Kantam R.R.L., (2005), Acceptance sampling based on the inverse Rayleigh distribution, *Economic Quality Control*, **20**, 277-286.
[49] Rosaiah K., Kantam R.R.L., Kumar C.S, (2006), Reliability test plans for exponentiated log-logistic distribution, *Economic Quality Control*, **21**(2), 165-175.
[50] Shanker R., Sharma S., Shanker U. and Shanker R. (2013), A Two-Parameter Lindley distribution for modeling waiting and survival times data, *Applied Mathematics*, **4**, 363-368.
[51] Shanker R., Fesshaye H.(2015), On Poisson-Lindley Distribution and Its Applications to Biological Sciences, *Biometrics and Biostatistics International Journal* **2(4)**,1-5.
[52] Wood A., (1996), Predicting software reliability, *IEEE Transactions on Software Engineering*, **22**, 69-77.
[53] Zakerzadeh Y., Dolati A. (2009), Generalized Lindley distribution, *Journal of Mathematical Extension*, **3**, 13-25.

RT&A, No 3 (63)
Volume 16, September 2021

AN OPTIMUM RESOURCE SCORE ESTIMATION METHOD USING BIPARTITE GRAPH MODEL AND SINGLE NODE SYSTEMATIC SAMPLING

DEEPIKA RAJORIYA AND D. SHUKLA

•

Dr. Harisingh Gour University, Sagar (M.P.) 470003
deepikarajoriya2112@gmail.com, diwakarshukla@rediffmail.com

Abstract

Consider a graphical population of vertices (nodes) and edges, where edges are connected with vertices to form a Bipartite graph. A complete Bipartite graph has vertices that can be partitioned into two subsets such that no edge has both endpoints in the same subset, and every possible edge connected to vertices in different subsets is a part of graph. In real life, there may hundreds of cities where at least one possible way exists reaching source to destinations. Several tourist places and small towns are the examples where the road transportation is available between origin and destination and these roads constitute Bipartite graph when they are like edges. The travel needs resource consumption who could be measured through resource-score. Walking at the hill station needs more energy consumption than at the plane area. This paper suggests an example to estimate the resource consumption by the values of score. Further, paper proposes a sample based methodology for calculating the average resource consumption between a pair of small town (city) and tourist place. Bipartite graph is used as a model tool. A single-node systematic sampling procedure is proposed under the Bipartite graph setup which is found useful for solution. The suggested estimation strategy is optimum at specific choice of parametric values. For quick selection, ready-reckoner tables are prepared who provide immediate optimum choice of constant. Results are numerically supported by the empirical study and proved by the calculation of confidence intervals.

Keywords: Graph; Bipartite Graph; Estimator; Bias; Mean Squared Error (MSE); Optimum Choice, Confidence intervals (C.I.).

1. Introduction

Traveling from one place to another, for pleasure and entertainment, is an integral part of human life. It is a diversified international phenomenon essential for recreation, interaction, exchange of thoughts and participation in events. Travel agencies offer world tour plans and holiday packages with maximum facilities at minimum cost. Private and public sector organizations, around the world, offer special leave to their employees for tour and travel event including financial support. For economists and scientists challenges are to get optimum estimates of traveling cost using appropriate models between any pair of tourist places. It is rather more difficult to pick up significant cost variables to incorporate in a cost model. Travel by road, by train or by flight need different types of resource consumption like time, fuel, energy, money etc. In a contribution of Moons[2], some equations for computing travel cost are suggested. More explicitly, travel cost methods are for estimation of economic use values associated with ecosystem or site who are frequently used for recreation. Such are important for valuation of economic benefits, useful in planning and decision making (see Ecosystem Valuation [3]) like :

(a) change in access cost of a recreational location,
(b) elimination of an existing tourist site from list,
(c) addition of new sites in tour plan lists.

The logical basis of a travel cost methods is cost and time, incurred to visit a site. This is analogous to estimating desire of people to pay for money to purchase goods based on quantities and prices (see Ecosystem Valuation[3]). Computing methodologies for travel expense, as suggested in [2] are 1 to 5 listed below:

1. $Travel\ Time = \sum_{i=1}^{8} \frac{distance(i)}{averagespeed(i)}$
2. $Fuel\ Cost = \frac{fuel\ cost(EURO/car\ per\ km) * distance(km)}{number\ of\ passengers}$
3. $Total\ car\ usage\ costs = \frac{total\ car\ usage\ cost(EURO/car\ pe\ rkm) * distance(km)}{number\ of\ passenger}$
4. $Total\ calculated\ costs = fuel\ cost + travel\ cost * 3$
5. $Total\ calulated\ cost\ = Total\ car\ usage\ cost + Trave\ time * 3.$

While travel by road, distinction may be among four wheelers, for example small car, gas car, petrol car and diesel car. Table 1.1 presents the calculated costs derived from [2].

Table 1.1 Car Type and Cost (see [2])

S. No.	Car Type	Small gasoline	Big gasoline	Small diesel	Big diesel
1.	Cylinder capacity(in cm^3)	< 1600	≥ 1600	< 2000	≥ 2000
2.	Fuel consumption(liter/km)	0.0716	0.0771	0.0534	0.0758
3.	Price of fuel(EURO/liter)	0.89	0.89	0.61	0.61
4.	Fuel cost(EURO/liter)	0.064	0.068	0.033	0.046
5.	Total car-usage cost(EURO/liter)	0.33	0.52	0.25	0.38

. Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

2. Mathematical Setup for Travel Cost

Let A, B, C, D are four locations at distances Φ_1 and Φ_2 scattered apart geographically. While reaching from A to B, the resource consumption is Φ_1' whereas from C to D it is Φ_2'. Figure 2.1 reveals graphical relationship to travel from origin A and C to destinations B and D with distances and resource consumptions.

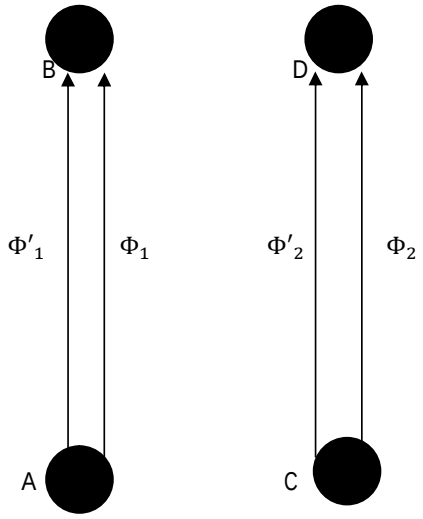


Fig 2.1 Resource consumption and distance for travel

Average resource consumption is $\bar{\Phi} = \frac{\Phi_1+\Phi_2}{2}$ which is an unknown quantity but useful for those who work as travel advisors. Government employees, in many countries, are entitled for availing the Leave Travel Concession (LTC) repeatedly in a block period of few years. They need to prepare and place an estimate of the likely expenditure to the government department for prior sanction of the advance money before the start of journey. This requires calculation of resource consumption in different segments of likely expenditure relating to travel plan. Table 2.1 presents an example of calculation of resource scores.

Table 2.1 Resource Score Traveling Plan from A to B and C to D

Segments	A to B Resource Score(0-25)	C to D Resource Score(0-25)
Fare(Bus/Train/Air)	x_{11}	x_{12}
Food	x_{21}	x_{22}
Boarding/Loading	x_{31}	x_{32}
Local Convenience	x_{41}	x_{42}
Energy Consumption	x_{51}	x_{52}
Time Consumption	x_{61}	x_{62}

Class-intervals for resource scores are as per perception like very low (0-5), low(5-10), medium(10-15), high(15-20), very high(20-25). Energy and time consumption in walking at hill station are higher than at cities located in plain area. Let x_{ij} be the resource score of i^{th} segment of j^{th} travel plan $(i = 1, 2, 3, 4, 5, 6, j = 1, 2)$ as per table 2.1. Average resource score of j^{th} plan is:

$$\bar{\Phi}'_j = \frac{\sum_{i=1}^{6} x_{ij}}{6}$$

In this study $\bar{\Phi}'_j$ is taken as a variable of main interest for different travel plans and packages and needs to be estimated regularly on sample basis over time domain. While travel plans are large in number, $\bar{\Phi}'_j$ are useful unknown parameters for travel agents to prepare the estimate of travel advance money for employees of private or public sectors where tour packages are part of privilege of the annual salary.

3. Graphical Model for $\bar{\Phi}_j$ Estimation

Let there are two groups of vertices, each having well connected edges. Every vertex of first group is connected to all vertices of other group once and only once. This constitutes a Bipartite graph and the mean-edge-length estimation using sampling techniques can be used with graph as a model tool for obtaining solution.

To extend further, assume that first group has even number of vertices, divided into two

subgroups such that each contains equal number of vertices. First subgroup is of main interest while second contains well known correlated information, for example distances between cities. Figure 2.1 is a Bipartite graph structure useful as a model for average travel cost estimation problem considered in this paper.

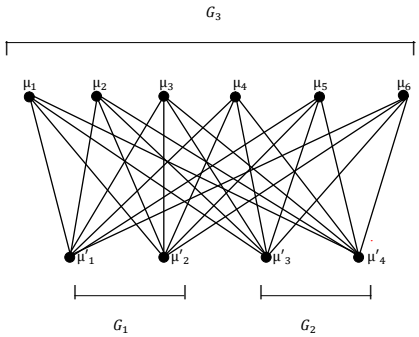


Fig 3.1 Bipartite Type Graph Structure

Remark 1. If the vertices of any graph G (say) can be split into two disjoint subsets V_1 and V_2 in such a way that each edge in the graph joins a vertex in V_1 to a vertex in V_2, and there are no edge in graph G that connect two vertices in V_1 or two vertices in V_2, then the graph is called Bipartite graph. For example, every tree structure is a bipartite graph (see [1]).

Remark 2. Suppose a graph G contains x vertices in the first set and y vertices in the second set . A complete Bipartite graph is a graph in which each vertex in the first set is joined with every single vertex in the second set (see fig 3.1 and [8] [14] [15][18] [19]).

3.1. Example of Bipartite Graph Application (as per fig 3.1)

Let there are six tourist places (destinations) and four origins. For each origin, the travel connectivity exists to all tourist places. Four origins form the first group and six tourist places constitute the second group. First group is further divided into two subgroups, each of two vertices (origins). Edge lengths of first subgroup reveal the total resources likely to consume to travel from origin to the tourist places. Second subgroup is a prototype of the first subgroup where edges represent travel distances who are well known through the travel booklets. Larger distance traveled leads to higher consumption of resources and corresponding requirement of more travel cost. The focus of this study is to know about what an average amount of resource consumption likely to occur while traveling between any pair of origins and tourist places.

Table 3.1 Node-Edge Matrix for Bipartite Graph (Fig 3.1)

	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	*row total*
μ'_1	1	1	1	1	1	1	6
μ'_2	1	1	1	1	1	1	6
μ'_3	1	1	1	1	1	1	6
μ'_4	1	1	1	1	1	1	6
							36

Table 3.1 displays the edge connectivity with vertices of groups where unit number denotes existence of an edge.

3.2. General Bipartite Graph Model

Let k origins (vertices) of even number denoted as $\{\mu'_1, \mu'_2, \mu'_3, ...\mu'_k\}$ divided into groups G_1 and G_2 each strictly of size $t = \frac{k}{2}$. The G_1 contains half of vertices (denoted as cities) while G_2 has remaining who are prototype of G_1(like taxi-stand of the same city in G_1).

$$G_1 : \{\mu'_1, \mu'_2, \mu'_3, ...\mu'_t\} \tag{1}$$

$$G_2 : \{\mu'_{t+1}, \mu'_{t+2}, \mu'_{t+3}, ...\mu'_k\} \tag{2}$$

Further, let there exist r tourist places (vertices) marked as third group G_3 .

$$G_3 : \{\mu_1, \mu_2, \mu_3, ...\mu_r\} \tag{3}$$

Each vertex of G_1 and G_2 is connected to every vertex of G_3 one and only once which constitutes a model like a Bipartite graph among groups G_1, G_2 and G_3. Similarity of identities exist in sequential pair $\{\mu'_1, \mu'_{t+1}\}$, $\{\mu'_2, \mu'_{t+2}\}$, $\{\mu'_3, \mu'_{t+3}\}$,...$\{\mu'_t, \mu'_k\}$ just as same origins or same objects or same organizations.
Assume ϵ_{ij} denotes edge with weights connecting to i^{th} vertex of G_1 to the j^{th} vertex of G_3 and ϵ'_{ij} the edge with weights of i^{th} vertex of G_2 to the j^{th} vertex of G_3 $(i = 1, 2, 3, ..., t, j = 1, 2, 3, ..., r)$. Further, k and r both are large integers and $r > k$ holds for an even k. Moreover, ϵ_{ij} represents weights as resource consumption score (unknown and to estimate) while ϵ'_{ij} are the known weights in advance like road distances. The problem undertaken in this paper is to estimate the average amount of resource consumption likely to occur between any pair of vertices of G_1 and G_3 with the help of edge weights prototype pair of vertices of G_2 and G_3 where information are priorly known.

Table 3.2 General Node-Edge Matrix

	Group (G_3)					
	Node	μ_1	μ_2	μ_3	--------- μ_r	Total
	μ'_1	1	1	1	1	r
	μ'_2	1	1	1	1	r
Group	μ'_3	1	1	1	1	r
(G_1)	⋮					
	μ'_t	1	1	1	1	r
	μ'_{t+1}	1	1	1	1	r
Group	μ'_{t+2}	1	1	1	1	r
(G_2)	μ'_{t+3}	1	1	1	1	r
	⋮					
	μ'_k	1	1	1	1	r

In table 3.2, if an edge exists between $(i, j)^{th}$ pair then $\epsilon_{ij} = \epsilon'_{ij} = 1$ else zero everywhere.

4. Single-Node-Systematic Random Sampling

A Single-Node Systematic Sampling scheme without replacement is proposed as under:

Step I For given r, even k, $t = \frac{k}{2}$, groups G_1, G_2, G_3, large k and r in population, construct population matrix as stated in table 3.2. Vertex-edge connectivity structure constitutes a model like a Bipartite graph.

Step II Assign labels to vertices in G_1 and G_2 like $1, 2, 3, ..., t, t+1, t+2, t+3, ...k$.

Step III Draw one vertex randomly from label $1, 2, 3, ..., t$ (table 3.2). Assume it is l^{th} vertex $(l = 1, 2, 3, ..., t)$ from G_1. Using systematic sampling concept, the unit labeled $(l + t)$ is automatically selected in the sample from G_2. The resultant is single node (vertex) systematic random selection of units (μ'_l, μ'_{l+t}) from G_1 and G_2, each having r edges.

Step IV Note down weights, marked as edge-lengths, of μ'_l and μ'_{l+t} connecting to all vertices of G_3. Sampled weight values of r edges are as under:

$$\mu'_l : (s\epsilon_{lj},\ j = 1, 2, 3, ..., r)\ sample\ from\ G_1,\ G_3 \tag{4}$$

$$\mu'_{l+t} : (s\epsilon'_{l+t},\ j = 1, 2, 3, ..., r)\ sample\ from\ G_2,\ G_3 \tag{5}$$

Step V Apply an appropriate estimation procedure to obtain the estimate of unknown parameter (average resource consumption score).

5. Estimation

For a large Bipartite graph, the population parameters (means) are:

$$\bar{\Delta_1} = \frac{1}{rt}\sum_{i=1}^{t}\sum_{j=1}^{r}\epsilon_{ij} \tag{6}$$

$$\bar{\Delta_2} = \frac{1}{rt}\sum_{i=1}^{t}\sum_{j=1}^{r}\epsilon'_{ij} \tag{7}$$

Using step I to step V of Single-Node-Systematic Sampling, the two sample means are:

$$\bar{\delta_1} = \frac{1}{r}\sum_{j=1}^{r}s\epsilon_{ij} \tag{8}$$

$$\bar{\delta_2} = \frac{1}{r}\sum_{j=1}^{r}s\epsilon'_{ij} \tag{9}$$

where $s\epsilon_{ij}$ and $s\epsilon'_{ij}$ denote weights like edge-lengths appeared in sample. For very small numbers h_1, h_2, $|h_1| < 1$, $|h_2| < 1$, one can use large sample approximations discussed in [5],[6],[7].

$$\bar{\delta_1} = \Delta_1(1 + h_1)\ \ \bar{\delta_2} = \Delta_2(1 + h_2)\ \ with\ \ E(h_1) = E(h_2) = 0 \tag{10}$$

$$E(h_1^2) = \frac{t-1}{rt}(C^*\epsilon)^2,\ \ E(h_2^2) = \frac{t-1}{rt}(C^*\epsilon')^2,\ E(h_1h_2) = \frac{t-1}{rt}\rho(C^*\epsilon)(C^*\epsilon') \tag{11}$$

$$(C^*\epsilon) = \frac{(S\epsilon)}{\bar{\Delta_1}} \tag{12}$$

$$(C^*\epsilon') = \frac{(S\epsilon')}{\bar{\Delta_2}} \tag{13}$$

$$(C^*\epsilon\epsilon') = \frac{(S\epsilon\epsilon')}{(\bar{\Delta_1}.\bar{\Delta_2})} \tag{14}$$

$$(S\epsilon)^2 = \frac{1}{rt-1}\sum_{i=1}^{t}\sum_{j=1}^{r}(\epsilon_{ij} - \bar{\Delta_1})^2 \tag{15}$$

$$(S\epsilon')^2 = \frac{1}{rt-1}\sum_{i=1}^{t}\sum_{j=1}^{r}(\epsilon'_{ij} - \bar{\Delta_2})^2 \tag{16}$$

$$(S\epsilon\epsilon') = \frac{1}{rt-1}\sum_{i=1}^{t}\sum_{j=1}^{r}(\epsilon_{ij}-\bar{\Delta_1})(\epsilon'_{ij}-\bar{\Delta_2}) \tag{17}$$

$$\rho = \frac{S\epsilon\epsilon'}{[(S\epsilon)(S\epsilon')]} \tag{18}$$

Equation (5.13) is correlation coefficient between ϵ_{ij} and ϵ'_{ij} in Bipartite population.

5.1. Proposed method of estimation

Deriving motivation from [9],[10], [12], [14], [16], [17] and [18], estimation strategy E is proposed as under:

$$E = \bar{\delta_1}\left[f_1(\bar{\Delta_2},\bar{\delta_2})\right]\left[f_2(\bar{\Delta_2},\bar{\delta_2})\right]^{-1} \tag{19}$$

where

$$f_1(\bar{\Delta_2},\bar{\delta_2}) = \left[(A+C)\bar{\Delta_2} + gB\bar{\delta_2}\right]$$

$$f_2(\bar{\Delta_2},\bar{\delta_2}) = \left[(A+gB)\bar{\Delta_2} + C\bar{\delta_2}\right]$$

and $A = (P-1)(P-2)$; $B = (P-1)(P-4)(P-5)$; $C = (P-2)(P-3)(P-4)$; $g = \frac{r}{rt} = \frac{1}{t}$; $0 < P < \infty$.

The proposed method of estimation is in accordance with Shukla et al. [9], [10],[12] with change in the term B which is now cubic in this paper but was considered quadratic in earlier contributions.

Theorem 1. Proposed method (5.14) could be expressed as
$E = \bar{\Delta_1}\left[(1+h_1) + \frac{(gB-C)}{(A+gB+C)}\left\{h_2 + h_1h_2 - \frac{Ch_2^2}{A+gB+C}\right\}\right]$
Proof. $E = \bar{\delta_1}\left[f_1(\bar{\Delta_2},\bar{\delta_2})\right]\left[f_2(\bar{\Delta_2},\bar{\delta_2})\right]^{-1}$

Using (5.5), $|h_1| < 1$, $|h_2| < 1$

$$f_1(\bar{\Delta_2},\bar{\delta_2}) = [(A+C)\bar{\Delta_2} + gB\Delta_2(1+h_2)] \tag{20}$$

$$f_2(\bar{\Delta_2},\bar{\delta_2}) = [(A+gB)\bar{\Delta_2} + C\Delta_2(1+h_2)] \tag{21}$$

The (5.15) expressed using (5.5) as

$$f_1(\bar{\Delta_2},\bar{\delta_2}) = \bar{\Delta_2}\left[(A+gB+C) + gB(h_2)\right] = \bar{\Delta_2}\left[(A+gB+C)\right]\left[1 + \frac{gBh_2}{(A+gB+C)}\right] \tag{22}$$

Since $|h_2| < 1$, therefore $|\frac{gBh_2}{(A+gB+C)}| < 1$, due to $\forall\ g > 0,\ P > 0$
For (5.16), the expansion of $(1+x)^{-1}$ is used when $|x| < 1$ as under:

$$\left[f_2(\bar{\Delta_2},\bar{\delta_2})\right]^{-1} = (\bar{\Delta_2})^{-1}\left[(A+gB+C) + Ch_2\right]^{-1}$$

$$= (\bar{\Delta_2})^{-1}\left[(A+gB+C)\right]^{-1}\left[1 + \frac{Ch_2}{(A+gB+C)}\right]^{-1}$$

$$Then, E = \bar{\Delta_1}(1+h_1)\left[1+\frac{gBh_2}{(A+gB+C)}\right]\left[1+\frac{Ch_2}{(A+gB+C)}\right]^{-1}$$

$$= \bar{\Delta_1}(1+h_1)\left[1+\frac{gBh_2}{(A+gB+C)}\right]\left[1-\frac{Ch_2}{(A+gB+C)}+\frac{C^2h_2^2}{(A+gB+C)^2}-\frac{C^3h_2^3}{(A+gB+C)^3}\cdots\right]$$

$$= \bar{\Delta_1}(1+h_1)\left[1-\frac{Ch_2}{(A+gB+C)}+\frac{C^2h_2^2}{(A+gB+C)^2}-\frac{C^3h_2^3}{(A+gB+C)^3}\cdots\right]$$
$$+\bar{\Delta_1}(1+h_1)\left[\left\{\frac{gBh_2}{(A+gB+C)}-\frac{gBCh_2^2}{(A+gB+C)^2}+\frac{gBC^2h_2^3}{(A+gB+C)^3}\cdots\right\}\right]$$

The terms $[(h_1)^u(h_2)^v]$ could be ignored for $(u+v) > 2$, $u,v = 0,1,2,3,4,5...$ because of their low impact in ultimate estimate of parameter due to $|h_1| < 1$, $|h_2| < 1$, and $\left[\frac{1}{(A+gB+C)}\right] < 1$ (see [9] [10] [12])

Then above reduces to:

$$= \bar{\Delta_1}\left[(1+h_1)+\frac{(gB-C)h_2}{(A+gB+C)}-\frac{(gB-C)Ch_2^2}{(A+gB+C)^2}-\frac{(gB-C)h_1h_2}{(A+gB+C)}\right] \tag{23}$$

$$= \bar{\Delta_1}\left[(1+h_1)+\frac{(gB-C)}{(A+gB+C)}\left\{h_2+h_1h_2-\frac{Ch_2^2}{(A+gB+C)}\right\}\right] \tag{24}$$

■

Remark 3. Define Z = $[\frac{(gB-C)}{(A+gB+C)}]$.

Theorem 2. Using theorem 5.1, the bias of the proposed method is
$Bias(E) = B(E) = \bar{\Delta_1}(\frac{t-1}{rt})\left[\rho(C\epsilon)(C\epsilon') - \frac{C}{(A+gB+C)}(C\epsilon')^2\right]$
where ρ denotes correlation coefficients between ϵ_i and ϵ_i' at the Bipartite graph population level.

Proof. Let $E(.)$ denotes the expected value and $E(h_1) = 0$; $E(h_2) = 0$ (as in theorem 5.1 and using equation (5.5));

$Bias(E) = B(E) = E[E - \bar{\Delta_1}]$ (see [5],[6],[7])
$= E\left[\bar{\Delta_1}\left\{(1+h_1)+Z\left(h_2+h_1h_2-\frac{Ch_2^2}{(A+gB+C)}\right)\right\}-\bar{\Delta_1}\right]$ theorem 5.1, remark 5.1
$= E\left[\bar{\Delta_1}\left\{h_1+Z\left(h_2+h_1h_2-\frac{Ch_2^2}{(A+gB+C)}\right)\right\}\right]$
$= \bar{\Delta_1}\left[E(h_1)+Z\left\{E(h_2)+E(h_1h_2)-\frac{CE(h_2^2)}{(A+gB+C)}\right\}\right]$ [using (5.5), (5.6)]
$= \bar{\Delta_1}\left[Z\left\{E(h_1h_2)-\frac{CE(h_2^2)}{(A+gB+C)}\right\}\right]$
$Bias(E) = \bar{\Delta_1}Z(\frac{t-1}{rt})\left[\rho(C\epsilon)(C\epsilon') - \frac{C}{(A+gB+C)}(C\epsilon')^2\right]$ [using (5.5), (5.6)] ■

Theorem 3. The proposed methodology is almost unbiased for a given pair of (g,R) when following equation is satisfied at suitable choice of P
$R(A+gB)+C(R-1) = 0$, *where* $R = \frac{\rho(C\epsilon)}{(C\epsilon')}$, $g = \frac{1}{t}$

Proof. $Bias(E) = 0 \implies \left[\rho(C\epsilon)(C\epsilon') - \frac{C(C\epsilon')^2}{(A+gB+C)}\right] = 0$

For given (g,R) , re-arranging above, one can get equation

$R(A+gB)+C(R-1) = 0$ ■

Theorem 4. For a given pair of (g, R), on maximum of three values of P, the proposed strategy E is almost unbiased.
Proof. Condition of unbiasedness is $R(A + gB) + C(R - 1) = 0$

This could be expressed in the power of P as:

$$(R-1)P^3 + [R + gR + g]\,P^2 + [23R + 299R - 26]\,P - [22R + 20gR - 24] = 0 \tag{25}$$

Above equation is of degree three in P and has maximum of three roots at which E is almost unbiased. ■
Corollary 1. Few of maximum three roots of (5.20) may imaginary or overlapping depending upon given (g, R).
Corollary 2. The best choice of P, among maximum of three, is that bearing lowest mean squared error.
Theorem 5. The mean squared error of E under (5.5) to (5.13) and using (5.20) is

$MSE[E] = (\bar{\Delta_1})^2(\frac{t-1}{rt})\left[(C^*\epsilon)^2 + Z^2(C^*\epsilon')^2 + 2Z\rho(C^*\epsilon)(C^*\epsilon')\right]$
Proof. $MSE[E] = E[E - \bar{\Delta_1}]^2$ (see [], [], [])
$= E\left[\bar{\Delta_1}\left\{(1 + h_1) + \frac{(gB-C)h_2}{(A+gB+C)} + ...\right\} - \bar{\Delta_1}\right]^2$ using theorem 5.1 and ignoring terms $[(h_1)^u(h_2)^v]$, $u, v = 0, 1, 2, 3, 4$ of higher order $(u + v) > 2$ as adopted in theorem 5.1

$= (\bar{\Delta_1})^2 E\left[h_1 + Zh_2\right]^2$
$= (\bar{\Delta})^2\left[E(h_1^2) + (Z)^2E(h_2^2) + (2Z)E(h_1h_2)\right]$

$MSE[E] = (\bar{\Delta})^2(\frac{t-1}{rt})\left[(C^*\epsilon)^2 + Z^2(C^*\epsilon')^2 + 2Z\rho(C^*\bar{\epsilon})(C^*\bar{\epsilon'})\right]$ using (5.6). ■
Theorem 6. For a given pair of (g, R), the minimum mean squared error occurs at suitable choice of P when $RA + (R + 1)gB + (R - 1)C = 0$ is satisfied.
Proof. Differentiating $MSE[E]$, in theorem 5.5, with respect to the term Z and equating to zero, one can get,

$$2Z = -2\left[\frac{\rho(C^*\epsilon)}{(C^*\epsilon')}\right] = -2R$$

$$\implies (gB - C) + R(A + gB + C) = 0$$

$$\implies RA + (R + 1)gB + (R - 1)C = 0 \tag{26}$$

■
Theorem 7. For given (g, R), the maximum three values of P exist at which MSE[E] is minimum.
Proof. The condition for minimum MSE is $RA + (R + 1)gB + (R - 1)C = 0$
which could be expressed in the power of P as:

$$P^3[R(g+1) + (g-1)] - P^2[2R(5g+4) + (10g-9)] + P[R(29g+23) + (29g-26)] - 2[R(10g+11) + 2(5g-6)] = 0$$

Above equation is of degree three in P, therefore, has at most the three roots. ■
Corollary 3. One can get three values of P, using theorem 5.7, having the minimum MSE. The best P is that having the lowest bias.

6. Best Selection Procedure of P for Almost Optimum MSE Strategy

Step I Find P for given pair of (g, R) producing almost unbiased strategy.

Step II Plot the range of P over the X-axis obtained by step I.

Step III Find P for given pair of (g, R) producing minimum MSE.

Step IV Plot the range of P over X-axis under step III.

Step V The overlapping range is the best choice of P for efficient estimation.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

7. Empirical Study

Consider six tourist places $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)$, two remote small towns (cities) (μ_1', μ_2'). The μ_3' is prototype of μ_1' and μ_4' is prototype of μ_2' like taxi stands of μ_1' and μ_2' cities respectively. The weight of edges ϵ_{ij}' are assumed known like distances while weight of edges ϵ_{ij} are resource consumption scores likely during traveling by road and unknown.

Table 7.1 Population

Tourist Places		μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
	μ_1'	$\epsilon_{11} = 09$	$\epsilon_{12} = 13$	$\epsilon_{13} = 21$	$\epsilon_{14} = 15$	$\epsilon_{15} = 14$	$\epsilon_{16} = 17$
Cities and	μ_2'	$\epsilon_{21} = 15$	$\epsilon_{22} = 07$	$\epsilon_{23} = 16$	$\epsilon_{24} = 19$	$\epsilon_{25} = 11$	$\epsilon_{26} = 14$
Prototype	μ_3'	$\epsilon_{31}' = 04$	$\epsilon_{32}' = 06$	$\epsilon_{33}' = 11$	$\epsilon_{34}' = 08$	$\epsilon_{35}' = 07$	$\epsilon_{36}' = 09$
	μ_4'	$\epsilon_{41}' = 07$	$\epsilon_{42}' = 03$	$\epsilon_{43}' = 08$	$\epsilon_{44}' = 09$	$\epsilon_{45}' = 05$	$\epsilon_{46}' = 06$

Table 7.3 First Sample (Random selection of μ_1')

Tourist Places		μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Cities and	μ_1'	$\epsilon_{11} = 09$	$\epsilon_{12} = 13$	$\epsilon_{13} = 21$	$\epsilon_{14} = 15$	$\epsilon_{15} = 14$	$\epsilon_{16} = 17$
Prototype	μ_3'	$\epsilon_{31}' = 04$	$\epsilon_{32}' = 06$	$\epsilon_{33}' = 11$	$\epsilon_{34}' = 08$	$\epsilon_{35}' = 07$	$\epsilon_{36}' = 09$

Table 7.4 Second Sample (Random selection of μ_2')

Tourist Places		μ_1	μ_2	μ_3	μ_4	μ_5	μ_6
Cities and	μ_2'	$\epsilon_{21} = 15$	$\epsilon_{22} = 07$	$\epsilon_{23} = 16$	$\epsilon_{24} = 19$	$\epsilon_{25} = 11$	$\epsilon_{26} = 14$
Prototype	μ_4'	$\epsilon_{41}' = 07$	$\epsilon_{42}' = 03$	$\epsilon_{43}' = 08$	$\epsilon_{44}' = 09$	$\epsilon_{45}' = 05$	$\epsilon_{46}' = 06$

Table 7.5 Bipartite Graph Population Parameters [table 7.1]

S.No.	Parameter	Value	Description/Equation no.
1.	rt	12	Population size
2.	t	2	Sample size
3.	$\bar{\Delta}_1$	14.25	Using (5.1)
4.	$\bar{\Delta}_2$	6.91	Using (5.2)
5.	$S\epsilon$	3.95	Using (5.10)
6.	$S\epsilon'$	2.27	Using(5.11)
7.	$C^*\epsilon$	0.27	Using (5.7)
8.	$C^*\epsilon'$	0.32	Using (5.8)
9.	ρ	0.98	(5.13)
10.	R	0.8285	Using Theorem (5.6)
11.	r	6	Large Integer

Table 7.6 Almost Unbiased Choice of P for given (g, R)

S.No.	R	g	Choice of P	Bias	MSE
1.	0.8285	0.5	$P_1 = 0.61$	-0.0001	0.1308
2.	0.8285	0.5	$P_2 = --$	–	–
3.	0.8285	0.5	$P_3 = --$	–	–

Table 7.7 Choice of P for Minimum MSE for given (g, R)

S.No.	R	g	Choice of q	MSE	Bias
1.	0.8285	0.5	$P_{1(opt)} = 0.8830$	0.0462	0.0100
2.	0.8285	0.5	$P_{2(opt)} = --$	–	–
3.	0.8285	0.5	$P_{3(opt)} = --$	–	–

Remark 4. As per data of table 7.1, and calculation of table 7.6 and table 7.7 when R=0.8285, g=0.5, the most suitable range of P is P= 0.61 to P=0.8830 to produce the best estimate of average resource consumption score using proposed E. This range of consistent P provides almost unbiased optimal estimate of average resource score.

Table 7.8 Comparison with Particular Cases of Proposed E (Using PRE)

S.No.	Choice of P	Bias(E) (Theorem 5.2)	MSE(E) (Theorem 5.5)	Comparison (PRE)
1.	At P=1	0.0220	0.1001	53.1468%
2.	At P=2	0.1066	6.1768	99.24407%
3.	At P=3	0.0533	3.2824	98.5711%
4.	At P=4	0.0000	1.3049	96.4058%
5.	At P=5	-0.0212	0.4958	90.5405%

Where, Percentage Relative Efficiency (PRE) at $P_{opt} = 0.8830$

$$= \frac{[MSE(E)_{P=1,2,3,4,5}] - [MSE(E)_{P_{(opt)}}]}{MSE(E)_{P=1,2,3,4,5}} X100$$

Remark 5. As per table 7.8, for given pair of (g,R)=(0.5, 0.8285), the best choice of P is $P = 0.8830$ which provides lowest bias and minimum MSE for the proposed strategy E in light of given data set of table 7.1.

Remark 6. The proposed strategy E is constantly better over particular cases at P= 1, 2, 3, 4, 5 as evident from table 7.8 using PRE.

Table 7.9 Calculation for First Sample Parameter and Confidence Interval (C.I.)

$\bar{\delta_1}$	$\bar{\delta_2}$	$(s_\epsilon)^2$	$(s'_\epsilon)^2$	(c_ϵ)	(c'_ϵ)	$(s_{\epsilon\epsilon'})$	ρ'	C.I.
14.8333	7.5	16.1660	5.9000	0.2711	0.3238	3.1144	0.3189	at P =0.61 (unbiasedness) [12.47, 17.18]
	–	–	–	–	–	–	–	at P =0.883 (Minimum MSE) [12.11, 17.47]

Table 7.10 Calculation for Second Sample Parameter and Confidence Interval (C.I.)

$\bar{\delta_1}$	$\bar{\delta_2}$	$(s_\epsilon)^2$	$(s'_\epsilon)^2$	(c_ϵ)	(c'_ϵ)	$(s_{\epsilon\epsilon'})$	ρ'	C.I.
13.6666	6.3333	17.4666	4.6666	0.3058	0.3410	2.9888	0.3310	at P =0.61 (unbiasedness) [11.28, 16.04]
	–	–	–	–	–	–	–	at P= 0.883 (minimum MSE) [11.03, 16.30]

Remark 7. $\bar{\delta_1}$ and $\bar{\delta_2}$ are in (5.3) and (5.4), and others are defined as

$(c_\epsilon) = \frac{(s_\epsilon)}{\bar{\delta_1}}$

$(c_{\epsilon'}) = \frac{(s_{\epsilon'})}{\bar{\delta_2}}$

$(c\epsilon\epsilon') = \frac{(s_{\epsilon\epsilon'})}{(\bar{\delta_1}.\bar{\delta_2})}$

$(s_\epsilon)^2 = \frac{1}{r-1}\sum_{j=1}^{r}(s_{\epsilon_{ij}} - \bar{\delta_1})^2$

$(s\epsilon')^2 = \frac{1}{r-1}\sum_{j=1}^{r}(s_{\epsilon'_{ij}} - \bar{\delta_2})^2$

$(s_{\epsilon\epsilon'}) = \frac{1}{r-1}\sum_{i=1}^{t}\sum_{j=1}^{r}(s_{\epsilon_{ij}} - \bar{\delta_1})(s_{\epsilon'_{ij}} - \bar{\delta_2})$

$\rho' = \frac{s_{\epsilon\epsilon'}}{[(s_\epsilon)(s_{\epsilon'})]}$

Remark 8. Let P[A] denotes the probability of event A then a 95% confidence interval in defined as:

$$P\left[(E)_P - 1.96\sqrt{MSE(E)_P},\ (E)_P + 1.96\sqrt{MSE(E)_P}\right] = 0.95$$

This interval indicates that there is 95% chance, the true value of population mean lies in the range $P\left[(E)_P - 1.96\sqrt{MSE(E)_P},\ (E)_P + 1.96\sqrt{MSE(E)_P}\right] = 0.95$. If the value of $P = P_{opt}$ then this interval will be the optimal confidence interval with respect to minimum MSE (or with respect to unbiasedness as the case may be)

Table 7.11 Ready Reckoner for P Value for Unbiasedness for given (g, R)

S.No.	R	g	Choice of P	Bias	MSE	S.No.	R	g	Choice of P	Bias	MSE
1.	0.2	0.3	$P_1 = 4.5641$	-0.0182	0.7089	46.	1.2	0.3	$P_1 = 1.2549$	0.0016	0.7877
2.	0.2	0.3	$P_2 = 4$	0.0000	1.3049	47.	1.2	0.3	$P_2 = 4$	0.0000	1.3049
3.	0.2	0.3	$P_3 = --$	–	–	48.	1.2	0.3	$P_3 = --$	–	–
4.	0.2	0.6	$P_1 = 0.45410$	-0.0202	0.6601	49.	1.2	0.6	$P_1 = 1.335$	0.0682	0.7085
5.	0.2	0.6	$P_2 = 4$	0.0000	1.3049	50	1.2	0.6	$P_2 = 4$	0.0000	1.3049
6.	0.2	0.6	$P_3 = --$	–	–	51.	1.2	0.6	$P_3 = --$	–	–
7.	0.2	0.9	$P_1 = 4.5264$	-0.0224	0.6019	52.	1.2	0.9	$P_1 = 1.0900$	0.0675	0.6823
8.	0.2	0.9	$P_2 = 4$	0.0000	0.1304	53.	1.2	0.9	$P_2 = 4$	0.0000	0.1304
9.	0.2	0.9	$P_3 = --$	–	–	54.	1.2	0.9	$P_3 = --$	–	–
10.	0.4	0.3	$P_1 = 5.2910$	-0.0209	0.4150	55.	1.4	0.3	$P_1 = 1.3560$	0.1407	2.2086
11.	0.4	0.3	$P_2 = 4$	0.0000	1.3049	56.	1.4	0.3	$P_2 = 4$	0.0000	1.3049
12.	0.4	0.3	$P_3 = --$	–	–	57.	1.4	0.3	$P_3 = --$	–	–
13.	0.4	0.6	$P_1 = 5.3389$	-0.0195	0.4573	58.	1.4	0.6	$P_1 = 1.2048$	0.1366	1.9884
14.	0.4	0.6	$P_2 = 4$	0.0000	1.3049	59.	1.4	0.6	$P_2 = 4$	0.0000	1.3049
15.	0.4	0.6	$P_3 = --$	–	–	60.	1.4	0.6	$P_3 = --$	–	–
16.	0.4	0.9	$P_1 = 5.4050$	-0.0177	0.5174	61.	1.4	0.9	$P_1 = 1.1430$	0.1345	1.9025
17.	0.4	0.9	$P_2 = 4$	0.0000	1.3049	62.	1.4	0.9	$P_2 = 4$	0.0000	1.3049
18.	0.4	0.9	$P_3 = --$	–	–	63.	1.4	0.9	$P_3 = --$	–	–
19.	0.6	0.3	$P_1 = 7.0810$	-0.0144	0.2561	64.	1.6	0.3	$P_1 = 1.4121$	0.2341	4.3451
20.	0.6	0.3	$P_2 = 4$	-0.0000	1.3049	65.	1.6	0.3	$P_2 = 4$	-0.0000	1.3049
21.	0.6	0.3	$P_3 = --$	–	–	66.	1.6	0.3	$P_3 = --$	–	–
22.	0.6	0.6	$P_1 = 12.0001$	-0.0085	0.5738	67.	1.6	0.6	$P_1 = 1.2490$	0.2259	3.9084
23.	0.6	0.6	$P_2 = 4$	0.0000	1.309	68.	1.6	0.6	$P_2 = 4$	0.0000	1.309
24.	0.6	0.6	$P_3 = --$	–	–	69.	1.6	0.6	$P_3 = --$	–	–
25.	0.6	0.9	$P_1 = 0.0001$	-0.0044	0.8903	70.	1.6	0.9	$P_1 = 1.1781$	0.2232	3.7648
26.	0.6	0.9	$P_2 = 4$	0.0000	1.3049	71.	1.6	0.9	$P_2 = 4$	0.0000	1.3049
27.	0.6	0.9	$P_3 = --$	–	–	72.	1.6	0.9	$P_3 = --$	–	–
28.	0.8	0.3	$P_1 = 0.0001$	0.0027	0.1015	73.	1.8	0.3	$P_1 = 1.4480$	0.3497	7.1827
29.	0.8	0.3	$P_2 = 4$	0.0000	1.3049	74.	1.8	0.3	$P_2 = 4$	0.0000	1.3049
30.	0.8	0.3	$P_3 = --$	–	–	75.	1.8	0.3	$P_3 = --$	–	–
31.	0.8	0.6	$P_1 = 0.6354$	-0.0020	0.1752	76.	1.8	0.6	$P_1 = 1.2800$	0.3391	6.5374
32.	0.8	0.6	$P_2 = 4$	0.0000	1.3049	77.	1.8	0.6	$P_2 = 4$	0.0000	1.3049
33.	0.8	0.6	$P_3 = --$	–	–	78.	1.8	0.6	$P_3 = --$	–	–
34.	0.8	0.9	$P_1 = 0.8201$	-0.0015	0.1456	79.	1.8	0.9	$P_1 = 1.2030$	0.3331	6.2609
35.	0.8	0.9	$P_2 = 4$	0.0000	1.3049	80.	1.8	0.9	$P_2 = 4$	0.0000	1.3049
36.	0.8	0.9	$P_3 = --$	–	–	81.	1.8	0.9	$P_3 = --$	–	–
37.	1.0	0.3	$P_1 = 0.9008$	0.0158	0.0592	82.	2.0	0.3	$P_1 = 1.4738$	0.4928	10.8517
38.	1.0	0.3	$P_2 = 4$	0.0000	1.3049	83.	2.0	0.3	$P_2 = 4$	0.0000	1.3049
39.	1.0	0.3	$P_3 = --$	–	–	84.	2.0	0.3	$P_3 = --$	–	–
40.	1.0	0.6	$P_1 = 0.9800$	0.0186	0.0766	85.	2.0	0.6	$P_1 = 1.3020$	0.4704	9.7348
41.	1.0	0.6	$P_2 = 4$	0.0000	1.3049	86.	2.0	0.6	$P_2 = 4$	0.0000	1.3049
42.	1.0	0.6	$P_3 = --$	–	–	87.	2.0	0.6	$P_3 = --$	–	–
43.	1.0	0.9	$P_1 = 0.9500$	0.0109	0.0470	88.	2.0	0.9	$P_1 = 1.2220$	0.4678	9.4728
44.	1.0	0.9	$P_2 = 4$	0.0000	1.3049	89.	2.0	0.9	$P_2 = 4$	0.0000	1.3049
45.	1.0	0.9	$P_3 = --$	–	–	90.	2.0	0.9	$P_3 = --$	–	–

Table 7.12 Ready Reckoner for P Value for Minimum MSE for given (g, R)

S.No.	R	g	Choice of P	Bias	MSE	S.No.	R	g	Choice of P	Bias	MSE
1.	0.2	0.3	$P_1 = 4.5001$	-0.0171	0.7561	46.	1.2	0.3	$P_1 = 1.1452$	0.0396	0.2993
2.	0.2	0.3	$P_2 = --$	–	–	47.	1.2	0.3	$P_2 = --$	–	–
3.	0.2	0.3	$P_3 = --$	–	–	48.	1.2	0.3	$P_3 = --$	–	–
4.	0.2	0.6	$P_1 = 4.5005$	-0.0196	0.6861	49.	1.2	0.6	$P_1 = 1.0743$	0.0408	0.2994
5.	0.2	0.6	$P_2 = --$	–	–	50.	1.2	0.6	$P_2 = --$	–	–
6.	0.2	0.6	$P_3 = --$	–	–	51.	1.2	0.6	$P_3 = --$	–	–
7.	0.2	0.9	$P_1 = 0.1964$	-0.0052	0.7840	52.	1.2	0.9	$P_1 = 1.0495$	0.0410	0.2969
8.	0.2	0.9	$P_2 = --$	–	–	53.	1.2	0.9	$P_2 = --$	–	–
9.	0.2	0.9	$P_3 = --$	–	–	54.	1.2	0.9	$P_3 = --$	–	–
10.	0.4	0.3	$P_1 = 5.5014$	-0.0203	0.3744	55.	1.4	0.3	$P_1 = 1.2340$	0.0621	0.6498
11.	0.4	0.3	$P_2 = --$	–	–	56.	1.4	0.3	$P_2 = --$	–	–
12.	0.4	0.3	$P_3 = --$	–	–	57.	1.4	0.3	$P_3 = --$	–	–
13.	0.4	0.6	$P_1 = 0.2300$	-0.0052	0.3826	48.	1.4	0.6	$P_1 = 1.1270$	0.0644	0.6452
14.	0.4	0.6	$P_2 = --$	–	–	59.	1.4	0.6	$P_2 = --$	–	–
15.	0.4	0.6	$P_3 = --$	–	–	60.	1.4	0.6	$P_3 = --$	–	–
16.	0.4	0.9	$P_1 = 0.6381$	-0.0057	0.3830	61.	1.4	0.9	$P_1 = 1.1201$	0.0991	1.2263
17.	0.4	0.9	$P_2 = --$	–	–	62.	1.4	0.9	$P_2 = --$	–	–
18.	0.4	0.9	$P_3 = --$	–	–	63.	1.4	0.9	$P_3 = --$	–	–
19.	0.6	0.3	$P_1 = 0.0001$	0.0023	0.1094	64.	1.6	0.3	$P_1 = 1.2928$	0.0888	1.1370
20.	0.6	0.3	$P_2 = --$	–	–	65.	1.6	0.3	$P_2 = --$	–	–
21.	0.6	0.3	$P_3 = --$	–	–	66.	1.6	0.3	$P_3 = --$	–	–
22.	0.6	0.6	$P_1 = 0.6900$	-0.0009	0.1416	67.	1.6	0.6	$P_1 = 1.1664$	0.0928	1.1367
23.	0.6	0.6	$P_2 = --$	–	–	68.	1.6	0.6	$P_2 = --$	–	–
24.	0.6	0.6	$P_3 = --$	–	–	69.	1.6	0.6	$P_3 = --$	–	–
25.	0.6	0.9	$P_1 = 0.8501$	-0.0001	0.1098	70.	1.6	0.9	$P_1 = 1.0990$	0.0756	0.8146
26.	0.6	0.9	$P_2 = --$	–	–	71.	1.6	0.9	$P_2 = --$	–	–
27.	0.6	0.9	$P_3 = --$	–	–	72.	1.6	0.9	$P_3 = --$	–	–
28.	0.8	0.3	$P_1 = 0.7102$	0.0093	0.0473	73.	1.8	0.3	$P_1 = 1.3360$	0.1203	1.7742
29.	0.8	0.3	$P_2 = --$	–	–	74.	1.8	0.3	$P_2 = --$	–	–
30.	0.8	0.3	$P_3 = --$	–	–	75.	1.8	0.3	$P_3 = --$	–	–
31.	0.8	0.6	$P_1 = 0.8863$	0.0081	0.0477	76.	1.8	0.6	$P_1 = 1.1970$	0.1259	1.7725
32.	0.8	0.6	$P_2 = --$	–	–	77.	1.8	0.6	$P_2 = --$	–	–
33.	0.8	0.6	$P_3 = --$	–	–	78.	1.8	0.6	$P_3 = --$	–	–
34.	0.8	0.9	$P_1 = 0.9301$	0.0079	0.0476	79.	1.8	0.9	$P_1 = 1.1395$	0.1282	1.7790
35.	0.8	0.9	$P_2 = --$	–	–	80.	1.8	0.9	$P_2 = --$	–	–
36.	0.8	0.9	$P_3 = --$	–	–	81.	1.8	0.9	$P_3 = --$	–	–
37.	1.0	0.3	$P_1 = 0.9700$	0.0198	0.0831	82.	2.0	0.3	$P_1 = 1.3690$	0.1567	2.5605
38.	1.0	0.3	$P_2 = --$	–	–	83.	2.0	0.3	$P_2 = --$	–	–
39.	1.0	0.3	$P_3 = --$	–	–	84.	2.0	0.3	$P_3 = --$	–	–
40.	1.0	0.6	$P_1 = 0.9801$	0.0187	0.0767	85.	2.0	0.6	$P_1 = 1.2215$	0.1638	2.5523
41.	1.0	0.6	$P_2 = --$	–	–	86.	2.0	0.6	$P_2 = --$	–	–
42.	1.0	0.6	$P_3 = --$	–	–	87.	2.0	0.6	$P_3 = --$	–	–
43.	1.0	0.9	$P_1 = 1.001$	0.0223	0.1022	88.	2.0	0.9	$P_1 = 1.1585$	0.1671	2.5642
44.	1.0	0.9	$P_2 = --$	–	–	89.	2.0	0.9	$P_2 = --$	–	–
45.	1.0	0.9	$P_3 = --$	–	–	90.	2.0	0.9	$P_3 = --$	–	–

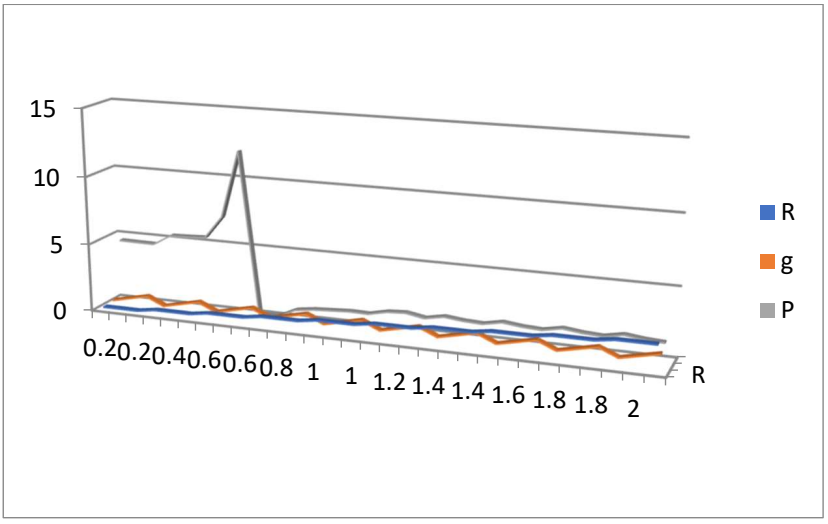


Fig 7.1 Three Dimensional Graph for Almost Unbiasedness

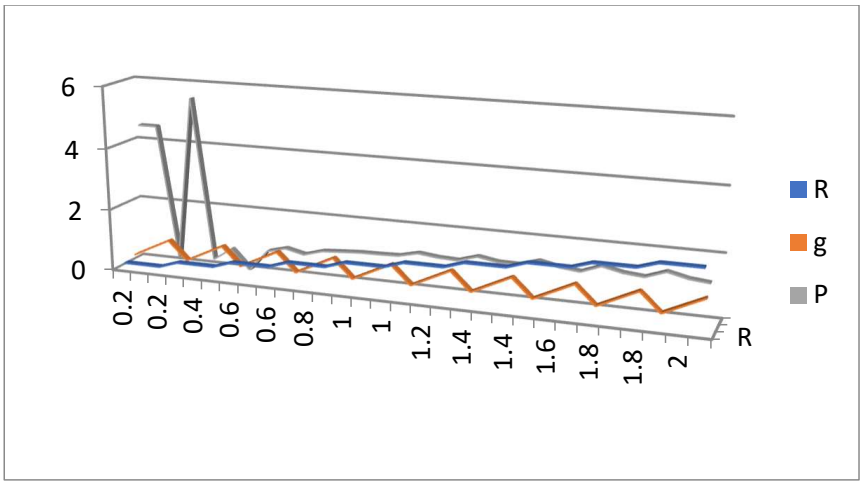


Fig 7.2 Three Dimensional Graph for Optimum MSE

7.1. About given (g, R)

To observe that $g = \frac{1}{t}$, where t is size of the group G_1 who is known and fixed before the draw of random sample using the single-node-systematic sampling procedure. The quantity R is R = $\frac{\rho(C^*\epsilon)}{(C^*\epsilon')}$ as described in theorem 5.6. The R is a ratio of $C^*\epsilon$ and $C^*\epsilon'$ who are coefficients of variations and remain almost stable over time occasions. Therefore, R could be guessed from the past data or by past reports or by the current sample. While minor fluctuations in R, the table 7.13 helps to find out most suitable values of P.

Table 7.13 Best Choice of P Under Fluctuations of R (independent of g)

Range of R	Best range of P for unbiasedness	Best range of P for opt. MSE
$0 < R \leq 0.6$	$0 < P < 5.4$	$0 < P < 5.5$
$0.6 < R \leq 1.0$	$0 < P < 0.95$	$0.71 < P < 1.001$
$1.0 < R \leq 1.6$	$1.09 < P < 1.7$	$1.0 < P < 1.29$
$1.6 < R \leq 2.0$	$1.22 < P < 1.45$	$1.15 < P < 1.36$

Note 7.2 The table 7.13 is independent of variations of g which is created using the Ready-reckoner tables 7.11 and 7.12. However, for any given (g, R), one can choose the best P using Ready-reckoner table 7.11 and 7.12.

Note 7.3 The most suitable value of P,for almost unbiased estimation, is ($P = 0.10$ to $P = 5.4$) for all $(g, 0 < g < 1)$ and for all R, $(0 < R \leq 2)$ (see table 7.13).

Note 7.4 The most suitable value of P for low MSE is ($P = 0.001$ to $P = 5.5$) for all g and $0 < R \leq 2$ (see table 7.13)

Note 7.5 The general recommended P is $P \in (0.1, 5.5)$ where one can get low bias and low MSE by the proposed strategy, whatever be the g, $(0 < g < 1)$ and whatever be R, $0 < R \leq 2$.

Note 7.6 This is beauty of the proposed strategy E because it is now independent to the known pair (g, R) to produce good estimate of average consumption of score in a Bipartite graph population when user has chosen $0.1 < P < 5.5$.

Note 7.7 The figure 7.2 presents three dimensional aspect of choice of P for given pair (g, R). The X- axis has R, Z- axis has g and Y- axis has calculated value of P.

8. Conclusion

On recapitulation, the content of the paper has a sample based methodology for evaluating the average resource consumption scores with the help of Bipartite graph. A single node systematic sampling procedure is suggested in the content which is a graph sampling based procedure, useful for parameter estimation in places where similar situation exists. This procedure opens up avenues for further researches in the area of sampling where population is synonymous to the graphical structure. The Bipartite graph is used for getting solution of the problem of estimation of the travel resource consumption parameter. An estimation strategy is proposed and its properties are derived. It is proved that they are bias and MSE controlled both, at the same time due to cubic equation. The proposed also converts to an almost unbiased optimum strategy at some appropriate choice of P.

The main difficulty occurs with the strategy is the selection of suitable value of constant incorporated in its structure. Two Ready-reckoner tables have been prepared who are useful for the quick selection of constant for give pair of (g, R) values. These tables provide to users, the population independency just as to utilize only (g, R), irrespective of the other population characteristics. Whatever may the distance and resource consumption, if (g, R) are given (or guessed or estimated or calculated from past data) ,the bias and MSE can be predicted through these Ready-reckoner tables who made easy to apply the suggested. It can be used for developing computer algorithms and softwares for enhancing the applications.

Two systematic samples are taken into account and their computed confidence intervals are $(12.47 - 17.18)$, $(12.11 - 17.47)$ for first sample , $(12.28 - 16.04)$, $(11.03 - 16.30)$ for second sample. All intervals are catching the true values of average resource consumption which is 14.25. The interval based prediction is sound enough indicating the efficiency of the suggested strategy. For $0 < R \leq 2$, a table has been developed, for the ease of users, towards rapid selection of best value of constant P whatever be the value of g and whatever be the population values in terms of distance and resource consumption scores . The general recommendation for best choice of P is the range $(0.1 - 5.5)$ irrespective of other population features and dependencies. Bipartite graph is used as a model tool for developing the single node systematic sampling procedure who can be extended further using other kinds of estimation procedures exiting in the concerned literature.

References

[1] Deo, N. (2001). *Graph theory with application to engineering and computer science*. Eastern Economy Edn, New Delhi, India .

[2] Moons, E., Loomis,J., Proost, S. Eggermont, K. and Hermy,M. (2001). *Travel cost and time measurement in travel cost models*. Katholieke Universiteit Leuven.

[3] Wikipedia, the Free Encyclopedia. *https://www.google.com/search?client=firefox-b-dq=Travel+Cost+Analysis*.

[4] Wikipedia, the Free Encyclopedia. *https://www.google.com/search?client=firefox-b-dq=ecosystem+valuation*.

[5] Shukla, D. (2002). *F-T estimator under two-phase sampling*. Metron, 60:97-106.

[6] Singh, S. (2003). *Advanced sampling theory with applications*. Kluwer Academic Publisher, Netherland, Springer, Dordrecht. htttps://doi.org/10.1007/978-94-007-0789-4.

[7] Cochran, W.G. (2005). *Sampling Techniques*. Jhon Willy and Sons.

[8] Zhang,B. , Chambers M. C. and Tabb, D.L. (2007). *Proteomic parsimony through Bipartite graph analysis improves accuracy and tranparency* Journal of Research Articles Proteome Research, 6.

[9] Shukla,D., Rajput, Y.S. and Thakur, N.S. (2009). *Estimation of spanning tree mean-edge using node sampling*. Model Assisted Statistics and Applications, 4:23-37.

[10] Shukla,D., Rajput, Y.S. and Thakur, N.S. (2010). *Edge estimation in population of planer graph using sampling*. Journal of Reliability and Statistical Studies, 3:13-29.

[11] Koutra,D. , Tong, H. and Lubensky, D. (2013). *Big-Align: fast Bipartite graph alignment*. 13th International Conference on Data Mining (IEEE), 389-398.

[12] Shukla,D., Rajput, Y.S. and Thakur, N.S.(2014). *Edge estimation in the population of a binary tree using node-sampling*. Communication in Statistics Theory and Methods, 43:2815“2829.

[13] F. Serratosa (2014). *Speeding up fast Bipartite graph matching through a new cost matrix*. International Journal of Pattern Recognition and Artificial Intellingence, 29, http://dx.doi.org/10.1016/j.patrec.2014.04.015.

[14] Shukla, D., Pathak, S. and Trivedi, S.K. (2016). *Graph sampling by Spanning tree under stratified setup*. Research and Review: Journal of Statistics, 5:11-31.

[15] Xiong, H., Zhau,Y., Hu,C., Wei, X. and Li,L. (2017). *A novel recommendation algorithm frame for tourist spots based on multi clustering Bipartite graph*. 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, 276-282.

[16] Shukla,D., Trivedi,S.K., Pathak, S. and Rajoriya, D. (2018). *Mean edge estimation in population of Binary tree using stratified sampling procedure*. Research and Reviews:Journal of Statistics, 7:60-73.

[17] Shukla,D. Trivedi,S.K., Pathak, S. and Rajoriya,D. (2019). *Mean edge estimation in population of Planar graph using stratified sampling procedure*. Research and Reviews: Journal of Statistics, 8:11-31.

[18] Rajoriya, D. and Shukla, D. (2020). *Optimal estimation strategies of parameters in Hamiltonian circuit type population*. Journal of Critical Reviews, 7:3561-3573.

[19] Donga,J., Bhojak, P., Patel, K.and Shah, V. (2021). *An analysis of different computer science algorithms with the graph theory of mathematics*. Reliability: Theory and Methods, 16:376-383.

ISSN 1932-2321