

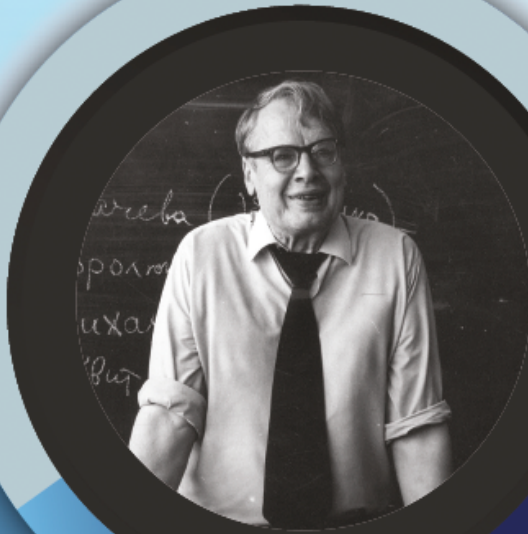
RTA

ISSN 1932-2321

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

RELIABILITY:
THEORY & APPLICATIONS

INTERNATIONAL
GROUP ON
RELIABILITY



GNEDENKO FORUM PUBLICATIONS

#2

(73) VOL.18

JUNE
2023

SAN DIEGO

RELIABILITY

RISK ANALYSIS

MAINTENANCE

SAFETY

ISSN 1932-2321

© "Reliability: Theory & Applications", 2006, 2007, 2009-2023

© " Reliability & Risk Analysis: Theory & Applications", 2008

© I.A. Ushakov

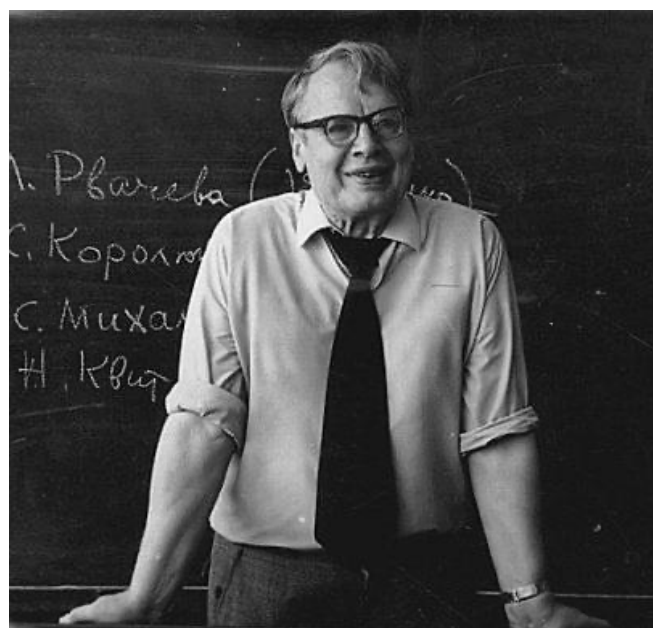
© A.V. Bochkov, 2006-2023

© Kristina Ushakov, Cover Design, 2023

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

All rights are reserved

The reference to the magazine "Reliability: Theory & Applications"
at partial use of materials is obligatory.



RELIABILITY: THEORY & APPLICATIONS

Vol.18 No.2 (73),
June 2023

San Diego
2023

Editorial Board

Editor-in-Chief

Rykov, Vladimir (Russia)

Doctor of Sci, Professor, Department of Applied Mathematics & Computer Modeling, Gubkin Russian State Oil & Gas University, Leninsky Prospect, 65, 119991 Moscow, Russia.

e-mail: vladimir_rykov@mail.ru

Managing Editors

Bochkov, Alexander (Russia)

Doctor of Technical Sciences, Scientific Secretary JSC NIIAS, Scientific-Research and Design Institute Informatization, Automation and Communication in Railway Transport, Moscow, Russia, 107078, Orlikov pereulok, 5, building 1
e-mail: a.bochkov@gmail.com

Gnedenko, Ekaterina (USA)

PhD, Lecturer Department of Economics Boston University, Boston 02215, USA
e-mail: gnedenko@bu.edu

Deputy Editors

Dimitrov, Boyan (USA)

Ph.D., Dr. of Math. Sci., Professor of Probability and Statistics, Associate Professor of Mathematics (Probability and Statistics), GMI Engineering and Management Inst. (now Kettering)
e-mail: bdimetro@kettering.edu

Gnedenko, Dmitry (Russia)

Doctor of Sci., Assos. Professor, Department of Probability, Faculty of Mechanics and Mathematics, Moscow State University, Moscow, 119899, Russia
e-mail: dmitry@gnedenko.com

Kashtanov, Victor A. (Russia)

PhD, M. Sc (Physics and Mathematics), Professor of Moscow Institute of Applied Mathematics, National Research University "Higher School of Economics" (Moscow, Russia)
e-mail: VAKashtan@yandex.ru

Krishnamoorthy, Achyutha (India)

M.Sc. (Mathematics), PhD (Probability, Stochastic Processes & Operations Research), Professor Emeritus, Department of Mathematics,

Cochin University of Science & Technology, Kochi-682022, INDIA.

e-mail: achyuthacusat@gmail.com

Recchia, Charles H. (USA)

PhD, Senior Member IEEE Chair, Boston IEEE Reliability Chapter A Joint Chapter with New Hampshire and Providence, Advisory Committee, IEEE Reliability Society
e-mail: charles.recchia@macom.com

Shybinsky Igor (Russia)

Doctor of Sci., Professor, Division manager, VNIIS (Russian Scientific and Research Institute of Informatics, Automatics and Communications), expert of the Scientific Council under Security Council of the Russia
e-mail: igor-shubinsky@yandex.ru

Yastrebenetsky, Mikhail (Ukraine)

Doctor of Sci., Professor. State Scientific and Technical Center for Nuclear and Radiation Safety (SSTC NRS), 53, Chernishevskaya str., of.2, 61002, Kharkov, Ukraine
e-mail: ma_yastrebenetsky@sstc.com.ua

Associate Editors

Aliyev, Vugar (Azerbaijan)

Doctor of Sci., Professor, Chief Researcher of the Institute of Physics of the National Academy of Sciences of Azerbaijan, Director of the AMIR Technical Services Company
e-mail: prof.vugar.aliyev@gmail.com

Balakrishnan, Narayanaswamy (Canada)

Professor of Statistics, Department of Mathematics and Statistics, McMaster University
e-mail: bala@mcmaster.ca

Carrión García, Andrés (Spain)

Professor Titular de Universidad, Director of the Center for Quality and Change Management, Universidad Politécnica de Valencia, Spain
e-mail: acarrion@eio.upv.es

Chakravarthy, Srinivas (USA)

Ph.D., Professor of Industrial Engineering & Statistics, Departments of Industrial and Manufacturing Engineering & Mathematics, Kettering University (formerly GMI-EMI) 1700,

University Avenue, Flint, MI48504
e-mail: schakrav@kettering.edu

Cui, Lirong (China)

PhD, Professor, School of Management & Economics, Beijing Institute of Technology, Beijing, P. R. China (Zip:100081)
e-mail: lirongcui@bit.edu.cn

Finkelstein, Maxim (SAR)

Doctor of Sci., Distinguished Professor in Statistics/Mathematical Statistics at the UFS. Visiting researcher at Max Planck Institute for Demographic Research, Rostock, Germany and Visiting research professor (from 2014) at the ITMO University, St Petersburg, Russia
e-mail: FinkelM@ufs.ac.za

Kaminsky, Mark (USA)

PhD, principal reliability engineer at the NASA Goddard Space Flight Center
e-mail: mkaminskiy@hotmail.com

Krivtsov, Vasiliy (USA)

PhD. Director of Reliability Analytics at the Ford Motor Company. Associate Professor of Reliability Engineering at the University of Maryland (USA)
e-mail: VKrivtso@Ford.com_krivtsov@umd.edu

Lemeshko Boris (Russia)

Doctor of Sci., Professor, Novosibirsk State Technical University, Professor of Theoretical and Applied Informatics Department
e-mail: Lemeshko@ami.nstu.ru

Lesnykh, Valery (Russia)

Doctor of Sci. Director of Risk Analysis Center, 20-8, Staraya Basmannaya str., Moscow, Russia, 105066, LLC "NIIGAZECONOMIKA" (Economics and Management Science in Gas Industry Research Institute)
e-mail: vvlesnykh@gmail.com

Levitin, Gregory (Israel)

PhD, The Israel Electric Corporation Ltd. Planning, Development & Technology Division. Reliability & Equipment Department, Engineer-Expert; OR and Artificial Intelligence applications in Power Engineering, Reliability.
e-mail: levitin@iec.co.il

Limnios, Nikolaos (France)

Professor, Université de Technologie de Compiègne, Laboratoire de Mathématiques, Appliquées Centre de Recherches de Royallieu, BP 20529, 60205 COMPIEGNE CEDEX, France
e-mail: Nikolaos.Limnios@utc.fr

Papic, Ljubisha (Serbia)

PhD, Professor, Head of the Department of Industrial and Systems Engineering Faculty of Technical Sciences Cacak, University of Kragujevac, Director and Founder the Research Center of Dependability and Quality Management (DQM Research Center), Prijedor, Serbia
e-mail: dqmcenter@mts.rs

Ram, Mangey (India)

Professor, Department of Mathematics, Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun, India. Visiting Professor, Institute of Advanced Manufacturing Technologies, Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia.
e-mail: mangeyram@gmail.comq

Zio, Enrico (Italy)

PhD, Full Professor, Direttore della Scuola di Dottorato del Politecnico di Milano, Italy.
e-mail: Enrico.Zio@polimi.it

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

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

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

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

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

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

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

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

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

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

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

Table of Contents

ON THE TESU-G FAMILY OF DISTRIBUTIONS APPLIED TO LIFE DATA ANALYSIS 24

Idzhar A. Lakibul, Bernadette F. Tubo

This paper derives distributions from the U-quadratic and the T-X family of distributions labeled as the T-extended Standard U-quadratic. G family of distributions or simply, TeSU-G family. In particular, the TeSU-Weibull distribution (TeSU-W) is explored with respect to some statistical properties such as its limiting distribution, moment, mean and variance and moment generating function. Also, the statistical properties of the TeSU-Exponential distribution (TeSU-E) which is a special case of the TeSU-W are also derived. The Weibull and Exponential distributions are mostly used in life data analysis because of its ability to adapt to different situations. Moreover, the formula for the median is derived via a proposed algorithm. Simulation study is conducted to verify the performance of the ML estimates of the TeSU-W distribution for varied sample sizes. Further, real life data analysis reveals that derived extended distribution can provide a better fit than several well-known distributions.

EXPONENTIATED ADYA DISTRIBUTION: PROPERTIES AND APPLICATIONS 39

Rashid A. Ganaie, Aafaq A. Rather, D. Vedavathi Saraja, Mahfooz Alam, Aijaz Ahmad, Arif Muhammad Tali, Rifat Nisa, Berihan R. Elemary

In this article, we introduce a new generalization of Adya distribution known as Exponentiated Adya distribution. We have also derived and discussed its different statistical properties. The Exponentiated Adya distribution has two parameters (scale and shape). The different structural properties of the proposed distribution have been obtained. The maximum likelihood estimation technique is also used for estimating the parameters of the proposed distribution. Finally, a real lifetime data set is used for examining the superiority of the proposed distribution.

A REVIEW ON QUANTILE FUNCTIONS, INCOME DISTRIBUTIONS, AND INCOME INEQUALITY MEASURES 50

Ashlin Varkey, Haritha N Haridas

The modeling of income data has originated a century ago with the work of Vilfredo Pareto. Since then several authors have added immense literature about income distributions and income inequality measures. In the present paper, we have pointed out some recent works in income distributions and income inequality measures. Recently the potential of the quantile function has been discussed in parallel with the distribution function for modeling reliability and income data. So we have also included some quantile functions existing in the literature and their potentiality to model income data in the review. We derived the Lorenz curve, Gini index, Pietra index, Bonferroni index, Bonferroni curve, and Zenga curve for six distributions and checked the model adequacy of three distributions using real income data.

THE EXTENDED ERLANG TRUNCATED EXPONENTIAL DISTRIBUTION: PROPERTIES AND APPLICATION TO STATISTICAL PREDICTION PROBLEM..... 63

Imtiyaz A. Shah

The Erlang Truncated (ETE) distribution is modified and the new lifetime distribution is called the Extended Erlang Truncated Exponential (EETE) distribution. Some statistical and reliability properties of the new distribution functions have been characterized based on two non-adjacent generalized and dual generalized order statistics. Moreover, we show that these characterization properties provide a beneficial strategy to predict future events, which are based on past or current events and on an arbitrary distribution function. A characterization in statistics is a specific distributional property of a statistic that uniquely identify related parametric family of distributions. In statistical applications, the researchers usually want to verify whether the data that they are dealing with belong to a certain family of DFs. Therefore, the researchers have to rely on a characterization of the assumed distribution and check if the corresponding conditions are satisfied.

INTERVAL PATTERN RECOGNITION IN RELATION TO INFORMATION ABOUT THE TREATMENT OF COVID 19 IN PATIENTS WITH BLOOD DISEASES 73

G. Tsitsiashvili, V. Nevzorova, A. Talko, M. Osipova

Since the problem of coronavirus infection, especially in vulnerable groups, remains relevant in the healthcare system, it is necessary to recognize the risk of fatal events during the treatment of patients from COVID-19. Patients with hemoblastosis are exposed to a more severe course of coronavirus infection than the general population, as well as an increased risk of fatal events. The aim of our study was to determine the effect of signs on the risk outcomes of fatal events among patients with diseases of the blood system, both of a tumor and non-tumor nature. According to the values of the signs, it was necessary to recognize survivors and those who died during treatment from COVID-19. The method of interval pattern recognition was chosen due to the presence of big data, it is described in detail in the article. The patients were broken down by gender and age. The signs that significantly affect the forecast according to the developed algorithm were identified. This is especially evident in groups of men and women with malignant diseases of the blood system over the age of 60 years. In these groups, a positive outcome of treatment is detected due to the presence of a large set of uninformative signs. This phenomenon is closely related to the tasks of technical gerontology.

STUDY OF OPERATING MODES OF SMALL HPSS' HYDROELECTRIC UNITS EQUIPPED WITH SYNCHRONOUS GENERATOR..... 79

L.H. Hasanova, H.S. Aliyev, M.M. Musayev

A universal analytical expression is obtained for the power of hydraulic turbines of small hydropower plants such as Francis, Pelton and Kaplan with fixed blades as a function of water flow. This expression allows us to predict the generation of active power of each of the turbines, depending on the adjusted water flow rate, which varies in the seasonal, monthly and daily periods of the year. It also allows you to solve the inverse problem, when it is necessary, depending on the load schedule dictated by the electrical system to which these small hydroelectric power stations are connected, to regulate the amount of water flow by changing the opening angles of the guiding devices of the hydraulic turbines. Studies on the proposed mathematical model containing a synchronous generator with electromagnetic excitation coupled to a Francis hydraulic turbine with a 60% change in water flow from $q=1$ (r.u) to $q=0.4$ (r.u). The value of the required active power of a small hydroelectric power station, i.e. power output of the generator will jam (either by the dispatcher, or by the value of the load schedule), i.e. it is an input value, and as the output value of the controller is the opening angle of the guiding apparatus of the turbine.

A COMPOUND OF GAMMA AND SHANKER DISTRIBUTION 87

Mousumi Ray, Rama Shanker

This paper considers a new lifetime distribution called Gamma-Shanker distribution which is a compound of Gamma and Shanker distribution. Many important properties of the suggested distribution including its shape, Inverse moments, hazard rate function, reversed hazard rate function, quantile function and stress-strength reliability have been discussed. The estimation of its parameters has been discussed using maximum likelihood estimation. Goodness of fit of the proposed distribution has been explained with two examples of real lifetime data from biomedical sciences and it shows that the proposed distribution gives much closure fit over the considered distributions.

DIFFERENT ESTIMATION METHODS FOR TOPP-LEONE-(A) MODEL: INFERENCE AND APPLICATIONS TO COMPLETE AND CENSORED DATASETS 100

Adubisi O. D., Adubisi C. E.

This work proposes a new two-parameter model, titled Topp-Leone (A) model. The main benefit of the new model is that it has an inverted bathtub shaped curve, increasing and decreasing hazard rate function quite dependent on the shape parameter. Its structural properties including the ordinary moments, quantiles, probability weighted moment, median, entropy and order statistics are derived. More so, the survival, failure rate, reversed failure rate and cumulative failure rate functions are also derived. Six classical estimation methods are discussed for estimating the parameters of the new model. Monte Carlo experiments and real datasets analyses are conducted to examine the classical estimators performance of this model. Finally, the usefulness of the Topp-Leone (A) model demonstrated with different applications to complete and type-II right censored data proves its more flexible when compared to well-known models in statistical literature.

RELIABILITY AND AVAILABILITY ANALYSIS OF AN 8-STEP AUTO UNIT MANUFACTURING PROCESS HAVING A FAULT IN MAINTENANCE 118

Surabhi Sengar

Human plays a vital role in the manufacturing process of a product during the planning, design, assembly, production, and maintenance phase. This paper presents a systematic method of determining the reliability and availability of an 8-step auto unit manufacturing system taking into consideration the case of occurring human error during maintenance. The whole process involves eight major units as Combustion engine, power train unit, fuel feed unit, fuel injection unit, drain or exhaust unit, engine cooling unit, brake unit, and body frame. With the failure of any of these units, the whole process can fail. Also, a constant failure rate and a general repair time are taken into the consideration for each operative unit that makes up the process. The integral differential equations were generated based on Markov modeling of the process and solution derived by considering repair time distribution using the Laplace transform. Calculations of different reliability aspects like process availability, mean time to failure steady-state nature, and profit analysis are done using supplementary variable technique and copula methodology.

ESTIMATION OF A PARAMETER OF FARLIE-GUMBEL-MORGENSTERN BIVARIATE BILAL DISTRIBUTION BY RANKED SET SAMPLING 129

M. R. Irshad, R. Maya, A.I. Al-Omari, Ahmad A. Hanandeh, S. P. Arun

A bivariate version of the Bilal distribution has been proposed in the literature, called the Farlie-Gumbel-Morgenstern bivariate Bilal (FGMBB) distribution. In this article, we have dealt with the problem of estimation of the scale parameter associated with the study variable Z of primary interest, based on the ranked set sample defined by ordering the marginal observations on an auxiliary variable W, when (W, Z) follows a FGMBB distribution. When the dependence parameter is known, we have proposed the following estimators, viz., an unbiased estimator based on the Stokes ranked set sample and the best linear unbiased estimator based on the Stokes ranked set sample for the scale parameter of the variable of primary interest. The efficiency comparison of the proposed estimators with respect to the maximum likelihood estimator have been carried out.

PROCESS MODELING AND NUMERICAL INVESTIGATION OF VENEER CUTTING SYSTEM OF A PLYWOOD PLANT WITH STOCHASTIC APPROACH 141

Dr. Subhash Malik, Dr. Narendra Kumar, Er. Sudhir Kumar

Present paper covers performance modelling and performability evaluation of a veneer-cutting system of plywood manufacturing industry. The performability is evaluated as a function of availability. In this system the different subsystems are connected in hybrid mode. Markovian Approach was used for developing the process modeling of the subsystems and to evaluate the performability of said system. MATLAB software was used to perform the numerical computations as well as simulation of results. The current work examines the impact of varied failure rates and repair rates on the long-term availability of the system. A Particle Swarm Optimization (PSO) based technique was used to optimize the results. A Decision Support System (DSS), which can be helpful for making strategic decisions on financial investments in managing the maintenance priorities, spare part management, and human resource requirements, among other things, has been recommended based on the numerical investigation.

A NEW CONTROL CHART FOR PROCESS DISPERSION BASED ON RANKED SET SAMPLING 154

Chandrakant Gardi, Vikas Ghute

In this paper, we propose a new control chart based on Downton's estimator (D) for monitoring the process dispersion using ranked set sampling design and runs rules. The performance of the proposed control chart is compared with the originally proposed D chart based on simple random sampling method when underlying process distribution is normal and non-normal. The average run length is used to evaluate the performance of the proposed control chart. It is observed that the proposed control chart is efficient in detecting shifts in process dispersion as compared with the chart using simple random sampling method. The performance of the proposed chart is further investigated using runs rules. The efficiency of the designed runs rules RSS-D chart is compared with its existing counterparts and is found to be superior.

**MODELING THE RELIABILITY OF TRANSPORT UNDER
EXTREME CONDITIONS OF OPERATION AS
A QUEUING SYSTEM WITH PRIORITIES 167**

M.D. Katsman, V.I. Matsiuk, V.K. Myronenko

The article presents a simulation model of a queuing system (QS) with a queue and relative priority, which can be used to manage the reliability of transport systems under resource constraints. The developed simulation model combines agent and discrete-event simulation principles and allows studying queuing systems in terms of establishing regularities: probabilities (service, failure, push-out), time delays (waiting in a queue, under service), queue sizes, order of queue formation upon arrival of consumers of different priority. As a result of the research, dependencies were obtained for the probability of servicing higher priority consumers depending on the intensities of their arrival and service; probabilities of servicing lower priority consumers depending on the intensity of service and servicing of higher priority consumers; the probability of "pushing out" lower priority consumers from the QS by higher priority consumers depending on service intensities and the arrival of high priority consumers.

**SINGLE SAMPLING PLANS FOR VARIABLE INSPECTION BASED ON
BIRNBAUM – SAUNDERS DISTRIBUTION 180**

Saranya C R, Vijayaraghavan R, Sathya Narayana Sharma K

Sampling inspection by the method of variables is a well-known category of product control in which the discretion of acceptance or rejection is made hinge on some specific rule, which is framed according to the measurement of a quality characteristic under study. In this scenario, the quality feature under study is considered a continuous random variable that can be demonstrated using any continuous type statistical distribution. Birnbaum and Saunders distribution is a continuous probability distribution, which is having numerous applications in various fields. Application of Birnbaum and Saunders distribution is considered in this paper for establishing acceptance sampling inspection plans for variables based on the examination of units from a single sample. Numerical illustrations are given to demonstrate the application of proposed sampling plan. In addition, the results of numerical illustrations are explained with the help of simulated data.

**OPTIMAL SPARE-SWITCHING TIMES IN SERIES SYSTEMS UNDER
A GENERAL FRAMEWORK..... 191**

Hamideh Jeddi, Mahdi Doostparast

Spare parts are commonly used to improve system performances. They are allocated to original components during system missions. The optimal allocations depend on system configurations and lifetimes of components and spares. Various methods for finding optimal allocations have been proposed in the literature. For sake of brevity, lifetimes of components are commonly assumed to be independent. This paper deals with series systems, a common configuration, under a general setting, i.e. component lifetimes are dependent and heterogeneous. Moreover, the spare is also allowed to switch among original components to impose more flexibility for spare managements. This allowance occurred usually in network servers and electrical generators which manage by a dispatching center. Explicit expressions for system reliability functions are derived in detail. Since system lifetimes are random phenomena, stochastic orders are utilized for comparison purposes. Various illustrative examples are also given.

DESIGNING OF ACCEPTANCE SAMPLING PLAN BASED ON PERCENTILES FOR TOPP-LEONE GOMPERTZ DISTRIBUTION 200

SJayalakshmi, Aleesha A

Acceptance sampling is a statistical technique used to inspect the quality of a batch of products. An acceptance sampling plan under which sampling inspection is performed by conducting life test upon the sampled products is termed as reliability sampling plan. In this paper, a single acceptance sampling plan based on percentile is presented for Topp-Leone Gompertz (TL-G) distribution when the life test is truncated at a pre-specified time. The minimum sample size necessary to ensure the specified life percentile is obtained under a given consumerCs risk. The operating characteristic values (and curves) of the sampling plans as well as the producerTs risk are presented.

BAYESIAN AND CLASSICAL ESTIMATIONS OF TRANSMUTED INVERSE GOMPERTZ DISTRIBUTION 207

T. M. Adegoke, K.O. Obisesan, O. M. Oladoja, G.K. Adegoke X

In this article, the use of the transmuted inverse Gompertz distribution in modeling lifetime data is investigated particularly in cases where standard probability distributions are not able to properly handle complex datasets. The quadratic rank transmutation map scheme is utilized to obtain the distribution. The study explores several characteristics of the transmuted inverse Gompertz model, including the estimation of parameters through classical approaches such as maximum likelihood estimation, least-squares estimation, Crammer-Von Misses estimation, and maximum product spacing estimation. Additionally, the Bayesian techniques is used under different loss functions, such as the Linex loss function, square error loss function, and general entropy loss function. The estimates obtained from both classical and Bayesian techniques are evaluated using simulation. To illustrate the potential benefits of the transmuted inverse Gompertz model, a dataset on the strength of aircraft window glass is employed. The results obtained from the application of the new distribution to the real-life dataset demonstrate that it yields superior fits in comparison to other well-known distributions. The studyTs findings suggest that the transmuted inverse Gompertz model can provide a useful alternative for modeling lifetime data. The research offers valuable insights into the distributionTs properties and estimation techniques, as well as its superiority over other commonly used distributions. The new model can contribute to the development of more accurate and efficient models for analyzing lifetime dataset. Overall, the study highlights the importance of exploring new statistical models and techniques to improve data analysis and decision-making.

ON THE BAYESIAN MODELING OF ROAD ACCIDENTS DUE TO VEHICLE TYPE IN OYO STATE NIGERIA 223

Oladapo M. Oladoja, K.O. Obisesan, Sururat A. Adebayo, T. M. Adegoke

Transportation plays an important role in the day to day activities of human race. Road, rail, water and air are the four major forms of transportation in Nigeria. Road transportation is the most common means of movement in Oyo State and this has raise the likelihood of the occurrence of road traffic accident even posing a serious problem that needs serious attention. Road traffic accidents would Continually increase if tangible efforts are not made to tackle is problem. In Oyo state, the epicenter of the western Nigeria, road transportation is the most popular means of transportation. Of concern is the accidents recorded daily on the roads. This study takes into account the series of accidents caused on the state roads due to vehicle types. Seven vehicle types were considered due to the data available. Bayesian Model Averaging (BMA), a variable selection approach, was used to handle uncertainty in the model selection process. Several classical approach like time series have been used in analysing road accidents but few had explored the Bayesian approach via BMA route. A uniform model prior was used with a numeric g-prior to improve the predictive performance. The trend of accidents were observed for a period of 15 years from (2006 -2020). 78% of the models were visited and using the Posterior Inclusion Probability results

(99.546%), it was seen that the accidents that occurred via private cars was the predominant. It is the most important in modeling vehicle type accidents in Oyo State Nigeria. Also, accidents for 2023 and 2024 were predicted using the posterior predictive distribution. It is imperative for concerned authorities (Federal Road Safety Corps amongst others) in the state to look into the issue of private car owners in the state.

THE APPLICATION OF QUEUEING THEORY FOR MAXIMIZING SYSTEM SIZE USING ENCOURAGED ARRIVAL 233

S. Immaculate, P. Rajendran

Customers are frequently drawn in by lucrative deals and discounts offered by businesses. These interested customers are referred to as encouraged arrivals. The major goal of this study is to evaluate the performance of the automobile assembly line in order to decrease waiting time by coordinating the activities at each workstation using stopwatch time study approach. The novelty of this research is to convert poisson arrival to encouraged arrival with some discounts (like, 10%, 20%). The queuing problem is represented by the notation $M/M/1:FCFS/./.$ in Kendall's notation. It is a single channel, multi-server service with infinite system capacity and an infinite number of calling population. Data concerning the system's encouraged arrival and service distribution were established. These data were used to calculate the system performance parameter. The finding of the study was used to predict the system's performance and effectiveness and to make logical recommendations for possible future improvements. According to the results, it is possible to conclude that increasing the level of automation reduces part waiting time, decreasing the cost of waiting. When compared to the poisson arrival system, the size of the Markovian encouraged arrival queuing system is increased as shown in the table. Little's law is verified that system size and queue size is same as in length. Little's law is used to predict lead time based on production rate and work-in-process. Here it is verified as shown in table.

RELIABILITY ANALYSIS OF THIN -WALLED PRESSURE VESSELS USING ADVANCED FIRST ORDER SECOND MOMENT (AFOSM) METHOD 244

K. Sunitha, T. Sumathi Uma Maheshwari, M. Tirumala Devi

Pressure vessels are highly used in commercial purposes and industries such as boiling, softening and hot water storing tanks. Pressure vessels are subjected to its internal and external pressure. In this paper, thin-walled pressure vessels are taken for analysis. Pressure, radius, thickness and strength of the material are considered as random variables. The random variables follow normal distribution. The reliability index of the pressure vessel made with different materials such as 6061 aluminum alloy and SA 516 70 stainless steel has been found. Reliability analysis has been done for the pressure vessel by using AFOSM with MATLAB. It is observed that strength of the materials influences more on reliability of the vessels.

WEIGHTED INTERVENED EXPONENTIAL DISTRIBUTION AS A LIFETIME DISTRIBUTION 253

Vilayat Ali Bhat, Sudesh Pundir

This study proposes and investigates the weighted intervened exponential distribution, which is demonstrated as a generalized extension of the intervened exponential distribution. The form of the weighted intervened exponential distribution is obtained by considering a specific non-negative weighted function. The probability density function and cumulative density function of the proposed model are given, and its generalized form of reliability function and the hazard rate function is also derived. By choosing a different set of parametric values, the graphical demonstrations of the probability density function of weighted intervened exponential distribution are given where it acquires different curve shapes. The weighted intervened exponential distribution density

function is then further studied in the limited form as a special case called the length-biased intervened exponential distribution. Along with the distribution of order statistics, stochastic ordering, stress-strength reliability, and entropy measure, several distributional and reliability aspects of the length-biased intervened exponential distribution are derived. For estimating the unidentified parameters of the length-biased variant, the most suggested approach known as the maximum likelihood estimation technique is implemented. To explore the behavior of the parameter estimates for various sample sizes, a sample data generation technique is required to carry out the process. Since the quantile function of the length biased intervened exponential distribution is not in closed form. So, the alternative data generation algorithm is employed which is known as the acceptance-rejection algorithm technique, and a Monte-Carlo simulation study is done. The absolute average bias and mean square error of the estimated parameters of the length-biased version model are calculated and it is noticed that both the calculated measures decrease simultaneously on increasing the sample size. In order to determine if the model is appropriate, a real-life time-to-event data set is examined as an example, and length biased distribution is juxtaposed with several other common available lifetime distributions for comparison purposes.

**TRADE CREDIT FINANCING SCHEME ON RETAILERS
ORDERING QUANTITY FOR IMPERFECT QUALITY ITEM WITH
LEARNING EFFECTS AND STOCKING STRATEGIES 267**

A. R. Nigwal, U. K. Khedlekar, N. Gupta, L. Sharma

Today's life is a age of modern life, and in the modern life for any kind of business setup, customer service, pricing, stocking strategies and trade credit financing schemes are effective, essential and survival parameters to grow the business. In this paper, we have developed, an economical order quantity model for imperfect quality product by considering retailer's stock sensitive demand of product under trade credit financing policy. Further in this paper we have studied the Learning effect on screening process on every batch of imperfect quality product. Under the trade credit financing scheme, we have considered that, the supplier proposes to the retailer, a fixed credit time period for payment and retailer also offers to his customers to a fixed credit time period of payment. Finally an appropriate total profit function per unit time has been derived under the various trade credit financing periods of payment including various expenditure and other related parameters. A sensitivity analysis has been done to verify the optimum results and also a numerical example has been given to verify the model's outputs.

**A NOVEL METHOD TO GENERATE A FAMILY OF
BATHTUB-SHAPED FAILURE RATES FROM A FAMILY
OF UPSIDE DOWN BATHTUB-SHAPED FAILURE RATES
AND VICE-VERSA 286**

R.L. Giri, Subarna Bhattacharjee, Suchandan Kayal, S. K. Misra

It is indeed a matter of great significance for system engineers and scientists to derive new classes of lifetime distributions for providing a better statistical model which will fit a given lifetime data set. It is known that many real time data have varied characteristics and can be modeled by distributions with bathtub and upside down bathtub failure rates viz., Weibull, Modified Weibull, Inverse Weibull. This paper proposes a method which generates a family of distributions having bathtub (BT)-shaped failure rate from a distribution having upside down bathtub (UBT)-shaped failure rate and vice-versa. The proposed method is validated with the help of a few statistical distributions. The closure properties of the proposed model under various reliability operations are studied.

**AN APPROACH FOR ASSESSMENT OF RELIABILITY INDICES
CONSIDERING OMISSION OF FIXED REPAIR TIME FOR ELECTRIC
TRACTION SYSTEM APPLYING MONTE CARLO SIMULATION..... 298**

Aditya Tiwary

Assessment of numerous reliability indices is essential when availability and unavailability of supply in any electric power system is talked about. The reliability index which are very important for overall performance of any complex engineering system are mean up time, mean down time and unavailability. In this paper, assessment of various reliability indices for the electric traction system is done based on Monte Carlo simulation. If an engineering system fails then its repair is required to be performed at proper time interval. Omission of a threshold value of fixed repair time will not have so much impact on the overall reliability of the engineering system taken together. Electric traction system is very important as it is utilized for operation of passenger trains and freight trains across a large rail network throughout the continents. In view of above, modified values of mean up time, mean down time and unavailability have been obtained accounting fixed repair time omission for the electrical traction system taken under consideration.

**BAYESIAN INTERVAL ESTIMATION FOR THE PARAMETERS OF
POISSON TYPE LENGTH BIASED EXPONENTIAL CLASS MODEL 307**

Rajesh Singh, Preeti A. Badge, Pritee Singh

In this research paper, two sided Bayesian interval is proposed for Poisson type length biased exponential class software reliability growth model. The failure intensity function, mean time to failure function and likelihood function are derived. Bayesian interval estimation has been done for the parameters using non informative priors. The performance of proposed Bayesian interval is obtained by using Monte Carlo simulation technique. Average length and coverage probability of Bayesian interval for the parameters are calculated. From the obtained intervals it is concluded that Bayesian interval of parameters perform better for appropriate choice of execution time and certain values of parameters.

**RESILIENCE OF A TELECOMMUNICATIONS NETWORK
SUBJECTED TO CORRELATED GEOGRAPHICAL FAILURES 315**

Dora Jimenez, Abigail Medina

Due to the COVID-19 pandemic, the way in which routine activities are carried out changed, taking a leading role telecommunications networks, it is important to evaluate their operation, service interruption and progressive deterioration, especially those generated by natural disasters, we have focused on earthquakes. Venezuela is a seismic country, being vulnerable to economic and human losses caused by this disaster. Many of the infrastructures that are used by both public and private institutions may not follow the current laws on earthquake-resistant structures. We seek to evaluate the damage by correlated geographic faults produced by earthquakes in a telecommunications network using probabilistic seismic risk analysis.

A NEW ZERO-TRUNCATED DISTRIBUTION AND ITS APPLICATIONS TO COUNT DATA 327

Na Elah, Peer Bilal Ahmad, Muneeb Ahmad Wani

Numerous disciplines, including engineering, public health, sociology, psychology, and epidemiology, are particularly interested in the analysis and modelling of zero truncated count data. As a result, we suggest a novel and straightforward structural model in this study called zero truncated new discrete distribution. We examine its statistical properties including probability mass function, cumulative function, and moments. The parametric estimation of the zero-truncated new discrete distribution is explained by Maximum Likelihood Estimation method and, to investigate its performance, a simulation study is proposed. The importance of the distribution is evaluated using two real-world data sets as well as one simulated data set and the model comparison is made on the basis of AIC and BIC criterions.

USE OF TRIBODIAGNOSTICS IN PRACTICE 340

Pavol Mikus, Alena Breznicka

Tribodiagnosics deals with the problems of lubrication, friction and analysis of oils in technical fluids. Based on the results of parameter monitoring and chemical analysis of the oil, it is possible to determine the impending failure of the entire system very accurately. Today, this relatively young field of technical diagnostics is gradually becoming very viable and its results are fully in line with classical vibroacoustic diagnostics or thermodiagnosics. It is used in all mechanical systems containing oil systems. This is one of the methods of non-disassembly technical diagnostics, which is based on the knowledge that the lubricant after a certain period of use in the lubrication system reflects the condition of the equipment and the conditions in which this equipment was operated.

MOMENTS PROPERTIES OF CONCOMITANTS OF GENERALIZED ORDER STATISTICS FROM FGMTBM EXPONENTIAL DISTRIBUTION 348

Mustafa Kamal, Nayabuddin, Intekhab Alam, Ahmadur Rahman,
Abdul Salam, Shazia Zarrin

Concurrent or induced order statistics are produced when individuals in a random sample are ordered in compliance with the corresponding values of some other random sample. Concomitants are most helpful when $k(n)$ individuals are to be chosen using a selection technique based on their X -values. The relevant Y -values are then used to reflect how well a characteristic has performed. In this paper, concomitants of generalized order statistics (GOS) from Farlie Gumbel Morgenstern type bivariate moment (FGMTBM) exponential distribution are obtained. Additionally, distribution function (df) and probability density function (pdf) r -th generalized order statistics and a joint pdf of r -th and s -th GOS were also obtained. Furthermore, we provide the minimum variance linear unbiased estimator (MVLUE) of the position and scale parameters of the concomitants of the k -th upper record values and order statistics for the distribution under consideration. Finally, an implementation of the suggested methodology has been taken into account.

PROJECT CHARACTERISTICS WITH TRIANGULAR FUZZY NUMBER 359

Adilakshmi Siripurapu, Ravi Shankar Nowpada

The Critical Path Method (CPM) is required to plan, organize, and arrange for major project networks. A clear estimate of the time duration will help in the successful execution of the CPM. However, the time duration cannot be precisely specified in real life. As a result, there is always uncertainty about the duration of activities, leading to the invention of the fuzzy critical path method. This study proposes a simple method for critical path analysis and project characteristics in a project network with a triangular fuzzy number. Furthermore, this study defines the most critical path and the relative path degree of criticality, both theoretically valid and practical. The suggested method can determine the critical path, project characteristics and critical degree of an activity as shown by an example discussed in some earlier studies. The proposed approach is very simple to apply and does not require knowing the explicit form of the membership functions of the fuzzy activity times.

RELIABILITY FACTORS OF SOFTWARE FOR MICROPROCESSOR PROTECTIONS 370

M. I. Uspensky

The features of the program functioning reliability that are typical for critical applications operating in real mode for microprocessor protections are considered. Among the main characteristics of the relay protection functioning are reliability indicators. With the transition to the execution of such protections on a microprocessor-based basis, in addition to hardware reliability, it became necessary to characterize its operation by software reliability. The importance of the solve tasks in the operation process refers it to programs used in safety-critical software and operating in real mode. This, in turn, tightens the requirements for their reliability evaluation. Comparative analysis made it possible to assess the generality of the considered reliability factors for the functioning of the application software under consideration. At the same time, we were able to highlight some of the features that are typical for the implementation of relay protection on microprocessors. An example of such an assessment is given, showing that despite all the difficulties of complete testing for microprocessor protection programs the software erroneous contribution still amounts to 2.5% of the total contribution. It is shown that microprocessor protection programs are related to programs of critical applications operating in real mode, which makes it possible to use the experience and characteristics of such software in solving relay protection challenges. Nevertheless, some features are given that are specific for relay protection based on a microprocessor. Further tasks are defined.

ANALYSIS OF M[X1], M[X2]/G1, G2/1 RETRIAL QUEUE WITH PRIORITY SERVICES, DIFFERENTIATE BREAKDOWN, REPAIR, SYNCHRONIZED RENEGING AND OPTIONAL VACATION 376

G. Ayyappan, S. Nithya

This study deals with the steady-state analysis of single-server retrial non-preemptive priority queue with differentiate breakdown, repair, synchronized renegeing and optional vacation. For this purpose, two categories of customers are considered, priority and ordinary customers, who arrive as per Poisson arrival process. The server consistently affords single service for these customers based on general distribution. The server randomly fails while providing service to the customer. Hard failure and soft failure are the two kinds of system failure. Hard failure is defined as an equipment failure that requires a repairman with specialized knowledge to be physically present, which is a time-consuming process. Whereas soft failure is defined as failure caused by events rather than physical condition and is usually resolved rebooting the system. Ordinary customers may renege the orbit if the server is engaged or unavailable. Furthermore, once the service of all priority customers is completed by the server, the server goes for a vacation or becomes idle. In this study, we used probability generating function and supplementary variable technique to solve the Laplace transforms of time-dependent probabilities of system states. Finally, we evaluated performance measures and expressed the results in numerical values.

MULTI-COMPONENT CONDITIONAL STRESS-STRENGTH PARAMETER ... 392

Kavoos Khorshidian, Morteza Taheri Saif Abad

There are situations in which the experimenter has some information about the components of the operating system and he/she wants to use this information for better assessment or operating of the underlying system. In such cases the notion of conditional probability may help the operator to use that information and improve his/her task. In the present study this notion has been examined, and some conditional stress-strength parameters have been introduced for s of k systems. The multi-component conditional stress-strength parameter (MCCSSP) and its maximum likelihood estimator have been calculated when the strength and stress random variables are exponentially distributed. In the case of having extra information about the parameters, a closed form has been derived for the Bayes estimator of MCCSSP and has been calculated by using an algorithm together with Monte Carlo method. For the case of non-exponential stress or strengths, the nonparametric estimator of the defined parameter has also been derived. Finally, some simulation study on the MLE and Bayes estimator, as well as real data analysis for nonparametric estimators have been done to verify the analytic results.

BAYESIAN ANALYSIS OF ARMA AND BSTS MODELS FOR COVID-19 DATA USING R AND STAN 406

Muhammed Navas T, Athar Ali Khan

This study compares the performance of Bayesian ARIMA and BSTS models for COVID-19 data using Bayesian approach. Many studies in the literature have compared the BSTS model and classical ARIMA models for infectious disease modelling, and the BSTS model performs well. Apart from the literature, this study is trying to prove the Bayesian ARIMA model gives a better result than the BSTS model. This study uses a different modelling and model comparison method to compare widely used autoregressive integrated moving average (ARIMA) models with their Bayesian structural time series (BSTS) models for COVID-19 data using the Bayesian approach. It is essential to find the order of the ARIMA model before doing bayesian analysis. We find the order of the ARIMA model using measurement LOO information criteria, using the Hamiltonian Montecarlo algorithm and rstan estimate the parameters of ARIMA and BSTS models for COVID-19 data. Furthermore, compare both models using Looic and Waic values; Bayesian ARIMA models outperform in this study.

ANALYSIS OF PROJECTED PROFIT IN AN M/M/K ENCOURAGED ARRIVAL QUEUEING MODEL USING CHI-SQUARE TEST 417

V. Narmadha, P. Rajendran

Nowadays, queues be seen in fast food restaurants and in all service-based businesses. This study is a mathematical analysis of such business firms with the help of Queueing Theory. The discounts and promotions entice customers to the firm and in this study such attracted customers are referred to as Encouraged Arrivals. The Chi-square test is used to determine the kind of encouraged arrival pattern that adheres to the data observed from a fastfood outlet. We introduce the encouraged arrivals in an M/M/k queueing model for the analysis of performance metrics. The performance metrics of the various encouraged arrival patterns are compared and the ideal one is chosen for the firm. The economic analysis shows that with encouraged arrivals, the cost associated with the time lost due to waiting is reduced gradually with increasing number of servers. Thus the firm increases its projected profit with encouraged arrivals. This study helps the entrepreneurs to decide the kind of discounts that would attract the customers simultaneously improving the firm's profit. Little Tslawis also verified.

MOMENTS OF GENERALIZED RECORD VALUES FROM MODIFIED FRECHET DISTRIBUTION AND ITS CHARACTERIZATION 429

Zaki Anwar, Abdul Nasir Khan, Rafiqullah Khan

The aim of this paper is to introduce the relations for moments and characterizing results for the newly introduced modified Fréchet distribution based on generalized record values. Here, we used an ordered random variable approach like generalized record values for generating the results. We have established the recurrence relations for single and product moments of generalized record values from modified Fréchet distribution. These relations are also deduced for the lower record values and some specific distributions, which are the special cases of modified Fréchet distribution. Further, the characterization results for this distribution have been established by using recurrence relations for single and product moments and conditional expectation of a function of generalized record values and truncated moments.

PROFITABILITY ANALYSIS OF A FOOD INDUSTRIAL SYSTEM HAVING MAKE-AND-PACK PRODUCTION STRATEGY WITH PRIORITY BASIS REPAIR 441

Monika, Garima Chopra

The present investigation is concerned with the profitability analysis of a food industrial system where production is based on the make-and-pack strategy. The system is assumed to have two subsystems: first subsystem is for making while the second is for packing the product so formed. As per the gathered information about the production procedure in the food industrial plant, the priority of repair is given to the making subsystem over the packing subsystem. Here, failure of either subsystem leads to a complete breakdown of the system. Also, two types of failures are considered in the packing subsystem i.e. minor failures and major failures. Two kinds of repairers (operator and fitter) are appointed to tackle the failures in the subsystems. For minor and major failures in the packing subsystem, the operator and fitter respectively are responsible for repairs. However, any failure in making subsystem is repaired by the combined efforts of the operator and fitter. Reliability characteristics such as mean time to system failure (MTSF), system availability, and expected busy period of the repair persons are studied by employing the semi-Markov process and regenerative point technique. The system profitability is graphically analyzed concerning to failure rates of both subsystems.

HYBRID AND BLIND WATERMARKING FRAMEWORK FOR PRIVACY PROTECTION AND CONTENT AUTHENTICATION OF DIGITAL MULTIMEDIA 456

Swati J. Patel, Dr. Mehul C. Parikh

Nowadays, due to inexpensive and conveniently available internet access at the fingertip, the illegitimate sharing of digital multimedia i.e. image, audio, and video is becoming a universal and significant threat. Illegal transmission of digital multimedia through the internet creates an issue of authentication and copyright protection; hence, piracy protection is a vital need for protecting digital media. Digital watermarking is a method of preventing digital theft in which additional information, known as a watermark, is inserted into digital multimedia. This technology was originally designed for still photos, but it has subsequently been expanded to include additional multimedia artifacts such as audio and video, due to its countless use in today's era. Digital watermarking is an effective method of limiting piracy and providing authenticity and copyright ownership to digital content. Watermarking can be performed either in the spatial or in the transform domain. In this paper, a hybrid digital video watermarking technique for copyright protection, data security, and content authentication of multimedia, based on Discrete Wavelet Transform, Discrete Cosine Transform, and Singular Value

Decomposition is presented. The authenticity of the content has been ensured by embedding a watermark in the transform domain, while copyright protection has been provided by a strong watermark. The experimental results show that the proposed schemes achieve a PSNR greater than 51 dB on average, which illustrates that the proposed method gives excellent performance for robustness, authentication, and security. A comparison of the proposed framework to various cutting-edge techniques illustrates its effectiveness and superiority.

EMPIRICAL STUDY ON ROBUST REGRESSION ESTIMATORS AND THEIR PERFORMANCE 466

Lakshmi R, Dr.Sajesh TA

Regression Analysis is statistical technique to model data. But the presence of outliers and influential points affect data modelling and its interpretation. Robust regression analysis is an alternative choice to this. Here we made an attempt to study different robust estimators and propose a new robust reweighted Sn covariance based regression estimator. We have evaluated the performance empirically and the simulation study shows our proposed estimator is preferable to OLS and other robust regression estimators in terms of the MSE criteria. Also, proposed robust Sn covariance regression estimator produce outperforming results for regression equivariance and breakdown criterion. Robustness of the proposed estimator is proved empirically. The proposed method is innovatively used to model fluid data. R software is used for simulation and study.

RATIO ESTIMATOR OF POPULATION MEAN USING A NEW LINEAR COMBINATION UNDER RANKED SET SAMPLING 479

Saba Riyaz, Khalid Ul Islam Rather, Showkat Maqbool, T. R. Jan

Ranked set sampling is an approach to data collection originally combines simple random sampling with the field investigator's professional knowledge and judgment to pick places to collect samples. Alternatively, field screening measurements can replace professional judgment when appropriate and analysis that continues to stimulate substantial methodological research. The use of ranked set sampling increases the chance that the collected samples will yield representative measurements. This results in better estimates of the mean as well as improved performance of many statistical procedures. Moreover, ranked set sampling can be more cost-efficient than simple random sampling because fewer samples need to be collected and measured. The use of professional judgment in the process of selecting sampling locations is a powerful incentive to use ranked set sampling. This paper is devoted to the study, we introduce an approach to the mean estimators in ranked set sampling. The amount of information carried by the auxiliary variable is measured with the on populations and samples and to use this information in the estimator, the basic ratio and the generalized exponential ratio estimators are as an improved form of a difference cum exponential ratio type estimator under the ranked set sampling in order to estimate the population mean of study variate Y using single auxiliary variable X. The expressions for the mean squared error of propose estimator under ranked set sampling is derived and theoretical comparisons are made with competing estimators. We show that the proposed estimator has a lower mean square error than the existing estimators. In addition, these theoretical results are supported with the aid of some real data sets using R studio. Therefore, Under RSS architecture, a better difference cum exponential ratio type estimator has been suggested. The estimator's mathematical form has been developed, and its efficiency requirements have been developed in relation to various already-existing estimators from the literature. By imputing various values for the constants used in the creation of our proposed estimator, we also provide several specific situations of our estimator.

**COMBINED REDUNDANCY OPTIMIZATION
FOR A SYSTEM COMPRISING OPERATIVE, COLD
STANDBY AND WARM STANDBY UNITS..... 486**

Lalji Munda, Gulshan Taneja

The company or industry can increase system reliability and provide stress-free operation by adding redundant equivalent subsystems to the active unit. The warm standby system is accessible if the operational unit malfunctions, and the cold standby system can take over. This paper aims to analyze a system comprised of one operative unit, cold standby unit, and one warm standby unit. Cold standby is activated to become warm standby when the operative unit fails, and warm standby becomes operational immediately. A minor defect causes the warm standby unit to fail, whereas a major fault causes the operational unit to fail. Such systems are used by many businesses, sectors, and facilities to prevent operational and reputational losses. Cut-off values for the failure rate, activation rate, revenue cost, and cost per repairman visit have been calculated to determine when the system is profitable. Various system performance measures have been defined by using the Markov process and regeneration point method.

**A TWO-STATE FEEDBACK RETRIAL QUEUEING SYSTEM
HAVING TWO HETEROGENEOUS PARALLEL SERVERS
AND IMPATIENT CUSTOMERS 498**

Neelam Singla, Harwinder Kaur. X

Objective: In the present paper we consider a two-state retrial queueing system with feedback having two heterogeneous parallel servers and impatient customers. Transient state probabilities for exact number of arrivals and departures from the system will be obtained when both, one or none of the servers is busy. Numerical and graphical solutions will also be obtained. Methods: The difference-differential equations governing the system are solved recursively, Laplace transform is then used to obtain the transient state probabilities for exact number of arrivals and departures from the system. Findings: Time dependent probabilities are obtained when both, one and none of the servers is busy. Numerical and Graphical solutions are also obtained using MATLAB programming. Novelty: In past research, models considered arrivals and departures from the orbit whereas in present model arrivals and departures from the system are studied along with the concept of feedback. Applications: This type of model is implemented in computer systems.

**ANALYSIS OF A NON-IDENTICAL COMPONENT
STRENGTHS SYSTEM BASED ON LOWER RECORD DATA 513**

Amal S. Hassan, Doaa M. Ismail, Heba F. Nagy

In engineering applications and reliability literature, stress-strength models play a crucial role. The goal of this study is to develop more accurate stress-strength models by addressing the reliability estimation in multi-component systems with non-identical component strengths and stress. In the context of lower record values, the system's reliability is assessed using both classical and Bayesian approaches. In classical estimation, the maximum likelihood estimator of the reliability function is constructed, and a simulation study based on measurements of precision is used to assess the behavior of various estimates. The Bayesian estimators of reliability under general entropy, logarithmic and precautionary loss functions are computed. The suggested Bayesian estimates are calculated using the Markov Chain Monte Carlo method through a simulation study because there is no one particular way to do it. We found through simulated research that the accuracy of measurements decreases as the number of records rises. The theoretical results are validated using an example from actual data sets.

OPTIMAL SOFTWARE RELIABILITY PREDICTION USING CRITERIA WEIGHTS UNDER FUZZY DECISION-MAKING APPROACH 529

H. D. Arora, Anjali Naithani, Surbhi Gupta

Multi Criteria Decision Making techniques often face the challenge of determining criteria weights. The weights of criteria can significantly impact the outcomes of the decision-making process. Therefore, it is crucial to pay close consideration to the objectivity characteristics of criteria weights. Many weighting methods were discussed by various authors and utilized to solve various decision-making complications in Analytical Hierarchy Process (AHP), Entropy method, Weighing Score Method (WSM), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Best worst method (MWM), Vlsekriterijumska Optimizacija I Kompromisno Resenje (VIKOR), Criteria Importance Through Intercriteria Correlation (CRITIC) method, ELECTRE, etc. This research article gives an overview of various weighting strategies that can be used in multi-criteria optimization and proposes a novel approach to determine criteria weights using Pythagorean fuzzy sets to handle uncertainties in the decision maker's preferences for allocating software reliability. The comparative analysis shows that the proposed weighting method has the advantage of being simple and straightforward in comparison to the existing weighting methods. The evaluation confirms that this novel approach is effective enough to determine objective weights.

STRESS-STRENGTH RELIABILITY MODEL UNDER MULTIVARIATE NORMAL SETUP AND ITS APPLICATIONS 545

Babulal Seal, Anirban Goswami

We often see that in a system, the energy is supplied to the system by p_1 sources and its consumed through p_2 sources and the sources are linearly dependent with vector \mathbf{a}' and \mathbf{b}' . The overall representation of the two sets are related to vectors \mathbf{a} and \mathbf{b} , such that they are approximated by $\mathbf{a}'\mathbf{x}$ and $\mathbf{b}'\mathbf{y}$ as in principal component analysis. In this article, a stress strength reliability model $R = \Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$, when \mathbf{x} and \mathbf{y} are distributed independently multivariate normal distribution is proposed, with \mathbf{a} and \mathbf{b} are two known vectors. MVUE and MLE of R are obtained. Through simulation studies, their performances are compared using different measures. The two-sided confidence intervals and lower bounds of R are obtained through exact and asymptotic distribution of maximum-likelihood estimators and using bootstrap procedure. Through simulation studies, the performances of these confidence intervals are empirically checked using their coverage and the accuracy. In this study, we proposed to choose the optimal sample size for an experiment assures an adequate power and level. Finally, we applied these interval estimators to a real data set.

GENERALIZATION OF LENGTH BIASED WEIGHTED GENERALIZED UNIFORM DISTRIBUTION AND ITS APPLICATIONS 565

Jismi Mathew

In this article, a generalization of length biased weighted generalized uniform distribution called Marshall Olkin length biased weighted generalized uniform distribution is introduced and studied. Some of the statistical properties of the new distribution such as hazard rate function, compounding, quantile function, moments, Renyi and Shannon entropies are discussed. The maximum likelihood estimation of the model parameters is done and a simulation study is conducted for confirming the validity of the estimates and also introduced a minification process with respect to the model and explored its sample path behaviour for different combinations of parameters. Further, the stress strength analysis is carried out and the estimate of the reliability is obtained based on a simulation study.

A NOVEL METHOD FOR SOFTWARE BUG REPORT ASSIGNMENT 579

Lukasz Chmielowski, Pavlo Konstantynov, Ryszard Luczak, Michal Kucharzak, Robert Burduk

During the development of software and electronic devices, it is inevitable to make mistakes. In large, developed companies, assigning a request to the right development team or even a department is not an easy task. Often, the creation of software bug reports and assignment to groups is also formalized by appropriate processes. The paper presents a novel method of software bug report assignment to a group of developers or analysts. A specific usage of organizational structure at the company is a key component of the proposed approach. There are presented results from real use application including both machine learning predictions and human decisions. Human predictions are not independent, the issues are raised as to why comparing the results of machine learning models with those of humans may be inappropriate and what factors influence human decisions. The work also covers conclusive research about potential benefits of the application of automated assignment of bug reports.

A NEW TWO PARAMETRIC GENERALIZED DIVERGENCE MEASURE AND ITS RESIDUAL 589

Fayaz Ahmed, Khaliq Baig

In this research article, we proposed a new two-parametric divergence measure and developed its weighted version. We also looked at its properties and specific cases with examples and also obtained some results and bounds for new two parametric weighted generalized divergence measure. With the aid of a numerical example that determine the distribution function and also studied some inequality for the new proposed divergence measure. The known divergence measure is the particular case of our proposed measure. The proposed measure uniquely characterized the distribution function using the proportional hazard rate model (PHRM). Its residual function is also being worked on.

On the TeSU–G family of distributions applied to life data analysis

IDZHAR A. LAKIBUL¹ AND BERNADETTE F. TUBO²

•

^{1,2}Department of Mathematics and Statistics
Mindanao State University - Iligan Institute of Technology
Iligan City, Philippines

¹idzhar.lakibul@g.msuiit.edu.ph, ²bernadette.tubo@g.msuiit.edu.ph

Abstract

This paper derives distributions from the U-quadratic and the T-X family of distributions labeled as the T-extended Standard U-quadratic–G family of distributions or simply, TeSU–G family. In particular, the TeSU–Weibull distribution (TeSU–W) is explored with respect to some statistical properties such as its limiting distribution, moment, mean and variance and moment generating function. Also, the statistical properties of the TeSU–Exponential distribution (TeSU–E) which is a special case of the TeSU–W are also derived. The Weibull and Exponential distributions are mostly used in life data analysis because of its ability to adapt to different situations. Moreover, the formula for the median is derived via a proposed algorithm. Simulation study is conducted to verify the performance of the ML estimates of the TeSU–W distribution for varied sample sizes. Further, real life data analysis reveals that derived extended distribution can provide a better fit than several well-known distributions.

Keywords: U-quadratic distribution, T-X family of distributions, Weibull distribution

1. INTRODUCTION

Classical statistical distribution plays a vital role in many areas of science for describing the behavior of any data as well as for modelling data. But nowadays, due to the complexity of the data, the classical distribution needs to be modified in order to cater the complexity of the data. Up to this time, researchers are working in methodologies on statistical distribution theory in order to solve these types of problems.

In 1985, Azzalini [4] introduced a skewed family of distribution for generating a distribution with additional skewed parameter. Other identified family of distributions are the Marshall-Olkin extended (MOE) family [12] and the exponentiated family of distributions [10].

Moreover Eugene [9] in 2002 introduced a composite method of combining two or more known competing distributions through transformations, like the Gamma generated family [16], the Kumaraswamy–G (Kw–G) family [7], the Beta extended–G family [8], the Exponentiated Generalized family [5], the Kumarsway Marshall-Olkin–G family [1], the Generalized odd log-logistic family [6], the generalized transmuted–G family [13] and the Exponentiated Kumarasway–G class family [15].

This paper derives an extended or modified distribution named as TeSU–G family of distribution and explored a derived model using the Weibull as the baseline distribution. This is named as the TeSU–Weibull distribution (TeSU–W). The statistical properties like its limiting distribution, moment, mean and variance, and moment generating function are derived. Similarly, the properties of the TeSU– Exponential distribution (TeSU–E), which is a special case of the TeSU–W are obtained.

The rest of the paper is organized as follows: The extended Standard U-quadratic (eSU) distribution is derived in section 2; in section 3, the TeSU–G family of distribution is introduced; in section 4, the cdf and pdf of both TeSU–W and TeSU–E distributions are derived using the results in sections 2 and 3. Some statistical properties of TeSU–W are presented in section 5. In section 6, estimates of the TeSU–W parameters via the maximum likelihood estimation is generated. Simulation study is presented in section 7 while the application to real life dataset are discussed in section 8. Finally, some concluding remarks are presented in section 9.

2. THE ESU DISTRIBUTION

This section shows the derivation of the extended Standard U-quadratic distribution (eSU). Consider the special case of the T-X family which was introduced by Alzaatreh [2] in 2013. Accordingly, for any arbitrary baseline cumulative distribution function (cdf) $G(x)$, a new cdf $F(x)$ can be generated using the equation

$$F(x) = \int_0^{G(x)} f(t)dt \quad (1)$$

where $f(t)$ is a probability density function (pdf) of a random variable T with support on the interval $[0, 1]$. Also, consider the Transmuted–G family of distributions introduced by Shaw [14], that is, for any baseline cdf $G(x)$, we can define a new cdf $K(x)$ given by

$$K(x) = (1 + \lambda)G(x) - \lambda G^2(x), \quad (2)$$

where $\lambda \in [-1, 1]$. Note that (2) can be written as

$$K(x) = \int_0^{G(x)} f(t)dt$$

where

$$f(t) = 1 + \lambda - 2\lambda t I_{[0,1]}(t) = (1 - \lambda)f_1(t) + \lambda f_2(t) \quad (3)$$

with pdfs $f_1(t)$ and $f_2(t)$ are given as $f_1(t) = 1 I_{[0,1]}(t)$ and $f_2(t) = 2(1 - t) I_{[0,1]}(t)$, respectively. Hence, $f(t)$ can be written as a mixture of two pdfs with support set on the interval $[0, 1]$.

Now consider the pdf of the U-quadratic distribution. For a random variable T that follows a U-quadratic distribution, the pdf of T is given by

$$l(t) = m(t - n)^2, \quad (4)$$

where $t \in [a, b]$, $a < b$, $a, b \in \mathbb{R}$, $m = \frac{12}{(b - a)^3}$ and $n = \frac{a + b}{2}$.

To standardize equation (4), let $a = 0$ and $b = 1$. Then, equation (4) becomes

$$l(t) = 12\left(t - \frac{1}{2}\right)^2, \quad (5)$$

where $t \in [0, 1]$. Substituting $l(t)$ of (5) in equation (3) for $f_2(t)$ derives the pdf of the eSU-quadratic distribution denoted as $f_{eSU}(t)$ and is given by

$$f_{eSU}(t) = 1 - \lambda + 3\lambda(2t - 1)^2, \quad (6)$$

where $t \in [0, 1]$ and $\lambda \in [-\frac{1}{2}, 1]$.

3. THE TeSU–G FAMILY OF DISTRIBUTION

This section introduces a T-extended Standard U-quadratic (TeSU)– G Family of distribution. Using equation (1) and the pdf of eSU in (6) derives the *cdf* of TeSU–G family of distribution given by

$$F_{TeSU-G}(x) = (1 + 2\lambda)G(x) - 6\lambda G^2(x) + 4\lambda G^3(x), x \in \mathbb{R} \quad (7)$$

with corresponding *pdf*

$$f(x) = g(x)[1 - \lambda + 3\lambda(2G(x) - 1)^2], x \in \mathbb{R} \quad (8)$$

where $\lambda \in [-\frac{1}{2}, 1]$ and $g(x)$ is the *pdf* associated with a baseline *cdf* $G(x)$. Note that, if $\lambda = 0$, the *cdf* of TeSU-G reduces to the *cdf* of the baseline distribution.

4. THE TeSU–WEIBULL AND THE TeSU–EXPONENTIAL DISTRIBUTIONS

This section discusses the derivation of the *cdf* and *pdf* of the TeSU using the Weibull and Exponential as baseline distributions. Suppose that a random variable X has Weibull distribution with *cdf* $G_w(x)$ and *pdf* $g_w(x)$ given, respectively, as follows:

$$G_w(x) = 1 - e^{-\tau x^\beta} \text{ and} \quad (9)$$

$$g_w(x) = \tau \beta x^{\beta-1} e^{-\tau x^\beta}, \quad (10)$$

where $x \geq 0$ and with scale τ and shape β parameters.

The *cdf* of the TeSU–Weibull distribution (TeSU–W) is derived by substituting (9) in equation (7), so that we have

$$F_{TeSU-W}(x) = 1 - e^{-\tau x^\beta} (1 + 2\lambda - 6\lambda e^{-\tau x^\beta} + 4\lambda e^{-2\tau x^\beta}), \quad (11)$$

where $\tau > 0$, $\beta > 0$, $\lambda \in [-\frac{1}{2}, 1]$ and $x \geq 0$ with corresponding *pdf* given as

$$f_{TeSU-W}(x) = \tau \beta x^{\beta-1} e^{-\tau x^\beta} (1 + 2\lambda - 12\lambda e^{-\tau x^\beta} + 12\lambda e^{-2\lambda e^{-2\tau x^\beta}}). \quad (12)$$

Note that the exponential distribution is a special case of the classical Weibull distribution when $\beta = 1$. Thus, when $\beta = 1$, the TeSU–W reduces to the TeSU–Exponential distribution (TeSU–E).

The *cdf* of TeSU–E is given as

$$F_{TeSU-E}(x) = 1 - e^{-\tau x} (1 + 2\lambda - 6\lambda e^{-\tau x} + 4\lambda e^{-2\tau x}),$$

where $\tau > 0$, $\lambda \in [-\frac{1}{2}, 1]$ and $x \geq 0$ with corresponding *pdf*

$$f_{TeSU-E}(x) = \tau e^{-\tau x} (1 + 2\lambda - 12\lambda e^{-\tau x} + 12\lambda e^{-2\lambda e^{-2\tau x}}).$$

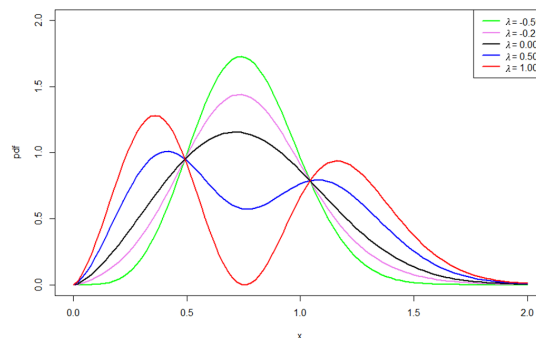


Figure 1: Plots of the *pdf* of the TeSU–W for $\tau = 1.4$, $\beta = 2.5$ and for some values of λ

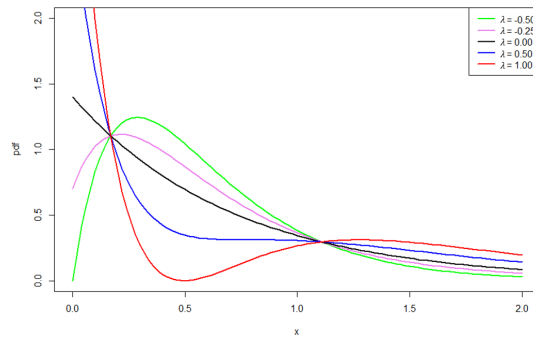


Figure 2: Plots of the pdf of the TeSU–E for $\tau = 1.4$ and for some values of λ

Figure 1 depicts the pdfs of TeSU–W at fixed values of $\tau = 1.4$ and $\beta = 2.5$ with varied values of $\lambda \in \{-0.5, -0.25, 0, 0.5, 1.0\}$. It can be observed that the TeSU–W displays a bimodal distribution when $0 < \lambda < 1$, while when $\lambda = 0$, it depicts the usual shape of the classical Weibull distribution. Moreover, for $-\frac{1}{2} \leq \lambda < 0$, a unimodal distribution which is leptokurtic in nature or a peaked top is observable.

A special case of the TeSU–W is the TeSU–Exponential distribution (TeSU–E), that is, when the parameter β is equal to one. Figure 2 shows the graph of TeSU–E with fixed value of $\tau = 1.4$ and varied values of λ stated previously. The following distribution shapes can be noticed: (1) when $\lambda = 0$, the graph of the TeSU–E is the same as the classical exponential distribution; (2) when $-\frac{1}{2} \leq \lambda < 0$, then it exhibits a unimodal distribution which is positively skewed, and (3) when $0 < \lambda \leq 1$, it follows an inverted skewed bathtub shape. These types of shape are important for describing the complex behavior of the data specially when data distribution reflects a bimodal shape. Hence, the next discussions are focused on the TeSU–W distribution which can cater the bimodal distribution.

5. SOME STATISTICAL PROPERTIES

5.1. Survival and Hazard Functions of TeSU–W

Let X be a random variable with *cdf* given in equation (11) and pdf given in equation (12). Then for $x > 0$, the survival function $S_{TeSU-W}(x)$ and hazard function $h_{TeSU-W}(x) = \frac{f_{TeSU-W}}{S_{TeSU-W}}$ of X are given, respectively, as follows:

$$\begin{aligned} S_{TeSU-W}(x) &= 1 - F_{TeSU-W}(x) \\ &= e^{-\tau x^\beta} (1 + 2\lambda - 6\lambda e^{-\tau x^\beta} + 4\lambda e^{-2\tau x^\beta}) \end{aligned}$$

and

$$h_{TeSU-W}(x) = \frac{\tau \beta x^{\beta-1} (1 + 2\lambda - 12\lambda e^{-\tau x^\beta} + 12\lambda e^{-2\tau x^\beta})}{1 + 2\lambda - 6\lambda e^{-\tau x^\beta} + 4\lambda e^{-2\tau x^\beta}}. \quad (13)$$

Note that if $\beta = 1$, then $S_{TeSU-W}(x) = S_{TeSU-E}(x)$ and $h_{TeSU-W}(x) = h_{TeSU-E}(x)$.

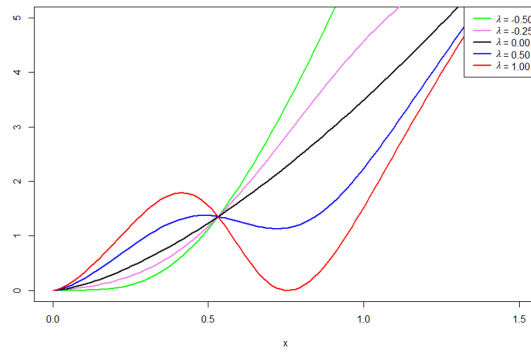


Figure 3: Plots of the $h(x)$ of the TeSU–W when $\tau = 1.4$, $\beta = 2.5$ with varied λ .

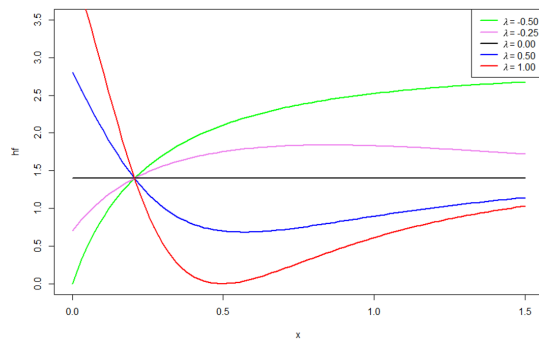


Figure 4: Plots of the $h(x)$ of the TeSU–E for $\tau = 1.4$ with varied λ .

Figure 3 shows that at fixed values of $\tau = 1.4$ and $\beta = 2.5$ and varied values $\lambda \in \{-0.5, -0.25, 0, 0.5, 1.0\}$, the hazard rate function $h(x)$ of the TeSU–W can model not only monotonic but also non-monotonic behavior of the failure rate of the observations, which are inherent in survival lifetime data. Moreover, Figure 4 reveals that the TeSU-E hazard rate function $h(x)$ can model complex data which are either non-monotonic decreasing, increasing or with constant rate.

5.2. The Limiting Distribution of TeSU–W

This section derives the limiting distribution of the probability distribution function (Theorem 1) and the hazard function (Theorem 2) of TeSU–W.

Theorem 1. (i) The limit of the probability density function $f(x)$ of the TeSU-Weibull distribution as $x \rightarrow \infty$ is equal to 0, that is,

$$\lim_{x \rightarrow \infty} f_{TeSU-W}(x) = 0.$$

(ii)

$$\lim_{x \rightarrow 0} f_{TeSU-W}(x) = \begin{cases} \infty & \text{if } \beta < 1 \\ \tau(1 + 2\lambda) & \text{if } \beta = 1 \\ 0 & \text{if } \beta > 1 \end{cases}.$$

Proof. Recall that the pdf of TeSU-W is given in equation (12). It is clear that

$$\lim_{x \rightarrow \infty} f_{TeSU-W}(x) = 0$$

since $\lim_{x \rightarrow \infty} e^{-\tau x^\beta} = \lim_{x \rightarrow \infty} \frac{1}{e^{\tau x^\beta}} = 0$. This proves (i).

To prove Theorem 1 (ii), we have

$$\lim_{x \rightarrow 0} f_{TeSU-W} = \tau\beta \lim_{x \rightarrow 0} x^{\beta-1} \lim_{x \rightarrow 0} e^{-\tau x^\beta} (1 + 2\lambda - 12\lambda \lim_{x \rightarrow 0} e^{-\tau x^\beta} + 12\lambda \lim_{x \rightarrow 0} e^{-2\tau x^\beta}).$$

Observe that for $\beta > 0$,

$$\lim_{x \rightarrow 0} e^{-\tau x^\beta} = \lim_{x \rightarrow 0} \frac{1}{e^{\tau x^\beta}} = 1.$$

It follows that

$$\lim_{x \rightarrow 0} f_{TeSU-W}(x) = \tau\beta(1 + 2\lambda) \lim_{x \rightarrow 0} x^{\beta-1}.$$

If $\beta = 1$, then we have

$$\lim_{x \rightarrow 0} f_{TeSU-W}(x) = \tau(1 + 2\lambda) \lim_{x \rightarrow 0} x^0 = \tau(1 + 2\lambda).$$

Next, for $\beta > 1$ we have

$$\lim_{x \rightarrow 0} f_{TeSU-W}(x) = \tau\beta(1 + 2\lambda) \lim_{x \rightarrow 0} x^{\beta-1} = 0.$$

Lastly, for $\beta < 1$ we get

$$\lim_{x \rightarrow 0} f_{TeSU-W}(x) = \tau\beta(1 + 2\lambda) \lim_{x \rightarrow 0} x^{\beta-1}$$

but $\beta - 1 < 0$ since $\beta < 1$. Also, $\beta - 1$ can be express as $\beta - 1 = -(1 - \beta) = -c$, where $c = 1 - \beta > 0$. It follows that

$$\lim_{x \rightarrow 0} f_{TeSU-W}(x) = \tau\beta(1 + 2\lambda) \lim_{x \rightarrow 0} x^{\beta-1} = \tau\beta(1 + 2\lambda) \left(\lim_{x \rightarrow 0} \frac{1}{x}\right)^c = \infty.$$

■

Theorem 2. The limit of the hazard rate function of the TeSU-W distribution is given by the following:

(i)

$$\lim_{x \rightarrow \infty} h_{TeSU-W}(x) = \begin{cases} 0 & \text{if } \beta < 1 \\ \tau & \text{if } \beta = 1. \\ \infty & \text{if } \beta > 1 \end{cases}$$

(ii)

$$\lim_{x \rightarrow 0} h_{TeSU-W}(x) = \begin{cases} \infty & \text{if } \beta < 1 \\ \tau(1 + 2\lambda) & \text{if } \beta = 1. \\ 0 & \text{if } \beta > 1 \end{cases}$$

Proof. By taking the limit of equation (13) as $x \rightarrow \infty$, we have the following results. It can be verified that

$$\lim_{x \rightarrow \infty} h_{TeSU-W}(x) = \frac{\tau\beta(\lim_{x \rightarrow \infty} x^{\beta-1})(1 + 2\lambda - 12\lambda \lim_{x \rightarrow \infty} e^{-\tau x^\beta} + 12\lambda \lim_{x \rightarrow \infty} e^{-2\tau x^\beta})}{1 + 2\lambda - 6\lambda \lim_{x \rightarrow \infty} e^{-\tau x^\beta} + 4\lambda \lim_{x \rightarrow \infty} e^{-2\tau x^\beta}}.$$

Observe that

$$\lim_{x \rightarrow \infty} e^{-\tau x^\beta} = \lim_{x \rightarrow \infty} \frac{1}{e^{\tau x^\beta}} = 0,$$

then it follows that

$$\lim_{x \rightarrow \infty} h_{TeSU-W}(x) = \tau\beta \lim_{x \rightarrow \infty} x^{\beta-1}.$$

If $\beta = 1$ then

$$\lim_{x \rightarrow \infty} h_{TeSU-W}(x) = \tau.$$

If $\beta > 1$ then

$$\lim_{x \rightarrow \infty} h_{TeSU-W}(x) = \tau\beta(\lim_{x \rightarrow \infty} x)^{\beta-1} = \infty.$$

If $\beta < 1$ then

$$\lim_{x \rightarrow \infty} h_{TeSU-W}(x) = \tau\beta \lim_{x \rightarrow \infty} x^{\beta-1} = 0.$$

This proves (i). The proof of Theorem 2 (ii) is as follows: By definition of the hazard function, we have

$$\lim_{x \rightarrow 0} h_{TeSU-W}(x) = \frac{\lim_{x \rightarrow 0} f_{TeSU-W}(x)}{\lim_{x \rightarrow 0} S_{TeSU-W}(x)}.$$

Observed that,

$$\lim_{x \rightarrow 0} S_{TeSU-W}(x) = (\lim_{x \rightarrow 0} e^{-\tau x^\beta})(1 + 2\lambda - 6\lambda \lim_{x \rightarrow 0} e^{-\tau x^\beta} + 4\lambda \lim_{x \rightarrow 0} e^{-2\tau x^\beta}).$$

But $\lim_{x \rightarrow 0} e^{-\tau x^\beta} = 1$. Hence, it follows that $\lim_{x \rightarrow 0} S_{TeSU-W}(x) = 1$. Thus,

$$\lim_{x \rightarrow 0} h_{TeSU-W}(x) = \lim_{x \rightarrow 0} f_{TeSU-W}(x).$$

By Theorem 1, we have

$$\lim_{x \rightarrow 0} h_{TeSU-W}(x) = \lim_{x \rightarrow 0} f_{TeSU-W}(x) = \lim_{x \rightarrow 0} f_{TeSU-W}(x) = \begin{cases} \infty & \text{if } \beta < 1 \\ \tau(1 + 2\lambda) & \text{if } \beta = 1 \\ 0 & \text{if } \beta > 1 \end{cases}.$$

■

5.3. Moment and Moment Generating Function of TeSU-W

This section derives the r th moment (Theorem 3), the mean and variance (Corollary 1), and the moment generating function (Theorem 4) of TeSU–W.

Theorem 3. The r th moment of TeSU-W distribution, with pdf given in (12) is given by

$$\mu'_r = \tau^{-\frac{r}{\beta}} \Gamma\left(\frac{r}{\beta} + 1\right) [1 + 2\lambda - 6\lambda 2^{-\frac{r}{\beta}} + 4\lambda 3^{-\frac{r}{\beta}}], \quad (14)$$

where $r = 1, 2, \dots, n$ and $\Gamma(\cdot)$ is a gamma function.

Proof. The r th raw moment is defined by $\mu'_r = E(X^r) = \int_0^\infty x^r f(x) dx$. Thus, using the pdf $f(x)$ in equation (12) and simplifying, we have

$$\begin{aligned} \mu'_r &= \int_0^\infty x^r \tau \beta x^{\beta-1} e^{-\tau x^\beta} (1 + 2\lambda - 12\lambda e^{-\tau x^\beta} + 12\lambda e^{-2\tau x^\beta}) dx \\ &= (1 + 2\lambda) \tau^{-\frac{r}{\beta}} \Gamma\left(\frac{r}{\beta} + 1\right) - 6\lambda 2^{-\frac{r}{\beta}} \tau^{-\frac{r}{\beta}} \Gamma\left(\frac{r}{\beta} + 1\right) + 4\lambda 3^{-\frac{r}{\beta}} \tau^{-\frac{r}{\beta}} \Gamma\left(\frac{r}{\beta} + 1\right) \\ &= \tau^{-\frac{r}{\beta}} \Gamma\left(\frac{r}{\beta} + 1\right) [1 + 2\lambda - 6\lambda 2^{-\frac{r}{\beta}} + 4\lambda 3^{-\frac{r}{\beta}}]. \end{aligned}$$

■

Corollary 1. The mean and variance of the TeSU-W distribution are, respectively, given by

$$\mu = \mu'_1 = \tau^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta} + 1\right) [1 + 2\lambda - 6\lambda 2^{-\frac{1}{\beta}} + 4\lambda 3^{-\frac{1}{\beta}}] \text{ and}$$

$$\sigma^2 = \tau^{-\frac{2}{\beta}} \left[\Gamma\left(\frac{2}{\beta} + 1\right) (1 + 2\lambda - 6\lambda 2^{-\frac{2}{\beta}} + 4\lambda 3^{-\frac{2}{\beta}}) - \Gamma^2\left(\frac{1}{\beta} + 1\right) (1 + 2\lambda - 6\lambda 2^{-\frac{1}{\beta}} + 4\lambda 3^{-\frac{1}{\beta}})^2 \right].$$

Proof. The mean of the TeSU-W distribution is obtained when $r = 1$ in (14). Thus,

$$\mu = \mu'_1 = \tau^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta} + 1\right) [1 + 2\lambda - 6\lambda 2^{-\frac{1}{\beta}} + 4\lambda 3^{-\frac{1}{\beta}}].$$

It is to note that $\sigma^2 = \mu'_2 - (\mu'_1)^2$. Now, the 2nd raw moment μ'_2 of the proposed distribution is obtained using equation (14) when $r = 2$. It follows that

$$\mu'_2 = \tau^{-\frac{2}{\beta}} \Gamma\left(\frac{2}{\beta} + 1\right) [1 + 2\lambda - 6\lambda 2^{-\frac{2}{\beta}} + 4\lambda 3^{-\frac{2}{\beta}}].$$

Therefore, the variance σ^2 of TeSU-W distribution is derived as

$$\begin{aligned} \sigma^2 &= \mu'_2 - (\mu'_1)^2 \\ &= \tau^{-\frac{2}{\beta}} \left[\Gamma\left(\frac{2}{\beta} + 1\right) \left(1 + 2\lambda - 6\lambda 2^{-\frac{2}{\beta}} + 4\lambda 3^{-\frac{2}{\beta}}\right) - \Gamma^2\left(\frac{1}{\beta} + 1\right) \left(1 + 2\lambda - 6\lambda 2^{-\frac{1}{\beta}} + 4\lambda 3^{-\frac{1}{\beta}}\right)^2 \right]. \end{aligned}$$

■

Theorem 4. Let X follows the TeSU-W distribution, then its moment generating function, $M_{X_{TeSU-W}}(t)$ is given as

$$M_{X_{TeSU-W}}(t) = \sum_{r=0}^{\infty} \frac{t^r \tau^{-\frac{r}{\beta}}}{r!} \Gamma\left(\frac{r}{\beta} + 1\right) (1 + 2\lambda - 6\lambda 2^{-\frac{r}{\beta}} + 4\lambda 3^{-\frac{r}{\beta}}),$$

where $t \in \mathbb{R}$.

Proof. By definition of moment generating function and using equation (14), we have

$$\begin{aligned} M_{X_{TeSU-W}}(t) &= \mathbb{E}(e^{tX}) \\ &= \int_0^{\infty} e^{tx} f_{TeSU-W}(x) dx. \end{aligned}$$

Recall that $e^{tX} = \sum_{r=0}^{\infty} \frac{t^r}{r!} x^r$. Hence, we have

$$\begin{aligned} M_{X_{TeSU-W}}(t) &= \int_0^{\infty} \sum_{r=0}^{\infty} \frac{t^r}{r!} x^r f_{TeSU-W}(x) dx \\ &= \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^{\infty} x^r f_{TeSU-W}(x) dx \\ &= \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu'_r. \end{aligned}$$

Thus,

$$M_{X_{TeSU-W}}(t) = \sum_{r=0}^{\infty} \frac{t^r \tau^{-\frac{r}{\beta}}}{r!} \Gamma\left(\frac{r}{\beta} + 1\right) (1 + 2\lambda - 6\lambda 2^{-\frac{r}{\beta}} + 4\lambda 3^{-\frac{r}{\beta}}).$$

■

5.4. The Median of TeSU–G family and TeSU–W distribution

This section described the process of the derivation of the median of the TeSU–G family of distributions. Consider the Structured Set of Skew–Kurtotic Transmutations proposed by Shaw [14], that is, for parameters α_1, α_2 we shall consider the polynomial family given by

$$P(z, \alpha_1, \alpha_2) = z - z(1 - z) \left[\alpha_1 + \alpha_2 \left(z - \frac{1}{2} \right) \right],$$

where $z \in [0, 1]$ and the non-negativity of the pdf P' at the end points should satisfy

$$-1 - \frac{\alpha_2}{2} \leq \alpha_1 \leq 1 + \frac{\alpha_2}{2}.$$

Let u follows a uniform distribution $(0, 1)$. Then the solution for the equation $P(z, \alpha_1, \alpha_2) = u$ is as follows:

$$z = \begin{cases} u, & \text{if } \alpha_1 = \alpha_2 = 0 \\ \frac{\alpha_1 - 1 + \sqrt{1 + \alpha_1(\alpha_1 + 4u - 2)}}{2\alpha_1}, & \text{if } \alpha_2 = 0 \\ \sqrt[3]{u}, & \text{if } \alpha_1 = \frac{3}{2}, \alpha_2 = 1 \\ 1 - \sqrt[3]{1 - u}, & \text{if } \alpha_1 = -\frac{3}{2}, \alpha_2 = 1 \\ C(u, \alpha_1, \alpha_2), & \text{otherwise} \end{cases}$$

where $C(\cdot)$ is a function that denotes the general cubic (GC) solver for other cases. This function is processed by the following algorithm.

Step 1. Compute

$$Q = \frac{4\alpha_1^2 + 3(\alpha_2 - 4)\alpha_2}{36\alpha_2^2},$$

$$R = \frac{4\alpha_1^3 - 9\alpha_2\alpha_1(\alpha_2 + 2) + 27(1 - 2u)\alpha_2^2}{108\alpha_2^3}.$$

Step 2. If $R^2 > Q^3$, the equation has one real and two complex roots. In this case we have,

$$A = -\text{sign}(R) \left(|R| + \sqrt{R^2 - Q^3} \right)^{\frac{1}{3}};$$

$$B = \begin{cases} A, & \text{if } A = 0 \\ \frac{Q}{A}, & \text{otherwise} \end{cases};$$

$$C(u, \alpha_1, \alpha_2) = A + B - \frac{1}{3} \left(\frac{\alpha_1}{\alpha_2} - \frac{3}{2} \right).$$

Otherwise, the cubic has three real roots and this is done by setting

$$\theta = \arccos \left(\frac{R}{\sqrt[3]{Q}} \right);$$

$$C(u, \alpha_1, \alpha_2) = -2\sqrt[3]{Q} \cos \left(\frac{\theta - 2\pi}{3} \right) - \frac{1}{3} \left(\frac{\alpha_1}{\alpha_2} - \frac{3}{2} \right).$$

Observed that the cdf (7) of the TeSU-G family can be rewritten as

$$F(x) = z - z(1 - z) \left[\alpha_1 + \alpha_2 \left(z - \frac{1}{2} \right) \right],$$

where $z = G(x)$, $\alpha_1 = 0$ and $\alpha_2 = 4\lambda$, $\lambda \in [-0.5, 1]$. The inverse of $F(x)$ is a solution to the following equation

$$z = C(u, \alpha_1, \alpha_2) = G(x) = C(u, 0, 4\lambda).$$

Hence, the given algorithm can be modified as follows. Let u follows a uniform distribution $(0, 1)$. Then,

Step 1*. Compute

$$Q = \frac{1}{2} \left(1 - \frac{1}{\lambda} \right), \lambda \neq 0;$$

$$R = \frac{1 - 2u}{16\lambda},$$

Step 2*. If $R^2 > Q^3$ then

$$A = -\text{sign}(R) \left(|R| + \sqrt{R^2 - Q^3} \right)^{\frac{1}{3}};$$

$$B = \begin{cases} A, & \text{if } A = 0 \\ \frac{Q}{A}, & \text{otherwise} \end{cases};$$

$$x = G^{-1} \left(A + B + \frac{1}{2} \right).$$

Otherwise,

$$\theta = \arccos \left(\frac{R}{\sqrt{Q^3}} \right);$$

$$x = G^{-1} \left(\frac{1}{2} - 2\sqrt{Q} \cos \left(\frac{\theta - 2\pi}{3} \right) \right),$$

where $G^{-1}(x)$ is the inverse function of any baseline distribution function $G(x)$. If $\lambda = 0$, then $x = G^{-1}(u)$. The updated algorithm can be used for generating random numbers that follows any TeSU distribution. Consequently, the median of TeSU–G family can be computed by taking $u = \frac{1}{2}$, that is,

$$x_{med} = \begin{cases} G^{-1} \left(\frac{1}{2} \right), & \text{if } \lambda = 0 \\ G^{-1} \left[\frac{1}{2} - 2\sqrt{\frac{1}{12} \left(1 - \frac{1}{\lambda} \right)} \cos \left(\frac{90 - 2\pi}{3} \right) \right], & \text{otherwise} \end{cases}. \quad (15)$$

Setting $G(x)$ to be the cdf in equation (9) of the Weibull distribution, the algorithm is then modified to generate random numbers from the TeSU–Weibull distribution. The modified algorithm is as follows:

Step 1**. Compute

$$Q = \frac{1}{2} \left(1 - \frac{1}{\lambda} \right), \lambda \neq 0;$$

$$R = \frac{1 - 2u}{16\lambda}.$$

Step 2**. If $R^2 > Q^3$ then

$$A = -\text{sign}(R) \left(|R| + \sqrt{R^2 - Q^3} \right)^{\frac{1}{3}};$$

$$B = \begin{cases} A & \text{if } A = 0 \\ \frac{Q}{A} & \text{otherwise} \end{cases};$$

$$x = \left(-\frac{1}{\tau} \log \left(\frac{1}{2} - A - B \right) \right)^{\frac{1}{\beta}}.$$

Otherwise,

$$\theta = \arccos \left(\frac{R}{\sqrt{Q^3}} \right);$$

$$x = \left(-\frac{1}{\tau} \log \left(\frac{1}{2} + 2\sqrt{Q} \cos \left(\frac{\theta - 2\pi}{3} \right) \right) \right)^{\frac{1}{\beta}}.$$

If $\lambda = 0$, then $x = \left(-\frac{1}{\tau} \log(u) \right)^{\frac{1}{\beta}}$. Hence, the median of the TeSU–W is solved using equation (15) as

$$x_{med(\text{TeSU-W})} = \begin{cases} \left(-\frac{1}{\tau} \log \left(\frac{1}{2} \right) \right)^{\frac{1}{\beta}}, & \text{if } \lambda = 0 \\ \left(-\frac{1}{\tau} \log \left(\frac{1}{2} + 2\sqrt{Q} \cos \left(\frac{90-2\pi}{3} \right) \right) \right)^{\frac{1}{\beta}}, & \text{otherwise} \end{cases}.$$

6. THE TeSU–W MODEL PARAMETER ESTIMATION

Let X_1, X_2, \dots, X_n be an independently and identically distributed random variables from a TeSU–Weibull distribution. Then the likelihood function of the TeSU–W is given by

$$\mathbb{L} = \prod_{i=1}^n \left[\tau \beta x_i^{\beta-1} e^{-\tau x_i^\beta} \left(1 + 2\lambda - 12\lambda e^{-\tau x_i^\beta} + 12\lambda e^{-2\tau x_i^\beta} \right) \right].$$

Then the log-likelihood function is given by

$$l = \sum_{i=1}^n \log \left[\tau \beta x_i^{\beta-1} e^{-\tau x_i^\beta} \left(1 + 2\lambda - 12\lambda e^{-\tau x_i^\beta} + 12\lambda e^{-2\tau x_i^\beta} \right) \right]. \quad (16)$$

The derivatives of (16) with respect to the parameters τ , β and λ are given as follow:

$$\frac{\partial l}{\partial \tau} = \frac{n}{\tau} - \sum_{i=1}^n x_i^\beta + 12\lambda \sum_{i=1}^n \frac{z_i x_i^\beta e^{-\tau x_i^\beta}}{y_i}; \quad (17)$$

$$\frac{\partial l}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n \tau x_i^\beta \log(x_i) + \sum_{i=1}^n \log(x_i) + 12\tau \lambda \sum_{i=1}^n \frac{z_i x_i^\beta \log(x_i) e^{-\tau x_i^\beta}}{y_i}; \quad (18)$$

$$\frac{\partial l}{\partial \lambda} = 2 \sum_{i=1}^n \frac{1 - 6e^{-\tau x_i^\beta} + 6e^{-2\tau x_i^\beta}}{y_i}, \quad (19)$$

where $y_i = 1 + 2\lambda - 12\lambda e^{-\tau x_i^\beta} + 12\lambda e^{-2\tau x_i^\beta}$ and $z_i = 1 - 2e^{-\tau x_i^\beta}$.

Setting equations (17), (18) and (19) equal to zero, the numerical maximum likelihood estimates $\hat{\tau}$, $\hat{\beta}$ and $\hat{\lambda}$ of the parameters can be obtained by any numerical method like the Newton-Raphson iterative method.

7. THE ASYMPTOTIC PROPERTIES OF TeSU–W ML ESTIMATES

This section presents the simulation study result conducted to verify the performance of the ML estimates of TeSU-W distribution when sample sizes are varied. The simulation process proceeded with 2 sets of data from TeSU-W distribution and considered the following sets of parameters values: $s_1 = \{\tau = 1.4, \beta = 2.5, \lambda = 0.5\}$ and $s_2 = \{\tau = 1.4, \beta = 1, \lambda = -0.5\}$. For each s_i , the study is processed for varied sample sizes $n \in \{50, 100, 200, 500, 1000\}$. Also, at each replication, the ML estimates $\hat{\tau}$, $\hat{\beta}$ and $\hat{\lambda}$ are computed. The process is repeated 1000 times for

each s_i , and some diagnostic statistics like the average estimate (AE), biases and mean squared errors (MSE) are determined and are summarized in Tables 1 and 2. These indicate that the MSE of $\hat{\tau}$, $\hat{\beta}$ and $\hat{\lambda}$ for sets s_i , $i = 1, 2$ decay toward zero as the sample size increases, that is, $\lim_{n \rightarrow \infty} MSE = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\hat{P}_i - P)^2 = 0$ where $P = \tau, \beta, \lambda$. This implies that the AE of the parameters for each s_i tend to be closer to the true parameters as sample sizes n increases.

Table 1: Some diagnostic statistics of TeSU-W for s_1 at varied n

n	MLE	AE	Bias	MSE
50	$\hat{\tau}$	1.446	0.046	0.047
	$\hat{\beta}$	2.716	0.216	0.124
	$\hat{\lambda}$	0.569	0.069	0.041
100	$\hat{\tau}$	1.440	0.040	0.022
	$\hat{\beta}$	2.694	0.194	0.070
	$\hat{\lambda}$	0.552	0.052	0.020
200	$\hat{\tau}$	1.433	0.033	0.010
	$\hat{\beta}$	2.683	0.183	0.051
	$\hat{\lambda}$	0.555	0.055	0.010
500	$\hat{\tau}$	1.429	0.029	0.005
	$\hat{\beta}$	2.679	0.179	0.039
	$\hat{\lambda}$	0.547	0.047	0.005
1000	$\hat{\tau}$	1.428	0.028	0.003
	$\hat{\beta}$	2.671	0.171	0.032
	$\hat{\lambda}$	0.544	0.044	0.003

Table 2: Some diagnostic statistics of TeSU-W for s_2 at varied n

n	MLE	AE	Bias	MSE
50	$\hat{\tau}$	1.509	0.109	0.044
	$\hat{\beta}$	1.107	0.107	0.036
	$\hat{\lambda}$	-0.486	0.014	0.007
100	$\hat{\tau}$	1.483	0.083	0.019
	$\hat{\beta}$	1.084	0.084	0.014
	$\hat{\lambda}$	-0.498	0.002	0.001
200	$\hat{\tau}$	1.473	0.073	0.012
	$\hat{\beta}$	1.080	0.080	0.009
	$\hat{\lambda}$	-0.500	0.000	0.000
500	$\hat{\tau}$	1.463	0.063	0.006
	$\hat{\beta}$	1.073	0.073	0.006
	$\hat{\lambda}$	-0.500	0.000	0.000
1000	$\hat{\tau}$	1.463	0.063	0.005
	$\hat{\beta}$	1.072	0.072	0.006
	$\hat{\lambda}$	-0.500	0.000	0.000

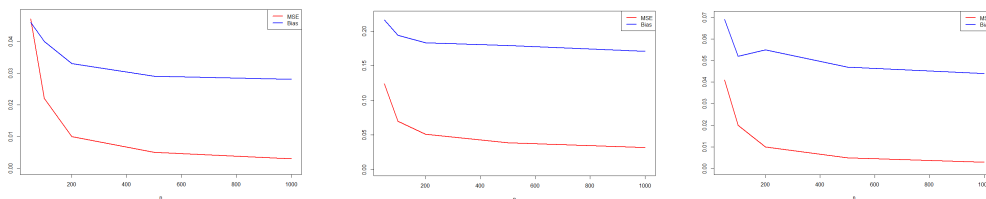


Figure 5: Plots of the MSE and Bias for $\hat{\tau}$ (left), $\hat{\beta}$ (center) and $\hat{\lambda}$ (right) in s_1

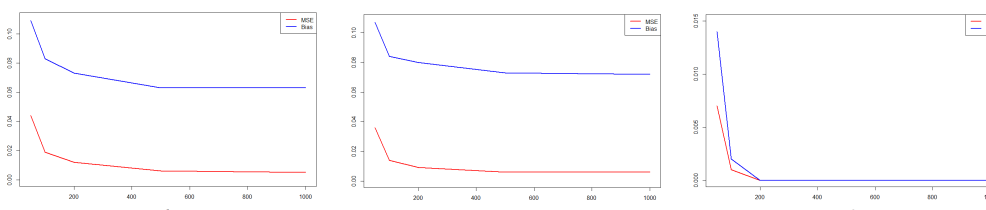


Figure 6: Plots of the MSE and Bias for $\hat{\tau}$ (left), $\hat{\beta}$ (center) and $\hat{\lambda}$ (right) in s_2

8. THE TeSU–W IN LIFE DATA ANALYSIS

This section illustrates the TeSU–W distribution when applied to real life dataset using a package "fitdistrplus" of the R software. The result of the TeSU–W distribution will then be compared to the recent work of Arif [3] on the New Extended Exponentiated Weibull (NEEW) distribution and the work of Malik [11] on the New Transmuted Weibull (NTW) distribution. The New NEEW

pdf is given by

$$f_{NEEW}(x) = \frac{\alpha \lambda x^{\lambda-1} e^{-\alpha x^\lambda} (1 - e^{-\alpha x^\lambda}) \left[e^{\theta(1-e^{-\alpha x^\lambda})} (2 + \theta - \theta e^{-\alpha x^\lambda}) + 2 \right]}{e^\theta + 1}, x \geq 0, \alpha, \lambda, \theta > 0$$

while the pdf of the New Transmuted Weibull (NTW) distribution is given by

$$f_{NTW}(x) = \theta \lambda x^{\lambda-1} e^{-\theta x^\lambda} \left[1 - \beta + \frac{2\beta}{2 - e^{-\theta x^\lambda}} \right], x, \theta, \lambda > 0, -1 \leq \beta \leq 1.$$

Model diagnostics are done with the determination of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Kolmogorov-Smirnov (K–S), Cramer-von Mises (W*) and the Anderson-Darling (A) statistics. As a rule of thumb, a smaller value of these statistics implies a better fit of the model using the proposed distribution to the given dataset.

The COVID-19 cases in India from May 1, 2020 to June 14, 2020 are used in this study. This data set can be accessed from the siteweb (Coronavirus Update (Live):7,114,524 Cases and 406,552 Deaths from COVID-19 Virus Pandemic - Worldometer). For calculation purpose, we consider data (10^{-2}). Table 3 lists the MLEs of the TeSU-W, NEEW, NTW and TeSU-E distributions fitted to the given dataset while Table 4 shows the different diagnostics statistics. Consistently in all diagnostic criterion, the TeSU–W gave the lowest values of the diagnostic statistics compared to NEEW, NTW and TeSU-E distributions. It may imply that the TeSU–W works well when fitted with the given dataset and that the ML estimates are asymptotically equal to the true values of the parameters. In addition, same result is observed from the plots of the fitted models and the histogram of the dataset given in Figure 7.

Table 3: ML estimates of the fitted models using the different distributions

Distribution	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\tau}$
TeSU–W		3.05948300		0.47928580	0.00000183
NEEW	0.00111771		0.00000082	1.69201100	
NTW		–0.99999996	0.00047039	1.86047516	
TeSU–E				–0.49999997	0.01248669

Table 4: Some diagnostic statistics of the fitted models using the different distributions

Distribution	AIC	BIC	K – S	A	W*
TeSU–W	428.9631	434.3830	0.1005729	0.4046896	0.0566358
NEEW	433.5504	438.9703	0.1271787	0.6914853	0.1042066
NTW	433.0563	438.4763	0.1255249	0.7159459	0.1097673
TeSU–E	446.9146	450.528	0.1660065	2.2543706	0.3196720

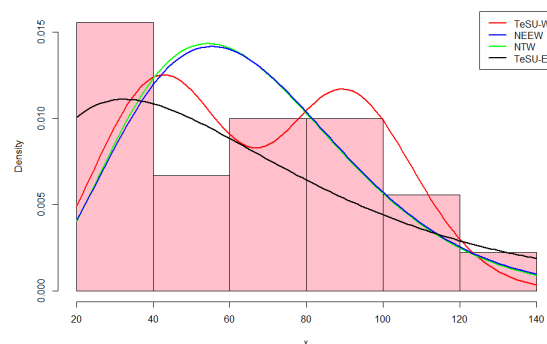


Figure 7: Plots of the models fitted to the COVID-19 data

9. CONCLUDING REMARKS

This paper derives a new family of distributions called the T-extended Standard U-quadratic–G family of distributions or simply, TeSU–G family. Derived models of the family called as TeSU–Weibull distribution (TeSU–W) and the TeSU-Exponential distribution (TeSU–E) are generated and its limiting behavior, moments, mean and variance, and moment generating function are computed. Also, formula of the median for the TeSU–G family as well as for TeSU–W distribution are derived. Furthermore, the Maximum Likelihood (ML) estimates of the TeSU–W distribution is derived. Simulation study shows that the ML estimates is asymptotically equal to the true value of the parameters as sample sizes increases. This can be observed by the values of MSE that goes to zero, on the average. Life data analysis using the TeSU–W distribution to a COVID-19 dataset provides better fit compared with the existing New Extended Exponentiated Weibull (NEEW) distribution and the New Transmuted Weibull (NTW) distribution as explored by Arif [3] in 2022 and Malik [11] in 2022, respectively.

ACKNOWLEDGEMENT

This study was supported by the Department of Science and Technology - Accelerated Science and Technology Human Resource Development Program (DOST-ASTHRDP).

REFERENCES

- [1] Alizadeh, M., Tahir, M. H., Zubair, M. and Hamedani, G. G. (2015). The Kumaraswamy Marshall-Olkin family of distributions. *Journal of the Egyptian Mathematical Society*, 23:546–557.
- [2] Alzaatreh, A. and Lee, C. and Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71:63–79.
- [3] Arif, M., Khan, D. M., Aamir, M., Khalil, U., Bantan, R. A. R. and Elgarhy, M. (2022). Modeling COVID-19 Data with a Novel Extended Exponentiated Class of Distributions. *Journal of Mathematics*, Article ID 1908161, <https://doi.org/10.1155/2022/1908161>.
- [4] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.
- [5] Cordeiro, G.M., Ortega, E. M. M. and Cunha, D. C. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science*, 11:1–27.
- [6] Cordeiro, G. M., Alizadeh, M., Tahir, M. H., Mansoor, M., Bourguignon, M. and G., H. G. (2017). The generalized odd log logistic family of distributions: properties, regression models and applications. *Journal of Statistical Computation and Simulation*, 87:908–932.
- [7] Cordeiro, G. M. and de Castro, M.. (2011). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81:883–893.
- [8] Cordeiro, G. M., Silva, G. O. and Ortega, E. M. M. (2012). The beta extended Weibull distribution. *Journal of Probability and Statistical Science*, 10:15–40.
- [9] Eugene, N., Lee, C. and Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics-Theory and Methods*, 31:497–512.
- [10] Gupta, R. C., Gupta, P. L. and Gupta, R. D. (1998). Modeling failure time data by Lehman alternatives. *Communication Statistics Theory and Methods*, 27:887–904.
- [11] Malik, A. S. and Ahmad, S. P. (2022). A New Transmuted Weibull Distribution: Properties and Application. *Pakistan Journal of Statistics and Operation Research*, 18:369–381.
- [12] 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*, 84:641–652.
- [13] Nofal, Z. M., Afify, A. Z., Yousof, H. M. and Cordeiro, G. M. (2017). The generalized transmuted - G family of distributions. *Communication Statistics Theory and Methods*, 49:4119–4136.

- [14] Shaw, W. T. and Buckley, I. R. C. (2007). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *Research report*.
- [15] Silva, R., Silva, F. G., Ramos, M., Cordeiro, G. M., Marinho, P. and De Andrade, T. A. N. (2019). The Exponentiated Kumaraswamy - G Class: General Properties and Application. *R.C de Estadística*, 42:1–33.
- [16] Zografos, K. J. and Balakrishnan, N. (2009). On families of beta and generalized gamma-generated distribution and associate inference. *Statistical Methodology*, 6:344–362.

EXPONENTIATED ADYA DISTRIBUTION: PROPERTIES AND APPLICATIONS

Rashid A. Ganaie¹, Aafaq A. Rather^{2,*}, D. Vedavathi Saraja³, Mahfooz Alam⁴, Aijaz Ahmad⁵, Arif Muhammad Tali⁶, Rifat Nisa⁷, Berihan R. Elemetry⁸

•

^{1,3}Department of Statistics, Annamalai University, Tamil nadu, India

^{2,*}Symbiosis Statistical Institute, Symbiosis International (Deemed University), Pune-411004, India

⁴Department of Statistics, Faculty of Science and Technology, Vishwakarma University, Pune, India

⁵Department of Mathematics, Bhagwant University, Ajmer, India

⁶Department of Mathematics, Sharda University, Noida, India

⁷Department of Statistics, Cluster University, Srinagar, Kashmir, India

⁸Department of Statistics and Insurance, Faculty of Commerce, Damietta University- 34511, Egypt

¹rashidau7745@gmail.com, ^{2,*}aafaq7741@gmail.com, ³sarajayoganand@gmail.com,

⁴mahfooz.alam@vupune.ac.in, ⁵aijazahmad4488@gmail.com, ⁶arif.tali@sharda.ac.in,

⁷rifatnisa1111@gmail.com, ⁸berihanelemetry@gmail.com

Abstract

In this article, we introduce a new generalization of Adya distribution known as Exponentiated Adya distribution. We have also derived and discussed its different statistical properties. The Exponentiated Adya distribution has two parameters (scale and shape). The different structural properties of the proposed distribution have been obtained. The maximum likelihood estimation technique is also used for estimating the parameters of the proposed distribution. Finally, a real lifetime data set is used for examining the superiority of the proposed distribution.

Keywords: Exponentiated distribution, Adya distribution, Order statistics, Entropies, Reliability analysis, Maximum likelihood Estimation.

1. Introduction

A new family of distributions namely the exponentiated exponential distribution was introduced by Gupta et al. [1]. The family has two parameters scale and shape, which are similar to the weibull or gamma family. Later Gupta and Kundu [2], studied some properties of the same distribution. They observed that many properties of the new family are similar to those of the weibull or gamma family. Hence the distribution can be used an alternative to a weibull or gamma distribution. The two-parameteric gamma and weibull are the most popular distributions for analyzing any lifetime data. The gamma distribution has a lot of applications in different fields other than lifetime distributions. The two parameters of gamma distribution represent the scale and the shape parameter and because of the scale and shape parameter, it has quite a bit of flexibility to analyze any positive real data. But one major disadvantage of the gamma distribution is that, if the shape parameter is not an integer, the

distribution function or survival function cannot be expressed in a closed form. This makes gamma distribution little bit unpopular as compared to the Weibull distribution, whose survival function and hazard function are simple and easy to study. Nowadays exponentiated distributions and their mathematical properties are widely studied for applied science experimental data sets. Exponentiated weibull family as an extension of weibull distribution studied by Pal et al. [3]. Exponentiated generalized Lindley distribution studied by Rodrigues et al. [4]. Hassan et al. [5] discussed Exponentiated Lomax geometric distribution with its properties and applications. Nasiru et al. [6] obtained exponentiated generalized power series family of distributions. Rather and subramanian [7] discussed the exponentiated Mukherjee-Islam distribution which shows more flexibility than the classical distribution. Rather and subramanian [8] discussed the exponentiated ishita distribution with properties and Applications. Subramanian and Rather [9] obtained the exponentiated version of power distribution with its properties and estimation. Rather and subramanian [10] discussed the exponentiated Garima distribution which shows more flexibility than the classical distribution. Ganie and Rajagopalan [11] obtained exponentiated Aradhana distribution with its properties and applications. Recently, Rather et al. [12] discussed the exponentiated Ailamujia distribution with statistical inference and applications of medical science which shows better performance than the classical distributions.

Adya distribution is a newly proposed one parameteric distribution formulated by Shanker et al. [13] for several engineering applications and calculated its various characteristics including stochastic ordering, moments, order statistics, Renyi entropy, stress strength reliability and ML estimation. The two parameters of an exponentiated Adya distribution represent the shape and the scale parameter. It also has the increasing or decreasing failure rate depending of the shape parameter. The density function varies significantly depending of the shape parameter.

2. Exponentiated Adya Distribution (EAD)

The probability density function of Adya distribution is given by

$$g(x) = \frac{\theta^3}{\theta^4 + 2\theta^2 + 2} (\theta + x)^2 e^{-\theta x}; x > 0, \theta > 0 \quad (1)$$

and the cumulative distribution function of Adya distribution is given by

$$G(x) = 1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x}; x > 0, \theta > 0 \quad (2)$$

A random variable X is said to have an exponentiated distribution, if its cumulative distribution function is given by

$$F_\alpha(x) = (G(x))^\alpha; x \in R^+, \alpha > 0 \quad (3)$$

Then X is said to have an exponentiated distribution.

The probability density function of X is given by

$$f_\alpha(x) = \alpha(G(x))^{\alpha-1} g(x) \quad (4)$$

By Substituting (2) in (3), we will obtain the cumulative distribution function of Exponentiated Adya distribution

$$F_\alpha(x) = \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^\alpha; x > 0, \theta > 0, \alpha > 0 \quad (5)$$

and the probability density function of Exponentiated Adya distribution can be obtained as

$$f_{\alpha}(x) = \frac{\alpha\theta^3(\theta+x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} \quad (6)$$

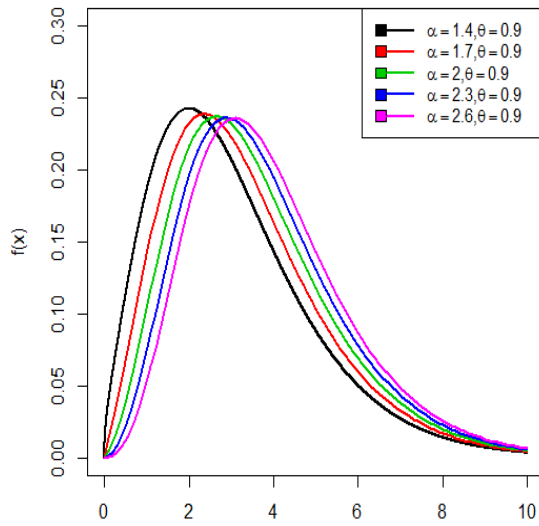


Fig 1: pdf plot of exponentiated Adya distribution

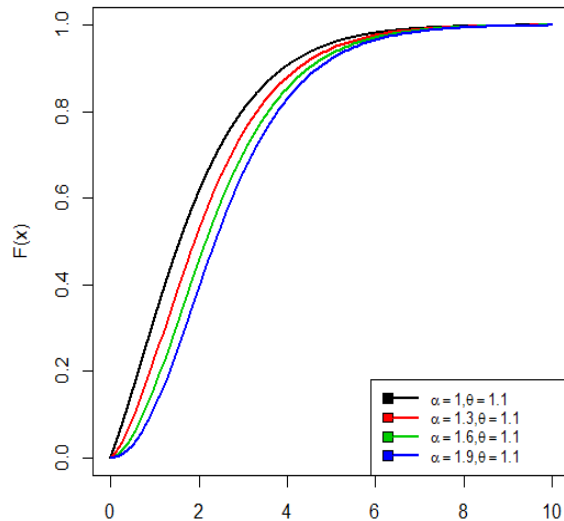


Fig 2: cdf plot of exponentiated Adya distribution

3. Reliability Analysis

In this section, we will obtain the survival function, hazard function and Reverse hazard rate function of the Exponentiated Adya distribution.

The survival function of Exponentiated Adya distribution is given by

$$S(x) = 1 - \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha} \quad (7)$$

The hazard function is also known as hazard rate, instantaneous failure rate or force of mortality and is given by

$$h(x) = \frac{\left(\frac{\alpha\theta^3(\theta+x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} \right)}{\left(1 - \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha} \right)} \quad (8)$$

The reverse hazard rate of exponentiated Adya distribution is given by

$$h_r(x) = \frac{\alpha\theta^3(\theta+x)^2 e^{-\theta x}}{\theta x(\theta x + 2\theta^2 + 2)e^{-\theta x}} \quad (9)$$

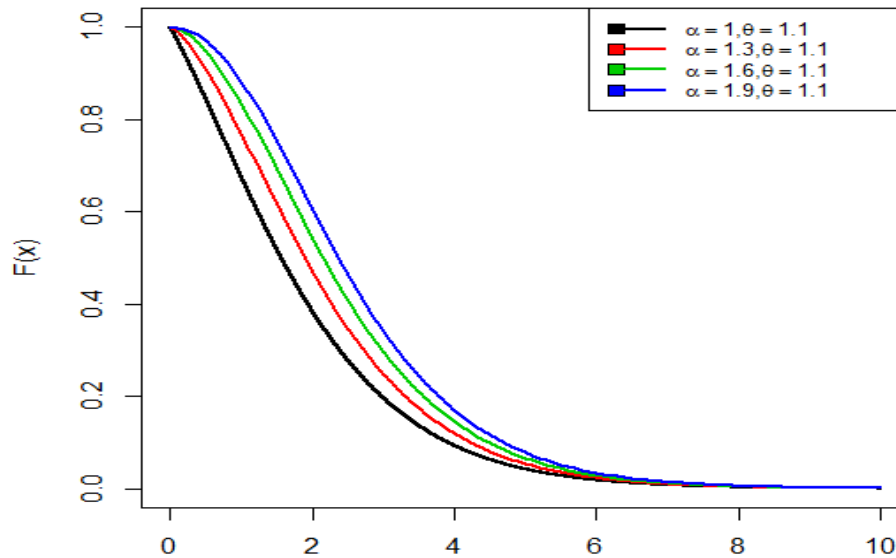


Fig 3: Survival plot of exponentiated Adya distribution

4. Statistical Properties

In this section, we will discuss the different statistical properties of the proposed Exponentiated Adya distribution.

4.1 Moments

Suppose X is a random variable following exponentiated Adya distribution with parameters α and θ , then the r th order moment $E(X^r)$ for a given probability distribution is given by

$$E(X^r) = \mu_r' = \int_0^{\infty} x^r f_{\alpha}(x) dx$$

$$E(X^r) = \frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \int_0^{\infty} x^r (\theta+x)^2 e^{-\theta x} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} dx \quad (10)$$

Using Binomial expansion of

$$\left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} = \sum_{i=0}^{\infty} \binom{\alpha-1}{i} \left\{ \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right\}^i (-1)^i$$

Equation (10) will become

$$E(X^r) = \frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \sum_{i=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \int_0^{\infty} x^r (\theta+x)^2 e^{-\theta x(1+i)} \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^i dx \quad (11)$$

Again using Binomial expansion of

$$\left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^i = \sum_{k=0}^{\infty} \binom{i}{k} \left(\frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^k$$

Equation (11), will reduce to

$$E(X^r) = \frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \left(\frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^k \int_0^{\infty} x^r (\theta + x)^2 e^{-\theta x(1+i)} dx \quad (12)$$

After simplification, we obtain

$$E(X^r) = \alpha\theta^3 \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \frac{(\theta^2 + 2\theta^3 + 2\theta)^k}{(\theta^4 + 2\theta^2 + 2)^{k+1}} \times \left(\frac{\theta^2(1+i)^2 \Gamma(r+4k+1) + \Gamma(r+4k+3) + 2\theta(\theta(1+i)\Gamma(r+4k+2))}{\theta(1+i)^{r+4k+3}} \right) \quad (13)$$

Since equation (13) is a convergent series for all $r \geq 0$, therefore all the moments exist. Therefore

$$E(X) = \alpha\theta^3 \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \frac{(\theta^2 + 2\theta^3 + 2\theta)^k}{(\theta^4 + 2\theta^2 + 2)^{k+1}} \left(\frac{\theta^2(1+i)^2 \Gamma(4k+2) + \Gamma(4k+4) + 2\theta(\theta(1+i)\Gamma(4k+3))}{\theta(1+i)^{4k+4}} \right) \quad (14)$$

$$E(X^2) = \alpha\theta^3 \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \frac{(\theta^2 + 2\theta^3 + 2\theta)^k}{(\theta^4 + 2\theta^2 + 2)^{k+1}} \left(\frac{\theta^2(1+i)^2 \Gamma(4k+3) + \Gamma(4k+5) + 2\theta(\theta(1+i)\Gamma(4k+4))}{\theta(1+i)^{4k+5}} \right) \quad (15)$$

Therefore, the Variance of X can be obtained as

$$V(X) = E(X^2) - (E(X))^2$$

4.2 Harmonic mean

The Harmonic mean for the proposed Exponentiated Adya distribution can be obtained as

$$H.M = E\left(\frac{1}{x}\right) = \int_0^{\infty} \frac{1}{x} f_{\alpha}(x) dx$$

$$H.M. = \frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \int_0^{\infty} \frac{1}{x} (\theta + x)^2 e^{-\theta x} \left(1 - \left(\frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} dx \quad (16)$$

Using Binomial expansion in equation (16), we get

$$H.M = \frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \sum_{i=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \int_0^{\infty} \frac{1}{x} (\theta + x)^2 e^{-\theta x(1+i)} \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^i dx \quad (17)$$

On using Binomial expansion in equation (17), we obtain

$$H.M = \frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \left(\frac{\theta^2 x^2 + 2\theta^3 x + 2\theta x}{\theta^4 + 2\theta^2 + 2} \right)^k \int_0^{\infty} \frac{1}{x} (\theta + x)^2 e^{-\theta x(1+i)} dx \quad (18)$$

After the simplification of equation (18), we obtain

$$H.M = \alpha \theta^3 \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \frac{(\theta^2 + 2\theta^3 + 2\theta)^k}{(\theta^4 + 2\theta^2 + 2)^{k+1}} \left(\frac{\theta^2(1+i)\Gamma(4k+1) + 2\theta(1+i)\Gamma(4k+1) + \Gamma(4k+2)}{\theta(1+i)^{4k+2}} \right) \quad (19)$$

4.3 Moment Generating Function and Characteristics Function

Let X have an exponentiated Adya distribution, then the moment generating function of X is obtained as

$$M_X(t) = E(e^{tx}) = \int_0^{\infty} e^{tx} f_{\alpha}(x) dx$$

Using Taylor's series, we get

$$M_X(t) = \int_0^{\infty} \left(1 + tx + \frac{(tx)^2}{2!} + \dots \right) f_{\alpha}(x) dx \quad (20)$$

$$M_X(t) = \alpha \theta^3 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \frac{t^j (\theta^2 + 2\theta^3 + 2\theta)^k}{(j! (\theta^4 + 2\theta^2 + 2)^{k+1}} \times \left(\frac{\theta^2(1+i)^2 \Gamma(j+4k+1) + \Gamma(j+4k+3) + 2\theta(\theta(1+i)\Gamma(j+4k+2))}{\theta(1+i)^{j+4k+3}} \right) \quad (21)$$

Similarly, the characteristic function of Exponentiated Adya distribution is given by

$$\varphi_X(t) = \alpha \theta^3 \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\alpha-1}{i} \binom{i}{k} \frac{mt^j (\theta^2 + 2\theta^3 + 2\theta)^k}{(j! (\theta^4 + 2\theta^2 + 2)^{k+1}} \times \left(\frac{\theta^2(1+i)^2 \Gamma(j+4k+1) + \Gamma(j+4k+3) + 2\theta(\theta(1+i)\Gamma(j+4k+2))}{\theta(1+i)^{j+4k+3}} \right) \quad (22)$$

5. Order Statistics

Order statistics represents the arranging of samples in an ascending order. Order statistics also has wide field in reliability and life testing. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the order statistics of a random sample X_1, X_2, \dots, X_n drawn from the continuous population with probability density function $f_x(x)$ and cumulative distribution function $F_x(x)$, then the pdf of r^{th} order statistics $X_{(r)}$ can be written as

$$f_{X(r)}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x) (F_X(x))^{r-1} (1 - F_X(x))^{n-r} \quad (23)$$

Substitute the values of equation (5) and (6) in equation (13), we will obtain the pdf of r^{th} order statistics $X_{(r)}$ for exponentiated Adya distribution and is given by

$$f_{X(r)}(x) = \frac{n!}{(r-1)!(n-r)!} \frac{\alpha \theta^3 (\theta + x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} \\ \times \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha(r-1)} \times \left(1 - \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^\alpha \right)^{n-r} \quad (24)$$

Therefore, the probability density function of higher order statistics $X_{(n)}$ for exponentiated Adya distribution can be obtained as

$$f_{X(n)}(x) = n \frac{\alpha \theta^3 (\theta + x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha(n-1)} \quad (25)$$

and the pdf of first order statistics $X_{(1)}$ for exponentiated Adya distribution can be obtained as

$$f_{X(1)}(x) = n \frac{\alpha \theta^3 (\theta + x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} \left(1 - \left(1 - \left(1 + \frac{\theta x (\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^\alpha \right)^{n-1} \quad (26)$$

6. Maximum Likelihood Estimation

In this section, we will discuss the maximum likelihood estimation for estimating the parameters of exponentiated Adya distribution. Let X_1, X_2, \dots, X_n be the random sample of size n from the Exponentiated Adya distribution, then the likelihood function can be written as

$$L(\alpha, \theta) = \frac{(\alpha \theta^3)^n}{(\theta^4 + 2\theta^2 + 2)^n} \prod_{i=1}^n \left((\theta + x_i)^2 e^{-\theta x_i} \left(1 - \left(1 + \frac{\theta x_i (\theta x_i + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x_i} \right)^{\alpha-1} \right) \quad (27)$$

The log likelihood function is given by

$$\log L(\alpha, \theta) = n \log \alpha + 3n \log \theta - n \log(\theta^4 + 2\theta^2 + 2) + 2 \sum_{i=1}^n \log(\theta + x_i) - \theta \sum_{i=1}^n x_i \\ + (\alpha - 1) \sum_{i=1}^n \log \left(1 - \left(1 + \frac{\theta x_i (\theta x_i + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x_i} \right) \quad (28)$$

The maximum likelihood estimates of α, θ which maximizes (28), must satisfy the normal equations given by

$$\frac{\partial \log L}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log \left(1 - \left(1 + \frac{\theta x_i (\theta x_i + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x_i} \right) = 0 \quad (29)$$

$$\Rightarrow \hat{\alpha} = \frac{n}{\sum_{i=1}^n \log \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)} \quad (30)$$

$$\frac{\partial \log L}{\partial \theta} = \frac{3n}{\theta} - n \left(\frac{4\theta^3 + 4\theta}{\theta^4 + 2\theta^2 + 2} \right) + 2 \sum_{i=1}^n \left(\frac{1}{(\theta + x)} \right) - \sum_{i=1}^n x_i + (\alpha - 1) \psi \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right) = 0 \quad (31)$$

Where $\psi(\cdot)$ is the digamma function.

It is important to mention here that the analytical solution of the above system of non-linear equation is unknown. Algebraically it is very difficult to solve the complicated form of likelihood system of nonlinear equations. Therefore, we use R and wolfram mathematics for estimating the required parameters.

7. Information Measures of Exponentiated Adya Distribution

7.1 Renyi Entropy

The Renyi entropy is named after Alfred Renyi in the context of fractal dimension estimation, the Renyi entropy forms the basis of the concept of generalized dimensions. The Renyi entropy is important in ecology and statistics as index of diversity. The Renyi entropy is also important in quantum information, where it can be used as a measure of entanglement. Entropies quantify the diversity, uncertainty, or randomness of a system. For a given probability distribution, Renyi entropy is given by

$$e(\beta) = \frac{1}{1 - \beta} \log \left(\int_0^{\infty} f^{\beta}(x) dx \right)$$

Where, $\beta > 0$ and $\beta \neq 1$

$$e(\beta) = \frac{1}{1 - \beta} \log \left(\int_0^{\infty} \left\{ \frac{\alpha \theta^3 (\theta + x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha - 1} \right\}^{\beta} dx \right) \quad (32)$$

$$e(\beta) = \frac{1}{1 - \beta} \log \left(\left(\frac{\alpha \theta^3}{\theta^4 + 2\theta^2 + 2} \right)^{\beta} \int_0^{\infty} (\theta + x)^{2\beta} e^{-\theta \beta x} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\beta(\alpha - 1)} dx \right) \quad (33)$$

Using binomial expansion in (33), we get

$$e(\beta) = \frac{1}{1 - \beta} \log \left(\left(\frac{\alpha \theta^3}{\theta^4 + 2\theta^2 + 2} \right)^{\beta} \sum_{i=0}^{\infty} (-1)^i \binom{\beta(\alpha - 1)}{i} \int_0^{\infty} (\theta + x)^{2\beta} e^{-\theta x(\beta + i)} \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^i dx \right) \quad (34)$$

Again using binomial expansion in (34), we get

$$e(\beta) = \frac{1}{1 - \beta} \log \left(\left(\frac{\alpha \theta^3}{\theta^4 + 2\theta^2 + 2} \right)^{\beta} \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\beta(\alpha - 1)}{i} \binom{i}{k} \left(\frac{\theta^2 x^2 + 2\theta^3 x + 2\theta x}{\theta^4 + 2\theta^2 + 2} \right)^k \int_0^{\infty} (\theta + x)^{2\beta} e^{-\theta x(\beta + i)} dx \right) \quad (35)$$

After the simplification of (35) we obtain

$$e(\beta) = \frac{1}{1-\beta} \log \left((\alpha\theta^3)^\beta \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\beta(\alpha-1)}{i} \binom{2\beta}{k} \binom{2\beta}{j} \frac{(\theta^2 + 2\theta^3 + 2\theta)^k}{(\theta^4 + 2\theta^2 + 2)^{\beta+k}} \theta^{2\beta-j} \frac{\Gamma(4k+j+1)}{\theta(\beta+i)^{4k+j+1}} \right) \quad (36)$$

7.2 Tsallis Entropy

A generalization of Boltzmann-Gibbs (B-G) statistical mechanics initiated by Tsallis has gained a great deal to attention. This generalization of B-G statistics was proposed firstly by introducing the mathematical expression of Tsallis entropy for a continuous random variable it is defined as

$$S_\lambda = \frac{1}{\lambda-1} \left(1 - \int_0^\infty f^\lambda(x) dx \right)$$

$$S_\lambda = \frac{1}{\lambda-1} \left(1 - \int_0^\infty \left[\frac{\alpha\theta^3(\theta+x)^2 e^{-\theta x}}{\theta^4 + 2\theta^2 + 2} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\alpha-1} \right]^\lambda dx \right) \quad (37)$$

$$S_\lambda = \frac{1}{\lambda-1} \left(1 - \left(\frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \right)^\lambda \int_0^\infty (\theta+x)^{2\lambda} e^{-\lambda\theta x} \left(1 - \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right) e^{-\theta x} \right)^{\lambda(\alpha-1)} dx \right) \quad (38)$$

Using binomial expansion in (38), we get

$$S_\lambda = \frac{1}{\lambda-1} \left(1 - \left(\frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \right)^\lambda \sum_{i=0}^{\infty} (-1)^i \binom{\lambda(\alpha-1)}{i} \int_0^\infty (\theta+x)^{2\lambda} e^{-\theta x(\lambda+i)} \left(1 + \frac{\theta x(\theta x + 2\theta^2 + 2)}{\theta^4 + 2\theta^2 + 2} \right)^i dx \right) \quad (39)$$

Again using binomial expansion in (39), we obtain

$$= \frac{1}{\lambda-1} \left(1 - \left(\frac{\alpha\theta^3}{\theta^4 + 2\theta^2 + 2} \right)^\lambda \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\lambda(\alpha-1)}{i} \binom{i}{k} \left(\frac{\theta^2 x^2 + 2\theta^3 x + 2\theta x}{\theta^4 + 2\theta^2 + 2} \right)^k \int_0^\infty (\theta+x)^{2\lambda} e^{-\theta x(\lambda+i)} dx \right) \quad (40)$$

After the simplification of (40), we get

$$S_\lambda = \frac{1}{\lambda-1} \left(1 - (\alpha\theta^3)^\lambda \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-1)^i \binom{\lambda(\alpha-1)}{i} \binom{i}{k} \binom{2\lambda}{j} \frac{(\theta^2 + 2\theta^3 + 2\theta)^k}{(\theta^4 + 2\theta^2 + 2)^{\lambda+k}} \theta^{2\lambda-j} \frac{\Gamma(4k+j+1)}{\theta(\lambda+i)^{4k+j+1}} \right) \quad (41)$$

7. Applications

In this section, we use a real-life time data set in exponentiated Adya distribution and the model has been compared over Adya distribution.

The following data set in table 1 represents the bladder cancer patients ($n = 128$) of the remission times (in months) reported by Lee and Wang [14].

Table 1: Data represents the 123 blood cancer patients

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	6.31	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		

In order to compare the exponentiated Adya distribution with Adya distribution. We consider the Criteria like BIC (Bayesian information criterion), AIC (Akaike information criterion), AICC (Corrected Akaike information criterion) and $-2\log L$. The better distribution is which corresponds to lesser values of AIC, BIC, AICC and $-2\log L$. For calculating AIC, BIC, AICC and $-2\log L$ can be evaluated by using the formulas as follows.

$$AIC = 2k - 2 \log L \quad AICC = AIC + \frac{2k(k+1)}{n-k-1} \quad \text{and} \quad BIC = k \log n - 2 \log L$$

Where k is the number of parameters in the statistical model, n is the sample size and $-2\log L$ is the maximized value of the log-likelihood function under the considered model.

Table 2: Fitted distributions of the data set and criteria for comparison

Distribution	MLE	S.E	$-2\log L$	AIC	BIC	AICC
Exponentiated	$\hat{\alpha} = 0.3967$ $\hat{\theta} = 0.1924$	$\hat{\alpha} = 0.0491$ $\hat{\theta} = 0.0201$	829.448	833.448	839.152	833.544
Adya	$\hat{\theta} = 0.3212$	$\hat{\theta} = 0.015$	891.2774	893.2774	896.1295	893.3091

From table 2, it can be observed that the exponentiated Adya distribution have the lesser AIC, BIC, AICC and $-2\log L$ values as compared to Adya distribution. Hence we can conclude that the exponentiated Adya distribution leads to a better fit than the Adya distribution.

8. Conclusion

In this article, we have introduced a new generalization of Adya distribution called as exponentiated Adya distribution with two parameters (scale and shape). The subject distribution is generated by using the exponentiated technique and the parameters have been obtained by using the maximum likelihood estimator. Some statistical properties along with reliability measures are discussed. The new distribution with its applications in real life-time data has been demonstrated. Finally, the result of a real lifetime data set has been compared with Adya distribution and it has been found that the exponentiated Adya distribution provides a better fit than the Adya distribution.

References

- [1] Ganie, R. A. and Rajagopalan, V. (2022). Exponentiated Aradhana distribution with properties and applications in engineering sciences. *Journal of Scientific Research*, Vol 66, issue 1, pp 316-325.
- [2] 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.
- [3] Gupta, R.D. and Kundu, (2001). Exponentiated exponential family: an alternative to gamma and Weibull. *Biometrical Journal.*, 43(1), 117- 130.
- [4] Hassan, A. S. and Abdelghafar, M. A., (2017). Exponentiated Lomax Geometric Distribution: Properties and Applications. *Pakistan Journal of Statistics and Operation Research*, 13(3).
- [5] Lee, E. T and Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis*. 3rd ed. New York, NY, USA: John Wiley & Sons, 512.
- [6] Nasiru, S., Mwita, P. N. and Ngesa, O., (2018). Exponentiated Generalized Half Logistic Burr X Distribution. *Advances and Applications in Statistics.*, 52 (3), 145-169.
- [7] Pal, M., Ali, M. M. and Woo, J., (2006). Exponentiated weibull distribution. *Statistica, auno*, LXVI.,2, 139-147.
- [8] Rather, A. A. and Subramanian, C., (2020). A New Exponentiated Distribution with Engineering Science Applications. *Journal of Statistics, Applications and Probability*, 9(1), 127-137.
- [9] Rather, A. A. and Subramanian, C., (2019). Exponentiated Ishita Distribution with Properties and Application. *International Journal of Management, Technology and Engineering*, 9(5), 2473-2484.
- [10] Rather, A. A. and Subramanian, C., (2018). Exponentiated Mukherjee-Islam Distribution. *Journal of Statistics Applications & Probability*, 7(2), 357-361.
- [11] Rather, A.A., Subramanian, C., Al-Omari, A.I & Alanzi, A.R.A., (2022). Exponentiated Ailamujia distribution with statistical inference and applications of medical data. *Journal of Statistics and Management Systems*, DOI: 10.1080/09720510.2021.1966206
- [12] Rodrigues, J., Percontini, A. and Hamedani, G., (2017). The Exponentiated Generalized Lindley Distributio. *Asian Research Journal of Mathematics*, 5(3), 1-14.
- [13] Shanker, R., Shukla, K. K., Amaresh Ranjan and Ravi Shanker, (2021). Adya distribution with properties and Applications. *Biometrics & Biostatistics International Journal*, 10(3), 81-88.
- [14] Subramanian, C. and Rather, A A., (2019), Exponentiated Power Distribution: Properties and Estimation. *Science, Technology and Development*, 12(9), pp 213-223.

A review on Quantile functions, Income distributions, and Income inequality measures

Ashlin Varkey¹, Haritha N Haridas²

•

^{1,2}Farook College (Autonomous), Department of Statistics, Kozhikode, 673632, Kerala, India.

ashlinvarkey@gmail.com , haritasuhas@gmail.com

Abstract

The modeling of income data has originated a century ago with the work of Vilfredo Pareto. Since then several authors have added immense literature about income distributions and income inequality measures. In the present paper, we have pointed out some recent works in income distributions and income inequality measures. Recently the potential of the quantile function has been discussed in parallel with the distribution function for modeling reliability and income data. So we have also included some quantile functions existing in the literature and their potentiality to model income data in the review. We derived the Lorenz curve, Gini index, Pietra index, Bonferroni index, Bonferroni curve, and Zenga curve for six distributions and checked the model adequacy of three distributions using real income data.

Keywords: Income distribution, Income Inequality measures, Quantile based reliability models, Quantile function, Distribution function.

1. Introduction

We can consider the introduction of the Pareto model by Vilfredo Pareto as the beginning of the study of income distributions. A variety of models such as Lognormal, Dagum, Singh Maddala, etc., came into the literature for modeling income data. Several income inequality measures also exist in the literature. Kakwani [1] made a detailed study on income distributions, income inequalities, government policies affecting personal income distributions, and the measurement of poverty. Dagum [2] introduced the economic distance ratio, which is used to determine the degree of income inequality between two populations. For a detailed review of income distributions and income inequality measures since the beginning one can refer to [3]. This review tried to bring all the income distributions and income inequality measures that came in the literature under one umbrella.

Most of the income distributions adopted the distribution function approach. Only a very few have used the quantile function approach to model income data [3, 4]. So in the present review, we have considered quantile functions that can be used as income models.

The study of probability distribution in applied problems can be accomplished in two different ways; one by specifying the distribution function and the other through the quantile function. The quantile function is defined as,

$$Q(u) = F^{-1}(u) = \inf(x | F(x) \geq u), 0 \leq u \leq 1. \quad (1)$$

Since $F(x) \geq u$ iff $Q(u) \leq x$, the knowledge of the form of $Q(u)$ is equivalent to the knowledge of the functional form of $F(x)$. Gilchrist [5] provided a thorough explanation of quantile function and

model-building principles.

After this introduction in Section 2, we have reviewed some income distributions in which the distribution function is in the closed form. In Section 3 we have discussed some quantile functions which can be used as income models. We have reviewed some more quantile functions which have applications in reliability in Section 4. In Section 5 we have discussed income inequality measures existing in the literature and derived inequality measures for six distributions having closed-form quantile functions. We carried out a real data analysis in Section 6. Finally, the conclusions of the study are given in Section 7.

2. Distribution function based Income distributions

McDonald et al. [6] found that the Weibull, the Dagum, the Generalized beta of the second kind (GB2) gave the best fit among 5 parameter Generalized beta and its 10 special cases, that were fitted into 82 income data sets. These data sets comprise income data for 23 nations, covering both developed and emerging economies, and were taken from Luxembourg Income Study (LIS) database. He also observed that for almost every country, the Gini index increases monotonically over time.

Exponential Kumaraswamy - Dagum (EKD) distribution for analysis of income and lifetime data was introduced by [7]. The Cumulative Distribution Function (CDF) of EKD distribution is given by,

$$F_{EKD}(x; \alpha, \lambda, \delta, \phi, \theta) = \left\{ 1 - \left[1 - \left(1 + \lambda x^{-\delta} \right)^{-\alpha} \right]^{\phi} \right\}^{\theta} \quad (2)$$

where $\alpha, \lambda, \delta, \phi, \theta > 0$ and $x > 0$. Its basic statistical properties like moments, hazard functions, mean and median deviations, measures of income inequalities like Lorenz and Bonferroni curves, reliability measures, k^{th} order statistics, and Renyi entropy were derived. Maximum likelihood estimation was done to estimate the parameters of EKD and its application to real data was also presented.

An application of Gamma distribution to income distributions was explained by [8]. Its parameters were estimated based on the income quantile data, using non-linear optimization and the least square method. The future distribution of income and hunger issues were also discussed in this paper. They used the Gamma distribution to compare the estimates of people living in absolute poverty to the World Bank's report and estimated the number of people who are hungry in each country based on the results of a crop market model.

Fabio Clementi et al. [9] have surveyed k-generalized distribution for modeling income and wealth size distributions. They studied the k-generalized model [10, 11, 12] for the distribution of income, the k-generalized mixture model for the distribution of wealth, and the extended k-generalized distributions of the first and second kinds. They discussed its basic properties, interrelations with other distributions; income inequalities like Gini Index and Lorenz Curve. They concluded that a good fit for the distribution of income and wealth is given by k-generalized models.

A three-parameter Weibull-Pareto (WP) distribution with CDF,

$$F(x; b, \alpha, \theta) = 1 - e^{-\left[\left(\frac{x}{\theta} \right)^{\alpha} - 1 \right]^b}; x > \theta > 0, \alpha > 0, b > 0 \quad (3)$$

was proposed by [13]. Various properties of WP distribution including moments, incomplete moments, mean deviations, mode, reliability measures, generating functions, quantile functions, Bonferroni and Lorenz curves were derived. They also obtained density of order statistics, Renyi, and q-entropy of the above distribution. The parameters of the above distribution were estimated using the method of maximum likelihood. On two real-life data sets, WP distribution provides a better fit than comparable lifetime models.

Calderin-Ojeda et al. [14] proposed two extensions of the Exponential distribution ie, the Exponential Arc Tan (EAT) model and the composite EAT-Lognormal model. These models can

describe income distributions even with zero income. The adequacy of these models was evaluated using Australian income data for the years 2001-2012 and found that the proposed models provide a better fit for the data than exponential and gamma distribution.

Six mixture distributions based on Weibull ie, Weibull-Paralogistic, Weibull-Weibull, Weibull-Fisk, Weibull-Gamma, Weibull-Burr, and Weibull-Dagum to model income data were proposed by [15]. Here the method of maximum likelihood is used to estimate the parameters of Weibull mixture models and are evaluated with respect to average income per tax unit data for ten countries. Also, the application of income inequality measures like Gini, Bonferroni, and generalized index and mathematical expression of poverty measures like headcount ratio and poverty gap ratio was discussed in this paper.

A comparative study was taken by [16] on the distribution of income of poor households in Malaysia. They compared the models Reverse Pareto [17, 18] shifted reverse exponential, shifted reverse stretched exponential [19], and shifted reverse lognormal. They derived expressions for the Lorenz curve and Gini index for the four models. Results from Kolmogorov Smirnov (KS) test and R^2 coefficient conclude that the Reverse Pareto distribution adequately describes the distribution of income in poor households. Also, the Lorenz curve and Gini index based on Reverse Pareto models showed that poor households have evenly distributed incomes.

The historical evidence, empirical properties, the relationship between distribution functions, and model selection of top income distributions were reviewed by [20]. He summarized that for modeling income, the Generalized Pareto and Generalized Beta family of 3-4 parameters, which includes Dagum, Singh-Maddala, and GB2 distributions are invariably successful.

3. Quantile function based Income distributions

Sodomova et al. [21] did the statistical analysis and modeling on a sample of the yearly net income of 1566 households in the Slovak Republic in the year 2002. Results showed that except for the intervals having the lowest and highest income, the Weibull distribution with parameters from maximum likelihood estimation provides the best fit. Whereas for the intervals two fitted distributions from the Weibull-Pareto class, which is obtained by the method of modeling with quantile distributions provided the best fit.

Hankin and Lee [22] introduced and studied distribution with quantile function,

$$Q(u) = \frac{Cu^{\lambda_1}}{(1-u)^{\lambda_2}} \tag{4}$$

where $C, \lambda_1, \lambda_2 > 0$ and $0 \leq u \leq 1$. They studied the properties, shape, comparison with other distributions and calculated the moments of distribution in equation (4). The parameters of this distribution were estimated using maximum likelihood and the method of logged regression. In this paper, they compared the efficiencies of the above two estimation methods using simulation and found when the sample size is small and parameters are roughly equal we can use the regression method otherwise the maximum likelihood method. This distribution was also applied for modeling the toxic gas release. The quantile function in equation (4) is obtained as the product of quantile functions of Pareto and Power distribution, hence it has the potential for income modeling.

Haritha et al. [3] worked on the Modified Lambda family (MLF) with quantile function,

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2} \left[\frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4} \right] \tag{5}$$

where $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are real. The distributional characteristics of MLF were studied in detail. Major income distributions were obtained either as special or limiting cases or by approximation from MLF. They also expressed commonly used income inequality measures in quantile terms and calculated those measures for MLF. The parameters of MLF were evaluated using a new estimation

procedure involving location, dispersion, skewness, and kurtosis in quantile measures. Through a simulation study, it was shown that the above method of estimation is better than the method of percentiles and moments. They characterized income distributions using truncated Gini index and Income Gap Ratio at various ranges of poverty and affluence limit.

Using Dagum's three-parameter type 1 model [23] examined the change in personal income in Spain between 1995 and 2005. The quantile function of this distribution is given as,

$$Q(u) = \lambda^{\frac{1}{\delta}} \left[\left(\frac{1}{u} \right)^{\frac{1}{\beta}} - 1 \right]^{\frac{-1}{\delta}} \quad (6)$$

where $0 \leq u \leq 1$ and $\lambda, \beta, \delta > 0$. The model in equation (6) fits the empirical income distribution of Spain quite well. The economic interpretation of the Dagum model parameters was also studied. The data from the European Community Household Panel (ECHP) and the European Union Statistics on Income and Living Conditions (EU-SILC) is used to study the effect of parameter changes in the growth of inequality in Spain as well as on different income percentiles.

The distributional and geometric properties of partial moments of first and second order in quantile terms were studied by [24]. The r^{th} order partial moment is given by,

$$P_r(u) = \int_u^1 (Q(p) - Q(u))^r dp \quad (7)$$

where $Q(\cdot)$ is a quantile function. Stop-loss transform based on quantile function is mainly discussed in this paper. Relationships of income inequality measures like Lorenz, Gini, Bonferroni, and Leimkuhler curves with scaled stop-loss transform curves were also developed.

The capabilities of the Zenga curve as an inequality measure were explored by [25]. They made a detailed study on the properties of the Zenga curve and stochastic orders based on this curve were used to prove some results. They established a relationship between the Zenga curve and inequality measures like the Bonferroni curve and the Leimkuhler curve. Similarly, they established an association between the Zenga curve and reliability measures like mean residual quantile function and reversed mean residual quantile function. A study on quantile-based income distributions like the Govindarajulu distribution, quantile model with linear hazard quantile form, and Power \times Pareto distribution was made and derived measures of income inequality like the Lorenz curve, Bonferroni curve, etc. In this thesis, they also studied the relationship between L-moments and income inequality measures. Bivariate reliability concepts based on copula were also discussed.

Ekum et al. [26] proposed a six-parameter distribution named the exponentiated-exponential Dagum {Lomax} (EEDL) from T- Dagum{Y} family using T-R{Y} framework. Its quantile function is given by

$$Q_x(u) = \phi^{(\frac{1}{p})} \left\{ \left[1 - \left(1 - \frac{1}{\beta} \log \left(1 - u^{(\frac{1}{\alpha})} \right) \right)^{-\gamma} \right]^{-\frac{1}{q}} - 1 \right\}^{-(\frac{1}{p})} \quad (8)$$

where α, γ, q, p defines the shape and ϕ, β defines the spread of the distribution and $0 \leq u \leq 1$. The basic distributional and reliability characteristics including stochastic ordering, asymptotes, analysis of stress-strength, and Shannon entropy were studied. The parameters of EEDL distribution were estimated using the method of maximum likelihood. On two real data sets, the EEDL distribution did well in comparison with the Exponential Kumaraswamy Dagum (EKD), the Exponentiated Generalized Exponential Dagum (EKD), and the Mc Dagum (McD) distributions.

4. New quantile functions in reliability analysis

Nair et al. [27] discussed the method of constructing quantile functions for lifetime models by utilizing the relationship between the hazard quantile function and the Parzens score function [28]. Three models based on score function were illustrated in this paper and many known distributions exist as its special case. Various reliability properties of the Parzen score function were also studied. The reliability ideas in quantile terms were broadly explained by [29]. They also discussed distributions having closed-form quantile functions, ageing concepts, total time on test transforms, L-moments of residual life, hazard quantile function, stochastic orders, and modeling in a quantile framework.

Thomas et al. [30] introduced a software reliability model with a quantile function,

$$Q(u) = k\beta(u, a+1, b+1) \quad (9)$$

where $k > 0$, a, b are real numbers, $0 \leq u \leq 1$ and $\beta(u, a+1, b+1)$ is an incomplete beta function. Various distributional and reliability characteristics of the above quantile function were studied. They approximated equation (9), to two well-known distributions like Inverse Gaussian and Weibull. The method of L-moments was used to estimate the parameters of the model and applied the model in (9) to a real data set.

The reliability properties of the quantile-based proportional hazard model (PHM) were studied by [31]. The ageing properties and characterizations for the PHM were derived and demonstrated with examples. Certain important stochastic orders in the context of PHM were discussed. They also proposed the quantile-based dynamic cumulative residual Kullback-Leibler divergence of PHM.

A new class of distribution as the product of quantile functions of Weibull and Pareto distributions was proposed by [32] and it is given as,

$$Q(u) = \sigma(1-u)^{-\alpha} (-\log(1-u))^{\beta} \quad (10)$$

where $0 < u < 1$; $\alpha, \beta, \sigma > 0$. The distributional characteristics and reliability properties of distribution in equation (10) were studied in detail. The inference was done using L-moments and the model was applied to two real datasets.

Kumar and Paduthol [33] introduced a new class of distributions with quantile functions,

$$Q(u) = \sigma \left(\log \left(1 + \frac{u}{\beta(1-u)} \right) + \frac{\gamma u}{\beta + (1-\beta)u} \right) \quad (11)$$

where $\sigma > 0, \beta > 0$ and $\gamma > -1$, this class of distributions was obtained as an extension of distributions with linear mean residual quantile function. The distributional characteristics, reliability properties, and L-moments of equation (11) were calculated. The method of percentiles was employed to estimate the parameters of the model in (11). As an application, the above model was applied to real data reported in [34], consisting of the strength of glass fibers.

A class of distributions with quadratic hazard quantile function was developed by [35] and is given as,

$$Q(u) = \frac{1}{2(\beta-\alpha)} \log(1+u) - \frac{1}{2(\beta+\alpha)} \log(1-u) - \frac{\alpha}{(\beta^2-\alpha^2)} \log\left(\frac{\beta+\alpha u}{\beta}\right) \quad (12)$$

where, $\beta > 0, \beta > |\alpha|, \beta \neq \alpha$. They studied distributional properties, reliability characteristics, and characterization of the distribution. Estimation was done using the method of least squares and demonstrated the models' utility using real data.

Ghosal et al. [36] used subject-specific quantile functions to capture the distributional nature of wearable data. They used these quantile functions L-moment representations in Scalar-On-Function Regression (SOFR) model [37], Functional Generalized Additive Model (FGAM) [38], and

Joint and Individual Variation Explained(JIVE) method [39]. As an application, they illustrated the proposed method in the study of Alzheimer's Disease(AD).

Power exponential geometric distribution was introduced by [40] as the sum of the quantile function of the power and the exponential geometric distributions. The distributional and reliability properties were studied. A simulation study was done and applied the model to real data.

5. Income inequality measures

A book on modeling income distributions and the Lorenz curves was edited by [41]. A compilation of five major papers in this area makes the first part and four survey papers on Lorenz functions, as well as generalizations and extensions of a few income distributions, are included in Part two. Eight papers on recent research and advancement in this field are included in the last section.

The distinction between the Lorenz curve in economics and the Leimkuhler curve in information science is that the Lorenz curve is used to arrange sources in increasing productivity order whereas for Leimkuhler is arranged in decreasing order. A general definition for the Leimkuhler curve was introduced by [42] in terms of theoretical CDF and is given as,

$$K_X(u) = \frac{1}{\mu_X} \int_{1-u}^1 F_X^{-1}(y) dy \quad (13)$$

where $0 \leq u \leq 1$. The discrete, continuous, and mixed random variables are covered by the equation (13). In this paper, they derived the Leimkuhler curves expressions for five continuous, one mixed, and one discrete distribution.

The Bonferroni curve and Bonferroni index have applications not only in the field of economics to study poverty and income but also in the field of insurance, medicine, demography, and reliability. For thirty-five continuous distributions, [43] provided explicit expressions for Bonferroni Curve, Bonferroni Index, Lorenz curve, and Gini index.

Fellman [44] studied two optimal cases, where the transformed variable Lorenz dominates the initial variable and the initial variable Lorenz dominates the transformed one. The first case has more practical application than the second because it results in policies that decrease inequality. The properties and limits of the transformed Lorenz curve were also analyzed. In this work, the limits found are valid for a wide range of distributions and transformations but the inequalities that result from pursuing general conditions cannot be improved.

Chotikapanich et al. [45] discussed poverty measures like the head-count ratio, the Foster-Greer-Thorbecke (FGT) measure, the Atkinson index, the Watts index, the Sen index, and the Gini index and derived their expression for GB2 distribution. An analysis of poverty trends in South and South East Asian nations is done using beta 2 distribution, which is a special case of GB2.

Using the semiparametric method, [46] estimated the Lorenz curve and Gini index for the exponential distribution. The above estimation was done under type 1, type 2, and interval censoring. From Monte Carlo simulation studies they found, that as sample size increases, the mean square error (MSE) of the estimator decreases.

Fellman [47] comprehensively explained the Lorenz curve and gave a brief description of the Gini index and Pietra index. As an application, the changes in Lorenz curves, Gini, and Pietra indices with respect to model parameters of Pareto [48], the simplified Rao Tam [49], and the Chotikapanich [50] distributions were also given. When the above three models have the same Gini index, the Lorenz curves for the simplified Rao Tam and the Chotikapanich models are rather similar but Pareto's is different.

Behdani & Mohtashami Borzadaran [51] reformulated certain income inequality measures using quantiles. They used the relationship between the Lorenz curve and reliability concepts like mean residual quantile function and reversed mean residual quantile function to characterize probability distributions. They also studied ageing concepts using the Lorenz curve and quantile function.

Kattumannil et al. [52] proposed a non-parametric estimation of the Gini index for right-censored observations in the sample. The propound estimator has an asymptotic normal distribution and is consistent. Monte Carlo simulation was used to determine the potential of the above estimator. The simulation study showed that the confidence interval of the Gini index based on the proposed estimator has good coverage probability and can be implemented very easily.

A detailed literature review on the association between income inequality and economic growth was done by [53]. To comprehend how income inequality and growth are related, theoretical and empirical literature is studied and analyzed.

Table 1 lists major income inequality measures and curves of the six distributions covered in this review. The distributions having explicit quantile functions and those for which income inequality measures exist in closed form are taken for the calculation of inequality measures.

Expressions used in Table 1

- (i) LC, GI, PI, BC, BI, and ZC denote the Lorenz curve, Gini index, Pietra index, Bonferroni curve, Bonferroni index, and Zenga curve respectively.
- (ii) $\beta_u(.,.)$ denotes the incomplete beta function.
- (iii) u_0 in Pietra indexes can be obtained by solving u the equation $\mu = Q(u)$.
- (iv) $\text{csc}(\cdot)$ denotes the *coscant*(\cdot) function in trigonometry.
- (v) $Har.no(n)$ denotes the n^{th} harmonic number and is given as $Har.no(n) = \sum_{k=1}^n k^{-1}$.
- (vi) $\Gamma(\cdot)$ denotes the gamma function.
- (vii) $\Gamma(.,.)$ denotes incomplete gamma function.
- (viii) $I_u(.,.)$ is the regularized incomplete beta function and is given as $I_u(.,.) = \frac{\beta_u(.,.)}{\beta(.,.)}$.
- (ix) $\frac{1}{\beta(a+1, b+2)} \left\{ \frac{\pi \text{csc}(b\pi) \Gamma(a+2) [Har.no(a+1) - Har.no(a+b+2)]}{\Gamma(-b) \Gamma(a+b+3)} \right\}$
- (x) $\frac{2ku_0(1-u_0)^{-\frac{1}{\alpha}}(1-(1-u_0)^{2k})^{\frac{1}{\alpha}} - \beta_{1-(1-u_0)^{2k}}(\frac{1}{\alpha}+1, \frac{1}{2k} - \frac{1}{2\alpha})}{\beta(\frac{1}{\alpha}+1, \frac{1}{2k} - \frac{1}{2\alpha})}$
- (xi) $\frac{1}{\mu} \left\{ (\gamma-1)^{-1}(\gamma-2)^{-1}(1-u)^{-\gamma} \left[(1-u)^\gamma + u(u+\gamma-u\gamma) - 1 \right] \eta + \lambda u \right.$
 $\left. + 2^{-(\beta+1)} \eta \left[\Gamma(\beta+1) - \Gamma(\beta+1, -2\ln(1-u)) \right] \right\}$
- (xii) $\frac{-3^{-(\beta+1)} \eta \left[6^{\beta+1}(\gamma+1) + (2^{\beta+2} - 3^{\beta+1})(\gamma-1)(\gamma-2)(\gamma-3)\Gamma(\beta+1) \right]}{(\gamma-3) \left\{ 2^{\beta+1} \left[\eta + (\gamma-1)(\gamma-2)\lambda \right] + \eta(\gamma-1)(\gamma-2)\Gamma(\beta+1) \right\}}$
- (xiii) $\frac{\eta}{\mu} \left\{ u_0(1-u_0) \left[-\ln(1-u_0) \right]^\beta + (\gamma-1)^{-1}(\gamma-2)^{-1}(1-u_0)^{-\gamma} \left[u_0^2(\gamma-1)^2 - u_0\gamma + 1 - (1-u_0)^\gamma \right] \right.$
 $\left. - 2^{-(\beta+1)} \left[\Gamma(\beta+1) - \Gamma(\beta+1, -2\ln(1-u_0)) \right] \right\}$
- (xiv) $\frac{1}{\mu} \left\{ (\gamma-1)^{-1}(\gamma-2)^{-1}u^{-1}(1-u)^{-\gamma} \left[(1-u)^\gamma + u(u+\gamma-u\gamma) - 1 \right] \eta + \lambda \right.$
 $\left. + 2^{-(\beta+1)} \eta u^{-1} \left[\Gamma(\beta+1) - \Gamma(\beta+1, -2\ln(1-u)) \right] \right\}$

$$(xv) \quad 1 - \int_0^1 \frac{1}{\mu} \left\{ (\gamma - 1)^{-1} (\gamma - 2)^{-1} u^{-1} (1 - u)^{-\gamma} \left[(1 - u)^\gamma + u(u + \gamma - u\gamma) - 1 \right] \eta + \lambda \right. \\ \left. + 2^{-(\beta+1)} \eta u^{-1} \left[\Gamma(\beta + 1) - \Gamma(\beta + 1, -2 \ln(1 - u)) \right] \right\} du$$

(xvi) In expressions (xi), (xiii), (xiv), and (xv), μ is the mean of the Weibull-Pareto distribution and is given as $\mu = \lambda + \frac{\eta}{(\gamma - 1)(\gamma - 2)} + 2^{-(\beta+1)} \eta \Gamma(\beta + 1)$

Table 1: Income inequality measures and inequality curves

Quantile function		Income inequality measures/ inequality curves
		$\frac{[u\beta_u(a + 1, b + 1) - \beta_u(a + 2, b + 1)]}{\beta(a + 1, b + 2)}$
	LC	
	GI	$(a + 1)(a + b + 3)^{-1}$
	PI	$\frac{\beta_{u_0}(a + 2, b + 1)}{\beta(a + 1, b + 2)}$
Bijamma Thomas distribution	BC	$\frac{1}{\beta(a + 1, b + 2)} \left[\beta_u(a + 1, b + 1) - \frac{1}{u} \beta_u(a + 2, b + 1) \right]$
$Q(u) = k\beta(u, a + 1, b + 1)$	BI	(ix)
	ZC	$\frac{\beta(a + 1, b + 2) - \beta_u(a + 1, b + 1) + \frac{1}{u} \beta_u(a + 2, b + 1)}{\beta(a + 1, b + 2) - u\beta_u(a + 1, b + 1) + \beta_u(a + 2, b + 1)}$
		$\frac{u(\ln u + \lambda x_0 - 1)}{\lambda x_0 - 1}$
	LC	$\frac{1}{2(\lambda x_0 - 1)}$
	GI	$\frac{u_0}{\lambda x_0 - 1}$
Shifted reverse exponential distribution	PI	$\frac{\ln u + \lambda x_0 - 1}{\lambda x_0 - 1}$
$Q(u) = x_0 + \frac{1}{\lambda} \ln(u)$	BC	$\frac{1}{\lambda x_0 - 1}$
	BI	$\frac{1}{\ln u}$
	ZC	$\frac{1 + u(\ln u - 1) - \lambda x_0(1 - u)}{\lambda x_0 - 1}$
	LC	$\frac{x_0 u - \lambda \Gamma(1 + \frac{1}{\rho}, -\ln u)}{x_0 - \lambda \Gamma(1 + \frac{1}{\rho})}$
	GI	$(1 - 2^{-\frac{1}{\rho}}) \frac{\lambda \Gamma(1 + \frac{1}{\rho})}{\left[x_0 - \lambda \Gamma(1 + \frac{1}{\rho}) \right]}$
Shifted reverse stretched exponential distribution	PI	$\frac{\lambda \Gamma(1 + \frac{1}{\rho}, -\ln u_0) - \lambda u_0 (-\ln u_0)^{\frac{1}{\rho}}}{x_0 - \lambda \Gamma(1 + \frac{1}{\rho})}$
$Q(u) = x_0 - \lambda(-\ln u)^{\frac{1}{\rho}}$	BC	$\frac{x_0 - \lambda u^{-1} \Gamma(1 + \frac{1}{\rho}, -\ln u)}{x_0 - \lambda \Gamma(1 + \frac{1}{\rho})}$

	BI	$\frac{\lambda\Gamma(1 + \frac{1}{\rho})}{x_0\rho - \lambda\Gamma(\frac{1}{\rho})}$
	ZC	$\frac{\lambda\left[\Gamma(1 + \frac{1}{\rho}, -\ln u) - u\Gamma(1 + \frac{1}{\rho})\right]}{u\left\{x_0(1 - u) + \lambda\left[\Gamma(1 + \frac{1}{\rho}, -\ln u) - \Gamma(1 + \frac{1}{\rho})\right]\right\}}$
	LC	$I_{1-(1-u)^{2k}}\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right)$
	GI	$1 - \frac{2\beta\left(\frac{1}{\alpha} + 1, \frac{1}{k} - \frac{1}{2\alpha}\right)}{\beta\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right)}$
K-Generalized distribution	PI	$\frac{1}{u} I_{1-(1-u)^{2k}}^{(x)}\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right)$
$Q(u) = \beta\left[\ln_K\left(\frac{1}{1-u}\right)\right]^{\frac{1}{\alpha}}$	BC	$1 - \int_0^1 \frac{1}{u} I_{1-(1-u)^{2k}}\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right) du$
	BI	$\frac{u\beta\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right) - \beta_{1-(1-u)^{2k}}\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right)}{u\left[\beta\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right) - \beta_{1-(1-u)^{2k}}\left(\frac{1}{\alpha} + 1, \frac{1}{2k} - \frac{1}{2\alpha}\right)\right]}$
	ZC	$\frac{\beta + 2}{(\beta + 2)\theta + 2\sigma} \left[\theta u + \sigma u^{\beta+1} \left(\frac{\beta + 2 - \beta u}{\beta + 2} \right) \right]$
		LC
Govindarajulu distribution	PI	$\frac{\sigma\beta u_0^{\beta+1}}{[(\beta + 2)\theta + 2\sigma]} [(\beta + 2) - (\beta + 1)u_0]$
$Q(u) = \theta + \sigma[(\beta + 1)u^\beta - \beta u^{\beta+1}]$	BC	$\frac{\beta + 2}{(\beta + 2)\theta + 2\sigma} \left[\theta + \sigma u^\beta \left(\frac{\beta + 2 - \beta u}{\beta + 2} \right) \right]$
	BI	$\frac{\sigma\beta(2\beta + 3)}{(\beta + 1)(\beta + 2)[(\beta + 2)\theta + 2\sigma]}$
		$\frac{(\beta + 2)}{(\beta + 2)\theta + 2\sigma} - \left[\theta + \sigma u^\beta \left(\frac{\beta + 2 - \beta u}{\beta + 2} \right) \right]$
	ZC	$\frac{(\beta + 2)}{(\beta + 2)\theta + 2\sigma} - u \left[\theta + \sigma u^\beta \left(\frac{\beta + 2 - \beta u}{\beta + 2} \right) \right]$
Weibull-Pareto distribution	LC	(xi)
	GI	(xii)
$Q(u) = \lambda + \eta \left\{ (1-u)[- \ln(1-u)]^\beta + \frac{u}{(1-u)^\gamma} \right\}$	PI	(xiii)
	BC	(xiv)
	BI	(xv)
	ZC	(xvii)

$$(xvii) \quad \frac{\eta \left\{ \frac{2(1-u)^{-\gamma+1} [(1-u)^\gamma + \gamma u - u - 1]}{(\gamma-1)(\gamma-2)} + 2^{-\beta} [\beta\Gamma(\beta) - u\Gamma(\beta+1) - \Gamma(\beta+1, -2\ln(1-u))] \right\}}{u(1-u) \left[\frac{(1-u)^{-\gamma} (\gamma u - u - 1)\eta - \lambda}{(\gamma-1)(\gamma-2)} - 2^{-(\beta+1)} \eta u \Gamma(\beta+1, -2\ln(1-u)) \right]}$$

6. Data Analysis

This section analyses 2020's per capita personal income (in dollars) of 254 counties in Texas, US. This data is available from <https://www.bea.gov> and is used for studying the potential of Shifted Reverse Exponential, Govindarajulu, and Weibull Pareto distribution in income modeling. Here we use the method of percentiles for estimating the parameters of the above three distributions.

Here, the Chi-square (χ^2) test and Q-Q plot are used to determine model adequacy. Table 2 provides the parameter estimates, χ^2 test statistics, and p-values of Shifted Reverse Exponential, Govindarajulu, and Weibull Pareto distributions. Table 2 and the Q-Q plot given in Figure 1 make it evident that the Weibull Pareto distribution provides the best fit for the real data.

Table 2: Parameter Estimates, χ^2 Statistic, p-value

Distribution	Parameter Estimates	χ^2 Statistic	p-value
Shifted Reverse Exponential	$x_0 = 5.472113 \times 10^4$ $\lambda = 8.922956 \times 10^{-5}$	68.39832	8.41825×10^{-8}
Govindarajulu	$\theta = 3.85826 \times 10^4$ $\sigma = 2.605615 \times 10^4$ $\beta = 2.943906$	78.84319	1.36464×10^{-9}
Weibull Pareto	$\lambda = 3.429882 \times 10^4$ $\eta = 1.177798 \times 10^4$ $\beta = 5.374932 \times 10^{-1}$ $\gamma = 4.088145 \times 10^{-1}$	16.73783	0.54119

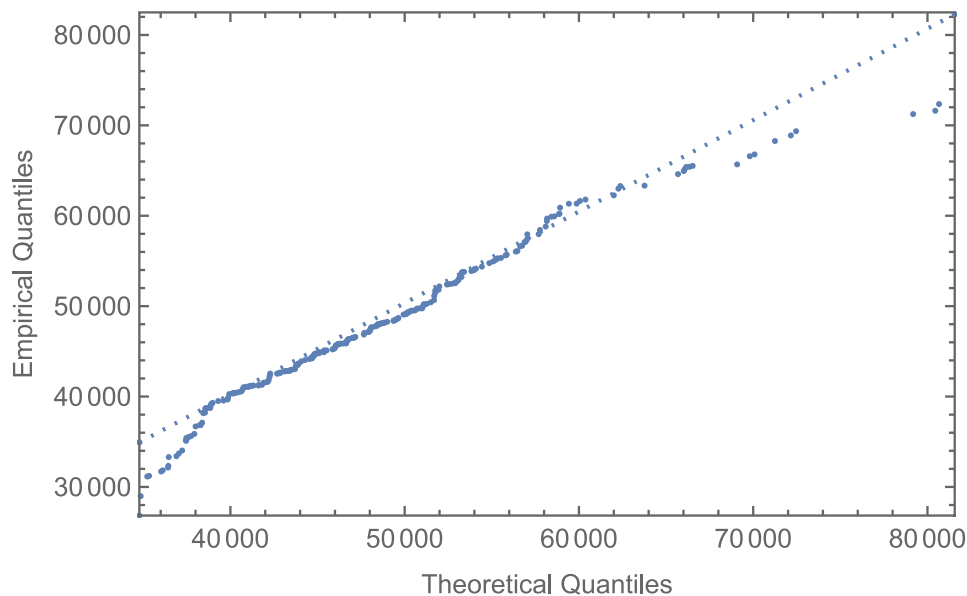


Figure 1: Q-Q plot corresponding to Weibull Pareto distribution

7. Conclusion

In this work, we reviewed recent income distributions, income inequalities, and quantile functions that appeared in the literature. This study is carried out in five sections comprising income distributions, income models based on quantiles, new quantile functions in reliability analysis, income inequality measures, and data analysis. For the six distributions examined in this work, the Lorenz curve, Gini index, Pietra index, Bonferroni index, Bonferroni curve, and Zenga curve were determined. Three models were applied to the per capita personal income data of 254 counties in Texas State and found that the Weibull Pareto distribution provides the best fit. In future works, we can check whether more quantile functions used in reliability analysis have potential in income modeling.

Acknowledgment

The first author is thankful to Kerala State Council for Science, Technology and Environment (KSCSTE) for the financial support.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- [1] N. C. Kakwani, *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. 1980.
- [2] C. Dagum, "Inequality Measures between Income Distributions with Applications," *Econometrica*, vol. 48, no. 7, pp. 1791, 1980, doi: 10.2307/1911936.
- [3] H. Haritha N., K. R. M. Nair, and N. U. Nair, "Income Modeling Using Quantile functions," 2007.
- [4] A. Tarsitano, "Fitting the Generalized Lambda," *Compstat 2004 Symposium*, 2004, pp. 1861–1867.
- [5] W. G. Gilchrist, *Statistical Modeling with Quantile Functions*. Chapman and Hall/CRC, New York, 2000.
- [6] J. B. McDonald, R. Bandourian, and R. S. Turley, "A Comparison of Parametric Models of Income Distribution Across Countries and Over Time," *SSRN Electron. J.*, 2002.
- [7] S. Huang and B. O. Oluyede, "Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data," *J. Stat. Distrib. Appl.*, vol. 1, 2014, doi: 10.1186/2195-5832-1-8.
- [8] S. Mori, D. Nakata, and T. Kaneda, "An Application of Gamma Distribution to the Income Distribution and the Estimation of Potential Food Demand Functions," *Mod. Econ.*, vol. 06, no. 09, pp. 1001–1017, 2015, doi: 10.4236/me.2015.69095.
- [9] F. Clementi, M. Gallegati, G. Kaniadakis, and S. Landini, " κ -generalized models of income and wealth distributions: A survey," *Eur. Phys. J. Spec. Top.*, vol. 225, no. 10, pp. 1959–1984, Oct. 2016, doi: 10.1140/epjst/e2016-60014-2.
- [10] F. Clementi, M. Gallegati, and G. Kaniadakis, " κ -generalized statistics in personal income distribution," *Eur. Phys. J. B*, vol. 57, no. 2, pp. 187–193, 2007, doi: 10.1140/epjb/e2007-00120-9.
- [11] F. Clementi, T. Di Matteo, M. Gallegati, and G. Kaniadakis, "The κ -generalized distribution: A new descriptive model for the size distribution of incomes," *Phys. A Stat. Mech. its Appl.*, vol. 387, no. 13, pp. 3201–3208, 2008, doi: 10.1016/j.physa.2008.01.109.
- [12] F. Clementi, M. Gallegati, and G. Kaniadakis, "A κ -generalized statistical mechanics approach to income analysis," *J. Stat. Mech. Theory Exp.*, vol. 2009, no. 2, 2009, doi: 10.1088/1742-5468/2009/02/P02037.
- [13] M. H. Tahir, G. M. Cordeiro, A. Alzaatreh, M. Mansoor, and M. Zubair, "A New Weibull–Pareto Distribution: Properties and Applications," *Commun. Stat. Simul. Comput.*, vol. 45, no. 10, pp. 3548–3567, 2016, doi: 10.1080/03610918.2014.948190.

- [14] E. Calderín-Ojeda, F. Azpitarte, and E. Déniz, "Modelling income data using two extensions of the exponential distribution," *Phys. A Stat. Mech. its Appl.*, vol. 461, pp. 756–766, 2016, doi: 10.1016/j.physa.2016.06.047.
- [15] S. A. A. Bakar and D. Pathmanathan, "Income modeling with the Weibull mixtures," *Commun. Stat. - Theory Methods*, 2020, doi: 10.1080/03610926.2020.1800737.
- [16] M. A. M. Safari, N. Masseran, K. Ibrahim, and S. I. Hussain, "Modeling the income distribution of poor households in Malaysia," *AIP Conf. Proc.*, vol. 2266, no. October, 2020, doi: 10.1063/5.0018066.
- [17] M. A. M. Safari, N. Masseran, K. Ibrahim, and S. I. Hussain, "A robust and efficient estimator for the tail index of inverse Pareto distribution," *Phys. A Stat. Mech. its Appl.*, vol. 517, pp. 431–439, 2019, doi.org/10.1016/j.physa.2018.11.029.
- [18] C. Kleiber and S. Kotz, *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons, Ltd, 2003.
- [19] M. Brzezinski, "Power laws in citation distributions: evidence from Scopus," *Scientometrics*, vol. 103, no. 1, pp. 213–228, 2015, doi: 10.1007/s11192-014-1524-z.
- [20] V. Hlasny, "Parametric representation of the top of income distributions: Options, historical evidence, and model selection," *J. Econ. Surv.*, vol. 35, no. 4, pp. 1217–1256, 2021, doi.org/10.1111/joes.12435.
- [21] E. Sodomova, B. Pacak, and L. Sipkova, "Models of households in the Slovak Republic .," *Statistics in management of social and economic development: 11th Ukrainian-Polish-Slovak scientific conference, 20–22 October 2004*, 2005, pp. 80–91.
- [22] R. K. S. Hankin and A. Lee, "A new family of non-negative distributions," *Aust. New Zeal. J. Stat.*, vol. 48, no. 1, pp. 67–78, 2006, doi: 10.1111/j.1467-842X.2006.00426.x.
- [23] C. G. Pérez and M. P. Alaiz, "Using the Dagum model to explain changes in personal income distribution," *Appl. Econ.*, vol. 43, no. 28, pp. 4377–4386, 2011, doi: 10.1080/00036846.2010.491459.
- [24] N. U. Nair, P. G. Sankaran, and S. M. Sunoj, "Quantile based stop-loss transform and its applications," *Stat. Methods Appl.*, vol.22, no.2, pp.167–182, 2013,doi:10.1007/s10260-012-0213-4.
- [25] N. Sreelakshmi and K. Nair, "A quantile based analysis of income data," 2014.
- [26] M. I. Ekum, M. O. Adamu, and E. E. Akarawak, "T-Dagum: A Way of Generalizing Dagum Distribution Using Lomax Quantile Function," *J. Probab. Stat.*, vol. 2020, 2020, doi: 10.1155/2020/1641207.
- [27] N. U. Nair, P. G. Sankaran, and B. V. Kumar, "Modelling lifetimes by quantile functions using Parzen's score function," *Statistics (Ber.)*, vol. 46, no. 6, pp. 799–811, 2012, doi: 10.1080/02331888.2011.555551.
- [28] E. Parzen, "Nonparametric Statistical Data Modeling," *J. Am. Stat. Assoc.*, vol. 74, no. 365, pp. 105–121, Mar. 1979, doi: 10.1080/01621459.1979.10481621.
- [29] N. Nair, P. G. Sankaran, and N. Balakrishnan, *Quantile-Based Reliability Analysis*. 2013.
- [30] B. Thomas, M. N. Nellikkattu, and S. Godan Paduthol, "A software reliability model using quantile function," *J. Probab. Stat.*, vol. 2014, no. March, 2014, doi: 10.1155/2014/951608.
- [31] N. U. Nair, P. G. Sankaran, and S. M. Sunoj, "Proportional hazards model with quantile functions," *Commun. Stat. - Theory Methods*, vol. 47, no. 19, pp. 4710–4723, 2018, doi: 10.1080/03610926.2018.1445858.
- [32] S. Paduthol and D. Kumar, "Power Pareto quantile function," *J. Appl. Probab. Stat.*, vol. 13, pp. 81–95, 2018.
- [33] D. Kumar Maladan and P. G. Sankaran, "A new family of quantile functions and its applications," *Commun. Stat. - Theory Methods*, vol. 50, no. 18, pp. 4216–4235, 2020, doi: 10.1080/03610926.2020.1713368.
- [34] R. L. Smith and J. C. Naylor, "A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull Distribution," *J. R. Stat. Soc. Ser. C (Applied Stat.)*, vol. 36, no. 3, pp. 358–369, 1987, doi.org/10.2307/2347795.

- [35] I. C. Aswin, P. G. Sankaran, and S. M. Sunoj, "A Class of Distributions with Quadratic Hazard Quantile Function," *J. Indian Soc. Probab. Stat.*, vol. 21, no. 2, pp. 409–426, 2020, doi: 10.1007/s41096-020-00088-6.
- [36] R. Ghosal *et al.*, "Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in Alzheimer's Disease," *Biostatistics*, 2021, doi: 10.1093/biostatistics/kxab041.
- [37] J. S. Morris, "Functional Regression," *Annu. Rev. Stat. Its Appl.*, vol. 2, no. 1, pp. 321–359, Apr. 2015, doi: 10.1146/annurev-statistics-010814-020413.
- [38] M. W. McLean, G. Hooker, A.-M. Staicu, F. Scheipl, and D. Ruppert, "Functional Generalized Additive Models," *J. Comput. Graph. Stat.*, vol. 23, no. 1, pp. 249–269, Jan. 2014, doi: 10.1080/10618600.2012.729985.
- [39] E. F. Lock, K. A. Hoadley, J. S. Marron, and A. B. Nobel, "Joint and individual variation explained (JIVE) for integrated analysis of multiple data types," *Ann. Appl. Stat.*, vol. 7, no. 1, pp. 523–542, Mar. 2013, doi: 10.1214/12-AOAS597.
- [40] J. Joseph and A. A.P., "Power -Exponential Geometric Quantile Function," *Reliab. Theory Appl.*, vol. 16, no. 4, pp. 294–307, 2021.
- [41] D. Chotikapanich, Ed., *Modeling Income Distributions and Lorenz Curves*. Springer, 2008.
- [42] J. M. Sarabia, "A general definition of the Leimkuhler curve," *J. Informetr.*, vol. 2, no. 2, pp. 156–163, 2008, doi: 10.1016/j.joi.2008.01.002.
- [43] G. M. Giorgi and S. Nadarajah, "Bonferroni and Gini indices for various parametric families of distributions," *Metron*, vol. 68, no. 1, pp. 23–46, 2010, doi: 10.1007/BF03263522.
- [44] J. Fellman, "Properties of Lorenz Curves for Transformed Income Distributions," *Theor. Econ. Lett.*, vol. 02, no. 05, pp. 487–493, 2012, doi: 10.4236/tel.2012.25091.
- [45] D. Chotikapanich, W. Griffiths, W. Karunarathne, and D. S. Prasada Rao, "Calculating Poverty Measures from the Generalised Beta Income Distribution," *Econ. Rec.*, vol. 89, no. S1, pp. 48–66, 2013, doi.org/10.1111/1475-4932.12031.
- [46] P. P. C. Pillai, G. Rajesh, and E. I. Abdul-Sathar, "Semi-Parametric Estimation of Lorenz Curve and Gini-index in an Exponential Distribution," *Sri Lankan J. Appl. Stat.*, vol. 15, no. 1, 2014, doi: 10.4038/sljastats.v15i1.6749.
- [47] J. Fellman, "Income Inequality Measures," *Theor. Econ. Lett.*, vol. 08, no. 03, pp. 557–574, 2018, doi: 10.4236/tel.2018.83039.
- [48] R. H. Rasche, J. Gaffney, A. Y. C. Koo, and N. Obst, "Functional Forms for Estimating the Lorenz Curve," *Econometrica*, vol. 48, no. 4, pp. 1061–1062, 1980, [Online]. Available: <http://www.jstor.org/stable/1912948>.
- [49] U. L. Gouranga Rao and A. Yuk-Pang Tam, "An empirical study of selection and estimation of alternative models of the Lorenz curve," *J. Appl. Stat.*, vol. 14, no. 3, pp. 275–280, Jan. 1987, doi: 10.1080/02664768700000032.
- [50] D. Chotikapanich, "A comparison of alternative functional forms for the Lorenz curve," *Econ. Lett.*, vol. 41, no. 2, pp. 129–138, 1993, doi.org/10.1016/0165-1765(93)90186-G.
- [51] Z. Behdani and G. R. Mohtashami Borzadaran, "Measures of Income Inequality Based on Quantile Function," 2019.
- [52] S. K. Kattumannil, I. Dewan, and S. N., "Non-parametric estimation of Gini index with censored observations," *Stat. Probab. Lett.*, vol. 175, no. April, 2021, doi: 10.1016/j.spl.2021.109113.
- [53] K. Mdingi and S.-Y. Ho, "Literature review on income inequality and economic growth," *MethodsX*, vol. 8, p. 101402, 2021, doi.org/10.1016/j.mex.2021.101402.

THE EXTENDED ERLANG TRUNCATED EXPONENTIAL DISTRIBUTION: PROPERTIES AND APPLICATION TO STATISTICAL PREDICTION PROBLEM

Imtiyaz A. Shah

•

Department of Community Medicine (Bio Statistics), Sher-I-Kashmir
Institute of Medical Sciences, Srinagar (J&K) India
driashah03@gmail.com

Abstract

The Erlang Truncated (ETE) distribution is modified and the new lifetime distribution is called the Extended Erlang Truncated Exponential (EETE) distribution. Some statistical and reliability properties of the new distribution functions have been characterized based on two non-adjacent generalized and dual generalized order statistics. Moreover, we show that these characterization properties provide a beneficial strategy to predict future events, which are based on past or current events and on an arbitrary distribution function. A characterization in statistics is a specific distributional property of a statistic that uniquely identify related parametric family of distributions. In statistical applications, the researchers usually want to verify whether the data that they are dealing with belong to a certain family of DFs. Therefore, the researchers have to rely on a characterization of the assumed distribution and check if the corresponding conditions are satisfied.

Keywords: Generalized order statistics, Dual generalized order statistics, Dilation Characterization of distributions, Point prediction

1. Introduction

Generalized order statistics introduced by [15]. A concept of generalized order statistics (GOSs) as a unified approach to a variety of models of ascendingly ordered random variables (RVs) introduced by [15]. The concept of dual GOSs, denoted by DGOSs, was introduced by [14] as a parallel concept of GOSs to enable a common approach to descendingly ordered RVs. [14] has shown that (cf. Theorem 3.3) there is a direct link between DGOSs and GOSs.

The subclasses m -GOSs and m -DGOSs of GOSs and DGOSs, respectively, contain many important models of ordered RVs such as ordinary order statistics (OOSs), lower and upper record values, k -records, sequential order statistics (SOSs) and type II censored OOSs. For any $1 \leq r \leq n$, the marginal probability density functions (PDFs) of the r th m -GOS $X(r, n; m, k)$ and m -DGOS $X^*(r, n; m, k)$, based on a continuous distribution function (DF) $F_X(x) = P(X \leq x)$ with a PDF $f_X(x)$, are given, respectively, by (cf. [15] and [28]).

$$f_{X(r,n,m,k)}(x) = \frac{c_{r-1}^{(n)}}{(r-1)!} [\bar{F}_X(x)]^{r^{(n)}-1} \left[\frac{1-[\bar{F}_X(x)]^{m+1}}{m+1} \right]^{r-1} f_X(x) , \quad m \neq -1 \quad (1)$$

and

$$f_{X^*(r,n,m,k)}(x) = \frac{c_{r-1}^{(n)}}{(r-1)!} F_X^{r^{(n)}-1} \left[\frac{1-[F_X(x)]^{m+1}}{m+1} \right]^{r-1} f_X(x) , \quad m \neq -1 \quad (2)$$

where $\bar{F} = 1 - F$, $\gamma_r^{(n)} = k + (n - r)(m + 1)$ and $C_{r-1}^{(n)} = \prod_{i=1}^r \gamma_i^{(n)}$, $1 \leq r \leq n$.

Classical results in characterizations can be found in [2], [3], [4] and [5]. Different results of characterization and its applications in terms of GOSs and DGOSs are derived by many authors. Among these authors are [6], [7], [8], [9], [10], [11], [12] and [13].

In this paper, we prove some new characteristic properties of the Erlang truncated exponential DF $\exp(\beta\alpha_\lambda)$, with mean $\frac{1}{(\beta\alpha_\lambda)}$, $\beta > 0, \alpha_\lambda > 0$. The Erlang truncated exponential distribution is prominent in life testing experiments and reliability problems.

The result of this paper enables us to predict the time at which some survived components will have failed or to predict the mean failure time of unobserved lifetimes in a lifetime experiment by using the result of another independent lifetime experiment. Throughout this paper, " $X \stackrel{d}{=} Y$ " means that the RVs X and Y have the same DFs and " $X \sim F$ " means that the RV X has the DF F .

The rest of this paper is organized as follows. In Section 2, we reveal some characterization properties for the Erlang truncated exponential distribution based on two nonadjacent m -GOSs (consequently m -DGOSs) from two independent Erlang truncated exponential distributions. In Section 3, we use the results of Section 2 in an application of the prediction problem concerning the lifetime experiments. If support of the distribution $F_X(x)$ be over (a, b) , then by convention, we will write

$$X(0, n, m, k) = a \text{ and } X^*(0, n, m, k) = b$$

It may be seen that if Y is a measurable function of X with the relation

$$Y = h(X)$$

Then

$$Y(r, n, m, k) = h[X(r, n, m, k)]$$

if h is an increasing function and

$$Y^*(r, n, m, k) = h[X(r, n, m, k)]$$

if h is a decreasing function

where $X(r, n, m, k)$ is the r^{th} m -GOS and $X^*(r, n, m, k)$ is the r^{th} m - DGOS.

Erlang-Truncated Exponential (ETE) distribution was originally introduced by [1] as an extension of the standard one parameter exponential distribution. The Erlang-Truncated Exponential (ETE) distribution results from the mixture of Erlang distribution and the left truncated one-parameter exponential distribution. The cumulative distribution function CDF $F_X(x)$ and probability density function PDF $f_X(x)$ of the Erlang-Truncated Exponential (ETE) distribution are given by

$$F_X(x) = [1 - e^{-\beta(\alpha_\lambda)x}], \quad 0 \leq x < \infty, \beta, \lambda > 0 \tag{3}$$

where $\alpha_\lambda = (1 - e^{-\lambda})$

and

$$f_X(x) = \beta (\alpha_\lambda) e^{-\beta(\alpha_\lambda)x}, \quad 0 \leq x < \infty, \beta, \lambda > 0 \tag{4}$$

respectively, where β is the shape parameter and λ is the scale parameter. The Erlang-Truncated Exponential (ETE) distribution collapses to the classical one-parameter exponential distribution with parameter β and $\lambda \rightarrow \infty$.

$$X \sim \text{Par}(\beta(\alpha_\lambda))$$

if X has a Pareto distribution with the DF

$$F_X(x) = [1 - x^{-\beta(\alpha\lambda)}], 1 < x < \infty, \beta > 0, \alpha > 0, \lambda > 0 \quad (5)$$

$$X \sim \text{pow}(\beta(\alpha\lambda))$$

if X has a power function distribution with the DF

$$F_X(x) = x^{\beta(\alpha\lambda)}, 0 < x < 1, \beta > 0, \alpha > 0, \lambda > 0 \quad (6)$$

It may further be noted that

if $\log X \sim \text{Erlang-truncated exp}(\beta(\alpha\lambda))$ then $X \sim \text{Par}(\beta(\alpha\lambda))$

if $-\log X \sim \text{Erlang-truncated exp}(\beta(\alpha\lambda))$ then $X \sim \text{pow}(\beta(\alpha\lambda))$

It has been assumed here throughout that the DF is differentiable *w. r. t.* its argument.

The new distribution is called the Extended Erlang-Truncated Exponential (EETE) distribution. The Extended Erlang-Truncated Exponential (EETE) distribution has a tractable PDF whose shape is either decreasing or unimodal. The failure rate function (FRF) is characterized by decreasing, constant and increasing shapes and the new three-parameter distribution demonstrates a superior fit when compared with some other well-known three parameter distributions, as we shall see later. Related works are: the Transmuted Erlang Truncated Exponential distribution, due to [25], Marshall–Olkin generalized Erlang-truncated exponential distribution, due to [26] and the generalized Erlang-Truncated Exponential distribution, due to [27].

2. MODEL

The cumulative distribution function CDF $F_X(x)$ and probability density function PDF $f_X(x)$ of the Extended Erlang-Truncated Exponential (EETE) distribution are given by

$$F_X(x) = [1 - e^{-\beta(\alpha\lambda)x}]^\alpha, 0 \leq x < \infty, \alpha, \beta, \lambda > 0 \quad (7)$$

and

$$f_X(x) = \alpha \beta (\alpha\lambda) e^{-\beta(\alpha\lambda)x} [1 - e^{-\beta(\alpha\lambda)x}]^{\alpha-1}, 0 \leq x < \infty, \alpha, \beta, \lambda > 0 \quad (8)$$

where α and β are the shape parameters and λ is the scale parameter.

The Extended Erlang-Truncated Exponential (EETE) distribution reduces to Erlang-Truncated Exponential (ETE) when $\alpha = 1$.

3. RELIABILITY CHARACTERISTICS

The reliability function $R(x)$ is an important tool for characterizing life phenomenon. $R(x)$ is analytically expressed as $R(x) = 1 - F(x)$. Under certain predefined conditions, the reliability function $R(x)$ gives the probability that a system will operate without failure until a specified time x . The reliability function of the Extended Erlang-Truncated Exponential (EETE) distribution is given by

$$R(x) = 1 - (1 - e^{-\beta(\alpha\lambda)x})^\alpha, 0 \leq x < \infty, \alpha, \beta, \lambda > 0 \quad (9)$$

Another important reliability characteristics is the failure rate function. The failure rate function gives the probability of failure for a system that has survived up to time x . The failure rate function $h(x)$ is mathematically expressed $h(x) = f(x)/R(x)$. The failure rate function the Extended Erlang-Truncated Exponential (EETE) distribution is given by:

$$h(x) = \frac{\alpha \beta (\alpha_\lambda) e^{-\beta(\alpha_\lambda)x} [1 - e^{-\beta(\alpha_\lambda)x}]^{\alpha-1}}{1 - [1 - e^{-\beta(\alpha_\lambda)x}]^\alpha}, \quad 0 \leq x < \infty, \quad \alpha, \beta, \lambda > 0 \quad (10)$$

4. CHARACTERISTION RESULTS

We assume that all considered DFs are differentiable with respect to their arguments. Moreover, all the considered RVs are non-negative

THEOREM 4.1 :-

Let $X(r, n; m, k)$, $m \neq -1$ be the r^{th} m - GOS from a sample of size n drawn from a continuous DF $F_X(x)$ with PDF $f_X(x)$. Furthermore, let $Y(r, n, m, k)$, $m \neq -1$ be the r^{th} m - GOS based on a sample of size n , which is drawn from a continuous DF $F_Y(y) = P(Y \leq y)$, where Y is independent of X . Finally, let the relation

$$X(s, n, m, k) \stackrel{d}{=} X(r, n, m, k) + \tilde{Y} \quad (11)$$

be satisfied for all $1 \leq r < s \leq n$, Then, $\tilde{Y} \stackrel{d}{=} X(s - r, n - r, m, k)$ and $Y \sim \exp(\beta\alpha_\lambda)$ if and if $X \sim \exp(\beta\alpha_\lambda)$, $\beta > 0, \alpha_\lambda > 0$.

Proof. We first prove the necessary part. Let the moment generating function (MGF) of $X(s, n, m, k)$ be $M_{X(s,n,m,k)}(t)$. Then, (11) implies that

$$M_{X(s,n,m,k)}(t) = M_{X(r,n,m,k)}(t) \cdot M_{\tilde{Y}}(t) \quad (12)$$

Let us now derive the MGF of the r^{th} m -GOS $X(s, n, m, k)$ based on Erlang truncated $\exp(\beta\alpha_\lambda)$. Clearly, in view of (1), we get

$$M_{X(s,n,m,k)}(t) = \frac{\beta(\alpha_\lambda) c_{s-1}^{(n)}}{(r-1)!(m+1)^{s-1}} \int_0^\infty e^{-x((\beta\alpha_\lambda)\gamma_s^{(n)} - t)} [1 - e^{-\beta(\alpha_\lambda)(m+1)x}]^{s-1} dx$$

Which by using the transformation $y = e^{-\beta\alpha_\lambda(m+1)x}$ takes the form

$$\begin{aligned} M_{X(s,n,m,k)}(t) &= \frac{C_{s-1}^{(n)}}{(s-1)!(m+1)^s} \int_0^1 y^{\left(\frac{\gamma_s^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)}\right)^{-1}} (1-y)^{s-1} dy \\ &= \frac{C_{s-1}^{(n)}}{(m+1)^s} \frac{\Gamma\left(\frac{\gamma_s^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)}\right)}{\Gamma\left(\frac{\gamma_s^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)} + s\right)} = \prod_{i=1}^s \left(\frac{\frac{\gamma_s^{(n)}}{m+1}}{\frac{\gamma_s^{(n)}}{m+1} - \frac{1}{\beta(\alpha_\lambda)(m+1)} + s - i} \right) \\ &= \prod_{i=1}^s \left(\frac{\frac{\gamma_s^{(n)}}{m+1}}{\frac{\gamma_s^{(n)}}{m+1} - \frac{1}{\beta(\alpha_\lambda)(m+1)} + s - i} \right) = \prod_{i=1}^s \left(1 - \frac{t}{\beta(\alpha_\lambda)\gamma_i^{(n)}} \right)^{-1} \end{aligned} \quad (13)$$

Where $\Gamma(\cdot)$ is the usual gamma function. On the other hand, in view of (12)

$$M_{\tilde{Y}}(t) = \frac{M_{X(s,n,m,k)}(t)}{M_{X(r,n,m,k)}(t)} = \frac{(m+1)^r C_{s-1}^{(n)} \Gamma\left(\frac{\gamma_s^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)}\right) \Gamma\left(\frac{\gamma_r^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)} + r\right)}{(m+1)^s C_{r-1}^{(n)} \Gamma\left(\frac{\gamma_s^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)} + s\right) \Gamma\left(\frac{\gamma_r^{(n)}}{m+1} - \frac{t}{\beta(\alpha_\lambda)(m+1)}\right)}$$

$$= \prod_{i=r+1}^s \left(1 - \frac{t}{\beta(\alpha_\lambda)\gamma_i^{(n)}}\right)^{-1} = \prod_{j=1}^{s-r} \left(1 - \frac{t}{\beta(\alpha_\lambda)\gamma_{r+j}^{(n)}}\right)^{-1} = \prod_{j=1}^{s-r} \left(1 - \frac{t}{\beta(\alpha_\lambda)\gamma_j^{(n-r)}}\right)^{-1} \quad (14)$$

Since, $\gamma_{r+j}^{(n)} = k + (n - r - j)(m + 1) = \gamma_j^{(n-r)}$. On comparing (14) with (13), we deduce that $M_{\tilde{Y}}(t)$ is the MGF of $Y(s - r, n - r, m, k)$, i.e., the $(s - r)^{th}$ m -GOS from a sample of size $(n - r)$ drawn from the DF Erlang truncated $\exp(\beta(\alpha_\lambda))$. This completes the proof of the necessity part.

We now turn to prove the sufficiency part. Let the representation (11) be satisfied with $\tilde{Y} \stackrel{d}{=} Y(s - r, n - r, m, k)$ and $Y \sim \exp(\beta(\alpha_\lambda))$. Furthermore, let $X(s, n, m, k)$ and $X(r, n, m, k)$ in (11) be m -GOSs, which are based on an unknown DF $F_X(x)$ and they are independent of $Y(s, n, m, k)$. Therefore, the convolution relation (11) implies that

$$\begin{aligned} f_{X(s,n,m,k)}(x) &= \int_0^\infty f_{X(r,n,m,k)}(y) f_{Y((s,n,m,k))}(x - y) dy \\ &= \frac{\beta(\alpha_\lambda) c_{(s-r-1)}^{(n-r)}}{(s-r-1)!(m+1)^{s-r-1}} \int_0^\infty [e^{-\beta(\alpha_\lambda)(x-y)}] \gamma_s^{(n)} \\ &\quad \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{s-r-1} f_{X(r,n,m,k)}(y) dy \end{aligned} \quad (15)$$

as $\gamma_{s-r}^{(n-r)} = \gamma_s^{(n)}$

Differentiating both the sides of (15) w. r. t. x , we get

$$C_{s-r-1}^{(n-r)} = \gamma_{s-r}^{(n-r)} C_{s-r-2}^{(n-r)} = \gamma_s^{(n)} C_{s-r-2}^{(n-r)}$$

and, $\gamma_s^{(n)} + (m + 1) = \gamma_{s-1}^{(n)}$, we get

$$\begin{aligned} \frac{d}{dx} f_{X(s,n,m,k)}(x) &= \frac{(\beta(\alpha_\lambda))^2 \gamma_s^{(n)} C_{s-r-2}^{(n-r)}}{(s-r-2)!(m+1)^{s-r-2}} \int_0^x [e^{-\beta(\alpha_\lambda)(x-y)}] \gamma_{s-1}^{(n)+(m+1)} \\ &\quad \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{s-r-2} f_{X(r,n,m,k)}(y) dy \\ &\quad - \frac{(\beta(\alpha_\lambda))^2 \gamma_s^{(n)} C_{s-r-1}^{(n-r)}}{(s-r-1)!(m+1)^{s-r-1}} \int_0^x [e^{-\beta(\alpha_\lambda)(x-y)}] \gamma_s^{(n)} \\ &\quad \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{s-r-1} f_{X(r,n,m,k)}(y) dy \end{aligned} \quad (16)$$

On the other hand, by using the obvious relation

$$\begin{aligned} e^{-\beta(\alpha_\lambda)\gamma_s^{(n)}z} (1 - (e^{-\beta(\alpha_\lambda)(z)})^{m+1})^{s-r-1} &= e^{-\beta(\alpha_\lambda)\gamma_s^{(n)}z} (1 - e^{-\beta(\alpha_\lambda)(m+1)z})^{s-r-2} \\ &\quad - e^{-\beta(\alpha_\lambda)\gamma_s^{(n)+(m+1)z}} (1 - e^{-\beta(\alpha_\lambda)(m+1)z})^{s-r-2} \end{aligned}$$

and by using the representation (3.5), we get

$$\begin{aligned} &\frac{\beta(\alpha_\lambda) c_{s-r-1}^{(n-r)}}{(s-r-2)!(m+1)^{s-r-2}} \int_0^x [e^{-\beta(\alpha_\lambda)(x-y)}] \gamma_{s-r}^{(n)+(m+1)} \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{s-r-2} f_{X(r,n,m,k)}(y) dy \\ &= \frac{c_{s-r-1}^{(n-r)}}{c_{s-r-2}^{(n-1)}} f_{X(s-r-1,n-1,m,k)}(x) - (m + 1)(s - r - 1) f_{X(s-r,n,m,k)}(x) \end{aligned} \quad (17)$$

Thus, by using the relation $\gamma_{s-r}^{(n)} + (m + 1)(s - r - 1) = \gamma_1^{(n)}$ and

$$\frac{C_{s-r-1}^{(n-r)}}{C_{s-r-2}^{(n-1)}} = \frac{\prod_{i=1}^{s-r} \gamma_i^{(n)}}{\prod_{i=2}^{s-r} \gamma_i^{(n)}} = \gamma_s^{(n)}$$

and by combing (23) and (16), we get

$$\frac{d}{dx} f_{X(s,n,m,k)}(x) = \beta(\alpha_\lambda) \gamma_s^{(n)} [f_{X(s-1,n-1,m,k)}(x) - f_{X(s,n,m,k)}(x)]$$

or equivalently, by integrating from 0 to x

$$f_{X(s,n,m,k)}(x) = \beta(\alpha_\lambda) \gamma_s^{(n)} [F_{X(s-1,n-1,m,k)}(x) - F_{X(s,n,m,k)}(x)] \quad (18)$$

Now, by using the relation (II) of [15] on page 75, we get

$$F_{X(s-1,n-1,m,k)}(x) - F_{X(s,n,m,k)}(x) = \frac{C_{s-2}^{(n)}}{(s-1)!(m+1)^{s-1}} [\bar{F}_X(x)]^{\gamma_s^{(n)}-1} [1 - (\bar{F}_X(x))^{m+1}]^{s-1} \quad (19)$$

Therefore, by combing (1), (18) and (19), we have

$$\frac{f_X(x)}{\bar{F}_X(x)} = \beta(\alpha_\lambda)$$

Which implies that $F_X(x) = [1 - e^{-\beta(\alpha_\lambda)x}]$, $x > 0, \beta > 0, \alpha > 0, \lambda > 0$. This completes the proof of the sufficient part, as well as the proof of theorem 3.1.

Corollary 4.1. Assume that the RVs X and Y are independent, as we assumed in Theorem 4.1. By replacing the additive relation (11) by the multiplication relation

$$X(s, n, m, k) \stackrel{d}{=} X(r, n, m, k) \cdot \tilde{Y} \quad (20)$$

Then, $\tilde{Y} \stackrel{d}{=} Y(s - r, n - r, m, k)$ and $Y \sim \text{Pareto}(\beta(\alpha_\lambda))$ (i.e., $F(x) = [1 - y^{-\beta(\alpha_\lambda)}, y > 1]$) if and if $X \sim \text{Pareto}(\beta(\alpha_\lambda))$, $\beta > 0, \alpha > 0, \lambda > 0$.

Proof. Here the proof immediately follows, by noting that if $X \sim \text{Pareto}(\beta(\alpha_\lambda))$, then $\log X \sim \text{exp}(\beta(\alpha_\lambda))$ and

$$\log X(s, n, m, k) \stackrel{d}{=} \log X(r, n, m, k) + \log \tilde{Y}$$

which implies

$$X(s, n, m, k) \stackrel{d}{=} X(r, n, m, k) \cdot \tilde{Y}$$

Remark 1. In (20), the product $X(r, n, m, k) \times \tilde{Y}$ is called random dilation of $X(r, n, m, k)$, cf. [7]. Moreover, at $s = r + 1$, the representation (20) gives

$$X(r + 1, n, m, k) \stackrel{d}{=} X(r, n, m, k) \cdot Y(1, n - s + 1, m, k) \quad (21)$$

as was obtained by [7] for $X \sim \text{Pareto}(\beta(\alpha_\lambda))$. Also, at $r = s - 1$, the relation (20) gives

$$X(r + 1, n, 0, 1) \stackrel{d}{=} X(r, n - 1, 0, 1) \cdot Y(1, n, 0, 1)$$

(i.e., for the OOSs model), which was obtained by [8], for $X(1, n, 0, 1) \sim \text{Pareto}(\beta(\alpha_\lambda)n)$. Finally, the representation (21) can be written as (for OOSs model) $X(s, n, 0, 1) \stackrel{d}{=} X(r, n, 0, 1) \cdot V$, $1 \leq r < s$, which was an unsolved problem due to [6]).

Corollary 4.2. Assume that the RVs X and Y are independent, Let $X^*(r, n, m, k)$ and $Y^*(r, n, m, k)$ be the r^{th} m- DGOS based on a sample of size n drawn from F_X and F_Y , respectively.

By replacing the additive relation (11) by the multiplicative relation

$$X^*(s, n, m, k) \stackrel{d}{=} X^*(r, n, m, k) \cdot Y^*$$

Then, $Y^* \stackrel{d}{=} Y^*(s - r, n - r, m, k)$ and $Y^* \sim \text{Power}(\beta(\alpha_\lambda))$, $\beta > 0, \alpha > 0$ (i.e., $F_Y(y) = y^{\beta(\alpha_\lambda)}$) if and if $X^* \sim \text{Power}(\beta(\alpha_\lambda))$.

Proof. The proof immediately follows from the simple relation between the GOSs and DGOSs, by noting that if, by noting that if $X \sim \text{Power}(\beta(\alpha_\lambda))$, then $-\log X \sim \text{exp}(\beta(\alpha_\lambda))$ and

$$-\log X^*(s, n, m, k) \stackrel{d}{=} -\log X^*(r, n, m, k) - \log Y^*$$

which implies

$$X^*(s, n, m, k) \stackrel{d}{=} X^*(r, n, m, k) \cdot Y^*$$

THEOREM 4.2 :-

Let $X(r, n, m, k)$, $m \neq -1$ be the r^{th} m - GOS from a sample of size n drawn from a continuous DF $F_X(x)$ with PDF $f_X(x)$. Furthermore, let $Y(r, n, m, k)$, $m \neq -1$ be the r^{th} m - GOS based on a sample of size n , which is drawn from a continuous DF $F_Y(y)$, where Y is independent of X . Finally, let the relation

$$X(s, n, m, k) \stackrel{d}{=} X(s - r, n - r, m, k) + \tilde{Y} \tag{22}$$

be satisfied for all $1 \leq r < s$, Then, $\tilde{Y} \stackrel{d}{=} X(r, n, m, k)$ and $Y \sim \exp(\beta\alpha_\lambda)$ if and if $X \sim \exp(\beta\alpha_\lambda)$, $\beta > 0, \alpha > 0, \lambda > 0$.

Proof. Clearly, the proof of the necessity part follows from Theorem 4.1, while the proof of the sufficiency part follows closely as the sufficiency part of Theorem 4.1. Namely, let the representation (22) be satisfied with $\tilde{Y} \stackrel{d}{=} X(r, n, m, k)$ and $Y \sim \exp(\beta\alpha_\lambda)$. Furthermore, let $X(s, n, m, k)$ and $X(s - r, n - r, m, k)$ in (22) be m -GOSs, which are based on an unknown DF $F_X(x)$ and they are independent of $Y(r, n, m, k)$. Therefore, the convolution relation (22) implies that

$$\begin{aligned} f_{X(s,n,m,k)}(x) &= \int_0^x f_{X(s-r,n-r,m,k)}(y) f_{Y(r,n,m,k)}(x-y) dy \\ &= \frac{\beta(\alpha_\lambda) C_{r-1}^{(n)}}{(r-1)!(m+1)^{r-1}} \int_0^x e^{-\beta(\alpha_\lambda) \gamma_r^{(n)}(x-y)} \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{r-1} f_{X(s-r,n-r,m,k)}(y) dy \end{aligned} \tag{23}$$

By differentiating both the sides of (23) with respect to x , we get

$$\begin{aligned} \frac{df_{X(s,n,m,k)}(x)}{dx} &= \frac{(\beta(\alpha_\lambda))^2 C_{r-1}^{(n)}}{(r-2)!(m+1)^{r-2}} \int_0^x [e^{-\beta(\alpha_\lambda)(\gamma_r^{(n2)} + (m+1))(x-y)} \\ &\quad \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{r-2} f_{X(s-r,n-r,m,k)}(y) dy \\ &\quad - \frac{(\beta(\alpha_\lambda))^2 \gamma_r^{(n)} C_{r-1}^{(n)}}{(r-1)!(m+1)^{r-1}} \int_0^x e^{-\beta(\alpha_\lambda)(\gamma_r^{(n)}(x-y)} \\ &\quad \times [1 - (e^{-\beta(\alpha_\lambda)(x-y)})^{m+1}]^{r-1} f_{X(s-r,n-r,m,k)}(y) dy \\ &= \beta(\alpha_\lambda) \gamma_1^{(n)} [f_{X(s-1,n,m,k)}(x) - f_{X(s,n,m,k)}(x)] \end{aligned}$$

Or equivalently, by integrating from 0 to x ,

$$f_{X(s,n,m,k)}(x) = \beta(\alpha_\lambda) \gamma_1^{(n)} [F_{X(s-1,n-1,m,k)}(x) - F_{X(s,n,m,k)}(x)] \tag{24}$$

Now, by using the relation of [15] on page 75, we get

$$F_{X(s-1,n-1,m,k)}(x) - F_{X(s,n,m,k)}(x) = \frac{C_{s-2}^{(n-1)}}{(s-1)!(m+1)^{s-1}} [\bar{F}_X(x)]^{\gamma_s^{(n)}-1} [1 - (\bar{F}_X(x))^{m+1}]^{s-1} \tag{25}$$

Therefore, by combing (1), (24) and (25), we get

$$\frac{f_X(x)}{\bar{F}_X(x)} = \beta(\alpha_\lambda)$$

which implies that

$$F_X(x) = [1 - e^{-\beta(\alpha_\lambda)y}], \beta > 0, \alpha > 0, \lambda > 0, x > 0$$

This complete the proof of the sufficiency part, as well as the proof of Theorem 4.2.

Corollary 4.3 Assume that the RVs X and Y are independent, as we assumed in Theorem 4.2. By replacing the additive relation (22) by the multiplicative relation

$$X(s, n, m, k) \stackrel{d}{=} X(s - r, n - r, m, k) \cdot \tilde{Y} \tag{26}$$

Then, $\tilde{Y} \stackrel{d}{=} Y(r, n, m, k)$ and $Y \sim \text{Pareto}(\beta(\alpha_\lambda))$ if and only if $X \sim \text{Pareto}(\beta(\alpha_\lambda))$

Proof. The proof follows exactly as the proof of Corollary 4.1.

Remark 2. For OOSs model the relation (26) takes the form

$$X(s, n; 0, 1) \stackrel{d}{=} X(s - r, n - r; 0, 1) \cdot Y(r, n; 0, 1)$$

Which implies the relation $X(s, n; 0, 1) \stackrel{d}{=} X(r, n; 0, 1) \cdot Y(s - r, n - r; 0, 1)$ that is belonging to [8]).

Corollary 4.4. Assume that the RVs X and Y are independent, Let $X^*(r, n, m, k)$ and $Y^*(r, n, m, k)$ be the r^{th} m - DGOs based on a sample of size n drawn from F_X and F_Y , respectively. By replacing the additive relation (22) by the multiplicative relation

$$X^*(s, n, m, k) \stackrel{d}{=} X^*(s - r, n - r, m, k) \cdot Y^*$$

Then, $Y^* \stackrel{d}{=} Y^*(r, n, m, k)$ and $Y^* \sim \text{Power}(\beta(\alpha_\lambda))$ if and if $X^* \sim \text{Power}(\beta(\alpha_\lambda))$, $\beta > 0, \alpha > 0, \lambda > 0$.

Proof. The proof follows as the proof of Corollary 4.2.

5. APPLICATIONS TO THE PREDICTION PROBLEM

Prediction problem usually arises in life-testing experiments of medical and industrial applications. Often, in the life-testing experiments, the observations arrive in ascending order of magnitude. Consequently, in reliability theory, especially for OOSs and SOSs, $X(r, n, m, k)$ represents the life length of a $n - r + 1$ - out-of- n system made up of n independent life lengths (these components are identical for OOSs and non identical for SOSs). Motivation for the prediction problems arises when the experiment is terminated before its conclusion by stopping after a given time (Type I censoring) or after a given number of failures (Type II censoring). Several authors have considered prediction problems involving GOSs, see for example [30], [31], [32], [33] and [34].

Theorems 4.1 and 4.2 suggest a new method for treating two prediction problems of different types. Namely, Theorem 4.2 treats a classical prediction problem, that predicting $X(s, n, m, k)$, $1 \leq r < s \leq n$, based on the observed m -GOSs $X(1, n, m, k) \leq X(2, n, m, k) \leq \dots \leq X(r, n, m, k)$. On the other hand, Theorem 3.1 considers the prediction problem of $X(r, n, m, k)$, when the sample size of the test is enlarged from n to N , by adding some extra items X_{n+1}, \dots, X_N after observing $X(r, n, m, k)$. Clearly, the sequence $\{X(r, n, m, k)\}$ is non-increasing in n . For example, if $F_X(x)$ is continuous and for any fixed value $r < n$, the observed value of $X(r, n, 0, 1)$, denoted by $x(r, n, 0, 1)$, did not change if $\min(x_{n+1}, \dots, x_N) > x(r, n, 0, 1)$, otherwise we get $x(r, n, 0, 1) < x(r, N, 0, 1)$. In the preceding two prediction problems, the failure times of the unobserved lifetimes in a lifetime experiment are predicted by using the result of another independent lifetime experiment.

References

- [1] Alosey-EL, R. A. (2007). Random sum of new type of mixture of distribution. Int. J. Stat. Syst. 2: 49-57.
- [2] El-Adll, M. E. (2018), 'Characterization of Distributions by Equalities of Two Generalized or Dual Generalized Order Statistics', Communications in Statistics - Theory and Methods 26(3), 522-528.
- [3] Galambos, J. & Kotz, S. (1978), Characterizations of Probability Distributions, Cambridge University Press, Berlin Heidelberg, New York.
- [4] Nagaraja, H. N. (2006), Characterizations of Probability Distributions -Handbook of Engineering Statistics, 79-87(In Book Chapter), Springer, Editor: Hoang Pham, Uk.
- [5] Rao, C. R. & Chanabhng, D. N. (1998), Recent Approaches to Characterizations Based on Order Statistics and Record Values (Handbook of Statistics, 16, 231- 256), Elsevier.

- [6] Arnold, B. C., Castillo, E. & Sarabia, J. (2008), 'Some Characterizations Involving Uniform and Powers of Uniform Random Variables', *Statistics* 42(6), 527–534.
- [7] Beutner, E. and Kamps, U. (2008). Random contraction and random dilation of generalized order statistics. *Comm. Statist. Theory Methods* 37: 2185-2201.
- [8] Castaño-Martínez, A., López-Bilazquez, F. and Salamanca-Miño, B. (2012). Random translations, contractions and dilations of order statistics and records. *Statistics* 46(1): 57-67.
- [9] Khan, A. H., Shah Imtiyaz, A. and Ahsanullah, M. (2012). Characterization through distributional properties of order statistics. *J. Egyptian Math. Soc.* 20 :211-214.
- [10] Öncel, S. Y., Ahsanullah, M., Aliev, F. A. & Aygun, F. (2005), 'Switching Record and Order Statistics via Random Contraction', *Statistical Probability Letters* 73(3), 207–217.
- [11] Samuel, P. (2008), 'Characterization of Distributions by Conditional Expectation of Generalized Order Statistics', *Statistical Papers* 49, 101–108.
- [12] Tavangar, M. & Hashemi, M. (2013), 'On Characterizations of the Generalized Pareto Distributions Based on Progressively Censored Order Statistics', *Statistical Papers* 54, 381–390.
- [13] Wesolowski, J. & Ahsanullah, M. (2004), 'Switching Order Statistics Through Random Power Contractions', *Australian & New Zealand Journal of Statistics* 46(2), 297–303
- [14] Burkschat, M., Cramer, E. & Kamps, U. (2003), 'Dual Generalized Order Statistics', *Metron* LXI(I), 13–26.
- [15] Kamps, U. (1995), *A Concept of Generalized Order Statistics*, B. G. Teubner, Stuttgart.
- [16] Ahsanullah, M. (2000). Generalized order statistics from exponential distribution. *J. Statist. Plan. Inf.* 85: 85-91.
- [17] Ahsanullah, M. (2006). The generalized order statistics from exponential distribution. *J. Statist. Research* 40(2): 21-27.
- [18] Aitchison, J. and Dunsmore, I. (1975). *Statistical Prediction Analysis*. Cambridge University Press, Cambridge.
- [19] AL-Hussaini E. K. and Al-Awadi, F. (2010). Bayes two-sample prediction of generalized order statistics with fixed and random sample size. *J. Statist. Comput. Sim.* 80(1): 13-28.
- [20] Arnold, B. C., Castillo, E. and Sarabia, J. M. (2008). Some characterizations involving uniform and powers of uniform random variables. *Statistics* 42(6): 527-534.
- [21] Fagioli, E., Pellerey, F. and Shaked, M. (1999). A characterization of the dilation order and its applications. *Statist. Papers* 40: 393-406.
- [22] Kaminsky, K. S. and Nelson, P. I. (1998). Prediction Intervals. In: *Handbook of Statistics*,
- [23] Balakrishnan, N. and C. R. Rao (Eds.), Amsterdam, North Holland, 431-450.
- [24] Lawless, J. F. (1971). A prediction problem concerning samples from the exponential distribution with applications in life testing. *Technometrics* 13: 725-730.
- [25] Okorie, I. E., Akpanta, A.C and Ohakwe J. , Transmuted Erlang-truncated exponential distribution, *Econ. Qual. Control* 31 (2) (2016) 71–84.
- [26] Okorie, I.E, Akpanta,A.C. and Ohakwe, J. Marshall–Olkin generalized Erlang truncated exponential distribution: properties and applications, *Cogent Math.* 4 (1) (2017) 1285093.
- [27] Nasiru, S., Luguterah, A. and Iddrisu, M.M. Generalized Erlang-truncated exponential distribution, *Adv. Appl. Stat.* 48 (4) (2016) 273–301.
- [28] Ahsanullah, M. (2004), 'A characterization of the Uniform Distribution by Dual Generalized Order Statistics', *Communications in Statistics - Theory and Methods* 33(12), 2921–2928.
- [29] Cramer, E. (2003). *Contributions to Generalized Order Statistics*. Habilitationsschrift, Reprint, University of Oldenburg.
- [30] Aitchison, J. & Dunsmore, I. (1975), *Statistical Prediction Analysis*, Cambridge University Press, Cambridge, U. K.
- [31] Lawless, J. F. (1971), 'A Prediction Problem Concerning Samples from the Exponential Distribution with Applications in Life Testing', *Technometrics* 13(4), 725–730.
- [32] Nagaraja, H. N. (1986), 'Comparison of Estimators and Predictors from Two Parameter Exponential Distribution', *Sankhya, Series B* 48(1), 10–18.

- [33] Raqab, M. Z. & Barakat, H. M. (2018), 'Prediction Intervals for Future Observations Based on Samples of Random Sizes', *Journal of Mathematics and Statistics* 14(1), 16–28.
- [34] Raqab, Z. M. (2001), 'Optimal Prediction-Intervals for the Exponential Distribution Based on Generalized Order Statistics', *IEEE Transactions on Reliability* 50(1), 112–115.

INTERVAL PATTERN RECOGNITION IN RELATION TO INFORMATION ABOUT THE TREATMENT OF COVID 19 IN PATIENTS WITH BLOOD DISEASES

¹ G. TSITSIASHVILI, ² V. NEVZOROVA, ² A. TALKO, ^{1,3} M. OSIPOVA



¹ Russia, 690041, Vladivostok, Radio street 7, Institute for Applied Mathematics
Far Eastern Branch of Russian Academy Sciences
² Russia, 690002, Vladivostok, 2 Ostryakova Ave.,
Pacific State Medical University,
³ Russia, 690002, Vladivostok, FEFU Campus 10 Ajax Bay,
Russky Island, Far Eastern Federal University,
guram@iam.dvo.ru, nevzorova@inbox.ru, talkang92@mail.ru, mao1975@list.ru

Abstract

Since the problem of coronavirus infection, especially in vulnerable groups, remains relevant in the healthcare system, it is necessary to recognize the risk of fatal events during the treatment of patients from COVID-19. Patients with hemoblastosis are exposed to a more severe course of coronavirus infection than the general population, as well as an increased risk of fatal events. The aim of our study was to determine the effect of signs on the risk outcomes of fatal events among patients with diseases of the blood system, both of a tumor and non-tumor nature. According to the values of the signs, it was necessary to recognize survivors and those who died during treatment from COVID-19. The method of interval pattern recognition was chosen due to the presence of big data, it is described in detail in the article. The patients were broken down by gender and age. The signs that significantly affect the forecast according to the developed algorithm were identified. This is especially evident in groups of men and women with malignant diseases of the blood system over the age of 60 years. In these groups, a positive outcome of treatment is detected due to the presence of a large set of uninformative signs. This phenomenon is closely related to the tasks of technical gerontology.

Keywords: interval pattern recognition, big data, informative signs, technical gerontology, coronavirus infection, hemoblastosis, tumor.

INTRODUCTION

According to world studies, it is known that patients with hemoblastosis are susceptible to a more severe course of coronavirus infection than the general population, as well as an increased risk of fatal complications [1, 2, 3]. In the era of the ongoing coronavirus pandemic, it is an urgent task for the population and the healthcare system to recognize the risk of fatal events during the treatment of COVID-19 patients with benign and malignant diseases of the blood system based on the results of their examination.

The initial sample contains data from medical records of the medical history of 221 patients with diseases of the blood system of Primorsky Krai who were treated for COVID-19 coronavirus infection in the infectious department of the KKB2 in Vladivostok, among which 48 patients had non-tumor (benign) diseases of the blood system, 173 - tumor (malignant) diseases of the blood system. Patients with blood tumor diseases were divided depending on the age category into groups: women at least 60 years of age - 60 patients, under 60 years of age - 31 patients, men at

least 60 years of age - 53 patients, under 60 years of age - 29 patients. The use of medical data was made after the signing of informed voluntary consent from the patient.

As a result of this partitioning, the original rather heterogeneous sample was divided into five relatively homogeneous subsamples. Each of these samples was characterized by about 40 different characteristics. Some of these features have the character of Boolean variables of the yes - no type, and some are the results of laboratory studies. Despite the relatively small number of patients in each of the selected groups, the number of signs is quite large. This property allows the source information to be characterized as Big Data. The task was to investigate the effect of signs on the result of treatment in the selected groups. According to the values of the signs, it was necessary to recognize survivors and those who died during treatment from coronavirus infection. The presence of Big Data in the problem required solving by an algorithm with the lowest possible computational complexity. The method of interval pattern recognition was chosen as such a method. This method [4] was developed to predict outbreaks of tick-borne encephalitis by meteorological characteristics of the winter period and later extended to the analysis of extreme meteorological phenomena during last twenty years. The issues of medical information processing considered in the paper are closely related to the tasks of technical gerontology [5] – [7].

I. INTERVAL PATTERN RECOGNITION AND ITS PROPERTIES

Let there be two sets of m -dimensional vectors characterizing some objects:

$$X^0 = \{(x_{11}^0, \dots, x_{1m}^0), \dots, (x_{n_01}^0, \dots, x_{n_0m}^0)\},$$

$$X^1 = \{(x_{11}^1, \dots, x_{1m}^1), \dots, (x_{n_11}^1, \dots, x_{n_1m}^1)\}.$$

We assume that the components of these vectors with numbers $k = 1, \dots, m' < m$ are Boolean variables, and the components with numbers $k = m' + 1, \dots, m$ are real variables. Each vector from the sets X^0 , X^1 describes Boolean/real features of a separate object. Our task is to recognize whether objects belong to the set of X^0 or to the set of X^1 by the vector of signs characterizing these objects.

The essence of the interval pattern recognition method is as follows. An object defined by the vector $(x_{k1}^1, \dots, x_{km}^1)$ is recognized as an element of the set X^0 , if all inequalities are met

$$x_i^- = \min_{1 \leq k \leq n_0} x_{ki}^0 \leq x_{ki}^1 \leq \max_{1 \leq k \leq n_0} x_{ki}^0 = x_i^+, \quad i = 1, \dots, m. \quad (1)$$

Otherwise, the object characterized by the vector $(x_{k1}^1, \dots, x_{km}^1)$ is recognized as an element of the set X^1 , denote their number s (the number of correctly recognized objects of the set X^1). The quality of this method is determined by the ratio

$$\rho = \frac{n_0 + s}{n_0 + n_1}, \quad (2)$$

since the number of correctly recognized objects of the set X^0 is n_0 .

Let's list some properties of the value ρ .

1. All objects of the set X^0 are correctly recognized.
2. The recognition quality of ρ increases with the number of m features.
3. The number of arithmetic operations for implementing interval pattern recognition depends linearly on m and on n .
4. The Boolean attribute i , such that not all numbers $x_{k,i}^0$ match, does not affect the quality of recognition ρ in any way.

Remark 1. In multidimensional statistics, the recognition quality increases with the growth of n and with the decrease of m . In turn, in many real observations, on the contrary: n is small (about 30), and m is quite large (at least 5). This circumstance in conjunction with the properties {2, 3} makes it convenient to use the interval pattern recognition method when analysing Big Data.

But to use this method to medical information, some adaptation is required. Particular attention should be paid to the signs determined by Boolean variables, since doctors consider these signs to be more reliable. Therefore, first a set of I_1 Boolean features is allocated (conditionally denote them $1, \dots, m'$), according to which all objects of the set X^0 take the same values. Next is s_1 of objects, (conditionally denote them $1, \dots, n_1^1$) of the set X^1 , which are correctly recognized by the pain signs of the set I_1 . Let X_1^1 be the collection of all such objects in the set X^1 .

The next step is to build segments of the type (1) based on laboratory signs and check objects $X^1 \setminus X_1^1$ for the correctness of their recognition by signs $m' + 1, \dots, m$. Denote s_2 the number of objects of the set $X^1 \setminus X_1^1$ correctly recognized by these signs, then the equality $s = s_1 + s_2$ is fulfilled. For each feature $i, i = 1, \dots, m$, the number of elements of the set X^1 correctly recognized by it, divided by the number of elements of this set n_1 , is compared as its informativeness.

II. RESEARCH RESULTS

For calculations from the initial medical information, the following signs were identified:

1. Phase of hematological disease at the time of diagnosis of COVID-19: 1 - remission induction, 2 - first remission, 3 - second and subsequent remission, 4 -relapse therapy, 5 - refractory course, 6 - no data;
2. Number of comorbidities (such as obesity, diabetes mellitus, cardiovascular disease, respiratory disease, liver disease, chronic kidney disease);
3. ECOG stage (from 0 to 4) at the time of admission (ECOG is a scale for assessing the general condition of a cancer patient);
4. ECOG maximum stage (that is, how the general condition of the patient has changed from the moment of infection with coronavirus to the moment of recovery/death);
5. Indicators of a clinical blood test: 5.1.hemoglobin (g/l), 5.2.leukocytes (g/l), 5.3. erythrocytes (g/l), 5.4. platelets (g/l); 5.5. ESR (erythrocyte sedimentation rate, mm/h), 5.6. eosinophils (%), 5.7. P / I (stab neutrophils, %), 5.8. c / i (segmented neutrophils, %), 5.9. lymphocytes (%), 5.10. monocytes (%);
6. Parameters of blood biochemistry: 6.1. ALT (IU/l), 6.2. Ast (IU/l), 6.3. GFR (glomerular filtration rate), 6.4. creatinine (mmol/l), 6.5. urea (mmol/l), 6.6.total protein (g/l), 6.7. total bilirubin (mmol/l);
7. Indicators of the blood coagulation system: 7.1. APTT (activated partial thromboplastin time, sec), 7.2. PT (prothrombin time, sec), 7.3. fibrinogen (g/l).

The rest of the blood signs are not informative (few of the deceased have values in the table). Some of the Boolean features allocated to doctors do not affect the quality of recognition due to property 5.

Table 1. Quality of interval recognition

	n_0	n_1	$n_0 + n_1$	s	ρ
Benign blood diseases (group 1)	10	38	48	38	1
Malignant blood diseases, women at least 60 years old (group 2)	24	36	60	22	0,77
Malignant blood diseases, women under 60 (group 3)	15	16	31	11	0,84
Malignant blood diseases, men at least 60 years old (group 4)	28	25	53	15	0,81
Malignant blood diseases, men under 60 (group 5)	12	16	28	15	0,96

The interval pattern recognition algorithm described above (in the presence of Boolean signs) calculated the indicator ρ , characterizing the quality of interval recognition for all selected groups of patients (see Table 1). In our notation n_0 is the number of patients who died from coronavirus

infection, n_1 is the number of patients who recovered from coronavirus infection, s is the number of recovered patients recognized as recovered (correctly recognized).

Table 2. Informative value of recognizing signs

feature number	Group 1	Group 2	Group 3	Group 4	Group 5
1	32/38	0	1/16	0	0
2	11/38	4/36	1/16	0	0
3	0	0	0	0	6/16
4	0	0	0	3/25	7/16
5.1	10/38	1/36	1/16	1/25	1/16
5.2.	33/38	2/36	2/16	1/25	1/16
5.3.	13/38	1/36	0	1/25	10/16
5.4.	18/38	0	0	7/25	3/16
5.5.	4/38	1/36	2/16	0	1/16
5.6.	15/38	4/36	1/16	2/25	1/16
5.7.	2/38	0	0	0	0
5.8.	35/38	0	0	0	3/16
5.9.	6/38	2/36	2/16	1/25	2/16
5.10.	1/38	0	6/16	0	0
6.1	7/38	4/36	2/16	0	1/16
6.2.	3/38	1/36	0	0	0
6.3.	2/38	1/36	1/16	1/25	3/16
6.4.	7/38	1/36	0	1/25	3/16
6.5.	28/38	2/36	0	0	2/16
6.6.	9/38	3/36	0	0	0
6.7.	6/38	1/36	0	0	2/16
7.1.	10/38	3/36	7/16	1/25	4/16
7.2.	4/38	0	0	0	0
7.3.	9/38	1/36	1/16	1/25	1/16

From Table 2, it follows that in group 1 of patients with benign blood diseases, three signs are distinguished with an informativeness greater than 0.70, the maximum of them is 0.92. In the group 2 female patients at least 60 years old with malignant blood diseases, the maximum informative value of signs is 0.11, but there are many signs with small values. In the group 3 female patients under 60 years of age with malignant blood diseases, the maximum informative value of signs is 0.38, however, there are many signs with zero informative value. In a group 4 male patients at least 60 years old with malignant blood diseases, the maximum informative value of signs is 0.28, but there are many signs with zero informative value. In a group 5 male patients under 60 years of age with malignant blood diseases, the maximum informative value of signs is 0.62. From a comparison of signs with maximum effectiveness, it can be seen that they are different in different groups (see Figure 1). In conclusion, it should be noted that due to its speed, the interval pattern recognition method proved to be convenient when processing big data in the treatment of coronavirus infection in various age groups of patients with diseases of the blood system. This is especially evident in groups of men and women with malignant diseases of the blood system over the age of 60 years. In these groups, a positive outcome of treatment is detected due to the presence of a large set of uninformative signs. The research was carried out within the state assignment for IAM FEB RAS (N 075-01290-23-00).

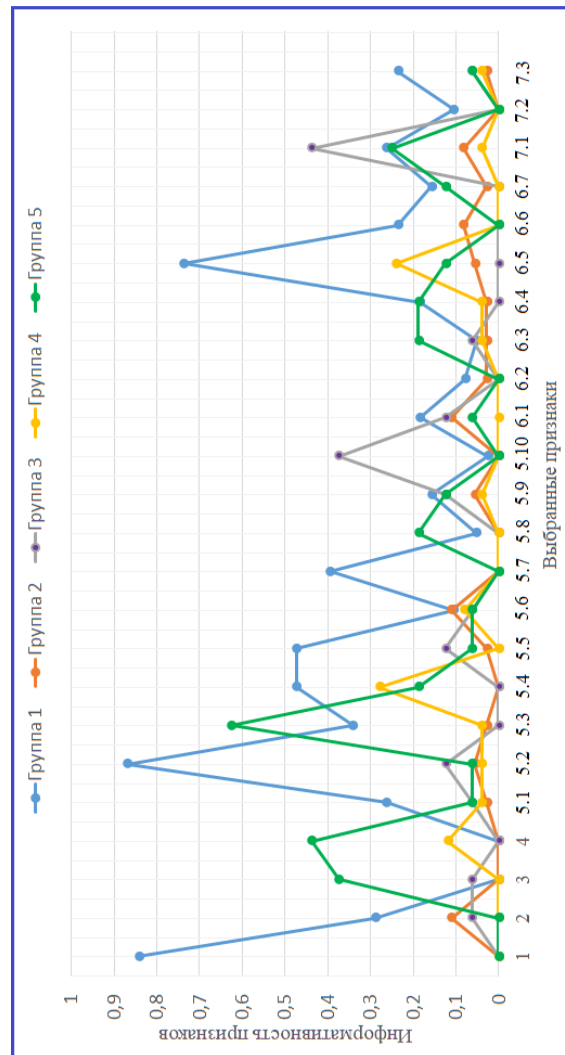


Figure 1: Graphic representation of informative features in groups of patients

REFERENCES

- [1] Pinana, J. L. et al. (2020) Risk factors and outcome of COVID-19 in patients with hematological malignancies. *Experimental hematology and oncology*, 9: 21–25. DOI: 10.1186/s40164-020-00177-z
- [2] Gossell-Williams, M. (2020)
Faculty Opinions recommendation of Case Fatality Rate of Cancer Patients with COVID-19 in a New York Hospital System. Faculty Opinions. Post-Publication Peer Review of the Biomedical Literature. DOI:10.3410/f.737865566.79357670
- [3] Kaprin, A. D. et al. (2020) The impact of the COVID-19 pandemic on cancer practice Siberian *Journal of Oncology*, 19(3): 5–22 (In Russian). DOI: 10.21294/1814-4861-2020-19-3-5-22.
- [4] Tsitsiashvili, G. Sh.: editor. Efficient algorithms of time series processing and their applications. Nova Publishers Science. N.Y., 2009.
- [5] Abramov, O. V. (2013) Algorithm for estimating and predicting the remaining resource of complex technical systems *Reliability and Quality of Complex Systems*, I, 5-?6. (In Russian).

- [6] Kanade, T.: Guest Editor. (2012) Quality of life technology Proceedings of the IEEE, 100, 2394?-2396. DOI: 10.1109/JPROC.2012.2200555
- [7] Schulz, R.et al (2015) Advancing the Aging and Technology Agenda in Gerontology Gerontologist, 55 (5), 724-?734. DOI: 10.1093/geront/gnu071

STUDY OF OPERATING MODES OF SMALL HPSS' HYDROELECTRIC UNITS EQUIPPED WITH SYNCHRONOUS GENERATOR

L.H. Hasanova¹, H.S. Aliyev¹, M.M. Musayev²

•

¹Azerbaijan Technical University, Baku, Azerbaijan
AZ1073, 25 Hussein Javid pros.
e-mail: gasanovalg@mail.ru, hikmetaliyev@aztu.edu.az

²Azerbaijan Scientific-Research and Designed-Prospecting
Institute of Energetics Baku, Azerbaijan
AZ1012, Aven. H.Zardabi-94
e-mail: musahil@gmail.com

Abstract

A universal analytical expression is obtained for the power of hydraulic turbines of small hydropower plants such as Francis, Pelton and Kaplan with fixed blades as a function of water flow. This expression allows us to predict the generation of active power of each of the turbines, depending on the adjusted water flow rate, which varies in the seasonal, monthly and daily periods of the year. It also allows you to solve the inverse problem, when it is necessary, depending on the load schedule dictated by the electrical system to which these small hydroelectric power stations are connected, to regulate the amount of water flow by changing the opening angles of the guiding devices of the hydraulic turbines. Studies on the proposed mathematical model containing a synchronous generator with electromagnetic excitation coupled to a Francis hydraulic turbine with a 60% change in water flow from $q=1$ (r.u) to $q=0.4$ (r.u). The value of the required active power of a small hydroelectric power station, i.e. power output of the generator will jam (either by the dispatcher, or by the value of the load schedule), i.e. it is an input value, and as the output value of the controller is the opening angle of the guiding apparatus of the turbine.

Keywords: small hydropower plants, Francis turbine, Pelton turbine, axial turbines, synchronous generator with electromagnetic excitation, electrical grid.

1. Introduction

In recent years small HPSs (hydroelectric power station) are successfully adopted as the electric power producers at regional level in many world countries. One of the main advantages of this power type is its environmental purity, and a disadvantage is the demand for significant investments in their construction.

In small HPSs, whose power varies from several kW to 25–30 MW, in most cases as the hydraulic turbines the Francis and Pelton turbines are used, and a series of hydraulic units with axial flow turbines.

The purpose of this paper is the modeling of behavior of stated turbines in general analytic form and the solution of some issues of their combined action with electromechanical converter connected to electric power network of power system.

2. Materials and methods

To deduce the analytical relations, describing the processes in above turbines, as a basis the approach presented in [1] is assumed.

It is known, that the active power of hydraulic turbines is described by the expression of [2] form:

$$P = g \cdot \eta \cdot Q \cdot H \quad (1)$$

where P – power in [kW]; η – efficiency, expressed in relative units; Q – water discharge [m^3/s], H – head [m], $g=9.81$ – acceleration of gravity [m/s^2].

It is clear that the rated power of hydraulic turbine will be:

$$P_{rat} = 9,81 \cdot \eta_{rat} \cdot Q_{rat} \cdot H_{rat} \quad (2)$$

If to write down the current power in relative units, taking the rated values as the basis, we will obtain:

$$p^* = \eta^* \cdot q^* \cdot h^* \quad (3)$$

where $p^* = \frac{P}{P_{rat}}$; $\eta^* = \frac{\eta}{\eta_{rat}}$; $q^* = \frac{Q}{Q_{rat}}$; $h^* = \frac{H}{H_{rat}}$.

The expression for efficiency of hydraulic turbines is presented in [1] in the form of:

$$\eta^* = a \cdot q^{*2} + b \cdot q^* + c \quad (4)$$

In (4) equation the a , b and c indexes depend on turbine's type. For example for the Francis hydraulic turbines they are equal to $a=-0.537$; $b=1.047$; $c=0.49$ with $\eta_{rat}=0.9$ for Pelton turbines they are equal to $a=-0.224$; $b=0.483$; $c=0.741$ also with; and finally for axial flow turbines $a=-0.219$; $b=0.476$; $c=0.743$ with $\eta_{rat}=0.9$ [1].

It is also known, that in general case the rated head is described by the expression of following form:

$$H_{rat} = H_0 \cdot (1 - \lambda) \quad (5)$$

where H_0 – head corresponded to theoretical fall, i.e. without taking into account the friction loss, λ – index, taking into account the water friction when passing from top level to hydraulic turbine.

It is natural that a value of this λ index depends on a value of water flow q^* , this dependence is imaged in the expression for current value of head, which is presented in the form of:

$$H = H_0(1 - \lambda \cdot q^{*2}) \quad (6)$$

Thus with taking into account (5) and (6) expressions the expression of current head in relative units has the form:

$$h^* = \frac{H}{H_{rat}} = \frac{1 - \lambda \cdot q^{*2}}{1 - \lambda} \quad (7)$$

Inserting the (4) and (7) expressions into (3) expression and making the simple conversions we finally obtain the expression for relative power of hydraulic turbine so small HPS in the form of:

$$p^* = -A \cdot q^{*5} - B \cdot q^{*4} + C \cdot q^{*3} + D \cdot q^{*2} + E \cdot q^* \quad (8)$$

where $A = \frac{\lambda \cdot a}{1 - \lambda}$; $B = \frac{\lambda \cdot b}{1 - \lambda}$; $C = \frac{a - c \cdot \lambda}{1 - \lambda}$; $D = \frac{b}{1 - \lambda}$; $E = \frac{c}{1 - \lambda}$.

Thus the expression (8) for power of hydraulic turbines is the universal analytic expression, linking a power value of three types of turbines –Francis, Pelton and axial flow with water discharge q , in this process only change the values of a , b , c indexes and friction factor λ .

Table 1. Calculated relationship $p^*=f(q^*)$ for Francis turbine

q^*	0.4	0.5	0.6	0.7	0.8	0.9	1	1.2	1.4
p^*	0.33	0.47	0.59	0.7	0.816	0.915	1	1.11	1.12

Table 2. Calculated relationship $p^*=f(q^*)$ for Pelton turbine

q^*	0.4	0.5	0.6	0.7	0.8	0.9	1	1.2	1.4
p^*	0.39	0.502	0.61	0.718	0.819	0.916	1	1.14	1.22

Table 3. Calculated relationship $p^*=f(q^*)$ for axial (Kaplan) turbine water rate

q^*	0.4	0.5	0.6	0.7	0.8	0.9	1	1.2	1.4
p^*	0.39	0.5	0.61	0.718	0.818	0.914	1	1.1	1.2

If to assume the friction factor λ is equal to $\lambda=0.1$, then the calculated dependences $p^*=f(q^*)$ for Francis turbines will be presented in Table 1, for Pelton turbines in Table 2 and for axial flow turbines in Table 3.

Analyzing the results, given in 1÷3 Tables, the single-valued conclusions can be drawn, that the power factors of Pelton turbine and axial flow turbine in turndown of flow rate q^* from 0.4 to 1.4 are fully coincide in practice. For Francis hydraulic turbines in the range of water discharge from 0.4 to 0.7 the power value p^* as a function of flow rate q^* is a little bit less than for above turbines, which indicates that the efficiency of this turbine for small flow rates is lower than the one of Pelton's and axial flow turbines (the less flow rate the less output power value). The curves of $p^*=f(q^*)$ change for above hydraulic turbines of small HPSs are shown in Fig. 1 (further "*" indexes are removed).

And also it needs to point to the circumstance, that all above mentioned correlations were obtained for constant (rated) rotational frequency of hydraulic turbines. Thus if the hydraulic turbines were jointed with uncontrolled electric generators – synchronous or asynchronous, whose rotational frequency is constant in steady-state mode, then (8) equation becomes automatically the one of driving torque, developed by hydraulic turbine as:

$$m_{ht} = \frac{p^*}{n^*} = p_{ht} \quad (9)$$

where $n^*=1$.

3. Results and discussion

The issues of operating modes study of hydraulic units with adjustable rotational frequency are presented in [3, 4].

Let's consider the equations of state of hydraulic unit, composed of Francis turbine and classical synchronous generator with electromagnetic excitation, operating to electric network.

The equations of synchronous generator with electromagnetic excitation are presented in [5]. Thus with consideration for the equations of hydraulic turbine the general equations will be presented in the form of:

$$\begin{aligned}
 p\Psi_{ds} &= U_{ds} - \omega_r \cdot \Psi_{qs} - r_s \cdot i_{ds} \\
 p\Psi_{qs} &= U_{qs} + \omega_r \cdot \Psi_{ds} - r_s \cdot i_{qs} \\
 p\Psi_{dr} &= -\frac{r_{dr}}{x_{dr}} \cdot \Psi_{dr} + \frac{r_{dr} \cdot x_{ad}}{x_{dr}} \cdot i_{ds} + \frac{r_{dr} \cdot x_{ad}}{x_{dr}} \cdot i_{df} \\
 p\Psi_{qr} &= \frac{r_{qr}}{x_{qr}} \cdot \Psi_{qr} + \frac{r_{qr} \cdot x_{aq}}{x_{qr}} \cdot i_{qs} \\
 p\Psi_{df} &= \frac{r_{df}}{x_{ad}} \cdot U_{df}^* - r_{df} \cdot i_{df} \\
 p\omega_r &= \frac{1}{T_j} \cdot m_{ht} - \frac{1}{T_j} \cdot m_{em} \\
 i_{ds} &= \frac{x_{dr}}{\Delta d} \cdot \Psi_{ds} - \frac{\Delta d_1}{\Delta d} \cdot i_{df} - \frac{x_{ad}}{\Delta d} \cdot \Psi_{dr} \\
 i_{qs} &= \frac{x_{qr}}{\Delta q} \cdot \Psi_{qs} - \frac{x_{aq}}{\Delta q} \cdot \Psi_{qr} \\
 i_{df} &= \frac{x_{dr}}{\Delta d_2} \cdot \Psi_{df} - \frac{\Delta d_1}{\Delta d_2} \cdot i_{ds} - \frac{x_{ad}}{\Delta d_2} \cdot \Psi_{dr} \\
 m_{em} &= \Psi_{ds} \cdot i_{qs} - \Psi_{qs} \cdot i_{ds} \\
 \omega_r &= p\theta - 1 \\
 m_{ht} &= k_{tran} \cdot (-A \cdot q^5 - B \cdot q^4 + C \cdot q^3 + D \cdot q^2 + E \cdot q)
 \end{aligned} \tag{10}$$

where $U_{ds} = -U_s \cdot \sin\theta$, $U_{qs} = U_s \cdot \cos\theta$

$$\Delta d = x_{ds} \cdot x_{dr} - x_{ad}^2; \Delta q = x_{qs} \cdot x_{qr} - x_{aq}^2$$

$$\Delta d_1 = x_{dr} \cdot x_{ad} - x_{ad}^2; \Delta d_2 = x_{dr} \cdot x_{df} - x_{ad}^2$$

U_{df}^* – excitation voltage as a fraction of excitation voltage of no-load operation [6]; m_{ht} – driving torque of hydraulic turbine; k_{tran} – index of transfer of hydraulic turbine's basic units to the basic units of synchronous generator; q – water discharge in relative units. .

If to automatize the system, then it becomes necessary to link the water discharge q , which determines identically a value of turbine's driving torque m_{ht} , with the output power of generator p_{get} , i.e. besides the generator's equations it needs to take into account the equation of controller. For Francis hydraulic turbines and propeller (axial flow) ones a water discharge is controlled with the help of opening of wicket gates, which in one's turn are controlled by servomotors. A time constant of control tract is rather significant and reaches 1–2 seconds.

If to take for simplicity of analysis the inertia governor as a controller (link of first order), then its equation will be in the form of:

$$U_{out} = \frac{K}{T_p + 1} U_{in} \tag{11}$$

As it was noted, as the output power of controller U_{out} the angle of opening of wicket gates is used, which determine a water discharge q flowing through turbine, and the input power is the active power at the output of generator, which is set either by dispatcher or by a value of load diagram p_{gset} .

Thus the equations of controller and expressions for current value of active power on the terminals of generator will add to the equations of generator and hydraulic turbine (10):

$$\left. \begin{aligned} pq &= \frac{k_{12}}{T_p} (p_{gset}) - \frac{q}{T_p} \\ p_{gen} &= U_{ds} \cdot i_{ds} + U_{qs} \cdot i_{qs} \end{aligned} \right\} \quad (12)$$

where p – symbol of differentiation with respect to synchronous time $\tau=314 \cdot t$, T_p – time constant of controller in [rad], k_{12} – amplification factor (of transfer) of controller, p_{gset} – set value of power at generator's output.

In accordance with Fig. 1 k_{12} can be determined by approximation of curve $q=f(p)$, in this process at least two approximations are needed: up to $q \leq 1$ $k_{12}=k_1$ (for example for axial flow turbines $k_1 \approx 1$), and for $q \geq 1$ $k_{12}=k_2$, (rarely carried out mode).

$$k_{12} = \begin{cases} k_1 & npu \quad q \leq 1 \\ k_2 & npu \quad q \geq 1 \end{cases} \quad (13)$$

Let's perform the approximate calculation of small hydraulic unit with Francis turbine, which parameters are equal to: rated head $H_{rat}=50$ m, rated water discharge for diameter of hydraulic turbine $D=2.45$ m is equal to $Q=56$ m³/s, then the power of small HPS in nominal conditions will be equal to:

$$P_{rat.ht} = 9.81 \cdot 0.9 \cdot 56 \cdot 50 = 24.72 \text{ MW}$$

Rotational frequency of hydraulic turbine with reduced rotational frequency equal to $n_{rat}^1 = 130$ rpm [2] is determined by the formula [3].

$$n_{ht} = \frac{n_{rat}^1 \cdot \sqrt{H}}{D} = \frac{130 \sqrt{50}}{2.45} \approx 375 \text{ rpm}$$

The generator is chosen with the power of $P_{rat}=25$ MW, total power of generator S , which is taken as a basis one, is equal to $S_{rat}=31.25$ MW with rated revolutions equal to $n_{rat}=375$ rpm. In this case an index of transfer is $k_{trat}=0.8$.

The parameters of steady-state mode of the system are given in Table 4.

Table 4. Parameters of steady-state mode of the system

q	rel.unit	0.4	0.6	0.8	1
p	rel.unit	0.33	0.59	0.816	1
m_{ht}	$0.8 \cdot m_{ht}$	0.264	0.47	0.65	0.8
p_{gset}	rel.unit	0.258	0.46	0.642	0.787
$q_{st-state}$	rel.unit	0.328	0.585	0.816	1

Two first rows of the table were determined in accordance with Fig. 1 for Francis hydraulic turbine (i.e. data of Table 1). Third row displays a value of driving torque of hydraulic turbine m_{ht} with taking into account the index of transfer $k_{trat}=0.8$ according to (9) expression. The data of set active power at the output of generator p_{gset} (in relative units) were placed in the fourth row. And finally in the fifth row the data of controller output were placed, i.e. the values $q_{st-state}$ in steady-state mode (equation (12))

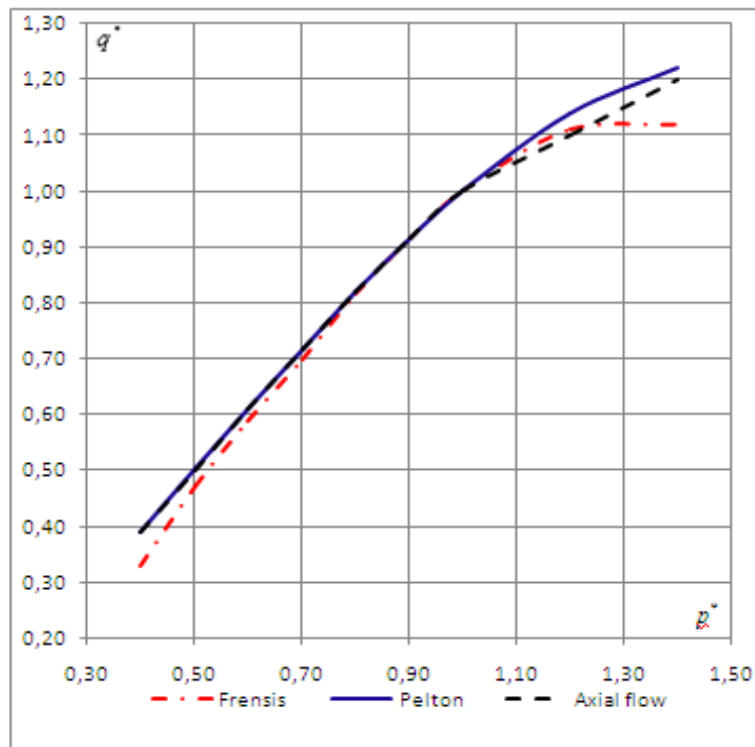


Figure 1. The relationship between turbine output power and water flow rate

In this section, we design a tracking controller for a The chosen parameters of controller are the following: time constant of controller $T_{gen}=1.5$ (471.22 rad.), amplification factor (of transfer) $k_{12}=k_1=1.27$.

The fluktoqrammas of mode parameters' change of small HPS's generator with automatic control system of water discharge (and therefore of the hydraulic turbine's power) are presented in Fig. 2. The data of generator and controller and algorithm of modeling are given in Appendix. The fluktoqramma of change of generator's electromagnetic torque $m_{em}=f(\tau)$ is given in Fig. 2(a). According to the modeling algorithm at the first stage the asynchronous start is carried out without load and with short-circuited excitation winding within the period up to $\tau=5000$ radian. A slip s changes in accordance with Fig 2(b). At 5000 radian the open-circuit voltage $U_f=1$ is supplied to the excitation winding, and the machine locks in synchronism, the rotational frequency of generator sets at the mark $\omega=1$ (Fig.2, c).

At 15000 radian the set value of active power $p_{set1}=0.787$ inputs to the equation of controller, a torque of hydraulic turbine in this process is formed according to expression $m_{ht}=k_{tran} \cdot q_1=-0.8 \cdot q_1$, here with a value of water discharge q sets at a level of $q_1=1$ (Fig.2, d) and the relevant to it value of driving torque of turbine is equal to $m_{ht}=-0.8$ (minus sign indicates the generator mode) (Fig.2, a). Further in accordance with the set values of output power of generator $p_{set2}=0.642$, $p_{set3}=0.46$, $p_{set4}=0.258$ the q values, accordingly equal to $q_2=0.815$, $q_3=0.584$ and $q_4=0.327$ (Fig.2, d), are set at the output of controller. In accordance with these q values the torque values m_{ht} and m_{em} automatically form, the values of last one are accordingly equal to $m_{em1}=-0.799$, $m_{em2}=-0.652$, $m_{em3}=-0.467$ and $m_{em4}=-0.262$. These values of electromagnetic torque are corresponded to the values of active powers at the generator's output $p_{gen1}=-0.784$, $p_{gen2}=-0.64$, $p_{gen3}=-0.45$ and $p_{gen4}=-0.254$ (Fig. 2, e). Comparing the values of current powers of generator $p_{gen1} \dots p_{gen4}$ with the set values it can be stated, that the error amount because of approximation and offset of chosen type of controller doesn't exceed 3%.

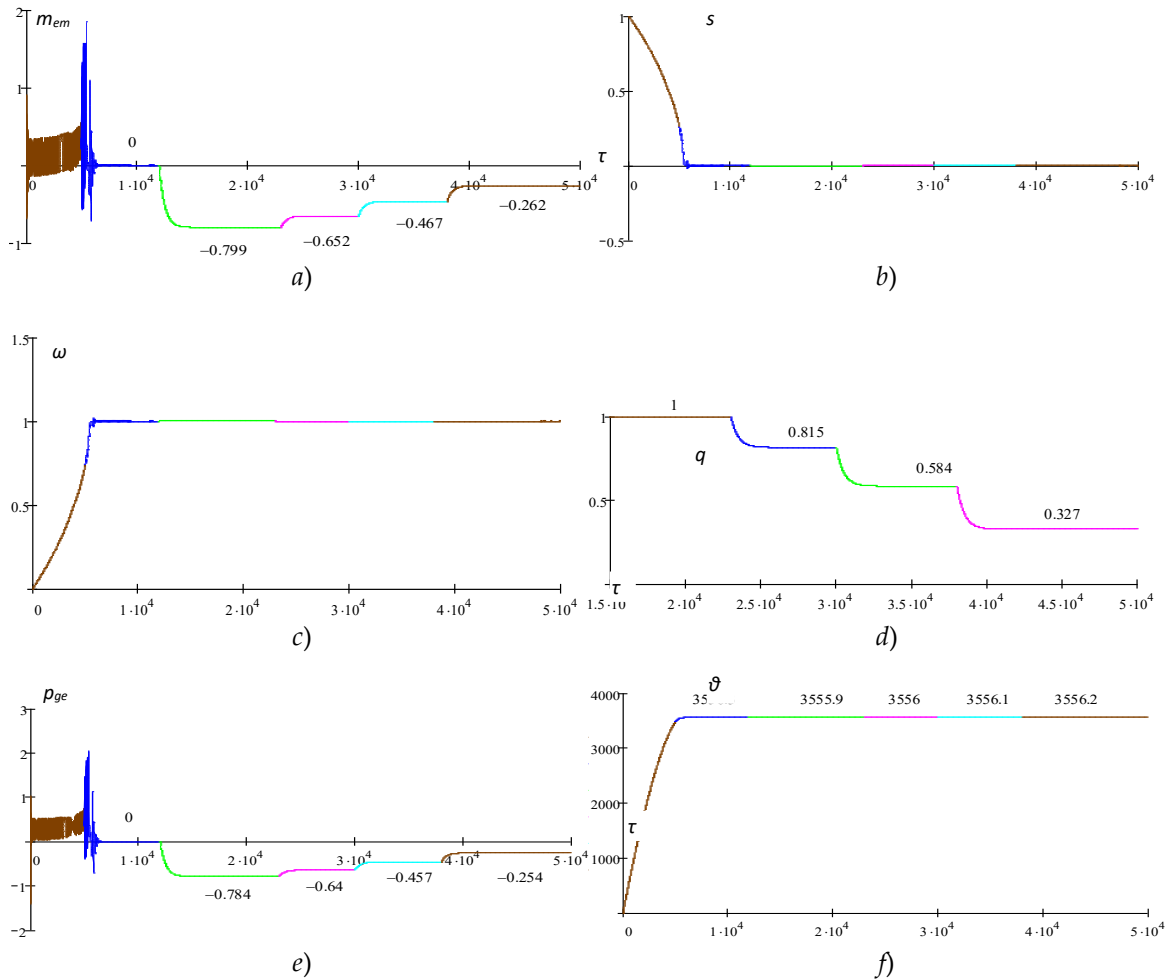


Figure 2. Fluctograms of operating parameters variations of small hydroelectric power generator

The curve of change of synchronous machine's interior angle θ is given in Fig.(2, f); for almost rated active load ($p_{gen1}=-0.784$) this angle is equal to $\theta_{rat}=0.4$ radian (~ 23 deg.), and for minimum active load ($p_{gen4}=-0.254$) is equal to $\theta_4=0.1$ radian (~ 5.7 deg.).

In conclusion it needs to note again, that the controller in these researches is chosen as the relaxation circuit of first order, which is a simplification, needs for more visual demonstration of system's principle of operation. The choice of more complicated controller will not influence on the idea and algorithm of construction of general modeling system.

Conclusions

1. The mathematical model is developed for study of operating modes of small HPSS' hydroelectric units, equipped with Frensis, Pelton and axial flow turbines (Kaplan turbines with antideflexion mounting of blades). For three above mentioned types of turbines the analytic expression was obtained in a form of multinomial, which connects the power values of turbines, expressed in relative units, with a value of flow rate of energy carrier – water; the structure of multinomial is invariable and only the values of its factors change according to turbines' types.

2. Collaborate research of hydraulic turbines with synchronous generators on full equations and also of the controller has demonstrated the working capacity, accuracy and effectiveness of developed mathematical model, which allows its using for oriented calculations of operating modes of small HPSS both at the designing stage and in operating conditions.

Appendices

Appendix 1. Parameters of Synchronous generator

$x_{ds}=0.986$	$x_{ad}=0.787$
$x_{qs}=0.63$	$x_{dr}=1$
$x_{aq}=0.435$	$r_s=0.02$
$x_{qr}=0.7$	$r_{qr}=0.019$
$x_{df}=1.1$	$r_{dr}=0.028$
$T_f=1000$ rad. (~ 1.5 s.)	$U_s=1$

Calculated value $\Delta d=0.367$; $\Delta d_1=0.167$; $\Delta d_2=0.48$; $\Delta q=0.255$.

Controller parameters: gain ratio $k_{12}=k_1=1.27$; time constant $T_{gen}=471.22$ rad.

Appendix 2. Algorithm of the mathematical model

$$D(t, Y) = \begin{bmatrix} -\sin(Y_6) + Y_1 - Y_1 \cdot Y_5 - 0.02 \cdot i_{sd} \\ \cos(Y_6) - Y_0 + Y_0 \cdot Y_5 - 0.02 \cdot i_{sq} \\ -0.028 \cdot Y_2 - 0.022(3.23 \cdot Y_0 - 1.12 \cdot Y_4 - 1.65 \cdot Y_2) + 0.002 \cdot i_{df} \\ -0.027 \cdot Y_3 - 0.0117 \cdot i_{sq} \\ 0.057 \cdot U_{df}^* - 0.045 \cdot i_{df} \\ 0.001 \cdot (-k_{ran} \cdot Y_7) - 0.001[Y_0 \cdot (i_{sq}) - Y_1 \cdot (i_{sd})] \\ Y_5 \\ 0.0027 \cdot (p_{set}) - 0.002126 \cdot Y_7 \end{bmatrix} \quad D(t, Y) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$Y_0=\Psi_{ds}; Y_1=\Psi_{qs}; Y_2=\Psi_{dr}; Y_3=\Psi_{qr}; Y_4=\Psi_{df}; Y_5=s; Y_6=\theta; Y_7=q.$$

$$\omega = p\theta - 1 = s - 1$$

$$i_{ds} = 3.23 \cdot Y_0 - 1.12 \cdot Y_4 - 1.65 \cdot Y_2$$

$$i_{qs} = 2.745 \cdot Y_1 - 1.7 \cdot Y_3$$

$$i_{df} = 2.47 \cdot Y_4 - 1.12 \cdot Y_0 - 1.07 \cdot Y_2$$

p_{set} – active power preset value in generator output

References

[1] J.N.G.Voros, C.T.Kiranoudis, Z.B.Maroulis. Short-cut design of small hydroelectric plans. <http://bse.srv214.bse.vt.edu/Grisso/Ethiopia/BooksResources/Hydro-electric>.

[2] G.I.Krivchenko, Hydraulic machines (turbines, pumps). Moscow, "Energy" journal pg. 236. 1978.

[3] R.I.Mustafayev, L.G.Hasanov, M.M.Musayev, E.M.Mamedov, Kh.I.Nabiyev. "Modeling and analysis of operating regimes of hydraulic units of small hydropower plant with double fed machine", Electromechanics, No. 6, pp. 79-86, Moscow, Russian, 2015.

[4] R.I.Mustafayev, L.G.Hasanov, M.M.Musayev. "Modeling and study of hydropower units of small hydro power plants with frequency-controlled permanent magnet synchronous generators", HEI of CISEnergetics, No. 2, pp. 102-117, Minsk, Belarus, 2016.

[5] R.I.Mustafayev, L.H. Hasanova. "Modeling and study state of working of wind-electric installations synchronous generator under frequency steering", Electricity, No. 7, pp. 34-40, Moscow, Russian, 2010.

[6] N.I.Sokolov. Application of analog computers in power systems. Under editorship of. Prof. Publishing house "Energy" Moscow, Russia, p. 408, 1964.

[7] V.L.Anchimyuk. Automatic control theory. Publishing "Graduate school", p.350. Minsk, 1979.

A COMPOUND OF GAMMA AND SHANKER DISTRIBUTION

Mousumi Ray*

•

Department of Statistics, Assam University, Silchar, India
mousumiray616@gmail.com

Rama Shanker

•

Department of Statistics, Assam University, Silchar, India
shankerrama2009@gmail.com

*Corresponding Author

Abstract

This paper considers a new lifetime distribution called Gamma-Shanker distribution which is a compound of Gamma and Shanker distribution. Many important properties of the suggested distribution including its shape, Inverse moments, hazard rate function, reversed hazard rate function, quantile function and stress-strength reliability have been discussed. The estimation of its parameters has been discussed using maximum likelihood estimation. Goodness of fit of the proposed distribution has been explained with two examples of real lifetime data from biomedical sciences and it shows that the proposed distribution gives much closure fit over the considered distributions

Keywords: Lifetime distribution, Statistical Properties, Stress-strength reliability, Maximum Likelihood estimation, Goodness of fit.

1. Introduction

The real lifetime data from different fields of knowledge are generally stochastic in nature and requires a distribution which can capture the variation to a great extent. There were two classical one parameter lifetime distributions namely exponential and Lindley by Lindley [1] in use for analyzing the lifetime data. Shanker et al [2] have detailed comparative study on the goodness of fit of exponential and Lindley distributions and observed that in some data sets exponential gives better fit than Lindley whereas in some datasets Lindley gives much closure fit than exponential and there were some datasets where neither exponential nor Lindley gives good fit. One of the important advantages of Lindley distribution over exponential distribution is that hazard rate for exponential distribution is constant while the hazard rate of Lindley distribution is not constant. The gamma and Weibull distributions which contain exponential distribution as particular case are the classical two-parameter lifetime distributions for the analysis and modeling of lifetime data. Shanker et al [3] have detailed comparative study on modeling of lifetime data using gamma and Weibull distributions and observed that both gamma and Weibull distributions are competing and each has some advantages over the others and there were some datasets where both gamma and Weibull did not give good fit.

Recently, Abdi et al [4] proposed gamma-Lindley distribution (G-LD) by compounding gamma distribution with Lindley distribution assuming that the scale parameter of gamma distribution follows Lindley distribution.

As we know that the nature of lifetime data is in general stochastic in nature and thus have different failure rates and to capture the analysis of such lifetime data different distributions are required. The decreasing and unimodal (upside down bathtub) failure rates have a lot of applications in survival analysis including the situation where the probability of an event in a fixed time interval in the future decreases over time and can be observed in case of infant mortality rate where earlier failures are eliminated or corrected, as observed by Finkelstein [5]. One of the practical examples of decreasing failure rate is the failure in the air conditioning systems explained by Proschan [6]. It has been observed by Lie and Xie [7] that if the main reasons of the failures of products are caused by fatigue and corrosion, the failure rates of those products exhibit unimodal shapes. For example, the data relating to breast cancer and infection in biomedical sciences with new viruses are generally of unimodal shape, as observed by Demicheli et al [8].

Recently, Abdi et al [4] derived gamma-Lindley distribution (G-LD) by compounding gamma(α, λ) with Lindley (β) distribution when the scale parameter λ of gamma distribution follows Lindley distribution. The probability density function (pdf) and cumulative distribution function (cdf) of G-LD are given by

$$f(x; \alpha, \beta) = \frac{\alpha \beta^2 (1 + \alpha + \beta + x) x^{\alpha-1}}{(\beta + 1)(\beta + x)^{\alpha+2}}; x > 0, \alpha > 0, \beta > 0 \tag{1}$$

$$F(x; \alpha, \beta) = \frac{x^\alpha [(\beta + 1)x + (1 + \alpha + \beta)\beta]}{(\beta + 1)(\beta + x)^{\alpha+1}}; x > 0, \alpha > 0, \beta > 0. \tag{2}$$

The pdf and cdf of Shanker distribution obtained by Shanker [9] are given by

$$f(x; \theta) = \frac{\theta^2}{\theta^2 + 1} (\theta + x) e^{-\theta x}; x > 0, \theta > 0 \tag{3}$$

$$F(x; \theta) = 1 - \left[\frac{\theta^2 + 1 + \theta x}{\theta^2 + 1} \right] e^{-\theta x}; x > 0, \theta > 0 \tag{4}$$

It should be noted that Shanker distribution is a two-component mixture of exponential distribution having scale parameter (θ) and a gamma distribution having shape parameter 2 and scale parameter (θ) with mixing proportion $\frac{\theta^2}{\theta^2 + 1}$. The statistical properties, estimation of parameter and application of Shanker distribution are available in Shanker [9].

The main motivations for proposing Gamma-Shanker distribution are as follows:

The first motivation G-SD lies in the fact that if X is the lifetime of component and λ is the scale parameter of the distribution of X and suppose that in the population from which the sample is being drawn has some variability in the scale parameter, then that variability can be described by the distribution of λ . The second motivation is that in real life situation components in a certain population differs sustainability from each other and this heterogeneity can easily be taken into consideration for the analysis of such population using compound distribution. In fact, the G-SD distribution can be shown as mixture representation like G-LD which is recommended for such variation in the population. The third motivation for proposing G-SD as a compound of Gamma and Lindley lies in context of Bayesian inference is that G-SD arises when Gamma $f(x|\alpha, \lambda)$ represents the distribution of future observations and the Lindley $f(\lambda|\beta)$ is the posterior distribution of the parameters of $f(x|\alpha, \lambda)$, given the information in a sample of observed data. The fourth motivation is that there are several lifetime data which have long right tail and the G-SD is most suitable for long right tail data. The fifth and the final motivation is that as the Shanker distribution provides much closure fit than exponential and Lindley distribution and G-LD distribution provides better fit than Gamma, Weibull and other two-

parameter distribution, it is expected and hoped that G-SD would provide much better fit than G-LD and other two-parameter distributions. In the present paper, statistical properties, estimation of parameters and applications of G-SD have been discussed.

2. Compound of Gamma and Shanker Distribution

Following the approach of obtaining G-LD distribution, the pdf and the cdf of gamma-Shanker distribution (G-SD) are obtained as

$$f(x; \alpha, \beta) = \frac{\alpha\beta^2(1 + \alpha + \beta x + \beta^2)x^{\alpha-1}}{(1 + \beta^2)(\beta + x)^{2+\alpha}}; x > 0, \alpha > 0, \beta > 0 \quad (5)$$

$$F(x; \alpha, \beta) = \frac{x^\alpha [x(1 + \beta^2) + (1 + \alpha + \beta^2)\beta]}{(1 + \beta^2)(\beta + x)^{1+\alpha}}; x > 0, \alpha > 0, \beta > 0 \quad (6)$$

The shapes of the pdf and the cdf of G-SD for varying values of parameters are shown in the following figures 1 and 2 respectively. It is obvious that the pdf of G-SD is unimodal and positively skewed.

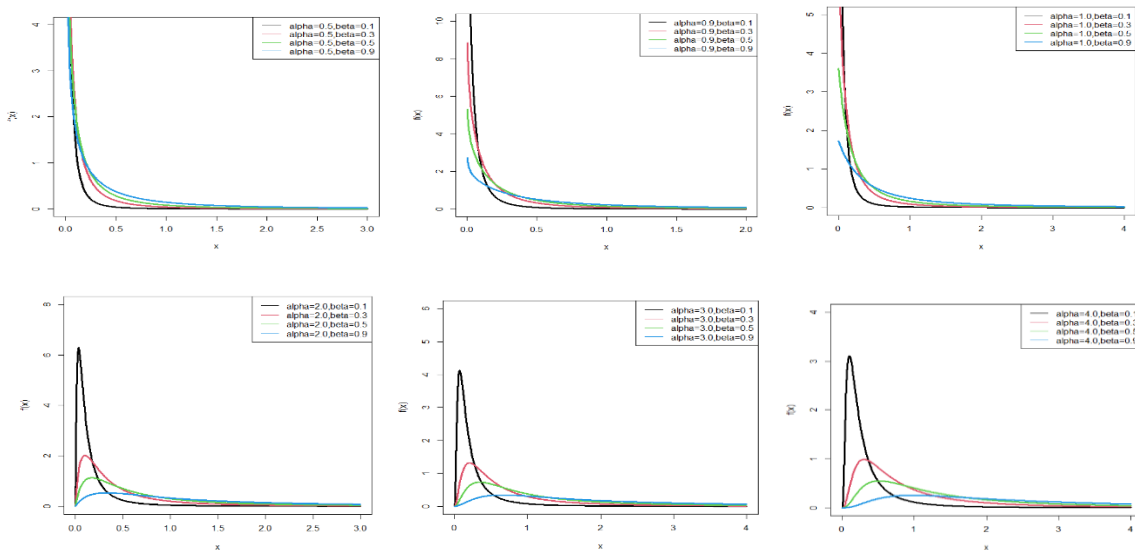


Figure 1: pdf plots of G-SD for some selected values of parameters

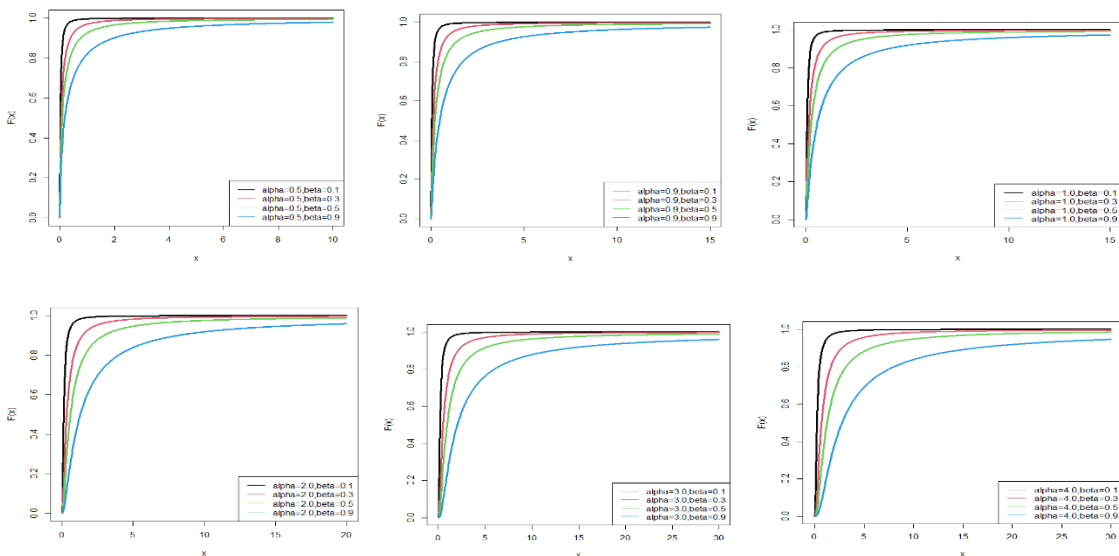


Figure 2: cdf plots of G-SD for some selected values of parameters

In the following theorem an attempt has been made to prove the decreasing nature of G-SD and the unimodality.

Theorem 1: The pdf of G-SD distribution is decreasing for $\alpha \leq 1$ and unimodal for $\alpha > 1$

Proof: We have,

$$f(x) = \frac{\alpha\beta^2(1+\alpha+\beta x+\beta^2)x^{\alpha-1}}{(1+\beta^2)(\beta+x)^{2+\alpha}}$$

$$\log f(x) = C + \log(1+\alpha+\beta^2+\beta x) + (\alpha-1)\log(x) - (\alpha+2)\log(\beta+x).$$

where C is a constant. We have

$$\frac{d}{dx} \log f(x) = \frac{\alpha-1}{x} - \frac{(\alpha+1)(\beta^2+\beta x+\alpha+2)}{(\beta+x)(1+\alpha+\beta^2+\beta x)}$$

For $\alpha \leq 1$, $\frac{d}{dx} \log f(x) < 0$ and this means that $f(x)$ is decreasing for all x .

For $\alpha > 1$

$$\frac{d}{dx} \log f(x) = \frac{\alpha-1}{x} - \frac{(\alpha+1)(\beta^2+\beta x+\alpha+2)}{(\beta+x)(1+\alpha+\beta^2+\beta x)} = 0$$

This gives the following quadratic equation

$$2\beta x^2 - (\alpha\beta^2 - 3\alpha - 3\beta^2 - 3)x + (\beta + \alpha\beta + \beta^3 - \alpha\beta - \alpha^2\beta - \alpha\beta^3) = 0$$

A real root of the above equation is given by

$$x_0 = \frac{(\alpha\beta^2 - 3\alpha - 3\beta^2 - 3) + \sqrt{\alpha^2\beta^4 + 9\alpha^2 + \beta^4 + 2\alpha^2\beta^2 + 2\alpha\beta^4 + 12\alpha\beta^2 + 18\alpha + 10\beta^2 + 9}}{4\beta},$$

which is also the mode of G-SD.

Since, $\lim_{x \rightarrow 0} f(x) = 0$ and $\lim_{x \rightarrow \infty} f(x) = 0$, the pdf $f(x)$ is unimodal for $\alpha > 1$.

3. Hazard rate function and Reversed hazard rate function

The hazard rate function and the reverse hazard rate function are two important functions of a distribution. The reliability (survival) function of G-SD is given by

$$R(x; \alpha, \beta) = 1 - F(x; \alpha, \beta) = \frac{(\beta+x)^{\alpha+1}(1+\beta^2) - x^\alpha [x(\beta^2+1) + (1+\alpha+\beta^2)\beta]}{(\beta+x)^{\alpha+1}(1+\beta^2)} \quad (7)$$

The corresponding Hazard rate and Reversed Hazard rate function of G-SD are obtained as

$$h(x; \alpha, \beta) = \frac{f(x; \alpha, \beta)}{R(x; \alpha, \beta)} = \frac{\alpha\beta^2 x^{\alpha-1} (1+\alpha+\beta x+\beta^2)}{(\beta+x) [(\beta+x)^{\alpha+1} (1+\beta^2) - x^\alpha (1+\beta^2)(\beta+x) - \alpha\beta x^\alpha]} \quad (8)$$

$$r(x; \alpha, \beta) = \frac{f(x; \alpha, \beta)}{F(x; \alpha, \beta)} = \frac{\alpha\beta^2 (1+\alpha+\beta x+\beta^2)}{x(\beta+x) [x(1+\beta^2) + (1+\alpha+\beta^2)\beta]} \quad (9)$$

The behavior of $h(x)$ when $x \rightarrow 0$ and $x \rightarrow \infty$, respectively are given by

$$\lim_{x \rightarrow 0} h(x) = \begin{cases} \infty, \alpha < 1 \\ \frac{(2 + \beta^2)}{\beta(1 + \beta^2)}, \alpha = 1 \\ 0, \alpha > 1 \end{cases} \text{ and } \lim_{x \rightarrow \infty} h(x) = 0$$

$$\lim_{x \rightarrow 0} r(x) = \infty \text{ and } \lim_{x \rightarrow \infty} r(x) = 0.$$

The natures of hazard rate and the reversed hazard rate function of G-SD are shown in the figures 3 and 4 respectively.

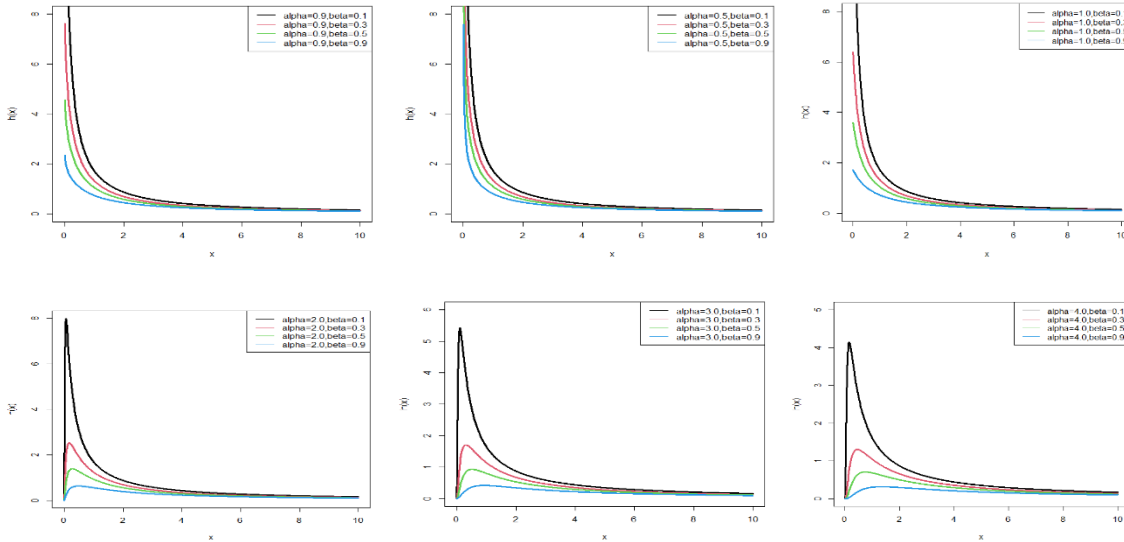


Figure 3: Plots of hazard rate function of G-SD for some parameter values

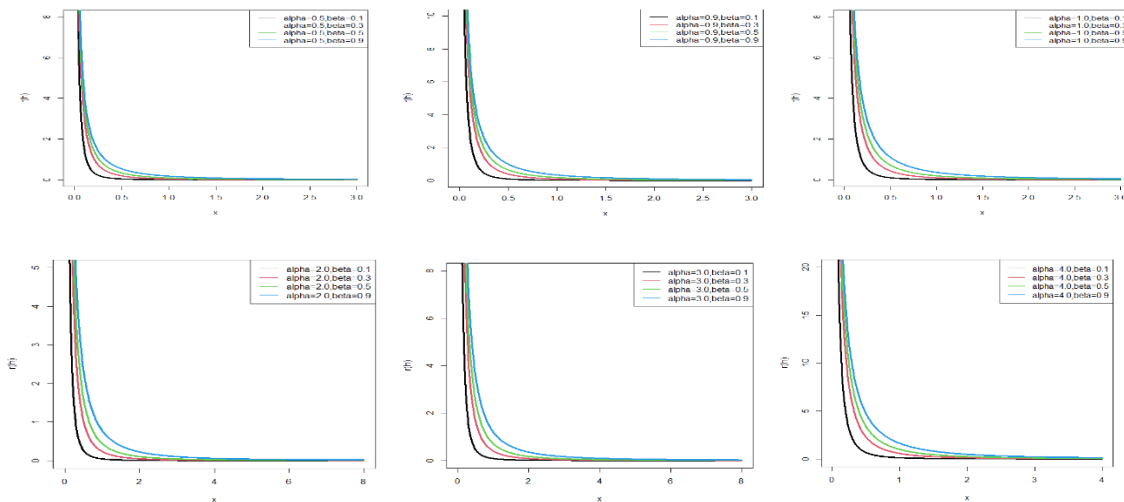


Figure 4: Plots of Reversed hazard rate function of G-SD for some parameter values

From figures 3 and 4 it is quite clear that the hazard rate function of G-SD is decreasing for $\alpha \leq 1$ and unimodal for $\alpha > 1$ and the reversed hazard rate function of the G-SD is also decreasing, which is also shown in the following theorems 2 and 3, respectively.

Theorem 2: The hazard rate function of the G-SD is decreasing for $\alpha \leq 1$ and unimodal for $\alpha > 1$

Proof: We have

$$f(x) = \frac{\alpha\beta^2(1+\alpha+\beta x+\beta^2)x^{\alpha-1}}{(1+\beta^2)(\beta+x)^{2+\alpha}}, \text{ and}$$

$$f'(x) = \frac{\alpha\beta^2x^{\alpha-1}[(\beta+x)(\alpha-1)(1+\alpha+\beta^2+\beta x)+x\beta(\beta+x)-x(\alpha+2)(1+\alpha+\beta^2+\beta x)]}{x(1+\beta^2)(\beta+x)^{3+\alpha}}$$

Now, suppose that

$$\phi(x) = -\frac{f'(x)}{f(x)} = -\frac{(\alpha-1)}{x} + \frac{(\alpha+1)(2+\alpha+\beta x+\beta^2)}{(\beta+x)(1+\alpha+\beta x+\beta^2)}.$$

This gives

$$\phi'(x) = \frac{(\alpha-1)}{x^2} + \frac{(\alpha+1)(6\beta^2+6\alpha\beta^2+3\beta^4+4\beta^3x+6\beta x+6\alpha\beta x+3\beta^2x^2+3\alpha+\alpha^2+2)}{(\beta+x)^2(1+\alpha+\beta x+\beta^2)^2}$$

It is quite obvious that for $\alpha \leq 1$, $\phi'(x) < 0$ and for $\alpha > 1$, $\phi(x) < 0$ has a global maximum at mode (say x_0).

Theorem 3: The reversed hazard rate function of the G-SD is decreasing

Proof: We have,

$$r(x) = \frac{\alpha\beta^2(1+\alpha+\beta x+\beta^2)}{x(\beta+x)[x(1+\beta^2)+(1+\alpha+\beta^2)\beta]}.$$

This gives

$$\frac{d}{dx} \log r(x) = \frac{-1-\alpha-\beta^2}{(1+\alpha+\beta x+\beta^2)[x(1+\beta^2)+(1+\alpha+\beta^2)\beta]} - \frac{1}{x} - \frac{1}{(\beta+x)} < 0 \text{ for all } \alpha, \beta$$

Therefore, reversed hazard function is decreasing for any value of the parameters α, β .

4. Quantiles and Moments

The p th quantiles x_p of G-SD. defined by $F(x_p) = p$, is the root of the equation

$$\frac{x_p^\alpha [x_p(\beta^2+1)+(1+\alpha+\beta^2)\beta]}{(\beta^2+1)(\beta+x_p)^{\alpha+1}} = p \tag{10}$$

This gives

$$x_p = \frac{\alpha\beta - \beta(1+\beta^2) \left[\left(1 + \frac{\beta}{x_p}\right)^\alpha p - 1 \right]}{(1+\beta^2) \left[\left(1 + \frac{\beta}{x_p}\right)^\alpha p - 1 \right]} \tag{11}$$

It should be noted that this x_p may be used to generate G-SD random variates. Further, the median of G-SD can be obtained from above equation by taking $p = \frac{1}{2}$.

The moments of G-SD can be obtained as follows:

If $X \sim \text{G-SD}(\alpha, \beta)$ then,

$$E(X) = E(E(X | \lambda)) = E\left(\frac{\alpha}{\lambda}\right) = \alpha E\left(\frac{1}{\lambda}\right) = \infty$$

Thus, in general, $E(X^r) = \infty$ for $r \geq 1$. This means that all moments of G-SD are infinite and hence G-SD has no mean. As G-SD has no mean, if we take a sample (X_1, X_2, \dots, X_n) from G-SD, then mean \bar{X} does not tend to a particular value. Since G-SD has no raw and central moments, we have to derive inverse moments. Negative moments are useful in several real life applications, such as life testing problems and estimation purpose. The negative moments for G-SD can be obtained as follows:

The r^{th} negative moment about origin, $\mu_{(-r)}'$, of the G-SD can be obtained as

$$\begin{aligned} \mu_{(-r)}' &= E(X^{-r}) = E(E(X^{-r} | \lambda)) \\ &= \int_0^\infty \left[\int_0^\infty x^{-r} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx \right] \frac{\beta^2}{\beta^2 + 1} (\beta + \lambda) e^{-\beta \lambda} d\lambda \\ &= \frac{\Gamma(\alpha - r)}{\Gamma(\alpha)} \cdot \frac{r!(\beta^2 + r + 1)}{\beta^r (\beta^2 + 1)}; r = 1, 2, 3, \dots \end{aligned} \tag{12}$$

Thus, for $r = 1, 2, 3, 4$, we have

$$\mu_{(-1)}' = E\left(\frac{1}{X}\right) = \frac{(\beta^2 + 2)}{\beta(\beta^2 + 1)(\alpha - 1)}, \alpha > 1 \tag{13}$$

$$\mu_{(-2)}' = E\left(\frac{1}{X^2}\right) = \frac{2(\beta^3 + 3)}{\beta^2(\alpha - 1)(\alpha - 2)(\beta^2 + 1)}, \alpha \geq 2 \tag{14}$$

$$\mu_{(-3)}' = E\left(\frac{1}{X^3}\right) = \frac{6(\beta^3 + 4)}{\beta^3(\alpha - 1)(\alpha - 2)(\alpha - 3)(\beta^2 + 1)}, \alpha \geq 3 \tag{15}$$

$$\mu_{(-4)}' = E\left(\frac{1}{X^4}\right) = \frac{24(\beta^3 + 5)}{\beta^4(\alpha - 1)(\alpha - 2)(\alpha - 3)(\alpha - 4)(\beta^2 + 1)}, \alpha \geq 4 \tag{16}$$

It is obvious from the above expressions for negative moments that negative moments are not defined for $\alpha \leq 1$.

5. Extreme Order Statistics

Let, $X_{1:n}, \dots, X_{n:n}$ be the order statistics of a random sample of size n from the G-SD(α, β) distribution with distribution function $F(x)$. The cdf of the minimum order statistic $X_{1:n}$ is given by

$$F_{X_{1:n}}(x) = 1 - [1 - F(x)]^n = 1 - \left[\frac{(1 + \beta^2)(\beta + x)^{1+\alpha} - x^\alpha(\beta + x)(1 + \beta^2) - \alpha\beta x^\alpha}{(1 + \beta^2)(\beta + x)^{1+\alpha}} \right] \tag{17}$$

The cdf of the maximum order statistic $X_{n:n}$ is given by

$$F_{X_{n:n}}(x) = [F(x)]^n = \left[\frac{x^\alpha}{(\beta + x)^\alpha} + \frac{\alpha\beta x^\alpha}{(1 + \beta^2)(\beta + x)^{1+\alpha}} \right]^n \tag{18}$$

6. Stochastic Orderings

In probability theory and Statistics, a stochastic order quantifies the concept of one random variable being “bigger” than other. In many problems, it becomes necessary to compare two lifetime

distributions with reference to some of their characteristics. Stochastic orders provide the necessary tools in such case.

A random variable X is said to be smaller than a random variable Y in the

- i. Stochastic order ($X \prec_{st} Y$) if $F_X(x) \geq F_Y(y)$ for all x
- ii. Hazard rate order ($X \prec_{hr} Y$) if $h_X(x) \geq h_Y(y)$ for all x
- iii. Mean residual life order ($X \prec_{mrl} Y$) if $m_X(x) \geq m_Y(y)$ for all x
- iv. Likelihood ratio order ($X \prec_{lr} Y$) if $\frac{f_X(x)}{f_Y(y)}$ decrease in x

The following results due to Shaked and Shantikumar [10] are well known for establishing stochastic ordering of distributions:

$$X \prec_{lr} Y \Rightarrow X \prec_{hr} Y \Rightarrow X \prec_{mrl} Y$$

$$\Downarrow$$

$$X \prec_{st} Y$$

Theorem 4: Let $X_1 \sim \text{G-SD}(\alpha_1, \beta_1)$ and $X_2 \sim \text{G-SD}(\alpha_2, \beta_2)$. If $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 \leq \beta_2$ and if $\beta_1 = \beta_2 = \beta \geq 1$ with $\alpha_1 \leq \alpha_2$, then $X_1 \prec_{lr} X_2 \Rightarrow X_1 \prec_{hr} X_2 \Rightarrow X_1 \prec_{st} X_2$.

Proof: We have

$$\frac{f_{X_1}(x)}{f_{X_2}(x)} = \frac{\alpha_1 \beta_1^2 (\beta_2^2 + 1) (1 + \alpha_1 + \beta_1 x + \beta_1^2) (\beta_2 + x)^{\alpha_2 + 2}}{\alpha_2 \beta_2^2 (\beta_1^2 + 1) (1 + \alpha_2 + \beta_2 x + \beta_2^2) (\beta_1 + x)^{\alpha_1 + 2}} x^{\alpha_1 - \alpha_2}$$

If $\alpha_1 = \alpha_2 = \alpha$, we get

$$g_1(x) = \frac{\beta_1^2 (\beta_2^2 + 1) (1 + \alpha + \beta_1 x + \beta_1^2) (\beta_2 + x)^{\alpha + 2}}{\beta_2^2 (\beta_1^2 + 1) (1 + \alpha + \beta_2 x + \beta_2^2) (\beta_1 + x)^{\alpha + 2}}$$

$$\frac{d \log g_1(x)}{dx} = \left(\frac{\alpha + 2}{\beta_2 + x} - \frac{\beta_2}{1 + \alpha + \beta_2 x + \beta_2^2} \right) - \left(\frac{\alpha + 2}{\beta_1 + x} - \frac{\beta_1}{1 + \alpha + \beta_1 x + \beta_1^2} \right)$$

$$= q(\beta_2) - q(\beta_1),$$

where

$$q(\beta) = \left(\frac{\alpha + 2}{\beta + x} - \frac{\beta}{1 + \alpha + \beta x + \beta^2} \right)$$

$$\frac{d}{d\beta} q(\beta) = \frac{-(\alpha + 2)}{(\beta + x)^2} - \frac{1 + \alpha + \beta x}{(1 + \alpha + \beta x + \beta^2)^2} < 0$$

If $\alpha_1 = \alpha_2 = \alpha$, then X_1 is stochastically smaller than X_2 with respect to the likelihood ratio if and only if $\beta_1 \leq \beta_2$

Case II: If $\beta_1 = \beta_2 = \beta \geq 1$, we get

$$g_2(x) = \frac{\alpha_1 (1 + \alpha_1 + \beta x + \beta^2)}{\alpha_2 (1 + \alpha_2 + \beta x + \beta^2)} \left(\frac{x}{\beta + x} \right)^{\alpha_1 - \alpha_2}$$

$$\frac{d \log g_2(x)}{dx} = \left(\frac{\beta}{1 + \alpha_1 + \beta x + \beta^2} + \frac{\alpha_1}{x} - \frac{\alpha_1}{\beta + x} \right) - \left(\frac{\beta}{1 + \alpha_2 + \beta x + \beta^2} + \frac{\alpha_2}{x} - \frac{\alpha_2}{\beta + x} \right)$$

$$= u(\alpha_1) - u(\alpha_2)$$

Where

$$u(\alpha) = \left(\frac{\beta}{1 + \alpha + \beta x + \beta^2} + \frac{\alpha}{x} - \frac{\alpha}{\beta + x} \right)$$

$$\frac{d}{d\alpha} u(\alpha) = \frac{-1}{(1 + \alpha + \beta x + \beta^2)^2} + \frac{1}{x} - \frac{1}{\beta + x} > 0 \text{ for } \beta \geq 1.$$

Thus, it is obvious that $\alpha_1 \leq \alpha_2$, $\frac{d \log g_2(x)}{dx} < 0$. Hence, if $\beta_1 = \beta_2 = \beta \geq 1$ then X_1 is stochastically smaller than X_2 with respect to the likelihood ratio if and only if $\alpha_1 \leq \alpha_2$.

7. Estimation of parameters

Let (x_1, x_2, \dots, x_n) be the observed values of a random sample (X_1, X_2, \dots, X_n) from the G-SD. Then the Likelihood function is given by

$$L(\alpha, \beta) = \left(\frac{\alpha\beta^2}{\beta^2+1}\right)^n \frac{\prod_{i=1}^n (1+\alpha+\beta x_i+\beta^2) \left(\prod_{i=1}^n x_i\right)^{\alpha-1}}{\prod_{i=1}^n (\beta+x_i)^{\alpha+2}} \tag{19}$$

The log-likelihood function of G-SD is thus obtained as

$$\begin{aligned} \ln L(\alpha, \beta) &= n \ln \alpha + 2n \ln \beta - n \ln(\beta^2+1) + \sum_{i=1}^n \ln(1+\alpha+\beta x_i+\beta^2) \\ &+ (\alpha-1) \sum_{i=1}^n \ln(x_i) - (\alpha+2) \sum_{i=1}^n \ln(\beta+x_i) \end{aligned} \tag{20}$$

The maximum likelihood estimators (MLEs) of α and β are the simultaneous solutions of the following log-likelihood equations

$$\begin{aligned} \frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} &= \frac{n}{\alpha} + \sum_{i=1}^n \frac{1}{(1+\alpha+\beta x_i+\beta^2)} + \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \ln(\beta+x_i) = 0 \\ \frac{\partial \ln L(\alpha, \beta)}{\partial \beta} &= \frac{2n}{\beta} - \frac{2n\beta}{(\beta^2+1)} + \sum_{i=1}^n \frac{(x_i+2\beta)}{(1+\alpha+\beta x_i+\beta^2)} - (\alpha+2) \sum_{i=1}^n \frac{1}{(\beta+x_i)} = 0 \end{aligned}$$

It is very difficult to solve these two log-likelihood equations directly, so we will use Fisher's scoring method. We have

$$\begin{aligned} \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \alpha^2} &= \frac{-n}{\alpha^2} - \sum_{i=1}^n \frac{1}{(1+\alpha+\beta x_i+\beta^2)^2} \\ \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \alpha \partial \beta} &= -\sum_{i=1}^n \frac{2\beta+x_i}{(1+\alpha+\beta x_i+\beta^2)^2} - \sum_{i=1}^n \frac{1}{\beta+x_i} = \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \beta \partial \alpha} \\ \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \beta^2} &= \frac{-2n}{\beta^2} + 2n \left[\frac{(\beta^2-1)}{(\beta^2+1)^2} \right] + \sum_{i=1}^n \left[\frac{2+2\alpha-2\beta^2-x_i^2-2\beta x_i}{(1+\alpha+\beta x_i+\beta^2)^2} \right] + \sum_{i=1}^n \frac{\alpha+2}{(\beta+x_i)^2} \end{aligned}$$

The following equation can be solved for MLE's of α and β of G-SD

$$\begin{pmatrix} \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \alpha^2} & \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \beta \partial \alpha} & \frac{\partial^2 \ln L(\alpha, \beta)}{\partial \beta^2} \end{pmatrix}_{\hat{\alpha}=\alpha_0, \hat{\beta}=\beta_0} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} = \begin{pmatrix} \frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} \\ \frac{\partial \ln L(\alpha, \beta)}{\partial \beta} \end{pmatrix}_{\hat{\alpha}=\alpha_0, \hat{\beta}=\beta_0},$$

where α_0 and β_0 are initial value of α and β respectively. The initial values of the parameters taken in this paper for estimating parameters are $\alpha_0 = 0.5$ and $\beta_0 = 0.5$.

8. Estimation of the Stress-Strength parameter $R = P(X > Y)$

In Reliability, the Stress-Strength model describes the life of a component which has a random Strength X subjected to a 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 $X > Y$. In this section our objective is to estimate $R = P(X > Y)$ when $X \sim G\text{-SD}(\alpha_1, \beta_1)$ and $Y \sim G\text{-SD}(\alpha_2, \beta_2)$ and X and Y are independently distributed. The, the Stress- Strength Parameter is given by

$$\begin{aligned}
 R = P(X > Y) &= \int_0^\infty P(X > Y | Y = y) f_Y(y) dy \\
 &= \int_0^\infty [1 - F_X(y)] f_Y(y) dy \\
 &= 1 - \int_0^\infty \frac{y^{\alpha_1} \left[y(\beta_1^2 + 1) + (1 + \alpha_1 + \beta_1^2)\beta_1 \right] \alpha_2 \beta_2^2 (1 + \alpha_2 + \beta_2 y + \beta_2^2) y^{\alpha_2 - 1}}{(\beta_1^2 + 1)(\beta_1 + y)^{\alpha_1 + 1} (\beta_2^2 + 1)(\beta_2 + y)^{\alpha_2 + 2}} dy \\
 &= 1 - \int_0^\infty \frac{\alpha_2 \beta_2^2}{(\beta_1^2 + 1)(\beta_2^2 + 1)} \\
 &\quad \times \frac{y^{\alpha_1 + \alpha_2 - 1} \left[y(\beta_1^2 + 1) + (1 + \alpha_1 + \beta_1^2)\beta_1 \right] (1 + \alpha_2 + \beta_2 y + \beta_2^2)}{(\beta_1 + y)^{\alpha_1 + 1} (\beta_2 + y)^{\alpha_2 + 2}} dy \\
 &= H(\alpha_1, \alpha_2, \beta_1, \beta_2)
 \end{aligned} \tag{21}$$

Let, (x_1, x_2, \dots, x_n) be the observed value of a random sample of size n from $G\text{-SD}(\alpha_1, \beta_1)$ and (y_1, y_2, \dots, y_m) be the observed value of a random sample of size m from $G\text{-SD}(\alpha_2, \beta_2)$.

The log-likelihood function of $\alpha_1, \alpha_2, \beta_1$ and β_2 is given by

$$\begin{aligned}
 \ln L(\alpha_1, \alpha_2, \beta_1, \beta_2) &= n \ln(\alpha_1) + 2n \ln(\beta_1) - n \ln(\beta_1^2 + 1) + \sum_{i=1}^n \ln(1 + \alpha_1 + \beta_1 x_i + \beta_1^2) \\
 &\quad + (\alpha_1 - 1) \sum_{i=1}^n \ln(x_i) - (\alpha_1 + 2) \sum_{i=1}^n \ln(\beta_1 + x_i) + m \ln(\alpha_2) + 2m \ln(\beta_2) - m \ln(\beta_2^2 + 1) \\
 &\quad + \sum_{i=1}^m \ln(1 + \alpha_2 + \beta_2 y_i + \beta_2^2) + (\alpha_2 - 1) \sum_{i=1}^m \ln(y_i) - (\alpha_2 + 2) \sum_{i=1}^m \ln(\beta_2 + y_i)
 \end{aligned}$$

The maximum likelihood estimates of $\alpha_1, \alpha_2, \beta_1$ and β_2 are the solutions of following log-likelihood equations

$$\begin{aligned}
 \frac{\partial}{\partial \alpha_1} (\ln L(\alpha_1, \alpha_2, \beta_1, \beta_2)) &= \frac{n}{\alpha_1} + \sum_{i=1}^n \frac{1}{(1 + \alpha_1 + \beta_1 x_i + \beta_1^2)} + \sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \ln(\beta_1 + x_i) = 0 \\
 \frac{\partial}{\partial \alpha_2} (\ln L(\alpha_1, \alpha_2, \beta_1, \beta_2)) &= \frac{m}{\alpha_2} + \sum_{i=1}^m \frac{1}{(1 + \alpha_2 + \beta_2 y_i + \beta_2^2)} + \sum_{i=1}^m \ln(y_i) - \sum_{i=1}^m \ln(\beta_2 + y_i) = 0 \\
 \frac{\partial}{\partial \beta_1} (\ln L(\alpha_1, \alpha_2, \beta_1, \beta_2)) &= \frac{2n}{\beta_1} - \frac{2n\beta_1}{(\beta_1^2 + 1)} + \sum_{i=1}^n \frac{x_i}{(1 + \alpha_1 + \beta_1 x_i + \beta_1^2)} - (\alpha_1 + 2) \sum_{i=1}^n \frac{1}{(\beta_1 + x_i)} = 0 \\
 \frac{\partial}{\partial \beta_2} (\ln L(\alpha_1, \alpha_2, \beta_1, \beta_2)) &= \frac{2m}{\beta_2} - \frac{2m\beta_2}{(\beta_2^2 + 1)} + \sum_{i=1}^m \frac{y_i}{(1 + \alpha_2 + \beta_2 y_i + \beta_2^2)} - (\alpha_2 + 2) \sum_{i=1}^m \frac{1}{(\beta_2 + y_i)} = 0
 \end{aligned}$$

Solving these non-linear equations using any iterative methods available in R packages we can obtain the MLEs of the parameters as $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2)$ and hence the MLE of R can thus be obtained as

$$\hat{R} = H(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2) \tag{22}$$

9. Applications

In this section, we present the goodness of fit of the G-SD to two real lifetime datasets to illustrate its applications. The goodness of fit of G-LD, Weibull, gamma, Shanker, Lindley and exponential have also been given for ready comparison. The datasets that we considered to demonstrate the applications of the proposed distribution are as follows:

Dataset 1: The dataset consists of plasma concentrations of indomethacin (mcg/ml) given by Team R.C. (2014) [11] are

1.50, 0.94, 0.78, 0.48, 0.37, 0.19, 0.12, 0.11, 0.08, 0.07, 0.05, 2.03, 1.63, 0.71, 0.70, 0.64, 0.36, 0.32, 0.20, 0.25, 0.12, 0.08, 2.72, 1.49, 1.16, 0.80, 0.80, 0.39, 0.22, 0.12, 0.11, 0.08, 0.08, 1.85, 1.39, 1.02, 0.89, 0.59, 0.40, 0.16, 0.11, 0.10, 0.07, 0.07, 2.05, 1.04, 0.81, 0.39, 0.30, 0.23, 0.13, 0.11, 0.08, 0.10, 0.06, 2.31, 1.44, 1.03, 0.84, 0.64, 0.42, 0.24, 0.17, 0.13, 0.10, 0.09

Dataset 2: The dataset is the survival times (in days) of 73 patients who diagnosed with acute bone cancer [12], available in <https://doi.org/10.22436/jnsa.013.05.01>

0.09, 0.76, 1.81, 1.10, 3.72, 0.72, 2.49, 1.00, 0.53, 0.66, 31.61, 0.60, 0.20, 1.61, 1.88, 0.70, 1.36, 0.43, 3.16, 1.57, 4.93, 11.07, 1.63, 1.39, 4.54, 3.12, 86.01, 1.92, 0.92, 4.04, 1.16, 2.26, 0.20, 0.94, 1.82, 3.99, 1.46, 2.75, 1.38, 2.76, 1.86, 2.68, 1.76, 0.67, 1.29, 1.56, 2.83, 0.71, 1.48, 2.41, 0.66, 0.65, 2.36, 1.29, 13.75, 0.67, 3.70, 0.76, 3.63, 0.68, 2.65, 0.95, 2.30, 2.57, 0.61, 3.93, 1.56, 1.29, 9.94, 1.67, 1.42, 4.18, 1.37

In order to compare lifetime distributions, values of $-2\log L$, AIC (Akaike information criterion), BIC (Bayesian information criterion), Kolmogorov – Smirnov (K-S) statistics with their P- values for the considered datasets has been computed. The formulae for computing AIC, AICC, BIC and K-S Statistics are as follows:

$$AIC = -2\log L + 2k, \quad AICC = AIC + \frac{2k(k+1)}{n-k-1}, \quad BIC = -2\log L + k \log n,$$

$$D = \sup_x |F_n(x) - F_0(x)| \text{ where } k = \text{number of parameter, } n = \text{sample size.}$$

The distribution corresponding to the lower values of $-2\log L$, AIC and K-S is the best fit distribution. The standard errors of estimate of parameters are given in the parenthesis along with the ML estimates.

Table1: ML estimates, $-2\log L$, AIC, BIC and K-S statistics with their P-values of the distributions for first data set 1

Distributions	ML estimates $\hat{\alpha}$ (S.E) $\hat{\beta}$ (S.E)	$-2\log L$	AIC	BIC	K-S	P-value
G-SD	2.2087 (0.6607) 0.2677 (0.0986)	60.72	64.72	65.51	0.15	0.09
G-LD	2.5615 (0.8798) 0.2034(0.0837)	61.09	65.09	65.88	0.19	0.03
Weibull	1.6857 (0.2078) 0.9545 (0.0903)	62.51	66.51	67.31	0.42	0.00
Gamma	1.6513 (0.3257) 0.9772 (0.1495)	62.74	66.74	67.53	0.41	0.00
Shanker	2.0294 (0.1916)	63.27	67.27	68.06	0.17	0.08
Lindley	2.2152 (0.2208)	64.28	66.28	66.67	0.17	0.08
Exponential	1.6897 (0.2080)	62.76	64.76	65.15	0.17	0.15

Table 2: ML estimates, $-2\log L$, AIC, BIC and K-S statistics with their P-values of the distributions for second data set

Distributions	ML estimates $\hat{\alpha}$ (S.E) $\hat{\beta}$ (S.E)	$-2\log L$	AIC	BIC	K-S	P-value
G-SD	4.8969 (1.3904) 0.4968 (0.1360)	282.82	286.81	296.62	0.09	0.63
G-LD	5.1601 (1.8468) 0.4376 (0.1602)	284.31	288.31	298.13	0.13	0.23
Weibull	0.4395 (0.0687) 0.7656 (0.0568)	322.80	326.80	336.62	0.31	0.00
Gamma	0.1985 (0.0389) 0.7457 (0.1058)	334.53	338.53	348.35	0.78	0.00
Shanker	1.9473 (0.2707)	310.45	312.45	317.36	0.46	0.00
Lindley	0.4499 (0.0382)	374.77	376.77	381.67	0.31	0.00
Exponential	0.2663 (0.0312)	339.18	341.18	346.09	0.19	0.08

Based on the values of $-2\log L$, AIC, BIC and K-S statistics with their P-values it is obvious from tables 1 and 2 shows that among all considered distributions, gamma Shanker distribution (G-SD) give much closer fit.

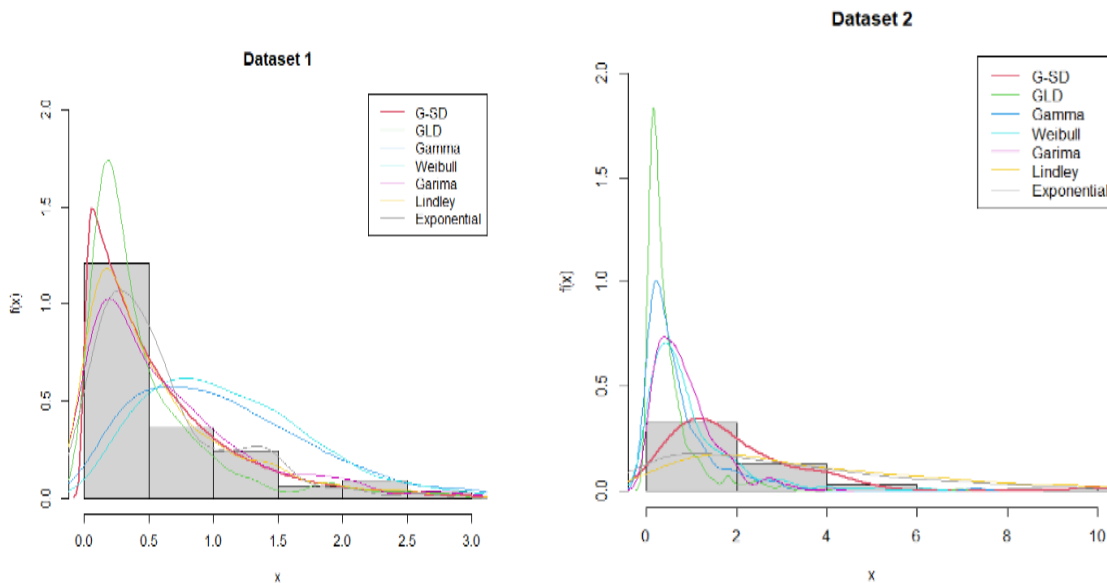


Figure 5: Fitted Plots for Dataset 1 and 2

In dataset 2 there are four values which act as an outlier, so for better graphical representation we have exclude the values from the dataset.

10. Concluding Remarks

In this paper, we propose a gamma-Shanker distribution by compounding gamma and Shanker distribution. Its statistical properties including shapes of hazard rate and reversed hazard rate

function, Quantile, negative moments, stochastic ordering, and Stress-Strength reliability have been discussed. Maximum Likelihood estimation has been discussed for estimating its parameter. The goodness of fit of G-SD over G-LD distribution, Weibull distribution, gamma distribution, Shanker distribution, Lindley distribution and exponential distribution shows that G-SD gives much closure fit than these distributions for the considered datasets.

Conflict of Interest

The Authors declare that there is no conflict of Interest.

References

- [1] Lindley, D.V. (1958), Fiducial Distribution and Bayes' Theorem, *Journal of The Royal Statistical Society*, 20(1), 102-107.
- [2] Shanker, R., Hagos, F. and Sujatha, S. (2015). On Modeling of Lifetimes data using Exponential and Lindley distributions, *Biometrics & Biostatistics International Journal*, 2(5), 140 -147
- [3] Shanker, R., Shukla, K.K., Shanker, R. and Tekie, A.L. (2016). On modelling of Lifetime data using two-parameter gamma and Weibull distributions, *Biometrics & Biostatistics International journal*, 4(5), 1 – 6.
- [4] Abdi, M., Asgharzadeh, A., Bakouch, H.S., Alipour, Z. (2019). A new compound Gamma and Lindley distribution with Application to failure data, *Austrian Journal of Statistics*, 48 ,54-75.
- [5] Finkelstein, M. (2008). Failure Rate Modeling for Reliability and Risk, *Springer series in Reliability Engineering*.
- [6] Proschan, F. (1963). Theoretical Explanation of Observed Decreasing Failure Rate, *Technometrics*,5(3), 375 – 383.
- [7] Lie, C.D. and Xie, M. (2006). Stochastic Ageing and Dependence for Reliability, *Springer, New York*.
- [8] Demicheli, R., Bonadonna, G., Hrushesky, W. J., Retsky, M.W., and Valgussa, P. (2004). Menopausal Status Dependence of the timing of Breast cancer Recurrence after Surgical Removal of the primary Tumour, *Breast Cancer Research*, 6(6), 689 – 696.
- [9] Shanker, R. (2015). Shanker Distribution and its Applications, *International Journal of Statistics and Applications*,5(6),338-348.
- [10] Shaked, M. and Shanthikumar, J.G. (1994). Stochastic Orders and Their Applications, *Academic Press New Work*.
- [11] Team, R.C. and R. (2014). A language and Environment for Statistical Computing Version 3.1.2, *R Foundation for Statistical Computing*.
- [12] Mansour, M. M., Yousof, H. M., Shehata, W. A. M., Ibrahim, M. (2020). A New Two Parameter Burr XII Distribution Properties Copula Different Estimation Methods and Modeling Acute Bone Cancer Data, *Journal of Nonlinear Science and Applications*, 13, 223-238.

DIFFERENT ESTIMATION METHODS FOR TOPP-LEONE-(A) MODEL: INFERENCE AND APPLICATIONS TO COMPLETE AND CENSORED DATASETS

ADUBISI O. D.¹

ADUBISI C. E.²



¹Department of Mathematics and Statistics, Federal University Wukari, Nigeria.

²Department of Physics, University of Ilorin, Nigeria.

adubisiobinna@fuwukari.edu.ng¹

chidiadubisi@gmail.com²

Abstract

This work proposes a new two-parameter model, titled Topp-Leone (A) model. The main benefit of the new model is that it has an inverted bathtub shaped curve, increasing and decreasing hazard rate function quite dependent on the shape parameter. Its structural properties including the ordinary moments, quantiles, probability weighted moment, median, entropy and order statistics are derived. More so, the survival, failure rate, reversed failure rate and cumulative failure rate functions are also derived. Six classical estimation methods are discussed for estimating the parameters of the new model. Monte Carlo experiments and real datasets analyses are conducted to examine the classical estimators performance of this model. Finally, the usefulness of the Topp-Leone (A) model demonstrated with different applications to complete and type-II right censored data proves its more flexible when compared to well-known models in statistical literature.

Keywords: (A)-model, Censoring, Estimation methods, Simulation, Topp-Leone-G class

1. INTRODUCTION

In the past decades, classical models have been utilized extensively for analysing various datasets in the fields of demography, engineering, finance, medical and social sciences, environmental science, biological and actuarial studies. In many workable circumstances the classical models do not give a sufficient fit to actual datasets. Therefore, various generalizations and extensions of the classical models have been proposed and studied. For example, inverse Gompertz model was pioneered by [1], odd Fréchet inverse exponential model was studied by [2], Kumaraswamy inverse Gompertz model was introduced by [3], Odd exponentiated half-logistic exponential model was studied by [4], Pareto exponential model was proposed by [5], odd exponentiated skew-t model was pioneered by [6], type-I half logistic skew-t model studied by [7], exponentiated half logistic skew-t model was introduced by [8], exponentiated odd lomax exponential model was studied by [9] and polynomial exponential model was studied by [10], among others. [11] proposed the (A) model having just a scale parameter which makes it unsuitable for modeling most real life circumstances, hence the need to extend the (A) model to increase its flexibility and capability. The novelty and input made by this study is the creation of a new two-parameter model

known as the Topp-Leone-(A) (TL_(A)) model reliant on Eqs (3) and (4). The principal focus in this work are: utilizing the TL-G class to improve the structural properties and flexibility of the (A) model, provide a new generalized version of the (A) model with a closed-form quantile function, investigate the important descriptive aspects of the TL_(A) model, such as the mode, median, mean, variance (VAR), skewness (SK), kurtosis (KU), moments, moment generating function, entropy, probability weighted moment and order statistics, investigate the statistical inference of the TL_(A) model using six different methods such as the maximum likelihood estimation (MLE), maximum product spacing estimation (MPS), Anderson Darling estimation (ADE) least square estimation (LSE), and weighted least square estimation (WLSE), Cramer Von Mises estimation (CVME) for complete datasets, and provide better fits than competing generalized statistical models and also the suitability for testing the goodness of fit of the TL_(A) to its sub-model, the (A) model. Suppose that Z is a random variable, the cumulative distribution function (CDF) and probability density function (PDF) of the (A) model with scale parameter $\kappa > 0$ are respectively, given by

$$G(z) = e^{-\frac{1}{\kappa}\left(e^{\frac{\kappa}{z}}-1\right)}; z > 0; \kappa > 0, \tag{1}$$

and

$$g(z) = \frac{1}{z^2}e^{\frac{\kappa}{z}-\frac{1}{\kappa}\left(e^{\frac{\kappa}{z}}-1\right)}; z > 0; \kappa > 0. \tag{2}$$

Recently, the Topp-Leone-G (TL-G) family is one essential generator that have increased the interest of researchers in distribution theory. [12] introduced the CDF of the TL-G family as

$$F(z) = \left\{1 - [1 - G(z)]^2\right\}^\eta, \tag{3}$$

and the corresponding PDF to Eq (3) takes the form

$$f(z) = 2\eta g(z) [1 - G(z)] \left\{1 - [1 - G(z)]^2\right\}^{\eta-1}. \tag{4}$$

where $\eta > 0$ is a shape parameter, G(z) and g(z) are considered as the CDF and PDF of a baseline r.v Z.

The remaining parts are outlined as follows: Part 2 introduces the CDF and PDF of the TL_(A) model. Part 3 presents several fundamental structural properties of the TL_(A) model. Some essential functions used in reliability analysis are introduced in Part 4. The six classical estimation approaches are discussed in Part 5 to appreciate the parameters of the TL_(A). The maximum likelihood estimator of TL_(A) for the type-II right censored are presented in Part 6. The performance of TL_(A) estimators is appreciated in Part 7 using Monte Carlo experiments. Three real datasets; two complete and one type-II right censored data are analysed and the empirical results presented in Part 8. Finally, in Part 9, discussions and conclusion are presented.

2. TOPP-LEONE-(A) MODEL

2.1. Genesis of TL_(A)

The non-negative r.v Z is said to have the TL_(A) model with parameters vector $\Psi = (\kappa, \eta)$, say $Z \sim TL_{(A)}(\Psi)$. The CDF of TL_(A) model takes the form

$$F(z) = \left\{1 - \left[1 - e^{-\frac{1}{\kappa}\left(e^{\frac{\kappa}{z}}-1\right)}\right]^2\right\}^\eta, \tag{5}$$

and the corresponding PDF to Eq (5) takes the form

$$f(z) = 2\eta z^{-2}e^{\frac{\kappa}{z}-\frac{1}{\kappa}\left(e^{\frac{\kappa}{z}}-1\right)} \left[1 - e^{-\frac{1}{\kappa}\left(e^{\frac{\kappa}{z}}-1\right)}\right] \left\{1 - \left[1 - e^{-\frac{1}{\kappa}\left(e^{\frac{\kappa}{z}}-1\right)}\right]^2\right\}^{\eta-1}. \tag{6}$$

where $\eta > 0, \kappa > 0$ are the shape and scale parameters, and $z \in \mathfrak{R}^+$. Figure 1 depicts the graphical shapes of the TL_(A) PDF with selected values for the parameters η and κ . The PDF is uni-modal, increasing-decreasing, right-skewness, decreasing, and heavy-tailed. The failure rate function of the new model in Figure 3 takes the form of "an inverted bathtub shaped, increasing and decreasing".

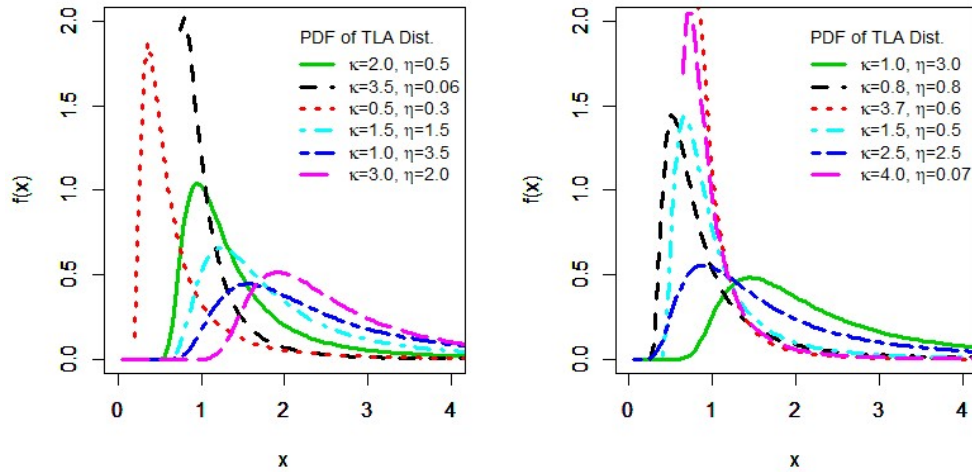


Figure 1: The density function (PDF) plots of the TL_(A) model.

3. STRUCTURAL PROPERTIES OF TL_(A) MODEL

This part inspects some fundamental structural properties of the TL_(A) model.

3.1. Quantiles function

The explicit forms of the quantile and median functions for the TL_(A) model are presented in this subpart. The quantile function found by inverting Eq (5) takes the form

$$z_u = \frac{\kappa}{\log \left\{ 1 - \kappa \log \left[1 - \left(1 - u^{\frac{1}{\eta}} \right)^{\frac{1}{2}} \right] \right\}}; \quad 0 < u < 1, \tag{7}$$

By setting $u = \frac{1}{2}$ in Eq (7), the median (M) function takes the form

$$M = \frac{\kappa}{\log \left\{ 1 - \kappa \log \left[1 - \left(1 - 0.5^{\frac{1}{\eta}} \right)^{\frac{1}{2}} \right] \right\}}. \tag{8}$$

3.2. Moments and moment generating function

If $Z \sim \text{TL}_{(A)}(\Psi)$, then the rth ordinary moment (OM) of Z is found using

$$\mu'_r = E(Z^r) = \int_0^\infty Z^r f(z; \Psi) dz \tag{9}$$

By substituting Eq (6) into Eq (9), expanding using the Taylor series and invoking the beta function. The OM of the TL_(A) model takes the form

$$\mu'_r = \sum_{i=0}^{\eta-1} \vartheta_i \Gamma(h-r+1), \tag{10}$$

where

$$\vartheta_i = 2\eta \sum_{j,g=0}^{\infty} \sum_{l=0}^g \sum_{h=0}^{\infty} \binom{\eta-1}{i} \binom{2i+1}{j} \frac{(-1)^{g+i+j+l+1} (1+j)^g (1+g)^h \kappa^{r-g-1} l^{r-h-1}}{l! h! (g-l)!}$$

Similarly, the moment generating function (MGF) of TL_(A) model, say $M_Z(t)$ is found using

$$M_Z(t) = E(e^{tz}) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^{\infty} Z^r f(z; \Psi) dz = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu'_r. \tag{11}$$

By substituting Eq (10) into Eq (11), the MGF takes the form

$$M_Z(t) = \sum_{r=0}^{\infty} \vartheta_r \Gamma(h-r+1), \tag{12}$$

where

$$\vartheta_r = 2\eta \sum_{i=0}^{\eta-1} \sum_{j,g=0}^{\infty} \sum_{l=0}^g \sum_{h=0}^{\infty} \binom{\eta-1}{i} \binom{2i+1}{j} \frac{(-1)^{g+i+j+l+1} (1+j)^g (1+g)^h \kappa^{r-g-1} l^{r-h-1}}{r! l! h! (g-l)!}$$

3.3. Entropy measures

Entropy performs a crucial part in computer science, information theory, probability theory and engineering. It is considered as a measure of dispersion for the uncertainty associated with a random variable Z ; see [13]. The Rényi entropy of Z , say $I_{\tau}(Z)$ is given by

$$I_{\tau}(Z) = \frac{1}{1-\tau} \log \int_0^{\infty} f^{\tau}(z) dz; \tau > 0 \text{ and } \tau \neq 1, \tag{13}$$

If $Z \sim \text{TL}_{(A)}(\Psi)$, substituting Eq (6) into Eq (13), expanding using the Taylor series and invoking the beta function. The $I_{\tau}(Z)$ takes the form

$$I_{\tau}(z) = \frac{1}{1-\tau} \log \left[\sum_{i=0}^{\tau(\eta-1)} \vartheta_i^* \Gamma(2i+h-1) \right], \tag{14}$$

where

$$\vartheta_i^* = (2\eta)^{\tau} \sum_{j,g=0}^{\infty} \sum_{l=0}^g \sum_{h=0}^{\infty} \binom{\tau(\eta-1)}{i} \binom{2i+1}{j} \frac{(-1)^{g+i+j+l+1} (\tau+j)^g (\tau+g)^h \kappa^{-2\tau-g-1} l^{-2\tau-h-1}}{l! h! (g-l)!}$$

3.4. Probability weighted moment

According to [14], the probability weighted moment (PWM) is a very useful quantity in mathematical statistics. The PWM of Z , say $\zeta_{r,s}$ is given by

$$\zeta_{r,s} = E[z^r F^s(z)] = \int_0^{\infty} z^r F^s(z) f(z) dz, \tag{15}$$

If $Z \sim \text{TL}_{(A)}(\Psi)$, substituting Eqs (5) and (6) into Eq (15), expanding using the Taylor series and invoking the beta function. The PWM $\zeta_{r,s}$ takes the form

$$\zeta_{s,r} = \sum_{a=0}^{\infty} \vartheta_a \Gamma(h-s+1), \tag{16}$$

where

$$\vartheta_a = 2\eta \sum_{b=0}^{\eta(a+1)-1} \sum_{c,i=0}^{\infty} \sum_{g=0}^i \sum_{h=0}^{\infty} \binom{r}{a} \binom{\eta(a+1)-1}{b} \binom{2b+1}{c} \times \frac{(-1)^{b+c+i+g+1} (1+c)^i (1+i)^h \kappa^{s-i-1} g^{s-h-1}}{g! h! (i-g)!}$$

3.5. Order statistics

Let z_1, z_2, \dots, z_n be a random sample from a continuous distribution, and the sequence $z_{1:n} < z_{2:n} < \dots < z_{n:n}$ are order statistics (O.S) obtained from the sample. According to [15], the p^{th} O.S is given by

$$f_{p:N}(z) = \frac{g(z)}{B(p, N-p+1)} [G(z)]^{p-1} [1-G(z)]^{N-p}, \quad (17)$$

where $G(z)$ and $g(z)$ are the CDF and PDF of the TL_(A) model, and $B(.,.)$ is the beta function. Expanding $[1-G(z)]^{N-p}$, the O.S takes the form

$$f_{p:N}(z) = \frac{1}{B(p, N-p+1)} \sum_{l=0}^{N-p} (-1)^l \binom{N-p}{l} [F(z)]^{p+l-1} f(z), \quad (18)$$

By substituting Eqs (5) and (6) into Eq (18), and then expanding. The O.S takes the form

$$f_{p:N}(z) = \frac{2\eta}{B(p, N-p+1)} \sum_{l=0}^{N-p} \vartheta_l \frac{e^{\frac{\kappa(1+g)}{z}}}{z^2}, \quad (19)$$

where

$$\vartheta_l = \sum_{i=0}^{\eta(p+l)-1} \sum_{j,g=0}^{\infty} \sum_{h=0}^g \binom{N-p}{l} \binom{\eta(p+l)-1}{b} \binom{2i+1}{j} \frac{(-1)^{g+h+i+j+l} (1+j)^g \kappa^{-g}}{h!(g-h)!}$$

3.6. Skewness, kurtosis, dispersion index and coefficient of variation

The quantile function of the TL_(A) presented in Eq (7) can be utilized in investigating the effect of the shape parameter on the mean (ME), variance (VAR), standard deviation (STD), median (M), skewness (S_k), kurtosis (K_u), dispersion index (DI) and coefficient of variation (CV). [16] proposed the skewness computational method using the quartiles, titled Bowley skewness. It is expressed as

$$S_k = \frac{Q\left(\frac{3}{4}; \Psi\right) - 2Q\left(\frac{1}{2}; \Psi\right) + Q\left(\frac{1}{4}; \Psi\right)}{Q\left(\frac{3}{4}; \Psi\right) - Q\left(\frac{1}{4}; \Psi\right)} \quad (20)$$

Likewise, [17] introduced the kurtosis computational method based on the octiles, titled Moor's kurtosis. It is expressed as

$$K_u = \frac{Q\left(\frac{7}{8}; \Psi\right) - Q\left(\frac{5}{8}; \Psi\right) - Q\left(\frac{3}{8}; \Psi\right) + Q\left(\frac{1}{8}; \Psi\right)}{Q\left(\frac{6}{8}; \Psi\right) - Q\left(\frac{2}{8}; \Psi\right)} \quad (21)$$

The DI shows whether a model is suitable for modeling equi, under or over-dispersed datasets. More so, a distribution is considered equi-dispersed if $DI = 1$, under-dispersed if $DI < 1$ and over-dispersed if $DI > 1$. The DI is expressed as

$$DI = \frac{Var(X)}{E(X)} = \frac{\frac{Q\left(\frac{3}{4}; \Psi\right) - Q\left(\frac{1}{4}; \Psi\right)}{1.35}}{\frac{Q\left(\frac{3}{4}; \Psi\right) + Q\left(\frac{1}{2}; \Psi\right) + Q\left(\frac{1}{4}; \Psi\right)}{3}}. \quad (22)$$

The CV is a relative measure of variability and generally utilized to compare independent samples based on their variability. A large CV value indicates a higher variability. The CV is expressed as

$$CV = \frac{(Var(X))^{\frac{1}{2}}}{E(X)} = \frac{\left(\frac{Q\left(\frac{3}{4}; \Psi\right) - Q\left(\frac{1}{4}; \Psi\right)}{1.35}\right)^{\frac{1}{2}}}{\frac{Q\left(\frac{3}{4}; \Psi\right) + Q\left(\frac{1}{2}; \Psi\right) + Q\left(\frac{1}{4}; \Psi\right)}{3}}. \quad (23)$$

where $Q(\cdot)$ is the quantile function. The numerical values of the descriptive measures for the TL_(A) model under selected values of η and κ are reported in Table 1. The following conclusions are reached:

1. The mean and standard deviation of the TL_(A) model increases as the values of κ and η increases. From the reported numerical values of the skewness and kurtosis, we can conclude that the TL_(A) model is positively skewed. Also, as the values of κ and η are increased the skewness and kurtosis values decreases.
2. The TL_(A) model is beneficial for under-dispersed datasets while the DI increases and the CV decreases, as the values of κ and η increases.

Table 1: The descriptive measures for the TL_(A) model.

Parameters		Descriptive measures							
η	κ	ME	VAR	STD	M	S _k	K _u	DI	CV
0.2	0.2	0.375	0.072	0.268	0.331	0.365	1.140	0.192	0.715
0.5	0.2	0.664	0.224	0.473	0.592	0.340	1.055	0.337	0.712
1.0	0.5	1.148	0.507	0.712	1.044	0.322	0.997	0.442	0.620
2.0	0.5	1.667	1.061	1.030	1.520	0.317	0.978	0.636	0.618
2.5	1.0	2.097	1.378	1.174	1.933	0.310	0.959	0.657	0.560
3.0	1.5	2.501	1.708	1.307	2.321	0.306	0.943	0.683	0.523
3.5	1.5	2.683	1.991	1.411	2.489	0.306	0.944	0.742	0.526
3.5	2.0	2.885	2.048	1.431	2.691	0.302	0.931	0.710	0.496
4.5	2.5	3.417	2.683	1.638	3.196	0.299	0.923	0.785	0.479
5.0	3.0	3.765	3.042	1.744	3.533	0.297	0.915	0.808	0.463
6.5	3.5	4.383	3.988	1.997	4.117	0.296	0.912	0.910	0.456
7.0	4.0	4.705	4.360	2.088	4.429	0.294	0.906	0.927	0.444

Fig 2 depicts the 3D plots of the mean, variance, skewness and kurtosis of the TL_(A) for some values of η and κ parameters. The plots in figure 2 reveal that as the values of η and κ increases, the skewness and kurtosis values decrease, and the mean and variance values increase, respectively.

4. RELIABILITY ANALYSIS

4.1. Survival and failure rate functions

The survival function (Reliability) of $Z \sim \text{TL}_{(A)}(\Psi)$, takes the form

$$R(z) = 1 - \left\{ 1 - \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right]^2 \right\}^\eta; \eta, \kappa > 0. \tag{24}$$

The failure (hazard) rate function (HRF) of $Z \sim \text{TL}_{(A)}(\Psi)$, takes the form

$$h(z) = 2\eta z^{-2} e^{\frac{\kappa}{z} - \frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right] \left\{ 1 - \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right]^2 \right\}^{\eta-1} \\ \times \left(1 - \left\{ 1 - \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right]^2 \right\}^\eta \right)^{-1}. \tag{25}$$

More so, if $Z \sim \text{TL}_{(A)}(\Psi)$, then the reversed HRF takes the form

$$r(z) = 2\eta z^{-2} e^{\frac{\kappa}{z} - \frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right] \left\{ 1 - \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right]^2 \right\}^{\eta-1} \\ \times \left(\left\{ 1 - \left[1 - e^{-\frac{1}{\kappa}(e^{\frac{\kappa}{z}} - 1)} \right]^2 \right\}^\eta \right)^{-1}. \tag{26}$$

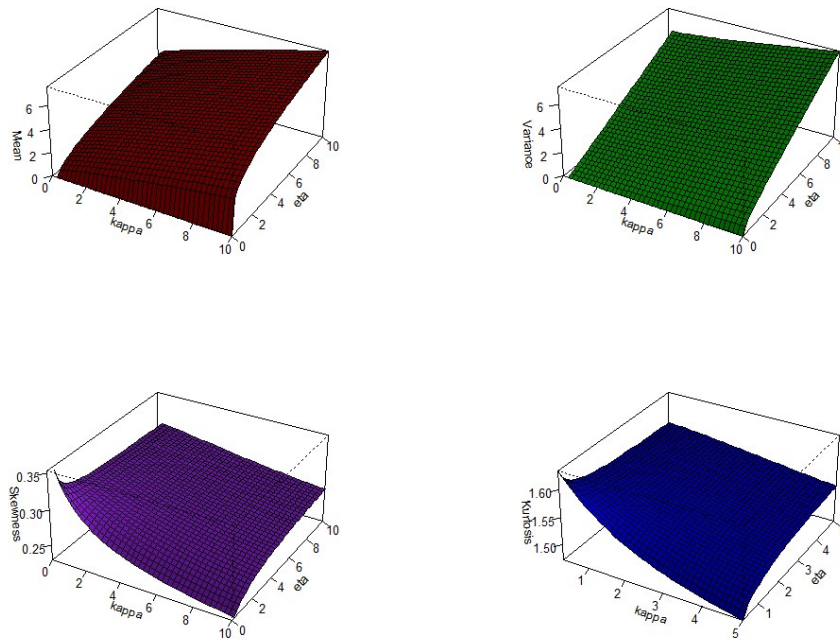


Figure 2: The Mean, Variance, skewness and kurtosis plots of the TL_(A) model.

and the cumulative HRF takes the form

$$H(z) = -\log \left(1 - \left\{ 1 - \left[1 - e^{-\frac{1}{\kappa} \left(e^{\frac{\kappa}{2} z} - 1 \right)} \right]^2 \right\}^\eta \right). \quad (27)$$

The graphical shapes of the HRF for TL_(A) with various selected values of η and κ are depicted in Fig 3. The model is characterized by an inverted bathtub shaped curve, increasing and decreasing HRF.

5. ESTIMATION METHODS

This part discusses estimating the TL_(A) parameters via different estimation methods. The method of maximum likelihood (MLE), method of maximum product of spacing (MPS), methods of ordinary least squares (OLS) and weighted least squares (WLS), method of Cramer-Von Mises (CVM) and method of Anderson Darling (ANDA) are considered for the complete data.

5.1. The MLE

The maximum likelihood (ML) method for estimating the unknown parameters of TL_(A)(Ψ) for complete samples is considered. Let z_1, z_2, \dots, z_s be the random observed values of size (s) from TL_(A)(Ψ). Hence, the log-likelihood function $L(\Psi)$ of Eq (6) takes the form

$$L(\Psi) = s \log(2\eta) + 2 \sum_{j=1}^s \log z_j + \kappa \sum_{j=1}^s \frac{1}{z_j} - \frac{1}{\kappa} \sum_{j=1}^s v_j + \sum_{j=1}^s \log \left(1 - e^{-\frac{1}{\kappa} v_j} \right) + (\eta - 1) \sum_{j=1}^s \log \left[1 - \left(1 - e^{-\frac{1}{\kappa} v_j} \right)^2 \right], \quad (28)$$

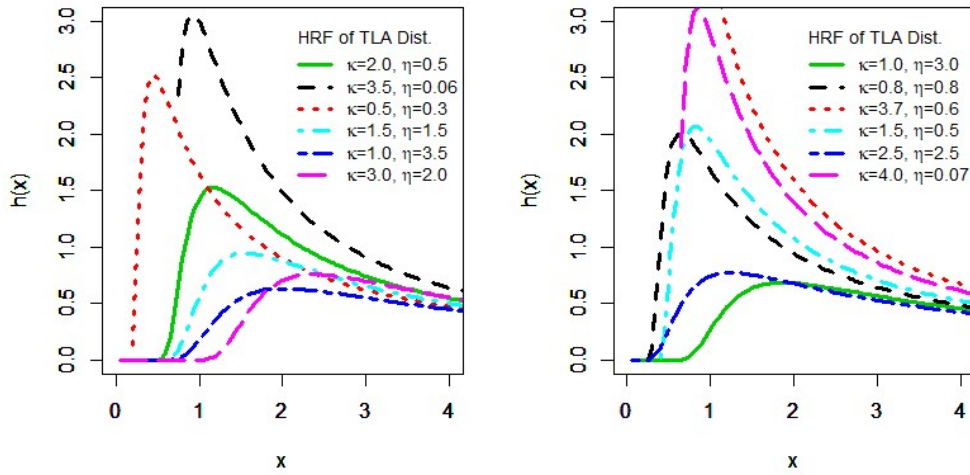


Figure 3: The HRF plots of the TL_(A) model.

where $v_j = e^{\frac{\kappa}{z_j}} - 1$. By deriving the first partial derivatives of $L(\Psi)$ and equating to zero. The associated score function $U(\Psi) = \left(\frac{\partial[L(\Psi)]}{\partial\eta}, \frac{\partial[L(\Psi)]}{\partial\kappa} \right)^T = 0$ are given by

$$U_\eta(\Psi) = \frac{s}{\eta} + \sum_{j=1}^s \log \left[1 - \left(1 - e^{-\frac{1}{\kappa}v_j} \right)^2 \right] = 0, \quad (29)$$

and

$$U_\kappa(\Psi) = \sum_{j=1}^s \left(\frac{1}{z_j} \right) + \frac{1}{\kappa^2} \sum_{j=1}^s (v_j) - \frac{1}{\kappa} \sum_{j=1}^s \left(\frac{e^{\frac{\kappa}{z_j}}}{z_j} \right) - \sum_{j=1}^s (\zeta_j) + 2(\eta - 1) \sum_{j=1}^s \frac{\left(1 - e^{-\frac{1}{\kappa}v_j} \right) \left(\frac{1}{\kappa^2} v_j - \frac{1}{\kappa z_j} e^{\frac{\kappa}{z_j}} \right) e^{-\frac{1}{\kappa}v_j}}{1 - \left(1 - e^{-\frac{1}{\kappa}v_j} \right)^2} = 0. \quad (30)$$

where $\zeta_j = \frac{\left(\frac{v_j}{\kappa^2} - \frac{e^{\frac{\kappa}{z_j}}}{\kappa z_j} \right) e^{-\frac{v_j}{\kappa}}}{1 - e^{-\frac{v_j}{\kappa}}}$.

The ML estimates $\hat{\eta}_{ML}$ and $\hat{\kappa}_{ML}$ of the parameters of TL_(A)(Ψ) are found by maximizing Eq (??) using the (*Optim function*) in R-programming software.

5.2. The OLS and WLS

Let $z_{(1:s)}, z_{(2:s)}, \dots, z_{(s:s)}$ be the ordered sample of size (s) from CDF of the TL_(A)(Ψ) in Eq (5). The ordinary least squares (OLS) estimates $\hat{\eta}_{OLS}$ and $\hat{\kappa}_{OLS}$ can be found by minimizing with respect to η and κ , the function

$$OL(\eta, \kappa) = \sum_{j=1}^s \left[F \left(z_{(j)} \mid \eta, \kappa \right) - \zeta(j, s) \right]^2, \quad (31)$$

where $\zeta(j, s) = j / (s + 1)$. Equivalently, the OLS estimates can be found by solving the following differential equation

$$\sum_{j=1}^s \left[F \left(z_{(j)} \mid \eta, \kappa \right) - \xi(j, s) \right] \Delta_i \left(z_{(j)} \mid \eta, \kappa \right) = 0; \quad i = 1, 2, \quad (32)$$

where

$$\Delta_1 \left(z_{(j)} \mid \eta, \kappa \right) = \frac{\partial}{\partial \eta} F \left(z_{(j)} \mid \eta, \kappa \right), \quad \Delta_2 \left(z_{(j)} \mid \eta, \kappa \right) = \frac{\partial}{\partial \kappa} F \left(z_{(j)} \mid \eta, \kappa \right). \quad (33)$$

The solutions of Δ_i for $i = 1, 2$ can be found numerically. For more details, see [18].

Likewise, the weighted least squares (WLS) estimates $\hat{\eta}_{\text{WLS}}$ and $\hat{\kappa}_{\text{WLS}}$ can be found by minimizing with respect to η and κ , the function

$$WL(\eta, \kappa) = \sum_{j=1}^s \Phi(j, s) \sum_{j=1}^s \left[F \left(z_{(j)} \mid \eta, \kappa \right) - \xi(j, s) \right]^2, \quad (34)$$

where $\Phi(j, s) = (s + 1)^2 (s + 2) / j (s - j + 1)$. The WLS estimates can also be found by solving the following differential equation

$$\sum_{j=1}^s \Phi(j, s) \left[F \left(z_{(j)} \mid \eta, \kappa \right) - \xi(j, s) \right] \Delta_i \left(z_{(j)} \mid \eta, \kappa \right) = 0; \quad i = 1, 2, \quad (35)$$

where $\Delta_1(\cdot \mid \eta, \kappa)$ and $\Delta_2(\cdot \mid \eta, \kappa)$ are given in Eq (33).

5.3. The MPS

The maximum product spacing (MPS) estimator proposed by [19, 20], for the estimation of unknown parameters with ordered sample $z_{(1:s)}, z_{(2:s)}, \dots, z_{(s:s)}$ from $\text{TL}_{(A)}(\Psi)$, and the uniform spacing for this random sample is given by

$$D_j(\eta, \kappa) = F \left(z_{(j:s)} \mid \eta, \kappa \right) - F \left(z_{(j-1:s)} \mid \eta, \kappa \right); \quad j = 1, 2, \dots, T + 1, \quad (36)$$

where

$$F \left(z_{(0:s)} \mid \eta, \kappa \right) = 0, \quad F \left(z_{(s+1:s)} \mid \eta, \kappa \right) = 1.$$

$$\sum_{j=1}^{s+1} D_j(\eta, \kappa) = 1.$$

The MPS estimates $\hat{\eta}_{\text{MPS}}$ and $\hat{\kappa}_{\text{MPS}}$ can be found by maximizing the geometric mean (GM) of the spacing given by

$$GM(\eta, \kappa) = \left[\prod_{j=1}^{s+1} D_j(\eta, \kappa) \right]^{1/s+1}, \quad (37)$$

relative to η and κ or maximizing the logarithm of GM of the spacing given by

$$LGM(\eta, \kappa) = \frac{1}{s+1} \sum_{j=1}^{s+1} \log D_j(\eta, \kappa), \quad (38)$$

The MPS estimates $\hat{\eta}_{\text{MPS}}$ and $\hat{\kappa}_{\text{MPS}}$ of $\text{TL}_{(A)}(\Psi)$ can also be found by solving the following differential equation

$$\sum_{j=1}^{s+1} \frac{1}{D_j(\eta, \kappa)} \left[\Delta_i \left(z_{(j:s)} \mid \eta, \kappa \right) - \Delta_i \left(z_{(j-1:s)} \mid \eta, \kappa \right) \right] = 0; \quad i = 1, 2, \quad (39)$$

where $\Delta_1(\cdot \mid \eta, \kappa)$ and $\Delta_2(\cdot \mid \eta, \kappa)$ are given in Eq (33).

5.4. The ANDA

The ANDA estimates $\hat{\eta}_{\text{ANDA}}$ and $\hat{\kappa}_{\text{ANDA}}$ can be found for TL_(A)(Ψ) by minimizing the function

$$AD(\eta, \kappa) = -s - \frac{1}{s} \sum_{j=1}^s (2j-1) \left[\log F(y_{(j:s)} | \eta, \kappa) + \log \bar{F}(y_{(s+1-j:s)} | \eta, \kappa) \right], \quad (40)$$

relative to η and κ . The ANDA estimates can also be found by solving the following non-linear equation

$$\sum_{j=1}^s (2j-1) \left[\frac{\Delta_i(z_{(j:s)} | \eta, \kappa)}{F(z_{(j:s)} | \eta, \kappa)} - \frac{\Delta_m(z_{(s+1-j:s)} | \eta, \kappa)}{\bar{F}(z_{(s+1-j:s)} | \eta, \kappa)} \right] = 0; \quad i, m = 1, 2. \quad (41)$$

where $\Delta_1(\cdot | \eta, \kappa)$ and $\Delta_2(\cdot | \eta, \kappa)$ are given in Eq (33).

5.5. The CVM

The CVM estimates $\hat{\eta}_{\text{CVM}}$ and $\hat{\kappa}_{\text{CVM}}$ of TL_(A)(Ψ) are found by minimizing the function

$$CV(\eta, \kappa) = \frac{1}{12s} + \sum_{j=1}^s \left[F(z_{(j:s)} | \eta, \kappa) - \frac{2(j-1)+1}{2s} \right]^2, \quad (42)$$

relative to η and κ . Solving the non-linear equation, the CVM estimates can also be found by solving the following non-linear equation

$$\sum_{j=1}^s \left[F(z_{(j:s)} | \eta, \kappa) - \frac{2(j-1)+1}{2s} \right] \Delta_i(z_{(j:s)} | \eta, \kappa) = 0; \quad i = 1, 2., \quad (43)$$

where $\Delta_1(\cdot | \eta, \kappa)$ and $\Delta_2(\cdot | \eta, \kappa)$ are given in Eq (33).

6. MLE FOR TYPE-II RIGHT CENSORED DATA

Experiments on life testing is terminated when a specified number of failed objects have been observed, then the objects remaining are designated to be a type-II-right censored W. Let $z_{(1)}, z_{(2)}, \dots, z_{(p)}$, $p \leq s$ denote the ordered values of a random sample z_1, z_2, \dots, z_s (failure times) and observations terminate after the p^{th} failure occurs, then the likelihood function (C_{t-II}) is

$$C_{t-II} = \frac{s!}{(s-p)!} [R(z_p; \Psi)]^{s-p} \prod_{j=1}^p f(z_j; \Psi). \quad (44)$$

If z_1, z_2, \dots, z_s is a random sample from the TL_(A)(Ψ), then the log-likelihood function $L^{**}(\Psi)$ of $z_{(1)}, z_{(2)}, \dots, z_{(p)}$, $p \leq s$ is

$$\begin{aligned} L^{**}(\Psi) = & p \log(2\eta) + \log\left(\frac{s!}{(s-p)!}\right) + (s-p) \log\left\{1 - \left[1 - \left(1 - e^{-\frac{2}{\kappa}v_p}\right)^2\right]^\eta\right\} \\ & - 2 \sum_{j=1}^p \log(z_j) + \kappa \sum_{j=1}^p \frac{1}{z_j} - \frac{1}{\kappa} \sum_{j=1}^p v_j + \sum_{j=1}^p \log\left(1 - e^{-\frac{1}{\kappa}v_j}\right) \\ & + (\eta - 1) \sum_{j=1}^p \log\left[1 - \left(1 - e^{-\frac{1}{\kappa}v_j}\right)^2\right] \end{aligned} \quad (45)$$

where $v_p = e^{\frac{\kappa}{z_p}} - 1$ and $v_j = e^{\frac{\kappa}{z_j}} - 1$. The ML estimates $\hat{\eta}_{\text{ML}}$ and $\hat{\kappa}_{\text{ML}}$ of the unknown parameters of TL_(A)(Ψ) is found by maximizing Eq (45) using the R-programming software (*Optim function*).

7. MONTE CARLO EXPERIMENTS

In this part, the average estimates (AEs), absolute biases (AbsBs), mean square errors (MSEs) and mean relative errors (MREs) are computed for the TL_(A) parameters (Pa.) using Monte Carlo experiments with complete samples.

7.1. Monte Carlo experiments based on complete data

These Monte Carlo experiments are executed in R-programming software and the sampling distributions are found for different sample sizes (T) from $s = 3000$ replications for various values of κ and η . The classical estimators discussed in Part 5 for complete data are assessed and the average estimates (AEs) for each estimator are presented in Tables 2, 3 and 4. The comparison of the estimators graphically according to the AbsBs, MSEs and MREs for the TL_(A) parameters (Pa.) are depicted in Figures 4, 5 and 6. Therefore, the following conclusions are reached utilizing the graphical plots.

1. The estimators are asymptotically unbiased given that their absolute biases converge to zero as the sample size increases. The estimators are consistent given that their MSEs tend to zero for large sample size.
2. The MLE and OLS performs better than the other estimators in terms of minimum AbsBs and MREs in most cases while the MPS has the largest absolute biases and MREs compared to other estimators in most cases. The results indicate that the MLE, OLS, WLS, ANDA, CVM and MPS perform quite well in estimating the TL_(A) model parameters.

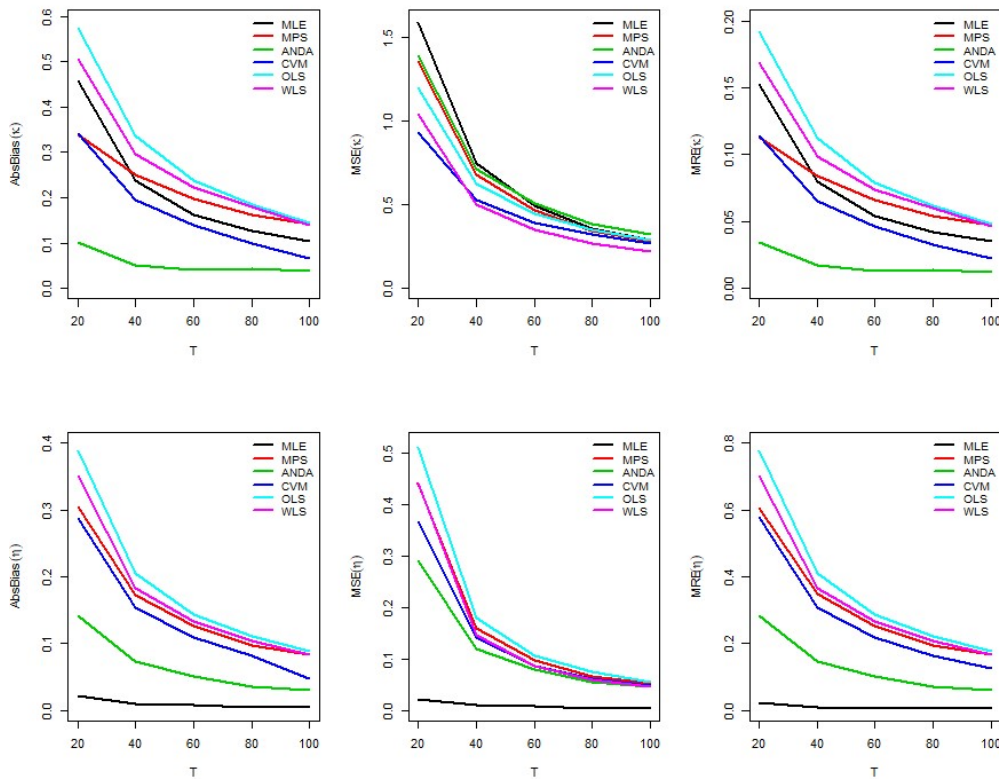


Figure 4: The estimators AbsBs, MSEs and MREs when $\kappa = 3.0$ and $\eta = 0.5$ (complete data).

Table 2: The estimators AEs when $\kappa = 3.0$ and $\eta = 0.5$ based on complete data.

T	Pa.	MLE	OLS	WLS	MPS	ANDA	CVM
20	κ	3.458	2.425	2.493	2.662	3.102	2.659
	η	0.522	0.889	0.850	0.804	0.642	0.788
40	κ	3.239	2.664	2.704	2.749	3.051	2.805
	η	0.510	0.705	0.683	0.673	0.573	0.653
60	κ	3.163	2.763	2.777	2.802	3.040	2.862
	η	0.507	0.643	0.633	0.626	0.551	0.609
80	κ	3.127	2.815	2.820	2.837	3.042	2.901
	η	0.504	0.611	0.604	0.597	0.535	0.582
100	κ	3.104	2.855	2.861	2.859	3.037	2.933
	η	0.504	0.588	0.583	0.583	0.530	0.563

Table 3: The estimators AEs when $\kappa = 3.5$ and $\eta = 2.0$ based on complete data.

T	Pa.	MLE	OLS	WLS	MPS	ANDA	CVM
50	κ	3.788	3.491	3.562	3.220	3.610	3.750
	η	2.003	2.286	2.200	2.460	2.161	2.100
100	κ	3.649	3.528	3.572	3.320	3.568	3.660
	η	2.004	2.134	2.080	2.270	2.081	2.040
150	κ	3.598	3.516	3.552	3.360	3.545	3.600
	η	2.005	2.091	2.051	2.190	2.056	2.030
200	κ	3.577	3.513	3.542	3.380	3.536	3.580
	η	1.999	2.064	2.033	2.150	2.037	2.020
250	κ	3.562	3.514	3.538	3.400	3.531	3.570
	η	2.000	2.051	2.025	2.120	2.030	2.010

Table 4: The estimators AEs when $\kappa = 2.0$ and $\eta = 2.5$ based on complete data.

T	Pa.	MLE	OLS	WLS	MPS	ANDA	CVM
200	κ	2.056	2.006	2.029	1.920	2.023	2.060
	η	2.492	2.562	2.530	2.650	2.536	2.510
400	κ	2.025	2.002	2.015	1.950	2.010	2.030
	η	2.502	2.535	2.517	2.590	2.522	2.510
600	κ	2.013	2.000	2.008	1.950	2.004	2.020
	η	2.506	2.526	2.515	2.570	2.518	2.510
800	κ	2.009	1.999	2.005	1.960	2.002	2.010
	η	2.504	2.521	2.511	2.550	2.514	2.510
1000	κ	2.008	2.000	2.005	1.970	2.003	2.010
	η	2.502	2.514	2.507	2.540	2.509	2.500

8. DATA APPLICATIONS

The flexibility of the $TL_{(A)}$ model is demonstrated here with three real datasets; two complete data and one type-II right censored data.

8.1. Applications for complete data

The first dataset corresponding to the relief times of twenty patients receiving an analgesic was previously analysed by [22]. 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. The second dataset corresponding to the scores of the general rating of affective symptoms for preschoolers (GRASP) which measures the emotional and behavioural problems of children was previously analysed by [1] and [11]. 19(16), 20(15), 21(14), 22(9), 23(12), 24(10), 25(6), 26(9), 27(8), 28(5), 29(6), 30(4), 31(3), 32(4), 33(1), 34(1), 35(4), 36(2), 37(2), 39(1), 42(1), 44(1). The MLE will be used to compare the goodness-of-fit of the $TL_{(A)}$ with the (A) model, inverse Gompertz (IG) model, lomax(LOMX) model, Pareto (PE) model, inverse Pareto (IPE) model, Pareto type-I (PETI) model, exponentiated inverse rayleigh (EIR) model, type-I half logistic skew-t (TIHLST) model, generalized inverse exponential (GIE) model and odd frechet inverse exponential

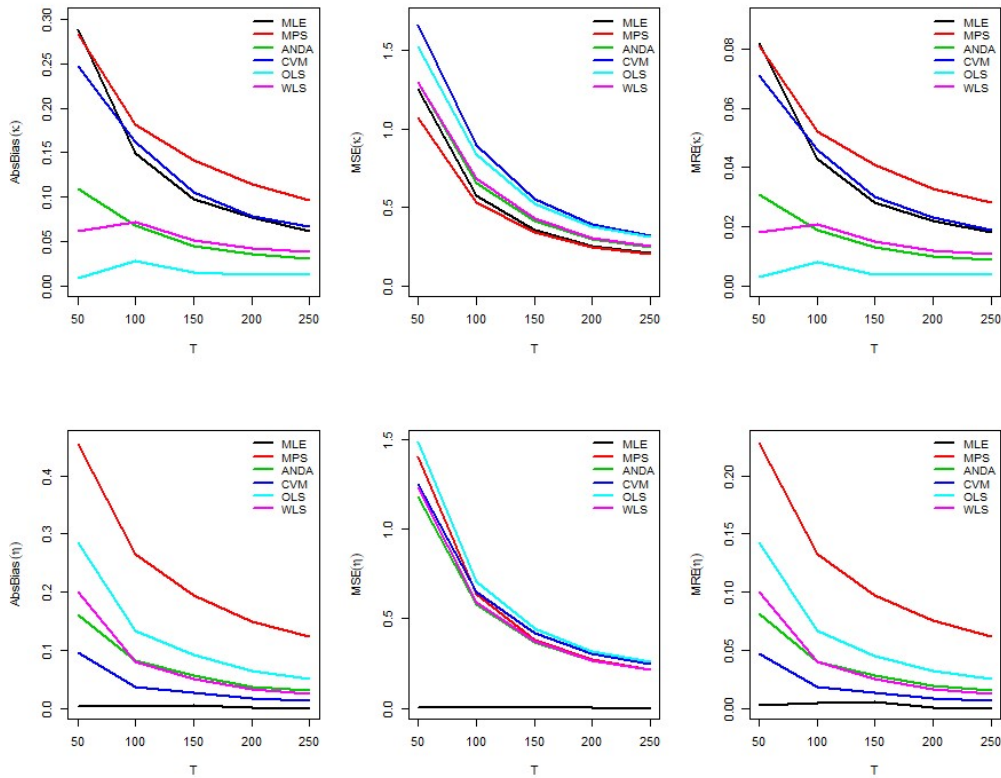


Figure 5: The estimators AbsBs, MSEs and MREs when $\kappa = 3.5$ and $\eta = 2.0$ (complete data).

(OFIE) model. These models will be fitted to the two complete datasets according to some criteria, namely, the Kolmogorov Smirnov test statistic (K.S) with its PVs. The Akaike information criterion (AIC), correct Akaike information criterion (CAIC), Hannan-Quinn information criterion (HQIC), Bayesian information criterion (BIC), Cramér-von Mises (W) statistic, Anderson-Darling (A) statistic and log-likelihood value (LL) will also be provided. The analysis is performed with the R-programming software using the *fitdistrplus*, *Optim* and *AdequacyModel* packages.

First dataset analysis

The MLEs, K.S and PVs for the first dataset are provided in Table 5 for all the studied models. The results show that the $TL_{(A)}$ has the least values for LL, AIC, CAIC, HQIC, BIC, and KS value with largest PV. This highlights that the $TL_{(A)}$ fits the first dataset better than (A), IG, LOMX, PE, IPE, PETI, OFIE, EIR, TIHLST and GIE models. This confirms that the $TL_{(A)}$ seems to be a very good model better than the other competing models. More so, the $TL_{(A)}$ model gives a more appropriate fit to the first data than the Kumaraswamy-transmuted exponentiated modified Weibull (KwTEXMW), McDonald log-logistic (McDLL), beta Weibull (BWE), modified Weibull (MWE), transmuted complementary Weibull geometric (TCWEG) and exponentiated transmuted generalized Rayleigh (ETGRH) models (see Table 5, [22]). Figure 7 depicts the fitted PDFs and fitted CDFs of all the models. The plots support the results presented in Table 5 that the $TL_{(A)}$ model provides the best goodness of fits to the first dataset.

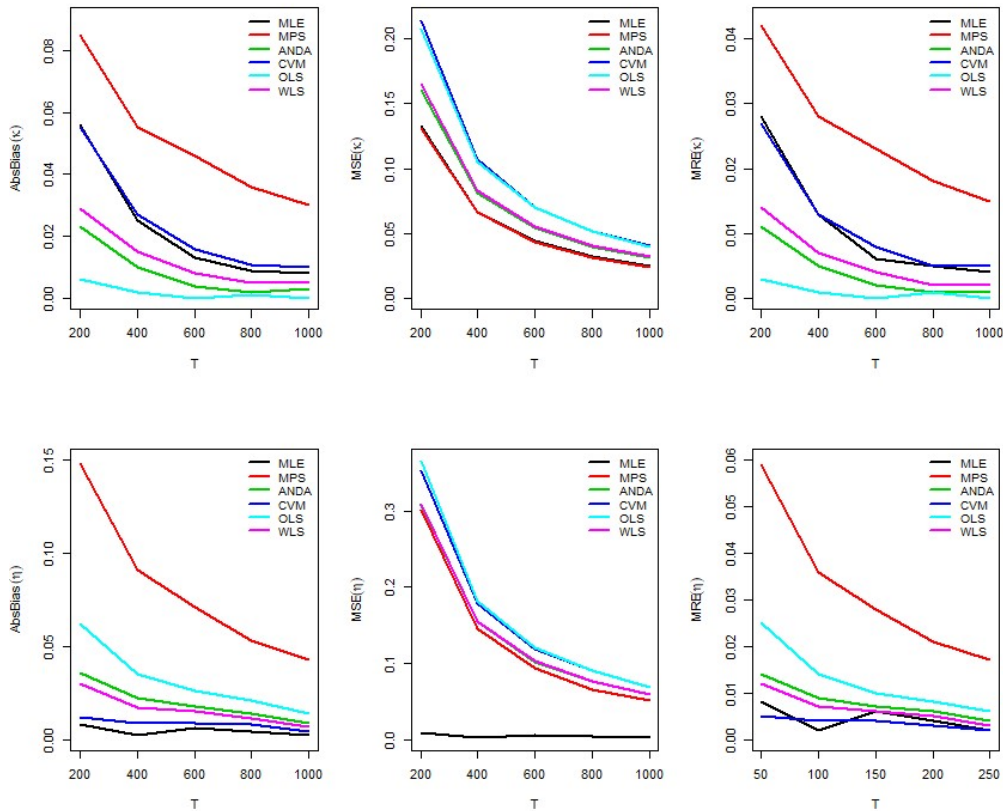


Figure 6: The estimators AbsBs, MSEs and MREs when $\kappa = 2.0$ and $\eta = 2.5$ (complete data).

Table 5: The MLEs, KS, PVs, LL, AIC, BIC, CAIC, HQIC, W^* and A^* values for the first dataset.

	Models										
	TLA	A	IG	LOMX	PE	IPE	PETI	GIE	TIHLST	EIR	OFIE
$\hat{\kappa}$	5.092	2.402	6.145	28.385	25.011	28.862	1.697	36.485	0.708	3.610	1.329
$\hat{\eta}$	0.228	-	0.110	0.019	47.333	0.059	-	2.232	2.347	2.336	0.906
K.S	0.116	0.385	0.142	0.444	0.437	0.380	0.285	0.134	0.504	0.127	0.358
PV	0.952	0.005	0.812	0.001	0.001	0.006	0.078	0.862	7.6E-05	0.906	0.012
LL	-15.650	-23.503	-16.392	-33.142	-33.182	-32.986	-21.207	-16.261	-38.265	-15.868	-26.591
AIC	35.300	49.006	36.783	70.283	70.364	69.972	44.414	36.521	80.529	35.736	57.181
CAIC	36.005	49.229	37.489	70.989	71.070	70.678	44.636	37.227	81.235	36.442	57.887
BIC	37.291	50.002	38.774	72.275	72.356	71.963	45.410	38.513	82.520	37.727	59.172
HQIC	35.688	49.201	37.172	70.672	70.729	70.361	44.609	36.910	80.918	36.125	57.570
W^*	0.025	0.028	0.055	0.102	0.102	0.050	0.038	0.054	0.090	0.042	0.032
A^*	0.105	0.162	0.332	0.607	0.605	0.293	0.219	0.319	0.533	0.244	0.179

Second dataset analysis

The MLEs, KS and PVs for the second dataset are provided in Table6 for all studied models. The results show that the $TL_{(A)}$ has the least values for LL, AIC, CAIC, HQIC, BIC, and KS value with largest PV. This highlights that the $TL_{(A)}$ fits the second dataset better than (A), IG, LOMX, PE, IPE, PETI, OFIE, EIR, TIHLST and GIE models. This confirms that the $TL_{(A)}$ seems to be a very good model better than the other competing models. More so, the $TL_{(A)}$ model gives a more appropriate fit to the second data than the generalized exponential (GEX), Gompertz (GTz), extended Gompertz (EGTz) and generalized Gompertz (GGTz) models (see Table 5, [1]). Figure 8 depicts the fitted pdfs and fitted CDFs plots of all the models. The plots support the results

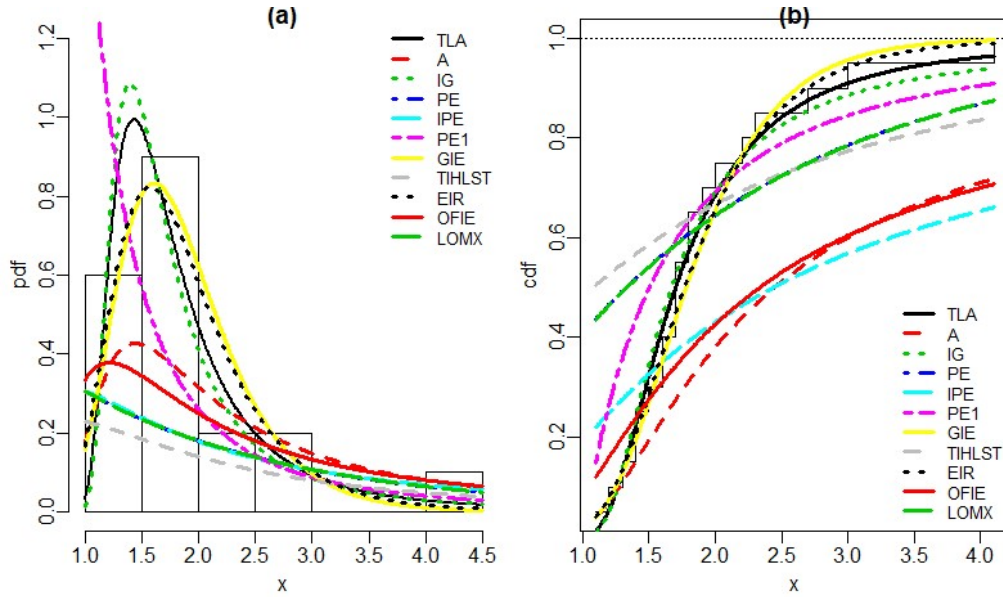


Figure 7: Fitted density function (pdf) plot (left panel), Fitted distribution function (CDF) plot (right panel) for the TL_(A) model (first dataset).

presented in Table 6 that the TL_(A) model provides the best goodness of fits to the second dataset.

Table 6: The MLEs, KS, PVs, LL, AIC, BIC, CAIC, HQIC, W* and A* values for the second dataset.

	Models										
	TLA	A	IG	LOMX	PE	IPE	PETI	GIE	TIHLST	EIR	OFIE
$\hat{\kappa}$	86.970	107.354	149.337	3.328	7.808	63.123	0.313	138.914	0.676	10.246	18.987
$\hat{\eta}$	4.809	-	0.175	0.012	185.974	0.379	-	0.218	359.391	40.286	0.943
K.S	0.100	0.130	0.102	0.495	0.532	0.452	0.602	0.108	0.572	0.109	0.446
PV	0.138	0.022	0.120	2.2E-16	2.2E-16	2.2E-16	2.2E-16	0.088	2.2E-16	0.085	2.2E-16
LL	-393.202	-404.277	-393.401	-582.812	-573.127	-565.908	-718.074	-399.265	-601.745	-401.694	-520.798
AIC	790.404	810.555	790.801	1169.625	1150.253	1135.816	1438.148	802.530	1207.490	807.388	1045.595
CAIC	790.496	810.585	790.893	1169.716	1150.345	1135.907	1438.179	802.622	1207.582	807.479	1045.687
BIC	796.200	813.453	796.597	1175.420	1156.049	1141.611	1441.046	808.326	1213.286	813.452	1051.391
HQIC	792.759	811.732	793.157	1171.980	1152.608	1138.171	1439.326	804.886	1209.846	809.743	1047.951
W	0.208	0.209	0.235	0.371	0.388	0.286	0.302	0.246	0.377	0.322	0.227
A	1.523	1.525	1.708	2.512	2.613	1.994	2.091	1.756	2.545	2.214	1.631

For the TL_(A), the approximate 95% two-sided confidence intervals (CIs) for the parameters κ and η are [2.276, 7.909] and [-0.176, 0.632] for the first dataset and [65.743, 108.197] and [-1.249, 10.868] for the second dataset, respectively. The likelihood ratio test (LRT) is normally used to test if the fit by TL_(A) model is statistically superior to the fit provided by the (A) model. Table 7 provides the values of the LRT, degree of freedom (d.f) and its PVs for the first and second datasets. Based on the PVs, the null hypothesis (H_0) is rejected at $\alpha = 0.05$ level of significance.

Table 7: The LR tests for the first and second datasets.

Model	Hypotheses	LR	PV
First dataset			
(A) vs. TL _(A)	$H_0 : \eta = 1$ vs. $H_1 : H_0$ is false	15.707	0.00074
Second dataset			
(A) vs. TL _(A)	$H_0 : \eta = 1$ vs. $H_1 : H_0$ is false	22.151	0.0000025

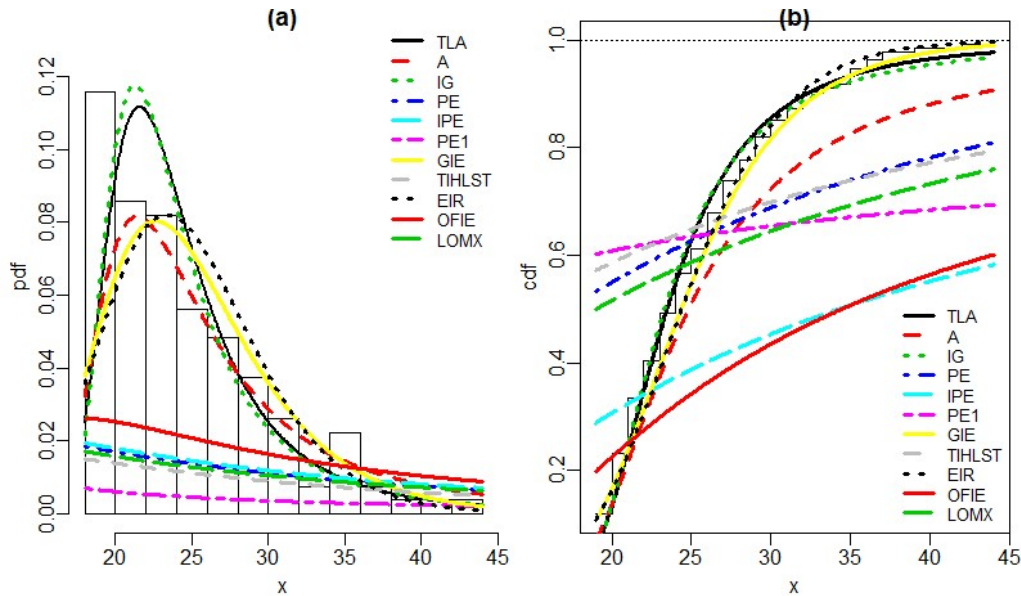


Figure 8: Fitted density function (pdf) plot (left panel), Fitted distribution function (CDF) plot (right panel) for the $TL_{(A)}$ model (second dataset).

8.2. Application for censored data

The censored data used here which represents the fatigue life for 10 bearings of a specific type in hours was introduced by DD. 152.7, 172, 172.5, 173.5, 193, 204.7, 216.5, 234.9, 262.6, 422.6. Assume a type II right censored sample of size $p = 8$ is taken from this data. Table 8 shows the MLEs, LL, K.S and PV for the $TL_{(A)}$ model. It is clear that the $TL_{(A)}$ well fitted to the data based on the K.S and its PV.

Table 8: The MLEs and performance measures for the type-II right censored data.

Models	MLEs	LL	K.S	PV
$TL_{(A)}$	$\kappa = 539.86, \eta = 946.82$	-41.129	0.263	0.550

9. DISCUSSIONS AND CONCLUSION

In this work, a new model titled $TL_{(A)}$ which is considered as an extension and generalization to the (A) model is proposed. The $TL_{(A)}$ is characterized by an inverted bathtub shaped curve, increasing and decreasing hazard rate function quite dependent on the shape parameter. More so, the $TL_{(A)}$ is appropriate for testing the goodness of fit of the sub-model, the (A) model. Some structural properties including the ordinary and incomplete moments, MGFs, PWMs, quantiles, Bonferroni and Lorenz curves, entropies, median and order statistics of $TL_{(A)}$ are derived. Likewise, basic functions utilized in reliability theory such as the survival function, HRF, reversed HRF, cumulative HRF, MTTF, MRL and MWT are derived. The Monte Carlo experiments are carried out to determine the performance of MLE, MPS, ANDA, CVM, OLS and WLS methods according to AbsBs, MSEs and MREs measures. The experiments results indicate that the estimators perform quite well in producing good parameter estimates for all the various parameter groups at different sample sizes. However, the MLE method produced closer estimates for $TL_{(A)}$ parameters. This conforms to the reports by [24, 25, 26, 27, 28]. Furthermore, the parameters of $TL_{(A)}$ are appreciated using the MLE in the case of complete and type-II-right censored data. The two complete data are analysed using the $TL_{(A)}$ and compared with ten other competing lifetime models. Likewise, a type-II-right censored data is analysed using the $TL_{(A)}$ model. The results indicate that the $TL_{(A)}$ has more flexibility for fitting the various datasets.

Futuristically, the bivariate extension of the TL_(A) model, the TL_(A)-G family of distributions and the discrete case of the TL_(A) model will be addressed.

REFERENCES

- [1] Eliwa, M. S., El-Morshedy, M and Ibrahim, M. (2009). Inverse Gompertz distribution: properties and different estimation methods with application to complete and censored data. *Annals of data science*, 6(2):321–339.
- [2] Alrajhi, S. (2019). The odd Fréchet inverse exponential distribution with application. *Journal of Nonlinear Sciences & Applications (JNSA)*, 12(8):1–24.
- [3] El-Morshedy, M., El-Faheem, A. A. and El-Dawoody, M. (2020). Kumaraswamy inverse Gompertz distribution: Properties and engineering applications to complete, type-II right censored and upper record data. *Plos one*, 15(12):e0241970.
- [4] Aldahlan, Maha, A. D. and Afify, A. Z. (2020). The odd exponentiated half-logistic exponential distribution: estimation methods and application to engineering data. *Mathematics*, 8(10):1684.
- [5] Rana, M. S. and Rahman, M. M. (2022). The Pareto-Exponential Distribution: Theory and Real-life Applications. *European Journal of Mathematics and Statistics*, 3(3):30–39.
- [6] Adubisi, O. D., Abdulkadir, A. and Chiroma, H. (2021). A Two Parameter Odd Exponentiated Skew-T Distribution With J-Shaped Hazard Rate Function. *Journal of Statistical Modeling & Analytics (JOSMA)*, 3(1): 26–46.
- [7] Adubisi, O. D., Abdulkadir, A., Chiroma, H. and Abbas, U. F. (2021). The type I half logistic skew-t distribution: A heavy-tail model with inverted bathtub shaped hazard rate. *Asian J Probab Stat*, 14:21–40.
- [8] Adubisi, O. D., Abdulkadir, A., Abbas, U. F. and Chiroma, H. (2021). Financial data and a new generalization of the skew-t distribution. *Covenant Journal of Physical and Life Sciences*, 9(2):1–17.
- [9] Dhungana, G. P. and Kumar, V. (2022). Exponentiated Odd Lomax Exponential distribution with application to COVID-19 death cases of Nepal. *PloS one*, 17(6):e0269450.
- [10] Beghriche, A., Zeghdoudi, H., Raman, V. and Chouia, S. (2022). New polynomial exponential distribution: properties and applications. *Statistics in Transition new series*, 23(3):95–112.
- [11] Alshenawy, R. (2020). A new one parameter distribution: properties and estimation with applications to complete and type II censored data. *Journal of Taibah University for Science*, 14(1):11–18.
- [12] Al-Shomrani, A., Arif, O., Shawky, A. Hanif, S. and Shahbaz, M. Q. (2016). Topp–Leone family of distributions: Some properties and application. *Pakistan Journal of Statistics and Operation Research*, 7:443–451.
- [13] Rényi, A. (1961). Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. *Contributions to Biology and Problems of Medicine*, 4:547-561.
- [14] Greenwood, J. A., Landwehr, J. M., Matalas, N. C and Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water resources research*, 15(5):1049-1054.
- [15] David, H. A. and Nagaraja, H. N. Order statistics, John Wiley & Sons, 2004.
- [16] Kenney, J. F. Mathematics of statistics, D. Van Nostrand, 1939.
- [17] Moors, J. J. A. (1988). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37(1):25–32.
- [18] Swain, J. J., Venkatraman, S. and Wilson, J. R. (1988). Least-squares estimation of distribution functions in Johnson’s translation system. *Journal of Statistical Computation and Simulation*, 29(4):271–297.
- [19] Cheng, R. C. H and Amin, N. A. K. (1979). Maximum product-of-spacings estimation with applications to the lognormal distribution. *Math report*, 791.

- [20] Cheng, R. C. H and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(3):394–403.
- [21] Zheng, G. and Park, S. (2004). On the Fisher information in multiply censored and progressively censored data. *Communications in Statistics-Theory and Methods*, 33(8):1821–1835.
- [22] Afify, A., Yousof, H. and Nadarajah, S. (2017). The beta transmuted-H family for lifetime data. *Statistics and its Interface*, 10(3):505–520.
- [23] McCool, J. I. (1978). Competing risk and multiple comparison analysis for bearing fatigue tests. *ASLE transactions*, 21(4):271–284.
- [24] Ramos, P. L., Louzada, F., Ramos, E. and Dey, S. (2020). The Fréchet distribution: Estimation and application-An overview. *Journal of Statistics and Management Systems*, 23(3):549-578.
- [25] Chesneau, C., Bakouch, H. S., Ramos, P. L. and Louzada, F. (2020). The polynomial-exponential distribution: A continuous probability model allowing for occurrence of zero values. *Communications in Statistics-Simulation and Computation*, 22:1–26.
- [26] Adubisi, O. D., Abdulkadir, A. and Adashu, D. J. (2022). Improved Parameter Estimators for the Flexible Extended Skew-t Model with Extensive Simulations, Applications and Volatility modeling. *Scientific African*, e01443.
- [27] Adubisi, O. D., Adubisi, C. E. and Abdulkadir, A. (2022). Laplace Transformed Properties of the Extended Power-Gompertz Model: Simulation and Applications. *Scientific African*, e01523.
- [28] Adubisi, O. D., Abdulkadir, A. and Adubisi, C. E. (2022). A new hybrid form of the skew-t distribution: estimation methods comparison via Monte Carlo simulation and GARCH model application. *Data Science in Finance and Economics*, 2(2):54-79.

RELIABILITY AND AVAILABILITY ANALYSIS OF AN 8-STEP AUTO UNIT MANUFACTURING PROCESS HAVING A FAULT IN MAINTENANCE

Surabhi Sengar

•

G. B. Pant University of Agriculture and Technology, Pantnagar (India)
sursengar@gmail.com

Abstract

Human plays a vital role in the manufacturing process of a product during the planning, design, assembly, production, and maintenance phase. This paper presents a systematic method of determining the reliability and availability of an 8-step auto unit manufacturing system taking into consideration the case of occurring human error during maintenance. The whole process involves eight major units as Combustion engine, power train unit, fuel feed unit, fuel injection unit, drain or exhaust unit, engine cooling unit, brake unit, and body frame. With the failure of any of these units, the whole process can fail. Also, a constant failure rate and a general repair time are taken into the consideration for each operative unit that makes up the process. The integral differential equations were generated based on Markov modeling of the process and solution derived by considering repair time distribution using the Laplace transform. Calculations of different reliability aspects like process availability, mean time to failure steady-state nature, and profit analysis are done using supplementary variable technique and copula methodology.

Keywords: Reliability, Availability, Failure probability, Supplementary variable technique, Markov model, Human error.

1. Introduction

For a very long period, a good amount of work has been done to find the reliability of various manufacturing and industrial systems. Many researchers applied different methods like the Hidden Markov models, state space method, K map approach, probabilistic rational model method, etc. to analyze the reliability and steady-state nature of some realistic manufacturing systems like Aircraft and sea-going ships, bufferless production system, Aircraft commutators, automobile assembly process, fiber plant unit, etc. [1, 4, 5, 12, 13, and 14]. Here the discussion is focused on the reliability and availability analysis of an 8-step auto unit manufacturing process.

The automobile sector has played a significant role in the economic growth and development of any country. Because of the steady growth in the GDP of a country due to the contribution of the auto industry, it is also called an "industry of industries". Although the industry started in Germany and France now it has become one of the global all over the world.

The automobile manufacturing process is a multifaceted process involving various steps starting from designing, building, quality checks, and shipping. The first phase of the process is to design, develop, and product analysis. Once this phase is completed the next step is fulfill the requirements of tools and apparatus for the bulk production of the vehicle. This phase includes everything right from engine assembly, welding and painting to stamping machines. Following this

phase, is production planning which decides the quantity needs to manufacture. Before launching the product, it has to undergo different quality and safety checks. The purpose of these checks is to ensure that the product is designed by considering all safety standards and will perform failure-free operation.

In this competitive scenario, to sustain itself in the market and to fulfill customer expectation it is required to increase the product performance, topographies, and quality, as a result, the complexity and mechanization of the product are increased and it has resulted in several issues related to maintenance and repair [8, 9]. When we talk about manufacturing lines, a repairman plays a vital role in this field during the design, installation, production, and maintenance. The probability of occurrence of human error is more in this maintenance phase only. Human error in repair is a concern, that has not gotten the proper attention of the researchers. It can be defined as failure to perform a particular operation that could result in a delay in the production or could damage a machine or equipment. These errors may occur due to various reasons like insufficient facilities and services, less technical skills, design errors or improper planning, etc. These errors may be classified into different categories like Design, operational, installation, quality check, maintenance, etc. Maintenance error is defined as an unintentional failure occurred during the maintenance of the product because of improper repair or proactive measures. The probability of occurrence of maintenance error increases as the product gets older [6, 10].

By considering all the above facts the present study proposes a methodology to assess the reliability and performance of a manufacturing process having a fault in maintenance due to human error. The assumed process involves eight major units as Combustion engine, power train unit, fuel feed unit, fuel injection unit, drain or exhaust unit, engine cooling unit, brake unit, and body frame. The main focus of the study is on failure during maintenance due to human error. The following assumptions have been made for the whole process:

- The combustion engines can fail due to operational errors.
 - The power train unit contains: a clutch, transmission gear, differential and final drive
 - It is assumed that the process can fail due to inappropriate fitting of the power train and transmission gear.
 - The process can be in a degraded state because of inappropriate working of differential.
 - Also, the process can completely fail due to final drive failure.
 - It is considered that the process can fail due mistake in fuel injection pump timings during an inspection in the fuel feed unit.
 - Furthermore, it is also assumed that the process can fail due to maintenance errors and mistakes in quality checks in the fuel injection unit because of misfiring and disturbed pressure levels.
 - Improper assembly of the exhaust units can cause process failure and overheating problems in the engine cooling units can take the process into a degraded state.
 - The process can fail due to design error in break unit.
 - Improper installation of axle and chassis can lead to process failure in the build unit.
- The whole process can fail due to the failure of any of these units. A joint repair policy is applied to repair the system in the power train when the failure occurs due to failure of the differential and final drive.
- Fuel injection unit when the failure occurs due to misfiring and disturbed pressure level.

Here, the joint probability distribution is applied with the help of the Gumbel-Hougaard Copula methodology [7, 11]. Also, failures follow exponential time distribution while general time distribution is applied for repairs. To help the production industry some parametric investigations for process reliability, availability, mean time to failure, and profit analysis has been made [2, 3]. The state transition figure is shown in figure-1.

2. Notations

Different notations used in the model are given in table 1 and the state specification of the model is described in table 2.

Table 1: Notations

$P_0(t)$:	Indicates the probability that initially the whole system is fully operational, denoted by state S_0
$P_i(j,t)$:	The probability that the process is in the failed state due to failure of the i^{th} unit at any time, here, $i=1,2,4,5,6,7,9,10,11$ and Elapsed repair time $j= x, y, v, w, m, r, n, v, k$
α_{EO}	:	Combustion engine failure rate due to operational error
$\alpha_{TI} / \alpha_{FSS}$:	Failure rates of transmission gear and fuel injection pump in fuel supply
α_D / α_{DS}	:	Differential unit failure rate and final drive failure rate
α_M / α_{LP}	:	The failure rate of misfiring and disturbed pressure level to be low pressure of fuel injection unit
α_{EA}	:	Exhaust unit failure rate
α_{OVH}	:	Failure rate due to overheating
α_{CS}	:	The cooling unit failure rate
α_{BR}	:	Break unit failure rate
α_{BU}	:	Build unit failure rate
$\varphi_k(j)$:	Shows the repair rate of k^{th} failure in the time interval $(j, j+\Delta)$, where, $k = EO, TI, D, DS, FSS, M, LP, EA, OVH, CS, BR, BU$, and $j = x, y, v, w, m, r, n, v, k$
K_1, K_2	:	Profit cost and service cost per unit of time respectively

Consider that, $u_1 = e^m$, $u_2 = \phi_{LP+M}(m)$ and $X_1 = e^w$, $X_2 = \phi_{D+DS}(w)$, then joint probability is given by the expression,

$$\phi_{LP+M}(m) = \exp[m^\theta + (\log \phi_{LP+M}(m))^\theta]^{1/\theta}, \phi_{D+DS}(w) = \exp[w^\eta + (\log \phi_{D+DS}(w))^\eta]^{1/\eta}$$

using Gumbel- Hougaard copula methodology.

3. State specification

The state specification of the process is given by the table 2.

Table 2: Process state specification

States	Description	System State
S_0	State in which the process is fully operational	G
S_1	State in which the process gets failed due to operational error in combustion engine	FR
S_2	State in which the process is failed due to inappropriate fitting of the power train and transmission gear	FR

S ₃	State in which the process is in a degraded state because of inappropriate working of differential	D
S ₄	State in which the process is completely failed due to final drive failure	FR
S ₅	State in which the process fails due mistakes in fuel injection pump timings during an inspection in the fuel feed unit	FR
S ₆	State in which the process fails due to maintenance error and mistake in quality check in fuel injection unit because of misfiring and disturbed pressure level	FR
S ₇	State in which process gets fail due to improper assembly of the exhaust unit	FR
S ₈	State in which the process gets degraded because of overheating problem in the engine cooling unit	D
S ₉	State in which process is failed due to failure of the cooling unit	FR
S ₁₀	State in which process gets failed due to design error in break unit	FR
S ₁₁	State in which the process gets failed due to improper installation of axle and chassis in the build unit	FR

Note: G= Good state; FR= Failed state under repair; D = Degraded state

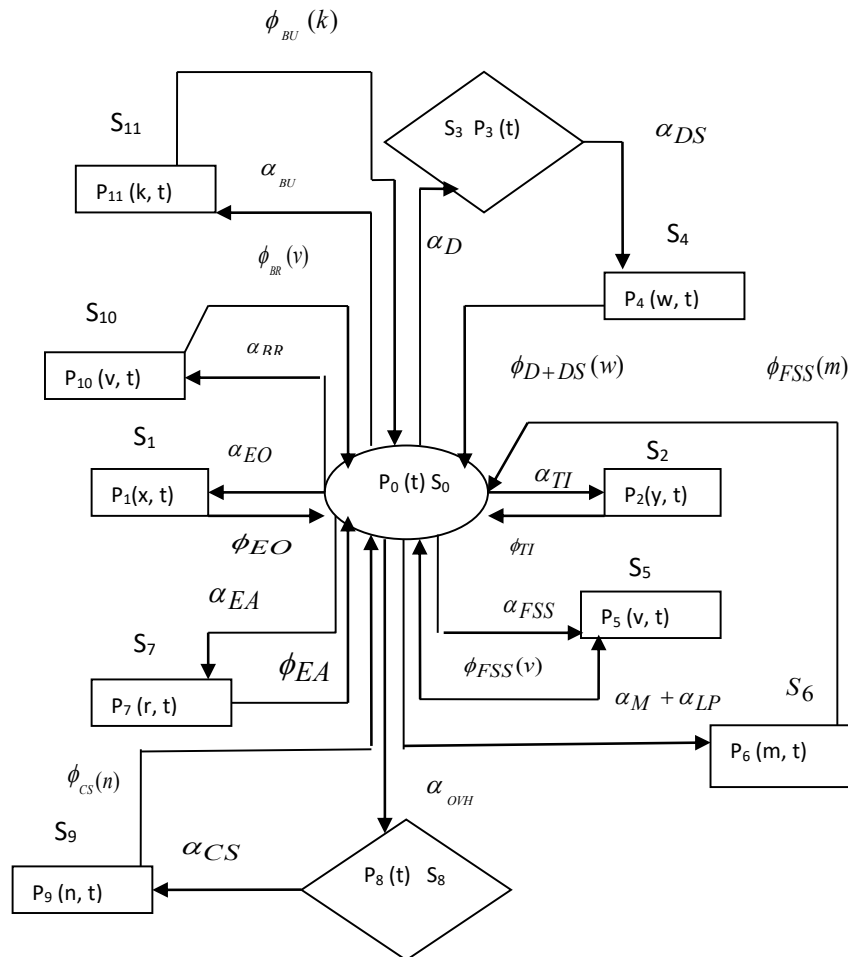


Figure 1: State transition diagram for 8-step auto manufacturing process

4. Mathematical formulation of the model

The differential transition state probabilities for different states of the model are given by the following equations

$$\left[\frac{d}{dt} + \alpha_{EO} + \alpha_{TI} + \alpha_D + \alpha_{FSS} + (\alpha_M + \alpha_{LP}) + \alpha_{EA} + \alpha_{OVH} + \alpha_{BR} + \alpha_{BU} \right] P_0(t) = \int_0^\infty \phi_{EO}(x) P_1(x,t) dx + \int_0^\infty \phi_{TI}(y) P_2(y,t) dy + \int_0^\infty \phi_{D+DS}(w) P_4(w,t) dw + \int_0^\infty \phi_{FSS}(v) P_5(v,t) dv + \int_0^\infty \phi_{M+LP}(m) P_6(m,t) dm + \int_0^\infty \phi_{EA}(r) P_7(r,t) dr + \int_0^\infty \phi_{CS}(n) P_9(n,t) dn + \int_0^\infty \phi_{BR}(l) P_{10}(l,t) dl + \int_0^\infty \phi_{BU}(k) P_{11}(k,t) dk \tag{1}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial x} + \phi_{EO}(x) \right] P_1(x,t) = 0 \tag{2}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial y} + \phi_{TI}(y) \right] P_2(y,t) = 0 \tag{3}$$

$$\left[\frac{d}{dt} + \alpha_{DS} \right] P_3(t) = \alpha_D P_0(t) \tag{4}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial w} + \phi_{D+DS}(w) \right] P_4(w,t) = 0 \tag{5}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial v} + \phi_{FSS}(v) \right] P_5(v,t) = 0 \tag{6}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial m} + \phi_{LP+M}(m) \right] P_6(m,t) = 0 \tag{7}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial r} + \phi_{EA}(r) \right] P_7(r,t) = 0 \tag{8}$$

$$\left[\frac{d}{dt} + \alpha_{CS} \right] P_8(t) = \alpha_{OVH} P_0(t) \tag{9}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial n} + \phi_{CS}(n) \right] P_9(n,t) = 0 \tag{10}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial v} + \phi_{BR}(v) \right] P_{10}(v,t) = 0 \tag{11}$$

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial k} + \phi_{BU}(k) \right] P_{11}(k,t) = 0 \tag{12}$$

Boundary Conditions:

$$P_i(0,t) = \alpha_k P_0(t) \tag{13}$$

$$P_4(0,t) = \alpha_{DS} P_3(t) \tag{14}$$

$$P_9(0,t) = \alpha_{CS} P_8(t) \tag{15}$$

Where, $i=1,2,5,6,7,10,11$ and $k=EO, TI, D, FSS, M, LP, EA, OVH, BR, BU$

Initial Condition:

$$P_0(0) = 1, \text{ otherwise zero.} \tag{16}$$

For finding the solution of the mathematical model, Taking Laplace transforms of equation (1) to (12), subject to the initial condition (16), and then solving them, we get the following transition state probabilities of the model:

$$\bar{P}_0(s) = \frac{1}{B(s)} \tag{17}$$

$$\bar{P}_i(s) = \frac{\alpha_k \times D_{\phi_k}(s)}{B(s)} \tag{18}$$

where $i=1, 2, 5, 6, 7, 10, 11$ and $k=EO, TI, FSS, LP+M, EA, BR, BU$

$$\bar{P}_3(s) = \frac{\alpha_D}{[s + \alpha_{DS}]} \times \frac{1}{B(s)} \tag{19}$$

$$\bar{P}_4(s) = \frac{\alpha_D \alpha_{DS}}{[s + \alpha_{DS}]} \times \frac{D\phi_{D+DS}(s)}{B(s)} \tag{20}$$

$$\bar{P}_8(s) = \frac{\alpha_{OVH}}{[s + \alpha_{CS}]} \times \frac{1}{B(s)} \tag{21}$$

$$\bar{P}_9(s) = \frac{\alpha_{CS} \times \alpha_{OVH} \times D_{\phi_{CS}}(s)}{(s + \alpha_{CS}) \times B(s)} \tag{22}$$

where,

$$B(s) = s + \alpha_{EO} + \alpha_{TI} + \alpha_D + \alpha_{FSS} + (\alpha_M + \alpha_{LP}) + \alpha_{EA} + \alpha_{OVH} + \alpha_{BR} + \alpha_{BU} \\
 - \alpha_{EO} \times \bar{S}_{\phi_{EO}}(s) - \alpha_{TI} \times \bar{S}_{\phi_{TI}}(s) - \frac{\alpha_D \alpha_{DS}}{[s + \alpha_{DS}]} \times \bar{S}_{\phi_{D+DS}}(s) - \alpha_{FSS} \times \bar{S}_{\phi_{FSS}}(s) \\
 - \alpha_{LP+M} \times \bar{S}_{\phi_{LP+M}}(s) - \alpha_{EA} \times \bar{S}_{\phi_{EA}}(s) - \frac{\alpha_{CS} \alpha_{OVH}}{(s + \alpha_{CS})} \times \bar{S}_{\phi_{CS}}(s) - \alpha_{BR} \times \bar{S}_{\phi_{BR}}(s) \\
 - \alpha_{BU} \times \bar{S}_{\phi_{BU}}(s) \tag{23}$$

$$\phi_{LP+M}(m) = \exp[m^\theta + (\log \phi_{LP+M}(m))^\theta]^{1/\theta} \tag{24}$$

$$\phi_{D+DS}(w) = \exp[w^\eta + (\log \phi_{D+DS}(w))^\eta]^{1/\eta} \tag{25}$$

Also, up-state and down-state probabilities of the system are given by:

$$\bar{P}_{up}(s) = \bar{P}_0(s) + \bar{P}_3(s) + \bar{P}_8(s) \\
 = \frac{1}{B(s)} \times \left[1 + \frac{\alpha_D}{[s + \alpha_{DS}]} + \frac{\alpha_{OVH}}{[s + \alpha_{CS}]} \right] \tag{26}$$

$$\bar{P}_{down}(s) = \bar{P}_1(s) + \bar{P}_2(s) + \bar{P}_4(s) + \bar{P}_5(s) + \bar{P}_6(s) + \bar{P}_7(s) + \bar{P}_9(s) + \bar{P}_{10}(s) + \bar{P}_{11}(s) \\
 = \frac{1}{B(s)} \times [$$

$$\alpha_{EO} \times D_{\phi_{EO}}(s) + \alpha_{TI} \times D_{\phi_{TI}}(s) + \frac{\alpha_D \alpha_{DS}}{[s + \alpha_{DS}]} D_{\phi_{D+DS}}(s) + \alpha_{FSS} \times D_{\phi_{FSS}}(s)$$

$$\begin{aligned}
 & + \alpha_{LP+M} \times D_{\phi_{LP+M}}(s) + \alpha_{EA} \times D_{\phi_{EA}}(s) + \frac{\alpha_{CS} \times \alpha_{OVH} D_{\phi_{CRSG}}(s)}{(s + \alpha_{CS})} + \\
 & \alpha_{BR} \times D_{\phi_{BR}}(s) + \alpha_{BU} \times D_{\phi_{BU}}(s)
 \end{aligned} \tag{27}$$

4.1. Steady-state behavior of the system

By using Abel's lemma,

$$\lim_{s \rightarrow 0} \{s \bar{F}(s)\} = \lim_{t \rightarrow \infty} F(t)$$

We get following up and down time independent operational probabilities:

$$\bar{P}_{up} = \frac{1}{B(0)} \times \left[1 + \frac{\alpha_D}{\alpha_{DS}} + \frac{\alpha_{OVH}}{\alpha_{CS}} \right] \tag{28}$$

$$\begin{aligned}
 \bar{P}_{down} = \frac{1}{B(0)} \times [& \alpha_{EO} \times \bar{M}_{\phi_{EO}} + \alpha_{TI} \times \bar{M}_{\phi_{TI}}(s) + \frac{\alpha_D}{\alpha_{DS}} + \frac{\alpha_D \times \alpha_{DS}}{[\alpha_{DS}]} \times M_{\phi_{D+DS}} \\
 & + \alpha_{FSS} \times M_{\phi_{FSS}} + \alpha_{LP+M} \times M_{\phi_{LP+M}} + \alpha_{EA} \times M_{\phi_{EA}} + \frac{\alpha_{OVH}}{\alpha_{CS}} \\
 & + \frac{\alpha_{CS} \alpha_{OVH}}{\alpha_{CS}} M_{\phi_{CS}} + \alpha_{BR} \times M_{\phi_{BR}} + \alpha_{BU} \times M_{\phi_{BU}}]
 \end{aligned} \tag{29}$$

where,

$$B(0) = \lim_{s \rightarrow 0} B(s) \tag{30}$$

$$\bar{M}_i = \lim_{s \rightarrow 0} \left\{ \frac{1 - \bar{S}_i(s)}{s} \right\}, \quad i = \phi_{EO}, \phi_{TI}, \phi_{D+DS}, \phi_{FSS}, \phi_{LP+M}, \phi_{EA}, \phi_{CS}, \phi_{BR}, \phi_{BU} \tag{31}$$

$$S_{\phi_i}(s) = \frac{\phi_i}{s + \phi_i}, \quad i = \phi_{EO}, \phi_{TI}, \phi_{D+DS}, \phi_{FSS}, \phi_{LP+M}, \phi_{EA}, \phi_{CS}, \phi_{BR}, \phi_{BU} \tag{32}$$

4.2. Special case

when repair rates follow exponential time distribution then,

$$\bar{S}_{\phi_i}(s) = \frac{\phi_i}{s + \phi_i}, \quad \forall i, \quad \bar{S}_{\phi_{LP+M}}(s) = \frac{\exp[m^\theta + [\log \phi_{LP+M}(m)]^\theta]^{1/\theta}}{s + \exp[m^\theta + [\log \phi_{LP+M}(m)]^\theta]^{1/\theta}} \quad \text{and}$$

$$\bar{S}_{\phi_{D+DS}}(s) = \frac{\exp[w^\theta + [\log \phi_{D+DS}(w)]^\eta]^{1/\eta}}{s + \exp[w^\theta + [\log \phi_{D+DS}(w)]^\eta]^{1/\eta}} \quad \text{in equations (17) to (22) we get,}$$

$$\begin{aligned}
 \bar{P}_{up}(s) = \frac{1}{B_1(s)} \times & \left[1 + \frac{\alpha_D}{[s + \alpha_{DS}]} + \frac{\alpha_{OVH}}{[s + \alpha_{CS}]} \right] \\
 \bar{P}_{down}(s) = \frac{1}{B_1(s)} \times & \left[\frac{\alpha_{EO}}{[s + \phi_{EO}]} + \frac{\alpha_{TI}}{[s + \phi_{TI}]} + \frac{\alpha_D}{[s + \phi_{DS}]} + \frac{\alpha_D \alpha_{DS}}{[s + \phi_{DS}][s + \phi_{D+DS}]} + \right. \\
 & \frac{\alpha_{FSS}}{[s + \phi_{FSS}]} + \frac{\alpha_{LP+M}}{[s + \phi_{LP+M}]} + \frac{\alpha_{EA}}{[s + \phi_{EA}]} + \frac{\alpha_{OVH}}{[s + \alpha_{CS}]} + \frac{\alpha_{CS} \alpha_{OVH}}{[s + \phi_{CS}][s + \alpha_{CS}]} \\
 & \left. + \frac{\alpha_{BR}}{[s + \phi_{BR}]} + \frac{\alpha_{BU}}{[s + \phi_{BU}]} \right]
 \end{aligned}$$

where,

$$\begin{aligned}
 B_1(s) = & s + \alpha_{EO} + \alpha_{TI} + \alpha_D + \alpha_{FSS} + (\alpha_M + \alpha_{LP}) + \alpha_{EA} + \alpha_{OVH} + \alpha_{BR} + \alpha_{BU} - \frac{\alpha_{EO} \times \phi_{EO}}{[s + \phi_{EO}]} \\
 & - \frac{\alpha_{TI} \times \phi_{TI}}{[s + \phi_{TI}]} - \frac{\alpha_D \times \alpha_{DS}}{[s + \alpha_{DS}]} \frac{\phi_{D+DS}}{[s + \phi_{D+DS}]} - \frac{\alpha_{FSS} \times \phi_{FSS}}{[s + \phi_{FSS}]} - \frac{\alpha_{LP+M} \times \phi_{LP+M}}{[s + \phi_{LP+M}]} - \frac{\alpha_{EA} \times \phi_{EA}}{[s + \phi_{EA}]} \\
 & - \frac{\alpha_{CS} \times \alpha_{OVH} \times \phi_{CS}}{[s + \phi_{CS}][s + \alpha_{CS}]} - \frac{\alpha_{BR} \times \phi_{BR}}{[s + \phi_{BR}]} - \frac{\alpha_{BU} \times \phi_{BU}}{[s + \phi_{BU}]}
 \end{aligned} \tag{33}$$

5. Numerical Computation

Considering different values of failure rates like $\alpha_{EO} = 0.006$, $\alpha_{TI} = 0.009$, $\alpha_D = 0.008$, $\alpha_{DS} = 0.01$, $\alpha_{FSS} = 0.07$, $\alpha_M = 0.008$, $\alpha_{LP} = 0.006$, $\alpha_{EA} = 0.002$, $\alpha_{OVH} = 0.005$, $\alpha_{CS} = 0.003$, $\alpha_{BR} = 0.009$, $\alpha_{BU} = 0.004$, $\Phi_i = 1$ for $i = EO, TI, D, DS, FSS, M, LP, EA, OVH, CS, BR, BU$, $\theta = \eta = 1$, and $x = y = w = v = m = r = n = k = 1$. Consider that all repair follows exponential time distribution then by substituting these values in equation (26) and taking inverse Laplace transformation, we have

$$\begin{aligned}
 P_{up}(t) = & -0.002992831292 e^{(-0.003000000000 t)} + 0.09089745085 e^{(-1.10098171 t)} - 0.1045972416 \\
 & 10^{(-7)} e^{(-.9969905879 t)} - 0.04625367733 e^{(-0.2043987365 t)} - 0.009328505569 e^{(-0.0047136733 t)} \\
 & + 0.9676775738
 \end{aligned} \tag{34}$$

By, putting $t=0,1, 2, \dots, 10$ in equation (34) we obtain the variation of availability for time shown in figure 2. Similarly, by considering different numerical values for failure rates we get the graphs of reliability and mean time to failure given in figures 3, 4 and 5.

5.1. Cost Analysis

If it is considered that the service facility is always available, then the expected profit function in the interval $(0, t]$ is given by

$$E_p(t) = K_1 \int_0^t P_{up}(t) dt - K_2 t$$

where, K_1 and K_2 are the revenue and service cost per unit of time respectively, then

$$\begin{aligned}
 E P(t) = & K_1 [0.2602109240 e^{(-0.5000000000 t)} + 0.003420894509 e^{(-0.9100000000 t)} + 0.3707166241 \\
 & e^{(-1.577110005 t)} + 0.2282704474 e^{(-0.8601482323 t)} \cos(0.007394121216 t) + 0.2901464146 \\
 & e^{(-0.8601482323 t)} \sin(0.007394121216 t) - 0.05867072106 e^{(-0.7862665523 t)} - 0.2283382234 \\
 & e^{(-0.7285677930 t)} - 0.003350319082 e^{(-0.6042513821 t)} - 9.234941008 e^{(-0.1268411365 t)} \\
 & + 8.996326343] - K_2 t
 \end{aligned} \tag{35}$$

Keeping $K_1 = 1$ and varying K_2 at 0.1, 0.2, 0.3, 0.4, 0.5 in equation (35), one can obtain Figure 6.

6. Results & Conclusion

For the more concrete behavior analysis of the process, a numerical calculation of availability, reliability, mean time to failure for various failure rates and cost function have been made. The following conclusions may be drawn based on the study conducted in the present paper.

Figure 2 shows a rapid decrease in the availability of the system for time initially because of an error in maintenance but later on, it becomes stable. Although, the reliability of the process shown in figure 3 has a constant decline and stabilizes at 0.4. Figures 4 and 5 show the variations in the meantime to failure (MTTF) for different failure rates. As the failure rates increase, the MTTF decreases and also it is observed that it is highest for fuel supply unit failure shown in figure 4 and 5. Now, cost analysis reveals that an increase in service cost results in decreased profit as shown in figure 6.

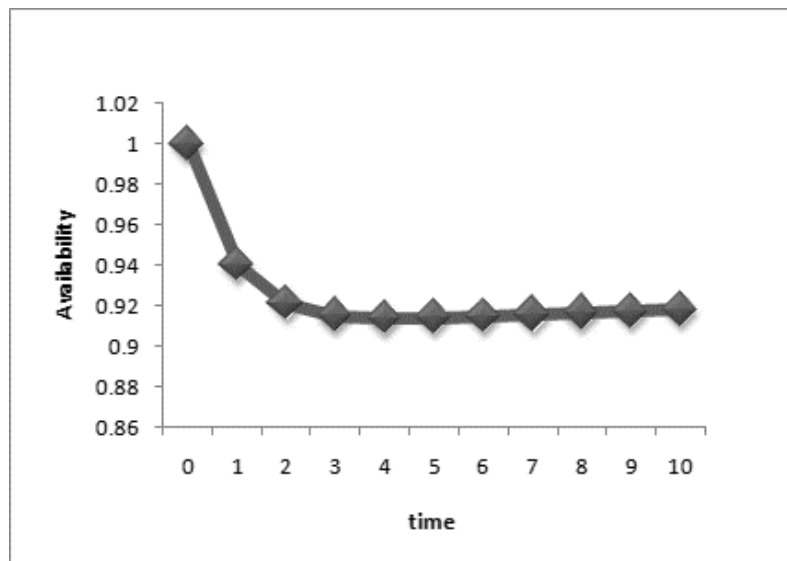


Figure 2: Time vs. Availability

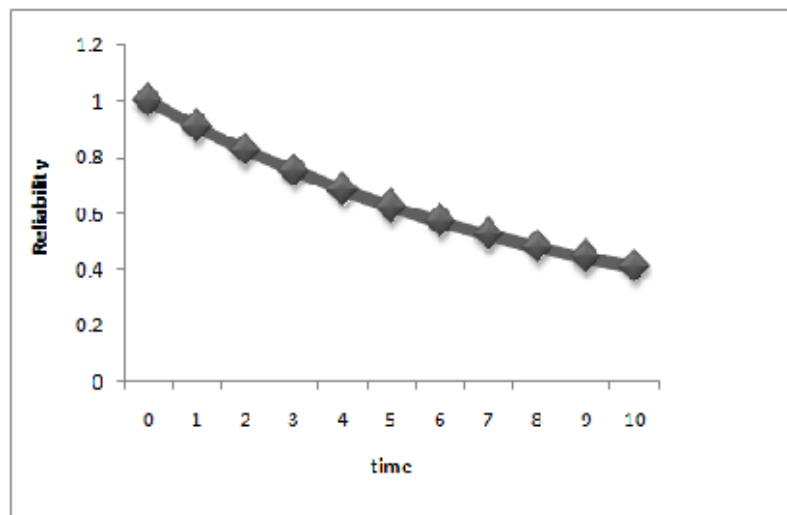


Figure 3: Time vs. Reliability

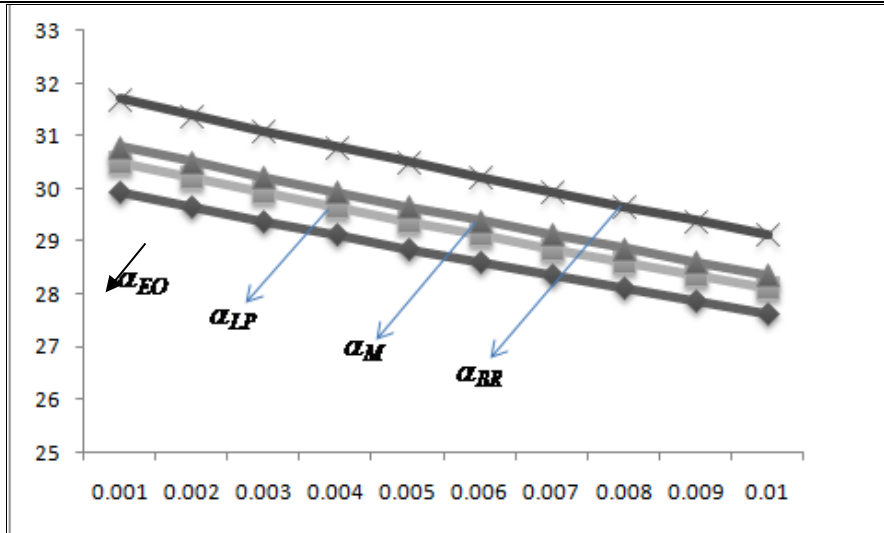


Figure 4: $MTTF$ vs. α_{EO} , α_{LP} , α_M , α_{BR}

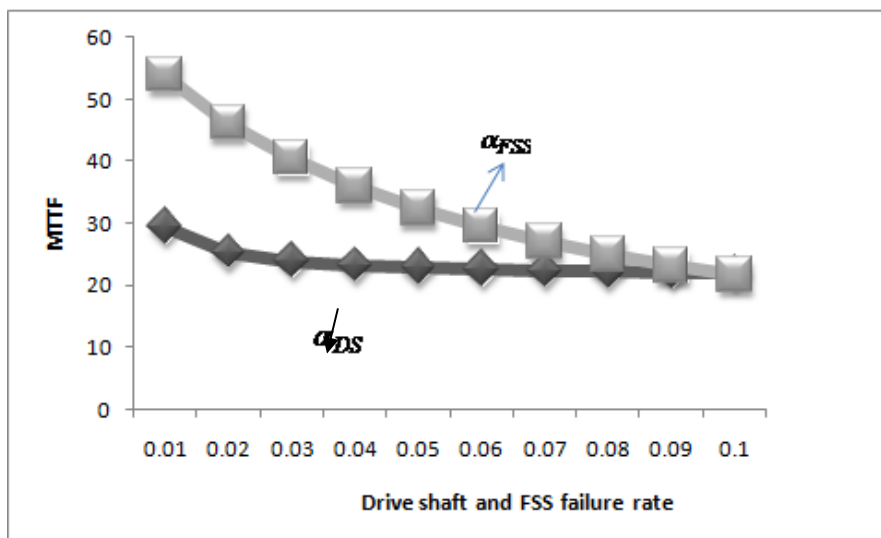


Figure 5: $MTTF$ vs. α_{DS} , α_{FSS}

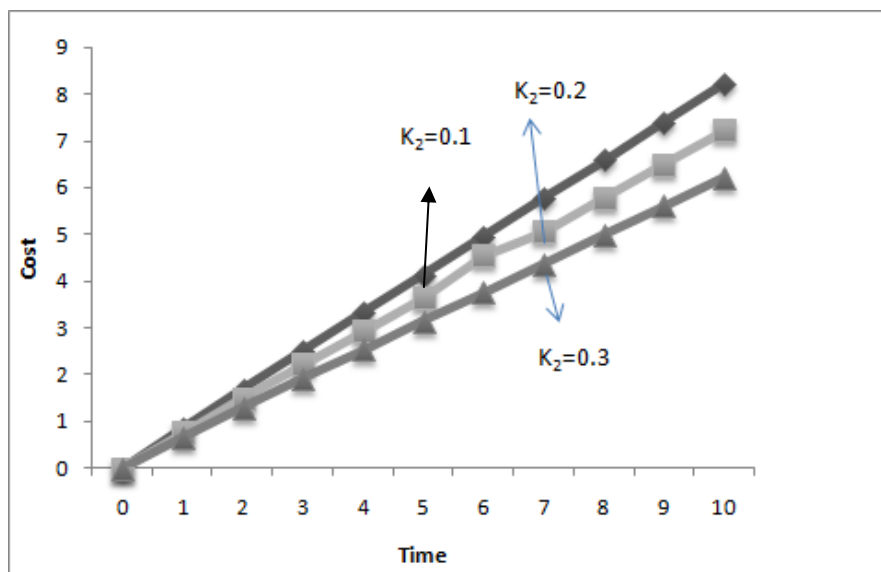


Figure 6: $Time$ vs. $cost$

References

- [1] Bao, Y., Mays, L. W. (1990). Model for water distribution system reliability. *Journal of Hydraulic Engineering*, 116, 1119-1137.
- [2] Gnedenko B. and Igor A.U. Probabilistic Reliability Engineering, 1995.
- [3] Gnedenko B, Igor V.P., Igor A. U. and Chakravarty S. Statistical Reliability Engineering, first edition, 1999.
- [4] Firas S. and Walid A. K. (2018). Reliability Assessment of Buffer less Production System: A Boolean Algebra and K-Map Approach, Proceedings of the International Conference on Industrial Engineering and Operations Management Washington DC, USA.
- [5] Henryk T., Mariusz Z., Mariusz W. (2016). A method for reliability assessment of structural components of aircraft and sea going ships with taking into account given failure generation model, *POLISH MARITIME RESEARCH*, 2 (90), 23: 83-90.
- [6] Ioannis S. T. and Chalikias M. (2021). On Reliability Structures with Two Common Failure Criteria Under Age-Based Maintenance Policy. *Reliability: theory & applications*, 2(64): 27-34.
- [7] Ram, M. and Singh, S. B. (2010). Availability, MTTF, and cost analysis of the complex system under pre-emptive-repeat repair discipline using Gumbel-Hougaard family copula. *International Journal of Quality & Reliability Management*, 27(5): 576-595.
- [8] Sengar, S. and Singh, S.B. (2012). Operational behavior and reliability measures of a viscose staple fiber plant including deliberate failures. *International journal of reliability and applications*, 13 (1): 1-17.
- [9] Sengar, S. (2020). Sensitivity analysis of a complex manufacturing system with the application of wait-in-line theory. *Reliability: theory & applications*, 4(59): 27-34.
- [10] Sengar, S., Mangay, R., & Yigit, K. (2021). Stochastic analysis of complex redundant system having problem of waiting line in repair using copula methodology. *Reliability: theory & applications*, 16:383-391.
- [11] Sengar S. Reliability analysis and cost optimization in manufacturing systems, System reliability Engineering, Walter de-Gruyter publication, 2021.
- [12] Sengar S. and Mangey R. (2022). Reliability and Performance Analysis of a Complex Manufacturing System with Inspection facility using Copula Methodology. *Reliability: Theory & Applications*. 4(71): 494-508.
- [13] Wawrzynski, W., Zieja, M., Tomaszewski, J., Michalski, M. (2021). Reliability Assessment of Aircraft Commutators. *Energies*, 14, 7404.
- [14] Wasiu A. A. and Olalekan O. (2014). Reliability Analysis of a 3-Machine Power Station Using State Space Approach. *International Journal of Engineering Research and Applications*, 4(7): 208-217.

ESTIMATION OF A PARAMETER OF FARLIE-GUMBEL-MORGENSTERN BIVARIATE BILAL DISTRIBUTION BY RANKED SET SAMPLING

M. R. IRSHAD¹, R. MAYA², A.I. AL-OMARI³, AHMAD A. HANANDEH⁴, S. P. ARUN⁵

•
¹Cochin University of Science and Technology, Kochi-22, Kerala.
irshadm24@gmail.com

²University College, Trivandrum-695 034.
publicationsofmaya@gmail.com

³Department of Mathematics, Faculty of Science, Al al-Bayt University, Mafraq, Jordan.
alomari_amer@yahoo.com

⁴Yarmouk University, Irbid 21163, Jordan.
ahmad.hanandeh@yu.edu.jo

⁵University Library, Research Centre, University of Kerala, Trivandrum-695 034.
arunsptvm@gmail.com

Abstract

A bivariate version of the Bilal distribution has been proposed in the literature, called the Farlie-Gumbel-Morgenstern bivariate Bilal (FGMBB) distribution. In this article, we have dealt with the problem of estimation of the scale parameter associated with the study variable Z of primary interest, based on the ranked set sample defined by ordering the marginal observations on an auxiliary variable W, when (W, Z) follows a FGMBB distribution. When the dependence parameter ϕ is known, we have proposed the following estimators, viz., an unbiased estimator based on the Stoke's ranked set sample and the best linear unbiased estimator based on the Stoke's ranked set sample for the scale parameter of the variable of primary interest. The efficiency comparison of the proposed estimators with respect to the maximum likelihood estimator have been carried out.

Keywords: Farlie-Gumbel-Morgenstern bivariate Bilal distribution, Concomitants of order statistics, Ranked set sampling, Best linear unbiased estimator

1. INTRODUCTION

The Bilal distribution was introduced by [1], as a member of the families of distributions for the median of a random sample arising from an arbitrary lifetime distribution. Also, he shows that, this distribution belongs to the class of new better than average renewal failure rates and its probability density function (pdf) is always unimodal and has less of skewness and kurtosis than the pdf of the exponential distribution by about 25% and 28% respectively. The cumulative distribution function (cdf) of the Bilal distribution with the scale parameter σ is given by

$$F(x; \sigma) = 1 - e^{-\frac{2x}{\sigma}} \left(3 - 2e^{-\frac{x}{\sigma}} \right); \sigma > 0, x > 0. \quad (1)$$

The corresponding pdf is given by

$$f(x; \sigma) = \frac{6}{\sigma} e^{-\frac{2x}{\sigma}} \left(1 - e^{-\frac{x}{\sigma}} \right); \sigma > 0, x > 0. \quad (2)$$

Furthermore, the author obtained the closed form expressions for the quantile function, the hazard rate function and simple expression for moments in terms of the exponential function. Even though the Bilal distribution has only one parameter, this distribution possess high fitting ability compared to other competing models for two different real datasets, namely, the dataset consisting of thirty successive values of precipitation (in inches) given by [14] and the data for waiting times before service of 100 bank customers reported by [13]. Based on type-2 censored sample, [2] provide certain estimators of the parameter of the Bilal distribution. According to [3], the one parameter Bilal model can be generalized into the two parameter Bilal model, whose applications are elaborately discussed. Now the Proficiency of univariate Bilal distribution compared to other competing models well established in the literature in the theoretical as well as applied perspective. But even a single work is not been seen so far in the available literature on bivariate Bilal model except the work of [17]. A bivariate extension of one parameter Bilal distribution using Morgenstern approach was proposed by [17], so-called the Farlie-Gumbel-Morgenstern Bivariate Bilal (FGMBB) distribution and elucidated its inferential aspects using concomitants of order statistics (COS).

A bivariate random variable (W, Z) is said to follow a FGMBB distribution, if its pdf is given by

$$f(w, z) = \begin{cases} \frac{36}{\sigma_1 \sigma_2} e^{-\frac{2w}{\sigma_1}} \left(1 - e^{-\frac{w}{\sigma_1}}\right) e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \\ \times \left[1 + \phi \left(2e^{-\frac{2w}{\sigma_1}} \left\{3 - 2e^{-\frac{w}{\sigma_1}}\right\} - 1\right) \left(2e^{-\frac{2z}{\sigma_2}} \left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\} - 1\right)\right], & (3) \\ w > 0, z > 0; \sigma_1 > 0, \sigma_2 > 0; -1 \leq \phi \leq 1. \\ 0, & \text{otherwise.} \end{cases}$$

Clearly the marginal distributions of W and Z variables are univariate Bilal distributions with pdf's are respectively given by

$$f_W(w) = \begin{cases} \frac{6}{\sigma_1} e^{-\frac{2w}{\sigma_1}} \left(1 - e^{-\frac{w}{\sigma_1}}\right); & \text{if } \sigma_1 > 0, w > 0, \\ 0, & \text{otherwise.} \end{cases}$$

and

$$f_Z(z) = \begin{cases} \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right); & \text{if } \sigma_2 > 0, z > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Clearly,

$$E(W) = \frac{5}{6}\sigma_1, \text{Var}(W) = \frac{13}{36}\sigma_1^2, \quad (5)$$

$$E(Z) = \frac{5}{6}\sigma_2, \quad (5)$$

$$\text{Var}(Z) = \frac{13}{36}\sigma_2^2. \quad (6)$$

The ranked set sampling (RSS) scheme was first developed by [19] as a process of increasing the precision of the sample mean as an estimator of the population mean. McIntyre's idea of ranking is possible whenever it can be done easily by a judgement method. For a detailed discussion on the theory and applications of RSS [11]. Basically the procedure involves choosing n sets of units, each of size n , and ordering the units of each of the set by judgement method or by applying some inexpensive method, without making actual measurement on the units. Then the unit ranked as one from the 1st set is actually measured, the unit ranked as two from the 2nd set is measured. The process continuous in this way until the unit ranked as n from the n^{th} set is measured. Then the observations obtained under the afore mentioned criterion is known as ranked set sample (*rss*) and the procedure is known as RSS. For recent developments in RSS, one can refer [6], [4] and [5].

In some practical problems, the variable of primary concern say Z , is more intricate to measure, but an auxiliary variable W related with Z is easily measurable and can be ordered exactly. In this case, [22] developed another scheme of RSS, which is as follows: Choose n independent bivariate sets, each of size n . In the first set of size n , the Z variate associated with smallest ordered W is measured, in the second set of size n , the Z variate associated with the second smallest, W is measured. This process is continued until the Z associated with the largest W from the n^{th} set is measured. The measurements on the Z variate of the resulting new set of n units chosen by the above method gives a rss as suggested by [22]. If $W_{(r:n)r}$ is the observation measured on the auxiliary variable W from the unit chosen from the r^{th} set, then we write $Z_{[r:n]r}$ to denote the corresponding measurement made on the study variable Z on this unit so that $Z_{[r:n]r}, r = 1, 2, \dots, n$ form the rss . $Z_{[r:n]r}$ was referred by [12] as the concomitant of the r^{th} order statistic arising from the r^{th} sample.

The rss mean as an estimator for the mean of the study variate Z , when an auxiliary variable W is used for ranking the sample units has suggested by [22], under the assumption that (W, Z) follows a bivariate normal distribution. Based on rss obtained on the study variate Z , [10] have improved the estimator of [22] by deriving the best linear unbiased estimator (BLUE) of the mean of the study variate Z . COS and its applications in RSS from Farlie-Gumbel-Morgenstern bivariate Lomax distribution is elaborately elucidated by [20]. The estimation of a parameter of Morgenstern type bivariate Lindley distribution by RSS has been discussed in [15]. Parameter estimation of Cambanis-type bivariate uniform distribution with RSS is studied by [16]. For review of various variants of RSS and their application in parameter estimation [11].

The remaining part of this paper is assembled as follows. In section 2, we have proposed an unbiased estimator σ_2^* of σ_2 using Stoke's rss . As mentioned earlier if (W, Z) has a FGMBB distribution as defined in (3), then the marginal distributions of both W and Z have Bilal distributions and the pdf of Z is given in (4). We have evaluated the Cramer-Rao Lower Bound (CRLB) for the variance of an unbiased estimator of σ_2 involved in (4) based on a random sample of size n and is given by $\frac{13}{25} \frac{\sigma_2^2}{n}$. In this section, we have also shown that the variance of proposed unbiased estimator σ_2^* is strictly less than $\frac{13}{25} \frac{\sigma_2^2}{n}$, the CRLB for the variance of an unbiased estimator of σ_2 involved in (4), for all $\phi \in B$, where $B = [-1, 1] - \{0\}$. In this section, we have further discussed an efficiency comparison between σ_2^* and the maximum likelihood estimator (MLE) $\hat{\sigma}_2$ of σ_2 based on a random sample of size n arising from (3). In section 3, we have derived the BLUE $\tilde{\sigma}_2$ of σ_2 involved in FGMBB distribution based on Stoke's rss and made an efficiency comparison of $\tilde{\sigma}_2$ relative to $\hat{\sigma}_2$.

2. AN UNBIASED ESTIMATOR OF σ_2 USING STOKE'S RSS.

Suppose the bivariate random vector (W, Z) follows a FGMBB distribution with pdf given in (3). Select a rss as per Stoke's RSS scheme. Let $W_{(r:n)r}$ be the observation obtained on the auxiliary variate W in the r^{th} unit of the rss and let $Z_{[r:n]r}$ be the measurement made on the variate related with $W_{(r:n)r}, r = 1, 2, \dots, n$. Clearly $Z_{[r:n]r}$ is the r^{th} COS of a random sample of size n arising from the FGMBB distribution. Using the results of [21], we obtain the pdf of $Z_{[r:n]r}, r = 1, 2, \dots, n$, and is given by

$$f_{[r:n]}(z) = \frac{6}{\sigma_2} e^{-\frac{2z}{\sigma_2}} \left(1 - e^{-\frac{z}{\sigma_2}}\right) \left[1 + \phi \frac{(n-2r+1)}{(n+1)} \left(2e^{-\frac{2z}{\sigma_2}} \left\{3 - 2e^{-\frac{z}{\sigma_2}}\right\} - 1\right)\right]. \quad (7)$$

The mean and variance of $Z_{[r:n]r}$ for $r = 1, 2, \dots, n$, is obtained as

$$E[Z_{[r:n]r}] = \sigma_2 \left[\frac{5}{6} - \frac{19}{60} \phi \frac{(n-2r+1)}{(n+1)}\right] \quad (8)$$

and

$$Var[Z_{[r:n]r}] = \sigma_2^2 \left[\frac{13}{36} - \frac{253}{1800} \phi \frac{(n-2r+1)}{(n+1)} - \frac{361}{3600} \phi^2 \frac{(n-2r+1)^2}{(n+1)^2}\right]. \quad (9)$$

Since $Z_{[r:n]r}$ and $Z_{[s:n]s}$ for $r \neq s$ are arising from two independent samples, we obtain

$$Cov[Z_{[r:n]r}, Z_{[s:n]s}] = 0, r \neq s.$$

Next, we derive an unbiased estimator of σ_2 and its variance using the *rss* observations $Z_{[r:n]r}$ for $r = 1, 2, \dots, n$, on the variable Z of primary interest and are given by the following theorem.

Theorem 1. Let (W, Z) follows a FGMBB distribution with pdf given by (3). Let $Z_{[r:n]r}$, $r = 1, 2, \dots, n$ be the *rss* observations on a study variate Z generated out of ranking made on an auxiliary variate W . Then

$$\sigma_2^* = \frac{6}{5n} \sum_{r=1}^n Z_{[r:n]r}$$

is an unbiased estimator of σ_2 and its variance is given by

$$Var[\sigma_2^*] = \frac{\sigma_2^2}{n} \left[\frac{13}{25} - \frac{361}{2500} \frac{\phi^2}{n} \sum_{r=1}^n \left(\frac{n-2r+1}{n+1} \right)^2 \right]. \quad (10)$$

Proof By using the definition, we have

$$\begin{aligned} E[\sigma_2^*] &= \frac{6}{5n} \sum_{r=1}^n E[Z_{[r:n]r}] \\ &= \frac{6}{5n} \sum_{r=1}^n \left[\frac{5}{6} - \frac{19}{60} \phi \frac{(n-2r+1)}{(n+1)} \right] \sigma_2. \end{aligned} \quad (11)$$

Using the result,

$$\sum_{r=1}^n (n-2r+1) = 0. \quad (12)$$

Applying (12) in (11) we get,

$$E[\sigma_2^*] = \sigma_2.$$

Therefore, σ_2^* is an unbiased estimator of σ_2 . The variance of σ_2^* is given by,

$$Var[\sigma_2^*] = \frac{36}{25n^2} \sum_{r=1}^n Var[Z_{[r:n]r}]. \quad (13)$$

Applying (9) and (12) in (13), we get

$$Var[\sigma_2^*] = \frac{\sigma_2^2}{n} \left[\frac{13}{25} - \frac{361}{2500} \frac{\phi^2}{n} \sum_{r=1}^n \left(\frac{n-2r+1}{n+1} \right)^2 \right].$$

Hence the proof.

As mentioned above, if (W, Z) has the FGMBB distribution as defined in (3), then the marginal distribution of both W and Z are Bilal distributions and the pdf of Z is given in (4). The CRLB for the variance of any unbiased estimator of σ_2 based on a random sample of size n drawn from (4) is obtained as $\frac{13}{25} \frac{\sigma_2^2}{n}$. Now we compare the the variance of σ_2^* with the CRLB for the variance of an unbiased estimator of σ_2 involved in (4). If we write $E_1(\sigma_2^*)$ to denote the ratio of $\frac{13}{25} \frac{\sigma_2^2}{n}$ with $Var(\sigma_2^*)$, then we have,

$$E_1(\sigma_2^*) = \frac{1}{\left[1 - \frac{361}{1300} \frac{\phi^2}{n} \sum_{r=1}^n \left(\frac{n-2r+1}{n+1} \right)^2 \right]}. \quad (14)$$

It is easily verified that

$$E_1(\sigma_2^*) \geq 1.$$

Thus we arrive at a conclusion that the estimator σ_2^* based on Stoke's *rss* is more efficient as it assert the statement that *rss* always provide more information than simple random sample even if ranking is imperfect [11]. It is very clear that $Var(\sigma_2^*)$ is a decreasing function of ϕ^2 and hence the gain in efficiency of the estimator σ_2^* increases as $|\phi|$ increases. Again on simplifying (14) we get,

$$E_1(\sigma_2^*) = \frac{1}{1 - \frac{361\phi^2}{1300} \left[\frac{2}{3} \left(\frac{2+1/n}{1+1/n} \right) - 1 \right]}.$$

Then,

$$\begin{aligned} \lim_{n \rightarrow \infty} E_1(\sigma_2^*) &= \lim_{n \rightarrow \infty} \frac{1}{1 - \frac{361\phi^2}{1300} \left[\frac{2}{3} \left(\frac{2+1/n}{1+1/n} \right) - 1 \right]} \\ &= \frac{1}{1 - \frac{361\phi^2}{3900}}. \end{aligned}$$

From the above expression it is clear that the maximum value for $E_1(\sigma_2^*)$ is attained when $|\phi| = 1$ and in this case $E_1(\sigma_2^*)$ tends to $3900/3539$.

Next we discuss the efficiency comparison of σ_2^* with the asymptotic variance of MLE of σ_2 involved in the FGMBB distribution. If (W, Z) follows a FGMBB distribution with pdf given in (3), then

$$\begin{aligned} \frac{\partial \log f(x, y)}{\partial \sigma_1} &= \frac{1}{\sigma_1} \left\{ -1 + \frac{2w}{\sigma_1} - \frac{we^{-\frac{w}{\sigma_1}}}{\sigma_1(1 - e^{-\frac{w}{\sigma_1}})} \right. \\ &\quad \left. + \frac{4\phi we^{-\frac{2w}{\sigma_1}} \left[-3 + 18e^{-\frac{2z}{\sigma_2}} - 12e^{-\frac{3z}{\sigma_2}} + 3e^{-\frac{w}{\sigma_1}} - 18e^{-\frac{w}{\sigma_1}} e^{-\frac{2z}{\sigma_2}} + 12e^{-\frac{w}{\sigma_1}} e^{-\frac{3z}{\sigma_2}} \right]}{\sigma_1 \left\{ 1 + \phi \left[1 - 2e^{-\frac{2w}{\sigma_1}} (3 - 2)e^{-\frac{w}{\sigma_1}} \right] \left[1 - 2e^{-\frac{2z}{\sigma_2}} (3 - 2)e^{-\frac{z}{\sigma_2}} \right] \right\}} \right\} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \log f(x, y)}{\partial \sigma_2} &= \frac{1}{\sigma_2} \left\{ -1 + \frac{2z}{\sigma_2} - \frac{ze^{-\frac{z}{\sigma_2}}}{\sigma_2(1 - e^{-\frac{z}{\sigma_2}})} \right. \\ &\quad \left. + \frac{4\phi ze^{-\frac{2z}{\sigma_2}} \left[-3 + 18e^{-\frac{2w}{\sigma_1}} - 12e^{-\frac{3w}{\sigma_1}} + 3e^{-\frac{z}{\sigma_2}} - 18e^{-\frac{z}{\sigma_2}} e^{-\frac{2w}{\sigma_1}} + 12e^{-\frac{z}{\sigma_2}} e^{-\frac{3w}{\sigma_1}} \right]}{\sigma_2 \left\{ 1 + \phi \left[1 - 2e^{-\frac{2w}{\sigma_1}} (3 - 2)e^{-\frac{w}{\sigma_1}} \right] \left[1 - 2e^{-\frac{2z}{\sigma_2}} (3 - 2)e^{-\frac{z}{\sigma_2}} \right] \right\}} \right\}. \end{aligned}$$

Then we have,

$$\begin{aligned} I_{\sigma_1}(\phi) &= E \left(\frac{\partial \log f(x, y)}{\partial \sigma_1} \right)^2 \\ &= \frac{36}{\sigma_1^2} \int_0^\infty \int_0^\infty e^{-2u} (1 - e^{-u}) \left\{ -1 + 4u + ue^{-u} (u - 2 + 2e^{-u}) (1 - e^{-u})^{-2} \right. \\ &\quad \left. - \frac{12\alpha u^2 e^{-2u} [-2 + 12e^{-2v} - 8e^{-3v} + 3e^{-u} - 18e^{-u} e^{-2v} + 12e^{-u} e^{-3v}]}{\{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\}} \right. \\ &\quad \left. + \frac{24\alpha ue^{-2u} [-1 + 6e^{-2v} - 4e^{-3v} + e^{-u} - 6e^{-u} e^{-2v} + 4e^{-u} e^{-3v}]}{\{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\}^2} \right\} \\ &\quad \times \{1 + \alpha[1 - 2e^{-2v}(3 - 2e^{-v})][1 - 6e^{-2u} - 6ue^{-2u} + 4e^{-3u} + 6ue^{-3u}]\} \\ &\quad \left. \right\} e^{-2v} (1 - e^{-v}) \{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\} dudv, \end{aligned}$$

$$\begin{aligned}
 I_{\sigma_2}(\phi) &= E \left(\frac{\partial \log f(x,y)}{\partial \sigma_2} \right)^2 \\
 &= \frac{36}{\sigma_2^2} \int_0^\infty \int_0^\infty e^{-2u} (1 - e^{-u}) \left\{ -1 + 4v + ve^{-v}(v - 2 + 2e^{-v})(1 - e^{-v})^{-2} \right. \\
 &\quad - \frac{12\alpha v^2 e^{-2v} [-2 + 12e^{-2u} - 8e^{-3u} + 3e^{-v} - 18e^{-v}e^{-2u} + 12e^{-v}e^{-3u}]}{\{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\}} \\
 &\quad + \frac{24\alpha v e^{-2v} [-1 + 6e^{-2u} - 4e^{-3u} + e^{-v} - 6e^{-v}e^{-2u} + 4e^{-v}e^{-3u}]}{\{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\}^2} \\
 &\quad \times \{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 6e^{-2v} - 6ve^{-2v} + 4e^{-3v} + 6ve^{-3v}]\} \\
 &\quad \left. \right\} e^{-2v} (1 - e^{-v}) \{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\} dudv
 \end{aligned}$$

and

$$\begin{aligned}
 I_{\sigma_1\sigma_2}(\phi) &= E \left(\frac{\partial^2 \log f(x,y)}{\partial \sigma_1 \partial \sigma_2} \right) \\
 &= \frac{36}{\sigma_1 \sigma_2} \int_0^\infty \int_0^\infty e^{-2u} (1 - e^{-u}) \left\{ 144\alpha u v e^{-2u} e^{-2v} [1 - e^{-u} - e^{-v} + e^{-u}e^{-v}] \right. \\
 &\quad - \frac{144\alpha^2 u v e^{-2u} e^{-2v} [-1 + 6e^{-2v} - 4e^{-3v} + e^{-u} - 6e^{-u}e^{-2v} + 4e^{-u}e^{-3v}]}{\{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\}^2} \\
 &\quad \times \{[1 - 2e^{-2u}(3 - 2e^{-u})][e^{-v} - 1]\} \\
 &\quad \left. \right\} e^{-2v} (1 - e^{-v}) \{1 + \alpha[1 - 2e^{-2u}(3 - 2e^{-u})][1 - 2e^{-2v}(3 - 2e^{-v})]\} dudv.
 \end{aligned}$$

Thus the Fisher information matrix associated with the random variable (W, Z) is given by,

$$I(\phi) = \begin{bmatrix} I_{\sigma_1}(\phi) & -I_{\sigma_1\sigma_2}(\phi) \\ -I_{\sigma_1\sigma_2}(\phi) & I_{\sigma_2}(\phi) \end{bmatrix}. \tag{15}$$

We have computed the values of $\sigma_1^{-2}I_{\sigma_1}(\phi)$ and $\sigma_1^{-1}\sigma_2^{-1}I_{\sigma_1\sigma_2}(\phi)$ numerically for $\phi = \pm 0.25, \pm 0.50, \pm 0.75, \pm 1$ (clearly $\sigma_1^{-2}I_{\sigma_1}(\phi) = \sigma_2^{-2}I_{\sigma_2}(\phi)$) and are given below:

ϕ	$\sigma_1^{-2}I_{\sigma_1}(\phi)$	$\sigma_1^{-1}\sigma_2^{-1}I_{\sigma_1\sigma_2}(\phi)$	ϕ	$\sigma_1^{-2}I_{\sigma_1}(\phi)$	$\sigma_1^{-1}\sigma_2^{-1}I_{\sigma_1\sigma_2}(\phi)$
0.25	1.9381	0.1373	-0.25	1.9381	-0.1373
0.50	1.9795	0.2772	-0.50	1.9795	-0.2773
0.75	2.0530	0.4230	-0.75	2.0530	-0.4236
1.00	2.1705	0.5815	-1.00	2.1705	-0.5841

Thus from (15), the asymptotic variance of the MLE $\hat{\sigma}_2$ of σ_2 involved in the FGMBB distribution under a bivariate sample of size n is obtained as

$$Var(\hat{\sigma}_2) = \frac{1}{n} I_{\sigma_2}^{-1}(\phi), \tag{16}$$

where $I_{\sigma_2}^{-1}(\phi)$ is the (2,2)th element of the inverse of $I(\phi)$ given by (15).

We have compute the efficiency $E(\sigma_2^*|\hat{\sigma}_2) = \frac{Var(\hat{\sigma}_2)}{Var(\sigma_2^*)}$ of σ_2^* relative to $\hat{\sigma}_2$ for $n = 2(2)20$; $\phi = \pm 0.25, \pm 0.50, \pm 0.75, \pm 1$ and are given in table 1. From the table, one can infer that the estimator σ_2^* is more efficient than $\hat{\sigma}_2$ and efficiency increases with n and $|\phi|$ for $n \geq 4$.

Remark 2.1. For given value of $\phi \in (0, 1]$, once the variance of σ_2^* is evaluated, then this variance is equal to the variance of σ_2^* for $-\phi$ because the variance given in (10) depends only on ϕ by a term containing ϕ^2 only.

3. BEST LINEAR UNBIASED ESTIMATOR OF σ_2 USING STOKE'S RSS

In this section we derive the BLUE of σ_2 provided the dependence parameter ϕ is known. Suppose $Z_{[r:n]r}$ for $r = 1, 2, \dots, n$, are the rss observation generated from (3) as per Stoke's RSS scheme. Let

$$\zeta_{r,n} = \frac{5}{6} - \frac{19}{60}\phi \frac{(n-2r+1)}{(n+1)}, \tag{17}$$

$$\psi_{r,r,n} = \frac{13}{36} - \frac{253}{1800}\phi \frac{(n-2r+1)}{(n+1)} - \frac{361}{3600}\phi^2 \frac{(n-2r+1)^2}{(n+1)^2}. \tag{18}$$

Using (17) and (18), we get

$$E[Z_{[r:n]r}] = \sigma_2 \zeta_{r,n}, \quad 1 \leq r \leq n \tag{19}$$

and

$$Var[Z_{[r:n]r}] = \sigma_2^2 \psi_{r,r,n}, \quad 1 \leq r \leq n. \tag{20}$$

Also we have

$$Cov[Z_{[r:n]r}, Z_{[s:n]s}] = 0, \quad r, s = 1, 2, \dots, n \text{ and } r \neq s. \tag{21}$$

Let $\mathbf{Z}_{[n]} = (Z_{[1:n]1}, Z_{[2:n]2}, \dots, Z_{[n:n]n})'$ denote the column vector of n rss observations. Then from (19), (20) and (21), we can write,

$$E[\mathbf{Z}_{[n]}] = \sigma_2 \boldsymbol{\zeta} \tag{22}$$

and the dispersion matrix of $\mathbf{Z}_{[n]}$,

$$D[\mathbf{Z}_{[n]}] = \sigma_2^2 \mathbf{G}, \tag{23}$$

where $\boldsymbol{\zeta} = (\zeta_{1,n}, \zeta_{2,n}, \dots, \zeta_{n,n})'$ and $\mathbf{G} = \text{diag}(\psi_{1,1,n}, \psi_{2,2,n}, \dots, \psi_{n,n,n})$, where $\zeta_{r,n}$ and $\psi_{r,r,n}$ for $r = 1, 2, \dots, n$ are respectively given by equations (17) and (18). If ϕ contained in $\boldsymbol{\zeta}$ and \mathbf{G} are known then (22) and (23) together defines a generalized Gauss-Markov setup and then the BLUE of σ_2 is given by

$$\tilde{\sigma}_2 = (\boldsymbol{\zeta}' \mathbf{G}^{-1} \boldsymbol{\zeta})^{-1} \boldsymbol{\zeta}' \mathbf{G}^{-1} \mathbf{Z}_{[n]}$$

and the variance of σ_2 is given by

$$Var(\tilde{\sigma}_2) = \frac{\sigma_2^2}{\boldsymbol{\zeta}' \mathbf{G}^{-1} \boldsymbol{\zeta}}.$$

On simplifying, we get

$$\tilde{\sigma}_2 = \frac{\sum_{r=1}^n \frac{\zeta_{r,n}}{\psi_{r,r,n}} Z_{[r:n]r}}{\sum_{r=1}^n \frac{\zeta_{r,n}^2}{\psi_{r,r,n}}} \tag{24}$$

and

$$Var(\tilde{\sigma}_2) = \frac{\sigma_2^2}{\sum_{r=1}^n \frac{\zeta_{r,n}^2}{\psi_{r,r,n}}}. \tag{25}$$

From (24), we have $\tilde{\sigma}_2$ is a linear functions of the rss observations $Z_{[r:n]r}, r = 1, 2, \dots, n$ and hence

$\tilde{\sigma}_2$ can be written as $\tilde{\sigma}_2 = \sum_{r=1}^n a_r Z_{[r:n]r}$, where

$$a_r = \frac{\frac{\zeta_{r,n}}{\psi_{r,r,n}}}{\sum_{r=1}^n \frac{\zeta_{r,n}^2}{\psi_{r,r,n}}}, \quad r = 1, 2, \dots, n.$$

We have evaluated the numerical values of means and variances using the expressions (17) and (18) respectively for $\phi = 0.25(0.25)1$ and for $n = 2(2)20$. Using these values we have evaluated

the variance of BLUE $\tilde{\sigma}_2$ for $\phi = 0.25(0.25)1$ and for $n = 2(2)20$. Also we have computed the ratio $E(\tilde{\sigma}_2|\hat{\sigma}_2) = \frac{Var(\tilde{\sigma}_2)}{Var(\hat{\sigma}_2)}$ to measure the efficiency of our estimator $\tilde{\sigma}_2$ relative to $\hat{\sigma}_2$ for $n = 2(2)20$ and $\phi = 0.25(0.25)1$ and are presented in table 1. From the table it is clear that, BLUE of σ_2 performs well compared to the MLE of σ_2 , namely $\hat{\sigma}_2$.

Remark 3.1. As in the case of variance of an unbiased estimator given in (10), for a given value of $\phi \in (0, 1]$, once the variance of $\tilde{\sigma}_2$ is evaluated, then there is no need to again evaluate the variance of $\tilde{\sigma}_2$ when $\phi = -\phi$. To establish this argument we prove the following theorem.

Theorem 2. Let (W, Z) follows a FGMBB distribution with pdf given by (3). For a given $\phi \in (0, 1]$, $Var[\tilde{\sigma}_2(\phi_0)]$ is the variance of the BLUE $\tilde{\sigma}_2$ of σ_2 involved in the FGMBB distribution, then

$$Var[\tilde{\sigma}_2(-\phi_0)] = Var[\tilde{\sigma}_2(\phi_0)]. \tag{26}$$

Proof The terms $\zeta_{r,n}$ and $\psi_{r,r,n}$ defined by (17) and (18) are functions of ϕ , r and n and hence $\zeta_{r,n}$ and $\psi_{r,r,n}$ can be denoted as $\zeta_{r,n}(\phi)$ and $\psi_{r,r,n}(\phi)$ respectively. From (17) and (18), it is clear that

$$\zeta_{r,n}(\phi) = \zeta_{n-r+1,n}(-\phi), \quad 1 \leq r \leq n \tag{27}$$

and

$$\psi_{r,r,n}(\phi) = \psi_{n-r+1,n-r+1,n}(-\phi), \quad 1 \leq r \leq n. \tag{28}$$

As a consequence of (27) and (28), we get

$$\begin{aligned} Var[\tilde{\sigma}_2(\phi)] &= \frac{\sigma_2^2}{\sum_{r=1}^n \frac{\zeta_{r,n}^2(\phi)}{\psi_{r,r,n}(\phi)}} = \frac{\sigma_2^2}{\sum_{r=1}^n \frac{\zeta_{n-r+1,n}^2(-\phi)}{\psi_{n-r+1,n-r+1,n}(-\phi)}} \\ &= Var[\tilde{\sigma}_2(-\phi)]. \end{aligned}$$

Hence the proof.

Remark 3.2. For FGMBB distribution defined in (3), we have evaluated the correlation coefficient between the two variables and is given by $\rho = \frac{361}{1300}\phi$. But in certain real life situations our assumption that ϕ is known may viewed as unrealistic. Hence if we have a situation with ϕ unknown, we compute the sample correlation τ from $(W_{r:n}, Z_{[r:n]})$ for $r = 1, 2, \dots, n$ and introduce a moment type estimator $\hat{\phi}$ for ϕ as,

$$\hat{\phi} = \begin{cases} -1, & \text{if } \tau < \frac{-361}{1300} \\ \frac{1300}{361}\tau, & \text{if } \frac{-361}{1300} \leq \tau \leq \frac{361}{1300} \\ 1, & \text{if } \tau > \frac{361}{1300}. \end{cases}$$

Table 1: Efficiencies of the estimators σ_2^* and $\tilde{\sigma}_2$ relative to $\hat{\sigma}_2$.

n	ϕ	$e(\sigma_2^* \hat{\sigma}_2)$	$e(\tilde{\sigma}_2 \hat{\sigma}_2)$	ϕ	$e(\sigma_2^* \hat{\sigma}_2)$	$e(\tilde{\sigma}_2 \hat{\sigma}_2)$
2	0.25	0.9992	0.9992	-0.25	0.9992	0.9992
	0.50	0.9984	0.9984	-0.50	0.9984	0.9984
	0.75	0.9957	0.9957	-0.75	0.9957	0.9957
	1.00	0.9849	0.9849	-1.00	0.9850	0.9853
4	0.25	1.0008	1.0008	-0.25	1.0008	1.0008
	0.50	1.0047	1.0047	-0.50	1.0029	1.0029
	0.75	1.0103	1.0111	-0.75	1.0047	1.0047
	1.00	1.0106	1.0139	-1.00	1.0110	1.0147
6	0.25	1.0012	1.0012	-0.25	1.0012	1.0012
	0.50	1.0082	1.0082	-0.50	1.0082	1.0082
	0.75	1.0168	1.0180	-0.75	1.0168	1.0180
	1.00	1.0223	1.0273	-1.00	1.0230	1.0286
	0.25	1.0015	1.0015	-0.25	1.0015	1.0015

n	ϕ	$e(\sigma_2^* \hat{\sigma}_2)$	$e(\bar{\sigma}_2 \hat{\sigma}_2)$	ϕ	$e(\sigma_2^* \hat{\sigma}_2)$	$e(\bar{\sigma}_2 \hat{\sigma}_2)$
8	0.50	1.0094	1.0094	-0.50	1.0094	1.0094
	0.75	1.0192	1.0209	-0.75	1.0192	1.0209
	1.00	1.0282	1.0351	-1.00	1.0300	1.0367
10	0.25	1.0019	1.0019	-0.25	1.0019	1.0019
	0.50	1.0098	1.0098	-0.50	1.0098	1.0098
	0.75	1.0221	1.0241	-0.75	1.0221	1.0241
12	1.00	1.0312	1.0398	-1.00	1.0330	1.0419
	0.25	1.0023	1.0023	-0.25	1.0023	1.0023
	0.50	1.0094	1.0094	-0.50	1.0094	1.0094
14	0.75	1.0242	1.0266	-0.75	1.0242	1.0266
	1.00	1.0376	1.0455	-1.00	1.0376	1.0455
	0.25	1.0000	1.0000	-0.25	1.0000	1.0000
16	0.50	1.0110	1.0110	-0.50	1.0110	1.0110
	0.75	1.0225	1.0254	-0.75	1.0225	1.0254
	1.00	1.0380	1.0472	-1.00	1.0380	1.0472
18	0.25	1.0031	1.0031	-0.25	1.0031	1.0031
	0.50	1.0126	1.0126	-0.50	1.0126	1.0126
	0.75	1.0258	1.0291	-0.75	1.0258	1.0291
20	1.00	1.0403	1.0473	-1.00	1.0403	1.0473
	0.25	1.0035	1.0035	-0.25	1.0035	1.0035
	0.50	1.0106	1.0106	-0.50	1.0106	1.0106
	0.75	1.0291	1.0291	-0.75	1.0291	1.0291
	1.00	1.0415	1.0534	-1.00	1.0415	1.0534
	0.25	1.0000	1.0000	-0.25	1.0000	1.0000
	0.50	1.0118	1.0157	-0.50	1.0118	1.0157
	0.75	1.0242	1.0283	-0.75	1.0242	1.0283
	1.00	1.0420	1.0508	-1.00	1.0420	1.0508

4. ESTIMATION OF σ_2 BASED ON CENSORED RANKED SET SAMPLE

In this section, we obtain some estimators of σ_2 using censored RSS scheme. Suppose k units are censored in the Stoke's RSS scheme, then we may represent the rss observations on the study variate Z as $\delta_1 Z_{[1:n]1}, \delta_2 Z_{[2:n]2}, \dots, \delta_n Z_{[n:n]n}$ where,

$$\delta_i = \begin{cases} 0, & \text{if the } i^{th} \text{ unit is censored,} \\ 1, & \text{otherwise.} \end{cases}$$

and hence $\sum_{i=1}^n \delta_i = n - k$. In this case the usual unbiased estimator of σ_2 is equal to $\frac{6 \sum_{i=1}^n \delta_i Z_{[i:n]i}}{5(n-k)}$. It may be noted that one need not get $\delta_i = 0$ for $i = 1, 2, \dots, k$ and $\delta_i = 1$ for $i = k + 1, k + 2, \dots, n$. Hence if we write $m_i, i = 1, 2, \dots, n - k$ as the integers such that $1 \leq m_1 < m_2 < \dots < m_{n-k}$ and for which $\delta_{m_i}=1$, then,

$$E \left[\frac{6 \sum_{i=1}^n \delta_i Z_{[i:n]i}}{5(n-k)} \right] = \sigma_2 \left[1 - \frac{19\phi}{50(n+1)(n-k)} \sum_{i=1}^{n-k} (n - 2m_i + 1) \right].$$

Thus it is clear that the in the censored case the usual unbiased estimator is not an unbiased estimator of σ_2 . However we can construct an unbiased estimator of σ_2 based on $\frac{6 \sum_{i=1}^n \delta_i Z_{[i:n]i}}{5(n-k)}$ is given in the following theorem.

Theorem 3. Suppose that the random variable (W, Z) has a FGMBB distribution as defined in (3). Let $Z_{[m_i]m_i}, i = 1, 2, \dots, n - k$ be the rss observations on the study variate Z resulting out of censoring applied on the auxiliary variable W . Then an unbiased estimator of σ_2 based on

$\frac{6}{5(n-k)} \sum_{i=1}^{n-k} Z_{[m_i]m_i}$ is given by

$$\sigma_2^*(k) = \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]} \sum_{i=1}^{n-k} Z_{[m_i]m_i}$$

and its variance is given by

$$Var[\sigma_2^*(k)] = \frac{3600(n+1)^2\sigma_2^2}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]^2} \sum_{i=1}^{n-k} \psi_{m_i}$$

where ψ_{m_i} is as defined in (18).

Proof We have

$$\begin{aligned} E[\sigma_2^*(k)] &= \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]} \sum_{i=1}^{n-k} E[Z_{[m_i]m_i}] \\ &= \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]} \\ &\quad \times \sum_{i=1}^{n-k} \left[\frac{5}{6} - \frac{19\phi}{60} \frac{(n-2m_i+1)}{(n+1)} \right] \sigma_2 \\ &= \frac{60(n+1)}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]} \\ &\quad \times \left[\frac{5(n-k)}{6} - \frac{19\phi}{60(n+1)} \sum_{i=1}^{n-k} (n-2m_i+1) \right] \sigma_2 \\ &= \sigma_2. \end{aligned}$$

Thus $\sigma_2^*(k)$ is an unbiased estimator of σ_2 . The variance of $\sigma_2^*(k)$ is given by

$$\begin{aligned} Var[\sigma_2^*(k)] &= \frac{3600(n+1)^2}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]^2} \sum_{i=1}^{n-k} Var(Z_{[m_i]m_i}) \\ &= \frac{3600(n+1)^2\sigma_2^2}{\left[50(n+1)(n-k) - 19\phi \sum_{r=1}^{n-k} (n-2m_r+1)\right]^2} \sum_{i=1}^{n-k} \psi_{m_i} \end{aligned}$$

where ψ_{m_i} is as defined in (18). Hence the theorem.

As a competitor of the estimator $\sigma_2^*(k)$, next we propose the BLUE of σ_2 based on the censored *rss*, resulting out of ranking of observations on W .

If $\mathbf{Z}_{[n]}(k) = (Z_{[m_1]m_1}, Z_{[m_2]m_2}, \dots, Z_{[m_{n-k}]m_{n-k}})'$, then the mean vector and the variance-covariance matrix of $\mathbf{Z}_{[n]}(k)$ are given by

$$E[\mathbf{Z}_{[n]}(k)] = \sigma_2 \boldsymbol{\zeta}(k), \tag{29}$$

$$D[\mathbf{Z}_{[n]}(k)] = \sigma_2 G(k), \tag{30}$$

where $\boldsymbol{\zeta}(k) = (\zeta_{m_1}, \zeta_{m_1}, \dots, \zeta_{m_{n-k}})'$, $G(k) = \text{diag}(\psi_{m_1}, \psi_{m_2}, \dots, \psi_{m_{n-k}})$.

if the parameter ϕ involved in $\boldsymbol{\zeta}(k)$ and $G(k)$ are known then (29) and (30) together defines a generalized Gauss-Markov setup and hence the BLUE $\tilde{\sigma}_2(k)$ of σ_2 is obtained as,

$$\tilde{\sigma}_2(k) = [(\boldsymbol{\zeta}(k))'(G(k))^{-1}\boldsymbol{\zeta}(k)]^{-1}(\boldsymbol{\zeta}(k))'(G(k))^{-1}\mathbf{Z}_{[n]}(k) \tag{31}$$

and the variance of σ_2 is given by

$$Var(\tilde{\sigma}_2(k)) = [(\boldsymbol{\zeta}(k))'(G(k))^{-1}\boldsymbol{\zeta}(k)]^{-1}\sigma_2^2. \tag{32}$$

On substituting the values of $\zeta(k)$ and $G(k)$ in (31) and (32) and simplifying we get,

$$\tilde{\sigma}_2(k) = \frac{\sum_{i=1}^{n-k} (\zeta_{m_i} / \psi_{m_i})}{\sum_{i=1}^{n-k} (\zeta_{m_i}^2 / \psi_{m_i})} Z_{[m_i]m_i} \quad (33)$$

and

$$Var(\tilde{\sigma}_2(k)) = \frac{1}{\sum_{i=1}^{n-k} (\zeta_{m_i}^2 / \psi_{m_i})} \sigma_2^2. \quad (34)$$

Remark 4.1. Since both the BLUE $\tilde{\sigma}_2(k)$ and the unbiased estimator $\sigma_2^*(k)$ based on the censored ranked set sample utilize the distributional property of the parent distribution they lose the usual robustness property. Hence in this case the BLUE $\tilde{\sigma}_2(k)$ shall be considered as a more preferable estimator than $\sigma_2^*(k)$.

REFERENCES

- [1] Abd-Elrahman, A. M. (2013). Utilizing ordered statistics in lifetime distributions production: A new lifetime distribution and applications. *Journal of Probability and Statistical Science*, 11:153–164.
- [2] Abd-Elrahman, A. M. and Niazi, S. F. (2016). Approximate Bayes estimators applied to the Bilal model. *Journal of the Egyptian Mathematical Society*, 1–6.
- [3] Abd-Elrahman, A. M. (2017). A new two-parameter lifetime distribution with decreasing, increasing or upside-down bathtub-shaped failure rate. *Communications in Statistics-Theory and Methods*, 46:8865–8880.
- [4] Al-Omari, A. I. (2021). Maximum likelihood estimation in location-scale families using varied L ranked set sampling. *RAIRO Operations Research*, 55:2759–2771.
- [5] Al-Omari, A. I. and Abdallah, M. S. (2021). Estimation of the distribution function using moving extreme and minimax ranked set sampling. *Communications in Statistics-Simulation and Computation*.
- [6] Al-Omari, A. I. and Almanjahie, I. M. (2021). New improved ranked set sampling design with an application to real data. *Computers, Materials and Continua*.
- [7] Al-Saleh, M. F. (2004). Steady-state ranked set sampling and parametric estimation. *Journal of Statistical planning and Inference*, 123:83–95.
- [8] Al-Saleh, M. F. and Al-Omari, A. I. (2002). Multistage ranked set sampling. *Journal of Statistical planning and Inference*, 273–286.
- [9] Al-Saleh, M. F. and Al-Kadiri, M. (2000). Double-ranked set sampling. *Statistics and Probability Letters*, 205–212.
- [10] Barnett, V. and Moore, K. (1997). Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering. *Communications in Statistics-Theory and Methods*, 24:697–710.
- [11] Chen, Z., Bai, Z. and Sinha, B. K. (2004). *Lecture Notes in Statistics, Ranked Set Sampling: Theory and Applications, Theory and Algorithms*, Springer, New York.
- [12] David, H. A. and Nagaraja, H. N. (2003). *Order statistics: Third edition.*, John Wiley and Sons, New York.
- [13] Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its applications. *Mathematics and Computers in Simulation*, 78:493–506.
- [14] Hinkley, D. (1977). On quick choice of power transformation. *Applied Statistics*, 67–69.
- [15] Irshad, M. R., Maya, R. and Arun S. P. (2019). Estimation of a parameter of Morgenstern type bivariate Lindley distribution by ranked set sampling. *iSTATISTIK: Journal of the Turkish Statistical Association*, 12:25–34.
- [16] Koshti, R. D. and Kamalja, K. K. (2020). Parameter estimation of Cambanis-type bivariate uniform distribution with ranked set sampling. *Journal of Applied Statistics*.

- [17] Maya, R., Irshad, M. R. and Arun, S. P. (2021). Farlie-Gumbel-Morgenstern bivariate Bilal distribution and its inferential aspects using concomitants of order statistics. *Journal of Probability and Statistical Science*, 19:1–20.
- [18] McGhilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, 47:461–466.
- [19] McIntyre, G. A. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3:385–390.
- [20] Philip, A. and Thomas, P. Y. (2017). On concomitants of order statistics and its application in defining ranked set sampling from Farlie-Gumbel Morgenstern bivariate Lomax distribution. *Journal of the Iranian Statistical Society*, 16:67–95.
- [21] Scaria, J. and Nair, N. U. (1999). On concomitants of order statistics from Morgenstern family. *Biometrical Journal*, 41:483–489.
- [22] Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics-Theory and Methods*, 6:1207–1211.

PROCESS MODELING AND NUMERICAL INVESTIGATION OF VENEER CUTTING SYSTEM OF A PLYWOOD PLANT WITH STOCHASTIC APPROACH

Dr. Subhash Malik, Dr. Narendra Kumar, Er. Sudhir Kumar

Maharishi Markandeshwar University, Mullana, India,

Technology Education & Research Integrated Institutions, Kurukshetra, India
subhashmalik604@gmail.com
narendraip005@gmail.com
sudhirtamak@gmail.com

Abstract

Present paper covers performance modelling and performability evaluation of a veneer-cutting system of plywood manufacturing industry. The performability is evaluated as a function of availability. In this system the different subsystems are connected in hybrid mode. Markovian Approach was used for developing the process modeling of the subsystems and to evaluate the performability of said system. MATLAB software was used to perform the numerical computations as well as simulation of results. The current work examines the impact of varied failure rates and repair rates on the long-term availability of the system. A Particle Swarm Optimization (PSO) based technique was used to optimize the results. A Decision Support System (DSS), which can be helpful for making strategic decisions on financial investments in managing the maintenance priorities, spare part management, and human resource requirements, among other things, has been recommended based on the numerical investigation..

Keywords: Availability, DSS, Maintenance, Markov Chain, PSO, RAM Tools

1. Introduction

The manufacturing of plywood is a complex engineering process. It undergoes through several stages like veneer cutting, laying up and gluing operation, hot/cold pressing and trimming process etc. With rapid increase of market competition manufacturers have ensure the progressively improve in their production processes. Use of human labor provides flexibility. Need of varying sizes of the required final product usually interrupt progression in the crucial stage of layout. The availability, cost of production, quality, and, in certain situations, the safety of the operator has all been negatively impacted by the condition. The modern business communities of these fields have taken a positive lesson from it. They grasped as an opportunity to learn from a long list of such failures and their impact in terms of economy and safety.

Regattieri and Bellomi [13] developed a system that reduces the manpower requirements and wastage of materials and improving the operational performance. Use of certain modeling tools in industrial practices can help them in making appropriate decisions on Reliability

Availability and Maintainability (RAM) issues. Various quantitative as well as qualitative tools and techniques are available for these types of analysis. Studies related to RAM facilitates in identifying several maintenance related issues and maintenance planning for smooth system working. Available literature on the subject shows that an equipment maintenance policy generally works in two ways: 1) Corrective Maintenance (CM) which is an offline activity where repair action is taken only after the equipment has failed; and 2) Preventive Maintenance (PM). This plan involves an online maintenance activity well in advance to avoid the frequent failure of system.

In the next sub section usefulness of the certain RAM tools and their typical applications in process industries are suitably discussed. RAM Approaches in Process Industries.

RAM tools reported in the literature in the present study may be classified into two categories namely (i) non state space and (ii) state space models.

The modeling techniques belonging to the non-state space category are Reliability-Block-Diagram, Fault Tree Analysis whereas Petri-Nets (PN) and Markovian processes are coming under the state space models [9,10]. These modeling tools are briefly described as under:

Fault Tree Analysis (FTA): At Bell Telephone Laboratory, this modelling method was created for the first time in 1962. In its original module, a combination of events that might cause a system failure was represented using a visual representation of logical links between events. The system represents top event in the modelling process. Dhillon and Rayapati [3] presented several examples describing successful applications of FTA to modelling of industrial systems. The RAM analysis of a RO desalination plant has also been done by Hajeer and Chaudhuri [4] using it. Its aptitude for managing complex maintenance operations, which are best handled by state-space approaches like Markov or PN formalisms, is one of its significant drawbacks when employed as a RAM analysis tool.

Reliability-Block-Diagram (RBD): Reliability Block Diagram represents the various connections between the components of system. The two forms of series in Reliability Block Diagram are Series and Parallel Configurations. Based upon the Operational Dependency each component in Reliability Bloc Diagram is represented with help of a block i.e. connected either in series, parallel or hybrid mode. RBD Techniques has proved its effectiveness so far in the analysis of reliability of system. The fundamental flow diagram of the process is used to create a high-level dependability block diagram in this.

Khan and Kabir [6] carried out availability analysis of ammonia plant using Reliability Block Diagram is an example of availability analysis of industrial process systems.

Petri Nets (PN) Model: PN Modeling Technique was first used in 1962 by Dr. Carl Adam Petri in his thesis of doctorate. PN uses bipartite directed graph for process modeling of systems having synchronization, randomness and concurrency simultaneously. It has circles to indicate places, bars to denote transitions, and black dots inside the circles to represent tokens [17]. Sachdeva et al [15] applied Petri Nets for the performance modelling and evaluate long run steady state availability of paper manufacturing plant.

Bahl [2] used PN approach for the availability assessment of various systems of a fertilizer industry. More recently, Angel and Jayaparvathy [1] applied PN approach for developing safety system against occurrence of fire. Kumar et al [7,8] performed availability analysis of different repairable industrial units producing different products however similar in operational nature such as randomness, synchronization and concurrency etc.

Markov Process: It is a great process of stochastic behavior used to develop the performance model of systems that exhibit probabilistic behavior. It has many important applications in time-based reliability as well as availability analysis. Here in Movkov state transition diagrams are used for modelling of stochastic behavior of system. The system is capable of a number of distinct states across time. It is possible to specify the speed at which transitions between these states happen. Regardless of the number of states the system passed through to

arrive at this state, the system's transition from one state to another solely depends on the prior one. There are two forms of Markovian Chain models: Discrete-Time-Markov-Chain (DTMC) and Continuous-Time-Markov-Chain (CTMC). Explosion of a number states is the main problem with the markovian chain models which makes difficult to deal with the tedious mathematical calculations.

The work of Singh et al. [16] and Kumar et al. [18,19] uses petri nets for modelling and performance evaluation of subsystems of a thermal power plant. More recently, Malik and Tewari [12] applied Markov Chains for modeling and evaluated availability for different power plants. Kempa [11] dealt with performance modelling of a production system with Markov chains.

Keeping in view of this, in the present study we have considered a Plywood plant facing several challenges as mentioned earlier. The system description and performance modeling is described in subsequent sections here after.

2. System Description

Usually, the plywood manufacturing has nine main steps. These are (i) log collection (ii) debarking (iii) steaming the blocks (iv) peeling blocks into veneers (v) drying veneers (vi) gluing and laying up the veneers one over the another (vii) pressing veneers in a hot press (viii) plywood trimming and (ix) finishing and stamping as shown in Fig.1. The veneer making system is responsible for around 35 to 40 % of the total production of the plant. The system under study is a poplar plywood manufacturer situated in the Ganga basin of Northern India. The various subsystems under consideration are as follows:

- **Debarking Machine:** It used to separate the tree bark and wood without damaging it. Further the logs extracted are into specified lengths. It consists of movable debarking head, roller table, hydraulics mechanism, horizontal and diagonal conveyors and electrically insulated control cabin etc.

- **Veneer Lathe:** In this subsystem a veneer knife cuts the steamed blocks into veneers of desired thickness usually 3mm. The veneer sheets are further clamped to a usable width, to allow for shrinkage and trim. Veneer peeling knives, mechanical drives, tool holders and chucks are the major components of this sub system.

- **Veneer Drier:** It is used to dry the veneers obtained from the barks maintaining the desired level of moisture contents (usually 1 to 15%). This subsystem comprises of heating and cooling components and process measuring devices. A veneer drier typically has three heating zones followed by a cooling section. Heating zone consists of source of hot air, circulating fans and the ports for exhaust which are used reduce the temperature veneer before exiting to the drier.

- **Plywood Scanner:** This subsystem is used to inspect, sort, grade and repair of plywood. It has (i) face and back scanners to detect visual 2-D and 3-D defects (ii) edge scanning for panel layup defects (iii) dimensional scanning for checking length and width and (iv) paralleling and guiding robotic movements for precise sorting and stacking etc.

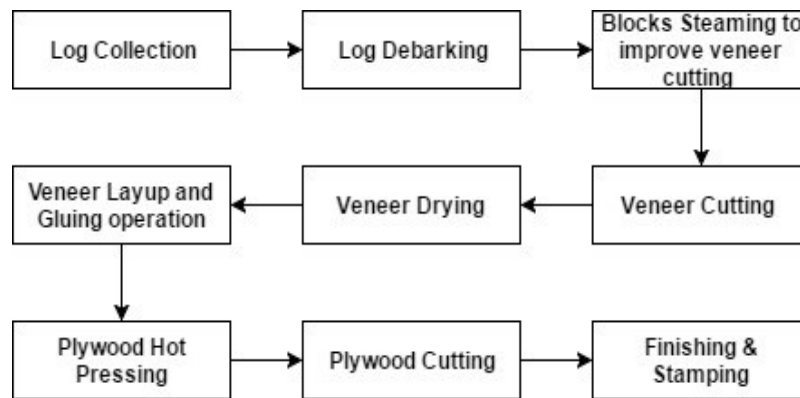


Figure 1: Flow diagram of Plywood Manufacturing Plant

3. Performance Modeling

In this study continuous time Markov Chains have been used to represent the transitions among various subsystems and to develop a performance model of the system. Fig. 2 shows the Markov model of the veneer cutting system. The failure and repair rates, among other variables, were taken into account when modelling the system's performance. The maintenance history books of the plant were obtained in discussion with plant's persons for the required data presented in table 1. Additionally, the following presumptions were used for system modelling and analysis.:

- Exponential distributions have been used to express failure and repair rates.
- A unit is as good as new after repair.
- Standby units have the same nature and capacity as active.
- Only the delay in the availability of repair facilities causes a delay in the start of repairs.
- The system can operate in reduced capacity mode as well.

Notations:

A, B, C and D: :All of the subsystems A, B, C, and D are in fine working order.

A^- :shows that subsystem A is functioning in a reduced state.

B^- :shows that subsystem B is functioning in a reduced state.

C^- :shows a reduced state of operation for subsystem C.

a, b, c and d : shows that A, B, C, and D are all in a failed state, correspondingly.

$\lambda_i, i=1,2,3,\dots,7$: Failure- Rates (FR) from states A, B, C, D, A^- , B^- and C^- to the states A^- , B^- , C^- , d, a, b and c respectively.

$\mu_i, i=1,2,3,\dots,7$: Repair- Rates (RR) from states A^- , B^- , C^- , d, a, b and c to the states A, B, C, D, A, B and C respectively.

$P_j(t), j=1,2,3,\dots,27$: Probability that all subsystems are functioning properly and the system is in the jth state at time t. $P_j'(t)$ represents the derivative of $P_j(t)$ with respect to time 't'.

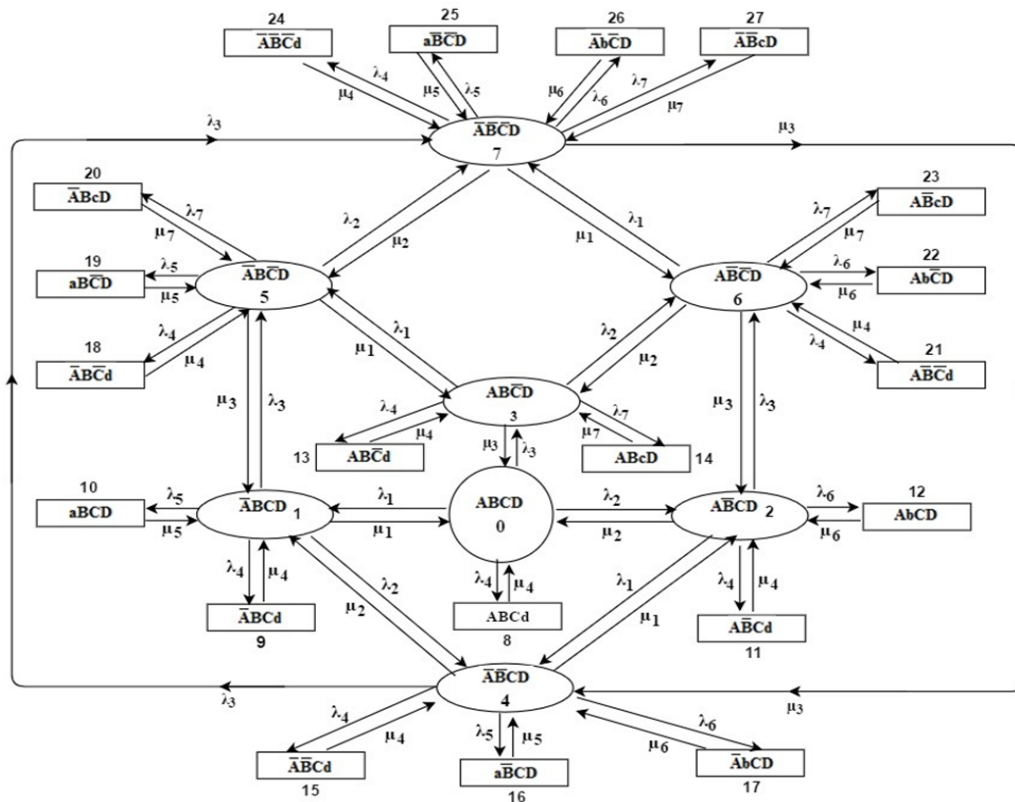


Figure 2: Performance Model of Veneer Cutting System

4. Performance Analysis

Using mnemonic rule a set of first order differential equations related to the transition diagram seen above (Fig.2) of the system at Time (t+Δt) may be written as follows:

$$P_0(t+\Delta t) - P_0(t) = [-\lambda_1 \Delta t - \lambda_2 \Delta t - \lambda_3 \Delta t - \lambda_4 \Delta t] P_0(t) + P_1(t) \mu_1 \Delta t + P_2(t) \mu_2 \Delta t + P_3(t) \mu_3 \Delta t + P_8(t) \mu_4 \Delta t$$

Dividing both sides by Δt, it becomes:

$$[P_0(t+\Delta t) - P_0(t)] / \Delta t = [-\lambda_1 - \lambda_2 - \lambda_3 - \lambda_4] P_0(t) + P_1(t) \mu_1 + P_2(t) \mu_2 + P_3(t) \mu_3 + P_8(t) \mu_4$$

On taking limit as Δt → 0, this obtained as:

$$P'_0(t) = -X_0 P_0(t) + \mu_1 P_1(t) + \mu_2 P_2(t) + \mu_3 P_3(t) + \mu_4 P_8(t) \tag{1}$$

$$P'_0(t) + X_0 P_0(t) = \mu_1 P_1(t) + \mu_2 P_2(t) + \mu_3 P_3(t) + \mu_4 P_8(t)$$

$$\text{Similarly,} \tag{2}$$

$$P'_1(t) + X_1 P_1(t) = \lambda_1 P_0(t) + \mu_2 P_4(t) + \mu_3 P_5(t) + \mu_4 P_9(t) + \mu_5 P_{10}(t) \tag{3}$$

$$P'_2(t) + X_2 P_2(t) = \lambda_2 P_0(t) + \mu_1 P_4(t) + \mu_3 P_6(t) + \mu_4 P_{11}(t) + \mu_6 P_{12}(t) \tag{4}$$

$$P'_3(t) + X_3 P_3(t) = \lambda_3 P_0(t) + \mu_1 P_5(t) + \mu_2 P_6(t) + \mu_4 P_{13}(t) + \mu_7 P_{14}(t) \tag{5}$$

$$P'_4(t) + X_4 P_4(t) = \lambda_2 P_1(t) + \lambda_1 P_2(t) + \mu_3 P_7(t) + \mu_4 P_{15}(t) + \mu_5 P_{16}(t) + \mu_6 P_{17}(t) \tag{6}$$

$$P'_5(t) + X_5 P_5(t) = \lambda_3 P_1(t) + \lambda_1 P_3(t) + \mu_2 P_7(t) + \mu_4 P_{18}(t) + \mu_5 P_{19}(t) + \mu_7 P_{20}(t) \tag{7}$$

$$P'_6(t) + X_6 P_6(t) = \lambda_3 P_2(t) + \lambda_2 P_3(t) + \mu_1 P_7(t) + \mu_4 P_{21}(t) + \mu_6 P_{22}(t) + \mu_7 P_{23}(t) \tag{8}$$

$$P'_7(t) + X_7 P_7(t) = \lambda_3 P_4(t) + \lambda_2 P_5(t) + \lambda_1 P_6(t) + \mu_4 P_{24}(t) + \mu_5 P_{25}(t) + \mu_6 P_{26}(t) + \mu_7 P_{27}(t) \tag{9}$$

$$\text{where, } X_0 = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4$$

$$X_1 = \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \mu_1$$

$$X_2 = \lambda_1 + \lambda_3 + \lambda_4 + \lambda_6 + \mu_2$$

$$X_3 = \lambda_1 + \lambda_2 + \lambda_4 + \lambda_7 + \mu_3$$

$$X_4 = \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \mu_1 + \mu_2$$

$$X_5 = \lambda_2 + \lambda_4 + \lambda_5 + \lambda_7 + \mu_1 + \mu_3$$

$$X_6 = \lambda_1 + \lambda_4 + \lambda_6 + \lambda_7 + \mu_2 + \mu_3$$

$$X_7 = \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7 + \mu_1 + \mu_2 + \mu_3$$

$$P'_8(t) + \mu_4 P_8(t) = \lambda_4 P_0(t) \tag{9}$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_1(t), \text{ where, } i=9,10; j=4,5 \quad (10)$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_2(t), \text{ here, } i=11,12; j=4,6 \quad (11)$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_3(t), \text{ here, } i=13,14; j=4,7 \quad (12)$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_4(t), \text{ here, } i=15,16,17; j=4,5,6 \quad (13)$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_5(t), \text{ here, } i=18,19,20; j=4,5,7 \quad (14)$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_6(t), \text{ here, } i=21,22,23; j=4,6,7 \quad (15)$$

$$P_i'(t) + \mu_j P_i(t) = \lambda_j P_7(t), \text{ here, } i=24,25,26,27; j=4,5,6,7 \quad (16)$$

Steady State

By imposing the condition, steady state probabilities of the system are derived that as $t \rightarrow \infty, d/dt \rightarrow 0$.

With this, equations (5.58) to (5.73) the following equation system is reduced to

$$X_0 P_0 = \mu_1 P_1 + \mu_2 P_2 + \mu_3 P_3 + \mu_4 P_8 \quad (17)$$

Similarly,

$$X_1 P_1 = \lambda_1 P_0 + \mu_2 P_4 + \mu_3 P_5 + \mu_4 P_9 + \mu_5 P_{10} \quad (18)$$

$$X_2 P_2 = \lambda_2 P_0 + \mu_1 P_4 + \mu_3 P_6 + \mu_4 P_{11} + \mu_6 P_{12} \quad (19)$$

$$X_3 P_3 = \lambda_3 P_0 + \mu_1 P_5 + \mu_2 P_6 + \mu_4 P_{13} + \mu_7 P_{14} \quad (20)$$

$$X_4 P_4 = \lambda_2 P_1 + \lambda_1 P_2 + \mu_3 P_7 + \mu_4 P_{15} + \mu_5 P_{16} + \mu_6 P_{17} \quad (21)$$

$$X_5 P_5 = \lambda_3 P_1 + \lambda_1 P_3 + \mu_2 P_7 + \mu_4 P_{18} + \mu_5 P_{19} + \mu_7 P_{20} \quad (22)$$

$$X_6 P_6 = \lambda_3 P_2 + \lambda_2 P_3 + \mu_1 P_7 + \mu_4 P_{21} + \mu_6 P_{22} + \mu_7 P_{23} \quad (23)$$

$$X_7 P_7 = \lambda_3 P_4 + \lambda_2 P_5 + \lambda_1 P_6 + \mu_4 P_{24} + \mu_5 P_{25} + \mu_6 P_{26} + \mu_7 P_{27} \quad (24)$$

$$\mu_4 P_8 = \lambda_4 P_0 \quad (25)$$

$$\mu_j P_i = \lambda_j P_1, \text{ where, } i=9, 10; j=4, 5 \quad (26)$$

$$\mu_j P_i = \lambda_j P_2, \text{ where, } i=11, 12; j=4, 6 \quad (27)$$

$$\mu_j P_i = \lambda_j P_3, \text{ where, } i=13, 14; j=4, 7 \quad (28)$$

$$\mu_j P_i = \lambda_j P_4, \text{ where, } i=15, 16, 17; j=4, 5, 6 \quad (29)$$

$$\mu_j P_i = \lambda_j P_5, \text{ where, } i=18, 19, 20; j=4, 5, 7 \quad (30)$$

$$\mu_j P_i = \lambda_j P_6, \text{ where, } i=21, 22, 23; j=4, 6, 7 \quad (31)$$

$$\mu_j P_i(t) = \lambda_j P_7(t), \text{ where, } i=24, 25, 26, 27; j=4, 5, 6, 7 \quad (32)$$

It can be found by recursively solving these equations as:

$$P_1 = (\lambda_1/\mu_1)P_0; P_2 = (\lambda_2/\mu_2)P_0; P_3 = (\lambda_3/\mu_3)P_0; P_4 = [(\lambda_1\lambda_2)/(\mu_1\mu_2)]P_0;$$

$$P_5 = [(\lambda_1\lambda_3)/(\mu_1\mu_3)]P_0; P_6 = [(\lambda_2\lambda_3)/(\mu_2\mu_3)]P_0; P_7 = [(\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)]P_0$$

On adding,

$$P_1 + P_2 + P_3 + \dots + P_7$$

$$= [(\lambda_1/\mu_1) + (\lambda_2/\mu_2) + (\lambda_3/\mu_3) + (\lambda_1\lambda_2)/(\mu_1\mu_2) + (\lambda_1\lambda_3)/(\mu_1\mu_3) + (\lambda_2\lambda_3)/(\mu_2\mu_3) + (\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)] P_0 = K P_0 \quad (33)$$

where, $K = [(\lambda_1/\mu_1) + (\lambda_2/\mu_2) + (\lambda_3/\mu_3) + (\lambda_1\lambda_2)/(\mu_1\mu_2) + (\lambda_1\lambda_3)/(\mu_1\mu_3) + (\lambda_2\lambda_3)/(\mu_2\mu_3) + (\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)]$;

$$\text{Similarly, } P_9 + P_{10} = (\lambda_4/\mu_4 + \lambda_5/\mu_5)(\lambda_1/\mu_1)P_0; P_{11} + P_{12} = (\lambda_4/\mu_4 + \lambda_6/\mu_6)(\lambda_2/\mu_2)P_0;$$

$$P_{13} + P_{14} = (\lambda_4/\mu_4 + \lambda_7/\mu_7)(\lambda_3/\mu_3)P_0; P_{15} + P_{16} + P_{17} = (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_6/\mu_6)(\lambda_1\lambda_2)/(\mu_1\mu_2)P_0;$$

$$P_{18} + P_{19} + P_{20} = (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_7/\mu_7)(\lambda_1\lambda_3)/(\mu_1\mu_3)P_0;$$

$$P_{21} + P_{22} + P_{23} = (\lambda_4/\mu_4 + \lambda_6/\mu_6 + \lambda_7/\mu_7)(\lambda_2\lambda_3)/(\mu_2\mu_3)P_0;$$

$$P_{24} + P_{25} + P_{26} + P_{27} = (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_6/\mu_6 + \lambda_7/\mu_7)(\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)P_0 \quad (34)$$

The sum of all probability must equal one under the normalizing condition., i.e.

$$\sum P_i = 1, \text{ Or, } P_0 + P_1 + P_2 + \dots + P_{27} = 1 \quad (35)$$

This implies,

$$[P_0 + (P_1 + P_2 + \dots + P_7) + P_8 + (P_9 + P_{10}) + (P_{11} + P_{12}) + (P_{13} + P_{14}) + (P_{15} + P_{16} + P_{17}) + (P_{18} + P_{19} + P_{20}) + (P_{21} + P_{22} + P_{23}) + P_{24} + P_{25} + P_{26} + P_{27}] = 1$$

Or,

$$P_0 [1 + K + \lambda_4/\mu_4 + (\lambda_4/\mu_4 + \lambda_5/\mu_5)(\lambda_1/\mu_1) + (\lambda_4/\mu_4 + \lambda_6/\mu_6)(\lambda_2/\mu_2) + (\lambda_4/\mu_4 + \lambda_7/\mu_7)(\lambda_3/\mu_3) + (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_6/\mu_6)(\lambda_1\lambda_2)/(\mu_1\mu_2) +$$

$$(\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_7/\mu_7)(\lambda_1\lambda_3)/(\mu_1\mu_3) + (\lambda_4/\mu_4 + \lambda_6/\mu_6 + \lambda_7/\mu_7)(\lambda_2\lambda_3)/(\mu_2\mu_3) + (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_6/\mu_6 + \lambda_7/\mu_7)(\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)] = 1$$

$$\text{Or, } P_0 = [1 + K + \lambda_4/\mu_4 + (\lambda_4/\mu_4 + \lambda_5/\mu_5)(\lambda_1/\mu_1) + (\lambda_4/\mu_4 + \lambda_6/\mu_6)(\lambda_2/\mu_2) + (\lambda_4/\mu_4 + \lambda_7/\mu_7)(\lambda_3/\mu_3) +$$

$$(\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_6/\mu_6)(\lambda_1\lambda_2)/(\mu_1\mu_2) + (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_7/\mu_7)(\lambda_1\lambda_3)/(\mu_1\mu_3) + (\lambda_4/\mu_4 + \lambda_6/\mu_6 + \lambda_7/\mu_7)(\lambda_2\lambda_3)/(\mu_2\mu_3) + (\lambda_4/\mu_4 + \lambda_5/\mu_5 + \lambda_6/\mu_6 + \lambda_7/\mu_7)(\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)]^{-1} \quad (36)$$

Now, it is possible to determine the system $A(\infty)$ availability by utilizing:

$$\begin{aligned}
 A(\infty) &= P_0 + P_1 + P_2 + P_3 + \dots + P_7 \\
 &= [1 + (\lambda_1/\mu_1) + (\lambda_2/\mu_2) + (\lambda_3/\mu_3) + (\lambda_1\lambda_2)/(\mu_1\mu_2) + (\lambda_1\lambda_3)/(\mu_1\mu_3) + (\lambda_2\lambda_3)/(\mu_2\mu_3) + (\lambda_1\lambda_2\lambda_3)/(\mu_1\mu_2\mu_3)]P_0 \\
 &= (1+K)P_0
 \end{aligned}
 \tag{37}$$

Using Eq. 37, it is possible to determine the long-term availabilities for a variety of permitted pairings of repair and failure rates of veneer manufacturing systems in steady state. Tables 2 provide a summary of the impact of failure and repair rates on system availability. Following is a discussion of how availability affects system performance in relation to the parameters of the subsystems under consideration.

Table 1: Data of failure and repair of Veneer System

Name of Sub-System	Mean Failure-Rate/hr (λ_i)	Mean Repair-Rate / hr (μ_i)
Debarking Machine (A)	2.0X10-3	1.3X10-2
Veneer cutting lathe (B)	6.0X10-2	14X10-2
Veneer Driers (C)	22.5X10-4	1.2X10-1
Optical Scanner (D)	7.5X10-5	1.2X10-3

Table 2: Effects of subsystem failure and repair rates variation on system performance

Subsystem	Variation in failure rates λ_i (Repair rates μ_i)	Effect of variation on system availability
1 Debarking machine	0.0016-0.0024 (0.017-0.009)	0.8342-0.8028 (2.14)
2 Veneer lathe	0.02-0.10 (0.20-0.08)	0.9156-0.5635 (35.21)
3 Drier	0.00025-0.00425 (0.20-0.04)	0.8301-0.8187 (1.14)
4 Plywood scanner	0.000035-0.000115 (0.0020-0.0004)	0.8578-0.6964 (16.14)

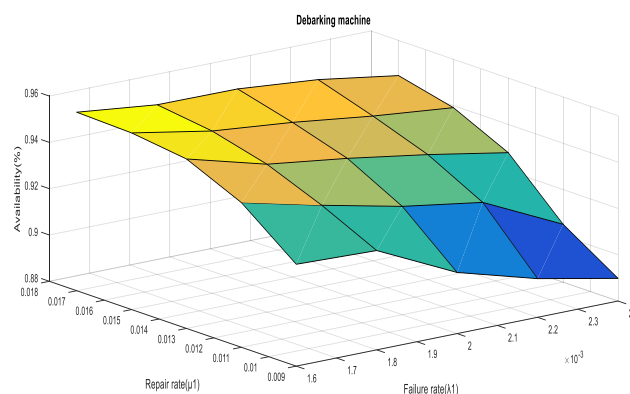


Figure 3: Effect on availability of the failure and repair rate of the debarking machine

Fig. 3 reveals that the variation in FRR of debarking machine has moderate impact on performance of system. Overall 2.14 % Changes have been noted in the system's availability, with the debarking machine's failure rate rising from 0.0016 to 0.0024 and its repair rate falling from 0.017 to 0.009.

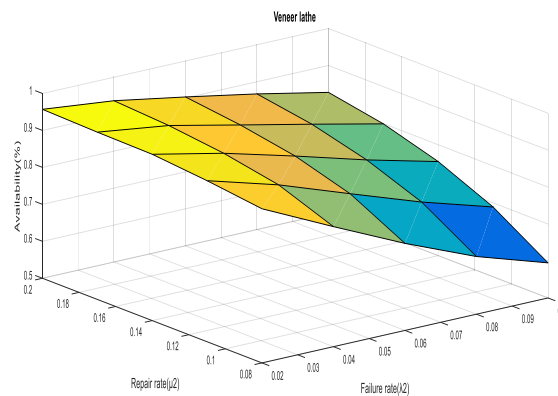


Figure 4: : Effect on availability of the failure and repair rate of the veneer lathe.

The failure and repair rates of the veneer lathe are shown to have a significant impact on the system's overall availability in Fig. 4 above. The overall availability varies by 35.20 percent, with the veneer lathe's failure rate rising from 0.02 to 0.10 and its repair rate falling.

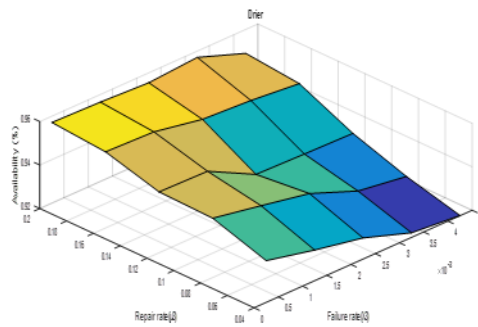


Figure 5: : Impact of the veneer dryer's repair and failure rates on availability.

As can be seen from Fig. 5, there is little to no impact on the system's availability when failure and repair rates vary within acceptable bounds. The system's availability varies 1.14 %, with the veneer dryer's failure rate rising from 0.00025 to 0.00425 and its repair rate falling from 0.20 to 0.04.

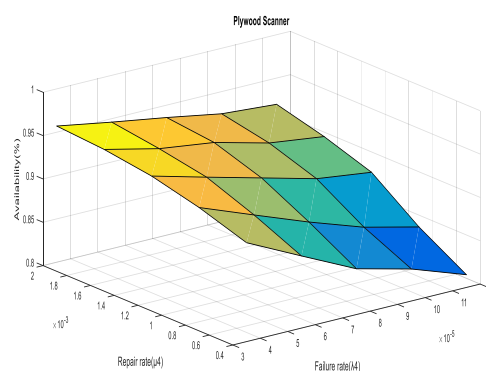


Figure 6: Impact on availability of the plywood scanner's failure and repair rates.

According to Fig. 6, changing the failure and repair rates within the given ranges has a significant impact on the system under consideration's overall availability. The plywood scanner's failure rate rose from 0.000035 to 0.000115, while its repair rates fell from 0.0020 to 0.0004, creating a 16.14% difference in the system's availability.

5. Performance Optimization

In the present study, to see the further enhancement in the availability of the system a performance optimization were carried out using Particle Swarm Optimization (PSO) algorithms. The PSO technique was first used by Dr. Kennedy [5] He proposed on the basis of the social behavior of birds or bees called ‘particles’ in their optimum search for food sources. In this, each bird has its own objective value at present, current position and current velocity. Ever experienced best value by the particle is called p-best i.e. personal best. It also considers the best objective value experienced by any particle ever called g-best i.e. global best.

Thomas Schoene [14] described a standard version of classical PSO which uses the following relations to determine velocity and position of the ith particle:

$$V_i(n+1) = w \cdot V_i(n) + C1(n) \cdot R1_i(n) \cdot \{p\text{-best}_i - X_i(n)\} + C2(n) \cdot R2_i(n) \cdot \{g\text{-best} - X_i(n)\};$$

$$n = 0, 1, \dots, N-1 \tag{38}$$

$$X_i(n+1) = X_i(n) + V_i(n+1); n = 0, 1, \dots, N-1; \tag{39}$$

where V_i is the velocity of i th particle, X_i is the position of i th particle. ‘ n ’ in parenthesis represents the iteration number, $n = 0$ refers to the initialization; N is the total no. of performed iterations, $C1$ and $C2$ are the personal weight and global weights respectively (preferably $C1 = C2 = 2$). $R1_i$ and $R2_i$ are random numbers distributed between 0 and 1 and ‘ w ’ the inertia weight that ranges from 0.4 to 1.4.

In PSO, the best solution represents the optimum position of a particle. There is random initialization of particles along with their velocity and position which were evaluated with equations (38) and (39). The main steps involved in optimization process may be depicted as shown in Fig.7 below:

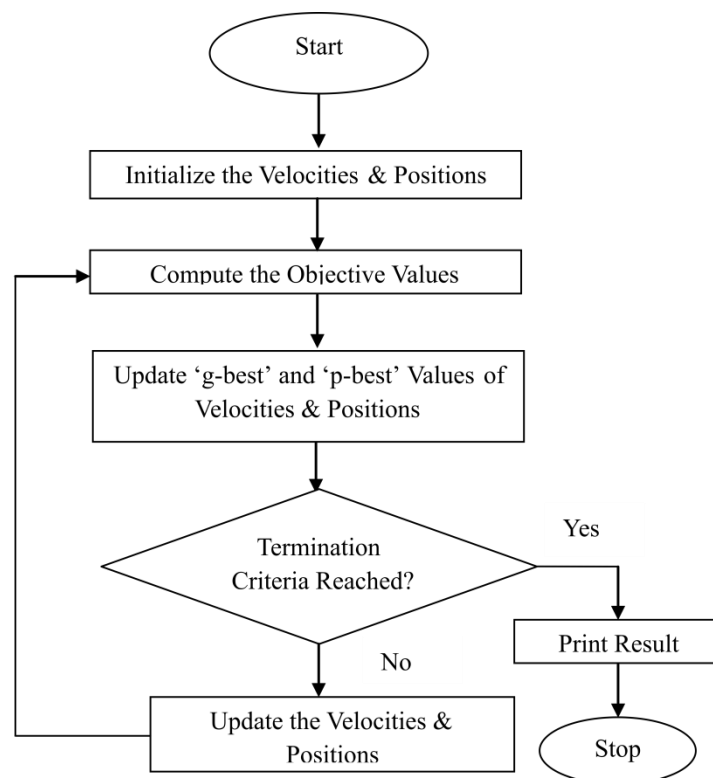


Figure 7: Flow Diagram PSO

The best position reached in each iteration is compared with the best previous position and similarly their position of the global best and personal best are updated. Each particle is updated to a new best position considering their previous experience after adjusting their velocities. After reaching to the new position, the particles of swarn are updated. Best optimal solution is obtained by repeating the process in same manner.

Computational Optimized Results of Veneer Making System

Following the procedure mentioned in Fig.7 by varying the failure and repair rate within the permissible limits the performance optimization of various subsystems has been carried out These are are shown below:

$\lambda_1 \in (0.0016-0.0024)$, $\mu_1 \in (0.009-0.017)$; $\lambda_2 \in (0.02-0.10)$, $\mu_2 \in (0.08-0.20)$;
 $\lambda_3 \in (0.00025-0.00425)$, $\mu_3 \in (0.04-0.20)$ and $\lambda_4 \in (0.000035-0.000115)$, $\mu_4 \in (0.0004-0.0020)$

The effect of population size and the number of iterations on the system performance is shown below in Tables 3 and 4 as given below.

Table 3: Effect of Population Size (PS) Variation on the Accessibility of Veneer Making System

Population Size@ no. of GS	Failure Rate				Repair rate				Optimum Availability (%)
	λ_1	λ_2	λ_3	λ_4	μ_1	μ_2	μ_3	μ_4	
100									
10	0.0019	0.02	0.00092	0.000109	0.010	0.17	0.15	0.0016	0.9123
20	0.0018	0.02	0.00343	0.000067	0.013	0.14	0.10	0.0019	0.9383
50	0.0018	0.02	0.00235	0.000048	0.013	0.16	0.10	0.0016	0.9492
100	0.0018	0.02	0.00246	0.000047	0.013	0.16	0.10	0.0016	0.9498
200	0.0016	0.02	0.00036	0.000041	0.016	0.17	0.10	0.0018	0.9611
1500	0.0018	0.02	0.00235	0.000040	0.015	0.18	0.05	0.0019	0.9614
3000	0.0016	0.02	0.00189	0.000038	0.015	0.20	0.04	0.0018	0.9638
4000	0.0016	0.02	0.00191	0.000038	0.015	0.20	0.04	0.0018	0.9639
6000	0.0016	0.02	0.00186	0.000037	0.015	0.20	0.04	0.0017	0.9641
8000	0.0016	0.02	0.00348	0.000035	0.015	0.19	0.08	0.0020	0.9669
15000	0.0016	0.02	0.00379	0.000035	0.017	0.20	0.10	0.0020	0.9685

Table 4: The impact of different iteration counts on the accessibility of veneer manufacturing systems

Generation Size @ no. of PS	Failure Rate				Repair rate				Optimum Availability (%)
	λ_1	λ_2	λ_3	λ_4	μ_1	μ_2	μ_3	μ_4	
50000									
10	0.0015	0.10	0.0051	0.0016	0.015	0.19	0.05	0.0016	0.9560
20	0.0016	0.02	0.00373	0.000035	0.017	0.20	0.11	0.0020	0.9682
30	0.0016	0.02	0.00379	0.000035	0.017	0.20	0.11	0.0020	0.9684
50	0.0016	0.02	0.00380	0.000035	0.017	0.20	0.11	0.0020	0.9685
80	0.0016	0.02	0.00380	0.000035	0.017	0.20	0.11	0.0020	0.9685
120	0.0016	0.02	0.00380	0.000035	0.017	0.20	0.11	0.0020	0.9685
200	0.0016	0.02	0.00380	0.000035	0.017	0.20	0.11	0.0020	0.9685
250	0.0016	0.02	0.00380	0.000035	0.017	0.20	0.11	0.0020	0.9685

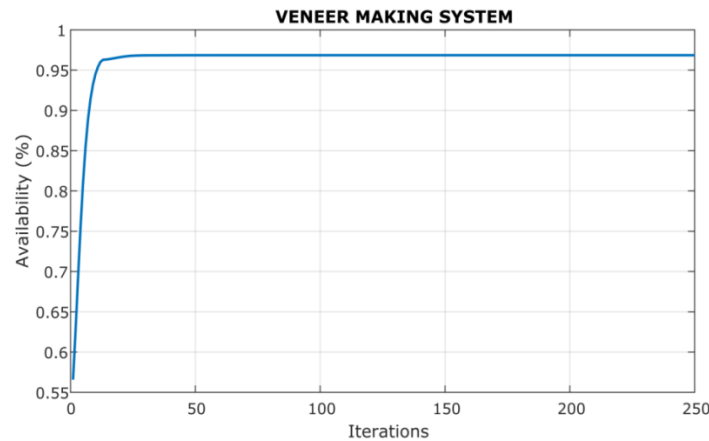


Figure 8: Optimum Availability of Veneer Making System using PSO

The Fig. 8 shows the highest achievable availability of system is as much as 96.85%. Based on the detailed investigation a comparative analysis of results is presented in Table 5 in the form of DSS.

6. Conclusions

The detailed investigation carried out on different subsystems here indicates that the veneer cutting lathe needs an utmost maintenance priority as it is the most critical subsystem. The failure of veneer knives is the main reason for the failure is in veneer cutting lathe. The effect of varying repair facilities on the availability of system were carried out which will further help in the allocation of repair facilities among the subsystems. The obtained outcomes also demonstrate the usefulness of the RAM tools. The analysis carried out will further help the maintenance engineers to optimize the overall maintenance cost and overall production cost. Thus it understood that appropriate RAM tools have the direct impact on the maintenance cost and overall production cost [15].

Table 5: Summary of Results and DSS for Veneer making system of Plywood Manufacturing Plant

Name of System	Name of subsystem	Varying Failure rates λ_i and μ_i (Repair Rates)	Impact of change on availability using Markov (%)	Optimized Availability based on PSO (%)	Suggested Maintenance Priorities*
Veneer Making System	Debarking Machine	0.0016-0.0024 (0.017-0.009)	0.8342-0.8028 (2.14)	96.85 %	III
	Veneer lathe	0.02-0.10 (0.2-0.08)	0.9156-0.5635 (35.21)		I
	Drier	0.00025-0.00425 (0.2-0.04)	0.8301-0.8187 (1.14)		IV
	Plywood Scanner	0.000035-0.000115 (0.0020-0.0004)	0.8578-0.6964 (16.14)		II

* At present being managed on the basis of intuitive decisions of the plant managers

In case it is desired to determine more possible efficient performance of such systems then we recommend the use of PSO type of optimization approach to be used for further optimizing the results obtained using Markov or any other convenient approaches as has been done in the present case study. Here a DSS is proposed (in Table 5) so that a significant amount of wastage in material and manpower involvement can be reduced. Hence, for researchers a lot of opportunity to work on the advancement of veneer layup process so as to achieve an optimum point between cost vs quality ensuring safety and volume of production. In future, It will be of great interest if Petri Nets approach is applied in such cases that supports non constant failure pattern also.

References

- [1] Angel, A.S. and Jayaparvathy, R. (2019), "Performance modeling of an intelligent emergency evacuation system in buildings on accidental fire occurrence". *Safety science* 12:196-205.doi.10.1016/j.ssci.2018.10.027.
- [2] Bahl, A., Sachdeva, A. and Garg, R.K. (2018), "Availability analysis of distillery plant using petri nets", *International Journal of Quality & Reliability Management*, Vol. 35 No. 10, pp. 2373-2387.
- [3] Dhillon, B. S., Rayapati, S. N. (1988), "Chemical-system reliability:a review", *IEEE Transactionson Reliability*,Vol. 37, pp.199-208.
- [4] Hajeer, M., Chaudhuri, D. (2000), "Reliability and availability assessment of reverse osmosis", *Desalination*, Vol.130, pp.185-192.
- [5] Kennedy,J. and Eberhart,R. (1995). "Particle swarm optimization", *Proceedings, IEEE International Conference on Neural Networks*, Vol. 4, pp.1942-1948.
- [6] Khan, M. R. R., Kabir and A. B. M. Z. (1995), "A vailability simulation of an ammonia plant", *Reliability Engineering and System Safety*, Vol.48, pp.217-227.
- [7] Kumar,N., Tewari, P.C. and Sachdeva, A. (2019), "Performance Modelling and Analysis of Refrigeration System of a Milk Processing Plant using Petri Nets", *International Journal of Performability Engineering*, Vol. 15, No. 7, pp. 1751-1759.
- [8] Kumar,N., Tewari, P.C. and Sachdeva, A. (2020) "Performance modelling and availability analysis of a milk pasteurising system using Petri nets formalism", *International Journal of Simulation and Process Modelling* , Vol.15, No.5, pp.401-408, 2020. DOI: 10.1504/IJSPM.2020.110915.
- [9] Kumar,N., Tewari, P.C. and Sachdeva, A. "Performance Assessment of Milk Processing System using Petri Nets" *SN Applied Sciences*, Vol. 2:1-12, DOI:10.1007/s42452-020-03282-0. 4.
- [10] Kumar,N., Tewari, P.C. and Sachdeva, A. "Petri Nets Modelling and Analysis of the Veneer Layup System of Plywood Manufacturing Plant", *Int. journal for engineering modeling*, DOI: 10.31534/engmod.2020.1-2.ri.07v.
- [11] Kempa, W. M. (2020). "On time-to-buffer overflow distribution in a single-machine discrete-time system with finite capacity", *Mathematical Modelling and Analysis*, 25(2),pp. 289-302. <https://doi.org/10.3846/mma.2020.10433>.
- [12] Malik, S and Tewari,PC 2020 Optimization of coal handling system performability for a thermal power plant using PSO algorithm*Grey Systems: Theory and Application* Vol. 10 No. 3, 2020, pp. 359-376.
- [13] Regattieri, A. and Bellomi, G. 2009. "Innovative lay-up system in plywood manufacturing process", *European Journal Wood Prod.* 67:55-62. DOI 10.1007/s00107-008-0282-0.
- [14] Schoene,T. (2011), "Step-Optimized Particle Swarm Optimization", M.S. thesis, University of Saskatchewan, Saskatoon, Canada.
- [15] Sachdeva,A., Kumar,D. Kumar,P. 2008. " Planning and optimizing the maintenance of paper production system in a paper plant", *Computers & industrial Engineering* 55:817-829.
- [16] Singh, J., Pandey, P.C. and Kumar, D. (1990), "Designing for reliable operation of urea

synthesis in the fertilizer industry”, *Microelectronics and Reliability*, Vol.30, pp.1021-1024.

[17] Viswanadham,N. and Narahari,Y. (2015), “Performance modeling of automated mfg. systems”, PHI learning Pvt. Ltd., Delhi.

[18] Er. Sudhir Kumar, Dr. P.C. Tewari Performability optimisation of multistate coal handling system of a thermal power plant having subsystems dependencies using pso and comparative study by petri nets. *Reliability: Theory & Applications*. 2023, March 1(72): 250-263. <https://doi.org/10.24412/1932-2321-2023-172-250-263>.

[19] Er. Sudhir Kumar, Dr. P.C. Tewari Performability analysis of multistate ash handling system of thermal power plant with hot redundancy using stochastic petrinets. *Reliability: Theory & Applications*. 2022, September 3(69): 190-201. <https://doi.org/10.24412/1932-2321-2022-369-190-201>

A NEW CONTROL CHART FOR PROCESS DISPERSION BASED ON RANKED SET SAMPLING

Chandrakant Gardi and Vikas Ghute*



Department of Statistics
Punyashlok Ahilyadevi Holkar Solapur University, Solapur (MS), India
chandrakant.gardi@gmail.com, vbghute_stats@rediffmail.com

*Corresponding author

Abstract

In this paper, we propose a new control chart based on Downton's estimator (D) for monitoring the process dispersion using ranked set sampling design and runs rules. The performance of the proposed control chart is compared with the originally proposed D chart based on simple random sampling method when underlying process distribution is normal and non-normal. The average run length is used to evaluate the performance of the proposed control chart. It is observed that the proposed control chart is efficient in detecting shifts in process dispersion as compared with the chart using simple random sampling method. The performance of the proposed chart is further investigated using runs rules. The efficiency of the designed runs rules RSS-D chart is compared with its existing counterparts and is found to be superior.

Keywords: Control chart, Average run length, Process dispersion, Ranked set sampling, Runs rules, Downton estimator.

1. Introduction

In statistical process control (SPC), control charts are the most popular tools used to monitor the quality characteristics in an industrial process. Shewhart R chart and S chart are the commonly used control charts for monitoring the process dispersion. The R chart is based on the sample range (R) whereas S chart is based on sample standard deviation (S). These charts are based on subgroups of sample size n using simple random sampling (SRS) method. Both R and S charts are easy to implement and are effective in the detection of large shifts, but are less sensitive in the detection of small and moderate shifts in the process dispersion (Montgomery[1]). Several alternatives have been proposed in the literature to improve the performance of Shewhart type control charts. One recent approach is the construction of control charts based on ranked set sampling (RSS) scheme. The RSS scheme was first suggested by McIntyre [2]. This scheme ensures increased precision through ranking of observations. He pointed out that this sampling scheme works superior to standard simple random sampling (SRS) for estimation of the population mean. Takahasi and Wakimoto [3] laid the necessary mathematical formulation for this sampling scheme. Several researchers have studied applications of ranked set sampling in different streams. Kvam and Samaniego [4] showed that ranked set sampling may occur naturally in survival analysis. Kvam and Samaniego [5] studied the applications of this sampling scheme in reliability. Recently, RSS scheme has got considerable attention in the construction of control charts. Salazar and Sinha [6] first suggested control charts for monitoring process mean using RSS scheme. It was shown that the RSS based control charts perform better than the charts based on SRS. Muttalak and Al-Sabah [7]

developed control charts for process mean based on different RSS schemes and showed that the proposed RSS based control charts perform better than the classical SRS charts. Al-Naseer and Al-Rawwash [8] developed the mean control chart using robust RSS method. Al-Omari and Al-Naseer [9] suggested a new quality control chart for mean using robust extreme ranked set sampling (RERSS) method. It was shown that RERSS chart performs better than all other charts based on SRS and RSS methods. Al-Omari and Haq [10] developed Shewhart-type control charts to improve the monitoring the process mean by using the double quartile-ranked set sampling, quartile double ranked set sampling and double quartile extreme ranked set sampling methods. Al-Rawwash et al. [11] developed control charts of the sample mean based on several sampling techniques including RSS as well as some of its recent modifications assuming that the underlying distribution is normal with mean μ and variance σ^2 . Mehmood et al. [12] discussed the details of different ranked set strategies and their applications in an industrial process using control charts. Abbasi [13] investigated the performance of the mean chart based on various sampling schemes under normal and non-normal process distributions. Tahir et al. [14] designed and investigated dispersion control charts based on different estimates of process dispersion under different RSS strategies for normal and non-normal processes.

Most of the papers published so far on the RSS based control charts have concentrated on monitoring the process location and few control charts has been developed for monitoring the process dispersion. Motivated by the attractive features of RSS scheme, in this paper, we propose a D chart based on RSS scheme (denoted as RSS-D chart) for monitoring process dispersion. It is shown that the proposed RSS-D chart performs uniformly better than the D chart without RSS scheme for detecting increasing shifts in the process dispersion. Although the RSS scheme improves the performance of the D chart in detecting shifts in process dispersion, further improvement can be achieved if runs rules scheme is implemented to the RSS-D chart. In case of SRS, a number of authors have defined runs rules schemes for different control charts to improve the detecting ability for small and moderate shifts in process parameters. Klein [15] proposed two control charts based on runs rule as efficient alternatives to Shewhart control chart. These control charts were named as 2-out-of-2 runs rule control chart and 2-out-of-3 runs rule control chart. In 2-out-of-2 runs rule control chart, if two successive points fall beyond a special control limit, then process is declared to be out of control. In 2-out-of-3 runs rule control chart, if two points among successive three points fall beyond a special control limit, then process is declared to be out of control. This special limit is determined in such a way that the in-control ARL equals with specified value ARL_0 . This chart has been proven to perform better than Shewhart chart. Antzoulakos and Rakitzis [16] have studied the control charts with supplementary runs rules for monitoring the process standard deviation. In general, the use of runs rules increases a chart's sensitivity but also produces more false alarms. To keep the same rate of false alarm, control limits have to be adjusted (Cheng and Chen [17]). Although runs rules have been widely used in practice to increase the sensitivity of control charts, to our best knowledge, little research has been done on the RSS based control charts.

The purpose of this paper is to improve the performance of the D chart proposed by Abbasi and Miller [18] to detect shifts in process dispersion by using RSS scheme and further to implement runs rule scheme to proposed RSS-D chart. The rest of the paper is organized as follows: In section 2, we outline the process of constructing the D control chart based on SRS (denoted as SRS-D chart) for monitoring the process dispersion. The description of ranked set sampling scheme used in this study is given in section 3. The charting structure of the proposed RSS-D chart is presented in section 4. RSS based D-chart with runs rules is discussed in section 5. The performance evaluation of proposed RSS-D chart and RSS-D chart with runs rules and its comparative study with SRS-D control chart is given in section 6. In section 7, an illustrative example based on simulated data sets is provided for a better understanding about implementation of the proposed control chart. A real life application of the proposed control chart is illustrated in section 8. Conclusions are given in the final section 9.

2. D Chart for Process Dispersion Based on SRS

Let X_1, X_2, \dots, X_n be a simple random sample of size n from a normal process whose mean is μ and standard deviation is σ . Further, let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the order statistics corresponding to this sample. Downton [19] proposed following statistic as an estimator for process standard deviation σ .

$$D = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^n \left[i - \frac{1}{2}(n+1) \right] X_{(i)} \quad (1)$$

which is an unbiased estimator of σ for normally distributed process data. Abbasi and Miller [18] proposed the control chart based on D to monitor the process variability. It was shown that for normally distributed process, the D chart is equally efficient to the Shewhart S chart for detecting shifts in the process standard deviation. Since the distribution of D is not symmetric at least for small to moderate values of n , they have used probability limits instead of three-sigma control limits in the construction of D chart. The probability limits of D chart are computed by using quantile points of the distribution of $\frac{D}{\sigma}$. The upper control limit (UCL) of the D chart is given as

$$UCL = Z_{1-\alpha} \bar{D} \text{ with } P(Z \geq Z_{1-\alpha}) = 1 - \alpha \quad (2)$$

where α is the specified probability of making Type-I error, Z_α is α^{th} quantile point of the distribution of Z . The process dispersion is then monitored by plotting values of the D statistic on the chart with UCL given by Eq. (2). If $D < UCL$, the process is regarded as in-control, otherwise there is an increase in process standard deviation and process is deemed out-of-control.

To design the D chart based on simple random sample (SRS), let X_1, X_2, \dots, X_n be independent and identically distributed $N(\mu_0, \sigma_0^2)$ random sample of size n , where μ_0 and σ_0^2 are the in-control mean and variance, respectively. As the concern of this paper is monitoring process dispersion, let us assume that process standard deviation has a shift of size δ . Hence, the new standard deviation for the process becomes $\sigma_1 = \delta\sigma_0$, where σ_1 is out-of-control value of process standard deviation. For positive shift ($\delta > 1$), that is, for increase in σ , $k^+\sigma_0$ is the upper control limit and process is declared to be out-of-control if $D > k^+\sigma_0$. For negative shift ($\delta < 1$), that is for decrease in σ , $k^-\sigma_0$ is the lower control limit and process is declared to be out-of-control if $D < k^-\sigma_0$. The values of k^+ and k^- are found for fixed in-control ARL value based on simulations.

A commonly used performance measure for a control chart is its ARL. For a given shift in process parameter, the ARL is the average number of points plotted on a control chart until an out-of-control signal is obtained. In case of Shewhart control chart, which signals as soon as a point falls beyond the control limits, $ARL = 1/p$, where p is the probability that a point exceeds the control limits.

For positive shift δ ,

$$\begin{aligned} p &= Pr(D > k^+\sigma_0 / \sigma = \delta\sigma_0) \\ &= Pr\left(\frac{D}{\delta\sigma_0} > \frac{k^+\sigma_0}{\delta\sigma_0} / \sigma = \delta\sigma_0\right) \\ &= Pr\left(Z > \frac{k^+}{\delta}\right) \\ &= 1 - F\left(\frac{k^+}{\delta}\right) \end{aligned}$$

In-control ARL is the reciprocal of the detecting power p .

$$ARL(\delta) = \frac{1}{p} = \frac{1}{1 - F\left(\frac{k^+}{\delta}\right)} \quad (3)$$

For negative shift δ ,

$$\begin{aligned} p &= Pr(D < k^-\sigma_0 / \sigma = \delta\sigma_0) \\ &= Pr\left(\frac{D}{\delta\sigma_0} < \frac{k^-\sigma_0}{\delta\sigma_0} / \sigma = \delta\sigma_0\right) \\ &= Pr\left(Z < \frac{k^-}{\delta}\right) \\ &= F\left(\frac{k^-}{\delta}\right) \end{aligned}$$

where $F(\cdot)$ is distribution function of Z .

In-control ARL is the reciprocal of the detecting power p .

$$ARL(\delta) = \frac{1}{p} = \frac{1}{F\left(\frac{k}{\delta}\right)} \quad (4)$$

3. Ranked Set Sampling Scheme

The mechanism of ranked set sampling (RSS) is as follows. Let X be the characteristic under study. In order to get a RSS sample of size n , total of n^2 units are drawn from the population. These n^2 units are randomly arranged in n sets each of size n . Let $\{X_{i,k}, i = 1, 2, \dots, n\}$ be the k^{th} set, where $k = 1, 2, \dots, n$. The observations in the k^{th} set $\{X_{i,k}, i = 1, 2, \dots, n\}$ are ranked either by visual magnitude of characteristic X or by using some other variable, which is significantly correlated with X . If the ranking of X is exact, then the sampling scheme is called RSS with perfect ranking, and if ranking is not exact, then the sampling scheme is called RSS with imperfect sampling. In the present study, we have assumed perfect RSS scheme. Let k^{th} ordered set be $\{X_{[i,k]}, i = 1, 2, \dots, n\}$, where $X_{[i,k]}, i = 1, 2, \dots, n; k = 1, 2, \dots, n$ is the i^{th} order statistic in k^{th} set. These sets are called the ranked sets. Then $i^{th}, i = 1, 2, \dots, n$ order statistic is picked from $i^{th}, i = 1, 2, \dots, n$ ranked set. That is, first order statistics from first ranked set, second order statistics from second ranked set and so on. Thus, a final selected sample of size n is $\{X_{[1,1]}, X_{[2,2]}, \dots, X_{[n,n]}\}$, which is termed as a ranked set sample (RSS).

4. RSS Based D Chart for Process Dispersion

In this section, a working mechanism of RSS-D chart is described. A subgroup sample from the underlying process is obtained using RSS scheme of sampling. It is assumed that ranking is perfect. Using this scheme, we generate subgroups of size n and repeat them for n cycles from the distribution that is under study. Downton statistic based on RSS sample $\{X_{[1,1]}, X_{[2,2]}, \dots, X_{[n,n]}\}$ of size n is given as

$$D = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{i=1}^n \left[i - \frac{1}{2}(n+1) \right] X_{[i,i]} \quad (5)$$

Since the exact distribution of D statistic based on ranked set sampling scheme is not known, simulation approach is used to compute ARL values of the proposed RSS-D chart. The control limits for RSS-D chart are found out using extensive simulation. 50000 ranked set samples of size n were simulated and quantile points of D statistic based on RSS scheme were realized. Using these quantile points, the values of k^+ and k^- are evaluated using equations (3) and (4) respectively for positive and negative shifts in process standard deviation σ . The estimated ARL values of the proposed RSS-D chart along with standard error of the chart are found with 50000 simulations for each shift of magnitude δ in process standard deviation σ . The ARL performance of the RSS-D chart under both normal and non-normal process distributions is presented in Section 6.

5. RSS-D Chart with Runs Rule

In this section, we study the performance of the proposed RSS-D chart using the following runs rules. Note that we would usually use only one of these rules at a time. Each rule would have different control limits.

The process is declared to be out of control when

- (i) Two consecutive points plot outside the control limits (2-of-2-rule).
- (ii) Any two of three successive points plot outside the control limit (2-of-3-rule)

The ARL performance of the RSS-D chart using these runs rules has been evaluated under both normal and non-normal process distributions using a simulation study in Section 6.

6. Performance Evaluation and Comparison

6.1. Performance evaluation under normality

This section evaluates and compares the performance of RSS-D chart and RSS-D chart using runs rules and D chart with simple random sampling. To compare the performance of these charts under normality, in-control observations are generated from $N(\mu_0, \sigma_0^2)$ distribution. Without loss of generality, we take $\mu_0 = 0$ and $\sigma_0^2 = 1$. For out-of-control process, observations are generated from $N(\mu_0, \sigma_1^2)$ distribution. Without RSS scheme, the values of the control chart statistic D for the D chart are computed using Eq. (1) for subgroups of size n . With the RSS scheme, the values of the control chart statistic D for the RSS- D chart are computed using Eq. (5) based on RSS sample of size n . For increase in standard deviation (that is, $\sigma_1 > \sigma_0$), where $\sigma_1 = \delta\sigma_0$, the magnitude of shift in process standard deviation considered are $\delta = 1.0, 1.1, 1.2, 1.3, 1.4, 1.5$ and 2.0 , while for decrease in process standard deviation (that is, $\sigma_1 < \sigma_0$), the magnitude of shift in process standard deviation considered are $\delta = 1.0, 0.9, 0.8, 0.7, 0.6, 0.5$ and 0.1 . All charts have calibrated in-control ARL of 200 and subgroup size of $n = 5, 8$ and 10 are considered. The values of k^+ and k^- are obtained by designing the chart to have an in-control ARL of 200. The out-of-control ARL values of the charts are found with 50000 simulations for each of shift of magnitude δ in the process standard deviation σ .

ARL values for positive shifts and negative shifts in process standard deviation with different sample sizes $asn = 5, 8, 10$ are given in Table 1 and Table 2 respectively.

Table 1: ARL comparison for process dispersion with positive shift under normal distribution.

Sample n	Shift δ	SRS-D	RSS-D	RSS-D (2-out-of-2)	RSS-D (2-out-of-3)
5	1.0	200	200.27(1.816)	200.39 (1.795)	200.25 (1.817)
	1.1	66.65	62.43 (0.561)	60.35 (0.540)	55.36 (0.491)
	1.2	29.67	25.90 (0.230)	25.76 (0.223)	23.22 (0.197)
	1.3	15.56	13.48 (0.117)	14.52 (0.119)	12.78 (0.102)
	1.4	9.59	8.35 (0.071)	9.50 (0.074)	8.26 (0.061)
	1.5	6.49	5.66 (0.048)	6.87 (0.050)	6.11 (0.042)
	2.0	2.31	2.05 (0.014)	3.22 (0.016)	2.98 (0.013)
	k^+	2.068	1.9985	1.5377	1.60286
8	1.0	200	199.65(1.822)	200.03 (1.779)	200.85(1.823)
	1.1	54.00	45.14 (0.407)	40.46 (0.358)	37.27 (0.326)
	1.2	20.62	15.75 (0.138)	14.63 (0.122)	13.14 (0.105)
	1.3	10.40	7.47 (0.064)	7.61 (0.057)	6.80 (0.048)
	1.4	6.05	4.33 (0.035)	4.91 (0.033)	4.44 (0.027)
	1.5	4.13	2.92 (0.022)	3.69 (0.021)	3.46 (0.018)
	2.0	1.55	1.27 (0.005)	2.24 (0.006)	2.19 (0.004)
	k^+	1.78	1.6296	1.33955	1.43986
10	1.0	200	199.55(1.852)	200.03 (1.809)	199.49 (1.792)
	1.1	48.33	36.44 (0.333)	32.07 (0.282)	29.49 (0.252)
	1.2	17.45	11.56 (0.101)	10.67 (0.085)	9.58 (0.074)
	1.3	8.27	5.20 (0.042)	5.39 (0.037)	5.02 (0.032)
	1.4	4.86	3.02 (0.023)	3.66 (0.021)	3.39 (0.017)
	1.5	3.28	2.08 (0.014)	2.88 (0.013)	2.74 (0.010)
	2.0	1.34	1.10 (0.003)	2.06 (0.003)	2.05 (0.002)
	k^+	1.676	1.508285	1.27474	1.307509

Table 2: ARL comparison for process dispersion with negative shift under normal distribution.

Sample n	Shift δ	SRS-D	RSS-D	RSS-D (2-out-of-2)	RSS-D (2-out-of-3)
5	1.0	201	200.64 (1.851)	200.77 (1.816)	200.98 (1.852)
	0.9	135.62	176.49 (1.592)	118.89 (1.055)	116.73 (1.067)
	0.8	86.12	152.17 (1.386)	62.21 (0.548)	62.80 (0.564)
	0.7	51.93	128.76(1.173)	30.37 (0.265)	30.30 (0.265)
	0.6	29.34	99.68 (0.917)	14.24 (0.118)	14.27 (0.117)
	0.5	15.42	72.75 (0.667)	6.58 (0.047)	6.53 (0.048)
	0.1	1.00	1.42 (0.007)	2.00(0.00)	2.00(0.00)
	k^-	0.240	0.1184	0.483308	0.483291
8	1.0	200	200.47(1.824)	199.59(1.786)	200.79(1.832)
	0.9	102.77	101.45(0.911)	61.40(0.548)	62.08(0.562)
	0.8	51.32	47.47(0.425)	20.01(0.169)	19.89(0.167)
	0.7	23.99	20.34(0.182)	7.24(0.0543)	7.22(0.054)
	0.6	10.72	8.29(0.071)	3.37(0.018)	3.40(0.019)
	0.5	4.63	3.20(0.024)	2.22(0.006)	2.23(0.006)
	0.1	1.00	1.00(0.00)	2.00(0.00)	2.00(0.00)
	k^-	0.388	0.4425	0.674418	0.67443
10	1.0	201.00	200.82(1.818)	201.01(1.859)	201.59 (1.823)
	0.9	91.63	72.29(0.658)	43.57(0.390)	43.54(0.379)
	0.8	39.72	24.76(0.220)	11.52(0.093)	11.81(0.095)
	0.7	16.37	8.50(0.072)	4.19(0.026)	4.23(0.026)
	0.6	6.77	3.06(0.023)	2.39(0.008)	2.40(0.008)
	0.5	2.81	1.43(0.007)	2.02(0.001)	2.02(0.002)
	0.1	1.00	1.00(0.00)	2.00(0.00)	2.00(0.00)
	k^-	0.449	0.549	0.7365543	0.736105

Following are the findings from Table 1 and Table 2.

- It is observed that with an increase in sample size n , the out-of-control ARL values of all the charts under investigation decreases. This shows that large sample sizes help control chart statistic to quickly detect the shifts in process dispersion of the production process.
- For a fixed value of sample size n , with an increase or decrease in the shift of process dispersion, the out-of-control ARL values of all the charts under investigation decreases. The same trend is observed in standard errors of the estimated ARL values.
- For increase in process dispersion, the proposed RSS-D chart consistently produces smaller out-of-control ARL values than SRS-D chart for the entire range of shifts in the process dispersion.
- The RSS-D chart with runs rule shows a better performance than SRS-D chart for smaller shifts ($\delta \leq 1.2$) and smaller sample size ($n = 5$), while for larger sample size ($n = 8$ or 10), chart shows better performance for small to moderate shifts.
- The RSS-D chart with runs rule shows better performance than RSS-D chart only for small shifts in process dispersion.
- For decrease in process dispersion, for smaller sample size ($n = 5$), the out-of-control ARL values of the proposed RSS-D chart are greater than that of SRS-D chart. As sample size increases, the RSS-D chart becomes more efficient than the SRS-D chart.
- RSS-D chart with runs rule shows better performance than RSS-D chart and SRS-D chart for decreasing shifts in process dispersion.
- The performance of RSS-D chart with 2-out-of-2 and 2-out-of-3 runs rule is similar.

6.2. Performance evaluation under non-normality

In this section, we study the performance of the RSS-D and RSS-D chart by implementing runs rules schemes under non-normal distributions and it is compared with the D chart based on SRS scheme. In order to study the performance of these charts under non-normality, we considered heavy-tailed symmetric and skewed distributions. Specifically, we simulated observations in the heavy-tailed symmetric case from double exponential distribution and in the skewed case from gamma distribution. The probability density functions of these non-normal distributions are given as

Double exponential distribution: The double exponential distribution, denoted by Double Exponential(μ, λ) has the probability density function,

$$f(x; \mu, \lambda) = \frac{1}{2\lambda} e^{-\lambda|x-\mu|}, -\infty < x < \infty; \lambda > 0, -\infty < \mu < \infty \quad (6)$$

where μ and λ are respectively the location and scale parameters. The mean and variance of a double exponential distribution are μ and $2\lambda^2$ respectively.

Gamma distribution: The gamma distribution, denoted by $Gamma(\alpha, \beta)$ has the probability density function,

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0; \alpha, \beta > 0 \quad (7)$$

where α and β are respectively the shape and scale parameters. The mean and variance of a gamma distribution are α/β and α/β^2 respectively.

To study the performance of the RSS-D control chart under non-normality, we have considered the process data from double exponential distribution with location parameter $\mu = 0$ and scale parameter $\lambda = 1$ and gamma distribution with shape parameter $\alpha = 2$ and scale parameter $\beta = 1$. Simulation approach is used to compute ARL values of these charts. The ARL values of the charts are found with 50000 simulations for each shift of magnitude δ in the process standard deviation. The standard errors of the estimated ARL values of these charts are given (in parentheses) along with ARL values. The value of k^+ is determined based on 100 iterations of 50,000 simulations from the underlying process with no shift, such that in-control ARL is 200. In this study we consider the case of increase in standard deviation only.

Tables 3 and 4 present the ARL values of the charts under investigation to detect increase in process standard deviation for different magnitudes for in-control ARL of 200 and sample sizes $n = 5, 8$ and 10, when underlying process distributions are gamma and double exponential.

Table 3: ARL comparison for process dispersion with positive shift under gamma distribution.

Sample n	Shift δ	SRS-D	RSS-D	RSS-D (2-out-of-2)	RSS-D (2-out-of-3)
5	1.0	200.23 (0.897)	200.61 (0.904)	199.50 (1.808)	200.26 (1.807)
	1.1	93.56 (0.415)	92.95 (0.41)	77.59 (0.703)	74.34 (0.668)
	1.2	50.63 (0.226)	48.38 (0.215)	37.54 (0.327)	34.74 (0.300)
	1.3	30.75 (0.135)	28.33 (0.126)	20.92 (0.178)	19.58 (0.163)
	1.4	20.19 (0.089)	17.99 (0.078)	13.36 (0.11)	12.38 (0.098)
	1.5	14.04 (0.060)	12.17 (0.052)	9.32 (0.073)	8.65 (0.063)
	2.0	4.46 (0.018)	3.533 (0.013)	3.59 (0.02)	3.41 (0.017)
	k^+	2.452	2.4051	1.635	1.73
8	1.0	200.32 (0.897)	201.30 (0.897)	201.5 (1.814)	200.46 (2.786)
	1.1	80.14 (0.357)	74.60 (0.330)	58.25 (0.517)	55.35 (0.788)
	1.2	38.73 (0.171)	22.76 (0.099)	23.21 (0.198)	21.53 (0.280)
	1.3	21.50 (0.094)	13.52 (0.058)	11.69 (0.095)	10.85 (0.084)
	1.4	13.31 (0.057)	8.81 (0.037)	7.16 (0.054)	6.76 (0.047)
	1.5	9.00 (0.038)	6.27 (0.026)	4.977 (0.033)	4.71 (0.029)
	2.0	2.75 (0.010)	1.82 (0.005)	2.364 (0.008)	2.34 (0.006)
	k^+	2.0315	1.8919	1.411	1.468

10	1.0	201.12 (0.892)	199.84 (0.887)	200.5 (1.814)	201.10 (1.809)
	1.1	74.03 (0.329)	65.79 (0.293)	47.66 (0.421)	46.55 (0.405)
	1.2	33.71 (0.149)	26.84 (0.119)	17.43 (0.148)	16.57 (0.134)
	1.3	17.92 (0.078)	12.85 (0.055)	8.418 (0.064)	7.98 (0.058)
	1.4	10.88 (0.046)	7.18 (0.029)	5.129 (0.035)	4.97 (0.031)
	1.5	7.11 (0.030)	4.49 (0.018)	3.708 (0.021)	3.65 (0.019)
	2.0	2.21 (0.007)	1.39 (0.003)	2.137 (0.005)	2.12 (0.003)
	k^+	1.885	1.717	1.33	1.378

Table 4: ARL comparison for process dispersion with positive shift under double exponential distribution.

Sample n	Shift δ	D	RSS-D	RSS-D (2-out-of-2)	RSS-D (2-out-of-3)
5	1.0	200.61 (0.889)	199.86 (0.891)	200.00 (1.825)	200.00 (1.801)
	1.1	94.25 (0.422)	91.15 (0.407)	83.25 (0.745)	79.02 (0.709)
	1.2	51.29 (0.228)	48.36 (0.214)	41.50 (0.370)	38.04 (0.333)
	1.3	31.22 (0.137)	28.46 (0.124)	24.08 (0.206)	21.82 (0.182)
	1.4	20.65 (0.089)	18.37 (0.079)	15.44 (0.129)	14.10 (0.115)
	1.5	14.54 (0.063)	12.64 (0.054)	11.20 (0.090)	9.88 (0.076)
	2.0	4.75 (0.019)	3.85 (0.014)	4.25 (0.026)	3.91 (0.021)
	k^+	2.521	2.5075	1.745	1.798
8	1.0	200.95 (0.893)	200.36 (0.901)	200.31 (1.778)	200.45 (1.817)
	1.1	83.00 (0.369)	75.99 (0.336)	62.75 (0.557)	60.27 (0.525)
	1.2	40.38 (0.178)	34.99 (0.154)	26.41 (0.227)	24.44 (0.208)
	1.3	22.65 (0.099)	18.41 (0.080)	13.94 (0.115)	12.95 (0.102)
	1.4	14.22 (0.061)	10.96 (0.046)	8.69 (0.068)	7.99 (0.058)
	1.5	9.65 (0.04)	7.20 (0.030)	6.088 (0.043)	5.67 (0.038)
	2.0	2.95 (0.01)	2.11 (0.007)	2.67 (0.011)	2.59 (0.009)
	k^+	2.095	1.975	1.457	1.523
10	1.0	199.84 (0.895)	199.48 (0.894)	200.40 (1.808)	200.33 (1.813)
	1.1	75.13 (0.332)	67.76 (0.301)	53.10 (0.476)	50.23 (0.439)
	1.2	34.70 (0.153)	28.44 (0.130)	20.59 (0.179)	19.13 (0.161)
	1.3	18.79 (0.082)	14.20 (0.060)	10.51 (0.085)	9.65 (0.073)
	1.4	11.49 (0.049)	8.19 (0.034)	6.33 (0.046)	6.00 (0.040)
	1.5	7.66 (0.032)	5.26 (0.021)	4.50 (0.029)	4.25 (0.025)
	2.0	2.38 (0.008)	1.61 (0.004)	2.28 (0.007)	2.26 (0.005)
	k^+	1.938	1.792	1.371	1.424

Following are the important findings based on Table 3 and Table 4.

- When underlying process distribution is gamma, the ARL values of RSS-D chart and RSS-D charts with runs rules are smaller than that of the SRS-D chart for all sample sizes n and entire range of shifts considered. The RSS-D charts with runs rules show a better performance than RSS-D and SRS-D charts. The RSS-D chart with 2-out-of-3 runs rule is very effective for detecting shifts than 2-out-of-2 chart.
- When underlying process distribution is double exponential, we see that the ARL values of RSS-D chart and RSS-D charts with runs rules are smaller than that of the SRS-D chart for all sample sizes n and entire range of shifts considered. The RSS-D charts with runs rules show a better performance than RSS-D and SRS-D charts. The RSS-D chart with 2-out-of-3 runs rule is very effective for detecting shifts than 2-out-of-2 chart.
- The ARL performance of RSS-D chart and SRS-D chart under non-normal process distributions is higher than that for the normal process distribution with sample sizes $n = 5, 8$ and 10

- In general, under skewed and heavy tailed process distributions, the RSS-D chart is more efficient than the SRS-D chart for monitoring shifts in process dispersion.

7. Illustrative Example

In this section, we provide an example to illustrate the proposed chart. For this purpose, two datasets are generated which will be later referred as dataset 1 and dataset 2.

Dataset 1 contains 40 simple random samples each of size $n = 5$, out of which first 10 samples are generated from $N(0, 1)$ referring to in-control situation and remaining 30 samples are generated from $N(0, 1.2)$ referring to a shift in a process standard deviation. The D-statistics of these 40 samples are computed and are presented in Table 5. The graphical display of the control chart applied to data set 1 with $UCL=1.676$ is given in Figure 1.

Table 5: D-values for simulated data with SRS.

In-control Process		Out of control process with Shift (δ) =1.2					
No.	D-value	No.	D-value	No.	D-value	No.	D-value
1	1.04679	11	1.285956	21	0.884477	31	1.100131
2	1.165552	12	1.513122	22	1.113954	32	1.534435
3	1.468238	13	0.722581	23	0.887508	33	1.196817
4	0.4986	14	1.658094	24	1.431795	34	1.590498
5	0.574144	15	1.40786	25	1.294181	35	1.173697
6	1.231162	16	1.263608	26	1.52249	36	0.677852
7	0.84646	17	1.518884	27	1.379621	37	0.991026
8	0.728805	18	0.896195	28	1.349834	38	1.412045
9	0.767625	19	1.431129	29	1.029293	39	1.136405
10	0.920877	20	1.032453	30	1.981427	40	1.138973

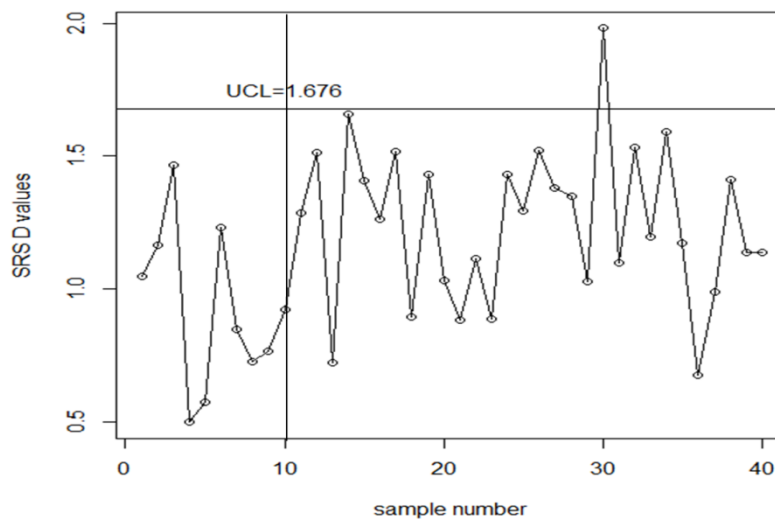


Figure 1: SRS-D control chart.

From Figure 1, we can see that D control chart gives first out of control signal at point no. 30, which is the only out of control signal.

Dataset 2 consists of 40 ranked set samples each of size $n = 10$ of which first 10 samples are generated from $N(0,1)$ referring to in-control situation and remaining 30 RSS samples are generated from $N(0, 1.2)$ referring to a shift in a process standard deviation. The D-statistics based on these 40 ranked samples are computed and are presented in Table 6. The graphical display of the control chart applied to data set 2 with $UCL=1.5082$ is given in Figure 2.

Table 6: *D-values for simulated data with RSS*

In-control Process		Out of control process with Shift (δ) =1.2					
No.	D-value	No.	D-value	No.	D-value	No.	D-value
1	0.76489	11	0.955223	21	1.658042	31	1.41809
2	1.11167	12	1.340428	22	1.200549	32	1.12013
3	0.928814	13	1.440937	23	1.057834	33	1.22928
4	1.029337	14	0.856949	24	1.827535	34	1.1807
5	1.046913	15	1.241251	25	1.452098	35	1.476113
6	1.197435	16	1.410242	26	0.963128	36	1.039491
7	1.258457	17	0.76041	27	1.096964	37	0.93336
8	1.291416	18	1.293803	28	0.947557	38	1.461749
9	1.262475	19	1.189676	29	1.917295	39	1.562749
10	0.871265	20	1.502025	30	1.601436	40	0.895917

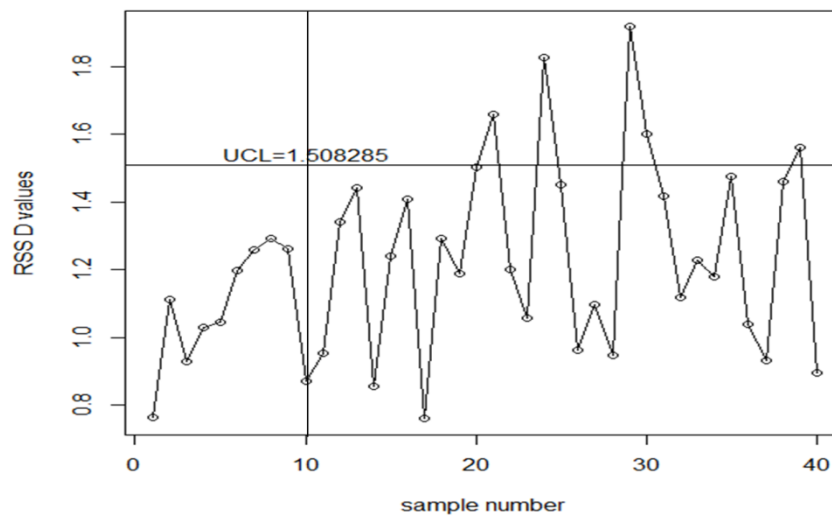


Figure 2: *RSS-D control chart*

From Figure 2, we can see that RSS D chart gives first out of control signal at point no.21 and total out of control signals given are 05. These signals are at point numbers 21, 24, 29, 30, 39.

From the above two figures, it is clear that RSS D chart signals earlier than D chart, also giving more number of signals than D chart. The above example clearly indicates that the proposed RSS-D chart performs better than D chart without RSS.

8. A Real Life Application

The use of RSS and other sampling schemes based on ranking of observations in SPC has increased in the recent years. As pointed out by Nawaz and Han [20], many a times, the actual measurement process has the constraints on the time and cost. Therefore, it is advisable to adopt such sampling designs which provide greater efficiency with relatively small sample sizes. Instead of taking physical measurements, sometimes it is easier to rank the sampling units on the basis of size, weight and volume by visual inspection or judgment. For example, in an automatic bottle filling plant, the actual quantification of filled volume of liquid in bottles, is costly and time consuming. Instead, it is easier and inexpensive to rank the bottles on the basis of the amount of filled volume of liquid in bottles by visual inspection before actual measurement.

Abujiya and Muttalak [21] constructed control charts for mean using double ranked set

sampling. To demonstrate the performance of proposed chart, they used a real dataset, which was originally taken by Muttalak and Al-Sabah [22]. The dataset consisted of 54 subgroups, each of size three. The subgroup data were collected using the SRS and RSS schemes. The data were taken from a Pepsi Cola production company in Khobar, Saudi Arabia. The quality characteristic under study was considered to be the unfilled part of the bottle, which was measured with the help of an instrument. By visual inspection of levels of the soft drinks, the bottles in each set were ranked. Then the samples were realized with newly proposed sampling techniques. Haq and Munir [23] also used the same dataset to demonstrate the performance of improved CUSUM charts for monitoring process mean based on RSS. The detection abilities of the CUSUM charts with RSS and ordered RSS schemes were compared.

Nawaz and Han [20] considered a real dataset taken from Tüfekci [24] for monitoring process location by using new RSS-based memory control charts. Since the original data was not collected from the quality control perspective, therefore, the available data was considered as a population and RSS samples were drawn from it to demonstrate the process monitoring.

In this section, we consider dataset on inside diameter measurements (mm) for automobile engine piston rings, given in Montgomery [1]. The dataset consists of total 40 subgroups each of size 5 collected using SRS method. Based on this dataset, we estimated the population mean (μ) and population standard deviation (σ) as $\hat{\mu} = 74.00361$ and $\hat{\sigma} = 0.010144$. To compare the performance of proposed RSS-D control chart with SRS-D chart, 8 RSS subgroups are formed using given 40 subgroups assuming that the ranking is perfect. To have fair comparison, out of available 40 SRS samples, 8 equidistant subgroups are selected. Using the estimated values of parameters, 22 SRS and 22 RSS samples are generated with a shift of size $\delta = 1.3$ in process dispersion. These values are tabulated in Table VII. The graphical display of the control charts applied to these data sets with $UCL=2.068$ for SRS and $UCL= 1.9985$ for RSS is given in Figure 3.

Table 7: D-values for SRS and RSS samples of Inside Diameter Measurements (mm) for Automobile Engine Piston Rings.

D values for SRS Samples							
In-control Process		Out of control process with Shift (δ) =1.3					
No.	D-value	No.	D-value	No.	D-value	No.	D-value
1	1.624979	9	1.373168	17	0.613449	25	2.546284
2	0.90859	10	1.278336	18	1.419153	26	2.014164
3	0.297039	11	0.993755	19	1.312875	27	1.862361
4	0.803753	12	0.417964	20	0.786504	28	1.006743
5	1.310467	13	0.778037	21	2.358553	29	0.645239
6	1.799708	14	1.593989	22	1.268609	30	1.791942
7	1.118265	15	1.286196	23	1.984099		
8	1.45025	16	0.627962	24	0.736714		

D values for RSS Samples							
In-control Process		Out of control process with Shift (δ) =1.3					
No.	D-value	No.	D-value	No.	D-value	No.	D-value
1	1.048373	9	0.749542	17	0.50747	25	2.456585
2	1.048373	10	1.469715	18	2.005007	26	1.333404
3	0.873645	11	1.827723	19	1.612167	27	1.005756
4	1.467723	12	1.485399	20	0.765135	28	1.214323
5	1.415304	13	1.57744	21	1.68864	29	0.801106
6	0.995955	14	1.691586	22	1.294557	30	1.63209
7	1.57256	15	2.131071	23	1.493243		
8	1.555087	16	1.51146	24	1.607816		

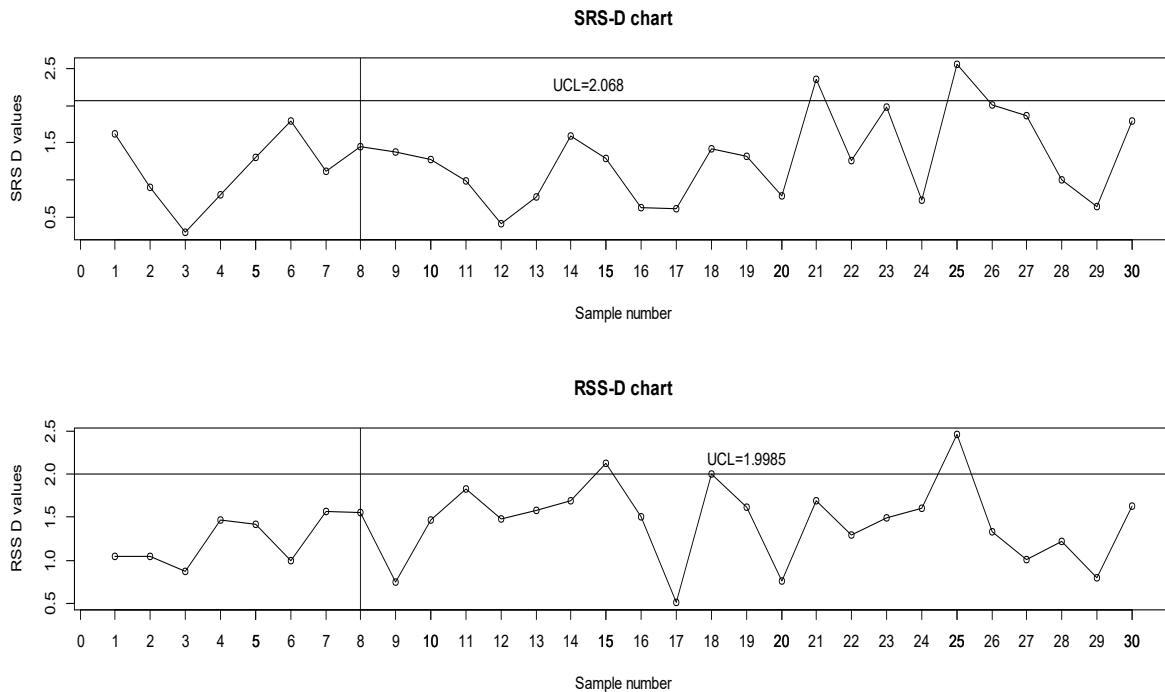


Figure 3: SRS-D and RSS-D control chart for a real dataset.

From Figure 3, we can see that the SRS D chart gives first out of control signal at point no 21 and total out of control signals given are 02, which are at point numbers 21,25. Whereas, RSS D chart gives first out of control signal at point no. 15 and total out of control signals given are 03. These signals are at point numbers 15, 18, 25.

9. Conclusions

In this paper, a control chart based on Downton’s estimator D is developed using ranked set sampling scheme for monitoring the process dispersion of normally and non-normally distributed processes. The performance of the proposed chart is investigated in terms of ARL and is compared with originally proposed D chart without RSS for monitoring the process dispersion. The proposed chart is found to be performing well for entire range of shifts in process dispersion under both normal and non-normal process distributions. The efficiency of proposed chart is further enhanced using two runs rules schemes. The chart with runs rules is found to be performing well for small and moderate shifts under normal process and performing uniformly better for entire range of shifts in dispersion under heavy tailed and skewed process distributions.

References

- [1] Montgomery, D. C. Introduction to Statistical Quality Control. *John Wiley & Sons* 2009; 6th edn. Inc. New York.
- [2] McIntyre, G. A. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3: 385–390.
- [3] Takahasi, K. and Wakimoto, K. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of Institute of Statistical Mathematics*, 20: 1–31.
- [4] Kvam, P. H. and Samaniego, F. J. (1993). On maximum likelihood estimation based on ranked set sampling with application to reliability. In: A. Basu (Ed.) *Advanced in Reliability*, 215–229.

- [5] Kvam, P. H. and Samaniego, F. J. (1994). Non-parametric maximum likelihood estimation based on ranked set samples. *Journal of American Statistical Association*, 89: 526–537.
- [6] Salazer, R. D. and Sinha, A. K.(1997). Control chart \bar{X} based on ranked set sampling. *Comunicacion Tecnica*, No. 1-97-09 (PE/CIMAT).
- [7]Muttalak, H. and Al-Sabah, W. (2003).Statistical quality control based on ranked set sampling. *Journal of Applied Statistics*, 30 (9):1055-1078.
- [8] Al-Naseer, A. D. and Al-Rawwash, M. A.(2007). A control chart based on ranked data. *Journal of Applied Science*, 7(14): 1936-1939.
- [9] Al-Omari, A. I. and Al-Naseer, A. D. (2011). Statistical quality control limits for the sample mean chart using robust extreme ranked set sampling. *Economic Quality Control* 2011; 26(1): 73-89.
- [10] Al-Omari, A. I. and Haq, A. (2012). Improved quality control charts for monitoring the process mean using double ranked set sampling methods. *Journal of Applied Statistics*, 39(4): 745-763.
- [11] Al-Rawwash, M., Al-Naseer, A. D., Al-Omari, M. I. (2014). Quality control charts using different ranked set sampling schemes. *Mathematical Methods in Science and Mechanics*, 173-177.
- [12] Mehmood, R., Riaz, M., Does, R. J. M. M. (2014). On the application of different ranked set sampling schemes. *Quality Engineering*, 26: 370-378.
- [13] Abbasi, S. A. (2019). Location charts based on ranked set sampling for normal and non-normal processes. *Quality and Reliability Engineering International*, 35 (6): 1603-1620.
- [14] Tahir, A., Zaman, B., Atir, A., Riaz M., Abbasi, S. A.(2019). On improved dispersion control charts under ranked set schemes for normal and non-normal processes. *Quality and Reliability Engineering International, Special Issue*: 1-29.
- [15] Klein, M. (2000). Two alternatives to the Shewhart \bar{X} control chart. *Journal of Quality Technology*, 32(4): 427-431.
- [16] Antzoulakos, D. L and Rakitzis, A. C. (2010). Runs rules schemes for monitoring process variability. *Journal of Applied Statistics*, 37 (97): 1231-1247.
- [17] Cheng, C. S. and Chen, P. W. (2011). An ARL unbiased design of time-between-events control charts with runs rules. *Journal of Statistical Computation and Simulation*, 81 (7): 857-871.
- [18] Abbasi, S. A. and Miller, A. (2011). D chart: An efficient alternative to monitor process dispersion. *Proceedings of the World Congress on Engineering and Computer Science (WCECS 2011)*, 2 San Francisco, USA.
- [19] Downton, F. (1966). Linear estimates with polynomial coefficients. *Biometrika*, 53: 129-141.
- [20] Nawaz, T. and Han, D. (2020). Monitoring the process location by using new ranked set sampling-based memory control charts. *Quality Technology & Quantitative Management*, 17(3), 255-284.
- [21] Abujiya, M. A. and Muttalak, H. (2004). Quality control chart for the mean using double ranked set sampling. *Journal of Applied Statistics*, 31(10): 1185-1201.
- [22] Muttalak, H. and Al-Sabah, W. (2003). Statistical quality control based on ranked set sampling. *Journal of Applied Statistics*, 30(9): 1055-1078.
- [23] Haq, A. and Munir, W. (2018). Improved CUSUM charts for monitoring process mean. *Journal of Statistical Computation and Simulation*, 88(9): 1684-1701.
- [24] Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems* 60: 126-140.

MODELING THE RELIABILITY OF TRANSPORT UNDER EXTREME CONDITIONS OF OPERATION AS A QUEUEING SYSTEM WITH PRIORITIES

M.D. Katsman

•

The Joint-Stock Company of Railway Transport of Ukraine "Ukrzaliznytsia", Kyiv, Ukraine
mdkatsman@gmail.com

V.I. Matsiuk

•

National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine
vimatsiuk@gmail.com

V.K. Myronenko

•

State University of infrastructure and technologies, Kyiv, Ukraine
pupil7591@gmail.com

Abstract

The article presents a simulation model of a queueing system (QS) with a queue and relative priority, which can be used to manage the reliability of transport systems under resource constraints. The developed simulation model combines agent and discrete-event simulation principles and allows studying queueing systems in terms of establishing regularities: probabilities (service, failure, push-out), time delays (waiting in a queue, under service), queue sizes, order of queue formation upon arrival of consumers of different priority.

As a result of the research, dependencies were obtained for the probability of servicing higher priority consumers depending on the intensities of their arrival and service; probabilities of servicing lower priority consumers depending on the intensity of service and servicing of higher priority consumers; the probability of "pushing out" lower priority consumers from the QS by higher priority consumers depending on service intensities and the arrival of high priority consumers.

Keywords: transport system, agent modeling, simulation modeling, queueing systems, absolute priority, arrival and service intensity, service and push-out probabilities, consumer service probabilities

1. Introduction

In mathematics and applied research, the queueing theory is widely used in modeling the functioning of real systems, information [1], computing, energy, medical [2], biological [3], transport [4][5] and others.

In the field of transport, the application of the queueing theory methods is quite traditional.

Recently, the events of the aggressive war of Russia, which it is waging against Ukraine during 2022-23, force us to formulate new practical problems that are solved with the use of these methods, in particular, in queuing system (QS) with priorities.

Military operations in Ukraine have repeatedly aimed at destroying the railway infrastructure, especially on electrified lines. For historical and technological reasons, almost 4 times more electric locomotives than diesel locomotives were used for transportation on Ukrainian railways. With such a disproportion, diesel locomotives cannot replace the entire fleet of electric locomotives, to serve all of transportation traffic on electrified lines, in the event that the power supply is stopped on these lines. Likewise, it is not possible to transfer all trains from destroyed/damaged electrified lines to non-electrified lines, due to the defined geography of transportation between consignors and consignees.

Then there were "problems with priorities" – for example, to which alternative route the train of a certain type should be directed, if the capacity of the route and the number of locomotives (diesel or electric) are limited. At the same time, the priority of dispatching the trains can be determined according to different criteria, separately for passenger and cargo service.

The priorities of passenger transportation can be assigned, for example, as a special military personnel transportation, the first term evacuation, and for the rest of passenger trains, it is possible, ranking them by the cost of all tickets sold up to an hour divided by the time remaining to the planned train arrival to final destination according to the timetable. In the event of cargo transportation, the priorities may be assigned to cargoes and trains especially important for the state and defense purposes, while the rest of cargoes/trains can be prioritized according to the value of cargoes and the time available for their delivery to the destination station. After ranking the priorities, from the highest to the lowest, the decision should be made on the appropriate use of locomotive fleet and railway line capacity on alternative routes. The priming and flexibility of these decisions are only growing, if the mathematical apparatus for optimizing the QS with priorities is applied.

In such QSs, if they have consumers (customers, trains, cargoes, shippers etc.) with different priorities, then the consumers with higher priority are served earlier. Priorities can be absolute. In QSs with absolute priorities, servicing the consumers of the lower priority is interrupted, when higher priority consumers arrive. While the lower priority consumer, for which the servicing which was interrupted, turns back to the queue. It is only serviced again when there are no consumers of higher priority in the queue. In QSs with relative priorities, the servicing of a consumer is always finished once it has started, even if a higher priority consumer arrives at that time.

Special attention should be paid to QSs with low rates of both arrival and servicing of higher priority customers. That is, the time intervals between high priority consumer arrivals and the time periods for servicing such consumers are long.

For example, with all the current non-predictability of warfare, it is known that the time intervals between infrastructure-destroying strikes and the duration of its recovery are relatively long compared to the intervals between trains and the duration of transportation in normal, peaceful conditions. Then the priority is given to transportation related to infrastructure restoration, and it is to ensure these transportations that locomotives and other resources are directed.

Each priority class of consumers can have a separate list of ranked consumers and its own queue. Consumers from the lower priority list are served only after the last consumer from the higher priority list has been served. For all arrivals, it is assumed that the processes of receipt of requirements are independent, Poisson and do not depend on service durations.

QSs with priorities have been sufficiently studied for a long time. With the help of Laplace transformations, dependences on the determination of the characteristics of such QSs for various service disciplines were obtained [6]. During the study of QS with different priorities – priorities without interruption of service and with interruption of service and additional after-service – dependences on finding the mean waiting time in the queue, the average time spent in the system

for k-priority consumers were obtained [7].

In the monograph [8], a partially modified QS was studied using the MathLab Simulinc simulation model development environment (with the SimEvent and StateFlow libraries). It is shown that with the help of dynamic priorities, the priority of a non-priority customer is increased once, while the probability of serving these customers increases. However, such a characteristic of QS as the probability of serving the flow of arriving consumers decreases, since non-priority consumers are served longer than priority consumers.

With the help of simulation model development environment (AnyLogic, with Java SE libraries) and tools of queuing theory, agent and discrete-event principles, technological risks and failures in transport and logistics systems were studied [9] [10]. As a result, the regularity of the influence of the number of service channels on the mean service rate with an unchanged total arrival and service rates in the system was established. However, the priority of applications was not considered in these studies.

In [11], methods of queuing theory of Markov and non-Markov types are applied to simulate the resistance of security personnel to a malicious group with a random number of criminals in the group and various ways of organizing the actions of such personnel. And in research [12], mathematical models of the queuing system are considered, simulating the processes of abnormal situations during railway transportation of dangerous goods, as well as the processes of elimination of ecologically dangerous consequences of such events.

In this work, a QS with a queue and relative priority is investigated, with customers of higher and lower priorities arriving at rates λ_H, λ_L , the customer servicing times of are exponential with rates μ_H, μ_L . A higher-priority customer that arrives to the QS "pushes" a lower-priority customer out after its service has finished and takes a place in front of it if it is in the queue for servicing. A lower priority customer that has been pushed out leaves the QS unserved if there are no places in the queue and is queued if a place exists.

The peculiarity of the proposed experiment, conducted for the purpose of studying the functioning of the model, is that its input receives a stream of events that have negative consequences that significantly affect the functioning of this model (for example, reduce the intensity of service in the system). The flow of these "negative" events (the higher priority flow) has low intensity compared to the lower priority flow and takes a significant amount of time to service the requirements of that flow.

The complexity of the implementation of the queuing theory analytical methods is determined by the voluminous forms of mathematical description for such a kind of QSs. Therefore, it is advisable to use computer simulation methods when solving the scientific and applied problems.

2. Methods

2.1 Development of simulation model for a QS with priority customers

The model is developed on the basis of discrete-event and agent principles. When developing the model, standard blocks of the Process Modeling Library were used.

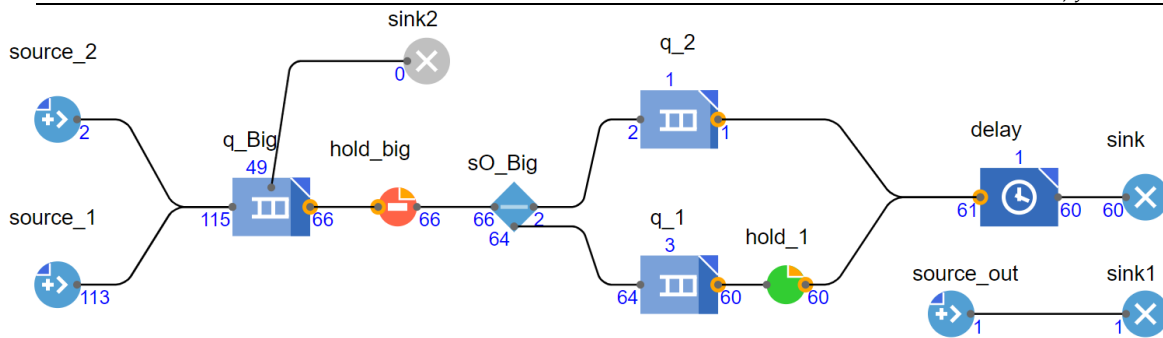


Figure 1: The main window of the simulation model in the AnyLogic environment

For the possibility of prioritizing customers based on the *Entity* base class of agents, a *MyAgent* population of agents (customers) was created with the *priority* parameter added. During modeling, each new requirement will be assigned a priority level:

priority = 2.0 – higher priority customers with the value of the parameter;

priority = 1.0 – lower priority customers with the parameter value.

The simulation of demand arrival and service starts with *source_1* and *source_2* blocks, which form the lower and higher priority demands, respectively.

When the generated agents (customers) exit from the source blocks, the priority parameter is assigned the value of the priority level of the customers: The simulation of customers arrival and service starts with *source_1* and *source_2* blocks, which form the lower and higher priority customers, respectively. When the generated agents (customers) exit from the source blocks, the priority parameter is assigned the value of the priority level of the requirements:

for lower priority agents (customers):

```
"agent.priority = 1;";
```

for higher priority agents (customers):

```
"agent.priority = 2;".
```

Next, the agents (customers) enter the general queue *q_Big*. This element is used to form a general queue of customers that will be accepted by the QS and processed in it. Queue *q_Big* is not an element of the QS under study and is outside the QS.

The queue in the *q_Big* element is formed according to the general principle of FIFO ("first in, first out"), taking into account the priority of customers and the model time of customers generation. That is, the FIFO principle is used within the customers of higher and lower priorities. To verify the process of queue formation, all information about the presence of customers is displayed in a text field: the higher the entry in the visual representation of the formation of the queue (Fig. 2), the closer the customer is to the exit from the *q_Big* block. Information about all customers is presented in the form:

```
"requirement generation serial number | priority level of the requirement | model time of requirement generation".
```

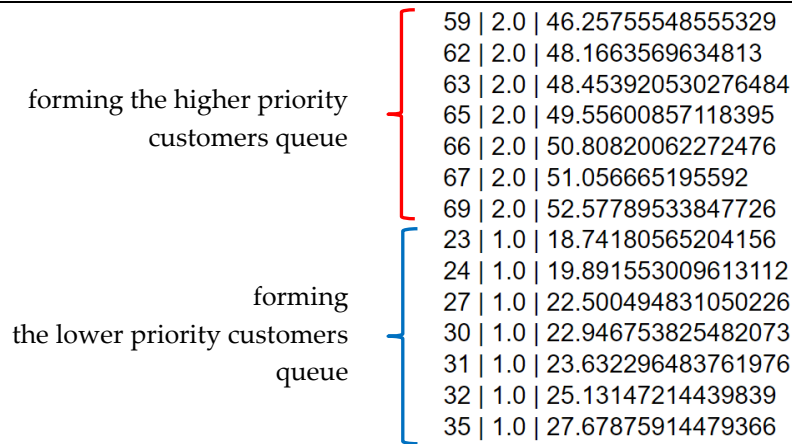


Figure 2: Visual display of queue formation in the q_Big block

Since the number of places in the queue is limited in the QS itself, the mechanism of separation of requirements by priority and formation of separate queues from requirements of higher and lower priorities, respectively (blocks q_2 and q_1) is used. Separation of requirements is carried out by the sO_Big element. Passage of requirements from the general q_Big queue to the QS queue is carried out using a block of the *hold type* ($hold_big$, Fig. 1). The algorithm is implemented through *Java* code using the built-in function $fun_hold_big()$:

```
"if ((q_2.size() + q_1.size()) < places){
hold_big.unblock();
}».
```

The function $fun_hold_big()$ is implemented every time any customer is received for the q_Big queue or the delay block. The $fun_hold_big()$ function algorithm checks the total number of requests in the QS queue, and if the number of customers is less than the set number of places in the queue ($places$), the $hold_big$ block is unblocked ($hold_big.unblock()$ procedure).

The model also implements the algorithm of passing agents (requirements) into the middle of the system "by priority" and pushing agents (requirements) of lower priority out of the system by agents (requirements) of higher priority. The specified algorithm is implemented using blocks of the hold type ($hold_1$) and the additional function $fun_Logic_HIghPr()$ with *Java* code:

```
«if ((q_2.size() + q_1.size()) == places && q_1.size() > 0){
if (q_Big.size() > 0 && q_Big.get(0).priority == 2){
q_1.remove(q_1.get(0));
source_out.inject(1);
hold_big.unblock();
}
}».
```

According to the algorithm, if an agent (customer) of higher priority arrives at the QS and there are no free places in the QS queue, the agent (customer) of lower priority is "pushed out" of the QS queue (block q_1) and is replaced by a corresponding agent (customer) of a higher priority from the general queue (block q_Big).

To simulate the processing of agents (customers), a *Delay* block with one location (one server device) is used. Agents (customers) of higher priority are served first and proceed immediately from the q_2 block to the *Delay* block without delay.

The arrival of lower priority agents (customers) is regulated functionally using the *hold_1* block and the *fun_Logic_LowPr()* function with *Java* code:

```
«if (q_2.size() == 0 && q_1.size() > 0){
    hold_1.unblock();
}».
```

When the agent (customer) passes the *hold_1* block using *Java* code, the *hold_1* block is blocked: «hold_1.block();»

During simulation, data is collected on the probability of servicing or pushing out the customers according to the general principle:

$$\xi = \frac{\sum N_{i,success}}{\sum N_i},$$

where $\sum N_{i,success}$ is the number of *i* - type events that were processed in the system or pushed out of the system;

$\sum N_i$ is the total number of *i* - type events generated during the simulation.

In addition, statistical information is generated regarding the average size of queues and the time that requests are in service or waiting in the QS.

2.2 Results of experiments and discussion of the results

To study the regularities of queue formation and service probabilities in the system, a series of experiments on the sensitivity of the model were conducted, where the variable parameter is the arrival flow of higher priority customers ($\lambda_H \in [0,001; 0,5]$) at different values of the service rate ($\mu_H \in [0,001; 0,5]$). Other raw data of the experiments are presented in Table 1.

Table 1: *Input parameters of the experiments*

Modeling parameter	Symbol	Range of values	Distribution density
Higher priority customers arrival rate (High)	λ_H	0,001 – 0,5; step 0,001	exponential
Lower priority customers arrival rate (Low)	λ_L	0,5 – const	exponential
Higher priority customers service rate (High)	μ_H	0,001 – 0,5; крок 0,1	exponential
Lower priority customers service rate (Low)	μ_L	0,5 – const	exponential
Number of servers		1	
Number of places in the QS queue		4	
Number of priorities		2	
Model time unit		hour	
Model time duration		6 months (4392 hrs.)	

The main simulation results are shown in Fig. 3 - 8.

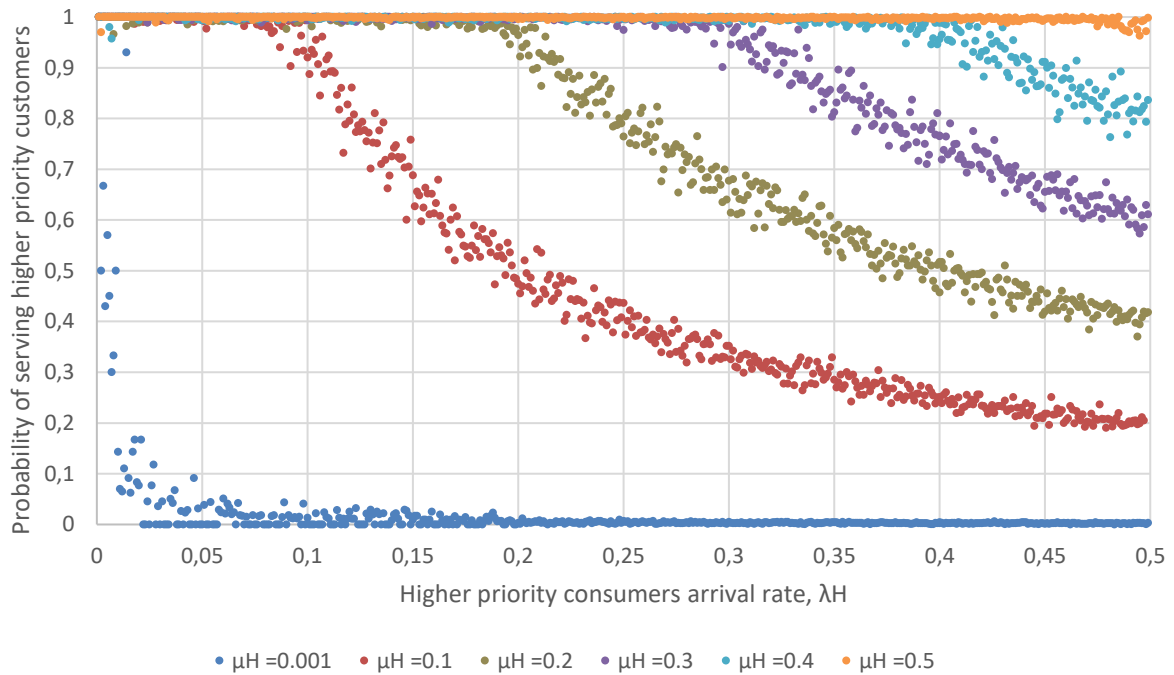


Figure 3: Probability of servicing higher priority consumers depending on arrival and service rates

The data of fig. 3 indicate that the probability of servicing higher priority customers decreases as the intensity of servicing higher priority customers approaches 0.5 value. Moreover, the "breaking point" at the beginning of the decrease of the probability from 1.0 begins at the moment of equilibrium of the intensities of arrival and service of higher priority customers, that is, at this point:

$$\lambda_H = \mu_H.$$

The descending part of the dependence of the probability of servicing higher priority customers (Fig. 4) is closely described by a power function of the form:

$$y = kx^{-c},$$

where k are c coefficients;

for $\mu_H = 0,1$:	$y = 0.0994x^{-0.998},$
for $\mu_H = 0,2$:	$y = 0.2032x^{-0.981},$
for $\mu_H = 0,3$:	$y = 0.3067x^{-0.97}.$

The result of approximating the periods of descending probability of service of higher priority customers depending on the intensities of service and arrival is presented in Fig. 4.

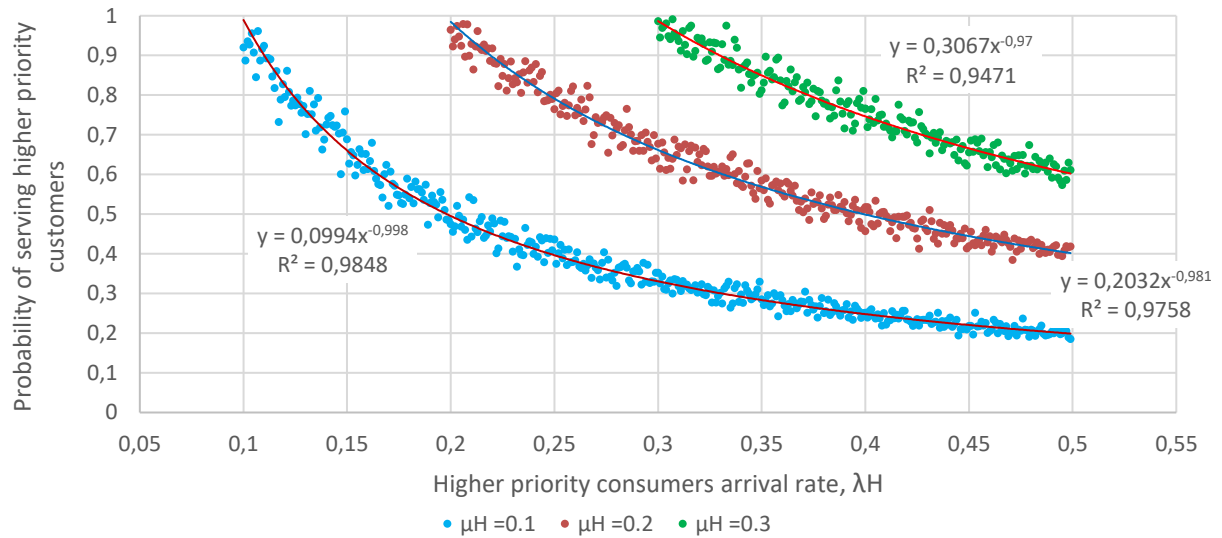


Figure 4: Approximation by power-law dependences of periods of descending probability of servicing higher priority customers depending on the rates of arrival and service

From those obtained in fig. 4 data can be recorded:

$$\begin{cases} P_{\mu_H} = \mu_H \lambda_H^{-1}, & \lambda_H > \mu_H, \\ P_{\mu_H} = 1, & \lambda_H \leq \mu_H. \end{cases}$$

If we round the value of k to tenths, we get:

$$P_{\mu_H} = \mu_H \lambda_H^{-C}, \quad \lambda_H \geq \mu_H,$$

then finally:

$$\begin{aligned} P_{0,1} &= 0,1 \lambda_H^{-1}, \quad \lambda_H \in [0,1, \dots, 0,5], \\ P_{0,2} &= 0,2 \lambda_H^{-1}, \quad \lambda_H \in [0,2, \dots, 0,5], \\ P_{0,3} &= 0,3 \lambda_H^{-1}, \quad \lambda_H \in [0,3, \dots, 0,5]. \end{aligned}$$

The dependence of higher priority customers servicing rate μ_H on the probability P_H and higher priority customers arrival rate λ_H is presented in the table. 2.

Table 2: Dependence of μ_H value on P_H and λ_H values

P_H	λ_H	0,1	0,2	0,3	0,4	0,5
1.00	$\mu_{H1,0}$	0.10	0.20	0.30	0.40	0.50
0.90	$\mu_{H0,9}$	0.09	0.18	0.27	0.36	0.45
0.80	$\mu_{H0,8}$	0.08	0.16	0.24	0.32	0.40
0.70	$\mu_{H0,7}$	0.07	0.14	0.21	0.28	0.35
0.60	$\mu_{H0,6}$	0.06	0.12	0.18	0.24	0.30
0.50	$\mu_{H0,5}$	0.05	0.10	0.15	0.20	0.25
0.40	$\mu_{H0,4}$	0.04	0.08	0.12	0.16	0.20
0.30	$\mu_{H0,3}$	0.03	0.06	0.09	0.12	0.15
0.20	$\mu_{H0,2}$	0.02	0.04	0.06	0.08	0.10
0.10	$\mu_{H0,1}$	0.01	0.02	0.03	0.04	0.05

Graphs of dependence of μ_H values on P_H and λ_H rate for $P_H = 1.0$; $P_H = 0.7$; $P_H = 0.5$ and $P_H = 0.3$ are presented in Fig. 5.

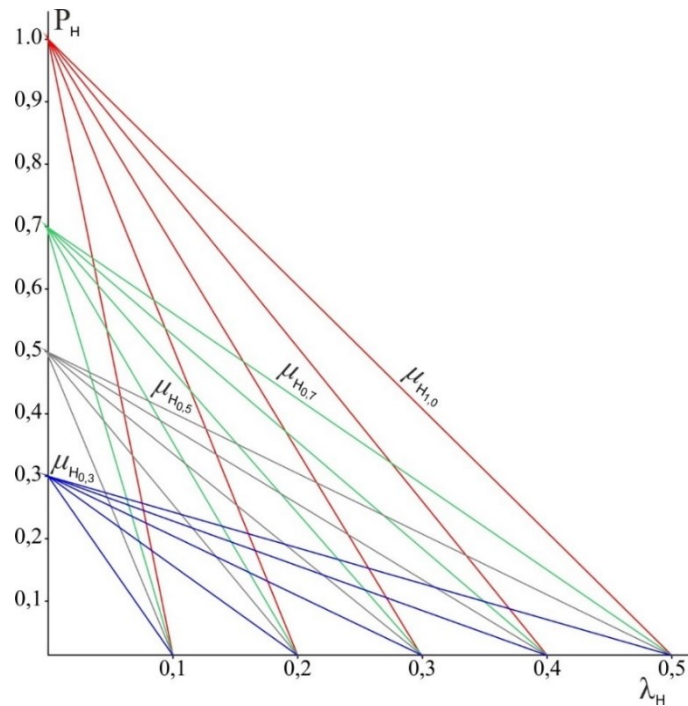


Figure 5: Graph of the dependence of μ_H values on λ_H values at $P_H = 1.0$; $P_H = 0.7$; $P_H = 0.5$ and $P_H = 0.3$

Table 2 and fig. 5 show that to increase the value of $P_H(\lambda_H, \mu_H)$ it is necessary to increase the values of μ_H . Thus, to increase P_H by 0.1, it is necessary to increase μ_H by 1% at $\lambda_H = 0.1$, at $\lambda_H = 0.2$ – by 2%, etc., at $\lambda_H = 0.5$ the increase is 5% from the previous values of μ_H .

The graph of P_L probabilities of servicing lower-priority customers depending on service μ_H and arrival λ_H intensities of higher-priority customers is presented in Fig. 6.

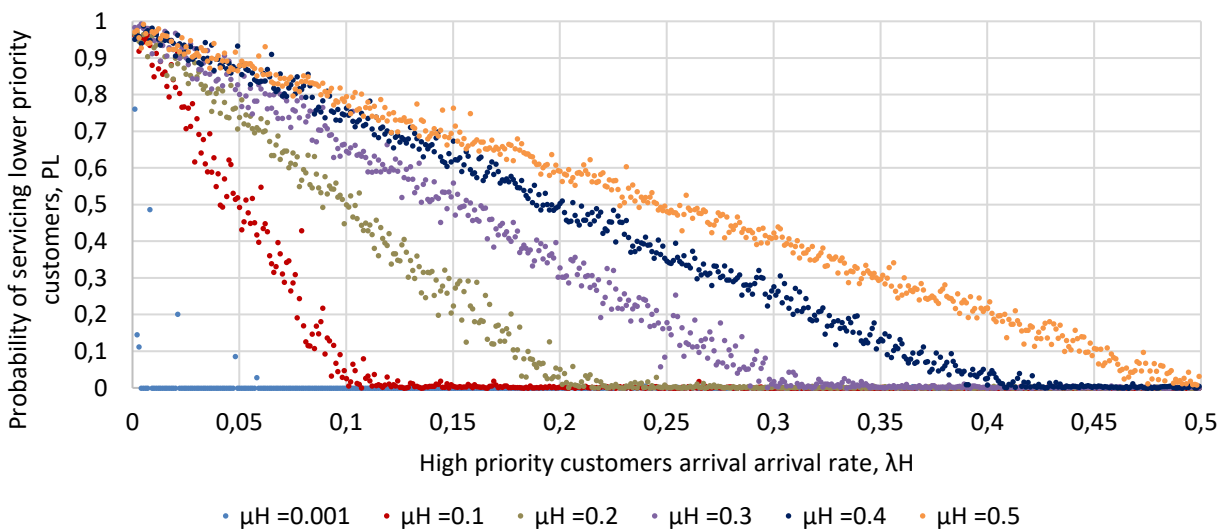


Figure 6: Probability of servicing P_L requirements of lower priority depending on higher priority customers arrival and servicing rate

From Fig. 6, it can be seen that the probability of servicing lower-priority customers has a decreasing nature, while the "sharpness" of the fall is inversely proportional to the intensity of servicing higher-priority customers.

Consider the function $P_L(t)$ depending on the intensity μ_H and the time $t_H = 1 / \lambda_H$ of the interval between the customers of the highest priority of the incoming flow. The function $P_L(t)$ is a function of the probability of failure-free service of lower priority consumers, which is linear in nature. The mean time of failure-free service of these consumers is equal to the area bounded by the line $P_L(t)$ and the coordinate axes $P_L(t)$ and t_H . The dependence of the probability value $P_L(t)$ of servicing lower priority consumers on the intensity μ_H of servicing higher priority consumers and the duration of the time interval t_H between these consumers in the arrival flow is shown in Fig. 7.

Indeed, with an increase in the value of μ_H , the average service duration decreases, at the same time, with an increase in the value of λ_H , the interval between consumers in the arrival flow of higher priority consumers decreases, which leads to an increase in the average number of such consumers, and therefore to a higher load on the service channel, which in turn, prevents consumers of lower priority from entering the service channel and reduces the service probability P_L .

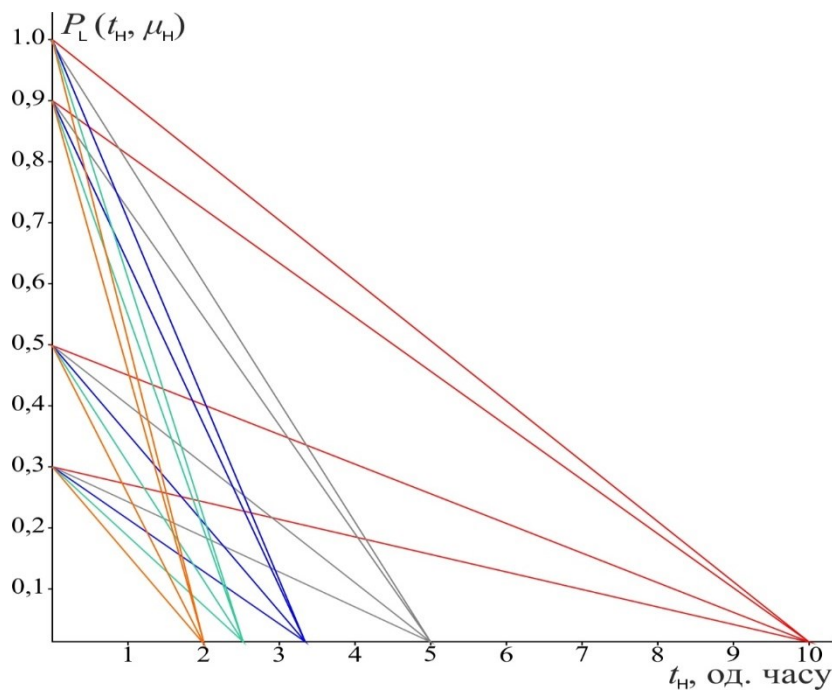


Figure 7: Graphs of the dependence of the probability of $P_L(t_H, \mu_H)$ on the values of μ_H and t_H

Fig. 7 shows that when the value of t_H decreases at $P_L(t_H, \mu_H) = \text{const}$, or the value of $P_L(t_H, \mu_H)$ at $t_H = \text{const}$, the value $\bar{t}_L = 0.5P_L(t_H, \mu_H)t_H$ decreases. To increase the value of $P_L(t_H, \mu_H)$ at a certain value of t_H , it is necessary to increase the value of μ_H , that is, to provide more intensive service of higher priority customers.

Fig. 8 shows graphs of the dependence of the probability of pushing out P_{LP} consumers of lower priority.

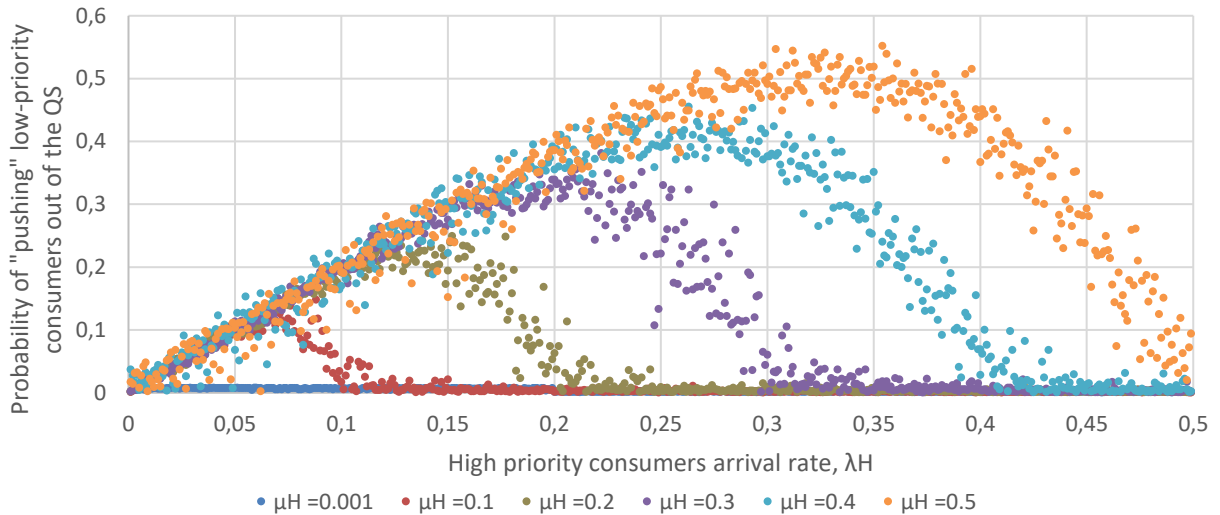


Figure 8: Probability P_{LP} of "pushing out" from the QS lower priority consumers by high priority consumers depending on high priority consumers arrival and service rates

The processes that are associated with the increase of P_{LP} are described in detail above, the decrease of P_{LP} after reaching its maximum is associated with a decrease in the number of consumers of lower priority and the load of the service channel. The value of $\lambda_{H,extr}$ corresponds to the maximum value of P_{LP} . To determine the value of $\lambda_{H,extr}$, dependences of the probability of P_{LP} were obtained, approximated by a polynomial of the sixth degree, of the form:

$$y = a_6x^6 + \dots + a_1x + C.$$

The coefficients of the polynomials approximating the probabilities of P_{LP} of the value $\lambda_{H,extr}$ are presented in the table. 3.

Table 3: Coefficients of polynomials approximating the probabilities of "pushing out" the QS of low-priority consumers

Coefficient of the polynomial	$\mu_H = 0.1$	$\mu_H = 0.2$	$\mu_H = 0.3$	$\mu_H = 0.4$	$\mu_H = 0.5$
a_6	-1767609	97550.53	4593.061	1980.917	983.323
a_5	431830.5	-54236	-3453.6	-2276.31	-1392
a_4	-42096	10862.98	857.173	966.0881	721.876
a_3	1791.001	-1046.51	-114.584	-212.099	-184.777
a_2	-34.143	48.5078	8.249455	23.6418	23.30407
a_1	2.292621	1.042772	1.748402	0.770417	0.615751
a_0	-0.00172	0.004007	-0.0002	0.017116	0.019575
Extremum, $\lambda_{H,extr}$	0.071	0.135	0.198	0.2628	0.328

Knowing the extrema of the probabilities of pushing out lower priority consumers and the corresponding rates of arrival flows and servicing higher priority consumers is of practical importance, for example, for making decisions about the distribution of trains of different priorities (repair, passenger, freight, other types) on different nodes and routes of the transport network.

Conclusions

1. Taking into account the practical experience of functioning and ensuring the reliability of the railway transport system in extreme conditions of military operations, destruction and restoration of infrastructure, shortage of time and fleet of vehicles, it is proposed to formalize the functioning of this system as a queuing system (QS) with priorities, with variable parameters of arrival and servicing the consumers of higher and lower priorities, which allows making more informed decisions for the management of such systems.

2. The necessity of using, along with analytical methods, simulation modeling tools has been proven, in particular, a simulation model has been developed and implemented, which combines agent and discrete-event simulation principles and allows studying Qs when consumers of different priority are arriving, namely in the part of establishing regularities: probabilities (service, refusal, push-out), time delays (waiting in a queue, under service), queue sizes, order of queue formation when customers of different priority arrive.

3. The resulting dependencies:

- the probability of servicing higher priority consumers depending on arrival and servicing rates;

- probabilities of servicing lower priority consumers depending on arrival and servicing rates of higher priority consumers;

- probability of "pushing" lower priority consumers out the QS by higher priority consumers, depending on arrival and service intensities of higher priority consumers;

- values of the intensity of service of higher priority consumers μ_H on the probability P_H and the intensity of their arrival λ_H ;

- values of the probability of P_L on the values of the intensity of service μ_H of the consumers of the highest priority and the duration of the interval t_H between these consumers in the arrival flow.

4. Experimental data of modeling the probability of servicing higher priority customers depending on their arrival and service rates, as well as the approximation of these data by empirical dependencies allow us to propose ratios that simplify the calculations of Qs with priorities, namely:

$$\begin{cases} P_{\mu_H} = \mu_H \lambda_H^{-1}, & \lambda_H > \mu_H, \\ P_{\mu_H} = 1, & \lambda_H \leq \mu_H. \end{cases}$$

References

[1] I. Makushenko, I. Usar, H. Livinska, and M. Sharapov, "Optimal threshold strategies for retrieval systems with queue," *J. Comput. Appl. Math.*, vol. 427, p. 115136, Aug. 2023, doi: 10.1016/J.CAM.2023.115136.

[2] A. Aghsami, S. R. Abazari, A. Bakhshi, M. A. Yazdani, S. Jolai, and F. Jolai, "A meta-heuristic optimization for a novel mathematical model for minimizing costs and maximizing donor satisfaction in blood supply chains with finite capacity queueing systems," *Healthc. Anal.*, vol. 3, p. 100136, Nov. 2023, doi: 10.1016/J.HEALTH.2023.100136.

[3] H. Jahani, A. E. Chaleshtori, S. M. S. Khaksar, A. Aghaie, and J. B. Sheu, "COVID-19 vaccine distribution planning using a congested queuing system—A real case from Australia," *Transp. Res. Part E Logist. Transp. Rev.*, vol. 163, p. 102749, Jul. 2022, doi: 10.1016/J.TRE.2022.102749.

[4] A. Mazaraki, V. Matsiuk, N. Ilchenko, O. Kavun-Moshkovska, and T. Grygorenko, "Development of a multimodal (railroad-water) chain of grain supply by the agent-based simulation method," *Eastern-European J. Enterp. Technol.*, vol. 6, no. 3 (108), pp. 14–22, Dec. 2020, doi: 10.15587/1729-4061.2020.220214.

[5] K. Abdulaziz Alnowibet, A. Khireldin, M. Abdelawwad, and A. Wagdy Mohamed, "Airport terminal building capacity evaluation using queuing system," *Alexandria Eng. J.*, vol. 61, no. 12, pp. 10109–10118, Dec. 2022, doi: 10.1016/J.AEJ.2022.03.055.

[6] "Вычислительные системы с очередями — Леонард Клейнрок." <https://www.livelib.ru/book/1001277388-vychislitelnye-sistemy-s-ocheredyami-leonard-klejnrok> (accessed Mar. 03, 2023).

[7] D. P. Bertsekas and R. G. Gallager, "Сети передачи данных," p. 486, 1986, Accessed: Mar. 03, 2023. [Online]. Available: <http://libarch.nmu.org.ua/handle/GenofondUA/60656>.

[8] "Управление моделью системы массового обслуживания с использованием динамических приоритетов - CORE Reader." <https://core.ac.uk/reader/80134930> (accessed Mar. 15, 2023).

[9] V. Matsiuk, N. Pchenko, O. Pryimuk, D. Kochubei, and A. Prokhorchenko, "Risk assessment of transport processes by agent-based simulation," *AIP Conf. Proc.*, vol. 2557, no. 1, p. 080003, Oct. 2022, doi: 10.1063/5.0105913.

[10] V. Matsiuk, O. Galan, A. Prokhorchenko, and V. Tverdomed, "An Agent-Based Simulation for Optimizing the Parameters of a Railway Transport System."

[11] M. D. Katsman, V. K. Myronenko, V. I. Matsiuk, and P. V. Lapin, "Approach to determining the parameters of physical security units for a critical infrastructure facility," *Reliab. Theory Appl.*, vol. 16, no. 1, pp. 71–80, 2021, doi: 10.24412/1932-2321-2021-161-71-80.

[12] M. D. Katsman, V. K. Myronenko, and V. I. Matsiuk, "Mathematical models of ecologically hazardous rail traffic accidents," *Reliab. Theory Appl.*, vol. 10, no. 1 (36), 2015, Accessed: Nov. 14, 2020. [Online]. Available: <https://cyberleninka.ru/article/n/mathematical-models-of-ecologically-hazardous-rail-traffic-accidents>.

SINGLE SAMPLING PLANS FOR VARIABLE INSPECTION BASED ON BIRNBAUM – SAUNDERS DISTRIBUTION

Saranya C R



Assistant Professor, Department of Statistics
Mar Ivanios College, Thiruvananthapuram 695 015, INDIA
saranyasreekumar17@gmail.com

Vijayaraghavan R



Professor & Head, Department of Statistics
Bharathiar University, Coimbatore 641 046, INDIA
rvijayrn@yahoo.com

Sathya Narayana Sharma K



Assistant Professor, Department of Mathematics
Vellore Institute of Technology, Vellore 632 014, INDIA
sharma14081992@gmail.com

Abstract

Sampling inspection by the method of variables is a well-known category of product control in which the discretion of acceptance or rejection is made hinge on some specific rule, which is framed according to the measurement of a quality characteristic under study. In this scenario, the quality feature under study is considered a continuous random variable that can be demonstrated using any continuous type statistical distribution. Birnbaum – Saunders distribution is a continuous probability distribution, which is having numerous applications in various fields. Application of Birnbaum – Saunders distribution is considered in this paper for establishing acceptance sampling inspection plans for variables based on the examination of units from a single sample. Numerical illustrations are given to demonstrate the application of proposed sampling plan. In addition, the results of numerical illustrations are explained with the help of simulated data.

Keywords: Sampling inspection plans, variable sampling, Birnbaum – Saunders distribution, non-normality, skewness.

I. Introduction

Sampling inspection plans or acceptance sampling plans play a cardinal role in determining the quality of products by reviewing the sample of units taken from the lot of products. More often, the product control methodologies are grouped into two categories, namely, sampling inspection by the method of attributes and sampling inspection by the method of variables. In sampling inspection by the method of attributes, selection of one or more samples is made from the manufactured lot of items and the sampled units are classified into defective (non-conforming) or non-defective (conforming) according to some prescribed decision criteria. The choice of accepting or rejecting the lot is taken based on some explicit criteria.

In sampling inspection by the method of variables, a sample of manufactured commodities is picked from the lot, and then the quality feature of interest is measured and recorded. The

decision about the lot acceptance or lot disposition is interpreted according to some specific criteria, which are framed based on the measurement of the quality variable under study. This method is applicable to the cases where the quality feature under study is a continuous random variable, which is quantifiable on a continuous scale.

A peculiar aspect of variable sampling is that the quality feature under study is a continuous random variable which can be demonstrated using any continuous type probability distribution from the literature of statistical theory. A noteworthy advantage of variable sampling is that the decision is made about the lot quality in accordance with the exact measurements of the quality variable. It provides more accurate information about the product quality than compared to the sampling inspection by attributes. Variable sampling plans can be used with same level of protection as attribute sampling plans, but with lesser number of sampling units, showing its efficiency with respect to the sample size.

Under normality assumption of the quality feature, many researchers have devised the variable sampling plans. When the normality assumption is sheered or when the quality variable shows some kind of skewness, the standard variable sampling plans based on normal distribution are no longer applicable. In such situations, the need for the application of non-normal distributions or skewed distributions arises. Many researchers have initiated the works on variable sampling plans for non-normal populations. Zimmer and Burr [1] considered Burr distribution for constructing variable sampling plans based on the measures of skewness and kurtosis. Takagi [2] designed sampling procedures for variable inspection based on non-normal distributions when the population variance is unknown. Geetha and Vijayaraghavan [3] considered Pareto distribution for the construction of variable sampling plans. Seifi and Nezhad [4] constructed variable sampling plan for resubmitted lots under Bayesian approach. Other major developments in variables sampling plans include the works of Srivastava [5], Owen [6], Duncan [7], Guenther [8, 9], Yeh [10, 11], Yeh *et.al.* [12], Aslam *et.al.* [13], Balamurali and Usha [14], Yen and Cheng [15], Wu *et.al.* [16-18], Geetha and Pavithra [19], and Rao *et.al.* [20].

The aspects of construction and assessment of variable sampling plans are quite easy when the quality feature under study stick to follow normality assumption. In industrial applications, the normality assumption of the quality variable may be sheered or the quality feature may exhibit non-normal patterns. In such situations, designing of sampling inspection plans by the method of variables becomes unwieldy.

In this paper, the designing procedure of sampling inspection plans by the method of variables based on a single sample is examined when the distribution of the quality feature of the manufactured commodity under study shows a similar pattern of Birnbaum – Saunders distribution.

II. Birnbaum – Saunders Distribution

One among the important probability distributions that have potential applications in life testing and reliability is the Birnbaum – Saunders distribution. It is a two - parameter continuous, unimodal and positively skewed probability distribution that is specifically used in the studies relating to modeling of fatigue life of metals, which are subject to periodic stress. [See, Balakrishnan and Kundu [21]]. It is also known as the fatigue-life distribution and has been widely applied to fatigue and reliability studies. Birnbaum and Saunders [22] derived an ingenious probability model to describe lifetimes associated with materials exposed to fatigue and tension. Birnbaum and Saunders [23, 24] formalized the fatigue-life distribution which is named after them. Since then, extensive work was carried out by many researchers on this distribution to provide generalizations, estimation and inferential procedures.

According to Leiva [25], its field of application has been extended beyond the context of fatigue and reliability analysis and is a model in situations where the accumulation of a certain factor forces a quantifiable characteristic to exceed a critical threshold. A detailed account of description, analysis and applications of the Birnbaum and Saunders distribution has been given in

the phenomenal paper by Balakrishnan and Kundu [21]. Another unparalled contribution on this distribution has been made by Leiva [25] in his book which provides an overview of the distribution, its probabilistic and statistical features, and its extension to regression analysis, diagnostics, etc.

The Birnbaum – Saunders distribution has a failure rate function which can take the three forms of failure rates, viz., increasing and decreasing failure rate, and unimodal or inverse bathtub. [See, Johnson, Kotz and Balakrishnan [26] and Nelson [27]]. When a probability distribution for life-time variable has a failure rate function that takes various shapes, it is the natural choice to adopt the distribution in practice. Considering the importance of Birnbaum and Saunders distribution, as pointed out by many research in the literature, its application in sampling inspection by variables is now considered.

Let X be a random variable representing the quality feature of a material and follows Birnbaum – Saunders distribution. The probability density function and cumulative distribution function of X are, respectively, given by

$$f(x; \theta, \delta) = \frac{\sqrt{x/\theta} + \sqrt{\theta/x}}{2\delta x} \varphi\left(\frac{\sqrt{x/\theta} - \sqrt{\theta/x}}{\delta}\right); x > 0, \theta > 0, \delta > 0 \tag{1}$$

and

$$F(x; \theta, \delta) = \Phi\left(\frac{\sqrt{x/\theta} - \sqrt{\theta/x}}{\delta}\right); x > 0, \theta > 0, \delta > 0 \tag{2}$$

where δ and θ are the shape and scale parameters, respectively, and φ and Φ are the density and distribution functions of standard normal distribution, respectively. The mean and variance of BSD are given by,

$$E(x) = \theta \left(1 + \frac{\delta^2}{2}\right), \tag{3}$$

and

$$V(x) = \theta^2 \delta^2 \left(1 + \frac{5}{4} \delta^2\right), \tag{4}$$

The reliability function and hazard function for specified time x under Birnbaum – Saunders distribution are, respectively, given by

$$R(x; \theta, \delta) = \Phi\left(\frac{\sqrt{\theta/x} - \sqrt{x/\theta}}{\delta}\right) \tag{5}$$

and

$$Z(x; \theta, \delta) = \frac{\frac{\sqrt{x/\theta} + \sqrt{\theta/x}}{2\delta x} \varphi\left(\frac{\sqrt{x/\theta} - \sqrt{\theta/x}}{\delta}\right)}{\Phi\left(\frac{\sqrt{\theta/x} - \sqrt{x/\theta}}{\delta}\right)}. \tag{6}$$

The coefficient of variation of BSD is,

$$\gamma(x) = \delta \left(\frac{\sqrt{5\delta^2 + 4}}{\delta^2 + 2} \right), \tag{7}$$

The coefficients of skewness and kurtosis of a random variable X which follows BSD are given by,

$$\alpha_3 = \frac{16\delta^2(11\delta^2 + 6)^2}{(5\delta^2 + 4)^3} \tag{8}$$

$$\alpha_4 = 3 + \frac{6\delta^2(93\delta^2 + 40)}{(5\delta^2 + 4)^2} \tag{9}$$

It can be identified that the coefficients of variation, skewness and kurtosis only depends on the shape parameter of BSD. It is also possible to note that, as the shape parameter, δ move towards zero, the BSD exhibit a symmetrical pattern around scale parameter, θ (the median of the BSD) and the variability decreases. Furthermore, notice that, as the shape parameter increases, the BSD has heavier tails. This means that δ modifies the skewness and kurtosis of the distribution.

III. Sampling Inspection by Variables

Acceptance sampling plans for variables based on a single sample can be stated based on the conditions mentioned hereafter.

- i. The random variable depicting the quality feature is quantifiable on a continuous scale that is having a recognized model of statistical distribution.
- ii. Each single item subjected to quality monitoring has provided with a single specification limit say, lower specification limit (LSL), L , or upper specification limit (USL), U . If a measurement on a particular item goes beyond the specification, the item is considered as an unsatisfactory component.

The working process of an acceptance sampling plan based on variable inspection for a single sample is described in the following manner:

Step 1: Randomly select a small group of elements of size n from the submitted population of manufactured items. Measure the quality variable of interest for each unit in the selected group and record the measurements.

Step 2: Approve the submitted population of manufactured commodities as accepted, if $\bar{x} + k\sigma \leq U$ or $\bar{x} - k\sigma \geq L$; and reject, otherwise. The conditions are chosen according to the given specification. If σ is unknown, s , an unbiased estimate of population standard deviation, is used in the place of σ .

The sample size n and acceptance constant k constitute the parameters of the acceptance sampling plans for variables based on a single sample.

IV. Operating Characteristic Function

The effectiveness of any acceptance sampling plan can be evaluated using an important measure, called operating characteristic (OC) function. It gives the probability of acceptance of a lot with a specified proportion of faulty items. It is denoted by $P_a(p)$, where p is the fraction of defective or nonconforming items in the bunch of commodities. When USL is provided, the proportion of nonconforming items and probability of accepting the lot are, respectively, given by

$$p = P(X > U | \mu) \tag{10}$$

and

$$P_a(p) = P(\bar{x} + k\sigma \leq U | \mu) \tag{11}$$

When σ is unknown, the sample standard deviation, s , which is an unbiased estimate of σ , is used.

A plot of probability of acceptance against the proportion of defective or nonconforming items is known as the operating characteristic (OC) curve, which is commonly used for comparing the efficiencies of sampling plans. A common procedure for designing a sampling plan is described by specifying two points on the OC curve, viz., $(p_0, 1-\alpha)$ and (p_1, β) , p_0 and p_1 denote the acceptable quality level (AQL) and limiting quality level (LQL), respectively; α and β are producer's risk and consumer's risk, respectively. The AQL and LQL are, respectively, defined by

$$AQL = p_0 = P(X > U | \mu_0) \tag{12}$$

and

$$LQL = p_1 = P(X > U | \mu_1), \tag{13}$$

where μ_0 and μ_1 are the means of probability distribution of the quality variable under study which results in AQL and LQL. The producer's risk and consumer's risk can be obtained using the given requirements of AQL and LQL as

$$\alpha = P(\bar{x} + k\sigma > U | \mu_0) \tag{14}$$

and

$$1 - \beta = P(\bar{x} + k\sigma > U | \mu_1) \tag{15}$$

When σ is unknown, the estimate s is used for the evaluation of α and β .

The probability of acceptance and the proportion of nonconforming items are defined based on the underlying distribution of the quality feature under study. They are obtained using the cumulative probability distribution function of BSD.

V. Designing Procedure of Variable Single Sampling Plan

In most of the practical scenarios, the quality feature of the product under study will show some kind of deviations from the normality assumption. In such situations, the standard variable sampling plans cannot be utilized. It is, also, the fact that population parameters of the statistical distribution, which is used to model the quality variable, are unidentifiable in most of the cases and hence, they are estimated using the sample statistics. If the distribution of quality variable under consideration is deviating from normality, the development phase of unknown σ sampling plans becomes more cumbersome.

Let $F(x; \theta, \delta)$ be the cumulative distribution function of BSD, which is also considered as the distribution function of the quality feature. From equation (12), it is easily observed that the acceptable quality level p_0 can be found using the cumulative distribution function of quality parameter, which can be expressed as

$$AQL = p_0 = 1 - F_x(x; \mu_0), \tag{16}$$

where μ_0 is the mean of BSD, which results in an acceptable quality level.

Similarly, from equation (13) the limiting quality level p_1 can be expressed as

$$LQL = p_1 = 1 - F_x(x; \mu_1), \tag{17}$$

where μ_1 is the mean of BSD, which results in limiting quality level.

Also, from equations (14) and (15), the producer's risk and consumer's risk can be obtained using the cumulative distribution function $F(x; \theta, \delta)$ as

$$\alpha = 1 - F_x(\bar{x}; \mu_0) \tag{18}$$

and

$$1 - \beta = 1 - F_x(\bar{x}; \mu_1) \tag{19}$$

When the underlying distribution of quality parameter is normal, the designing of variable sampling plan includes the process of determining the standard normal deviate value K_p^* . For designing a variable sampling plan for non-normal distributions, the value of K_p^* corresponds to the deviate value of underlying distribution. Hence, for obtaining the parameters of a variable sampling plan under BSD, the deviate values are obtained as follows.

$$K_{p_0}^* = \frac{x_{p_0} - M}{S}, \tag{20}$$

where x_{p_0} is the value of x for which the upper tail area of BSD is p_0 and $K_{p_0}^*$ is the standardized value of x_{p_0} . Here, M and S denotes, respectively, the mean and standard deviation of the BSD.

Similarly, the deviate value corresponding to the limiting quality level is obtained as

$$K_{p_1}^* = \frac{x_{p_1} - M}{S}, \tag{21}$$

where x_{p_1} is the value of x for which the upper tail area of BSD is p_1 and $K_{p_1}^*$ is the standardized value of x_{p_1} .

The optimum values of the parameters of a variable sampling plan for non-normal populations is given by Zimmer and Burr (1963) as

$$n_U = \left[\frac{K_\alpha + K_\beta}{K_{p_0}^* - K_{p_1}^*} \right]^2 \tag{22}$$

and

$$k_U = \frac{K_\alpha K_{p_1}^* + K_\beta K_{p_0}^*}{K_\alpha + K_\beta}. \tag{23}$$

Similarly, the constants of a variable sampling plan when the lower specification limit is specified, can be determined as

$$n_L = \left[\frac{K_\alpha + K_\beta}{K_{1-p_0}^* - K_{1-p_1}^*} \right]^2 \tag{24}$$

$$k_L = \frac{K_\alpha K_{1-p_1}^* + K_\beta K_{1-p_0}^*}{K_\alpha + K_\beta} \tag{25}$$

In 1972, Takagi introduced an approach for obtaining the sample size and the acceptance constant of single sampling inspection plans by variables based on a wide range of non-normal distributions and proposed an expansion component in the context of the measures of skewness and kurtosis. According to Takagi, when σ is unknown, $\bar{x} \pm ks$ will follow normal distribution asymptotically with parameters, $\mu_y = \mu \pm k\sigma$ and $\sigma_y^2 = \frac{\sigma^2}{n} \left[1 + \frac{k^2}{4}(\alpha_4 - 1) \pm k\alpha_3 \right]$, where μ, σ, α_3 and α_4 represent the mean, standard deviation, skewness and kurtosis of the underlying probability distribution, respectively. The operating characteristic function given in (11) can be rewritten, approximately, as

$$P_a(p) = P\left(Z \leq \frac{U - \mu_y}{\sigma_y} \mid \mu \right) \tag{26}$$

where μ_y and σ_y denote the mean and standard deviation of $\bar{x} \pm ks$ and Z is the standardized variable.

Let us denote $K_p^* = (U - \mu)/\sigma$. Then, for the designated points of acceptable and limiting quality levels, one may have

$$\mu_0 = U - K_{p_0}^* \sigma \tag{27}$$

and

$$\mu_1 = U - K_{p_1}^* \sigma. \tag{28}$$

Let α and β denote the producer's risk and consumer's risk, respectively. Then, the normal deviates corresponding to the specified risks are given by

$$K_\alpha = \frac{U - \mu_{x_0}}{\sigma_x} = \frac{U - (\mu_0 + k\sigma)}{\sqrt{(\sigma^2/n)[1 + (k^2/4)(\alpha_4 - 1) + k\alpha_3]}} = \frac{U - (\mu_0 + k_U\sigma)}{\sigma\sqrt{e_U/n_U}} \tag{29}$$

and

$$-K_\beta = \frac{U - \mu_{x_1}}{\sigma_x} = \frac{U - (\mu_1 + k\sigma)}{\sqrt{(\sigma^2/n)[1 + (k^2/4)(\alpha_4 - 1) + k\alpha_3]}} = \frac{U - (\mu_1 + k_U\sigma)}{\sigma\sqrt{e_U/n_U}}. \tag{30}$$

Using the expressions of μ_0, μ_1, K_α and K_β one can obtain the following:

$$K_{p_0}^* = k_U + K_\alpha \sqrt{\frac{e_U}{n_U}} \tag{31}$$

$$K_{p_1}^* = k_U - K_\beta \sqrt{\frac{e_U}{n_U}} \tag{32}$$

Solving the equations (31) and (32), the explicit expressions for the parameters of sampling inspection plan can be obtained as

$$n_U = e_U \left[\frac{K_\alpha + K_\beta}{K_{p_0}^* - K_{p_1}^*} \right]^2 \tag{33}$$

and

$$k_U = \frac{K_\alpha K_{p_1}^* + K_\beta K_{p_0}^*}{K_\alpha + K_\beta}, \tag{34}$$

where K_α and K_β denote the well-known standard normal deviates exceeding the probabilities α and β respectively, and $e_U = 1 + (k_U^2/4)(\alpha_4 - 1) + k_U \alpha_3$ is known as the expansion factor. The expansion factor is utilized in order to gain information about the parameters of known σ plans with the relations $n'_U = n_U / e_U$ and $k'_U = k_U$. In a similar manner, the constants of variable sampling plan, when the lower specification limit is specified, can be determined as

$$n_L = e_L \left[\frac{K_\alpha + K_\beta}{K_{1-p_0}^* - K_{1-p_1}^*} \right]^2 \tag{35}$$

and

$$k_L = \frac{K_\alpha K_{1-p_1}^* + K_\beta K_{1-p_0}^*}{K_\alpha + K_\beta}, \tag{36}$$

where $e_L = 1 + (k_L^2/4)(\alpha_4 - 1) + k_L \alpha_3$, making use of which the parameters of a known σ plan can be obtained as $n'_L = n_L / e_L$ and $k'_L = k_L$.

VI. Numerical Illustrations

I. Numerical Illustration 1

The thickness of silicon wafers is a vital quality feature in microelectronic circuits. The manufacturing firm specified the upper specification limit for the thickness as 0.02 mm. From the history, it was ascertained that the thickness of silicon wafer follows BSD having parameters $\delta = 0.25$ and $\theta = 0.0125$.

The mean and standard deviation were determined as $\mu = 0.01289$ and $\sigma = 0.0032$. Under the given conditions, for the specified values of $p_0 = 0.01, p_1 = 0.05, \alpha = 0.05$ and $\beta = 0.10$, it is required to obtain a single sampling plan by variables. Corresponding to the specified quality levels and the risks, one obtains the following: $K_{p_0}^* = 2.864411, K_{p_1}^* = 1.8225, K_\alpha = 1.644854$ and $K_\beta = 1.281552$. Thus, the optimum values of the parameters n and k of the sampling plan are determined from (21) and (22) as $n = 7.89$ or 8 and $k = 2.2788$.

In order to make the comparison of the results obtained in the illustration, data have been simulated based on 5000 runs using R programming. The simulated data of 8 observations from BSD having the specified parameters provide the sample average was found to be $\bar{x} = 0.01118$.

It can be observed that $\bar{x} + k\sigma = 0.0185$ which falls below the upper specification limit, $U = 0.02$. Hence, the lot would be accepted.

II. Numerical Illustration 2

A manufacturing company produces bottles for various purposes. For a particular make of bottle, a lower specification on the bursting strength of bottle is at 200 psi. The lot is accepted, if 1% or less of the bottles burst within this limit, with probability 0.95 (i.e., $p_0 = 0.01$, $\alpha = 0.05$) whereas if 6% or more of the bottles have bursting strength below this limit, the lot is rejected with probability 0.90 (i.e., $p_1 = 0.06$, $\beta = 0.10$).

Let X be the quality variable which follows BSD having the parameters, $\delta = 0.1$ and $\theta = 275$. The mean, variance, skewness and kurtosis are obtained as $\mu = 276.375$, $\sigma^2 = 765.7031$, $\alpha_3 = 01.1994$ and $\alpha_4 = 3.1497$, respectively. For the specified quality levels, say $p_0 = 0.01$ and $p_1 = 0.06$, and the associated producer's and consumer's risks, $\alpha = 0.05$ and $\beta = 0.10$, one can obtain the parameters of the sampling plan by variables using the expressions (21) and (22). Corresponding to $p_0 = 0.01$ and $p_1 = 0.06$, the deviate values $K_{p_0}^*$ and $K_{p_1}^*$ are obtained, from (19) and (20), as 2.1083 and 1.4793, respectively. Corresponding to the specified $\alpha = 0.05$ and $\beta = 0.10$, the values of K_α and K_β are obtained as 1.644854 and 1.281552. Given the lower specification limit and other requirements, the parameters of single sampling plan by variables are obtained from equations (21) and (22) as $n = 21.65$, which when rounded becomes 22 and $k = 1.7548$.

A simulation study is carried out for comparing the results arrived in the above illustration. Assume that the standard deviation is known. The simulated results are based on 5000 runs using R programming. By simulation based on 5000 runs, the randomly generated sample of 22 observations from BSD having the specified parameters $\delta = 0.1$ and $\theta = 275$ yields the sample mean, viz., $\bar{x} = 276.3843$.

It is known that the acceptance criterion under a variable sampling plan with specified lower specification limit is given as $\bar{x} - k\sigma \geq L$. It can be seen that $\bar{x} - k\sigma = 227.8266$, which is greater than the lower specification limit, i.e., $U = 200$. Hence, the lot would be accepted.

VII. Conclusion

Normal distribution and its wide range of properties play a prominent role in the theory and applications of statistics. Acceptance sampling plans for variable inspection consider normal distribution as an important model for the quality variable. Even though applications of normal distribution are a plenty in various domains, there are situations, where non-normality arises in the real-life data. In this paper, a Birnbaum – Saunders distribution (BSD) is considered for designing an acceptance sampling plan by variables based on the information acquired from a single sample. The strategy for the choice of variable single sampling inspection plans when the quality feature shows the behavior of BSD is discussed, when the producer's and consumer's requirements are specified. The numerical illustrations are given to demonstrate how the proposed plan could be executed in practical application. The simulation based on 5000 runs for generating samples from BSD has been done utilizing R programming and the simulated results are compared with the results arrived in the illustrations.

References

- [1] Zimmer WJ, Burr IW. (1963). Variables sampling plans based on non-normal populations. *Industrial Quality Control*. 20(1): 18-26.
- [2] Takagi K. (1972). On designing unknown-sigma sampling plans based on a wide class of non-normal distributions. *Technometrics*. 14(3): 669-678.
- [3] Geetha S, Vijayaraghavan R. (2013). A procedure for the selection of single sampling plans by variables based on Pareto distribution. *Journal of Quality and Reliability Engineering*. 2013. 5 pages.
- [4] Seifi S, Nezhad MS. (2017). Variable sampling plan for resubmitted lots based on process capability index and Bayesian approach. *International Journal of Advanced Manufacturing Technology*. 88(9-12): 2547-2555.
- [5] Srivastava AB. (1961). Variables sampling inspection for non-normal samples. *Journal of Science and Engineering Research*. 5: 145-152.
- [6] Owen DB. (1969). Summary of recent work on variables acceptance sampling with emphasis on non-normality. *Technometrics*. 11(4): 631-637.
- [7] Duncan AJ. *Quality Control and Industrial Statistics*, 5th. Ed. Irwin, Homewood, IL: Richard D Irvin, 1986.
- [8] Guenther WC. (1972). Variables sampling plans for the Poisson and the binomial. *Statistica Neerlandica*. 26(4): 113-120.
- [9] Guenther WC. (1985). LQL like plans for sampling by variables. *Journal of Quality Technology*. 17(3):155-157.
- [10] Yeh L. (1990). An optimal single variable sampling plan with censoring. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 39(1): 53-66.
- [11] Yeh L. (1994). Bayesian variable sampling plans for the exponential distribution with type I censoring. *Annals of Statistics*. 22(2): 696-711.
- [12] Yeh L, Hung L K, Cheng W, Wong H. (2006). Sequential variable sampling plan for normal distribution. *European Journal of Operation Research*. 172(1): 127-145.
- [13] Aslam M, Wu CW, Azam M, Jun CH. (2013). Variable sampling inspection for resubmitted lots based on process capability index C_{pk} for normally distributed items. *Applied Mathematical Modelling*. 37(3):667-675.
- [14] Balamurali S, Usha M. (2013). Optimal designing of variables chain sampling plan by minimizing the average sample number. *International Journal of Manufacturing Engineering*. 2013. 12 pages.
- [15] Yen CH, Chang CH. (2009). Designing variables sampling plans with process loss consideration. *Communications in Statistics - Simulation and Computation*. 38(8):1579-1591.
- [16] Wu CW, Pearn WL. (2008). A variables sampling plan based on C_{pmk} for product acceptance determination. *European Journal of Operation Research*. 184(2): 549-560.
- [17] Wu CW, Aslam M, Jun CH. (2012). Variables sampling inspection scheme for resubmitted lots based on the process capability index C_{pk} . *European Journal of Operation Research*. 217(3): 560-566.
- [18] Wu CW, Shu MH, Chang YN. (2018). Variable-sampling plans based on lifetime-performance index under exponential distribution with censoring and its extensions. *Applied Mathematical Modelling*. 55: 81-93.
- [19] Geetha S, Pavithra S. (2019). Two Point Acceptance Sampling Plan By Variable Using Range. *International Journal of Scientific and Technology Research*. 8: 3903-3905.
- [20] Rao GS, Aslam M, Jun CH. (2021). A variable sampling plan using generalized multiple dependent state based on a one-sided process capability index. *Communications in Statistics - Simulation and Computation*. 50(9): 2666-2677.
- [21] Balakrishnan, N., and Kundu, D. (2019). Birnbaum - Saunders Distribution: A review of Models, Analysis and Applications. *Applied Stochastic Models in Business and Industry*. 35(1): 4 – 49.

- [22] Birnbaum, Z. W. and Saunders, S. C. (1958). A Statistical Model for Life Lengths of Materials. *Journal of the American Statistical Association*. 53(281): 153 - 160.
- [23] Birnbaum, Z. W., and Saunders, S. C. (1969a). A New Family of Life Distributions, *Journal of Applied Probability*. 6(2): 319 - 327.
- [24] Birnbaum, Z. W., and Saunders, S. C. (1969b). Estimation for a Family of Life Distributions with Applications to Fatigue. *Journal of Applied Probability*. 6 (2): 328–347.
- [25] Leiva, V. *The Birnbaum - Saunders Distribution*, First Edition, Academic Press, Elsevier, London, UK, 2015.
- [26] Johnson, N. L., Kotz, S., and Balakrishnan, N. *Continuous Univariate Distribution - Vol. 2*, Second Edition, John Wiley & Sons, New York, US, 1995.
- [27] Nelson, W. B. *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, John Wiley & Sons, New York, US, 1990.

Optimal spare-switching times in series systems under a general framework

HAMIDEH JEDDI

Department of Statistics, Ferdowsi University of Mashhad, Iran
ha_je57@mail.um.ac.ir

MAHDI DOOSTPARAST

Department of Statistics, Ferdowsi University of Mashhad, Iran
doustparast@um.ac.ir

Abstract

Spares are commonly used to improve system performances. They are allocated to original components during system missions. The optimal allocations depend on system configurations and lifetimes of components and spares. Various methods for finding optimal allocations have been proposed in the literature. For sake of brevity, lifetimes of components are commonly assumed to be independent. This paper deals with series systems, a common configuration, under a general setting, i.e. component lifetimes are dependent and heterogeneous. Moreover, the spare is also allowed to switch among original components to impose more flexibility for spare managements. This allowance occurred usually in network servers and electrical generators which manage by a dispatching center. Explicit expressions for system reliability functions are derived in detail. Since system lifetimes are random phenomena, stochastic orders are utilized for comparison purposes. Various illustrative examples are also given.

Keywords: Allocation; Reliability; Stochastic orders; Redundancy; Switching.

1. INTRODUCTION

Implementing spare (redundant) is an effective and common method to attain high reliability systems. The redundancy allocation problem (RAP) has been extensively used in real-world applications such as circuit design, power plant, electrical power systems, transportation, safety, telecommunication, satellite, consumer-electronics industry, etc. Specifically redundants are also used in computer science and especially in network systems (servers), to guard the primary system against random failure as a backup system. In this case, redundant components can include both hardware elements of a system such as disk drives, peripherals, servers, switches, routers and software elements such as operating systems, applications, and databases. In the world of information technology (IT), redundancy is the “duplication of critical components or functions of a system with the intention of increasing reliability of the system, usually in the form of a backup or fail-safe, or to improve actual system performance, such as in the case of GNSS receivers (GNSS antennas), or multi-threaded computer processing” (See, e.g. [7]). Issues such as a hardware failure, network problems, or application faults could cause the primary servers to stop functioning correctly. This can leave users unable to access services, which poses a real barrier to productivity. Server redundancy helps businesses by protecting critical data by ensuring it exists in more than just one place. This means that the business can recover data if something happens to a live server. For applications where data integrity and access are vital, redundant servers are very important.

Additional components (spares) are used to improve engineering system performances. For more details, see Barlow and Proschan [2], Nakagawa [15]. There are many papers deal with redundancy allocation problem in reliability systems. Boland et al. [5] applied stochastic orders to consider this problem for series and parallel systems. Zhao et al. [24, 23] studied optimal allocation of redundancies with exponential components in the sense of various stochastic orders. Xie et al. [20] investigated the redundancy allocation problem in k -out-of- n hot standby systems to maximize the operational availability. But in the case of dependent components, there are not many works. Among a few works, Navarro et al. [17] studied the performance of a system composed by various kinds of units have dependent lifetimes to evaluate reliability. Navarro and Durante [16] studied the behaviour of the residual lifetimes of coherent systems with possibly dependent components. Belzunce et al. [3, 4] used the concept of "joint stochastic orders" and Jeddi and Doostparast [9, 10] studied this problems for series and parallel systems. Redundants are allocated to original components during system missions. Commonly, spares do not switch among the original components. Adding redundants to a system may be done with respect to some limitations such as cost, weight and volume. Switching a spare between original (primary) components is a possible way to overtake these limitations. For example, communication networks, managers may be able to control and switch spares (servers) among original components (or servers) to achieve more reliable connections among customers. This management is usually done by a dispatching center. In other words, redundants can change dynamically their respective original components. For more examples and recent developments, see Kim et al. [12], Li et al. [14], Jia et al. [11] and references therein. Notice that, the spare can not switch if the corresponding original component fails.

In the sequel, let $(\Omega, \mathbb{F}, \mathbb{P})$ be a probability space and $\mathbf{X} = (X_1, \dots, X_k) : \Omega \rightarrow \mathbb{R}_+^k$, for $k \geq 1$, be an absolutely continuous random vector with the joint distribution function and joint survival function

$$F_{X_1, \dots, X_k}(a_1, \dots, a_k) = P(X_1 \leq a_1, \dots, X_k \leq a_k), \quad \forall (a_1, \dots, a_k) \in \mathbb{R}_+^k,$$

and

$$\bar{F}_{X_1, \dots, X_k}(a_1, \dots, a_k) = P(X_1 > a_1, \dots, X_k > a_k), \quad \forall (a_1, \dots, a_k) \in \mathbb{R}_+^k,$$

respectively. Here, \mathbb{R}_+^k stands for the k -dimensional Euclidean space. Then, the density function of \mathbf{X} is given by $f_{X_1, \dots, X_k}(a_1, \dots, a_k) = \partial F_{X_1, \dots, X_k}(a_1, \dots, a_k) / (\partial a_1 \cdots \partial a_k)$. The marginal distribution of X_i ($1 \leq i \leq k$) is denoted by $F_{X_i}(x) = P(X_i \leq x), \forall x \in \mathbb{R}^+$. The random variables X_i is said to be smaller than X_j ($j \neq i$) in usual stochastic order denoted by $X_i \leq_{st} X_j$, if $F_{X_i}(x) \geq F_{X_j}(x), \forall x \in \mathbb{R}$. Equivalently, $\bar{F}_{X_i}(x) \leq \bar{F}_{X_j}(x)$ where $\bar{F}_{X_i}(x) = 1 - F_{X_i}(x), \forall x$ and for $1 \leq i \leq k$; See Shaked and Shanthikumar [19].

This paper is organized as follows. In Section 2, two possible spare allocations are described. Then, the improved system lifetimes by the two schemes for allocating the spare are also derived. In Section 3, a general form for the system reliability function is presented. Section 4 deals with comparing the two schemes. Indeed, The main result holds for systems with heterogeneous and dependent component lifetimes. Also, it provides the interval time for switching the spare among original components when the components and spare are independent. In Section 5, two general classes of lifetimes and some well known lifetimes are analysed in detail. Section 6 concludes the paper and provides further topics for future research.

2. SYSTEM LIFETIME WITH THE SPARE

Consider a 2-component series system consisting of a spare which can be added to the system configuration. The spare can switch only one-time. Let $\tau > 0$ be a preassigned deterministic constant and $T_i^{[0, \tau]}$ ($i = 1, 2$) denote the system lifetime when the spare is allocated (in parallel) to Component i during interval $[0, \tau]$ and then to Component j ($j \neq i$) beyond τ (> 0); See Figure 1. In Figure 1, the spare is allocated to Component 1 during the time interval $[0, \tau]$. Then, the spare

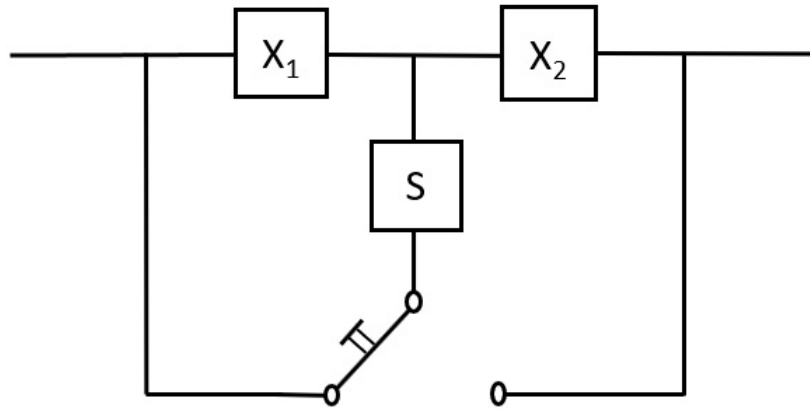


Figure 1: Allocation a spare under a switching scheme.

is allocated to Component 2. We call this allocation strategy as Scheme 1. Similarly, Scheme 2 is defined. Figure 2 pictures two possible schemes for allocating the spare.

Now, we obtain the system lifetime when the spare is added to the original components. To do this, let X_1 and X_2 denote the original component lifetimes, and S stands for the spare lifetime. Therefore, it can be seen that the improved system lifetime under Scheme I is

$$\begin{aligned}
 T_1^{[0,\tau]} &= \begin{cases} \wedge(\vee(X_1, S), X_2), & \text{if } X_1 \leq \tau, \\ X_2, & \text{if } X_1 > \tau, X_2 \leq \tau, \\ \tau + \wedge(X_1 - \tau, X_2 - \tau), & \text{if } X_1 > \tau, X_2 > \tau, S \leq \tau, \\ \tau + \wedge(X_1 - \tau, \vee(X_2 - \tau, S - \tau)), & \text{if } X_1 > \tau, X_2 > \tau, S > \tau, \end{cases} \\
 &= \begin{cases} \wedge(\vee(X_1, S), X_2), & \text{if } X_1 \leq \tau, \\ X_2, & \text{if } X_1 > \tau, X_2 \leq \tau, \\ \tau + \wedge(X_1 - \tau, \vee(X_2 - \tau, S - \tau)), & \text{if } X_1 > \tau, X_2 > \tau, \end{cases} \\
 &= \begin{cases} \wedge(\vee(X_1, S), X_2), & \text{if } X_1 \leq \tau, \\ X_2, & \text{if } X_1 > \tau, X_2 \leq \tau, \\ \wedge(X_1, \vee(X_2, S)), & \text{if } X_1 > \tau, X_2 > \tau, \end{cases} \quad (1)
 \end{aligned}$$

where $\vee(a_1, a_2) = \max\{a_1, a_2\}$ and $\wedge(a_1, a_2) = \min\{a_1, a_2\}$. Similarly, the improved system lifetime under Scheme II is

$$T_2^{[0,\tau]} = \begin{cases} \wedge(X_1, \vee(X_2, S)), & \text{if } X_2 \leq \tau, \\ X_1, & \text{if } X_2 > \tau, X_1 \leq \tau, \\ \wedge(\vee(X_1, S), X_2), & \text{if } X_2 > \tau, X_1 > \tau. \end{cases} \quad (2)$$

Equations (1) and (2) can be unified as

$$\begin{aligned}
 T_1^{[0,\tau]} &= \wedge(\vee(X_1, S), X_2)I(X_1 \leq \tau) + X_2I(X_1 > \tau, X_2 \leq \tau) \\
 &+ \wedge(X_1, \vee(X_2, S))I(X_1 > \tau, X_2 > \tau), \quad (3)
 \end{aligned}$$

and

$$\begin{aligned}
 T_2^{[0,\tau]} &= \wedge(X_1, \vee(X_2, S))I(X_2 \leq \tau) + X_1I(X_2 > \tau, X_1 \leq \tau) \\
 &+ \wedge(\vee(X_1, S), X_2)I(X_2 > \tau, X_1 > \tau), \quad (4)
 \end{aligned}$$

where $I_A(t)$ denotes the indicator function of the set A , i.e., $I_A(t) = 1$ for $t \in A$, and $I_A(t) = 0$ otherwise.

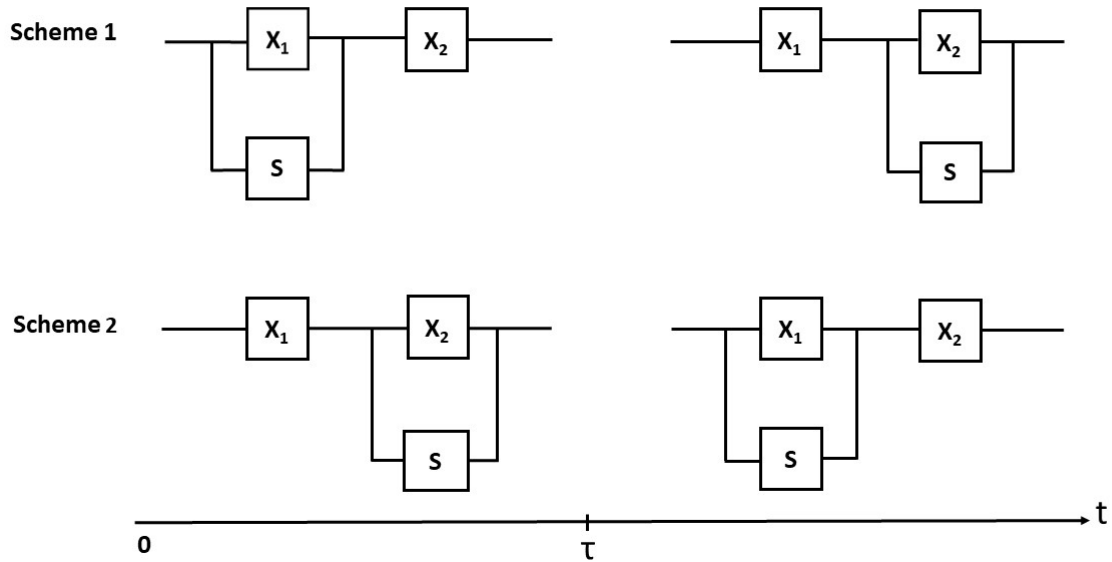


Figure 2: Scheme 1 with system lifetime $T_1^{[0,\tau]}$ and Scheme 2 with system lifetime $T_2^{[0,\tau]}$.

3. SYSTEM RELIABILITY WITH THE SPARE

In this section, the reliability function of the system is derived when the spare allows to switch among the system components one and only one time. To end this, from Equation (3), we have for $0 < t < \tau$,

$$\begin{aligned}
 P(T_1^{[0,\tau]} > t) &= P(T_1^{[0,\tau]} > t, X_1 \leq \tau) + P(T_1^{[0,\tau]} > t, X_1 > \tau, X_2 \leq \tau) \\
 &+ P(T_1^{[0,\tau]} > t, X_1 > \tau, X_2 > \tau) \\
 &= P(\wedge(\vee(X_1, S), X_2) > t, X_1 \leq \tau) + P(X_1 > \tau, X_2 \leq \tau, X_2 > t) \\
 &+ P(\wedge(X_1, \vee(X_2, S)) > t, X_1 > \tau, X_2 > \tau) \\
 &= P(\vee(X_1, S) > t, X_1 \leq \tau, X_2 > t) + P(X_1 > \tau, t < X_2 \leq \tau) \\
 &+ P(X_1 > t, \vee(X_2, S) > t, X_1 > \tau, X_2 > \tau) \\
 &= P(X_1 \leq \tau, X_2 > t) - P(X_1 \leq \tau, X_2 > t, X_1 \leq t, S \leq t) \\
 &+ P(X_1 > \tau, X_2 > t) - P(X_1 > \tau, X_2 > \tau) \\
 &+ P(X_1 > \tau, X_2 > \tau) - P(X_1 > \tau, X_2 > \tau, X_2 \leq t, S \leq t) \\
 &= P(X_1 \leq \tau, X_2 > t) - P(X_1 \leq t, X_2 > t, S \leq t) \\
 &+ P(X_1 > \tau, X_2 > t) \\
 &= P(X_2 > t) - P(X_1 \leq t, S \leq t) + P(X_1 \leq t, X_2 \leq t, S \leq t) \\
 &= 1 - F_{X_2}(t) - F_{X_1, S}(t, t) + F_{X_1, X_2, S}(t, t, t).
 \end{aligned} \tag{5}$$

Similarly, for $t \geq \tau$, we have

$$\begin{aligned}
 P(T_1^{[0,\tau]} > t) &= P(T_1^{[0,\tau]} > t, X_1 \leq \tau) + P(T_1^{[0,\tau]} > t, X_1 > \tau, X_2 \leq \tau) \\
 &+ P(T_1^{[0,\tau]} > t, X_1 > \tau, X_2 > \tau) \\
 &= P(\wedge(\vee(X_1, S), X_2) > t, X_1 \leq \tau) + P(X_1 > \tau, X_2 \leq \tau, X_2 > t) \\
 &+ P(\wedge(X_1, \vee(X_2, S)) > t, X_1 > \tau, X_2 > \tau) \\
 &= P(\vee(X_1, S) > t, X_1 \leq \tau, X_2 > t) \\
 &+ P(X_1 > t, \vee(X_2, S) > t, X_1 > \tau, X_2 > \tau) \\
 &= P(X_1 \leq \tau, X_2 > t) - P(X_1 \leq \tau, X_2 > t, X_1 \leq t, S \leq t) \\
 &+ P(X_1 > t, X_2 > \tau) - P(X_1 > t, X_2 > \tau, X_2 \leq t, S \leq t) \\
 &= P(X_1 \leq \tau, X_2 > t) - P(X_1 \leq \tau, X_2 > t, S \leq t) \\
 &+ P(X_1 > t, X_2 > \tau) - P(X_1 > t, \tau < X_2 \leq t, S \leq t) \\
 &= P(X_1 \leq \tau, X_2 > t, S > t) + P(X_1 > t, X_2 > \tau) \\
 &\quad - P(X_1 > t, X_2 > \tau, S \leq t) + P(X_1 > t, X_2 > t, S \leq t) \\
 &= P(X_2 > t, S > t) - P(X_1 > \tau, X_2 > t, S > t) \\
 &\quad + P(X_1 > t, X_2 > \tau, S > t) + P(X_1 > t, X_2 > t) - P(X_1 > t, X_2 > t, S > t) \\
 &= \bar{F}_{X_2, S}(t, t) - \bar{F}_{X_1, X_2, S}(\tau, t, t) + \bar{F}_{X_1, X_2, S}(t, \tau, t) + \bar{F}_{X_1, X_2}(t, t) \\
 &\quad - \bar{F}_{X_1, X_2, S}(t, t, t). \tag{6}
 \end{aligned}$$

Finally, from Equations (5) and (6), the next proposition is obtained.

Proposition 1. The system reliability functions of $T_1^{[0,\tau]}$ and $T_2^{[0,\tau]}$ are $\bar{F}_{T_1^{[0,\tau]}}(t) = y_1(t)I_{[0,\tau)}(t) + y_2(t)I_{[\tau,\infty)}(t)$, and $\bar{F}_{T_2^{[0,\tau]}}(t) = z_1(t)I_{[0,\tau)}(t) + z_2(t)I_{[\tau,\infty)}(t)$, respectively, where $y_i(t)$ and $z_i(t)$ ($i = 1, 2$) are defined by

$$y_1(t) = \bar{F}_{X_2}(t) - F_{X_1, S}(t, t) + F_{X_1, X_2, S}(t, t, t), \tag{7}$$

$$y_2(t) = \bar{F}_{X_2, S}(t, t) - \bar{F}_{X_1, X_2, S}(\tau, t, t) + \bar{F}_{X_1, X_2, S}(t, \tau, t) + \bar{F}_{X_1, X_2}(t, t) - \bar{F}_{X_1, X_2, S}(t, t, t), \tag{8}$$

$$z_1(t) = \bar{F}_{X_1}(t) - F_{X_2, S}(t, t) + F_{X_1, X_2, S}(t, t, t), \tag{9}$$

$$z_2(t) = \bar{F}_{X_1, S}(t, t) - \bar{F}_{X_1, X_2, S}(t, \tau, t) + \bar{F}_{X_1, X_2, S}(\tau, t, t) + \bar{F}_{X_1, X_2}(t, t) - \bar{F}_{X_1, X_2, S}(t, t, t),$$

for all $t > 0$.

Notice that $\lim_{t \rightarrow \tau^-} y_1(t) = \lim_{t \rightarrow \tau^+} y_2(t)$ and $\lim_{t \rightarrow \tau^-} z_1(t) = \lim_{t \rightarrow \tau^+} z_2(t)$. Therefore, the next corollary follows.

Corollary 1. The reliability functions of the lifetimes $T_1^{[0,\tau]}$ and $T_2^{[0,\tau]}$ are continuous in $t \in (0, +\infty)$.

4. COMPARISON AND OPTIMAL TIME TO SWITCH

System lifetimes are random variables and then partially orders should be considered for comparison purposes. Among various partially orders, stochastic orders are commonly used in reliability analyses. See, e.g. Boland et al. [5], Navarro et al. [17] and Belzunce et al. [3]. In this section, the main result of this paper is presented under a general setting for component and spare lifetimes. In the rest of this paper and for lifetimes U_1, U_2 and U_3 , let $\bar{F}_{U_1|(U_2, U_3)}(u_1|u_2, u_3) := P(U_1 > u_1 | U_2 > u_2, U_3 > u_3)$.

Proposition 2. Suppose that X_1, X_2 and S be dependent random variables and $[X_1 | S = s] \leq_{st} [X_2 | S = s]$ for all $s \geq 0$. If $\bar{F}_{X_1|(X_2, S)}(\tau|t, t) \leq 1/2$ and $\bar{F}_{X_2|(X_1, S)}(\tau|t, t) \geq 1/2$ for $t > \tau$, then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$.

Proof. [i] For $0 < t \leq \tau$, Equations (7) and (9) conclude

$$\begin{aligned} \bar{F}_{T_1^{[0,\tau]}}(t) - \bar{F}_{T_2^{[0,\tau]}}(t) &= \bar{F}_{X_2,S}(t,t) - \bar{F}_{X_1,S}(t,t) \\ &= \int_t^{+\infty} \left(P(X_2 > t|S = s) - P(X_1 > t|S = s) \right) dF_S(s) \geq 0, \end{aligned} \quad (10)$$

since $[X_1|S = s] \leq_{st} [X_2|S = s]$ for all $s > 0$. For $t > \tau$, Equations (8) and (10) imply

$$\begin{aligned} \bar{F}_{T_1^{[0,\tau]}}(t) - \bar{F}_{T_2^{[0,\tau]}}(t) &= \bar{F}_{X_2,S}(t,t) - \bar{F}_{X_1,S}(t,t) + 2\bar{F}_{X_1,X_2,S}(t,\tau,t) - 2\bar{F}_{X_1,X_2,S}(\tau,t,t) \\ &= \bar{F}_{X_2,S}(t,t)(1 - 2\bar{F}_{X_1|X_2,S}(\tau|t,t)) + \bar{F}_{X_1,S}(t,t)(2\bar{F}_{X_2|X_1,S}(\tau|t,t) - 1) \\ &\geq 0. \end{aligned} \quad (11)$$

Since $\bar{F}_{X_1|(X_2,S)}(\tau|t,t) \leq 1/2$ and $\bar{F}_{X_2|(X_1,S)}(\tau|t,t) \geq 1/2$ for $t > \tau$, the desired result follows. ■ There are situations in which components and spares are remote and hence they are approximately statistically independent. For example, suppose that there are two main servers in a network and the system administrator (in a dispatching center) wishes to improve system reliability by adding an extra server. Therefore, the three servers would be independent. In sequel, some conditions are assumed which simplify the main result is given in Proposition 2. First, assume that the spare is independent of original component lifetimes X_1 and X_2 while the original components may be dependent.

Corollary 2. Let S be independent of (X_1, X_2) . If $X_1 \leq_{st} X_2$ and $\bar{F}_{X_1|X_2}(\tau|t) \leq 1/2$ and $\bar{F}_{X_2|X_1}(\tau|t) \geq 1/2$ for $t > \tau$, then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$.

The next Proposition states that if the original components and the spare are independent, and the switch time lies between medians of the original component DFs, then Scheme 1 dominates Scheme 2 in st-order.

Proposition 3. Let X_1, X_2 and S be independent. If $X_1 \leq_{st} X_2$ and $m_1 < \tau < m_2$, where m_1 and m_2 stand for medians of X_1 and X_2 , respectively. Then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$.

Proof. For $0 < t \leq \tau$, Equations (7) and (9) conclude

$$\begin{aligned} \bar{F}_{T_1^{[0,\tau]}}(t) - \bar{F}_{T_2^{[0,\tau]}}(t) &= g_1(t) - z_1(t) \\ &= F_{X_1}(t) - F_{X_2}(t) + F_{X_2}(t)F_S(t) - F_{X_2}(t)F_S(t) \\ &= \bar{F}_S(t)(F_{X_1}(t) - F_{X_2}(t)) \geq 0, \end{aligned} \quad (12)$$

since $X_1 \leq_{st} X_2$. For $t > \tau$, Equations (8) and (10) imply

$$\begin{aligned} \bar{F}_{T_1^{[0,\tau]}}(t) - \bar{F}_{T_2^{[0,\tau]}}(t) &= g_2(t) - z_2(t) \\ &= \bar{F}_{X_2}(t)\bar{F}_S(t) - \bar{F}_{X_1}(t)\bar{F}_S(t) \\ &\quad + 2(\bar{F}_{X_1}(t)\bar{F}_{X_2}(\tau)\bar{F}_S(t) - \bar{F}_{X_1}(\tau)\bar{F}_{X_2}(t)\bar{F}_S(t)) \\ &= \bar{F}_{X_2}(t)\bar{F}_S(t)(1 - 2\bar{F}_{X_1}(\tau)) + \bar{F}_{X_1}(t)\bar{F}_S(t)(2\bar{F}_{X_2}(\tau) - 1) \\ &\geq 0, \end{aligned} \quad (13)$$

since $\tau > m_1$ and $\tau < m_2$. and the desired result follows. ■ Proposition 3 says that if component and spare lifetimes are independent, the spare should allocate to the weaker component at least up to its median lifetime and then before reaching to the median lifetime of the other component, the spare must switch.

Remark 1. The distribution of S in Proposition 3 is free and the given conditions do not rely on the DF of S .

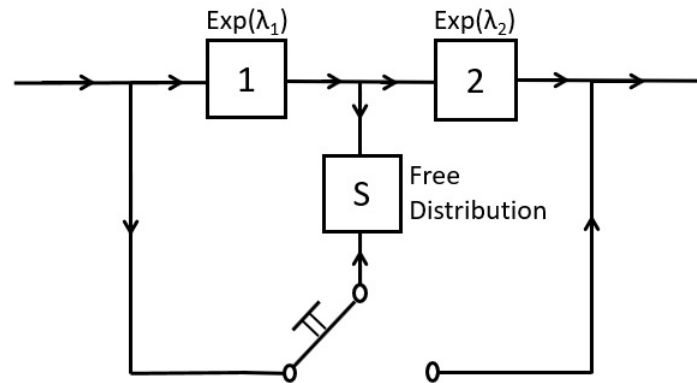


Figure 3: Allocation a spare under a switching scheme with independent exponential random variables.

5. EXAMPLES

In this section, some examples are analyzed to derived the optimal switching times for the spare.

Example 4. Let X_1 and X_2 , be independent exponential random variables with means $1/\lambda_1$ and $1/\lambda_2$, respectively. If $\lambda_1 > \lambda_2$, then Proposition 3 implies that $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$ provided that the switching occurs after $\frac{\ln 2}{\lambda_1}$ but before $\frac{\ln 2}{\lambda_2}$, that is $\frac{\ln 2}{\lambda_1} < \tau < \frac{\ln 2}{\lambda_2}$; See Figure 3. \square

Example 5. Let X_1, X_2 and S be independent and $X_i \sim Pa(\alpha_i, 1), i = 1, 2$ where $\alpha_i > 0$ and $Pa(a, b)$ stands for the Pareto distribution Type I with density $f(x) = \frac{a}{x^{a+1}}, x \geq 1$. It is easy to see that the medians of X_1 and X_2 , respectively, are given by $m_1 = \sqrt[a_1]{2}$ and $m_2 = \sqrt[a_2]{2}$. If $\alpha_1 > \alpha_2$ and $\sqrt[a_1]{2} < \tau < \sqrt[a_2]{2}$ then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$ from Proposition 3. \square

Example 6. The class of *newsboy distributions*, introduced by Braden and Freimer[6], includes absolutely continues distribution functions of the form $F_\theta(x) = 1 - e^{-\theta l(x)}$, where $l(x)$ is non-negative, increasing, differentiable and unbounded function with $l(0) = 0$. The newsboy distributions are used extensively in the modelling of excess demand of the inventory level that is lost and thus unobserved. Now assume that X_1, X_2 and S be independent and $X_i \sim F_{\theta_i}(x) = 1 - e^{-\theta_i l(x)}, i = 1, 2$ where $\theta_i > 0$. If $\theta_1 > \theta_2$ and $l^{-1}\left(\frac{\ln 2}{\theta_1}\right) < \tau < l^{-1}\left(\frac{\ln 2}{\theta_2}\right)$ then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$ by Proposition 3; See Table 1. Here, $l^{-1}(\cdot)$ denotes the inverse of the function $l(\cdot)$.

Example 7. Let Φ be the class of absolutely continues distribution function F_θ of the form $F_\theta(x) = 1 - e^{-K_\theta(x)}, x > 0$, where $K_\theta(x)$ is increasing in x and positive function $\theta \in \Theta$. Then the probability of density function is given by $f_\theta(x) = k_\theta(x)e^{-K_\theta(x)}, x > 0$, where $k_\theta(x) = \frac{\partial}{\partial x} K_\theta(x)$. This class include several important distribution such as exponential, Pareto, Weibull and has been studied in literature; See e.g, Al-Hussaini [1] for more details. Let $X_1 \sim F_{\theta_1}(x)$ and $X_2 \sim F_{\theta_2}(x)$. Then medians X_1 and X_2 are $m_1 = K_{\theta_1}^{-1}(\ln 2)$ and $m_2 = K_{\theta_2}^{-1}(\ln 2)$ respectively. If $K_{\theta_1}(x) \geq K_{\theta_2}(x)$ and $K_{\theta_1}^{-1}(\ln 2) \leq \tau \leq K_{\theta_2}^{-1}(\ln 2)$, then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$ by Proposition 3. For example, let $K_\theta(x) = \lambda x^\alpha$ and $\Theta = (\alpha, \lambda), \alpha, \lambda > 0$. Thus $X_i, i = 1, 2$, has the Weibull distribution with density function $f_{\alpha_i, \lambda}(x) = \alpha_i \lambda x^{\alpha_i - 1} e^{-\lambda x^{\alpha_i}}, \alpha_i, \lambda > 0$, therefore $m_i = K_{\theta_i}^{-1}(\ln 2) = \sqrt[\alpha_i]{\frac{\ln 2}{\lambda}}$ where $\theta_i = (\alpha_i, \lambda), i = 1, 2$. If $\alpha_1 > \alpha_2$ and $\sqrt[\alpha_1]{\frac{\ln 2}{\lambda}} < \tau < \sqrt[\alpha_2]{\frac{\ln 2}{\lambda}}$ then $T_1^{[0,\tau]} \geq_{st} T_2^{[0,\tau]}$ by Proposition 3. Table 1 presents some selected members in the class Φ and corresponding optimal switching times. \square

Table 1: Some well known members of Class Φ in Example 7

Distribution	Vector parameter $\vec{\theta}_i$	$K_{\theta_i}(x)(i = 1, 2)$	optimal switching time $K_{\theta_1}^{-1}(\ln 2) \leq \tau \leq K_{\theta_2}^{-1}(\ln 2)$
Exponential	λ_i	$\lambda_i x$	$\frac{\ln 2}{\lambda_1} < \tau < \frac{\ln 2}{\lambda_2}$
Pareto	(α_i, β)	$\alpha_i \ln\left(\frac{x}{\beta}\right)$	$\beta^{\alpha_1} \sqrt[2]{2} < \tau < \beta^{\alpha_2} \sqrt[2]{2}$
Weibull	(α_i, λ)	λx^{α_i}	$\sqrt[2]{\left(\frac{\ln 2}{\lambda}\right)^{\alpha_1}} < \tau < \sqrt[2]{\left(\frac{\ln 2}{\lambda}\right)^{\alpha_2}}$
Compound Weibull (Burr type XII)	(α_i, γ)	$\gamma \ln(1 + x^{\alpha_i})$	$\sqrt[2]{e^{\frac{\ln 2}{\gamma}} - 1} < \tau < \sqrt[2]{e^{\frac{\ln 2}{\gamma}} - 1}$
Rayleigh	σ_i	$\frac{x^2}{2\sigma_i^2}$	$\sqrt{2}\sigma_1 \ln 2 < \tau < \sqrt{2}\sigma_1 \ln 2$
Newsboy	θ_i	$\theta_i l(x)$	$l^{-1}\left(\frac{\ln 2}{\theta_1}\right) < \tau < l^{-1}\left(\frac{\ln 2}{\theta_2}\right)$

6. CONCLUSIONS

This paper derived the system reliability function consisting of two original components and a spare. The spare can switch among the original components. The finding of this paper hold under a general setting. The optimal scheme for switching was also provided. Some special cases which have also practical applications were studied in detail. It was shown that the best time for switching the spare falls between the median lifetimes of the original component lifetimes provided that all component and spare lifetimes be independent. This result does not depend on the distribution of spare lifetimes. The results of this paper may be extended in various directions. For example, one can study the system behaviour under some parametric conditions such as multivariate distribution functions for component lifetimes. The optimal switching times for the case of two and multiple redundancies may also be considered. Engineering systems including parallel-series, series-parallel and mixed systems are worth for consideration in details. Another problem is that the spare allows to switch at possible times $\tau_1, \dots, \tau_k (k \geq 1)$. This means that the spare can switch for k times. Finding optimal switching times τ_1, \dots, τ_k is essential in practice. Works on these topics are under consideration and we hope to report findings soon.

REFERENCES

- [1] AL-Hussaini, E.K., 1999. Predicting Observables from a General Class of Distributions, *Journal of Statistical Planning and Inference*, Vol. **79**, pp.79-91.
- [2] Barlow, R. E., Proschan, F., 1975. *Statistical theory of reliability and life testing*. Holt, Rinehart and Winston Inc. New York.
- [3] Belzunce, F., Martinez-Puertas, H. Ruiz, J. M., 2011. On optimal allocation of redundant components for series and parallel systems of two dependent components, *Journal of Statistical Planning and Inference*, Vol. **141**, pp. 3094-3104.
- [4] Belzunce, F., Martinez-Puertas, H. Ruiz, J. M., 2013. On allocation of redundant components for systems with dependent components, *European Journal of Operational Research*, Vol. **230**, pp. 573-580.
- [5] Boland, P. J., El-Newehi, E., Porschan, F., 1992. Stochastic order for redundancy allocations in series and parallel systems, *Advances in Applied Probability*, Vol. **24**, pp.161-171.
- [6] D.J. Braden, M. Freimer, 1991. Informational dynamics of censored observations, *Management Science*, Vol. **37**, pp.1390-1404.
- [7] Flammini, F., Marrone, S., Mazzocca, N., Vittorini, V., 2009. A new modeling approach to the safety evaluation of N-modular redundant computer systems in presence of imperfect maintenance, *Reliability Engineering & System Safety*, Vol. **94**, pp.1422-1432.

- [8] Hollander, M., Proschan, F., Sethuraman, J., 1977. Functions decreasing in transposition and their applications in ranking problems, *Annals of Statistics*, Vol. **5**, pp.722-733.
- [9] Jeddi, H., Doostparast, M., 2016. Optimal redundancy allocation problems in engineering systems with dependent component lifetime, *Applied Stochastic Models in Business and Industry*, Vol. **32**, pp.199-208.
- [10] Jeddi, H., Doostparast, M., 2022. Allocation of redundancies in systems: A general dependency-base framework, *Annals of Operations Research*, Vol. **312**, pp.259-273.
- [11] Jia, X., Chen, H., Cheng, Z., Gou, B., 2016. A comparison between two switching policies for two-unit standby system, *Reliability Engineering & System Safety*, Vol. **148**, pp.109-118.
- [12] Kim, T. H., Lee, S. H., Lim, J. T., 2001. Stability analysis of switched systems via redundancy factors. *International Journal of Systems Science*, Vol. **32**, pp.1199-1204.
- [13] Lehmann, E. L., 1966. Some concepts of dependence. *The Annals of Mathematical Statistics*, Vol. **37**, pp.1137-1153.
- [14] Li, X., Fang, R., Mi, J., 2015. On the timing to switch on the standby in k -out-of- n : G redundant systems, *Statistics and Probability Letters*, Vol. **96**, pp.10-20.
- [15] Nakagawa, T., 2005. *Maintenance theory of reliability*. Springer, New York.
- [16] J. Navarro, F. Durante, 2017. Copula-based representations for the reliability of the residual lifetimes of coherent systems with dependent components, *Journal of Multivariate Analysis*, Vol. **158**, pp.87-102.
- [17] J. Navarro, F. Pellerey, A. Di Crescenzo, 2015. Orderings of coherent systems with randomized dependent components, *European Journal of Operational Research*, Vol. **240**, pp.127-139.
- [18] Nelsen, R. B., 2006. *An introduction to copulas*. Springer, New York.
- [19] Shaked, M., Shanthikumar, J. G., 2007. *Stochastic orders*. Springer-Verlag, New York.
- [20] Wei Xie, Haitao Liao, Tongdan Jin, (2014). Maximizing system availability through joint decision on component redundancy and spares inventory. *European Journal of Operational Research*, **237**, 164-176.
- [21] Zhang, Y., E Amini-Seresht, Ding W., (2017). Component and system active redundancies for coherent systems with dependent components. *Applied Stochastic Models in Business and Industry*, **33**, 409-421.
- [22] Zhang, X., Zhang, Y., Fang, R., (2020). Allocations of cold standbys to series and parallel systems with dependent components. *Applied Stochastic Models in Business and Industry*, **36**, 432-451.
- [23] Zhao, P., Chan, P.S., Li, L., Ng, H.K.T., (2013). Allocation of two redundancies in two-component series systems. *Naval Research Logistics*, **60**, 588-598.
- [24] Zhao, P., Chan, P.S., Ng, H.K.T., (2012). Optimal allocation of redundancies in series systems. *European Journal of Operational Research*, **220**, 673-683.

DESIGNING OF ACCEPTANCE SAMPLING PLAN BASED ON PERCENTILES FOR TOPP-LEONE GOMPERTZ DISTRIBUTION

S JAYALAKSHMI¹, ALEESHA A²



1 Assistant Professor, Department of Statistics, Bharathiar University Coimbatore - 641 046

2 Research Scholar, Department of Statistics, Bharathiar University Coimbatore - 641 046

E-mail: ¹statjyalakshmi16@gmail.com,² aleeshaa992@gmail.com

Abstract

Acceptance sampling is a statistical technique used to inspect the quality of a batch of products. An acceptance sampling plan under which sampling inspection is performed by conducting life test upon the sampled products is termed as reliability sampling plan. In this paper, a single acceptance sampling plan based on percentile is presented for Topp-Leone Gompertz (TL-G) distribution when the life test is truncated at a pre-specified time. The minimum sample size necessary to ensure the specified life percentile is obtained under a given consumer's risk. The operating characteristic values (and curves) of the sampling plans as well as the producer's risk are presented.

Keywords: Acceptance sampling plan, Percentiles, Topp-Leone Gompertz (TL-G) distribution, Operating characteristic values, Producer's risk, Minimum sample size

1. INTRODUCTION

In statistical quality control, acceptance sampling for products is one aspect of quality assurance. If the quality characteristic is regarding the lifetime of the product, the acceptance sampling problem becomes a life test. Quality personnel would like to know whether the lifetimes of products reach the consumer's minimum standard or not. Traditionally, when the life test indicates that the mean life of products exceeds the specified one, the lot of products is accepted, otherwise it is rejected. For the purpose of reducing the test time and cost, a truncated life test may be conducted to determine the smallest sample size to ensure a certain mean life of products when the life test is terminated at a preassigned time t , and the number of failures observed does not exceed a given acceptance number ' c '.

Nzei et al. [13] developed the Topp-Leone Gompertz (TL-G) distribution. Studies regarding truncated life tests can be found in Epstein [4], Sobel and Tischendorf [20], Goode and Kao [6], Gupta and Groll [7], Gupta [8], Fertig and Mann [5], Kantam and Rosaiah [10], Baklizi [2], Wu and Tsai [22], Rosaiah et al. [19], Tsai and Wu [21], Balakrishnan et al. [3], Rao et al. [15], Aslam et al. [1], Rao et al. [17], . Mahmood et al. [12]. All these authors designed acceptance sampling plans based on the mean life time under a truncated life test using different distributions.

In contrast, Lio et al. [11] considered acceptance sampling plans for percentiles using Birnbaum-Saunders distribution. Srinivasa Rao et al. [16] studied acceptance sampling plans for percentiles based on the inverse Rayleigh distribution. Rao et al. [18] considered acceptance sampling plans for percentiles using Half Normal distribution. Pradeepa Veerakumari and Ponneeswari [14] designed acceptance sampling plan based on percentiles of exponentiated Rayleigh distribution Jayalakshmi and Vijilamery [9] studied Special Type Double Sampling Plan for truncate life test using Gompertz Frechet distribution.

2. TOPP-LEONE GOMPERTZ (TL-G) DISTRIBUTION

The TL-G distribution was developed by Nzei et al. in 2020. The CDF and PDF of the TL-G distribution are given by

$$F(t; \alpha, \delta, \gamma) = [1 - e^{(\frac{-2\delta(e^{\gamma t}-1)}{\gamma})}]^\alpha \quad (1)$$

$$f(t; \alpha, \delta, \gamma) = 2\alpha\delta e^{\gamma t} e^{(\frac{-2\delta(e^{\gamma t}-1)}{\gamma})} [1 - e^{(\frac{-2\delta(e^{\gamma t}-1)}{\gamma})}]^{(\alpha-1)}; t > 0 \quad (2)$$

For given $0 < q < 1$ the $100q^{\text{th}}$ actual percentile of the TL-G distribution can be given by

$$t_q = \frac{1}{\gamma} \ln(1 - \frac{\gamma}{2\delta} \ln(1 - q^{\frac{1}{\alpha}})) \quad (3)$$

The t_q increase as q increases Let

$$\eta = \ln(1 - \frac{\gamma}{2\delta} \ln(1 - q^{\frac{1}{\alpha}})) \quad (4)$$

Then from (3), $\gamma = \eta/t_q$

By letting $a = t/t_q$, $F(t)$ becomes

$$F(t; \alpha, \delta, \gamma) = [1 - e^{(\frac{-2\delta(e^{a\eta}-1)}{\gamma})}]^\alpha \quad (5)$$

Equation (5) gives the modified cdf and by partially differentiating the equation (4) w.r.t a we will get the modified pdf for percentiles of TL-G distribution where t_q is the 25^{th} percentile of the given distribution.

3. RELIABILITY ACCEPTANCE SAMPLING PLAN

A sampling plan in which a decision about the acceptance or rejection of a lot is based on a single sample that has been inspected is known as a Single Sampling Plan. For a single sampling plan, one sample of items is selected at random from a lot and the disposition of the lot is determined from the resulting information. Single Sampling Plans are the most common and easiest plans to use.

Reliability Single Sampling Plans are part of an inspection procedure used to determine whether to accept or reject a specific lot based on lifetime. The Reliability Single Sampling Plan can be represented as $(n, c, t/t_q^0)$. Here n and c are the sample size and acceptance number for the sampling plan. Assume that a life test is conducted and will be terminated at time t_q^0 .

3.1. Operating Procedure

The acceptance sampling plan based on truncated life tests consists of the following:

1. Take a random sample of size n from the lot and inspect them.
2. The maximum test duration time is t .
3. Count the number of defectives d in the sample of size n .
4. The benchmark of defective (d) units is c , where if $d \leq c$ defectives out of n occur at the end of the test period t_q^0 , the lot is accepted. Otherwise reject the lot.

3.2. Minimum sample size

For a fixed P^* our sampling plan is characterized by $(n, c, t/t_q^0)$. Here we consider sufficiently large sized lots so that the binomial distribution can be applied. The problem is to determine for given values of P^* ($0 < P^* < 1$), t_q^0 and c , the smallest positive integer, n required to assert that $t_q > t_q^0$ must satisfy

$$\sum_{i=0}^c p^i (1-p)^{(n-i)} \leq (1-P^*) \tag{6}$$

where $p=F(t, a_q)$, it is the probability of failure time during time t given a specified percentile of a lifetime t_q^0 and it depends on the $a = t/(t_q^0)$ since t_q^0 increases as q increases. Accordingly, we have

$$F(t, a) < F(t, \delta_0) \iff a \leq a_0$$

Or, equivalently

$$F(t; t_q) < F(t; t_q^0) \iff t_q \geq t_q^0$$

The smallest sample size n satisfying eq. (6) can be obtained for any given sampling plan $(n, c, t/t_q^0)$ is given in Table 1.

3.3. Operating Characteristic (OC) Function

The OC function $L(p)$ of the acceptance sampling plan $(n, c, t/t_q^0)$ is the probability of accepting a lot. It is given as

$$L(p) = \sum_{i=0}^c p^i (1-p)^{(n-i)} \tag{7}$$

where $p = F(t, a_q)$. It should be noticed that $F(t, a_q)$ can be represented as a function of $a_q = t/t_q$. Therefore, we have

$$p = F(t, a) = F\left(\frac{t}{t_q}, \frac{1}{d_q}\right)$$

where $d_q = t_q/t_q^0$

Using eq. (7) the OC values can be obtained for any sampling plan $(n, c, t/t_q^0)$. The OC values for the proposed sampling plan is presented in Table 3.

3.4. Producer's Risk (λ)

The producer's risk is defined as the probability of rejecting the lot when $t_q > t_q^0$. For a given value of the producer's risk, say λ , we are interested in knowing the value of d_q to ensure the producer's risk is less than or equal to λ if a sampling plan $(n, c, t/t_q^0)$ is developed at a specified confidence level P^* . Thus, one needs to find the smallest value d_q according to eq. (7).

$$L(p) \geq 1 - \lambda$$

Based on the sampling plans $(n, c, t/t_q^0)$ given in Table 2 the minimum ratios of $d_{0.25}$ at the producer's risk of $\lambda = 0.05$ are presented in Table 4.

4. ILLUSTRATION

Assume that the life distribution is TL-G distribution, and the experimenter is interested in showing that the true unknown 25th percentile life $t_{0.25}$ is at least 1000 hrs. Let $\alpha = 1.9, \delta = 0.125, \gamma = 1.7$ and $\lambda = 0.05$. It is desire to stop the experiment at time $t=3500$ hrs. For the acceptance number $c=1$ from the Table 1 one can obtain the Single Sampling Plan $(n, c, t/t_q^0) = (5, 1, 3.5)$. The optimum sample sizes needed for the given requirement is found to be as $n=5$.

The respective OC values for the proposed acceptance sampling plan $(n, c, t/t_q^0)$ with $P^* = 0.95$

Table 1

$t_q/(t_q^0)$	0.75	1	1.25	1.5	1.75	2	2.25	2.5
L(p)	0.000012	0.05241	0.37108	0.67370	0.83596	0.9137	0.95184	0.97148

for TL-G distribution from the Table 2 are given in above Table 1.

This shows that if the actual 25th percentile is equal to the required 25th percentile ($t_{0.25}/t_{0.25}^0 = 3.5$), the producer’s risk is approximately 0.94759 ($1 - 0.05241$). The producer’s risk almost equal to 0.03 or less when the actual 25th percentile is greater than or equal to 2.5 times the specified 25th percentile.

Table 4 gives the $d_{0.25}$ values for $c=1$ and $t/t_{0.25}^0 = 3.5$ to assure that the producer’s risk is less than or equal to 0.05.

In this example, the value of $d_{0.25}$ is 2.229656 for $c=1$, $t/t_{0.25}^0 = 3.5$ and $\lambda = 0.05$. This means the product can have a 25th percentile life of 2.229656 times the required 25th percentile lifetime. That is under the above Single Sampling Plan the product is accepted with probability of at least 0.95.

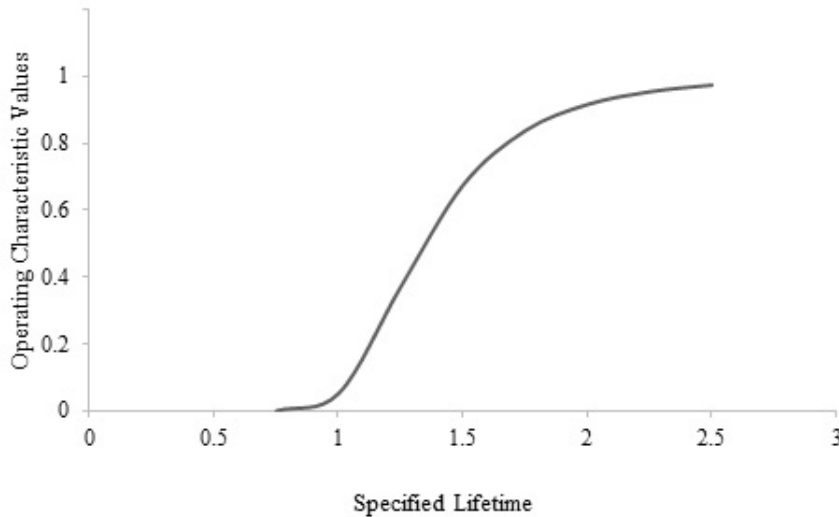


Figure 1: OC curve for the sampling plan ($n = 5, c = 1, t/t_{0.25}^0 = 3.5$)

5. CONSTRUCTION OF THE TABLE

Step 1: Find the value of η for the fixed values of $\alpha = 1.9, \delta = 0.125, \gamma = 1.7$ and $q=0.25$

Step 2: Set the value of $t/t_q^0 = 0.7, 0.9, 0.9, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5$

Step 3: Find the sample size n by satisfying $L(p) \leq 1 - P^*$ when $P^* = 0.99, 0.95, 0.90$ and 0.75 . Here P^* is the probability of rejecting a bad lot and

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

Step 4: for the n value obtained find the $d_{0.25}$ value such that $L(p) \geq 1 - \lambda$ where $\lambda = 0.05$ and $p = F(t/t_q^0, 1/d_q)$; $d_q = t_q/t_q^0$

Table 2: Minimum Sample Size values necessary to assure 25th percentile for TL-G distribution

p^*	c	$t/(t_q^0)$							
		0.7	0.9	1.0	1.5	2.0	2.5	3	3.5
0.75	1	261	144	112	39	18	8	6	3
	2	379	210	162	58	26	14	8	5
	3	495	274	211	75	33	18	10	7
	4	608	336	260	92	41	21	12	8
	5	719	398	309	109	48	25	15	11
0.90	1	377	208	160	57	25	13	7	5
	2	514	284	221	77	34	18	10	6
	3	647	357	276	97	43	21	13	9
	4	74	429	331	116	52	27	15	10
	5	897	498	384	135	60	31	19	12
0.95	1	458	253	196	69	29	15	10	5
	2	609	337	260	90	41	20	12	7
	3	749	414	320	112	50	26	15	9
	4	884	489	379	132	58	31	18	11
	5	1017	562	435	152	68	35	21	13
0.99	1	637	353	271	94	42	21	12	7
	2	803	446	343	121	53	27	15	9
	3	970	534	414	144	64	32	19	12
	4	1118	615	477	166	74	38	22	13
	5	1263	695	539	187	83	43	25	16

Table 3: Operating characteristic values of the sampling plan ($n, c = 1, t/(t_q^0)$) for a given P^* under TL-G distribution

p^*	$t/(t_q^0)$	n	$t_q/(t_q^0)$							
			0.75	1	1.25	1.5	1.75	2	2.25	2.5
0.75	0.7	261	0.0303	0.2489	0.5162	0.7027	0.8152	0.8815	0.9214	0.9462
	0.9	144	0.0255	0.2497	0.5292	0.7188	0.8293	0.8927	0.9299	0.9526
	1	112	0.0225	0.2466	0.5320	0.7238	0.8342	0.8967	0.9330	0.9550
	1.5	39	0.9497	0.9845	0.9938	0.9970	0.9984	0.9990	0.9994	0.9996
	2	18	0.0066	0.2254	0.5666	0.7751	0.8794	0.9315	0.9588	0.9739
	2.5	8	0.0104	0.2979	0.6584	0.8414	0.9216	0.9581	0.9675	0.9805
	3	6	0.0008	0.1566	0.5351	0.7774	0.8905	0.9424	0.9675	0.9805
	3.5	3	0.0046	0.2777	0.6755	0.8650	0.9395	0.9702	0.9840	0.9907
0.90	0.7	377	0.0038	0.0991	0.3191	0.5330	0.6879	0.7900	0.8559	0.8988
	0.9	208	0.0029	0.0995	0.3321	0.5536	0.7087	0.8079	0.8704	0.9103
	1	160	0.0026	0.1009	0.3407	0.5656	0.7202	0.8175	0.8779	0.9161
	1.5	57	0.0010	0.0934	0.3585	0.6009	0.7568	0.8487	0.9027	0.9352
	2	25	0.0005	0.0939	0.3889	0.6446	0.7963	0.8795	0.9256	0.9521
	2.5	13	0.0002	0.0851	0.3998	0.6692	0.8202	0.8983	0.9395	0.9622
	3	7	0.0001	0.0971	0.4461	0.7177	0.8563	0.9229	0.9560	0.9734
	3.5	5	0.00001	0.0524	0.3710	0.6737	0.8359	0.9137	0.9518	0.9714
0.95	0.7	458	0.00087	0.0502	0.2214	0.4298	0.6009	0.7227	0.8054	0.8611
	0.9	253	0.00062	0.0501	0.2324	0.4508	0.6246	0.7444	0.8237	0.8760
	1	196	0.0005	0.0495	0.2367	0.4598	0.6348	0.7538	0.8316	0.8823
	1.5	69	0.00018	0.0470	0.2583	0.5040	0.6833	0.7969	0.8172	0.9988
	2	29	0.00011	0.0555	0.3082	0.5732	0.7464	0.8466	0.8667	0.9375
	2.5	15	0.00003	0.0498	0.3201	0.6018	0.7757	0.8704	0.9039	0.9507
	3	10	0.00001	0.0213	0.2451	0.5458	0.7456	0.8555	0.8172	0.9471
	3.5	5	0.00001	0.0524	0.3710	0.6737	0.8359	0.9137	0.9518	0.9714
0.99	0.7	637	0.00003	0.0104	0.0933	0.2555	0.4306	0.5780	0.6896	0.7706
	0.9	353	0.00001	0.0101	0.0993	0.2733	0.4558	0.6047	0.7143	0.7920
	1	271	0.00001	0.0105	0.1049	0.2863	0.4725	0.6213	0.7290	0.8044
	1.5	94	0.000004	0.0105	0.1243	0.3370	0.5379	0.6855	0.7850	0.8505
	2	42	0	0.0092	0.1362	0.3750	0.5870	0.7324	0.8245	0.8822
	2.5	21	0	0.0093	0.1562	0.4226	0.6409	0.7793	0.8614	0.9103
	3	12	0	0.0074	0.1593	0.4450	0.6703	0.8059	0.8825	0.9262
	3.5	7	0	0.0085	0.1852	0.4945	0.7177	0.8422	0.9085	0.9444

Table 4: Minimum ratio of true $d_{0.25}$ for the acceptability of a lot for the TL-G distribution and producer's risk of $\lambda = 0.05$

p^*	$t/(t_q^0)$	n	$t_q/(t_q^0)$							
0.75	0.7	261	2.5451	2.5464	2.5454	2.5462	2.5467	2.5470	2.5499	2.5496
	0.9	144	2.4588	2.4600	2.4599	2.4611	2.4614	2.4622	2.4578	2.4750
	1	112	2.4264	2.4276	2.4277	2.4290	2.4293	2.4301	2.4273	2.4281
	1.5	39	2.2498	2.2507	2.2517	2.2433	2.2472	2.2521	2.2500	2.2600
	2	18	2.1487	2.1495	2.1506	2.1457	2.1494	2.1478	2.1416	2.1542
	2.5	8	1.9254	1.9191	1.9240	1.9192	1.9183	1.9201	1.9305	1.9229
	3	6	2.0570	2.0575	2.0514	2.0556	2.0526	2.0559	2.0600	2.0635
	3.5	3	1.8133	1.8095	1.8127	1.8122	1.8089	1.8143	1.8122	1.8124
0.90	0.7	377	3.0392	3.0515	3.0508	3.0483	3.0461	3.0442	3.0389	3.0529
	0.9	208	2.9272	2.9362	2.9356	2.934	2.9337	2.9324	2.9304	2.9324
	1	160	2.8693	2.8768	2.8763	2.8754	2.8753	2.8742	2.8734	2.8761
	1.5	57	2.6634	2.6585	2.6691	2.6692	2.6573	2.6571	2.6620	2.6675
	2	25	2.4689	2.4683	2.4662	2.4693	2.4698	2.4717	2.4658	2.4750
	2.5	13	2.3440	2.3433	2.3431	2.3452	2.3463	2.3419	2.3398	2.3515
	3	7	2.1876	2.1875	2.1887	2.1896	2.1861	2.1826	2.1875	2.1911
	3.5	5	2.2309	2.2296	2.2298	2.2313	2.2249	2.2319	2.2275	2.2250
0.95	0.7	458	3.3592	3.3560	3.3529	3.3445	3.3597	3.3578	3.3548	3.3483
	0.9	253	3.2281	3.2253	3.2241	3.2174	3.2289	3.2283	3.2264	3.2238
	1	196	3.1704	3.1678	3.1667	3.1612	3.1563	3.1710	3.1695	3.1681
	1.5	69	2.9000	2.9104	2.9096	2.9082	2.9081	2.9065	2.9054	2.9086
	2	29	2.6336	2.6372	2.6368	2.6370	2.6371	2.6260	2.6323	2.6365
	2.5	15	2.4881	2.4836	2.4807	2.4846	2.4853	2.4879	2.4804	2.5
	3	10	2.5268	2.5294	2.5290	2.5197	2.5206	2.5241	2.5288	2.5250
	3.5	5	2.2309	2.2296	2.2298	2.2313	2.2249	2.2319	2.2275	2.2250
0.99	0.7	637	3.9519	3.9484	3.9580	3.9540	3.9405	3.9577	3.9512	3.9411
	0.9	353	3.7928	3.7890	3.7985	3.7958	3.7860	3.7990	3.7958	3.7899
	1	271	3.7035	3.6996	3.7089	3.7065	3.6989	3.7096	3.7076	3.7032
	1.5	94	3.3554	3.3517	3.3462	3.3578	3.3558	3.3537	3.3496	3.3429
	2	42	3.1019	3.0989	3.0974	3.0906	3.1031	3.1027	3.1017	3.1013
	2.5	21	2.8693	2.8672	2.8660	2.8642	2.8648	2.8632	2.8635	2.8679
	3	12	2.7162	2.7239	2.7230	2.7227	2.7231	2.7232	2.7245	2.7183
	3.5	7	2.5509	2.5545	2.5539	2.5544	2.5546	2.5469	2.5534	2.5537

6. CONCLUSION

In this paper we have derived the acceptance sampling plans based on percentiles for the Topp-Leone Gompertz (TL-G) distribution when the life test is truncated at a pre-fixed time. The minimum sample size required to decide upon accepting or rejecting a lot based on its specified 25th percentile, the operating characteristic function values and corresponding producer's risk are obtained. Tables provided are helpful for the industrial use to save the cost and time of the experiment.

REFERENCES

- [1] Aslam, M., Jun, C. H., Ahmad, M. (2009)). A Group Sampling Plan based on Truncated Life Test for Gamma Distributed Items. *Pak. J. Statist*, Vol.25(3), pp.333-340.
- [2] Baklizi, A. (2003) Acceptance sampling based on truncated life tests in the Pareto distribution of the second kind. *Advances and Applications in Statistics*, Vol.3(1),pp 33-48.
- [3] Balakrishnan, N., Leiva, V., Lopez, J. (2007). Acceptance sampling plans from truncated life tests based on the generalized Birnbaum-Saunders distribution *Communications in Statistics-Simulation and Computation*, Vol.36(3),pp 643-656 .
- [4] Epstein, B. (1954) Truncated life tests in the exponential case. *The Annals of Mathematical Statistics*, Vol.25(3),pp555-564 .
- [5] Fertig, F.W. and Mann, N.R., (1980) Life-test sampling plans for two-parameter Weibull populations *Technometrics*, Vol.22(2),pp 165-177.
- [6] Goode, H. P., KAO, J. H. (1960). Sampling plans based on the Weibull distribution. *In Proceedings of Seventh National Symposium on Reliability and Quality Control, Philadelphia*,p 24-40.

- [7] S. S. Gupta and P. A. Groll, (1961) Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association* ,Vol.56(296),pp 942-970 .
- [8] Gupta, S. S. (1962) Life test sampling plans for normal and lognormal distributions *Technometrics*,Vol.4(2),pp 151-175 .
- [9] Jayalakshmi, S., and S. Vijilamery.(2022). Designing of Special Type Double Sampling Plan for Truncate Life Test using Gompertz Frechet Distribution *International Journal of Mathematics Trends and Technology*,Vol.68(1),pp 70-76 .
- [10] Kantam, R. R. L., and Rosaiah, K. (1998) Half logistic distribution in acceptance sampling based on life tests *IAPQR TRANSACTIONS*,Vol.23,pp 117-126.
- [11] Lio, Y. L., Tsai, T. R., and Wu, S. J. (2009). Acceptance sampling plans from truncated life tests based on the Birnbaum-Saunders distribution for percentiles *Communications in Statistics-Simulation and Computation* ,Vol.39(1),pp 119-136.
- [12] Mahmood, Y., Fatima, S., Khan, H., Amir, H., Khoo, M. B., and Teh, S. Y. (2021). Acceptance sampling plans based on Topp-Leone Gompertz distribution. *Computers and Industrial Engineering* ,159, 107526.
- [13] Nzei, L. C., Eghwerido, J. T., and Ekhosuehi, N. (2020). Topp-Leone Gompertz Distribution: Properties and Applications. *Journal of Data Science* ,vol.18(4), pp782-794.
- [14] PradeepaVeerakumari, K., and Ponneswari, P. (2016). Designing of acceptance sampling plan for life tests based on percentiles of exponentiated Rayleigh distribution. *International Journal of Current Engineering and Technology* ,vol.6(4), 1148-1153.
- [15] Rao, G. S., Ghitany, M. E., and Kantam, R. R. L. (2009). Reliability test plans for Marshall-Olkin extended exponential distribution. *Applied mathematical sciences*,vol.3(55), pp 2745-2755.
- [16] G. Srinivasa Rao, R.R.L. Kantam, K. Rosaiah and J. Pratapa Reddy. (2012). Acceptance sampling plans for percentiles based on the inverse Rayleigh distribution. *Electronic Journal of Applied Statistical Analysis* , vol.5(2), 164-177.
- [17] Rao, G. S., Ghitany, M. E., and Kantam, R. R. L. (2011). An economic reliability test plan for Marshall-Olkin extended exponential distribution. *Applied Mathematical Sciences*,vol.5(3), pp 103-112.
- [18] Rao, B. S., Kumar, C., and Rosaiah, K. (2013). Acceptance sampling plans from life tests based on percentiles of half normal distribution. *Journal of Quality and Reliability Engineering*, 2013.
- [19] Rosaiah, K., Kantam, R. R. L. and Kumar, Santosh.(2006). Reliability test plans for exponentiated log-logistic distribution. *Economic Quality Control* , vol. 21(2), pp 279-289.
- [20] Sobel, M., and Tischendorf, J. A. (1959, January). Acceptance sampling with new life test objectives. *In Proceedings of fifth national symposium on reliability and quality control* , vol. 108, pp. 118).
- [21] Tsai, T. R., and Wu, S. J. (2006). Acceptance sampling based on truncated life tests for generalized Rayleigh distribution. *Journal of Applied Statistics* , vol.33(6), pp. 595-600.
- [22] Wu, C. J., and Tsai, T. R. (2005). Acceptance sampling plans for Birnbaum-Saunders distribution under truncated life tests. *International Journal of Reliability, Quality and Safety Engineering* , vol. 12(06), pp 507-519.

BAYESIAN AND CLASSICAL ESTIMATIONS OF TRANSMUTED INVERSE GOMPERTZ DISTRIBUTION

*¹T. M. ADEGOKE, ²K.O. OBISESAN, ³O. M. OLADOJA & ⁴G.K. ADEGOKE



*^{1,2,3}Department of Mathematics and Statistics, First Technical University, Ibadan, Nigeria.

⁴Department of Computer Science, Nasarawa State University, Keffi, Nigeria.

*¹taiwo.adegoke@tech-u.edu.ng, ²obisesan.olalekan@tech-u.edu.ng,

³oladapo.oladoja@tech-u.edu.ng, ⁴gbolagade.adegoke@nsuk.edu.ng

Abstract

In this article, the use of the transmuted inverse Gompertz distribution in modeling lifetime data is investigated particularly in cases where standard probability distributions are not able to properly handle complex datasets. The quadratic rank transmutation map scheme is utilized to obtain the distribution. The study explores several characteristics of the transmuted inverse Gompertz model, including the estimation of parameters through classical approaches such as maximum likelihood estimation, least-squares estimation, Crammér-Von Misses estimation, and maximum product spacing estimation. Additionally, the Bayesian techniques is used under different loss functions, such as the Linex loss function, square error loss function, and general entropy loss function. The estimates obtained from both classical and Bayesian techniques are evaluated using simulation. To illustrate the potential benefits of the transmuted inverse Gompertz model, a dataset on the strength of aircraft window glass is employed. The results obtained from the application of the new distribution to the real-life dataset demonstrate that it yields superior fits in comparison to other well-known distributions. The study's findings suggest that the transmuted inverse Gompertz model can provide a useful alternative for modeling lifetime data. The research offers valuable insights into the distribution's properties and estimation techniques, as well as its superiority over other commonly used distributions. The new model can contribute to the development of more accurate and efficient models for analyzing lifetime dataset. Overall, the study highlights the importance of exploring new statistical models and techniques to improve data analysis and decision-making

Keywords: Classical method, Bayesian method, Posterior distribution, Loss functions, Transmuted inverse Gompertz distribution.

1. INTRODUCTION

The Gompertz distribution with two parameters is an expansion of the exponential distribution and was proposed by Gompertz [1]. It is widely used in survival analysis to construct accurate actuarial and human mortality tables. Additionally, it is a valuable tool for modeling survival distributions with increasing hazard rates and for describing the distribution of adult lifespans by demographers and actuaries (Willemse and Koppelaar [2]).

The inverse distribution was developed to model actuarial surveys biological and demography (El-Bassiouny *et al.* [3]). The inverse Gompertz distribution was proposed by Eliwa *et al.* [4] and it is useful in modeling lifetime observations. The cumulative probability density (CDF) and probability density function (PDF) are expressed as

$$G(x; \theta, \gamma) = e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \quad \theta > 0, \gamma > 0, x \geq 0 \quad (1)$$

$$g(x; \theta, \gamma) = \frac{\gamma}{x^2} e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right) + \frac{\theta}{x}} \quad \theta > 0, \gamma > 0, x \geq 0 \quad (2)$$

where the scale and shape parameters are presented as θ and γ .

Recently in literature the development of new family of lifetime distributions using transmuted method has recently been attempted to estimate model parameters efficiently for the subject model (see Aryal and Tsokos [5] and Khan *et al.* [6]). Using the techniques known as quadratic rank transmutation map, proposed by Shaw and Buckley [7], we are able to propose a new three parameters transmuted inverse Gompertz (TR-IG) distribution.

If a random variable X has a transmuted distribution, then its cumulative probability density function (CDF) and probability density function (PDF) satisfy the following relationship:

$$F(x) = (1 - \lambda)G(x) - \lambda G^2(x), \quad |\lambda| \leq 1 \quad (3)$$

$$f(x) = g(x) ((1 - \lambda) - 2\lambda G(x)) \quad (4)$$

where $G(x)$ is the CDF of the baseline model, $g(x)$ and $f(x)$ are the corresponding PDF associated with $G(x)$ and $g(x)$, respectively.

The motivation of this work is to proposed a model that can be used to model complex dataset which the standard probability distributions model can't handle properly. Also we examine the potential use of the TR-IG distribution and determine a number of the mathematical features.

2. TRANSMUTED INVERSE GOMPERTZ DISTRIBUTION

Consider a positive value of x with three parameters TR-IG distribution having a shape parameter γ and scale parameter θ and the transmuted parameter $|\lambda| \leq 1$, the CDF can be derived by substituting Equation (1) into Equation (3)

$$F(x; \gamma, \theta, \lambda) = (1 + \lambda) e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} - \lambda \left[e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right]^2 \quad \gamma > 0, \theta > 0, x \geq 0, |\lambda| \leq 1 \quad (5)$$

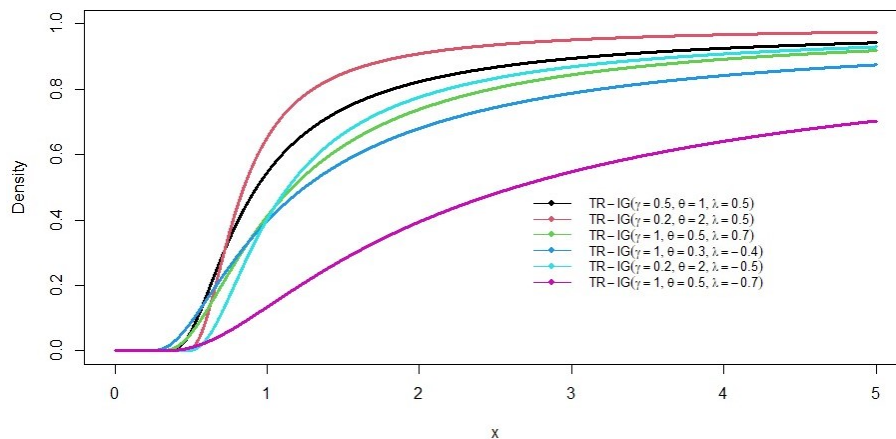


Figure 1: CDF Plot of TR-IG distribution for different parameter values .

Figure 1 shows the CDF plots of TR-IG distribution and we deduced that as x increases the CDF increase and remains constant as it tends to 1.

Substituting Equations (1) and (2) into Equation (4) produces the PDF of TR-IG distribution

$$f(x; \gamma, \theta, \lambda) = \frac{\gamma}{x^2} e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right) + \frac{\theta}{x}} \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right] \quad \gamma > 0, \theta > 0, x \geq 0, |\lambda| \leq 1 \quad (6)$$

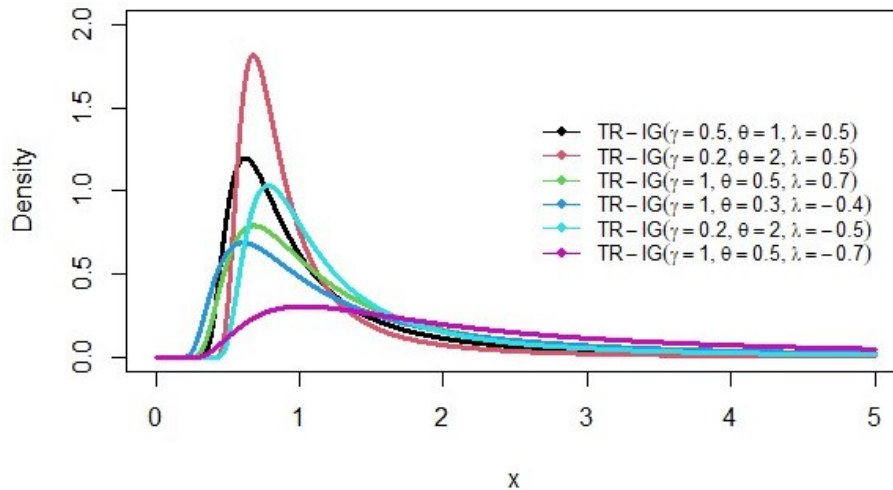


Figure 2: PDF plot of TR-IG distribution for different parameter values

Figure 2 shows that the PDF of TR-IG distribution is positively skewed. When $|\lambda| = 0$, the TR-IG distribution CDF and PDF (5) and (6) becomes inverse Gompertz distribution with CDF (1) and PDF (2).

3. STATISTICAL PROPERTIES

This section looks into some of the statistical characteristics of $TR - IG(\gamma, \beta, \lambda)$ such survival function, reversed hazard function, odds function, hazard function and moments.

3.1. Survival Function

A survival function of the TR-IG distribution can be expressed as

$$S(x) = 1 - (1 + \lambda)e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} - \lambda \left[e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} \right]^2 \quad \gamma > 0, \theta > 0, |\lambda| \leq 1, x > 0 \quad (7)$$

3.2. Hazard Function

The hazard function for the TR-IG distribution can be expressed as

$$h(x) = \frac{\frac{\gamma}{x^2} e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right) + \frac{\theta}{x}} \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} \right]}{1 - (1 + \lambda)e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} - \lambda \left[e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} \right]^2} \quad \gamma > 0, \theta > 0, |\lambda| \leq 1, x > 0 \quad (8)$$

3.3. Reversed Hazard Function (RHF)

The reversed hazard function for the TR-IG distribution can be expressed as

$$\tau(x) = \frac{\frac{\gamma}{x^2} e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right) + \frac{\theta}{x}} \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right]}{(1 + \lambda) e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} - \lambda \left[e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right]^2} \quad \gamma > 0, \theta > 0, |\lambda| \leq 1, x > 0 \quad (9)$$

3.4. Odd Function

The odd function for the TR-IG distribution can be expressed as

$$O(x) = \frac{(1 + \lambda) e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} - \lambda \left[e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right]^2}{1 - (1 + \lambda) e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} - \lambda \left[e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right]^2} \quad \gamma > 0, \theta > 0, |\lambda| \leq 1, x > 0 \quad (10)$$

3.5. Cumulative Hazard Function

The cumulative hazard function for the TR-IG distribution can be expressed as

$$H(x) = -\log \left(1 - (1 + \lambda) e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} - \lambda \left[e^{-\frac{\gamma}{\theta} \left(e^{\frac{\theta}{x}} - 1 \right)} \right]^2 \right) \quad \gamma > 0, \theta > 0, |\lambda| \leq 1, x > 0 \quad (11)$$

3.6. The Quantile Function

The expression for the quantile function is $Q(u) = F^{-1}(u)$. Consequently, the TR-IG distribution quantile function can be written as

$$Q(u) = \frac{\theta}{\log \left\{ 1 - \frac{\theta}{\gamma} \log \left[\frac{(1 + \lambda) - \sqrt{1 + (2 - 4u)\lambda + \lambda^2}}{2\lambda} \right] \right\}} \quad (12)$$

3.7. Moments

Moments can be used to analyze a variety of distributional features, including tendency, skewness, dispersion, and kurtosis. If $X \sim TR - IG(\gamma, \beta, \lambda)$, then the r^{th} moment expression of TR-IG distribution can be expressed as

$$\mu'_r = \int_0^\infty x^r f(x) dx \quad (13)$$

after performing some mathematical expressions, we obtained the moment generating function for the TR-IG distribution

$$\mu'_r = M_x(t) = \sum_{i,j,r=0}^\infty {}^i C_j (-1)^{i+j} \frac{1}{i!} \frac{(-t)^r}{r!} \frac{\Gamma(1-r)}{[\theta(j-i-1)]^{1-r}} \quad (14)$$

4. ESTIMATIONS OF TRANSMUTED INVERSE GOMPERTZ DISTRIBUTION

4.1. Maximum Likelihood Estimation (MLE)

Let x_1, x_2, \dots, x_n be random variables (rvs) drawn from $TR - IG(\gamma, \theta, \lambda)$. The log-likelihood function is defined as follows:

$$L(\gamma, \theta, \lambda | x) = \gamma^n \sum_{i=1}^n \frac{1}{x} e^{-\frac{\gamma}{\theta} \sum_{i=1}^n (e^{\frac{\theta}{x}} - 1) + \theta \sum_{i=1}^n \frac{1}{x}} \prod_{i=1}^n \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta} (e^{\frac{\theta}{x}} - 1)} \right] \quad (15)$$

taking the logarithm of Equation (15), we obtained the log-likelihood function

$$l = n \log \gamma - \log(x) - \frac{\gamma}{\theta} \sum_{i=1}^n (e^{\frac{\theta}{x}} - 1) + \theta \sum_{i=1}^n \frac{1}{x} + \sum_{i=1}^n \log \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta} (e^{\frac{\theta}{x}} - 1)} \right] \quad (16)$$

Maximizing $\log L(\gamma, \theta, \lambda)$ with respect to γ, θ and λ , the system with non-linear equations is obtained as:

$$\frac{\partial l}{\partial \gamma} = \frac{n}{\gamma} - \sum_{i=1}^n \left[\frac{(e^{\frac{\theta}{x}} - 1)}{\theta} \right] + \sum_{i=1}^n \left[\frac{2\lambda \mu (e^{\frac{\theta}{x}} - 1)}{\theta \omega} \right] \quad (17)$$

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^n \left[\frac{\gamma (e^{\frac{\theta}{x}} - 1)}{\theta^2} \right] - \sum_{i=1}^n \left[\frac{\gamma e^{\frac{\theta}{x}}}{\theta x} \right] + \sum_{i=1}^n \left[\frac{1}{x} \right] - \sum_{i=1}^n \left[\frac{2\lambda \rho \mu}{\omega} \right] \quad (18)$$

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^n \left[\frac{(1 - 2\mu)}{\omega} \right] \quad (19)$$

where $\rho = \left(\frac{\gamma (e^{\frac{\theta}{x}} - 1)}{\theta^2} - \frac{\gamma e^{\frac{\theta}{x}}}{\theta x} \right)$, $\mu = e^{-\frac{\gamma (e^{\frac{\theta}{x}} - 1)}{\theta}}$, $\omega = (1 + \lambda - 2\lambda \mu)$

To obtain the estimates of $\hat{\gamma}_{MLE}$, $\hat{\theta}_{MLE}$ and $\hat{\lambda}_{MLE}$, we equate expressions (17) - (19) to zero and solve the system of nonlinear equation. It was observed that solution of this system can not be obtained analytically so we employed a numerical approach known as Newton Raphson method.

4.2. Least-squares Method (OLS)

Let the related order statistics (or) of rvs x_1, x_2, \dots, x_n from the TR-IG distribution sorted in ascending order are denoted as $X_{1:n}, X_{2:n}, \dots, X_{n:n}$. According to Swain *et al.*[8], the OLS estimates of $\hat{\gamma}_{OLS}$, $\hat{\theta}_{OLS}$ and $\hat{\lambda}_{OLS}$ can be obtained by minimizing Equation 20 with respect to γ, θ and λ and equate the non-linear equations to zero.

$$S(\gamma, \theta, \lambda | x) = \sum_{i=1}^n \left[\left\{ (1 + \lambda) e^{-\frac{\gamma}{\theta} (e^{\frac{\theta}{x}} - 1)} - \lambda \left[e^{-\frac{\gamma}{\theta} (e^{\frac{\theta}{x}} - 1)} \right] \right\} - \frac{i}{n+1} \right]^2 \quad (20)$$

4.3. Cramér-von Mises (CVM)

Let the related os of rvs x_1, x_2, \dots, x_n from the TR-IG distribution sorted in ascending order are denoted as $X_{1:n}, X_{2:n}, \dots, X_{n:n}$. To obtain the CVM estimates of $\hat{\gamma}_{CVM}$, $\hat{\theta}_{CVM}$ and $\hat{\lambda}_{CVM}$, Equation 21 is minimized with respect to γ, θ and λ and equate the non-linear equations to zero.

$$C(\gamma, \theta, \lambda) = \frac{1}{12n} + \sum_{i=1}^n \left[\left\{ (1 + \lambda) e^{-\frac{\gamma}{\theta} (e^{\frac{\theta}{x}} - 1)} - \lambda \left[e^{-\frac{\gamma}{\theta} (e^{\frac{\theta}{x}} - 1)} \right] \right\} - \frac{2i-1}{2n} \right]^2 \quad (21)$$

4.4. Maximum Product Spacing Method (MPS)

Let the related os of rvs x_1, x_2, \dots, x_n from the TR-IG distribution sorted in ascending order are denoted as $X_{1:n}, X_{2:n}, \dots, X_{n:n}$. According to Cheng and Amin [9], the MPS estimates of $\hat{\gamma}_{MPS}$, $\hat{\theta}_{MPS}$ and $\hat{\lambda}_{MPS}$, can be obtained by minimizing Equation 22 with respect to γ, θ and λ and equate the non-linear equations to zero.

$$D = \left\{ (1 + \lambda)e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x_i}} - 1\right)} - \lambda \left[e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x_i}} - 1\right)} \right]^2 \right\} - \left\{ (1 + \lambda)e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x_{i-1}}} - 1\right)} - \lambda \left[e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x_{i-1}}} - 1\right)} \right]^2 \right\} \quad (22)$$

4.5. Bayesian Analysis

Let $x = (x_1, x_2, \dots, x_n)$ be a rv with parameters γ, θ and λ having a size n. The posterior probability density function of the parameters γ, θ and λ given x can be expressed as

$$\Pr(\gamma, \theta, \lambda | x) = \frac{\pi(\gamma, \theta, \lambda)l(\gamma, \theta, \lambda)}{\int \int \int \pi(\gamma, \theta, \lambda)l(\gamma, \theta, \lambda)d(\gamma, \theta, \lambda)} \quad (23)$$

where $l(\gamma, \theta, \lambda)$ is the likelihood function expressed in Equation (15) and $\pi(\gamma, \theta, \lambda)$ is the prior probability distribution, which is expressed in Equation (24).

$$\pi(\gamma, \theta, \lambda) = \frac{1}{\gamma\lambda\Gamma(a)b^a} \theta^{a-1} e^{-\frac{\theta}{b}} \quad \gamma, \theta, \lambda, a, b > 0 \quad (24)$$

where: $\gamma \sim Uniform(0, \gamma)$; $\lambda \sim Uniform(0, \lambda)$; $\theta \sim Gamma(a, b)$

substituting Equations (15) and (24), we obtained the posterior probability distribution

$$\Pr(\gamma, \theta, \lambda | x) = \frac{\frac{1}{\gamma\lambda\Gamma(a)b^a} \theta^{a-1} e^{-\frac{\theta}{b}} \frac{\gamma}{x^2} e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right) + \frac{\theta}{x}} \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} \right]}{\int \int \int \frac{1}{\gamma\lambda\Gamma(a)b^a} \theta^{a-1} e^{-\frac{\theta}{b}} \frac{\gamma}{x^2} e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right) + \frac{\theta}{x}} \left[1 + \lambda - 2\lambda e^{-\frac{\gamma}{\theta}\left(e^{\frac{\theta}{x}} - 1\right)} \right] d(\gamma, \theta, \lambda)} \quad (25)$$

4.6. Loss Functions

The Bayesian method was employed for the estimation of TR-IG distribution parameters, utilizing three different types of loss function. The first loss function considered was the LINEX loss function (LLF), also referred to as the linear-exponential loss function, initially proposed by Varian [11]. LLF is an asymmetric loss function that rises exponentially on one side of zero and linearly on the other, as described by Preda and Panaitescu [12].

The second loss function used was the General Entropy loss function (GELF), first introduced by Calabria and Pulcini [13]. GELF is also an asymmetric loss function that has been utilized by several authors, such as Dey and Liu [14], Sule and Adegoke [15], and Ogunsanya *et al.* [17], who used GELF in its original form by setting c to be equal to 1.

The third and final loss function considered was the Squared Error loss function (SELF), which is a common loss function in statistics and machine learning. SELF is also known as the L2 loss function and is used to measure the difference between the estimated and true values squared.

If θ is an estimator required to estimate the parameter $\hat{\theta}$, then the square error loss function can be defined as follows:

$$L_{SELF} \propto (\hat{\theta} - \theta)^2 \tag{26}$$

The LLF can be expressed as

$$L_{LLF} \propto \kappa(e^{c(\hat{\theta}-\theta)} - c(\hat{\theta} - \theta) - 1) \quad \kappa > 0, c \neq 0 \tag{27}$$

where c and κ are the scale and shape parameters of the LLF. In this study, we assume that $\kappa = 1$. Provided that $E_{\theta} [e^{-c\theta}]$ exists, Bayes estimator of the LLF is the value $\hat{\theta}$ that minimizes Equation (27) (Zeller [18]).

$$\hat{\theta} = -\frac{1}{c} \ln \left(E_{\theta} [e^{-c\theta}] \right) \tag{28}$$

The GELF is defined as

$$L_{GELF} \left(\frac{\hat{\theta}}{\theta} \right)^c \propto \left[\left(\frac{\hat{\theta}}{\theta} \right)^c - c \log \left(\frac{\hat{\theta}}{\theta} \right) - 1 \right] \tag{29}$$

where $c > 0$. The minimum occurred at $\hat{\theta} = \theta$. The GELF Bayes estimator is the value $\hat{\theta}$ that minimizes Equation (29) and can be expressed as

$$\hat{\theta} = [E(\theta^{-c})]^{-c} \tag{30}$$

4.7. Lindley Approximation Method

To estimate Equation (25), we will employ an iterative techniques known as Lindley’s approximation to estimate the parameter of interest for γ, θ and λ . According to Lindley [19], if n is large enough, any ratio of the integral of the form

$$I(x) = E[u(\gamma, \theta, \lambda)] = \frac{\int \int \int u(\gamma, \theta, \lambda) e^{l(\gamma, \theta, \lambda) + \rho(\gamma, \theta, \lambda)} \partial(\gamma, \theta, \lambda)}{\int \int \int e^{l(\gamma, \theta, \lambda) + \rho(\gamma, \theta, \lambda)} \partial(\gamma, \theta, \lambda)} \tag{31}$$

where $u(\gamma, \theta, \lambda)$ is a function of γ, θ and λ only, $l(\gamma, \theta, \lambda)$ is the log-likelihood and $\rho(\gamma, \theta, \lambda)$ is the log of prior distribution $\pi(\gamma, \theta, \lambda)$. Equation (31) can be evaluated as

$$\begin{aligned} I(x) = & u(\hat{\gamma}, \hat{\theta}, \hat{\lambda}) + (u_1 a_1 + u_2 a_2 + u_3 + a_3 + a_4 + a_5) + \frac{1}{2} A (u_1 \sigma_{11} + u_2 \sigma_{12} + u_3 \sigma_{13}) \\ & + \frac{1}{2} B (u_1 \sigma_{21} + u_2 \sigma_{22} + u_3 \sigma_{23}) + \frac{1}{2} C (u_1 \sigma_{31} + u_2 \sigma_{32} + u_3 \sigma_{33}) \end{aligned} \tag{32}$$

$$a_i = \rho_1 \sigma_{i1} + \rho_2 \sigma_{i2} + \rho_3 \sigma_{i3} \tag{33}$$

$$a_4 = u_{12} \sigma_{12} + u_{13} \sigma_{13} + u_{23} \sigma_{23} \tag{34}$$

$$a_5 = \frac{1}{2} (u_{11} \sigma_{11} + u_{22} \sigma_{22} + u_{33} \sigma_{33}) \tag{35}$$

$$A = \sigma_{11} L_{111} + 2\sigma_{12} L_{121} + 2\sigma_{13} L_{131} + 2\sigma_{23} L_{231} + \sigma_{22} L_{221} + \sigma_{33} L_{331} \tag{36}$$

$$B = \sigma_{11} L_{112} + 2\sigma_{12} L_{122} + 2\sigma_{13} L_{132} + 2\sigma_{23} L_{232} + \sigma_{22} L_{222} + \sigma_{33} L_{332} \tag{37}$$

$$C = \sigma_{11} L_{113} + 2\sigma_{12} L_{123} + 2\sigma_{13} L_{133} + 2\sigma_{23} L_{233} + \sigma_{22} L_{223} + \sigma_{33} L_{333} \tag{38}$$

$$\rho_i = \frac{\partial \rho}{\partial \theta_i}; \quad u_i = \frac{\partial u(\theta_1, \theta_2, \theta_3)}{\partial \theta_i}; \quad u_{i,j} = \frac{\partial^2 u(\theta_1, \theta_2, \theta_3)}{\partial \theta_i \partial \theta_j}; \quad L_{i,j,k} = \frac{\partial^3 l(\theta_1, \theta_2, \theta_3)}{\partial \theta_i \partial \theta_j \partial \theta_k} \tag{39}$$

where $\theta_1 = \gamma, \theta_2 = \theta$ and $\theta_3 = \lambda$. $\sigma_{i,j}$ is the $(i, j)^{th}$ element of the matrix’s inverse $L_{i,j}$ all evaluated at the MLE estimation and $i, j, k = 1, 2, 3$

$$\frac{d^2l}{d\gamma^2} = -\frac{n}{\gamma^2} - \sum_{i=1}^n \left[\frac{2\lambda\epsilon^2\mu}{\theta^2\omega} \right] - \sum_{i=1}^n \left[\frac{4\lambda^2\epsilon^2\mu^2}{\theta^2\omega^2} \right] \quad (40)$$

$$\begin{aligned} \frac{d^2l}{d\theta^2} = & -\sum_{i=1}^n \left[\frac{2\gamma\epsilon}{\theta^3} \right] + \sum_{i=1}^n \left[\frac{2\gamma e^{\frac{\theta}{x}}}{\theta^2 x} \right] - \sum_{i=1}^n \left[\frac{\gamma e^{\frac{\theta}{x}}}{\theta x^2} \right] - \sum_{i=1}^n \left[\frac{2\lambda\phi\mu}{\omega} \right] \\ & - \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi^2\mu}{\omega} \right] - \sum_{i=1}^n \left[\frac{4\lambda^2\gamma\psi^2\mu^2}{\omega^2} \right] \end{aligned} \quad (41)$$

$$\frac{d^2l}{d\lambda^2} = -\frac{nY^2}{\omega^2}; \quad \frac{\partial^2l}{\partial\theta\partial\lambda} = -\sum_{i=1}^n \left[\frac{2\gamma\psi\mu}{\omega} \right] + \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi\mu Y}{\omega^2} \right] \quad (42)$$

$$\frac{\partial^2l}{\partial\gamma\partial\theta} = \sum_{i=1}^n \left[\frac{\epsilon}{\theta^2} \right] - \sum_{i=1}^n \left[\frac{e^{\frac{\theta}{x}}}{\theta x} \right] - \sum_{i=1}^n \left[\frac{2\lambda\epsilon\mu}{\theta^2\omega} \right] + \sum_{i=1}^n \left[\frac{2\lambda e^{\frac{\theta}{x}}\mu}{\theta x\omega} \right] \quad (43)$$

$$+ \sum_{i=1}^n \left[\frac{2\lambda\epsilon\gamma\psi\mu}{\theta\omega} \right] + \sum_{i=1}^n \left[\frac{4\lambda^2\epsilon\mu^2\gamma\psi}{\theta\omega^2} \right] \quad (44)$$

$$\frac{\partial^2l}{\partial\gamma\partial\lambda} = \sum_{i=1}^n \left[\frac{2\epsilon\mu}{\theta\omega} \right] - \sum_{i=1}^n \left[\frac{2\lambda\epsilon\mu Y}{\theta\omega^2} \right]; \quad \frac{\partial^3l}{\partial\lambda^3} = \sum_{i=1}^n \left[\frac{2Y^3}{\omega^3} \right] \quad (45)$$

$$\frac{\partial^3l}{\partial\gamma^3} = \frac{2n}{\gamma^3} + \sum_{i=1}^n \left[\frac{2\lambda\epsilon^3\mu}{\theta^3\omega} \right] + \sum_{i=1}^n \left[\frac{12n\lambda^2\epsilon^3\mu^2}{\theta^3\omega^2} \right] + \sum_{i=1}^n \left[\frac{16n\lambda^3\epsilon^3\mu^3}{\theta^3\omega^3} \right] \quad (46)$$

$$\begin{aligned} \frac{\partial^3l}{\partial\theta^3} = & \sum_{i=1}^n \left[\frac{6\gamma\epsilon}{\theta^4} \right] - \sum_{i=1}^n \left[\frac{6\gamma e^{\frac{\theta}{x}}}{\theta^3 x} \right] + \sum_{i=1}^n \left[\frac{3\gamma n e^{\frac{\theta}{x}}}{\theta^2 x^2} \right] - \sum_{i=1}^n \left[\frac{\gamma e^{\frac{\theta}{x}}}{\theta x^3} \right] - \sum_{i=1}^n \left[\frac{2\lambda\tau\mu}{\omega} \right] - \sum_{i=1}^n \left[\frac{6\lambda\phi\gamma\psi\mu}{\omega} \right] \\ & - \sum_{i=1}^n \left[\frac{12\lambda^2\phi\mu^2\gamma\psi}{\omega^2} \right] - \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi^3\mu}{\omega} \right] - \sum_{i=1}^n \left[\frac{12\lambda^2\gamma\psi^3\mu^2}{\omega^2} \right] - \sum_{i=1}^n \left[\frac{16\lambda^3\gamma\psi^3\mu^3}{\omega^3} \right] \end{aligned} \quad (47)$$

$$\begin{aligned} \frac{\partial^3l}{\partial\gamma^2\partial\theta} = & \sum_{i=1}^n \left[\frac{4\lambda\epsilon^2\mu}{\theta^3\omega} \right] - \sum_{i=1}^n \left[\frac{4\lambda\epsilon\mu e^{\frac{\theta}{x}}}{\theta^2\omega x} \right] - \sum_{i=1}^n \left[\frac{2\lambda\epsilon^2\gamma\psi\mu}{\theta^2\omega} \right] - \sum_{i=1}^n \left[\frac{12\lambda^2\epsilon^2\mu^2\gamma\psi}{\theta^2\omega^2} \right] \\ & + \sum_{i=1}^n \left[\frac{8\lambda^2\epsilon^2\mu^2}{\theta^3\omega^2} \right] - \sum_{i=1}^n \left[\frac{8\lambda^2\epsilon\mu^2 e^{\frac{\theta}{x}}}{\theta^2\omega^2 x} \right] - \sum_{i=1}^n \left[\frac{16\lambda^3\epsilon^2\mu^3\gamma\psi}{\theta^2\omega^3} \right] \end{aligned} \quad (48)$$

$$\frac{\partial^3l}{\partial\gamma^2\partial\lambda} = -\sum_{i=1}^n \left[\frac{2\epsilon^2\mu}{\theta^2\omega} \right] + \sum_{i=1}^n \left[\frac{2\lambda\epsilon^2\mu Y}{\theta^2\omega^2} \right] - \sum_{i=1}^n \left[\frac{8\lambda\epsilon^2\mu^2}{\theta^2\omega^2} \right] + \sum_{i=1}^n \left[\frac{8\lambda^2\epsilon^2\mu^2 Y}{\theta^2\omega^3} \right] \quad (49)$$

$$\begin{aligned} \frac{\partial^3l}{\partial\theta^2\partial\lambda} = & -\sum_{i=1}^n \left[\frac{2\phi\mu}{\omega} \right] + \sum_{i=1}^n \left[\frac{2\lambda\phi\mu Y}{\omega^2} \right] - \sum_{i=1}^n \left[\frac{2n\gamma\psi^2\mu}{\omega} \right] \\ & + \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi^2\mu Y}{\omega^2} \right] - \sum_{i=1}^n \left[\frac{8\lambda\gamma\psi^2\mu^2}{\omega^2} \right] + \sum_{i=1}^n \left[\frac{8\lambda^2\gamma\psi^2\mu^2 Y}{\omega^3} \right] \end{aligned} \quad (50)$$

$$\begin{aligned} \frac{\partial^3l}{\partial\theta^2\partial\gamma} = & -\sum_{i=1}^n \left[\frac{2\epsilon}{\theta^3} \right] + \sum_{i=1}^n \left[\frac{2e^{\frac{\theta}{x}}}{\theta^2 x} \right] - \sum_{i=1}^n \left[\frac{e^{\frac{\theta}{x}}}{\theta x^2} \right] - \sum_{i=1}^n \left[\frac{2\lambda\phi\mu}{\omega} \right] + \sum_{i=1}^n \left[\frac{2\lambda\phi\epsilon\mu}{\theta\omega} \right] \\ & + \sum_{i=1}^n \left[\frac{4n\lambda^2\phi\mu^2\epsilon}{\omega^2\theta} \right] - \sum_{i=1}^n \left[\frac{4\lambda\gamma\psi\mu\gamma\psi}{\omega} \right] + \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi^2\epsilon\mu}{\theta\omega} \right] \\ & + \sum_{i=1}^n \left[\frac{12\lambda^2\gamma\psi^2\mu^2\epsilon}{\omega^2\theta} \right] - \sum_{i=1}^n \left[\frac{8\lambda^2\gamma\psi\mu^2\gamma\psi}{\omega^2} \right] + \sum_{i=1}^n \left[\frac{16\lambda^3\gamma\psi^2\mu^3\epsilon}{\omega^3\theta} \right] \end{aligned} \quad (51)$$

$$\frac{\partial^3 l}{\partial \lambda^2 \partial \gamma} = - \sum_{i=1}^n \left[\frac{4Y\epsilon\mu}{\omega^2 \theta} \right] + \sum_{i=1}^n \left[\frac{4Y^2 \lambda \epsilon \mu}{\omega^3 \theta} \right]; \quad \frac{\partial^3 l}{\partial \lambda^2 \partial \theta} = \sum_{i=1}^n \left[\frac{4Y\gamma\psi\mu}{\omega^2} \right] - \sum_{i=1}^n \left[\frac{4Y^2 \lambda \gamma \psi \mu}{\omega^3} \right] \quad (52)$$

$$\begin{aligned} \frac{\partial^3 l}{\partial \lambda \partial \theta \partial \gamma} = & - \sum_{i=1}^n \left[\frac{2\gamma\psi\mu}{\omega} \right] + \sum_{i=1}^n \left[\frac{2n\gamma\psi\epsilon\mu}{\theta\omega} \right] + \sum_{i=1}^n \left[\frac{8\gamma\psi\mu^2\lambda\epsilon}{\omega^2\theta} \right] \\ & + \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi\mu Y}{\omega^2} \right] - \sum_{i=1}^n \left[\frac{2\lambda\gamma\psi\epsilon\mu Y}{\theta\omega^2} \right] - \sum_{i=1}^n \left[\frac{8\lambda^2\gamma\psi\mu^2 Y \epsilon}{\omega^3\theta} \right] \end{aligned} \quad (53)$$

where:

$$\begin{aligned} Y &= (1 - 2\mu), \quad \Phi = \left(\frac{e^{\frac{\theta}{x}} - 1}{\theta^2} - \frac{e^{\frac{\theta}{x}}}{\theta x} \right), \quad \psi = \left(\frac{e^{\frac{\theta}{x}} - 1}{\theta^2} - \frac{e^{\frac{\theta}{x}}}{\theta x} \right), \quad \phi = \left(-\frac{2\gamma(e^{\frac{\theta}{x}} - 1)}{\theta^3} + \frac{2\gamma e^{\frac{\theta}{x}}}{\theta^2 x} - \frac{\gamma e^{\frac{\theta}{x}}}{\theta x^2} \right), \\ \varphi &= \left(-\frac{2(e^{\frac{\theta}{x}} - 1)}{\theta^3} + \frac{2e^{\frac{\theta}{x}}}{\theta^2 x} - \frac{e^{\frac{\theta}{x}}}{\theta x^2} \right), \quad \zeta = \left(\frac{6\gamma(e^{\frac{\theta}{x}} - 1)}{\theta^4} - \frac{6\gamma e^{\frac{\theta}{x}}}{\theta^3 x} + \frac{3\gamma e^{\frac{\theta}{x}}}{\theta^2 x^2} - \frac{\gamma e^{\frac{\theta}{x}}}{\theta x^3} \right), \quad \epsilon = \left(e^{\frac{\theta}{x}} - 1 \right) \end{aligned}$$

From the prior distribution in expression (24),

$$\begin{aligned} \rho &= \log(\pi(\gamma, \theta, \lambda)) = (a - 1) \ln \theta - \frac{\theta}{b} - \ln \gamma - \ln \lambda \\ \rho_1 &= \frac{\partial}{\partial \gamma} = -\frac{1}{\gamma}; \quad \rho_2 = \frac{\partial}{\partial \theta} = \frac{a - 1 - b\theta}{\theta}; \quad \rho_3 = \frac{\partial}{\partial \lambda} = -\frac{1}{\lambda} \end{aligned} \quad (54)$$

substituting Equations ((40)-(54)) into Equation (31) reduces the Lindley integral, therefore the Bayes estimates using SELF are thus

- i. If $u(\gamma, \theta, \lambda) = \hat{\gamma}$
 $\hat{\gamma}_{BS} = \hat{\gamma} - \frac{1}{\hat{\gamma}}\sigma_{11} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{12} - \frac{1}{\hat{\gamma}}\sigma_{13} + \frac{1}{2}(A\sigma_{11} + B\sigma_{12} + C\sigma_{13})$
- ii. If $u(\gamma, \theta, \lambda) = \hat{\theta}$ then
 $\hat{\theta}_{BS} = \hat{\theta} - \frac{1}{\hat{\gamma}}\sigma_{21} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{22} - \frac{1}{\hat{\gamma}}\sigma_{23} + \frac{1}{2}(A\sigma_{12} + B\sigma_{22} + C\sigma_{32})$
- iii. If $u(\gamma, \theta, \lambda) = \hat{\lambda}$ then
 $\hat{\lambda}_{BS} = \hat{\lambda} - \frac{1}{\hat{\gamma}}\sigma_{31} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{32} - \frac{1}{\hat{\gamma}}\sigma_{33} + \frac{1}{2}(A\sigma_{13} + B\sigma_{23} + C\sigma_{33})$

Also, the Bayes estimate using LINEX are thus

- i. If $u(\gamma, \theta, \lambda) = e^{-c\hat{\gamma}}$ then
 $\hat{\gamma}_{BL} = \hat{\gamma} + \log \left(1 - c \left(-\frac{1}{\hat{\gamma}}\sigma_{11} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{12} - \frac{1}{\hat{\gamma}}\sigma_{13} - \frac{c}{2}\sigma_{11} + \frac{1}{2}(A\sigma_{11} + B\sigma_{12} + C\sigma_{13}) \right) \right)$
- ii. If $u(\gamma, \theta, \lambda) = e^{-c\hat{\theta}}$ then
 $\hat{\theta}_{BL} = \hat{\theta} + \log \left(1 - c \left(-\frac{1}{\hat{\gamma}}\sigma_{21} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{22} - \frac{1}{\hat{\gamma}}\sigma_{23} - \frac{c}{2}\sigma_{22} + \frac{1}{2}(A\sigma_{12} + B\sigma_{22} + C\sigma_{32}) \right) \right)$
- iii. If $u(\gamma, \theta, \lambda) = e^{-c\hat{\lambda}}$ then
 $\hat{\lambda}_{BL} = \hat{\lambda} + \log \left(1 - c \left(-\frac{1}{\hat{\gamma}}\sigma_{31} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{32} - \frac{1}{\hat{\gamma}}\sigma_{33} - \frac{c}{2}\sigma_{33} + \frac{1}{2}(A\sigma_{13} + B\sigma_{23} + C\sigma_{33}) \right) \right)$

Finally, the Bayes estimate using the GELF are thus

- i. If $u(\gamma, \theta, \lambda) = \hat{\gamma}^{-c}$ then
 $\hat{\gamma}_{BG} = \left[\hat{\gamma}^{-c} \left[1 - \frac{c}{\hat{\gamma}} \left(-\frac{1}{\hat{\gamma}}\sigma_{11} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{12} - \frac{1}{\hat{\gamma}}\sigma_{13} - \frac{c+1}{2\hat{\gamma}} + \frac{1}{2}(A\sigma_{11} + B\sigma_{12} + C\sigma_{13}) \right) \right] \right]^{-\frac{1}{c}}$
- ii. If $u(\gamma, \theta, \lambda) = \hat{\theta}^{-c}$ then
 $\hat{\theta}_{BG} = \left[\hat{\theta}^{-c} \left[1 - \frac{c}{\hat{\theta}} \left(-\frac{1}{\hat{\gamma}}\sigma_{21} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{22} - \frac{1}{\hat{\gamma}}\sigma_{23} - \frac{c+1}{2\hat{\theta}} + \frac{1}{2}(A\sigma_{12} + B\sigma_{22} + C\sigma_{32}) \right) \right] \right]^{-\frac{1}{c}}$
- iii. If $u(\gamma, \theta, \lambda) = \hat{\lambda}^{-c}$ then
 $\hat{\lambda}_{BG} = \left[\hat{\lambda}^{-c} \left[1 - \frac{c}{\hat{\lambda}} \left(-\frac{1}{\hat{\gamma}}\sigma_{31} + \frac{a-1-b\hat{\theta}}{\hat{\theta}}\sigma_{32} - \frac{1}{\hat{\gamma}}\sigma_{33} - \frac{c+1}{2\hat{\lambda}} + \frac{1}{2}(A\sigma_{13} + B\sigma_{23} + C\sigma_{33}) \right) \right] \right]^{-\frac{1}{c}}$

5. RESULTS

5.1. Simulation Techniques

In this section, we consider a monte carlo simulation study to evaluate the performance of all the estimators using the MSE and biases with respect to different sample sizes $n = (20, 50, 100, 200, 500)$ for different parameters $TR - IG(\gamma, \theta, \lambda) = [(1,1,1), (0.7, 1, 0.5), (0.5, 1, -0.5)$ and $(0.5, 0.5, -0.2)]$ respectively. The results obtained from the analysis are displayed in Tables (1 - 7) and the results shown that the estimates using both the classical techniques and Bayesian methods performed excellently in estimating model parameters of the TR-IG distribution since the estimated results are closed to the true parameter values with small MSE and bias as the sample sizes increases for all the estimation techniques considered in this study.

Table 1: The MLE estimates, Bias and MSE for different parameter values.

n	Values	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20		1.3933	1.7840	-2.170	0.3933	0.7840	-3.1707	0.1547	0.6146	0.6146
50	$\gamma = 1$	1.0849	1.0267	2.0412	0.0849	0.0267	1.0412	0.0072	0.0007	0.0007
100	$\theta = 1$	1.2748	1.4741	1.5233	0.2748	0.4741	0.5233	0.0755	0.2248	0.2248
200	$\lambda = 1$	1.0842	0.9210	1.0594	0.0842	-0.0789	0.0594	0.0071	0.0062	0.0062
500		1.0406	0.9230	1.0752	0.0406	-0.0769	0.0752	0.0016	0.0059	0.0059
20		0.8490	0.9426	0.9628	0.1490	-0.0573	0.4628	0.0222	0.0032	0.0032
50	$\gamma = 0.7$	0.8012	0.4173	2.1917	0.1012	-0.5826	1.6917	0.0102	0.3395	0.3395
100	$\theta = 1$	0.9900	0.6209	1.5020	0.2900	-0.3790	1.0020	0.0841	0.1436	0.1436
200	$\lambda = 0.5$	0.8170	1.0175	0.5633	0.1170	0.01757	0.0633	0.0137	0.0003	0.0003
500		0.7696	0.9641	0.5986	0.0696	-0.0358	0.0986	0.0048	0.0012	0.0012
20		0.6706	1.2368	-1.7885	0.1706	0.2368	-1.2885	0.0291	0.0561	0.0561
50	$\gamma = 0.5$	0.4520	0.5452	-0.3392	-0.0479	-0.4547	0.1607	0.0023	0.2068	0.2068
100	$\theta = 1$	0.5724	0.9028	-0.5957	0.0724	-0.0971	-0.0957	0.0052	0.0094	0.0094
200	$\lambda = -0.5$	0.6274	0.9089	-0.3868	0.1274	-0.0910	0.1131	0.0162	0.0082	0.0082
500		0.5189	0.9966	-0.4976	0.0189	-0.0033	0.0023	0.0003	1.1E-05	1.1E-05
20		0.0679	-0.0188	-0.1817	-0.4320	-0.5188	0.0182	0.1866	0.2691	0.2691
50	$\gamma = 0.5$	0.2330	0.0920	0.2068	-0.2667	-0.4079	0.4068	0.0712	0.1664	0.1664
100	$\theta = 0.5$	0.3784	0.3681	-0.2509	-0.1215	-0.1318	-0.0509	0.0147	0.0173	0.0173
200	$\lambda = -0.2$	0.4593	0.4362	-0.1983	-0.0406	-0.0637	0.0016	0.0016	0.0040	0.0040
500		0.4179	0.5230	-0.3623	-0.0820	0.0230	-0.1623	0.0067	0.0005	0.0005

Table 2: The Bayesian estimate using GELF, Bias and MSE for different parameter values.

n	Values	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20	$\gamma = 1$ $\theta = 1$ $\lambda = 1$ $c = 1$	1.5700	1.9749	0.2172	-0.5700	-0.9749	0.7828	0.3249	0.9504	0.9504
50		1.0552	0.9894	1.6972	-0.0552	0.0106	0.6972	0.0030	0.0001	0.0001
100		1.2573	2.1781	1.8270	-0.2573	-1.1781	-0.8270	0.0662	1.3878	1.3878
200		1.0766	0.9160	1.0860	-0.0766	0.0840	-0.0860	0.0059	0.0071	0.0071
500		1.0377	0.9208	1.0854	-0.0377	0.0792	-0.0854	0.0014	0.0063	0.0063
20	$\gamma = 0.7$ $\theta = 1$ $\lambda = 0.5$ $c = 1$	0.7795	0.8400	-2.7292	-0.0795	0.1600	1.2292	0.0063	0.0256	0.0256
50		0.8106	0.4566	-0.8339	-0.1106	0.5434	1.3339	0.0122	0.2953	0.2953
100		0.9805	0.6114	8.3239	-0.2805	0.3886	-7.8239	0.0787	0.1510	0.1510
200		0.8090	1.0061	0.6476	-0.1090	-0.0061	-0.1476	0.0119	0.0000	0.0000
500		0.7667	0.9596	0.6238	-0.0667	0.0404	-0.1238	0.0044	0.0016	0.0016
20	$\gamma = 0.5$ $\theta = 1$ $\lambda = -0.5$ $c = -1$	0.6499	1.1989	-1.7308	-0.1499	-0.1989	1.2308	0.0225	0.0395	0.0395
50		0.4390	0.5267	-0.2967	0.0610	0.4733	-0.2033	0.0037	0.2240	0.2240
100		0.5669	0.8942	-0.5799	-0.0669	0.1058	0.0799	0.0045	0.0112	0.0112
200		0.6246	0.9049	-0.3776	-0.1246	0.0951	-0.1224	0.0155	0.0090	0.0090
500		0.5181	0.9953	-0.4949	-0.0181	0.0047	-0.0051	0.0003	0.0000	0.0000
20	$\gamma = 0.5$ $\theta = 0.5$ $\lambda = -0.2$ $c = -1$	0.0362	-0.0674	-0.0715	0.4638	0.5674	-0.1285	0.2151	0.3219	0.3219
50		0.2153	0.0708	0.2833	0.2847	0.4292	-0.4833	0.0811	0.1842	0.1842
100		0.3703	0.3577	-0.2202	0.1297	0.1423	0.0202	0.0168	0.0202	0.0202
200		0.4553	0.4314	-0.1808	0.0447	0.0686	-0.0192	0.0020	0.0047	0.0047
500		0.4167	0.5214	-0.3567	0.0833	-0.0214	0.1567	0.0069	0.0005	0.0005

Table 3: The Bayesian estimate using LINEX, Bias and MSE for different parameter values

n	Values	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20	$\gamma = 1$ $\theta = 1$ $\lambda = 1$ $c = 1$	0.3809	0.4126	2.2994	0.6191	0.5874	-1.2994	0.3833	0.3451	0.3451
50		1.2102	0.6444	-2.5123	-0.2102	0.3556	3.5123	0.0442	0.1264	0.1264
100		1.1813	0.1028	0.6599	-0.1813	0.8972	0.3401	0.0329	0.8050	0.8050
200		1.1288	1.0870	0.9989	-0.1288	-0.0870	0.0011	0.0166	0.0076	0.0076
500		1.0672	1.0210	1.0269	-0.0672	-0.0210	-0.0269	0.0045	0.0004	0.0004
20	$\gamma = 0.7$ $\theta = 1$ $\lambda = 0.5$ $c = 1$	0.9130	1.0132	0.2642	-0.2130	-0.0132	0.2358	0.0454	0.0002	0.0002
50		1.2158	0.9125	-0.3400	-0.5158	0.0875	0.8400	0.2661	0.0077	0.1685
100		1.1905	0.8456	-0.2682	-0.4905	0.1544	0.7682	0.2406	0.0239	0.0239
200		0.7941	0.9342	0.4847	-0.0941	0.0658	0.0153	0.0089	0.0043	0.0043
500		0.7636	0.9384	0.5599	-0.0636	0.0616	-0.0599	0.0040	0.0038	0.0038
20	$\gamma = 0.5$ $\theta = 1$ $\lambda = -0.5$ $c = -1$	0.6660	1.2612	-0.7200	-0.1660	-0.2612	-0.2200	0.0276	0.0682	0.4852
50		0.4446	0.5059	-0.3653	0.0554	0.4941	-0.1347	0.0031	0.2442	0.2442
100		0.5743	0.9084	-0.6158	-0.0743	0.0916	0.1158	0.0055	0.0084	0.0084
200		0.6285	0.9109	-0.3955	-0.1285	0.0891	-0.1045	0.0165	0.0079	0.0079
500		0.5193	0.9980	-0.5008	-0.0193	0.0020	0.0008	0.0004	0.0000	0.0000
20	$\gamma = 0.5$ $\theta = 0.5$ $\lambda = -0.2$ $c = -1$	0.0007	-0.3024	-0.2992	0.4993	0.8024	0.0992	0.2493	0.6438	0.6438
50		0.2015	0.0137	0.0638	0.2985	0.4863	-0.2638	0.0891	0.2365	0.2365
100		0.3741	0.3568	-0.2791	0.1259	0.1432	0.0791	0.0159	0.0205	0.0205
200		0.4584	0.4332	-0.2103	0.0416	0.0668	0.0103	0.0017	0.0045	0.0045
500		0.4181	0.5233	-0.3653	0.0819	-0.0233	0.1653	0.0067	0.0005	0.0005

Table 4: The Bayesian estimate using SELF, Bias and MSE for different parameter values

n		$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20		0.9476	1.2083	0.9019	0.0524	-0.2083	0.0981	0.0027	0.0434	0.0434
50	$\gamma = 1$	1.1452	0.8396	1.0749	-0.1452	0.1604	-0.0749	0.0211	0.0257	0.0257
100	$\theta = 1$	1.2268	0.9133	1.1424	-0.2268	0.0867	-0.1424	0.0514	0.0075	0.0075
200	$\lambda = 1$	1.1058	1.0044	1.0302	-0.1058	-0.0044	-0.0302	0.0112	0.0000	0.0000
500		1.0537	0.9721	1.0515	-0.0537	0.0279	-0.0515	0.0029	0.0008	0.0008
20		0.8749	0.9668	0.6708	-0.1749	0.0332	-0.1708	0.0306	0.0011	0.0011
50	$\gamma = 0.7$	1.0160	0.6759	0.8044	-0.3160	0.3241	-0.3044	0.0998	0.1051	0.1051
100	$\theta = 1$	1.0910	0.7341	0.8397	-0.3910	0.2659	-0.3397	0.1529	0.0707	0.0707
200	$\lambda = 0.5$	0.8050	0.9751	0.5277	-0.1050	0.0249	-0.0277	0.0110	0.0006	0.0006
500		0.7664	0.9509	0.5806	-0.0664	0.0491	-0.0806	0.0044	0.0024	0.0024
20		0.5120	1.0705	-0.6552	-0.0120	-0.0705	0.1552	0.0001	0.0050	0.0050
50	$\gamma = 0.5$	0.6015	0.8504	-0.4043	-0.1015	0.1496	-0.0957	0.0103	0.0224	0.0224
100	$\theta = 1$	0.5988	0.9835	-0.4885	-0.0988	0.0165	-0.0115	0.0098	0.0003	0.0003
200	$\lambda = -0.5$	0.6511	0.9742	-0.3346	-0.1511	0.0258	-0.1654	0.0228	0.0007	0.0007
500		0.5214	1.0081	-0.4678	-0.0214	-0.0081	-0.0322	0.0005	0.0001	0.0001
20		0.4064	0.5934	-0.6487	0.0936	-0.0934	0.4487	0.0088	0.0087	0.0087
50	$\gamma = 0.5$	0.4748	0.4553	-0.3847	0.0252	0.0447	0.1847	0.0006	0.0020	0.0020
100	$\theta = 0.5$	0.4836	0.5361	-0.4501	0.0164	-0.0361	0.2501	0.0003	0.0013	0.0013
200	$\lambda = -0.2$	0.5183	0.5333	-0.3103	-0.0183	-0.0333	0.1103	0.0003	0.0011	0.0011
500		0.4336	0.5511	-0.3879	0.0664	-0.0511	0.1879	0.0044	0.0026	0.0026

Table 5: OLS estimates, Bias and MSE for the simulated values

n	Values	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20		1.0444	0.7986	1.1393	-0.0444	0.2014	-0.1393	0.0020	0.0405	0.0405
50	$\gamma = 1$	0.9979	1.0074	1.0142	0.0021	-0.0074	-0.0142	0.0000	0.0001	0.0001
100	$\theta = 1$	1.0355	1.0912	0.9270	-0.0355	-0.0912	0.0730	0.0013	0.0083	0.0083
200	$\lambda = 1$	1.0684	1.1656	0.8362	-0.0684	-0.1656	0.1638	0.0047	0.0274	0.0274
500		1.0309	1.0772	0.9979	-0.0309	-0.0772	0.0021	0.0010	0.0060	0.0060
20		0.5445	1.0569	0.6307	0.1555	-0.0569	-0.1307	0.0242	0.0032	0.0032
50	$\gamma = 0.7$	0.7043	1.0029	0.5329	-0.0043	-0.0029	-0.0329	0.0000	0.0000	0.0000
100	$\theta = 1$	0.7956	1.0147	0.4648	-0.0956	-0.0147	0.0352	0.0091	0.0002	0.0002
200	$\lambda = 0.5$	0.8361	1.0624	0.3252	-0.1361	-0.0624	0.1748	0.0185	0.0039	0.0039
500		0.7327	1.0585	0.4770	-0.0327	-0.0585	0.0230	0.0011	0.0034	0.0034
20		0.4300	0.9754	-0.3656	0.0700	0.0246	-0.1344	0.0049	0.0006	0.0006
50	$\gamma = 0.5$	0.5046	1.0022	-0.4892	-0.0046	-0.0022	-0.0108	0.0000	0.0000	0.0000
100	$\theta = 1$	0.5923	1.0111	-0.5143	-0.0923	-0.0111	0.0143	0.0085	0.0001	0.0001
200	$\lambda = -0.5$	0.6208	1.0654	-0.5974	-0.1208	-0.0654	0.0974	0.0146	0.0043	0.0043
500		0.5374	1.0534	-0.5178	-0.0374	-0.0534	0.0178	0.0014	0.0029	0.0029
20		0.5315	0.3582	-0.1056	-0.0315	0.1418	-0.0944	0.0010	0.0201	0.0201
50	$\gamma = 0.5$	0.5138	0.4900	-0.1926	-0.0138	0.0100	-0.0074	0.0002	0.0001	0.0001
100	$\theta = 0.5$	0.5237	0.5591	-0.2425	-0.0237	-0.0591	0.0425	0.0006	0.0035	0.0035
200	$\lambda = -0.2$	0.5523	0.5970	-0.3090	-0.0523	-0.0970	0.1090	0.0027	0.0094	0.0094
500		0.5197	0.5442	-0.2383	-0.0197	-0.0442	0.0383	0.0004	0.0020	0.0020

Table 6: MPS estimates, Bias and MSE for the simulated values

n	Values	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20		1.0230	0.9286	0.7827	-0.0230	0.0714	0.2173	0.0005	0.0051	0.0051
50	$\gamma = 1$	1.0807	0.8271	0.9403	-0.0807	0.1729	0.0597	0.0065	0.0299	0.0299
100	$\theta = 1$	1.1303	0.9557	1.0101	-0.1303	0.0443	-0.0101	0.0170	0.0020	0.0020
200	$\lambda = 1$	1.0660	1.0193	0.9686	-0.0660	-0.0193	0.0314	0.0044	0.0004	0.0004
500		1.0181	0.9949	1.0038	-0.0181	0.0051	-0.0038	0.0003	0.0000	0.0000
20		0.9815	0.6772	0.6045	-0.2815	0.3228	-0.1045	0.0792	0.1042	0.1042
50	$\gamma = 0.7$	1.0213	0.6075	0.7410	-0.3213	0.3925	-0.2410	0.1032	0.1540	0.1540
100	$\theta = 1$	1.0915	0.6921	0.7970	-0.3915	0.3079	-0.2970	0.1532	0.0948	0.0948
200	$\lambda = 0.5$	0.8248	0.9325	0.5298	-0.1248	0.0675	-0.0298	0.0156	0.0046	0.0046
500		0.7810	0.9256	0.5875	-0.0810	0.0744	-0.0875	0.0066	0.0055	0.0055
20		1.6379	0.1379	0.5668	-1.1379	0.8621	-1.0668	1.2947	0.7432	0.7432
50	$\gamma = 0.5$	1.5484	0.1713	0.6225	-1.0484	0.8287	-1.1225	1.0991	0.6868	0.6868
100	$\theta = 1$	1.6676	0.2513	0.6615	-1.1676	0.7487	-1.1615	1.3633	0.5606	0.5606
200	$\lambda = -0.5$	0.6542	0.9326	-0.3554	-0.1542	0.0674	-0.1446	0.0238	0.0045	0.0045
500		0.9890	0.6552	0.2423	-0.4890	0.3448	-0.7423	0.2391	0.1189	0.1189
20		0.8555	1.1375	0.4301	-0.3555	-0.6375	-0.6301	0.1264	0.4065	0.4065
50	$\gamma = 0.5$	1.1167	0.0400	0.5593	-0.6167	0.4600	-0.7593	0.3803	0.2116	0.2116
100	$\theta = 0.5$	1.2114	0.0889	0.5987	-0.7114	0.4111	-0.7987	0.5062	0.1690	0.1690
200	$\lambda = -0.2$	0.5270	0.5020	-0.3156	-0.0270	-0.0020	0.1156	0.0007	0.0000	0.0000
500		1.1111	0.0000	0.6778	-0.6111	0.5000	-0.8778	0.3735	0.2500	0.2500

Table 7: The CVM estimates, Bias and MSE for the simulated values

n	Values	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$	$\hat{\gamma}_{Bias}$	$\hat{\theta}_{Bias}$	$\hat{\lambda}_{Bias}$	$\hat{\gamma}_{MSE}$	$\hat{\theta}_{MSE}$	$\hat{\lambda}_{MSE}$
20		1.0209	0.9660	1.1022	-0.0209	0.0340	-0.1022	0.0004	0.0012	0.0012
50	$\gamma = 1$	1.0513	1.0769	0.9379	-0.0513	-0.0769	0.0621	0.0026	0.0059	0.0059
100	$\theta = 1$	1.0598	1.1720	0.8364	-0.0598	-0.1720	0.1636	0.0036	0.0296	0.0296
200	$\lambda = 1$	1.1118	1.2235	0.7814	-0.1118	-0.2235	0.2186	0.0125	0.0500	0.0500
500		1.0926	1.0991	0.8669	-0.0926	-0.0991	0.1331	0.0086	0.0098	0.0098
20		0.6705	1.0208	0.5809	0.0295	-0.0208	-0.0809	0.0009	0.0004	0.0004
50	$\gamma = 0.7$	0.7955	1.0136	0.4675	-0.0955	-0.0136	0.0325	0.0091	0.0002	0.0002
100	$\theta = 1$	0.8104	1.0894	0.3628	-0.1104	-0.0894	0.1372	0.0122	0.0080	0.0080
200	$\lambda = 0.5$	0.8899	1.1010	0.2976	-0.1899	-0.1010	0.2024	0.0361	0.0102	0.0102
500		0.7668	1.0906	0.4191	-0.0668	-0.0906	0.0809	0.0045	0.0082	0.0082
20		0.4851	1.0183	-0.4424	0.0149	-0.0183	-0.0576	0.0002	0.0003	0.0003
50	$\gamma = 0.5$	0.5593	1.0397	-0.5468	-0.0593	-0.0397	0.0468	0.0035	0.0016	0.0016
100	$\theta = 1$	0.6335	1.0521	-0.5911	-0.1335	-0.0521	0.0911	0.0178	0.0027	0.0027
200	$\lambda = -0.5$	0.6754	1.0881	-0.6524	-0.1754	-0.0881	0.1524	0.0308	0.0078	0.0078
500		0.5966	1.0494	-0.5740	-0.0966	-0.0494	0.0740	0.0093	0.0024	0.0024
20		0.5134	0.4739	-0.1671	-0.0134	0.0261	-0.0329	0.0002	0.0007	0.0007
50	$\gamma = 0.5$	0.5278	0.5544	-0.2396	-0.0278	-0.0544	0.0396	0.0008	0.0030	0.0030
100	$\theta = 0.5$	0.5601	0.5904	-0.3026	-0.0601	-0.0904	0.1026	0.0036	0.0082	0.0082
200	$\lambda = -0.2$	0.6012	0.6143	-0.3492	-0.1012	-0.1143	0.1492	0.0102	0.0131	0.0131
500		0.5453	0.5665	-0.2794	-0.0453	-0.0665	0.0794	0.0021	0.0044	0.0044

5.2. Application

In this section, we analyzed the strength of glass of aircraft window datasets adopted by Fuller *et al.* [20] (whose dataset is displayed in Table (8) to ascertain that the TR-IG distribution is a good lifetime model, when compared with three known distribution like Inverse Gompertz (IGD), inverse Rayleigh (IR) and inverse Exponential distribution (IE). To assess the TR-IG distribution’s goodness-of-fit with these distributions, some criteria (such as the Kolmogorov-Smirnov test statistic (KS), log-likelihood (L) values, Akaike information criterion (AIC), Bayesian information criterion (BIC), and Cramér-von Mises statistic (W^*) and Anderson-Darling statistic (A^*)) were used to fit all of the above-mentioned distributions.

Figure 3 displayed the estimated PDFs, estimated CDFs of all tested distributions and the empirical and theoretical CDF. From the results displayed in Table 9, we deduced that the TR-IG distribution fits the data better than the other four model. Table 10, shows the classical and Bayesian estimates for the strength of glass of aircraft windows.

Table 8: The strength of glass of aircraft window.

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

Table 9: The estimates and goodness-of-fit measurements for glass strength data set.

Statistics	Model			
	IE	IR	IG	TR-IG
γ	29.215	810.504	1.249	0.6563
θ	-	- 0.6411	119.762	126.5584
λ	-	-	-	-
KS	0.477	0.325	0.139	0.1349
-L	137.262	118.201	107.884	94.5072
AIC	241.2363	208.4237	195.9174	193.0145
BIC	242.5321	209.7196	198.5090	196.902
A*	7.2108	3.5316	0.6034	0.5349
W*	1.5051	0.6627	0.0841	0.0748

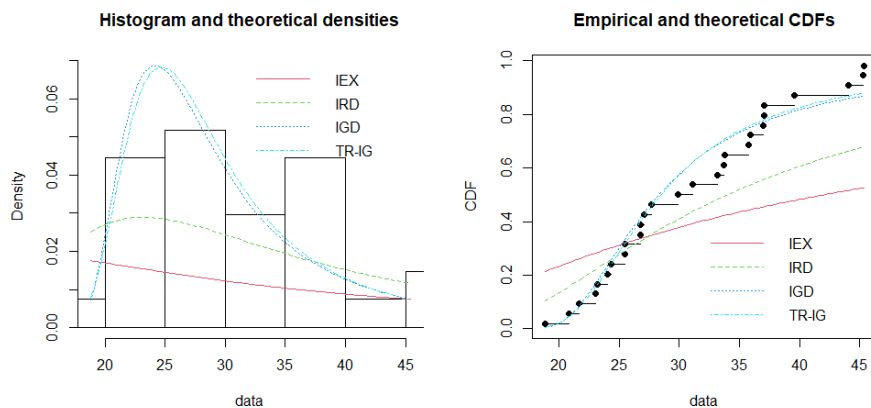


Figure 3: The Histogram, empirical and theoretical densities for Glass strength dataset .

Table 10: Estimated values of the strength of glass for aircraft window under different estimation techniques

Estimation Techniques	$\hat{\gamma}$	$\hat{\theta}$	$\hat{\lambda}$
CVM	0.4574	124.1316	-0.7821
MPS	0.9403	112.6889	-0.6803
SELF	0.5428	121.5032	-0.8790
LINEX (c =-1)	0.5467	129.2076	-0.9585
OLS	0.5663	122.5142	-0.5257
GELF (c =1)	0.7698	129.7899	-0.5152

6. CONCLUSION

We introduced a novel model called the TR-IG distribution, which expands upon the inverse Gompertz distribution for analyzing data with a real support. One clear motivation for extending a standard distribution is to increase the flexibility in modeling complex dataset. We derived some properties such as hazard function, survival function and etc. The parameters were estimated using both the classical and Bayesian estimation techniques. The utilization of the TR-IG distribution on actual data demonstrates that this new distribution can be employed with great effectiveness to yield superior fits in comparison IE, IR and IG distributions.

REFERENCES

- [1] Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies. . *Philosophical transactions of the royal society of London*, 115, 513–583. <https://doi.org/10.1098/rstl.1825.0026>.
- [2] Willemse, W. and Koppelaar, H. (2000). Knowledge elicitation of Gompertz law of mortality. *Scandinavian actuarial journal*, 2, 168–179. <https://doi.org/10.1080/034612300750066845>.
- [3] El-Bassiouny, A. H., El-Damcese, M. Mustafa, A. and Eliwa, M. S. (2017). Exponentiated generalized Weibull-Gompertz distribution with application in survival analysis. *Journal of statistics applications and probability*, 6(1), 7–16. <https://dx.doi.org/10.18576/jsap/060102>.
- [4] Eliwa, M., El-Morshedy, M. and Ibrahim, M. (2019). Inverse Gompertz Distribution: Properties and Different Estimation Methods with Application to Complete and Censored Data. *Annals Of Data Science*. 6, 321-339. <https://doi.org/10.1007/s40745-018-0173-0>
- [5] Aryal, G. and Tsokos, C. (2011). Transmuted Weibull Distribution: A Generalization of the Weibull Probability Distribution. *European Journal Of Pure And Applied Mathematics*. 4, 89-102.
- [6] Khan, M., King, R. and Hudson, I. (2016) Transmuted gompertz distribution: properties and estimation. *Pakistan Journal of Statistics*. 32, 161-182 .
- [7] Shaw, W. and Buckley, I. (2009). The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. *ArXiv*.
- [8] 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.
- [9] Cheng, R. and Amin, N. (1983). Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin. *Journal of The Royal Statistical Society: Series B (Methodological)*. 45, 394-403.
- [10] Kao, J. (1959). A Graphical Estimation of Mixed Weibull Parameters in Life Testing Electron Tube. *Technometrics*. 1, 389-407.
- [11] Varian, H. A. Bayesian Approach to Real Estate Assessment. North Holland, Amsterdam, 1975.

- [12] Preda, V. and Panaitescu, E. (2010). Bayes estimators of Modified-Weibull distribution parameters using Lindley's approximation. *WSEAS Transactions On Mathematics*. 9, 539-549.
- [13] Calabria, R. and Pulcini, G. (1994). An engineering approach to Bayes estimation for the Weibull distribution. *Microelectronics Reliability*. 34, 789-802, 1994.
- [14] Dey, D. and Liu, P. (1992) . On Comparison of Estimators in a Generalized Life Model. *Micro-electron. Reliab.* 45, 207-221.
- [15] Sule, B.O and Adegoke, T.M. (2020). Bayesian approach in estimation of shape parameter of an Exponential Inverse Exponential Distribution. *Asian Journal Of Probability And Statistics*. 9, 13 - 27.
- [16] Adegoke, T. M., Yahya, W. B and Adegoke, G. K (2018) A Study of the Bayes Estimator of the Parameters in an Inverted Generalized Exponential Distribution with Application. *Annals Of Statistical Theory And Applications*, 1, 1-10.
- [17] Ogunsanya, A. S, Yahya, W. B, Adegoke, T. M, Iluno, C., Aderere, O. and Iwada, E. A. (2021) New Three-Parameter Weibull Inverse Rayleigh Distribution: Theoretical Development and Applications. *Mathematics And Statistics*. 9, 249-272.
- [18] Zellner, A. (1986). Bayesian Estimation and Prediction using Asymmetric Loss Functions. *Jour. Amer. Statist. Assoc.* 446 - 451.
- [19] Lindley, D. Approximate Bayes Methods. Bayesian Statistics. Valency, 1980.
- [20] Fuller, R., Freiman, S., Quinn, J., Quinn, G. and Carter, W. (1994). Fracture Mechanics Approach to the Design of Glass Aircraft Windows: A Case Study. *Proceedings Of SPIE - The International Society For Optical Engineering*.7, 419-430.

ON THE BAYESIAN MODELING OF ROAD ACCIDENTS DUE TO VEHICLE TYPE IN OYO STATE NIGERIA

*¹OLADAPO M. OLADOJA, ²K.O. OBISESAN, ³SURURAT A. ADEBAYO & ⁴T. M. ADEGOKE

*^{1,2,3,4}Department of Mathematics and Statistics, First Technical University, Ibadan, Nigeria
*¹oladapo.oladoja@tech-u.edu.ng, ²obidairo@gmail.com,
³sururatadeboyin@gmail.com, ³taiwo.adegoke@tech-u.edu.ng

Abstract

Transportation plays an important role in the day to day activities of human race. Road, rail, water and air are the four major forms of transportation in Nigeria. Road transportation is the most common means of movement in Oyo State and this has raise the likelihood of the occurrence of road traffic accident even posing a serious problem that needs serious attention. Road traffic accidents would Continually increase if tangible efforts are not made to tackle is problem. In Oyo state, the epicenter of the western Nigeria, road transportation is the most popular means of transportation. Of concern is the accidents recorded daily on the roads. This study takes into account the series of accidents caused on the state roads due to vehicle types. Seven vehicle types were considered due to the data available. Bayesian Model Averaging (BMA), a variable selection approach, was used to handle uncertainty in the model selection process. Several classical approach like time series have been used in analysing road accidents but few had explored the Bayesian approach via BMA route. A uniform model prior was used with a numeric g-prior to improve the predictive performance. The trend of accidents were observed for a period of 15 years from (2006 - 2020). 78% of the models were visited and using the Posterior Inclusion Probability results (99.546%), it was seen that the accidents that occurred via private cars was the predominant. It is the most important in modeling vehicle type accidents in Oyo State Nigeria. Also, accidents for 2023 and 2024 were predicted using the posterior predictive distribution. It is imperative for concerned authorities (Federal Road Safety Corps amongst others) in the state to look into the issue of private car owners in the state.

Keywords: Bayesian Model Averaging, Posterior Inclusion Probability, Predictive, Trend, Transportation, Private Cars

1. INTRODUCTION

Transport plays a significant role in the day to day activities of man. The four main modes of transportation are air travel, road travel, river travel, and rail travel. Of these forms of transport, the most popular is road transport. As a country develops in economy, it proportionally becomes more motorized, as a result, more persons are expected to use road transport. This increase in the usage of road transport raises the probability of road traffic accident occurrence to a significant amount. Road traffic accidents are serious problems facing mankind today.

According to a 2018 World Health Organization publication on the state of global road safety (WHO), there have been increase in fatalities emanating from road traffic accident to 1.3 million annually (WHO [1]).

Nigeria is ranked as a lower middle-income economy, over the last decade an increase in Road transport accidents has been experienced. This is as a result of growing motorization and increase in urbanization in the country. As a low-income economy, road infrastructure development is still

lagging, as are policy challenges in meeting international safety standards. According to Ogbodo and Nduoma [2], the death rate from road traffic accidents is 162 per 100,000 people. This data compares to the global average of 22 deaths per 100,000 people (Sukhai *et al.* [3]).

Road traffic accident has become one of the issues of great concern in Nigeria. A road traffic accident occurs almost every day, resulting in an overall increase in morbidity and mortality rates, as well as financial costs to both society and the individual involved. It is disheartening that road accident is becoming a daily occurrence due to the fact that road traffic integration is treated as secondary issue in Nigeria.

According to Sanusi, *et al.* [4], road traffic accident has been classified using the International Statistical Classification of Diseases and Related Health Problems, as one of the leading causes of death worldwide. Road traffic accident was defined by Odugbemi [5] as anything that occur by chance, anything happening unexpectedly and undersigned. It has been noticed that there is an overall increase in the incidence, morbidity, and mortality rates of road traffic accidents around the world, with the majority of these fatalities and morbidities occurring in developing nations such as Nigeria. (Eze [6]; Agbonkhese *et al.* [7]).

Road traffic accidents are caused by one or more of three basic variables: environmental factors, mechanical factors, and human factors. Human factors are considered to be responsible for around 80% of all traffic accidents since drivers' operating ability is critical to the causes and prevention of traffic accidents Afolabi and Gbadamosi [8]. Agbonkhese, *et al.* [7] opined that on many Nigerian roads, deterioration often begins with the production of cracks or potholes on the road tarmacs, which vary depending on their configuration, traffic flow, forms, amplitude of stress, and rate of deformation. Aside from human and vehicle factors, the existence of these potholes is known to be one of the leading causes of road traffic accidents in Nigeria.

Deaths from reckless driving are the third biggest cause of mortality in Nigeria as a whole. According to Agbonkhese *et al.* [7], 473 people died in a total of 1,115 automotive accidents in Nigeria in 2012, while more people perished in vehicular accidents in 2013. They advocated for treating road traffic accidents as a serious issue that requires immediate attention in order to prevent early deaths and reduce the health, social, and economic consequences for the typical Nigerian.

Despite the scale of the problem and the loss of life, Oyo state has failed to receive the necessary attention. It is past time for the government to prioritize road traffic accidents in order to reduce their health, social, and economic consequences. In this project, an analysis of a 15-year (2006-2020) publically accessible data set will be sought to reveal patterns and traffic conditions on Oyo state's highways. The primary goal is to analyze the prevalence of accidents in Oyo state based on deaths, causative agent, and kind of vehicle involved.

Many safety strategies have been put in place by the Nigerian Government to check the menace of road traffic accidents through its agencies. For these policies to be effective continuous researches on the Road Traffic Accident (RTA) cases as well as on the implementation of the policies need to be carried out. Therefore, considering the importance of the road and the increased level of road traffic accidents in recent years, there is the need for this study which aimed at obtaining a model which would show the pattern of road traffic accidents. This would provide a parameter for assessing the effectiveness of current strategies for reducing accidents on our roads and for the development of new strategies where necessary by those responsible for maintaining safety on our local

2. NIGERIAN ROAD TRAFFIC ACCIDENTS

A crash is an unexpected event that causes damage, injuries, and, in extreme cases, death. A crash might be single (just one vehicle involved) or multiple (two or more vehicles involved). Accidents are classified as fatal, serious, or minor. A crash is considered fatal when there is a loss of human life. A serious crash occurs when someone is critically hurt and hospitalized, whereas a minor crash occurs when no injuries occur or when the victim is treated for minor injuries and immediately dismissed from the hospital. According to Chun *et al.* [9] and Abayomi [10], the

causes of road traffic accidents can be divided into three categories: human factors, mechanical factors, and environmental factors.

According to Oyenuga *et al.* [11], human factors account for approximately 80% of the causes of road traffic accidents, with the following being the leading causes: drunk driving, illiteracy, psychological factors, reliance, poor vision, temperaments, overconfidence, poor driving culture, economic factors, and underage drivers. Mechanical issues include the usage of motorized vehicles such as cars, trucks, buses, and motorcycles without adequate maintenance. Iwok [12] concurs that, while the driver (human factor) accounts for roughly 80% of the causal index of road crashes in Nigeria, vehicle conditions are a component that cannot be overlooked in this type of analysis. The geography of most of Nigeria's road network presents significant challenges to road development. Mountains, valleys, and rivers lead to acute bends, steep hills, and sharp slopes, all of which are potentially dangerous aspects for inexperienced highway motorists. Our tropical climate also presents difficulties for drivers. Heavy rainfall in the south and abnormally hot weather conditions, along with harmattan dust in northern Nigeria, have an impact on our road network. Potholes are easily formed, and deadly black's pots all pose obstacles to reading users.

In his study titled "Time Series Analysis of Road Accident in Osun State," Iwok [12] suggested that in order to better the work of the FRSC, there should be an emergency line for the commission to contact in case of an accident. Aluko [13] used multiple correlation coefficients to analyze her data and came to the conclusion that if the government improves the measures against motor vehicle accidents, the total number of accidents and adverse socioeconomic effects would be greatly reduced, allowing the country to move forward positively into an advanced or developing country. Trivedi and Rawal [14], evaluated the frequency of major and minor road traffic accidents among young drivers, as well as its relationship with driving practices. The study employed a cross-sectional design with young drivers drawn from tuition programs in Ahmedabad and Vadodara. The findings revealed that the prevalence of road traffic accidents is high among young drivers and is associated to driving at high speeds, using mobile phones, and failing to observe safety measures while driving.

According to Agbonkhese *et al.* [7], reckless driving deaths are the third biggest cause of death in Nigeria. According to Agbonkhese *et al.* [7], at least 473 people died in a total of 1,115 automobile incidents in Nigeria in 2012. More people perished in automotive accidents in 2013. They advocated for treating road traffic accidents as a serious issue that requires immediate attention in order to prevent early deaths and reduce the health, social, and economic consequences for the typical Nigerian. Using bus priority measures in Melbourne and analyzing resultant road safety performance and bus-involved accidents, Chun *et al.* [9], summarized their findings. An empirical investigation of accident types found a considerable reduction in the proportion of accidents involving buses colliding with stationary objects and cars, indicating the influence of bus priority in addressing bus maneuverability concerns. A main result of this study is that bus priority improves road safety in Melbourne and should be a major concern for road management organizations when implementing bus priority and road plans. Abdulkabir and Edem [15], investigated the pattern of accident occurrence in Ibadan, Nigeria. Their research revealed an increase tendency.

According to Sanusi *et al.* [4], vehicle traffic accidents in Nigeria have been increasing at an alarming rate. They examined road traffic accidents in Nigeria from 1960 to 2013 using time series analysis. They provided appropriate Autoregressive Integrated Moving Average (ARIMA) models for several kinds of road accidents, including minor cases, major cases, fatal cases, and total cases. They discovered that the ARIMA (1,1,1) model works best for minor and total cases, the ARIMA (1,1,0) model works best for serious instances, and the ARIMA (0,1,1) model works best for fatal cases. The forecast based on the various models suggests an average increase in the data for all scenarios. Iwok [12], utilized the ARIMA model to fit data from road traffic accidents in Port Harcourt. The seasonal-ARIMA model was found to fit the data in the study. Oyenuga *et al.* [11], also analyzed the pattern of monthly road accidents data along Oyo-Ibadan express road between 2004 to 2014. They employed moving average method to decompose the time series

using additive model approach. From the result they observed that accidents and deaths were higher at festive periods. Emenike and Kanu [18], examined how road traffic accidents can be caused by drivers distraction in Port Harcourt. The study also publicized that the use of mobile phones and gadgets in vehicles were blamable for most accidents involving commercial drivers.

Aluko [13], investigated the characteristics of road accident victims in Ado-Ekiti, Nigeria. According to the report, a bigger proportion of those involved in vehicle accidents were of working age. Macharia *et al.* [19] examined monthly road accident data for eight years, beginning in January 2010 and ending in December 2017. They used the Box-Jekins method for the analysis and Eviews as the statistical software. The study found no seasonality in the data, contradicting popular notion that accidents in this region of the world are seasonal, with increased incidence in September, October, November, and December. This implies that road accidents occur throughout the year, not just during specific seasons, within the research area. Feldkircher and Zeugner [20], utilized the ARIMA model to anticipate accidents up to ten years later (2030) and characterize the frequency of road traffic incidents that result in harm. A total of 70039 road traffic accidents were recorded during the 9-year observation period (from 2013 to 2021). The time plot shows a systematic shift, indicating a trend in the data. Furthermore, the trend exhibits an exponential decrease. This analysis concluded that traffic accidents are falling exponentially at a rate of 0.360 per year. The number of road accidents was predicted to fall to 47 (97%) by 2030, which is higher than the current objective of 50%, and that the time series for the yearly number of road traffic accidents did not exhibit substantial seasonality or moving average components.

The significance of this study is that it provides a method of minimizing the number of road traffic accidents/collisions in Oyo state and throughout Nigeria. It will assist road users and management of the Federal Road Safety Corps and the Nigeria police in determining the rate and working toward the reduction of accidents in Oyo state, particularly in the always busy and congested areas that have the highest reported cases and casualties of road traffic crash/accidents.

3. METHODOLOGY

3.1. Averaging in the Bayesian Model (BMA)

Canonical regression problems including model uncertainty are addressed with Bayesian Model Averaging (BMA) Akanbi and Oladoja, [21], (Akanbi and Oladoja [22], Obisesan and Oladoja [23]). Given a linear model, y is the response variable, μ_i is the intercept, and φ_j are the predictors,

$$y = \mu_i + X_j\varphi_j + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I) \quad (1)$$

where ε is independently and identically distributed with mean zero and variance σ^2 . The problem arises if matrix X has a large number of explanatory variables. It is inefficient or even impossible to draw conclusions from a single linear model that includes all variables when the sample size is small. An alternative strategy is to estimate models for all possible combinations of $[X]$ and then construct a weighted average of those models. The estimate of 2^K models is required if X consists of K potential variables. As a result of Bayes' theorem, posterior model probabilities are used for weighting the models.

$$p(B_j|y, X) = \frac{p(y|B_j, X)p(B_j)}{p(y|X)} = \frac{p(y|B_j, X)p(B_j)}{\sum_{s=1}^{2^K} p(y|B_s, X)p(B_s)} \quad (2)$$

The integrated likelihood over all models is denoted by the multiplicative term $p(y|X)$. In other words, the posterior model probability (PMP) ($p(B_j|y, X)$) can be determined as a function of the model's marginal likelihood (MLM) ($p(y|B_j, X)$), the probability of the data given the model multiplied by a prior model probability ($p(B_j)$) - a researcher's belief before looking at data is the likelihood of the model being accurate. For any statistic, renormalization yields the model

weighted posterior distribution (MWPDP) π .

$$p(\pi|y, X) = \sum_{j=1}^{2^k} p(\pi|B_j, X)p(B_j|X, y) \tag{3}$$

Priors on the model parameters must be specified in order to obtain posterior distributions. The constant variance and error variance here have ‘improper’ priors, which means their distribution is equal. The researcher constructs a normal distribution based on her prior beliefs on coefficients before analyzing the data. Often, it is assumed that the coefficients have zero prior means because little is known about them. Their variance structure is

$$\varphi_j|g \sim N\left(0, \sigma^2\left(\frac{1}{g}X_j'X_j\right)^{-1}\right) \tag{4}$$

Essentially, the researcher believes coefficients are zero and their variance-covariance structure closely matches the data’s. In relation to g , a researcher’s level of certainty is expressed by the degree to which he or she believes that coefficients are indeed zero. It implies that the researcher is confident (or conservative) that there are no prior coefficient variances. An increase in g , in contrast, indicates a researcher’s uncertainty about the coefficients.

3.2. Predictive Performance of Priors in BMA

Forecasting is a primary goal of statistical analysis. Similarly, there is always the argument that, everything else being equal, when comparing rival modeling systems, we are more impressed with a modeling strategy that consistently assigns greater probabilities to the events that really occur. Thus, one way to assess the efficacy of a BMA strategy is to measure how effectively a model predicts future data. The logarithmic scoring rule, which is based on the conditional predictive ordinate, is one measure of predicting ability. The Log Prediction Score (LPS) specifically gauges an individual model’s predictive skill by summing the logarithms of the observed ordinates of the predictive density for each observation in the test set.

$$- \sum_{d \in D^{test}} \log P(d|B, D^{train}) \tag{5}$$

The LPS is considered by splitting the data into two halves: training set, D^{train} (which will use the observation to estimate the BMA predictive distribution by obtaining the parameter posterior distributions, $P(\pi_j|y^*, B_j)$ and testing set, D^{test} (this formula forms the weights over the space of the model to measure its predictive ability). In order to find the posterior predictive density, you have to take the predictive likelihood and the posterior distribution of the parameters and multiply them.

$$p(\tilde{y}|y^*, B_j) = \int_{\pi_j} p(\tilde{y}|\pi_j, y^*, B_j)p(\pi_j|y^*, B_j)d\pi_j \tag{6}$$

It indicates how likely it is that the future observations \tilde{y} have been generated under model B_j given data y^* .

$$- \sum_{d \in D^{test}} \log \left[\sum_{B \in A} P(d|Z, D^{train})P(B|D^{train}) \right] \tag{7}$$

The better the prediction performance, the smaller the log predictive score for a specific model or model average. We can see that the logarithmic scoring rule is correct. There are two kinds of disparities between observed and expected values in probabilistic predictions: the predictor tends to overestimate or underestimate their predictive accuracy because of predictive biases (synthetically predicting on either side) and lack of calibration (synthetically underestimating or overestimating). The predicted log score is a bias and calibration metric.

4. RESULTS AND DISCUSSION

4.1. Exploratory Data Analysis

The data for this study is road accident data sourced from Oyo State Bureau of Statistics. It includes the types of vehicle (taxi, private car, bus, motor lorry, motor cycle, pedal cycle and hand manual) plying the road in the state. It is an annual data that spans from 2006 to 2020. Nigeria’s inland state of Oyo is located in the southwest. According to the 2016 estimate, Oyo State’s population is estimated at 7,840,864. Road is a major means of travelling in the state with, the state capital and previously the continent’s second most populous city, Ibadan a commercial hub linking the west to the northern axis of the country. On the average, the number of accidents caused by private car outweighs that of the other vehicle types while accidents caused by pedal cycle has the least number of accidents as displayed in Table 1 below in the state.

Table 1: Five-Figure Summary of Road Accidents Due to Vehicle Types in Oyo State from 2006 to 2020

Vehicle	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
Taxi	8.0	36.5	60.0	97.3	147.5	323.0
Private Car	14.0	152.5	192.0	218.9	250.0	677.0
Bus	13.0	119.5	163.0	197.1	193.5	695.0
Motor Lorry	3.0	100.0	130.0	160.1	189.0	426.0
Motor Cycle	10.0	98.0	129.0	179.3	242.0	473.0
Pedal Cycle	0.0	5.0	16.0	22.6	39.0	63.0
Hand Manual	0.0	13.0	50.0	53.53	65.0	152.0

The trend for accident due to vehicle types is displayed in Figure 1 below. The trend shows that the highest number of road accidents in the state occurred in 2013 and the number of road accidents begins to reduce in 2020 due to the COVID’19 outbreak.

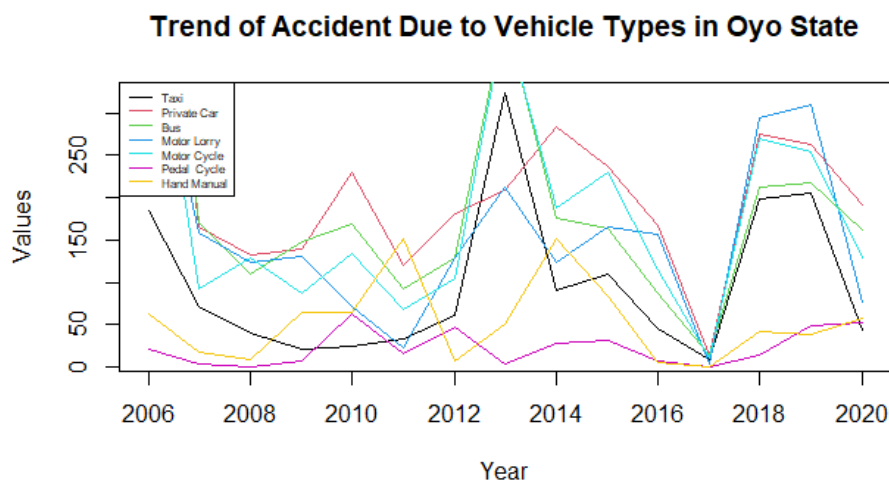


Figure 1: Time Plots for Vehicle Types of Road Accidents in Oyo State

4.2. Bayesian Model Averaging of Road Accidents

The study variable is the serious occurrences of accidents in the state from 2006 to 2020 and the predictor are the accidents caused by different type of vehicles namely taxi (TAXI), private

cars (PRIVATE.CAR), bus (BUS), motor lorry (MOTOR.LORRY), motor cycle (MOTOR.CYCLE), pedal cycle (PEDAL.CYCLE) hand manual (HAND.MANUAL). The Markov Chain Monte Carlo sampler uses 100,000 draws after the burn ins of 25,000 with uniform distribution as the prior model and the numeric modified g-priors for the parameters. Therefore the accident model due to vehicle types is given as

$$Accidents = \theta_0 + \theta_1 TAXI + \theta_2 PRIVATE.CAR + \theta_3 BUS + \theta_4 MOTOR.LORRY + \theta_5 MOTOR.CYCLE + \theta_6 PEDAL.CYCLE + \theta_7 HAND.MANUAL + \epsilon$$

where, ϵ is a stochastic error term, independent and identically distributed as $N(0, \sigma^2)$.

The posterior probabilities of incorporating each of the regressors are shown in Table 2. It can be seen that there are 7 explanatory factors and 15 years of observations of relevance. The model space is 128 and the number of models visited is 100, indicating that 78% of the models were visited. The modified g-prior employed for the application to significant accidents caused by vehicle types in Oyo State, Nigeria, established that the shrinkage factor was close to one, indicating overfitting. The averages and standard deviations of the Posterior Inclusion

Table 2: Summary of the Posterior Probabilities of Including each of the Regressors

Variables	7
Observations	15
Mean no. regressors	0.9973
Draw	100000
Burnings	25000
No. models visited	100
Model Space 2^k	128
% Visited	78
% Top Models	100
Corr PMP	1.000
Model Prior g-prior	Uniform/3.5 numeric
Shrinkage-Stats (Average)	1.0

Probabilities (PIP) of each regressor in the vehicle type accident model are shown in Table 3. Post Mean represents the coefficients averaged over all models, even those in which the variable was not present (implying that the coefficient is zero in this case). The covariate private automobiles with a PIP of 99% and a relatively big coefficient band appears to be the most relevant in modeling Nigeria vehicle type accidents in Oyo state. This demonstrates that private cars play an important part in the selection of any vehicle type accident model in Oyo state. Table 4 displays the posterior

Table 3: Posterior Probabilities of including the Regressors in the Accident Model

Regressors	PIP	Post Mean	Post SD	Cond.Pos.Sign	Index
PRIVATE.CAR	0.99546	1.4382	0.1462	1.0000	2
BUS	0.00147	0.10216	0.2980	0.96599	3
PEDAL.CYCLE	0.00033	-1.4630	0.9392	0.00000	6
TAXI	0.00002	-0.19399	0.2613	0.00000	1
MOTOR.CYCLE	0.00001	0.1553	0.2797	0.00000	5
MOTOR.LORRY	0.00000	0.00000	0.00000	0.00000	4
HAND.MANUAL	0.00000	0.00000	0.00000	0.00000	7

probability of the top five models out of the 128 visited for both the MCMC and precise samplers. This table shows that the best model, with a probability of 99.5%, depicts vehicle type accidents in Oyo State, with private automobile as the predictor. The chart shows that the genuine vehicle type accident model (20) is always preferred over any other model. Figure 2 depicts the prior

Table 4: Best 5 models of 128 models visited

Model	PMP (Exact)	PMP (MCMC)	Predictors
20	0.99492	0.99505	PRIVATE.CAR
00	0.00350	0.00312	
10	0.00118	0.00058	BUS
22	0.000204	0.000060	PRIVATE.CAR, PEDAL.CYCLE
28	0.000078	0.00026	PRIVATE.CAR, MOTOR.LORRY

and posterior distribution of model sizes, which helps to demonstrate the impact of the model prior assumption on the estimation outcomes. In accordance with the literature (Ley and Steel, 2009; Eicher *et al.*, 2011), the plot created allows for visual clarity of the choice of model prior on posterior outcomes.

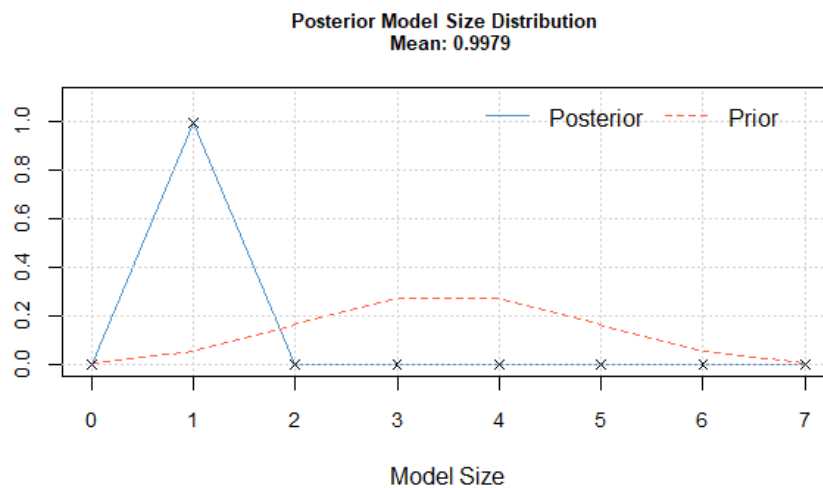


Figure 2: Posterior Model Size Distribution with Uniform Model Priors

Figure 3 depicts the mixed marginal posterior density for the significant regression coefficients. The red dotted vertical lines represent the equivalent standard deviation bounds from the MCMC technique, while the red and green vertical lines represent the conditional expected value and median, respectively. The charts show that the values of the conditional expected value and the median are very close to each other. Given that the related variable is included in the regression, the density in the graphs describes the posterior distribution of the regression coefficient.

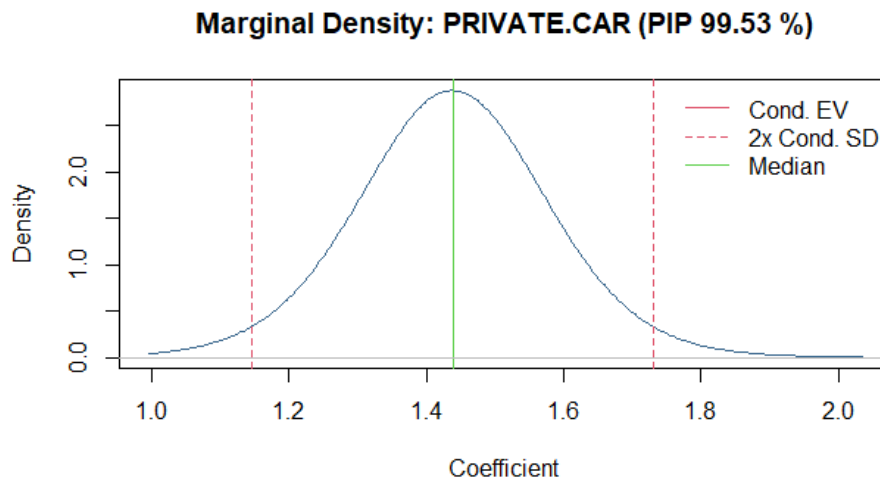


Figure 3: Posterior Density of Private Cars Accidents in Oyo State

Bayesian Model Averaging lends itself not only to inference, but also to prediction. The employed Bayesian regression models naturally give rise to predictive densities, whose mixtures yields the BMA predictive density. The predictive density represents the likelihood of future accidents based on our fitted model. Using the information from the first 15 years (2006-2020) to forecast vehicle type accidents for the next two years, namely 2023 and 2024 as displayed in Table 5 below. The 95% credible intervals were also estimated.

5. CONCLUSION

Because BMA provides researchers with a thorough framework for measuring model uncertainty, the theoretical and empirical evidence presented in this paper demonstrates the vital role of previous assumptions in BMA. In Oyo state, Nigeria, road accidents were on the high side in 2013. Series of vehicle types accident were monitored from 2006 to 2020 namely Taxi, Private Cars, Bus, Motor Lorry, Motor Cycle, Pedal Cycle and Hand Manuals. Using model averaging approach, the posterior probabilities of the explanatory variables was able to visit the 78% of the model space. Private Cars with a Posterior Inclusion Probability of 99% is important in modelling vehicle type accidents in Oyo state Nigeria. There is need for concerned authorities (Federal Road Safety Corps amongst others) in the state to look into the issue of private car owners in the state.

ACKNOWLEDGEMENTS

Acknowledgements goes to Oyo State Bureau of Statistics for providing road accidents data for this study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] WHO, (2018). Global Nutrition Report. <https://globalnutritionreport.org/0b98fb>
- [2] Ogbodo, D., and Nduoma, E. (2011). FRSC: Nigerian Roads Second Worst in the World. *This Day Live*.

- [3] Sukhai, A., Andrew, P.J, Barnaby, S.L and Rohin H. (2011). Temporal Variations in Road Traffic Fatalities in South Africa. *Accident Analysis & Prevention*, 43(1), 421-428, doi.org/10.1016/j.aap.2010.09.012.
- [4] Sanusi, R. A., Adebola, F. B. and Adegoke, N. A. (2016). Cases of Road Traffic Accident in Nigeria: A Time Series Approach, *Mediterranean Journal of Social Sciences*. 7(2):542-552.
- [5] Odugbemi, O. O. (2010). Road Transportation and Tourism in Nigeria *Joja Press, Lagos*
- [6] Eze, B. (2012). Road Traffic Accidents in Nigeria: A Public Health Problem. *Afrimedical Journal*, 3(2):34-36
- [7] Agbonkhese, O., Yisa, G. L., Agbonkhese, E. G., Akanbi, D.O., Aka, E. O. and Mondigha, E. B. (2013). Road Traffic Accidents in Nigeria: Causes and Preventive Measures. *Civil and Environmental Research*, 3(13), 90-99.
- [8] Afolabi J. O. and Gbadamosi K. T. (2017). Road Traffic Crashes in Nigeria: Causes and Consequences. *International Journal of Shipping and Transport Logistics*, 17(42).
- [9] Chun, K., Goh, K., Currie G., Sarvi, M., and Logan, D. (2014). Bus Accident Analysis of Routes with/without Bus Priority. *Accident Analysis & Prevention*, 65,18-27.
- [10] Abayomi, A. A. (2005). Road Traffic Accidents in Nigeria: Cases, Cause and Control, *Osogbo, D & G International Limited*
- [11] Oyenuga, I.F., Ayoola, F.J. and Shittu, O.I. (2016). Statistical Analysis of Pattern on Monthly Reported Road Accidents in Nigeria *Science Journal of Applied Mathematics and Statistics*.4(4):119-128.
- [12] Iwok, I.A. (2016). Seasonal Modeling of Road Traffic Accident. *Mathematical Theory and Modeling*, 6(4):83-94.
- [13] Aluko, O.O. (2018). Injury Pattern Among Road Traffic Accident Victims: Ado-Ekiti Case Study, *Global Scientific Journals*, 6(4), 1-5.
- [14] Trivedi, A., and Rawal, D. (2011) Prevalence of Road Traffic Accidents and Driving Practices Among Young Drivers. *Healthline*. 2(2):72-75.
- [15] Abdulkabir, M., Tunde, R.S. and Edem, U.A (2015). Trend Analysis on Road Traffic Accident in Nigeria. *Science Innovation*, 3(5), 52-57.
- [16] Fortunato, S., (2010) Community Detection in Graphs, *Phys. Rep.-Rev. Sec. Phys. Lett.*, 486, 75-174.
- [17] Newman, M. E. J. and Girvan, M. (2004). Finding and Evaluating Community Structure in Networks, *Phys. Rev. E.*, 69,026113.
- [18] Emenike, G.C and Kanu, C.A (2017). Driver's Distraction and Road Traffic Crashes in Port Harcourt Metropolis, Rivers State, Nigeria, *International Journal of New Technology and Research*, 3(4):01-07.
- [19] Macharia, W. M., Njeru, E. K., Muli-Musiime, F., and Nantulya, V. (2009). Severe Road Traffic Injuries in Kenya, Quality of Care and Access, *African Health Sciences (Pub Med, Web of Science)*, 9(2):118-124
- [20] Feldkircher, M. and Zeugner, S. (2022). Bayesian Model Averaging with BMS R Cran <http://bms.zeugner.eu/tutorials/bms.pdf>
- [21] Akanbi, O.B. and Oladoja, O.M. (2018), Bayesian Analysis of 2014 FIFA World Cup Matches Played and Goals Scored, *International Journal of Modern Mathematical Sciences, Florida, USA*, 16(1), 25 - 36.
- [22] Akanbi, O.B. and Oladoja, O.M. (2019), Application of A Modified g-parameter Prior in Bayesian Model Averaging to CO₂ Emissions in Nigeria, *Mathematical Theory and Modeling, IISTE*, 9(11), 57-71.
- [23] Obisesan, K.O. and Oladoja, O.M. (2022). On Normal Process of Diffusion Equation in Monitoring Carbon Monoxide Concentrations in Nigeria. *International Journal of Statistical Distributions and Applications*. 8(2)

THE APPLICATION OF QUEUEING THEORY FOR MAXIMIZING SYSTEM SIZE USING ENCOURAGED ARRIVAL

S. Immaculate, P. Rajendran*

•

Department of Mathematics, School of Advanced Sciences,
Vellore Institute of Technology, Vellore, India - 632014.

immaculate.s@vit.ac.in

Correspondence email: prajendran@vit.ac.in

Abstract

Customers are frequently drawn in by lucrative deals and discounts offered by businesses. These interested customers are referred to as encouraged arrivals. The major goal of this study is to evaluate the performance of the automobile assembly line in order to decrease waiting time by coordinating the activities at each workstation using stopwatch time study approach. The novelty of this research is to convert poisson arrival to encouraged arrival with some discounts (like., 10%, 20%). The queuing problem is represented by the notation $M/M/1: FCFS/\infty/\infty$ in Kendall's notation. It is a single channel, multi-server service with infinite system capacity and an infinite number of calling population. Data concerning the system's encouraged arrival and service distribution were established. These data were used to calculate the system performance parameter. The finding of the study was used to predict the system's performance and effectiveness and to make logical recommendations for possible future improvements. According to the results, it is possible to conclude that increasing the level of automation reduces part waiting time, decreasing the cost of waiting. When compared to the poisson arrival system, the size of the Markovian encouraged arrival queuing system is increased as shown in the table. Little's law is verified that system size and queue size is same as in length. Little's law is used to predicts lead time based on production rate and work-in-process. Here it is verified as shown in table.

Keywords: Encouraged arrival, Assembly Line, Queuing Analysis, Stopwatch time study, Steady state Solution.

1. Introduction

In public areas like post offices, banks and gas stations, waiting is a common occurrence. Not just people, but also machinery and moving vehicles at traffic lights experience the phenomena of waiting. Bottlenecks arise and assume the shape of lines when resources are limited and unable to fulfil demand. In an assembly line, to make a final product as soon as possible, pieces are systematically attached to a product utilizing the well-planned logistics. The sequential organization of employees, equipment or parts is the main goal of assembly lines. In flow-oriented production system, the productive units performing the operations are repeatedly linked to a service.

The work components are often conveyed along the line via, a transportation system such as a conveyor belt where they are delivered to stations one after another. The different types of assembly lines include single-model, batch-model and mixed-model lines. A few product models are

manufactured in batches, one at a time on the same line with time allocated for a changeover so that the line is set up for the production of a new system. The procedure for choosing the configuration of the goods to be produced on the line must minimize the number of workstations, balance the delay and fulfil other placement requirements such as production rate, variety, minimal distance moved, division of labor and quality.

Introduction to congestion theory in telephone systems was discussed in [1] effective implementation of cycle time reduction strategies. For semiconductor back-end manufacturing considered in [2]. The goal of traditional assembly line balancing procedures is to get us to the point of subdividing work so that the amount of time that stations are out of balance is kept to a minimum is studied in [3]. Queuing analysis to analyses patient load in outpatient and inpatient services to facilitate more realistic resource planning is found in [4]. The Application of queuing theory in multi-stage production line is studied in [5]. Discussion of operational transport analysis methods and the practical application of queuing theory to stationary traffic considered in [6]. Modelling and analysis of manufacturing systems considered in [7]. Queuing theory in solving automobile assembly line problems in [8]. Improving effectiveness and efficiency of assembly line with a stopwatch time study and balancing activity elements was discussed in [9]. Queuing theory and manufacturing systems modelling and analysis are done and they are developed. Few among them were found in [10 and 11]. Parallel tasks and stations are considered by Bard (1989), as is dead time, which is the time required for transporting workpieces from one station to the next while no tasks can be executed.

Maximization of system size in solving automobile assembly line problem using encouraged arrival proposed in this work. An introduction is described in Section 1. The Markovian queue with encouraged arrival for mathematical model formulation is described in Section 2. Numerical illustrations are provided in Section 3. Results and discussion are given in Section 4. Section 5 contains the Conclusion.

2. Mathematical Model

The mathematical model predicates to satisfy the following conditions:

- i) Customers arrive one by one to an encouraged arrival discipline process with rate $\lambda(1 + \chi)$, where χ represents the customer's previous or observed data. If a previous firm gave discounts and percentages, the number of consumers observed values ranging from $\chi = 0.1$ and $\chi = 0.2$ respectively.
- ii) Service time is symmetrically and exponentially distributed.
- iii) Customers adhere to the first in, first out principle.

2.1 Steady State Solution:

We obtain the following system of differential difference equations.

$$\frac{d}{dt}P_0(t) = -\lambda(1 + \chi)P_0(t) + \mu P_1(t) \quad (1)$$

$$\frac{d}{dt}P_n(t) = \lambda(1 + \chi)P_{n-1}(t) - \{\lambda(1 + \chi) + \mu\}P_n(t) + \mu P_{n+1}(t) \quad n \geq 1 \quad (2)$$

In the steady state, as $t \rightarrow \infty$, $P_n(t) = P_n$ and therefore $P_n^1(t) = 0$ as $t \rightarrow \infty$ then, the equations are,

$$0 = -\lambda(1 + \chi)P_0 + \mu P_1 \quad (3)$$

$$0 = \lambda(1 + \chi)P_{n-1} - \{\lambda(1 + \chi) + \mu\}P_n + \mu P_{n+1} \quad (4)$$

Now the value of P_n is obtained as, $P_n = \rho^n P_0$

$$P_n = \left(\frac{\lambda(1+\chi)}{\mu} \right)^n P_0 \quad (5)$$

The value P_0 can be computed by using the obvious requirement, that the sum of all probabilities must be equal to 1.

$$\begin{aligned} \sum_{n=0}^{\infty} P_n &= P_0 \sum_{n=0}^{\infty} \rho^n \\ &= P_0 \frac{1}{1-\rho} \\ \sum_{n=0}^{\infty} P_n &= 1 \end{aligned} \quad (6)$$

Where, $P_0 = 1 - \rho$

It is clear that the traffic rate ρ must be less than 1, otherwise the sum of probabilities would not be 1 (not even limited). From (6) to (5) gives the general formula for P_n :

$$\begin{aligned} P_n &= \rho^n (1 - \rho) \\ &= \left(\frac{\lambda(1+\chi)}{\mu} \right)^n \\ P_n &= \left(1 - \frac{\lambda(1+\chi)}{\mu} \right) \end{aligned} \quad (7)$$

The equation (7) represent a very important result used to obtain all the characteristics of the M/M/1 system.

3. Numerical Illustration

The performance of the M/M/1 queueing system is analysed numerically concerning the parameters Values $\lambda(1 + \chi)$, μ , χ represent discounts values 10% and 20% of the table and figure. The following table 1 displays the parameters for the number of customers in the queueing system for various values of $\lambda(1 + \chi)$, μ .

Labor level

An important component to take into account when conducting an assessment study of an automobile assembly plant is the number of personnel engaged in productive operations on the assembly line. Depending on how much automation is used throughout the line, different people are required. The encouraged arrival method is more efficient than the Poisson process.

Output of the Problem Evaluation Analysis

The Estimation Analysis was done by calculating the system performance parameters such as idle system, length in system, length in queue, waiting time in system, waiting time in queue and system utilization.

Table 1: Encouraged arrival 10% and utilization factor for Markovian model

Work Station	χ	Encouraged Arrival rate $\lambda(1 + \chi)/\text{Min}$	Mean Service Time (μ)/min	Utilization factor $\frac{\lambda(1+\chi)}{\mu}$
1	0.1	12.1	16.9	0.72
2	0.1	14.3	17.8	0.80
3	0.1	13.2	17.8	0.74
4	0.1	11.0	16.4	0.67
5	0.1	11.0	18.1	0.61
6	0.1	14.3	19.2	0.74
7	0.1	12.1	17.2	0.70
8	0.1	11.0	19.8	0.56
9	0.1	13.2	17.6	0.75
10	0.1	11.0	16.9	0.65
11	0.1	14.3	18.1	0.79
12	0.1	11.0	17.8	0.62
13	0.1	12.1	17.1	0.71
14	0.1	14.3	18.9	0.76
15	0.1	13.2	18.6	0.71
16	0.1	15.4	20.0	0.77

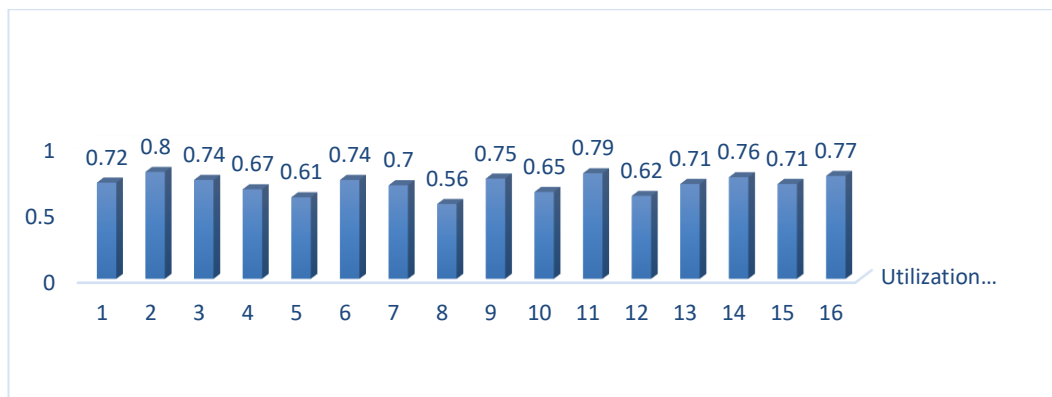


Figure 1: Usage of workstation graph

The utilization factor for each workstation is displayed in **Figure 1**. It is clear from the graph that certain workstations are operating below capacity. This might be as a result of the low amount of automation or labor at such a station.

3.1 Encouraged Arrival 10% of the Queuing System:

In station 1:

$$\lambda(1 + \chi) = 11(1 + 0.1) = 12.1 \text{ parts/min, } \mu = 16.9 \text{ parts/min}$$

(i) Traffic intensity $\rho = \frac{(\lambda(1+\chi))}{\mu} = \frac{12.1}{16.9} = 0.72$

- (ii) Idle system = $1 - \rho = 1 - 0.72 = 0.28$
- (iii) Length in system $L_s = \frac{(\lambda(1+\chi))}{\mu-\lambda(1+\chi)} = \frac{12.1}{16.9-12.1} = \frac{12.1}{4.8} = 2.52$
- (iv) Length in queue $L_q = \frac{(\lambda(1+\chi))^2}{\mu(\mu-\lambda(1+\chi))} = \frac{(12.1)^2}{16.9(4.8)} = \frac{146.41}{81.12} = 1.81$
- (v) Waiting time in system $w_s = \frac{1}{\mu-\lambda(1+\chi)} = \frac{1}{16.9-12.1} = \frac{1}{4.8} = 0.21 \times 60 = 12.6$ sec.
- (vi) Waiting time in queue $w_q = \frac{(\lambda(1+\chi))}{\mu(\mu-\lambda(1+\chi))} = \frac{12.1}{16.9(4.8)} = \frac{12.1}{81.12} = 0.14 \times 60 = 8.4$ sec

Table 2: Encouraged arrival for 10% discounts and utilization factor for a Markovian queuing.

Stations	χ	$\lambda(1 + \chi)$	μ	ρ	$1-\rho$	L_s	L_q	W_s	W_q
1	0.1	12.1	16.9	0.72	0.28	2.52	1.81	12.6	8.4
2	0.1	14.3	17.8	0.80	0.20	4.09	3.28	17.4	13.8
3	0.1	13.2	17.8	0.74	0.26	2.87	2.13	13.2	9.6
4	0.1	11.0	16.4	0.67	0.33	2.04	1.37	11.4	7.2
5	0.1	11.0	18.1	0.61	0.39	1.55	0.94	8.4	5.4
6	0.1	14.3	19.2	0.74	0.26	2.92	2.17	14.4	9.0
7	0.1	12.1	17.2	0.70	0.30	2.37	1.67	11.4	8.4
8	0.1	11.0	19.8	0.56	0.44	1.25	0.69	6.6	3.6
9	0.1	13.2	17.6	0.75	0.25	3.0	2.25	13.8	10.2
10	0.1	11.0	16.9	0.65	0.35	1.86	1.21	10.20	6.6
11	0.1	14.3	18.1	0.79	0.21	3.76	2.97	15.6	12.6
12	0.1	11.0	17.8	0.62	0.38	1.62	0.99	9.0	5.4
13	0.1	12.1	17.1	0.71	0.29	2.42	1.71	12.0	8.4
14	0.1	14.3	18.9	0.76	0.24	3.11	2.35	13.2	9.6
15	0.1	13.2	18.6	0.71	0.29	2.44	1.74	11.4	7.80
16	0.1	15.4	20.0	0.77	0.23	3.35	2.58	13.2	10.2

In M/M/1 automobile assembly line problem with 10% encouraged arrival for length in system and queue as well as waiting time in system and queue.

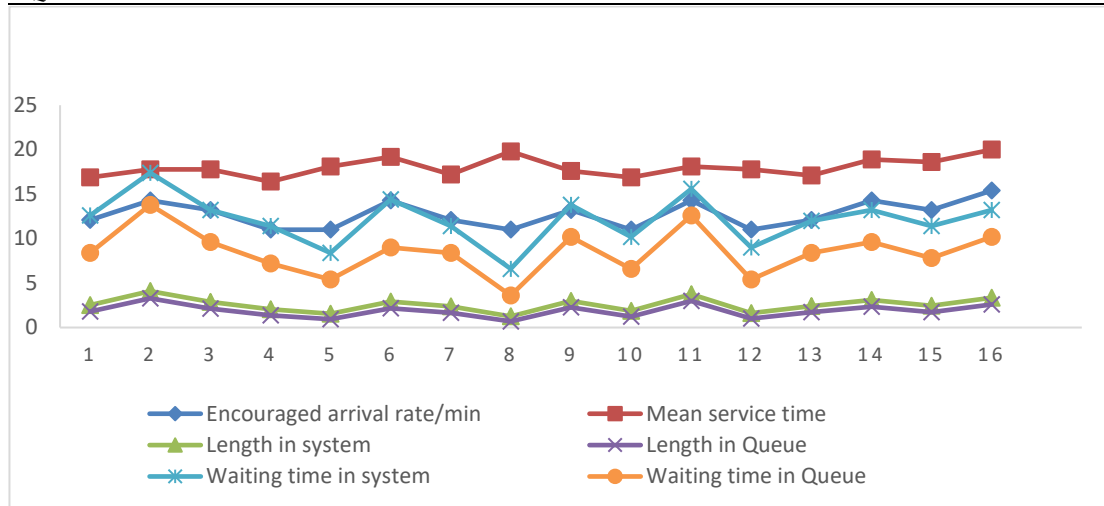


Figure 2: Encouraged arrival of 10% discounts system and queue Size.

Table 3: Encouraged Arrival 10% discounts comparing with the Poisson Arrival (PA) model.

Stations	(λ)	$\lambda(1 + \chi)$	PA (L_s)	EA (L_s)	PA (L_q)	EA (L_q)	PA (W_s)	EA (W_s)	PA (W_q)	EA (W_q)
1	11	12.1	1.86	2.52	1.21	1.81	10.2	12.6	6.62	8.4
2	13	14.3	2.71	4.09	1.98	3.28	12.5	17.4	9.13	13.8
3	12	13.2	2.07	2.87	1.39	2.13	10.3	13.2	6.97	9.6
4	10	11.0	1.56	2.04	0.95	1.37	9.4	11.4	5.72	7.2
5	10	11.0	1.23	1.55	0.68	0.94	7.4	8.4	4.09	5.4
6	13	14.3	2.10	2.92	1.42	2.17	9.7	14.4	6.55	9.0
7	11	12.1	1.77	2.37	1.12	1.67	9.7	11.4	6.19	8.4
8	10	11.0	1.02	1.25	0.52	0.69	6.1	6.6	3.09	3.6
9	12	13.2	2.14	3.0	1.46	2.25	10.7	13.8	7.31	10.2
10	10	11.0	1.45	1.86	0.86	1.21	8.7	10.20	5.15	6.6
11	13	14.3	2.55	3.76	1.83	2.97	11.8	15.6	8.45	12.6
12	10	11.0	1.28	1.62	0.72	0.99	7.7	9.0	4.32	5.4
13	11	12.1	1.80	2.42	1.16	1.71	9.8	12.0	6.33	8.4
14	13	14.3	2.20	3.11	1.51	2.35	10.2	13.2	6.99	9.6
15	12	13.2	1.82	2.44	1.17	1.74	9.1	11.4	5.87	7.80
16	14	15.4	2.33	3.35	1.63	2.58	10.0	13.2	7.00	10.2

Table 3 demonstrate that measuring each station has a substantial influence on the efficacy and efficiency of production operations in encouraged arrival by increasing system size and waiting time when compared to the poisson arrival for 10% discount by utilizing stopwatch time study methods.

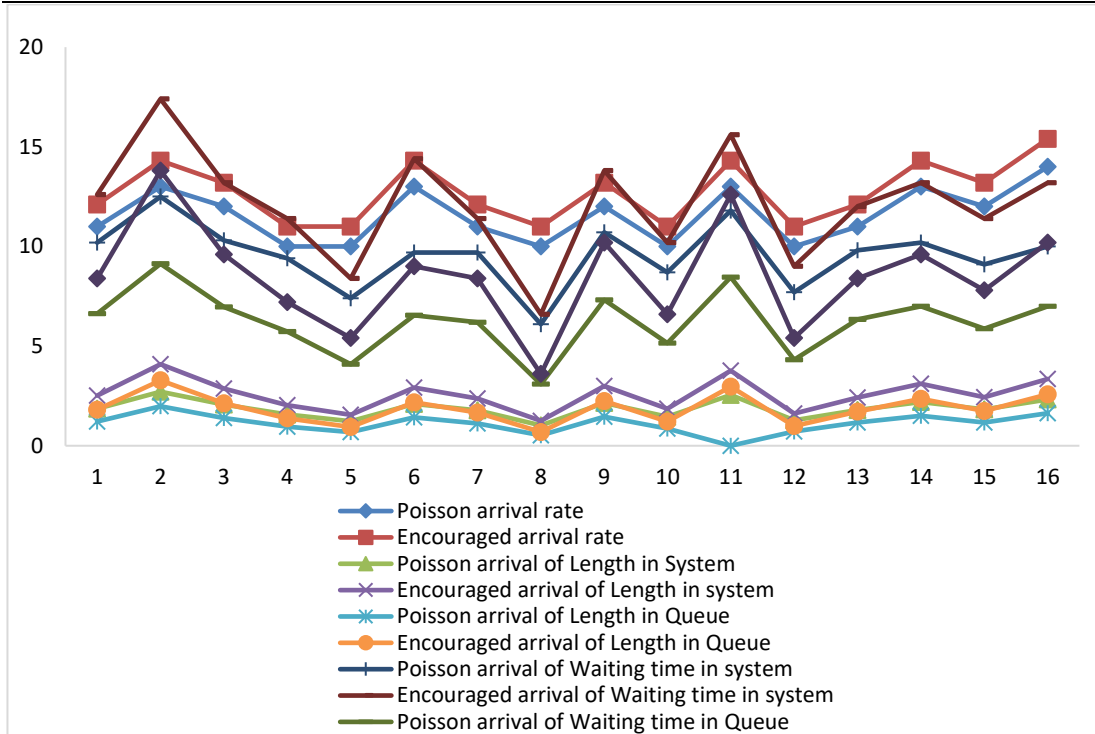


Figure 3: Comparison of poisson arrival to encouraged arrival in systems and queues for 10% discounts.

Table 4: Verification of Little's law

Stations	$\lambda(1 + \chi)$	L_q	L_s	W_q	W_s	$L_q = \lambda(1 + \chi)W_q$	$L_s = \lambda(1 + \chi) W_s$
1	12.1	1.81	2.52	0.14	0.21	1.7	2.54
2	14.3	3.28	4.09	0.23	0.29	3.28	4.14
3	13.2	2.13	2.87	0.16	0.22	2.11	2.90
4	11.0	1.37	2.04	0.12	0.19	1.32	2.09
5	11.0	0.94	1.55	0.09	0.14	0.99	1.54
6	14.3	2.17	2.92	0.15	0.204	2.15	2.92
7	12.1	1.67	2.37	0.14	0.19	1.69	2.29
8	11.0	0.69	1.25	0.06	0.11	0.66	1.21
9	13.2	2.25	3.0	0.17	0.23	2.24	3.04
10	11.0	1.21	1.86	0.11	0.17	1.21	1.87
11	14.3	2.97	3.76	0.21	0.26	3.0	3.72
12	11.0	0.99	1.62	0.09	0.15	0.99	1.65
13	12.1	1.71	2.42	0.14	0.20	1.7	2.42
14	14.3	2.35	3.11	0.16	0.22	2.28	3.15
15	13.2	1.74	2.44	0.13	0.19	1.72	2.51
16	15.4	2.58	3.35	0.17	0.22	2.62	3.38

3.2 Encouraged arrival 20% of the queuing system:

Table 5: Encouraged arrival for 20% discounts and utilization factor for a Markovian model.

Stations	χ	$\lambda(1+\chi)$	(μ)	(ρ)	$(1-\rho)$	(L_s)	(L_q)	(W_s)	(W_q)
1	0.2	13.2	16.9	0.78	0.22	3.57	2.79	16.2	12.6
2	0.2	15.6	17.8	0.87	0.13	7.09	0.98	27.0	23.4
3	0.2	14.4	17.8	0.80	0.20	4.24	3.43	17.4	14.4
4	0.2	12.0	16.4	0.73	0.27	2.73	1.99	13.8	10.20
5	0.2	12.0	18.1	0.66	0.34	1.97	1.30	9.6	6.6
6	0.2	15.6	19.2	0.81	0.19	4.33	3.52	16.8	15.6
7	0.2	13.2	17.2	0.77	0.23	3.3	2.53	15.0	11.4
8	0.2	12.0	19.8	0.61	0.39	1.54	0.93	7.8	4.8
9	0.2	14.4	17.6	0.82	0.18	4.5	3.68	13.2	15.6
10	0.2	12.0	16.9	0.71	0.29	2.45	1.74	12.0	8.4
11	0.2	15.6	18.1	0.86	0.14	6.24	5.38	24.0	20.4
12	0.2	12.0	17.8	0.67	0.33	2.07	1.39	10.20	7.2
13	0.2	13.2	17.1	0.77	0.23	3.38	2.61	15.6	11.4
14	0.2	15.6	18.9	0.83	0.17	4.73	3.90	18.0	15.0
15	0.2	14.4	18.6	0.77	0.23	3.43	2.65	14.4	10.8
16	0.2	16.8	20.0	0.84	0.16	5.25	4.41	18.6	15.6

In M/M/1 automobile assembly line problem using the stopwatch time study approach method with 20% encouraged arrival for length in system and queue, as well as waiting time in system and queue.

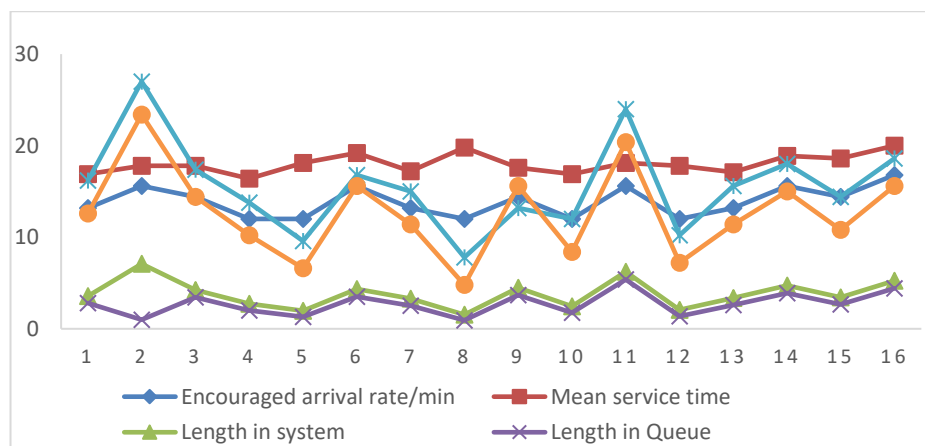


Figure 4: Shows that the encouraged arrival of 20% discounts systems and queues size.

The (table 6) demonstrate that measuring each station has a substantial influence on the efficacy and efficiency of production operations in encouraged arrival by increasing system size and waiting time

when compared to the poisson arrival for 20% discount by using stopwatch time study approach.

Table 6: Encouraged arrival 20% discounts model comparing with the Poisson arrival (PA) model.

Stations	(λ)	$\lambda(1 + \lambda)$	PA (L_s)	EA (L_s)	PA (L_q)	EA (L_q)	PA (W_s)	EA (W_s)	PA (W_q)	EA (W_q)
1	11	13.2	1.86	3.57	1.21	2.79	10.2	16.2	6.62	12.6
2	13	15.6	2.71	7.09	1.98	0.98	12.5	27.0	9.13	23.4
3	12	14.4	2.07	4.24	1.39	3.43	10.3	17.4	6.97	14.4
4	10	12.0	1.56	2.73	0.95	1.99	9.4	13.8	5.72	10.20
5	10	12.0	1.23	1.97	0.68	1.30	7.4	9.6	4.09	6.6
6	13	15.6	2.10	4.33	1.42	3.52	9.7	16.8	6.55	15.6
7	11	13.2	1.77	3.3	1.12	2.53	9.7	15.0	6.19	11.4
8	10	12.0	1.02	1.54	0.52	0.93	6.1	7.8	3.09	4.8
9	12	14.4	2.14	4.5	1.46	3.68	10.7	13.2	7.31	15.6
10	10	12.0	1.45	2.45	0.86	1.74	8.7	12.0	5.15	8.4
11	13	15.6	2.55	6.24	1.83	5.38	11.8	24.0	8.45	20.4
12	10	12.0	1.28	2.07	0.72	1.39	7.7	10.20	4.32	7.2
13	11	13.2	1.80	3.38	1.16	2.61	9.8	15.6	6.33	11.4
14	13	15.6	2.20	4.73	1.51	3.90	10.2	18.0	6.99	15.0
15	12	14.4	1.82	3.43	1.17	2.65	9.1	14.4	5.87	10.8
16	14	16.8	2.33	5.25	1.63	4.41	10.0	18.6	7.00	15.6

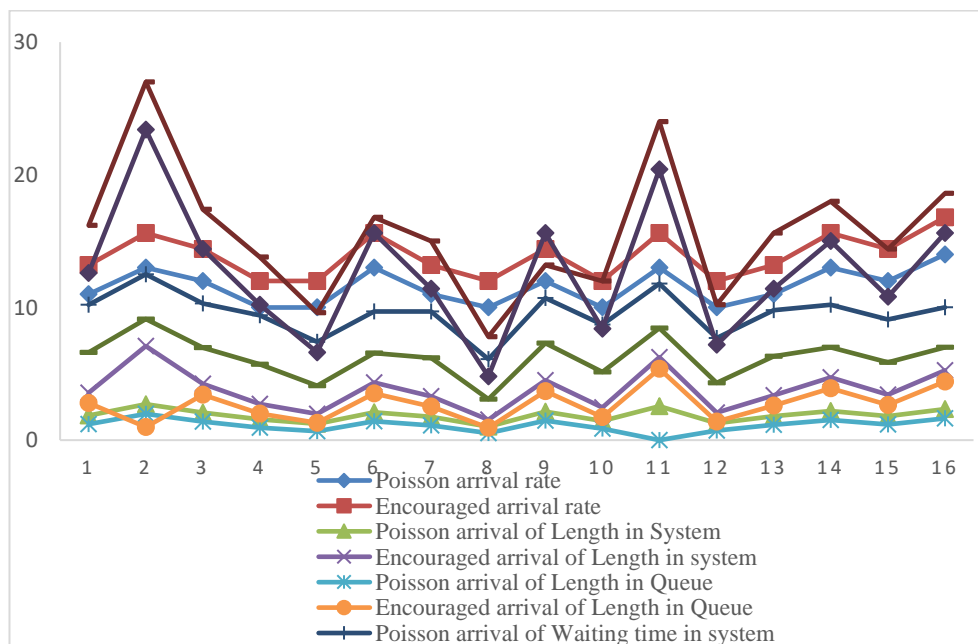


Figure 5: Comparison of poisson arrival to encouraged arrival in systems and queues for 20% discounts.

Table 7: Verification of Little's law:

Station	$\lambda(1 + \chi)$	L_q	L_s	W_q	W_s	$L_q = \lambda(1 + \chi)W_q$	$L_s = \lambda(1 + \chi) W_s$
1	13.2	2.79	3.57	0.21	0.27	2.77	3.56
2	15.6	0.98	7.09	0.39	0.45	7.02	7.02
3	14.4	3.43	4.24	0.24	0.29	3.46	4.18
4	12.0	1.99	2.73	0.17	0.23	2.04	2.76
5	12.0	1.30	1.97	0.11	0.16	1.92	1.92
6	15.6	3.52	4.33	0.26	0.28	4.06	4.37
7	13.2	2.53	3.3	0.19	0.25	2.51	3.3
8	12.0	0.93	1.54	0.08	0.13	0.96	1.56
9	14.4	3.68	4.5	0.26	0.31	3.74	4.5
10	12.0	1.74	2.45	0.14	0.20	1.7	2.4
11	15.6	5.38	6.24	0.34	0.4	5.30	6.24
12	12.0	1.39	2.07	0.12	0.17	1.44	2.04
13	13.2	2.61	3.38	0.19	0.26	2.51	3.43
14	15.6	3.90	4.73	0.25	0.30	3.9	4.7
15	14.4	2.65	3.43	0.18	0.24	2.6	3.46
16	16.8	4.41	5.25	0.26	0.31	4.37	5.21

4. Results and Discussions

This research provides recommendation to improve the standard of service to be more effective and efficient. The stopwatch time study approach in [9] was used to determine the average service time and distribution of encouraged arrival rates for each workstation. The degree of variation in the materials or components used to make the cars, as well as the manufacturing processes, affect service times. The timelines for the vehicle under consideration's arrival and assembly are summarized in the table. In comparison to the Poisson arrival system, the size of the Markovian encouraged arrival queuing system increased.

5. Conclusion

Based on the findings, the queuing problem in an automobile assembly line to improve the standard of service to be more effective and efficient. The encouraged arrival and service distribution data for the system were determined. These data were used to calculate the system performance parameter. The result shows that increasing automation will lead to quicker processing times for parts and lower lead costs. The company will benefit greatly from this study because it will make it easier for management to plan future production by providing them with all the data pertaining to the performance of the company's assembly line.

References

- [1] Syski, R. (1986) Introduction to Congestion Theory in Telephone Systems, 2nd Edition, Elsevier Science Publisher, Amsterdam.
- [2] Domaschke, J. and Brown, S. (1998) Effective implementation of cycle time reduction, proceeding of the 1998 Winter Simulation Conference, USA.
- [3] Sury, R.J. (1971) Aspects of assembly line balancing: *International Journal of Production Research*, Vol. 9 Issue 4, pp.501-512.
- [4] Mital, K.M. (2010) Queuing analysis for outpatient and inpatient services: a case study, *Management Decision* Vol.48 No. 3, pp. 419 – 439.
- [5] Muhammad Marsudi and Hani Shafeek (2014) The Application of Queuing Theory in Multi-Stage Production Line, *Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management*, Indonesia, pp 668 – 675.
- [6] Silver, E.A., Pyke, D and Peterson, R. (1998) *Inventory Management and Production Planning and Scheduling* (Third Edition): Wiley, New York.
- [7] Papadopolous H.T., Heavy, C. and Browne, J. (2013) *Queuing Theory in Manufacturing System Analysis and Design*, Springer Science & Business Media.
- [8] A.M. Adeyinka, B.Kareem (2018) Application of queuing theory in solving assembly line problem, *International journal of engineering research and technology* Vol. 7 Issue 06, pp.344-351.
- [9] Budiman, I., Sembiring, A. C., Tampubolon, J., Wahyuni, D., & Dharmala, A. (2019, July). Improving effectiveness and efficiency of assembly line with a stopwatch time study and balancing activity elements. In *Journal of Physics: Conference series* (Vol. 1230, No. 1, p.012041). IOP Publishing.
- [10] Gross, D. and Harris, C.M. (1985) *Fundamentals of Queuing Theory*, Wiley, New York.
- [11] Prabhu, N.U. (1994) *Foundations of Queuing Theory*: Ithaco, New York: Khuwar Academic Publisher.
- [12] Adam, E. E., & Ebert, R. J. (1992). *Production and Operation Management; Concepts, Models, and Behavior* (fifth edition ed.). Englewood Cliffs: Prentice Hall.
- [13] Banks J., Carson J.S, Nelson B. L. and Nicol D. M. (2000). *Discrete-Event System Simulation*, Prentice Hall, Upper Saddle River NJ. Bartholdi, J.J (1993). Balancing two-sided assembly lines: A case study. *International Journal of Production Research*. 31(10): 2447-2461.
- [14] Becker, C., & Scholl, A. (2006). A survey on problems and methods in generalized assembly line balancing. *European journal of operational research*, 168(3), 694-715.
- [15] Boysen, N., Fliedner, M., & Scholl, A. (2007). A classification of assembly line balancing problems. *European journal of operational research*, 183(2), 674-693.
- [16] Merengo, C., Nava, F., & Pozzetti, A. (1999). Balancing and sequencing manual mixed-model assembly lines. *International Journal of Production Research*, 37(12), 2835-2860.
- [17] Chakravarty, A. K., & Shtub, A. (1986). A cost minimization procedure for mixed model production lines with normally distributed task times. *European Journal of Operational Research*, 23(1), 25-36.
- [18] Baybars, I. (1986). A survey of exact algorithms for the simple assembly line balancing problem. *Management science*, 32(8), 909-932.
- [19] Bartholdi, J. J. (1993). Balancing two-sided assembly lines: a case study. *International Journal of Production Research*, 31(10), 2447-2461.
- [20] Bard, J. F. (1989). Assembly line balancing with parallel workstations and dead time. *The International Journal of Production Research*, 27(6), 1005-1018.

RELIABILITY ANALYSIS OF THIN -WALLED PRESSURE VESSELS USING ADVANCED FIRST ORDER SECOND MOMENT (AFOSM) METHOD

K. Sunitha, T. Sumathi Uma Maheshwari, M. Tirumala Devi

Department of Mathematics, Kakatiya University, Warangal, Telangana, India-506009
kadiresunitha79@gmail.com, tsumathiuma@gmail.com, oramdevi@yahoo.com

Abstract

Pressure vessels are highly used in commercial purposes and industries such as boiling, softening and hot water storing tanks. Pressure vessels are subjected to its internal and external pressure. In this paper, thin-walled pressure vessels are taken for analysis. Pressure, radius, thickness and strength of the material are considered as random variables. The random variables follow normal distribution. The reliability index of the pressure vessel made with different materials such as 6061 aluminum alloy and SA 516 70 stainless steel has been found. Reliability analysis has been done for the pressure vessel by using AFOSM with MATLAB. It is observed that strength of the materials influences more on reliability of the vessels.

Keywords: Thin-walled cylinder, reliability, stress, strength, normal distribution, pressure, thickness, inner radius of cylinder, mean, variance.

1. Introduction

Chemical Industry requires the handling and storing the large quantities of material such as liquids and gases in containers or vessels and therefore pressure vessels play an important role in the industry. Pressure vessels are leak-proof containers that store liquid or gas. Pressure vessels [1] of various sizes and shapes have been produced for different purposes, constructing a required pressure vessel is engineer's task. Engineer's aim is to design a high-level performance of a vessel. Reliability is the probability of a design, should satisfy certain needs under the prescribed environment for certain period. Reliability methods are used to get specified reliability of a structure. Finding reliability of the design is inescapable. In 1981, Cornell [2] called structural reliability is a healthy adolescent.

L. Cizelj et al. [3] applied First Order Reliability Method (FORM) and Second Order Reliability Method (SORM) in the safety assessment of steam generator tubes with through-wall axial stress corrosion cracks. Gunjan Agarwal and Baidurya Bhattacharya [4] studied partial safety factor design of rectangular partially prestressed concrete beams in ultimate flexural limit state. Antanas Kudzyis and Romualdas Kliukas [5] discussed the reliability index of the design of reinforced concrete structures of annular cross sections. P. Hari Prasad et al. [6] used FORM for the reliability analysis of a section of a structural beam and reliability index was found. Zheng Yulong et al. [7] used Hasofer-Lind method for thin -walled pressure vessels, compared with Second Moment Method and found that thickness of the pressure vessel was small. Devaraju A. and pazhanivel K. [8] studied

reliability of thin -walled pressure vessels with ANSYS software.

In this paper, thin-walled pressure vessel is considered for the analysis. Found reliability index of the pressure vessel which is made with different materials such as 6061 aluminum alloy and SA 516 70 stainless steel. Reliability analysis has been done for the pressure vessel by using AFOSM with MATLAB.

2. Methodology

If the system has a deterministic strength X , and a randomly developed stress Y then the reliability of the system is the probability that the resistance is greater than the stress, $P(X > Y)$. Failure probability is the probability that the stress is greater than the strength $P(Y > X)$. In the specific case of a Gaussian random variable, the stress Y can be reduced into standard normal variable y ,

$$Y = \mu_Y + y\sigma_Y \Rightarrow y = \frac{Y - \mu_Y}{\sigma_Y} \quad (1)$$

where μ_Y is the mean of Y and σ_Y is the standard deviation of Y , then the reliability of the system

$$\begin{aligned} R &= P(X > \mu_Y + y\sigma_Y) \\ R &= P\left(\frac{X - \mu_Y}{\sigma_Y} > y\right) \\ R &= P(\beta > y) \end{aligned} \quad (2)$$

where β is the reliability index, in the single variable case, this inequality is a safe region. This is the set of values of y for which the structure will not fail. The probability of failure is the complement of the reliability.

$$P_f = 1 - R = 1 - P(\beta > y) \quad (3)$$

Let $f_X(x)$ and $f_Y(y)$ are the probability density functions of strength X and stress Y . Then the distribution function F is

$$F_X(y) = P(Y \leq y) = \int_{-\infty}^y f_X(u) du \quad (4)$$

The probability of failure becomes

$$\begin{aligned} P_f &= P(Y > X) = \int_0^{\infty} F_X(y) f_Y(y) dy \\ &= \int_0^{\infty} \int_{-\infty}^y f_X(u) f_Y(y) du dy \end{aligned} \quad (5)$$

2.1. First Order Reliability Method

Let $(X_1, X_2, X_3, \dots, X_n)$ be the set of random variables (structural design variables). The limit state equation for the failure surface of the structure is

$$g(X_1, X_2, X_3, \dots, X_n) = 0 \quad (6)$$

Collapse of the structure or failure is defined by the failure condition as $g(X_1, X_2, X_3, \dots, X_n) < 0$. Probability of failure is $P_f = P(g(X_1, X_2, X_3, \dots, X_n) < 0)$ and reliability is $P(g(X_1, X_2, X_3, \dots, X_n) > 0)$. Methods for the determination of this probability depends on the complexity of the limit state function. The limit state function is the limit at which the performance transits from acceptable to

unacceptable.

There are different types of limit states: Ultimate limit state, serviceability limit state, etc. The limit state function is

$$g = X - Y \quad (7)$$

In the above expression, X and Y are random variables. X is the strength and Y is the stress developed in the structure. If $g < 0$ it leads to breakage of the structure. i.e., failure, and if $g > 0$ then the structure is safe.

If the random variables C, S for a linear performance function described by the equation (7) are normally distributed then C, S can be reduced as

$$C' = \frac{(C - \mu_C)}{\sigma_C} \text{ and } S' = \frac{(S - \mu_S)}{\sigma_S} \quad (8)$$

$$g = C - S = (C' \sigma_C + \mu_C) - (S' \sigma_S + \mu_S) = (\mu_C - \mu_S) + C' \sigma_C - S' \sigma_S \quad (9)$$

The line in reliability analysis is the line corresponding to $g(C', S') = 0$ because this line separates the safe and failure region in the standardized space. From the above, the reliability index β is the shortest distance from origin of reduced variables to the line of $g(C', S') = 0$

$$\beta = \frac{(\mu_C - \mu_S)}{\sqrt{\sigma_C^2 + \sigma_S^2}} \quad (10)$$

where μ_C and μ_S are the mean values of C and S respectively; σ_C^2 and σ_S^2 are their variance values. If the random variables C and S have the log-normal distribution, then the reliability index is given by

$$\beta = \frac{(\overline{W_C} - \overline{W_S})}{\sqrt{\sigma_{W_C}^2 + \sigma_{W_S}^2}} \quad (11)$$

where $W_C = \log C$, and $W_S = \log S$, $\overline{W_C}$, $\overline{W_S}$ are the mean values of W_C and W_S ; $\sigma_{W_C}^2$ and $\sigma_{W_S}^2$ are their variance values.

If the limit state surface is linear, then the Hasofer -Lind reliability index coincides with the reliability index computed from FOSM. However, Hasofer -Lind reliability method is used for non-linear limit state surfaces. If the limit state line is closer to the origin in the standardized coordinate system, the failure region is larger, and if it is farther away from the origin, the failure region is smaller. Thus, the position of the limit state surface relative to the origin in the standardized coordinate system is a measure of the reliability of the system. Then the probability of failure is

$$P_f = P((C - S) < 0) = P(g < 0) \quad (12)$$

$$P_f = \Phi(-\beta) \Rightarrow \beta = -\Phi^{-1}(P_f) \quad (13)$$

and reliability $R = 1 - P_f$, where Φ and Φ^{-1} are the cumulative distribution function and its inverse.

2.2. Advanced First Order Second Moment Method (AFOSM)

In this method, the assessment of the reliability index is based on the reduction of the limit state function to the standardized coordinate system Thus, the random variables X_i , which are normally distributed and are reduced as

$$X'_i = \frac{(X_i - \mu_{X_i})}{\sigma_{X_i}}, \quad (i = 1, 2 \dots n) \quad (14)$$

where X'_i is a standardized random variable with zero mean and unit standard deviation, Thus, Eq. (8) is used to transform the original limit state surface $g(X) = 0$ into a reduced limit state surface $g(X') = 0$. Where, X denotes 'original coordinate system'. In the standardized coordinate system, Hasofer- Lind reliability index β is equal to the minimum distance from the origin to the limit state surface

$$\beta = \sqrt{(x^{*'})^T(x^{*'})} \quad (15)$$

The minimum distance point on the limit state surface is called the design point. It is denoted by vector x^* in the original coordinate system and by vector $x^{*'}$ in the reduced coordinate system. These vectors represent the values of the random variables. For the general case of a non-linear limit state surface, the assessment of the minimum distance can be written as an optimization problem

$$\begin{aligned} \text{Minimize } D &= \sqrt{x'^T x'} \\ \text{Subject to } g(X') &= 0. \end{aligned} \quad (16)$$

By using Lagrange's multipliers, the minimum distance (for n variables) could be estimated as

$$\beta = - \frac{\sum_{i=1}^n x_i^{*' } \left(\frac{\partial g}{\partial x_i'} \right)^*}{\sqrt{\sum_{i=1}^n \left(\frac{\partial g}{\partial x_i'} \right)^{*2}}} \quad (17)$$

where $\left(\frac{\partial g}{\partial x_i'} \right)^*$ is the i^{th} partial derivative evaluated at the design point $(x_1^{*' }, x_2^{*' }, x_3^{*' }, \dots, x_n^{*' })$. The design point in the reduced coordinates is

$$x_i^{*' } = -\alpha_i \beta \quad (i = 1, 2, \dots, n) \quad (18)$$

where α_i are the direction cosines along the coordinate axes X_i' . They are given by

$$\alpha_i = \frac{\left(\frac{\partial g}{\partial x_i'} \right)^*}{\sqrt{\sum_{i=1}^n \left(\frac{\partial g}{\partial x_i'} \right)^{*2}}} \quad (19)$$

3. Results

3.1. Computation of reliability for thin-walled cylindrical shell pressure vessel

Thickness is very important in the design for the safety of the cylinder. Thickness of the cylinders is 20th part of diameter or even less. It is assumed that thin - walled cylindrical pressure vessel with SA516-70 stainless steel, and 6061 aluminum alloy. Yield strength, radius of the cylinder, pressure, thickness of the cylinder, are taken as random variables which follow normal distribution. Let $(\mu_x, \mu_r, \mu_p, \mu_t)$, $(\sigma_x, \sigma_r, \sigma_p, \sigma_t)$ be the mean and standard deviation design vector of the random variables.

Table 1: Design variables and materials properties

	SA51670stainless-steel		6061aluminumalloy	
	Mean	Standard deviation	Mean	Standard deviation
Yield strength (MPa)	335	16.56	276	16.56
Radius of cylinder(mm)	2000	100	2000	100
Inside pressure(MPa)	5	0.4	5	0.4
Thickness (mm)	50	2	50	2
Joint efficiency			0.85	

Let the thin cylindrical pressure vessel limit state function is [10]

$$g = X - Y = X - \frac{p(r-(0.6)t)}{Et} \quad (20)$$

where X is the strength of the cylinder, Y is stress of the cylinder, p is pressure in the cylinder, r is radius, t is thickness of the cylinder, E is joint efficiency. Therefore, the equation of failure surface is given by

$$g = X - \frac{p(r-(0.6)t)}{Et} = 0 \quad (21)$$

using eqn. (8), the failure surface function eqn. (20) in the standardized coordinate system is given by

$$g = (X'\sigma_X + \mu_X) - \frac{(p'\sigma_p + \mu_p)(r'\sigma_r + \mu_r - (0.6)(t'\sigma_t + \mu_t))}{E(t'\sigma_1 + \mu_t)} = 0 \quad (22)$$

where X', p', r', t' are reduced random variables.

Since $X' = \beta\alpha_1, p' = \beta\alpha_2, r' = \beta\alpha_3, t' = \beta\alpha_4$ then (23)

$$(\beta\alpha_1\sigma_X + \mu_X) - \frac{(\beta\alpha_2\sigma_p + \mu_p)(\beta\alpha_3\sigma_r + \mu_r - (0.6)(\beta\alpha_4\sigma_t + \mu_t))}{E(\beta\alpha_4\sigma_1 + \mu_t)} = 0 \quad (24)$$

$$(\beta\alpha_1\sigma_X + \mu_X)E(\beta\alpha_4\sigma_1 + \mu_t) - (\beta\alpha_2\sigma_p + \mu_p)((\beta\alpha_3\sigma_r + \mu_r) - (0.6)(\beta\alpha_4\sigma_t + \mu_t)) = 0 \quad (25)$$

$$\beta^2(E\alpha_1\alpha_3\sigma_X\sigma_t - \alpha_2\alpha_4\sigma_p\sigma_r + (0.6)\alpha_3\alpha_4\sigma_p\sigma_t) - \beta(E\alpha_1\sigma_X\mu_t + E\alpha_3\mu_X\mu_t - \alpha_4\sigma_p\mu_r - \alpha_2\mu_p\sigma_r + (0.6)\alpha_4\sigma_p\mu_t + (0.6)\alpha_3\mu_p\sigma_t) + (E\mu_X\mu_t - \mu_p\mu_r + (0.6)\mu_p\mu_t) = 0 \quad (26)$$

3.1. Mean thickness vs reliability

It is observed from Table 2 that if thickness of SA 516 70 Stainless steel increases from 50 mm to 140 mm, then reliability increases from 0.7882 to 0.9998 where as in 6061 aluminum alloy, the reliability of the material increases from 0.6627 to 0.9985. The reason for this is, with the increase of thickness, the stress is lowered at a given pressure.

Table 2: Mean thickness vs reliability

μ_t	SA51670 stainless-steel			6061aluminumalloy		
	β	P_f	R_1	β	P_f	R_2
50	0.8	0.2118	0.7882	0.42	0.3372	0.6627
60	1.29	0.0985	0.9014	0.85	0.1976	0.8023
70	1.68	0.0464	0.9535	1.21	0.1131	0.8868
82	2.03	0.0211	0.9788	1.51	0.0655	0.9344
100	2.56	0.0052	0.9947	2.05	0.0201	0.9798
120	3.08	0.0019	0.9989	2.53	0.0057	0.9942
140	3.58	0.0002	0.9998	2.97	0.0014	0.9985

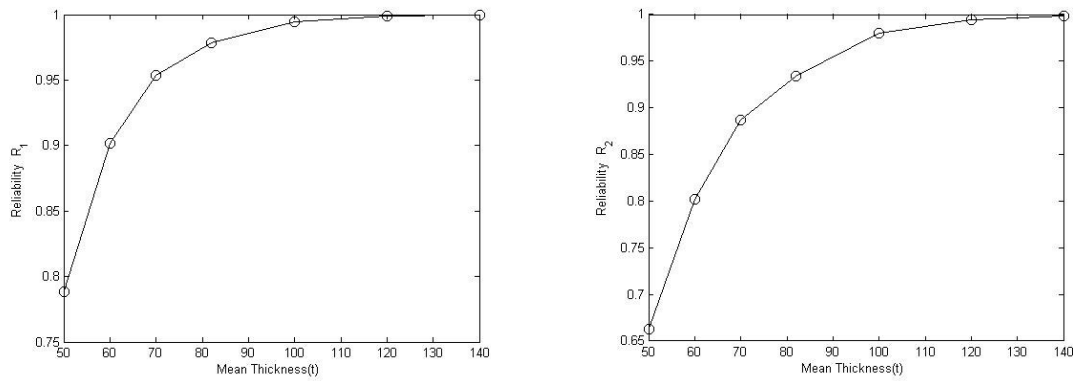


Figure 1: Variation of reliability as a function of mean thickness

3.2. Mean pressure vs reliability

It is observed from Table 3 that if μ_p increases from 1.9 MPa to 5 MPa of SA 516 70 stainless steel, then reliability decreases from 0.9996 to 0.7882 and the reliability of 6061 aluminum alloy decreases from 0.9971 to 0.6627. By the increase of pressure, the stress will be developed in the cylinder and the increment in the stress causes low.

Table 3: Mean pressure vs reliability

SA 51670 stainless steel				6061 aluminum alloy		
μ_p	β	P_f	R_1	β	P_f	R_2
1.9	3.35	0.0004	0.9996	2.77	0.0043	0.9971
2.5	3.19	0.0007	0.9993	2.04	0.0206	0.9793
3	2.06	0.0186	0.9804	1.59	0.0559	0.9440
4	1.38	0.0837	0.9163	0.93	0.1761	0.8238
5	0.8	0.2118	0.7882	0.42	0.3372	0.6627

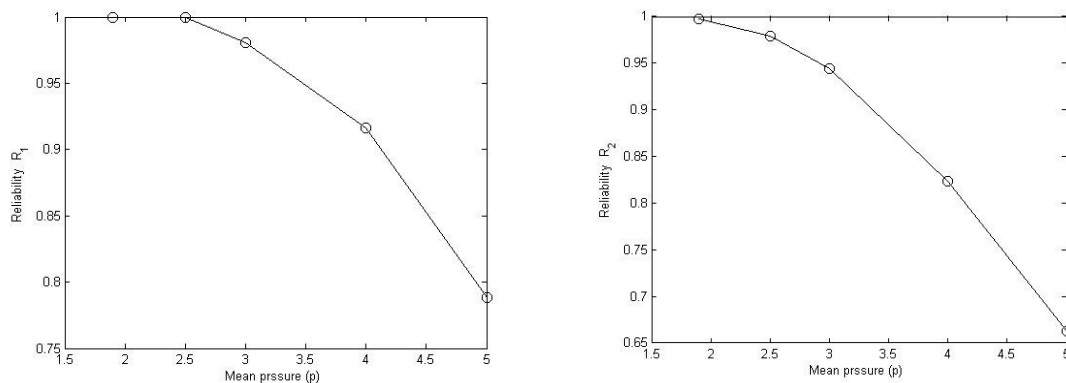


Figure 2: Variation of reliability as a function of mean pressure

3.3. Mean radius vs reliability

If mean radius of the design increases from 800 mm to 2000 mm, then reliability of SA 516 70 Stainless steel decreases from 0.9993 to 0.7882 and the reliability of 6061 aluminum alloy decreases from 0.9958 to 0.6627. With the increase of mean radius, the size of the cylinder increases, then the reliability decreases. The variation of reliability with mean diameter is shown below.

Table 4: Mean radius vs reliability

SA 516 70 Stainless steel				6061 aluminum alloy		
μ_r	β	P_f	R_1	β	P_f	R_2
800	3.21	0.0007	0.9993	2.64	0.0041	0.9958
1000	2.56	0.0053	0.9947	2.05	0.0201	0.9798
1200	2.07	0.0193	0.9807	1.61	0.0536	0.9463
1400	1.7	0.0446	0.9554	1.25	0.1056	0.8943
1600	1.28	0.0838	0.9162	0.94	0.1736	0.8263
1800	1.11	0.1335	0.8665	0.67	0.2514	0.7485
2000	0.8	0.2118	0.7882	0.42	0.3372	0.6627

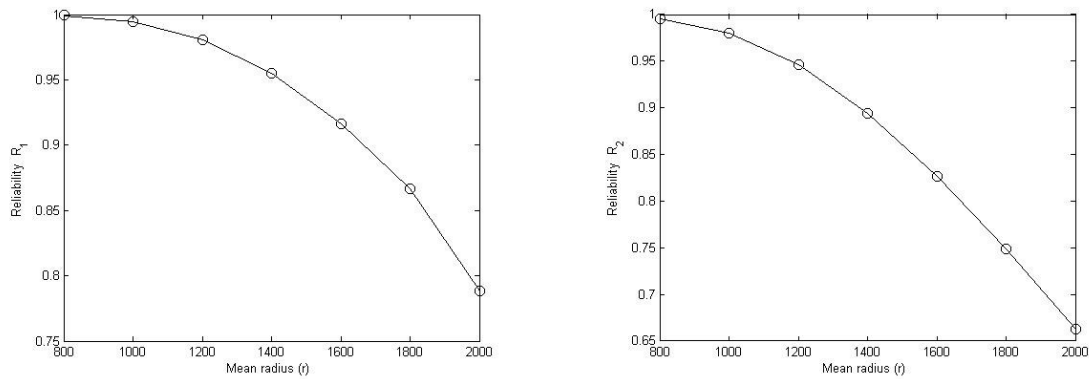


Figure 3: Variation of reliability as a function of mean radius

3.4. Mean strength vs Reliability

In this case, with the increment of mean strength, the deformation is lowered then the reliability of the cylinder enhances. The reliability of SA 516 70 Stainless steel increases from 0.7882 to 0.9846 with the increase of mean strength from 335 MPa to 580 MPa and in 6061 aluminum alloy, the reliability increases from 0.6628 to 0.7580 with the increase of mean strength from 276MPa to 312 MPa.

Table 5: Mean strength vs reliability

SA51670 Stainless steel				6061 aluminum alloy			
μ_c	β	P_f	R_1	μ_c	β	P_f	R_2
335	0.8	0.2118	0.7882	276	0.42	0.3372	0.6628
350	0.97	0.1669	0.8331	280	0.45	0.3263	0.6737
370	1.09	0.1378	0.8622	284	0.49	0.3120	0.6880
390	1.21	0.1131	0.8869	288	0.52	0.3015	0.6985
410	1.33	0.0917	0.9083	292	0.55	0.2911	0.7089
430	1.44	0.0749	0.9251	296	0.58	0.2809	0.7191

SA51670 Stainless steel				6061 aluminum alloy			
μ_c	β	P_f	R_1	μ_c	β	P_f	R_2
450	1.54	0.0617	0.9383	300	0.61	0.2709	0.7291
470	1.65	0.0494	0.9506	304	0.64	0.2610	0.7390
510	1.84	0.0328	0.9672	308	0.67	0.2514	0.7486
580	2.16	0.0154	0.9846	312	0.70	0.2419	0.7580

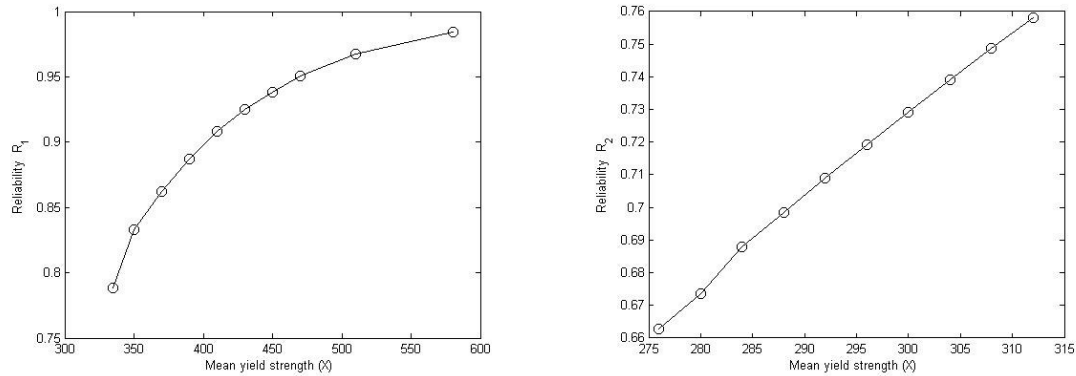


Figure 4: Variation of reliability as a function of mean strength

4. Conclusion

Prediction of the reliability of a pressure vessel leads to high performance of the vessel, several methods have been discussed to estimate the reliability of the vessel. In the analysis, thin-cylindrical pressure vessels with both ends closed are taken. Pressure, radius, thickness and strength of the material are considered as random variables and follow normal distribution. The reliability index of the pressure vessel made with different materials has been found. Reliability analysis has been done for the pressure vessel by using AFOSM with MATLAB.

The analysis shows that mean thickness of SA 516 70 stainless steel cylinder increases from 50 mm to 140 mm then reliability increases from 0.7882 to 0.9998 drastically and reliability of 6061 aluminum alloy cylinder increases from 0.6627 to 0.9985. If there is an increment in mean pressure of the cylinder from 1.9 MPa to 5 MPa then there is decrement in reliability from 0.9996 to 0.7882 and also a decrement in reliability of 6061 aluminum alloy cylinder from 0.6627 to 0.9971. Change in mean radius of the cylinder from 800 mm to 2000 mm causes change in reliability from 0.9996 to 0.7882. If mean strength of SA 516 70 stainless steel cylinder increases from 335 MPa to 630 MPa then reliability increases from 0.7882 to 0.9846. Thickness, diameter, pressure and strength of the cylinder influence the reliability of the pressure vessel. It is observed that strength of the materials has significant influence on reliability of the vessel.

References

- [1] S. Rahman, G. Chen, R. Firmature, probabilistic analysis of off-centre cracks in cylindrical structures, *International Journal of pressure Vessels and Piping* 77(2000)3-16
- [2] Cornell CA. Structural safety: some historical evidence that it is a healthy adolescent. In: *Proc. ICOSSAR' 81, Trondheim, Norway; 1981*. p. 19–31.

- [3] L. Cizelj, B. Mavko, H. Riesch- Opeermann, Application of first and second order reliability methods in the safety assessment of cracked steam generator tubing, *Nuclear Engineering and Design*, 147 (1994) 359-368.
- [4] Gunjan Agarwal and Baidurya Bhattacharya, Partial Safety factor design of rectangular partially pre- stressed concrete beams in ultimate flexural limit state, *Journal of structural Engineering*, Volume 37, No.4, October-November 2010 pp. 257-267.
- [5] Antanas Kudzys and Romualdas Kliukas, Reliability index design in Reinforced concrete structures of annular cross sections, *Modern- Building Materials and Techniques, the 10th International Conference*, 2010, Lithuania.
- [6] P. Hari prasad, M. Tirumala Devi, T. Sumathi Uma Maheshwari, Reliability Analysis of Structural Beam Section using First order Reliability Method, *International Journal of Scientific Research in Science and Technology*, March-April-2018 [(4)5:01-03.
- [7] Zheng Yulong, Lu Zhimin, Wang Linlin, Zhang Lin, Zhou Guangliang, The Application of Hasofer-Lind Method in Reliability Design of Thin-walled Cylindrical Shell, *Applied Mechanics and Materials*, 249-250, 303-306, 2013.
- [8] Devaraju A., and pazhanivel K., A study on stress analysis for design of pressure vessel, *international journal of mechanical and production engineering*, 3(11), pp.98-101, 2015.
- [9] Bhattacharya G.K. and Johnson R.A., Stress- Strength models for system reliability, *Proceedings of the symposium on reliability a fault tree analysis*, SIAM, 509-32, 1975.
- [10] James R. Farr and Maan H. Jawad, GUIDEBOOK FOR THE DESIGN OF ASME SECTION VIII PRESSURE VESSELS, Fourth Edition, 2010.
- [11] L.S. Srinath, Reliability Engineering, 2004.
- [12] S. Ramamrutham, R. Narayan, Strength of Materials, 2018.
- [13] R.S. Kurmi and N. Kurmi, Strength of Materials, S. Chand Publications, 2009.
- [14] R.K. Bansal, "Strength of Materials", Laxmi Publications, 2009.
- [15] K.C. Kapur and L.R. Lamberson, "Reliability in Engineering Design", John Wiley and Sons, Inc. U.K., 1997

WEIGHTED INTERVENED EXPONENTIAL DISTRIBUTION AS A LIFETIME DISTRIBUTION

VILAYAT ALI BHAT¹, SUDESH PUNDIR²

^{1,2} Department of Statistics, Pondicherry University, Kalapet, Puducherry-605014, INDIA

vilayat.stat@gmail.com¹

sudeshpundir19@gmail.com²

Abstract

This study proposes and investigates the weighted intervened exponential distribution, which is demonstrated as a generalized extension of the intervened exponential distribution. The form of the weighted intervened exponential distribution is obtained by considering a specific non-negative weighted function. The probability density function and cumulative density function of the proposed model are given, and its generalized form of reliability function and the hazard rate function is also derived. By choosing a different set of parametric values, the graphical demonstrations of the probability density function of weighted intervened exponential distribution are given where it acquires different curve shapes. The weighted intervened exponential distribution density function is then further studied in the limited form as a special case called the length-biased intervened exponential distribution. Along with the distribution of order statistics, stochastic ordering, stress-strength reliability, and entropy measure, several distributional and reliability aspects of the length-biased intervened exponential distribution are derived. For estimating the unidentified parameters of the length-biased variant, the most suggested approach known as the maximum likelihood estimation technique is implemented. To explore the behavior of the parameter estimates for various sample sizes, a sample data generation technique is required to carry out the process. Since the quantile function of the length biased intervened exponential distribution is not in closed form. So, the alternative data generation algorithm is employed which is known as the acceptance-rejection algorithm technique, and a Monte-Carlo simulation study is done. The absolute average bias and mean square error of the estimated parameters of the length-biased version model are calculated and it is noticed that both the calculated measures decrease simultaneously on increasing the sample size. In order to determine if the model is appropriate, a real-life time-to-event data set is examined as an example, and length biased distribution is juxtaposed with several other common available lifetime distributions for comparison purposes.

Keywords: Acceptance-rejection algorithm; Entropy; Weighted intervened distribution; Monte Carlo simulation

1. INTRODUCTION

The literature witnesses the dominance of exponential distribution among all existing lifetime models over data analysis in essential fields like reliability theory, survival analysis, and several other branches of statistical as well as the applied sciences. As it is well observed the exponential model has been extended in different ways with the help of new model development methods and transformation techniques, such as inverted gamma distribution by Lin *et al.*, [11], generalized exponential distribution by Gupta and Kundu [9], weighted exponential distribution by Gupta and Kundu [8], etc. are few examples. Apart from these extensions, a new development in statistics called the intervention was brought into existence by Shanmugam [21] in the form of a new discrete intervention model called the intervened Poisson distribution. Later on, in another

successful attempt Shanmugam *et al.* [22] developed the continuous intervened exponential model, which was later introduced as a lifetime distribution in the field of reliability theory and survival analysis by Bhat and Pundir [2]. The cumulative and probability density functions (*c.d.f* & *p.d.f*) of I_vED are given by

$$F_{I_vED}(x; \Theta) = \begin{cases} 1 - \frac{\rho e^{-(x-\sigma)/\rho\theta} - e^{-(x-\sigma)/\theta}}{(\rho-1)} & \rho \neq 1 \\ 1 - \left(1 + \frac{x-\sigma}{\theta}\right) e^{-(x-\sigma)/\theta} & \rho = 1 \end{cases} \quad (1)$$

and,

$$f_{I_vED}(x; \Theta) = \begin{cases} \frac{e^{-(x-\sigma)/\rho\theta} - e^{-(x-\sigma)/\theta}}{(\rho-1)\theta} & \rho \neq 1 \\ \frac{(x-\sigma)}{\theta^2} e^{-(x-\sigma)/\theta} & \rho = 1 \end{cases} \quad (2)$$

where, $\sigma < x < \infty$, and $\Theta = \{(\sigma, \theta, \rho) : \sigma > 0, \theta > 0, \rho > 0\}$ which is commonly known as the parameter space with intervention parameter ρ and rate parameter θ . The wider applicability of the model motivated us to extend the intervened exponential distribution (I_vED) further in the direction of the weighted distributions and the distribution obtained is known as weighted intervened exponential distribution (WI_vED). Thus, the newly proposed WI_vED is a type of generalization of I_vED . To trace the history of weighted distributions, Fisher [7] investigated the influence of ascertainment procedures on frequency estimates and developed the notion of weighted distributions. When diversifying Fisher's core theories, Rao ([18], [19]) realized the need for a unifying notion by attempting to identify numerous sampling conditions that might be handled according to what he named weighted distributions. To know the history and applicability of weighted distributions Patil published several articles that he wrote with co-authors refer Taillie *et al.* [23], Patil and Taillie [12], Patil and Rao [15], Denis and Patil [5], Laird *et al.* [10], Patil and Ord [14], etc. moreover, see more references Patil [16]. In a definition, Patil *et al.* [13] proposed the methodology used to determine the weighted version probability density function (*p.d.f*) is depicted below:

$$f(x; \Theta) = \frac{w(x)f_*(x; \Theta)}{E[w(X)]}$$

where, $f_*(x; \Theta)$ is the natural density function of existing distribution with parameter space Θ and $w(x)$ is the non-negative weighted function by choosing $w(X) = X^r$, then $E[w(X)] = E[X^r]$ is the r^{th} moment about origin. The r^{th} raw moment of I_vED derived by Bhat and Pundir [2] is presented in a simplified form as given by

$$\mu'_r = \frac{1}{\theta(\rho-1)} \sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+1} (\rho^{r-k+1} - 1). \quad (3)$$

Therefore, according to the definition of weighted distributions, the *c.d.f* and *p.d.f* of $WI_vED(\sigma, \theta, \rho)$ derived from a $r.v$ $X \sim I_vED(\sigma, \theta, \rho)$ are given by

$$F_{WI_vED}(x; \Theta) = \begin{cases} 1 - \frac{\theta^{r+1} \{ \rho^{r+1} e^{\sigma/\rho\theta} \Gamma(r+1, x/\rho\theta) - e^{\sigma/\theta} \Gamma(r+1, x/\theta) \}}{\sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+1} (\rho^{r-k+1} - 1) \Gamma(r-k+1)} & \rho \neq 1 \\ 1 - \frac{\theta^{r+1} \{ \theta \Gamma(r+2, x/\theta) - \sigma \Gamma(r+1, x/\theta) \}}{\sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+2} \Gamma(r-k+2)} e^{\sigma/\theta} & \rho = 1 \end{cases} \quad (4)$$

and,

$$f_{WI_vED}(x; \Theta) = \begin{cases} \frac{x^r e^{-(x-\sigma)/\rho\theta} - x^r e^{-(x-\sigma)/\theta}}{\sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+1} (\rho^{r-k+1} - 1) \Gamma(r-k+1)} & \rho \neq 1 \\ \frac{x^r (x-\sigma)}{\sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+2} \Gamma(r-k+2)} e^{-(x-\sigma)/\theta} & \rho = 1 \end{cases} \quad (5)$$

where, $\sigma < x < \infty$, and Θ is parameter space. The reliability function of the $WI_vED(\sigma, \theta, \rho)$ is obtained as

$$R_{WI_vED}(x; \Theta) = \begin{cases} \frac{\theta^{r+1} \{ \rho^{r+1} e^{\sigma/\rho\theta} \Gamma(r+1, x/\rho\theta) - e^{\sigma/\theta} \Gamma(r+1, x/\theta) \}}{\sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+1} (\rho^{r-k+1} - 1) \Gamma(r-k+1)} & \rho \neq 1 \\ \frac{\theta^{r+1} \{ \theta \Gamma(r+2, x/\theta) - \sigma \Gamma(r+1, x/\theta) \}}{\sum_{k=0}^r \binom{r}{k} \sigma^k \theta^{r-k+2} \Gamma(r-k+2)} e^{\sigma/\theta} & \rho = 1 \end{cases} \quad (6)$$

and, its hazard rate function is

$$h_{WI_vED}(x; \Theta) = \begin{cases} \frac{x^r e^{-(x-\sigma)/\rho\theta} - x^r e^{-(x-\sigma)/\theta}}{\theta^{r+1} \{ \rho^{r+1} e^{\sigma/\rho\theta} \Gamma(r+1, x/\rho\theta) - e^{\sigma/\theta} \Gamma(r+1, x/\theta) \}} & \rho \neq 1 \\ \frac{x^r (x-\sigma)}{\theta^{r+1} \{ \theta \Gamma(r+2, x/\theta) - \sigma \Gamma(r+1, x/\theta) \}} e^{-x/\theta} & \rho = 1 \end{cases} \quad (7)$$

further, we study the special case density function of WI_vED obtained at $r = 1$ from equation (5), that is, defined as Length(size)-biased intervened exponential distribution (LBI_vED). The mean, variance several statistical and reliability properties, parameter estimation, simulation, stochastic ordering, order statistics, and real-life applicability of LBI_vED are studied in proceeding sections and subsections.

2. DISTRIBUTION AND ITS PROPERTIES

The cumulative and probability density functions (*c.d.f* & *p.d.f*) of LBI_vED are given by

$$F_{LBI_vED}(x; \Theta) = \begin{cases} 1 - \frac{\{ \rho(x+\rho\theta)e^{-(x-\sigma)/\rho\theta} - (x+\theta)e^{-(x-\sigma)/\theta} \}}{(\rho-1)[\sigma+\theta(\rho+1)]} & \rho \neq 1 \\ 1 - \frac{\theta^2 + (x+\theta)(x-\sigma+\theta)}{\theta(\sigma+2\theta)} e^{-(x-\sigma)/\theta} & \rho = 1 \end{cases} \quad (8)$$

and,

$$f_{LBI_vED}(x; \Theta) = \begin{cases} \frac{xe^{-(x-\sigma)/\rho\theta} - xe^{-(x-\sigma)/\theta}}{\theta(\rho-1)[\sigma+\theta(\rho+1)]} & \rho \neq 1 \\ \frac{x(x-\sigma)}{\theta^2(\sigma+2\theta)} e^{-(x-\sigma)/\theta} & \rho = 1 \end{cases} \quad (9)$$

where, $0 < \sigma < x < \infty$ and Θ being its parameter space. The attempt is made to derive the expressions of mean (μ_x) and its variance (σ_x^2). Also, some of the measures of $LBI_vED(\sigma, \theta, \rho \neq 1)$ which are not in closed form are the quantile function, median and the mode. Note, for simplicity here on-wards, we take $\tau = (\rho - 1) [\sigma + \theta (\rho + 1)]$.

$$\mu_x = \frac{2\theta^2 (\rho^3 - 1) + 2\sigma\theta (\rho^2 - 1) + \sigma^2 (\rho - 1)}{\tau} \quad (10)$$

and

$$\sigma_x^2 = \frac{2\theta^3 (\rho^4 - 1) + 2\sigma\theta^2 (\rho^3 - 1) + (3\sigma + 2)\sigma\theta (\rho^2 - 1) + \sigma^3 (\rho - 1)}{\tau} - (\mu_x)^2. \quad (11)$$

In Figure 1 and Figure 2, the *p.d.f* of $LBI_vED(\sigma, \theta, \rho \neq 1)$ is plotted graphically. This is evident from the graphical representation that the distribution is positively skewed, where multiple *p.d.f* curves are displayed for various chosen sets of parameter values.

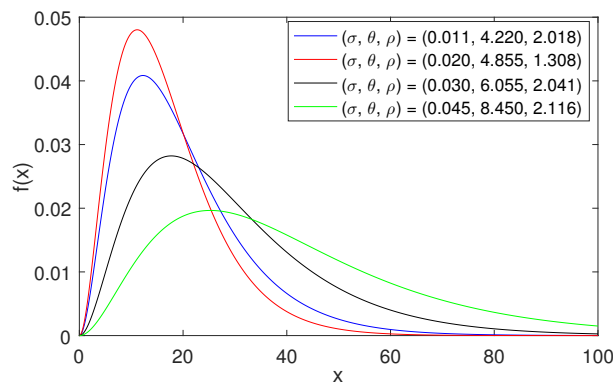


Figure 1: *P.d.f* Plot

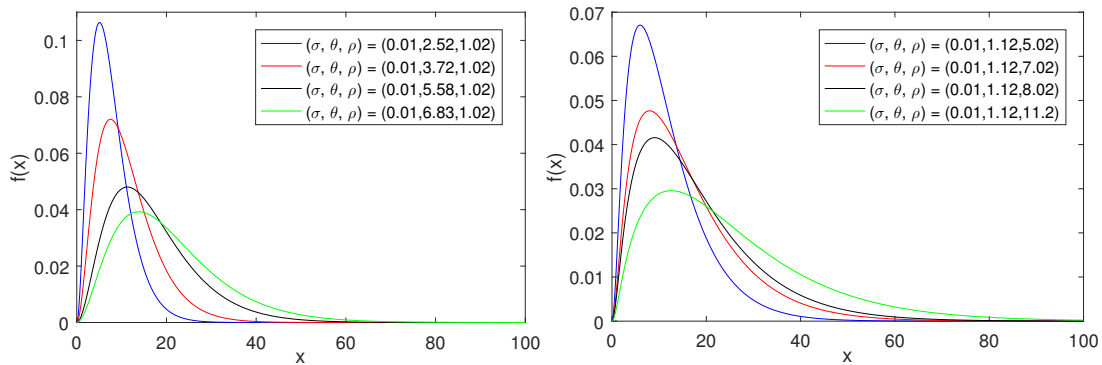


Figure 2: P.d.f Subplots

2.1. Moments

In this subsection, the derivation of moments particularly central and non-central are derived. So, if a random variable $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$, then central moments (μ_r) expression is obtained by

$$\mu_r = \frac{1}{\theta\tau} \int_{\sigma}^{\infty} (x - \mu_x)^r x \left\{ e^{-(x-\sigma)/\rho\theta} - e^{-(x-\sigma)/\theta} \right\} dx$$

to solve the integral, first, we expand the term $(x - \mu_x)^r$ by using binomial expansion. After, integration and simplification the resulting equation is

$$\mu_r = \frac{1}{\tau} \sum_{k=0}^r \binom{r}{k} (\sigma - \mu_x)^r \theta^{r-k} \Gamma(r - k + 1) \left\{ \theta (\rho^{r-k+2} - 1) (r - k + 1) - \sigma (\rho^{r-k+1} - 1) \right\}. \quad (12)$$

In a similar context, we derive the non-central moments (μ'_r) expression for a random variable $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$. The procedure to obtain the final resulting expression is given by

$$\mu'_r = \frac{1}{\theta\tau} \int_{\sigma}^{\infty} x^{r+1} \left\{ e^{-(x-\sigma)/\rho\theta} - e^{-(x-\sigma)/\theta} \right\} dx$$

after, simplifications the resulting equation is

$$\mu'_r = \frac{1}{\tau} \sum_{k=0}^{r+1} \binom{r+1}{k} \sigma^k \Gamma(r - k + 2) \theta^{r-k+1} (\rho^{r-k+2} - 1) \quad (13)$$

2.2. Generating Functions for Moments

In general, the moment generating function (*m.g.f*) denoted by $M_X(t)$ of a *r.v* X is obtained by,

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \quad (14)$$

so, to derive the *m.g.f* of a *r.v* $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$, we have

$$M_X(t) = \int_{\sigma}^{\infty} e^{tx} f_{LBI_vED}(x; \Theta) dx$$

on substituting the *p.d.f* of LBI_vED , when $\rho \neq 1$ and proceed with the transforming technique, we get

$$= \frac{e^{tx}}{\theta\tau} \int_0^\infty (x + \sigma) \left\{ e^{-\left(\frac{1}{\theta\rho} - t\right)x} - e^{-\left(\frac{1}{\theta} - t\right)x} \right\} dx$$

on solving this further by using the gamma function, the required resulting equation of *m.g.f* is obtained as

$$M_X(t) = \frac{\{\sigma(1 - \theta t)(1 - \theta\rho t) + \theta(1 + \rho - 2\theta\rho t)\} e^{\sigma t}}{(1 - \theta t)^2(1 - \theta\rho t)^2[\sigma + \theta(\rho + 1)]}. \tag{15}$$

Next, the characteristic function (*c.f*) of a *r.v* $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$ is derived by

$$\begin{aligned} \phi_X(t) &= E(e^{itX}) = \int_\sigma^\infty e^{itx} f_{LBI_vED}(x; \Theta) dx \\ &= \frac{\{\sigma(1 - \theta it)(1 - \theta\rho it) + \theta(1 + \rho - 2\theta\rho it)\} e^{\sigma it}}{(1 - \theta it)^2(1 - \theta\rho it)^2[\sigma + \theta(\rho + 1)]} \end{aligned} \tag{16}$$

3. DISCUSSION ON RELIABILITY PROPERTIES

Suppose a random variable X is such that it is non-negative, $X \sim LBI_vED(\sigma, \theta, \rho)$, the *p.d.f* of X is $f_{LBI_vED}(x; \Theta)$, and its *c.d.f* $F_{LBI_vED}(x; \Theta)$ which are given in equation (9) and (8) respectively. Then the reliability function of $LBI_vED(\sigma, \theta, \rho)$ is given by

$$R_{LBI_vED}(x; \xi) = \begin{cases} \frac{\rho(x + \rho\theta)e^{-(x-\sigma)/\rho\theta} - (x + \theta)e^{-(x-\sigma)/\theta}}{\theta^2 + (x + \theta)(x - \sigma + \theta)} e^{-(x-\sigma)/\theta} & \rho \neq 1 \\ \frac{x(x - \sigma)}{\theta(\sigma + 2\theta)} & \rho = 1 \end{cases} \tag{17}$$

the essential failure rate functions of a random variable $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$ are obtained. The hazard rate and reverse hazard rate functions are given by

$$h_{LBI_vED}(x; \xi) = \begin{cases} \frac{xe^{-(x-\sigma)/\rho\theta} - xe^{-(x-\sigma)/\theta}}{\theta\{\rho(x + \rho\theta)e^{-(x-\sigma)/\rho\theta} - (x + \theta)e^{-(x-\sigma)/\theta}\}} & \rho \neq 1 \\ \frac{x(x - \sigma)}{\theta\{\theta^2 + (x + \theta)(x - \sigma + \theta)\}} & \rho = 1 \end{cases} \tag{18}$$

in Figure 3 and Figure 4, the hazard rate function of $LBI_vED(\sigma, \theta, \rho \neq 1)$ is plotted graphically. It is shown in the graphical representation the distribution is having increasing multiple hazard rate function curves that are displayed for various chosen sets of parameter values.

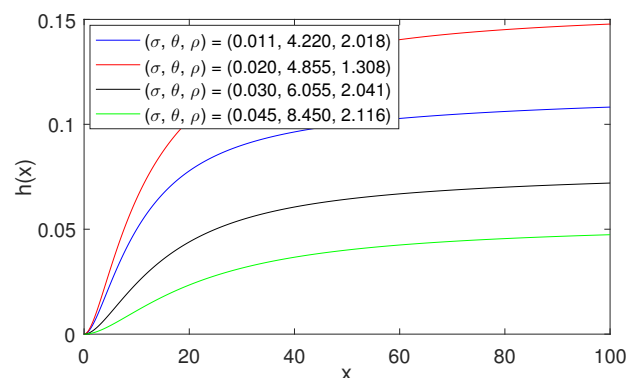


Figure 3: Hazard Rate Function Plot

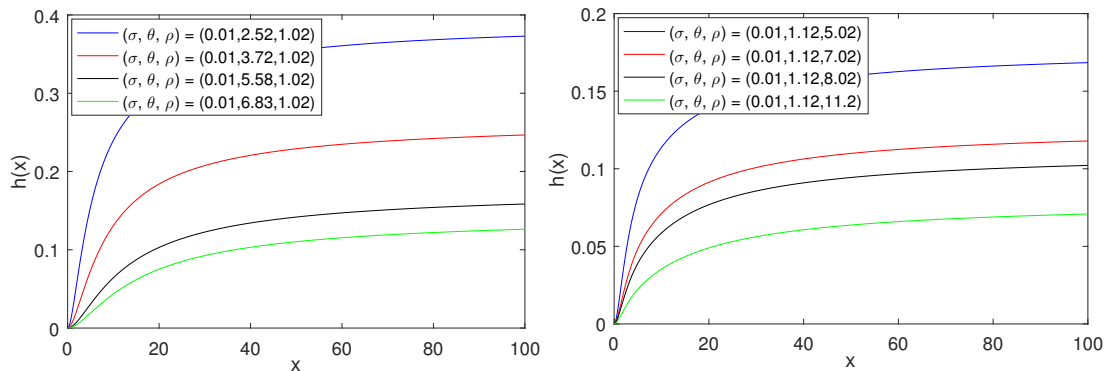


Figure 4: Hazard Rate Function Subplots

3.1. Mean Residual Life Function

The Mean Residual Life (*MRL*) function is used to investigate the aging process phenomenon of systems. To determine the *MRL* for a random variable X representing the component's life, is defined by

$$m_{LBI_vED}(x; \Theta) = E[X - x | X > x] = \frac{1}{R_{LBI_vED}(x; \Theta)} \int_x^{\infty} R_{LBI_vED}(x; \Theta) dx. \quad (19)$$

The function $m(x; \Theta)$ is defined over any domain of a random variable, it is of special significance for a non-negative random variable explaining the lifetime of the system, and at time x for an operational system, conditional expected residual life is determined by Finkelstein [6]. Furthermore, while the shape of the hazard rate function is essential, the *MRL* function is considered to be much more meaningful than the hazard rate function. Since the first describes the full residual life function, whilst the other simply evaluates the possibility of immediate failure at the time x . The *MRL* function of the $LBI_vED(\sigma, \theta, \rho \neq 1)$ is determined by using the relation (19) and is obtained as follows:

$$m_{LBI_vED}(x; \Theta) = \frac{1}{R_{LBI_vED}(x; \Theta)} \int_x^{\infty} \frac{\rho(x + \rho\theta) e^{-(x-\sigma)/\rho\theta} - (x + \theta) e^{-(x-\sigma)/\theta}}{\tau} dx.$$

On solving the above integral and further simplifications gives the following required *MRL* function of the model.

$$m_{LBI_vED}(x; \Theta) = \frac{\theta\rho^2(x + 2\rho\theta) e^{-(x-\sigma)/\rho\theta} - \theta(x + 2\theta) e^{-(x-\sigma)/\theta}}{\tau R_{LBI_vED}(x; \Theta)} \quad (20)$$

3.2. Stress-Strength Reliability

It is easily noticed, that a significant amount of research work had already been carried out in the field of stress-strength modeling on the estimation of reliability $R = Pr.(X_1 > X_2)$, where the random variables X_1 and X_2 represent the strength and stress factors of the system possessing the same distributions of uni-variate family. For the vast majority of the well-known standard distributions, the algebraic representation for R has been developed. Here, $X_1 \sim LBI_vED(\sigma, \theta_2, \rho_1 \neq 1)$ and $X_2 \sim LBI_vED(\sigma, \theta_2, \rho_2 \neq 1)$ are unrelated and have the same distribution, so-called LBI_vED ,

we calculate the reliability R . The procedure to measure R by having the *c.d.f* of X_2 and *p.d.f* of X_1 is described as follows

$$F_{LBI_vED}(x; \Theta) = 1 - \frac{\rho_2 (x + \rho_2 \theta_2) e^{-(x-\sigma)/\rho_2 \theta_2} - (x + \theta_2) e^{-(x-\sigma)/\theta_2}}{(\rho_2 - 1) [\sigma + \theta_2 (\rho_2 + 1)]} \quad \rho \neq 1 \quad (21)$$

and,

$$f_{LBI_vED}(x; \Theta) = \frac{x e^{-(x-\sigma)/\rho_1 \theta_1} - x e^{-(x-\sigma)/\theta_1}}{\theta_1 (\rho_1 - 1) [\sigma + \theta_1 (\rho_1 + 1)]} \quad \rho \neq 1 \quad (22)$$

then, R is derived by

$$\begin{aligned} R &= \int_{\sigma}^{\infty} \left\{ \int_{\sigma}^x f_{X_2}(x) dx \right\} f_{X_1}(x) dx = \int_{\sigma}^{\infty} F_{X_2}(x) f_{X_1}(x) dx \quad (23) \\ &= 1 - \frac{\theta_1 \theta_2 \rho_1 \rho_2^2}{B (\theta_1 \rho_1 + \theta_2 \rho_2)} \left\{ \sigma (\sigma + \theta_2 \rho_2) + \frac{\theta_1 \theta_2 \rho_1 \rho_2 (2\sigma + \theta_2 \rho_2)}{(\theta_1 \rho_1 + \theta_2 \rho_2)} + \frac{2 (\theta_1 \theta_2 \rho_1 \rho_2)^2}{(\theta_1 \rho_1 + \theta_2 \rho_2)^2} \right\} \\ &+ \frac{\theta_1 \theta_2 \rho_1}{B (\theta_1 \rho_1 + \theta_2)} \left\{ \sigma (\sigma + \theta_2) + \frac{\theta_1 \theta_2 \rho_2 (2\sigma + \theta_2)}{(\theta_1 \rho_1 + \theta_2)} + \frac{2 (\theta_1 \theta_2 \rho_2)^2}{(\theta_1 \rho_1 + \theta_2)^2} \right\} \\ &+ \frac{\theta_1 \theta_2 \rho_2^2}{B (\theta_1 + \theta_2 \rho_2)} \left\{ \sigma (\sigma + \theta_2 \rho_2) + \frac{\theta_1 \theta_2 \rho_2 (2\sigma + \theta_2 \rho_2)}{(\theta_1 + \theta_2 \rho_2)} + \frac{2 (\theta_1 \theta_2 \rho_2)^2}{(\theta_1 + \theta_2 \rho_2)^2} \right\} \\ &- \frac{\theta_1 \theta_2}{B (\theta_1 + \theta_2)} \left\{ \sigma (\sigma + \theta_2) + \frac{\theta_1 \theta_2 (2\sigma + \theta_2)}{(\theta_1 + \theta_2)} + \frac{2 (\theta_1 \theta_2)^2}{(\theta_1 + \theta_2)^2} \right\} \end{aligned}$$

where, $B = \theta_1 (\rho_1 - 1) (\rho_2 - 1) [\sigma + \theta_1 (\rho_1 + 1)] [\sigma + \theta_2 (\rho_2 + 1)]$.

4. ENTROPY MEASURES

In science and engineering, entropy measures have already been studied as a practical application of the system. The two uncertainty measures of variation so-called Rényi entropy by Rényi [17] and Tsallis entropy by Tsallis [24] are studied in this section, the expressions of both these measures are given in the following respective subsections:

4.1. Rényi Entropy

The Rényi entropy of order α , for a non-negative *r.v* $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$ is derived by

$$\hbar_R(\alpha) = \frac{1}{1-\alpha} \log \left[\int_{\sigma}^{\infty} \{f_{LBI_vED}(x; \Theta)\}^{\alpha} dx \right]; \quad \alpha \geq 0, \alpha \neq 1$$

on substituting the *p.d.f* of LBI_vED when $\rho \neq 1$ and making use of binomial expansion by treating α as a finite nature number. The simplified equation of Rényi entropy of order α is derived as given by

$$\hbar_R(\alpha) = \frac{1}{1-\alpha} \log \left\{ \sum_{r=0}^{\alpha} \sum_{k=0}^{\alpha} \frac{\binom{\alpha}{r} \binom{\alpha}{k} (-1)^r \sigma^k (\theta \rho)^{\alpha-k+1} \Gamma(\alpha-k+1)}{\{\theta \tau\}^r [\alpha+r(\rho-1)]^{\alpha-k+1}} \right\} \quad (24)$$

4.2. Tsallis Entropy

The Tsallis entropy also known as *q-entropy* of order q , for a non-negative *r.v* $X \sim LBI_vED(\sigma, \theta, \rho \neq 1)$ is derived by

$$T_q(x) = \frac{1}{1-q} \left[1 - \int_{\sigma}^{\infty} \{f_{LBI_vED}(x; \Theta)\}^q dx \right]; \quad q \geq 0, q \neq 1$$

on substituting the *p.d.f* of LBI_vED when $\rho \neq 1$ and making use of binomial expansion by treating α as a finite nature number. The simplified equation of Tsallis entropy of order q is derived as given by

$$T_q(x) = \frac{1}{1-q} \left[1 - \sum_{r=0}^q \sum_{k=0}^q \frac{\binom{q}{r} \binom{q}{k} (-1)^r \sigma^k (\theta\rho)^{q-k+1} \Gamma(q-k+1)}{\{\theta\tau\}^r [q+r(\rho-1)]^{q-k+1}} \right] \quad (25)$$

5. STOCHASTIC ORDERING AND ORDER STATISTICS

In this section, the stochastic ordering of the model and its order statistics are discussed.

5.1. Stochastic Ordering

Over the past several years, the usage of stochastic ordering has drastically increased across a wide range of statistical fields. These disciplines include reliability theory, queuing theory, survival analysis, and many more fields refer to Shaked and Shantikumar [20]. Let the *r.v/s* X_1 and X_2 , where $X_1 \sim LBI_vED(\sigma_1, \theta_1, \rho_1 \neq 1)$ and $X_2 \sim LBI_vED(\sigma_2, \theta_2, \rho_2 \neq 1)$, have *p.d.f/s* denoted by $f_{X_1}(x)$, $f_{X_2}(x)$ and *c.d.f/s* denoted by $F_{X_1}(x)$, $F_{X_2}(x)$ respectively. According to the model, it is believed that the random variable X_1 is smaller than X_2 in the,

- (a₁) Stochastic order (mathematically $X_1 \leq_{st} X_2$), if $F_{X_1}(x; \Theta_1) \geq F_{X_2}(x; \Theta_2) \forall x$.
- (a₂) Hazard rate order (mathematically $X_1 \leq_{hr} X_2$), if $h_{X_1}(x; \Theta_1) \geq h_{X_2}(x; \Theta_2) \forall x$.
- (a₃) Mean residual life order (mathematically $X_1 \leq_{MRL} X_2$), if $m_{X_1}(x; \Theta_1) \geq m_{X_2}(x; \Theta_2) \forall x$.
- (a₄) Likelihood ratio order (mathematically $X_1 \leq_{LR} X_2$), if $\frac{f_{X_1}(x; \Theta_1)}{f_{X_2}(x; \Theta_2)}$ decreases in x .

Since the following results show that, the four stochastic orders stated above are connected,

$$X_1 \leq_{MRL} X_2 \Leftrightarrow X_1 \leq_{hr} X_2 \Leftrightarrow X_1 \leq_{LR} X_2 \text{ and } X_1 \leq_{st} X_2 \Leftrightarrow X_1 \leq_{hr} X_2.$$

When the required conditions are met, the $LBI_vED(\sigma, \theta, \rho \neq 1)$ models are arranged *w.r.t* the strongest likelihood ratio ordering, stated in the following theorem.

Theorem 1. Suppose $X_1 \sim LBI_vED(\sigma_1, \theta_1, \rho_1)$, and $X_2 \sim LBI_vED(\sigma_2, \theta_2, \rho_2)$, with the condition, that if $\sigma_1 = \sigma_2 = \sigma$, $(\rho_1 > \rho_2) > 1$, and $(\theta_1 > \theta_2)$ then $(X_1 \leq_{st} X_2)$, $(X_1 \leq_{hr} X_2)$, $(X_1 \leq_{lr} X_2)$, and $(X_1 \leq_{MRL} X_2)$.

Proof. Proof of the specified ratio is sufficient to demonstrate the outcome.

$$\frac{f_{X_1}(x; \Theta_1)}{f_{X_2}(x; \Theta_2)} = \frac{\theta_2(\rho_2 - 1) [\sigma_2 + \theta_2(\rho_2 + 1)] \left\{ e^{-(x-\sigma_1)/\rho_1\theta_1} - e^{-(x-\sigma_1)/\theta_1} \right\}}{\theta_1(\rho_1 - 1) [\sigma_1 + \theta_1(\rho_1 + 1)] \left\{ e^{-(x-\sigma_2)/\rho_2\theta_2} - e^{-(x-\sigma_2)/\theta_2} \right\}}.$$

Now, applying log both sides and differentiate *w.r.t* x .

$$\frac{d}{dx} \log \left\{ \frac{f_{X_1}(x; \Theta_1)}{f_{X_2}(x; \Theta_2)} \right\} = \frac{\theta_1\rho_1 U (V_1 - \rho_2 V_2) - \theta_2\rho_2 V (U_1 - \rho_1 U_2)}{\theta_1\theta_2\rho_1\rho_2 UV}$$

where, $U = \left\{ e^{-(x-\sigma_1)/\rho_1\theta_1} - e^{-(x-\sigma_1)/\theta_1} \right\}$, $V = \left\{ e^{-(x-\sigma_2)/\rho_2\theta_2} - e^{-(x-\sigma_2)/\theta_2} \right\}$, $U_1 = e^{-(x-\sigma_1)/\rho_1\theta_1}$, $U_2 = e^{-(x-\sigma_1)/\theta_1}$, $V_1 = e^{-(x-\sigma_2)/\rho_2\theta_2}$, and $V_2 = e^{-(x-\sigma_2)/\theta_2}$. Hence, if $\sigma_1 = \sigma_2 = \sigma$, $(\rho_1 > \rho_2)$, and $(\theta_1 > \theta_2)$ then $\frac{d}{dx} \log \left\{ \frac{f_{X_1}(x; \Theta_1)}{f_{X_2}(x; \Theta_2)} \right\} \leq 0$, which implies that $(X_1 \leq_{st} X_2)$, $(X_1 \leq_{hr} X_2)$, $(X_1 \leq_{lr} X_2)$, and $(X_1 \leq_{MRL} X_2)$. ■

5.2. Order Statistics

In order to discuss the order statistics of n random observations $x = \{ x_1, x_2, x_3, \dots, x_n \}$ drawn from $LBI_vED(\sigma, \theta, \rho \neq 1)$. The random sample is arranged in ascending order such as $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$. Then, we denote the *p.d.f* of random variable which is of i^{th} order by $f_{i:n}(x_i; \Theta)$; $i = 1, 2, \dots, n$, and joint *p.d.f* of $(i, j)^{th}$ order variable pair by $f_{i:j:n}(x_i, x_j)$; $1 \leq i \leq j \leq n$, whose expressions are given by

$$f_{i:n}(x_i; \Theta) = \pi_1 [F_{LBI_vED}(x_i; \Theta)]^{i-1} [1 - F_{LBI_vED}(x_i; \Theta)]^{n-i} f_{LBI_vED}(x_i; \Theta) \quad (26)$$

and,

$$f_{i:j:n}(x_i, x_j) = \pi_2 [F_{LBI_vED}(x_i; \Theta)]^{i-1} [F_{LBI_vED}(x_j; \Theta) - F_{LBI_vED}(x_i; \Theta)]^{j-i-1} [1 - F_{LBI_vED}(x_j; \Theta)]^{n-j} f_{LBI_vED}(x_i; \Theta) f_{LBI_vED}(x_j; \Theta) \quad (27)$$

where $F(\cdot)$ and $f(\cdot)$ are *c.d.f* and *p.d.f*. The constants π_1 and π_2 are given by

$$\pi_1 = \frac{n!}{(i-1)!(n-i)!} \text{ and } \pi_2 = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}$$

on substituting $i = 1$ and $i = n$ in equation (26) the *p.d.f*s of 1^{st} and n^{th} order statistics obtained are given as follows:

$$\begin{aligned} f_{1:n}(x_{(1)}; \Theta) &= n \left[1 - F_{LBI_vED}(x_{(1)}; \Theta) \right]^{n-1} f_{LBI_vED}(x_{(1)}; \Theta) \\ &= \frac{n \left[\rho \phi_{(1)} e^{-(x_{(1)}-\sigma)/\rho\theta} - \psi_{(1)} e^{-(x_{(1)}-\sigma)/\theta} \right]^{n-1} \left[x_{(1)} \left\{ e^{-(x_{(1)}-\sigma)/\rho\theta} - e^{-(x_{(1)}-\sigma)/\theta} \right\} \right]}{\theta(\rho-1)^n [\sigma + \theta(\rho+1)]^n} \end{aligned} \quad (28)$$

and,

$$\begin{aligned} f_{n:n}(x_{(n)}; \Theta) &= n \left[F_{LBI_vED}(x_{(n)}; \Theta) \right]^{n-1} f_{LBI_vED}(x_{(n)}; \Theta) \\ &= \frac{n \left[\tau - \left\{ \rho \phi_{(n)} e^{-(x_{(n)}-\sigma)/\rho\theta} - \psi_{(n)} e^{-(x_{(n)}-\sigma)/\theta} \right\} \right]^{n-1} \left[x_{(n)} \left\{ e^{-(x_{(n)}-\sigma)/\rho\theta} - e^{-(x_{(n)}-\sigma)/\theta} \right\} \right]}{\theta(\rho-1)^n [\sigma + \theta(\rho+1)]^n}. \end{aligned} \quad (29)$$

Similarly, the joint order statistic density function of $LBI_vED(\sigma, \theta, \rho \neq 1)$ is given as

$$\begin{aligned} f_{i:j:n}(x_i, x_j) &= \frac{\pi_2}{\theta^2(\rho-1)^n [\sigma + \theta(\rho+1)]^n} \\ &\cdot \left[(\rho-1) [\sigma + \theta(\rho+1)] - \left\{ \rho \phi_{(i)} e^{-(x_i-\sigma)/\rho\theta} - \psi_{(i)} e^{-(x_i-\sigma)/\theta} \right\} \right]^{i-1} \\ &\cdot \left[\left\{ \rho \phi_{(i)} e^{-(x_i-\sigma)/\rho\theta} - \psi_{(i)} e^{-(x_i-\sigma)/\theta} \right\} - \left\{ \rho \phi_{(j)} e^{-(x_j-\sigma)/\rho\theta} - \psi_{(j)} e^{-(x_j-\sigma)/\theta} \right\} \right]^{j-i-1} \\ &\cdot \left[\rho \phi_{(j)} e^{-(x_j-\sigma)/\rho\theta} - \psi_{(j)} e^{-(x_j-\sigma)/\theta} \right]^{n-j} \left[x_{(i)} \left\{ e^{-(x_i-\sigma)/\rho\theta} - e^{-(x_i-\sigma)/\theta} \right\} \right] \\ &\cdot \left[x_{(j)} \left\{ e^{-(x_j-\sigma)/\rho\theta} - e^{-(x_j-\sigma)/\theta} \right\} \right] \end{aligned}$$

where, $\phi_{(r)} = (x_{(r)} + \rho\theta)$ and $\psi_{(r)} = (x_{(r)} + \theta)$; $r = 1, n, i, j$

6. PARAMETER ESTIMATION AND SIMULATION

6.1. Estimation Procedure of Model Parameters

Let a random sample x_1, x_2, \dots, x_n , consisting of n observations is drawn from $LBI_vED(\sigma, \theta, \rho \neq 1)$. The most frequently used technique called the method of maximum likelihood estimation

approach is chosen for the estimation of the parameters. So, we write the log-likelihood function of the complete data for $LBI_vED(\sigma, \theta, \rho \neq 1)$ as

$$\log L = \sum_{i=1}^n \log \left\{ x_i e^{-(x_i - \sigma)/\rho\theta} - x_i e^{-(x_i - \sigma)/\theta} \right\} - n \log \{ \theta(\rho - 1) [\sigma + \theta(\rho + 1)] \}. \quad (30)$$

Setting $\phi = e^{-(x_i - \sigma)/\rho\theta}$ and $\psi = e^{-(x_i - \sigma)/\theta}$, then (30) becomes

$$\log L = \sum_{i=1}^n \log \{ \phi - \psi \} + \sum_{i=1}^n \log x_i - n \log \theta - n \log(\rho - 1) - n \log [\sigma + \theta(\rho + 1)]. \quad (31)$$

On differentiating equation (31) with respect to (*w.r.t*) σ , θ , and ρ , the normal equations obtained are

$$\frac{\partial \log L}{\partial \sigma} = \sum_{i=1}^n \frac{\phi - \rho\psi}{\rho\theta(\phi - \psi)} - \frac{n}{[\sigma + \theta(\rho + 1)]} \quad (32)$$

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^n \frac{(x_i - \sigma)(\phi - \rho\psi)}{\rho\theta^2(\phi - \psi)} - \frac{n(\rho + 1)}{[\sigma + \theta(\rho + 1)]} - \frac{n}{\theta} \quad (33)$$

$$\frac{\partial \log L}{\partial \rho} = \sum_{i=1}^n \frac{\phi(x_i - \sigma)}{\rho\theta^2(\phi - \psi)} - \frac{n\theta}{[\sigma + \theta(\rho + 1)]} - \frac{n}{\rho - 1}. \quad (34)$$

To get the maximum likelihood estimates (MLE_s), we equate equations (32), (33), and (34) to zero. But the equations obtained are not in closed form. So, the recommended functions such as *nlm* or *optim* are used to maximize the log-likelihood function in the R programming language. Also, the alternative iterative technique known as the Newton-Raphson method could be employed to yield the solution for the parameters. Further to construct the Fisher information matrix there must exist the second-order partial differentials which do exist as can be proved by the continuity of first-order differentials. Also, let us suppose that the MLE_s of Θ are given by $\hat{\Theta} = \{ (\sigma, \theta, \rho) : \sigma > 0, \theta > 0, \rho > 0 \}$ then Fisher information matrix is defined as

$$I(\Theta) = -E \begin{bmatrix} \frac{\partial^2 \log L}{\partial \sigma^2} & \frac{\partial^2 \log L}{\partial \sigma \partial \theta} & \frac{\partial^2 \log L}{\partial \sigma \partial \rho} \\ \frac{\partial^2 \log L}{\partial \theta \partial \sigma} & \frac{\partial^2 \log L}{\partial \theta^2} & \frac{\partial^2 \log L}{\partial \theta \partial \rho} \\ \frac{\partial^2 \log L}{\partial \rho \partial \sigma} & \frac{\partial^2 \log L}{\partial \rho \partial \theta} & \frac{\partial^2 \log L}{\partial \rho^2} \end{bmatrix} \quad (35)$$

the partial derivatives of Fisher information matrix $I(\Theta)$ are given by

$$\frac{\partial^2 \log L}{\partial \sigma^2} = \sum_{i=1}^n \frac{(\phi - \rho\psi)^2 - (\phi - \psi)(\phi - \rho\psi)}{\rho^2\theta^2(\phi - \psi)^2} + \frac{n}{[\sigma + \theta(\rho + 1)]^2} \quad (36)$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = \sum_{i=1}^n \frac{(x_i - \sigma)(\phi - \rho\psi) \{ (1 + 3\rho)\psi - 2(1 + \rho)\phi \}}{\rho^2\theta^3(\phi - \psi)^2} + \frac{n(\rho + 1)^2}{[\sigma + \theta(\rho + 1)]^2} + \frac{n}{\theta^2} \quad (37)$$

$$\frac{\partial^2 \log L}{\partial \rho^2} = \sum_{i=1}^n \frac{\phi(x_i - \sigma) \{ (\sigma + 2\rho\theta)\phi - 2\rho\theta\psi \}}{\rho^4\theta^2(\phi - \psi)^2} + \frac{n\theta^2}{[\sigma + \theta(\rho + 1)]^2} + \frac{n}{(\rho - 1)^2} \quad (38)$$

$$\frac{\partial^2 \log L}{\partial \theta \partial \sigma} = \sum_{i=1}^n \frac{\rho\theta(\phi^2 + \rho\psi^2) - (\sigma + \rho\theta - 2\sigma\rho + \sigma\rho^2 + \rho^2\theta)\phi\psi}{\rho^2\theta^3(\phi - \psi)^2} + \frac{n(\rho + 1)}{[\sigma + \theta(\rho + 1)]^2} \quad (39)$$

$$\frac{\partial^2 \log L}{\partial \rho \partial \sigma} = \sum_{i=1}^n \frac{\rho\theta(\phi^2 + \rho\psi^2) - (\sigma - \sigma\rho + \rho\theta + \rho^2\theta)\phi\psi}{\rho^3\theta(\phi - \psi)^2} + \frac{n\theta}{[\sigma + \theta(\rho + 1)]^2} \quad (40)$$

$$\frac{\partial^2 \log L}{\partial \rho \partial \theta} = \sum_{i=1}^n \frac{(x_i - \sigma) \{ (\sigma - \sigma\rho + \rho^2\theta)\phi\psi - \rho\theta(\rho + 1)\psi^2 - \rho\theta\phi^2 \}}{\rho^3\theta^3(\phi - \psi)^2} - \frac{\sigma}{[\sigma + \theta(\rho + 1)]^2}. \quad (41)$$

We drop the expectation term as it is difficult to compute for the second-order partial differential elements of the Fisher information matrix see Cohen [4]. Therefore we write $I(\Theta)$ as

$$I(\hat{\Theta}) = - \begin{bmatrix} \frac{\partial^2 \log L}{\partial \sigma^2} & \frac{\partial^2 \log L}{\partial \sigma \partial \theta} & \frac{\partial^2 \log L}{\partial \sigma \partial \rho} \\ \frac{\partial^2 \log L}{\partial \theta \partial \sigma} & \frac{\partial^2 \log L}{\partial \theta^2} & \frac{\partial^2 \log L}{\partial \theta \partial \rho} \\ \frac{\partial^2 \log L}{\partial \rho \partial \sigma} & \frac{\partial^2 \log L}{\partial \rho \partial \theta} & \frac{\partial^2 \log L}{\partial \rho^2} \end{bmatrix}_{(\sigma, \theta, \rho) = (\hat{\sigma}, \hat{\theta}, \hat{\rho})} \quad (42)$$

now to construct the confidence intervals (C.Its) for parameters of $LBI_vED(\sigma, \theta, \rho \neq 1)$, we first find inversion matrix $I(\Theta)^{-1}$ consisting of the diagonal elements as variances whereas the off-diagonal elements represent the co-variances. The $100(1 - \zeta)\%$ C.Its for σ , θ , and ρ are $\hat{\sigma} \pm \chi_{\zeta/2} \sqrt{V(\hat{\sigma})}$, $\hat{\theta} \pm \chi_{\zeta/2} \sqrt{V(\hat{\theta})}$, and $\hat{\rho} \pm \chi_{\zeta/2} \sqrt{V(\hat{\rho})}$ respectively, where the term $\chi_{\zeta/2}$ is $(\zeta/2)th$ upper percentile of a normal variate.

6.2. Simulation Study

It is necessary for a statistical distribution, to explore the behavior of estimated parameters by performing a Monte-Carlo simulation study at various randomly selected sample sizes, as $n = \{30, 90, 150, 210, 350, 500\}$. Since the quantile function of LBI_vED does not exist in closed form. So, to simulate the data, an alternative technique is utilized known as the acceptance-rejection algorithm. In Table 1, the average bias ($Bias = \frac{1}{n} \sum_{i=1}^n (\hat{\Theta} - \Theta)$) and average mean square error ($MSE = \frac{1}{n} \sum_{i=1}^n (\hat{\Theta} - \Theta)^2$) of the estimated parameters are calculated. The simultaneous decrease of *Bias* and *MSE* of parameters σ , θ , and ρ are reported as the sample size increases. It is clear, that the consistency property of estimated parameters holds for LBI_vED .

Table 1: Parameter Bias and MSE of $LBI_vED(\sigma, \theta, \rho \neq 1)$

(σ, θ, ρ)	n	Bias			MSE		
		$\hat{\sigma}$	$\hat{\theta}$	$\hat{\rho}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\rho}$
(0.91, 0.82, 0.78)	30	0.16580	0.37902	1.7e+01	0.027489	0.14366	3.0e+02
	90	0.06044	0.27480	2.65334	0.003654	0.07551	7.04022
	150	0.04272	0.25322	1.11791	0.001825	0.06412	1.24971
	210	0.03215	0.23397	0.97621	0.001034	0.05474	0.95299
	350	0.01838	0.20611	0.67691	0.000338	0.04248	0.45821
	500	0.01439	0.20147	0.63070	0.000207	0.04059	0.39778
(1.01, 1.87, 0.83)	30	0.39676	0.73710	1.2e+01	0.157419	0.54332	1.4e+02
	90	0.16685	0.58280	2.36629	0.027840	0.33966	5.59931
	150	0.11098	0.48364	1.03614	0.012317	0.23391	1.07359
	210	0.08229	0.43797	0.69849	0.006772	0.19182	0.48789
	350	0.05865	0.39767	0.55490	0.003440	0.15814	0.30791
	500	0.04099	0.36097	0.46926	0.001681	0.13030	0.22020
(1.02, 1.59, 1.05)	30	0.37336	0.53592	1.2e+01	0.139400	0.28721	1.6e+02
	90	0.15147	0.32432	1.97947	0.022842	0.10519	3.91829
	150	0.09869	0.25286	0.69786	0.009740	0.06394	0.48701
	210	0.07163	0.22203	0.45958	0.005131	0.04930	0.21121
	350	0.03737	0.15245	0.27079	0.001397	0.02324	0.07333
	500	0.03229	0.11726	0.19951	0.001042	0.01375	0.03980

7. MODEL APPLICABILITY

7.1. Real-Life Data Based Applications

In this section, an illustration of the proposed methodology is given. Real-life examples are presented to demonstrate the superior performance of the proposed LBI_vED model. The real-life data sets that are analyzed in this study are given below:

The first data set of T8 fluorescent lamps analyzed by Ahmed [1] represents the lifetime (Hours) for 50 devices given by: 0.445, 0.493, 0.285, 0.564, 0.760, 0.381, 0.690, 0.579, 0.636, 0.238, 0.149, 0.244, 0.126, 0.796, 0.405, 0.553, 0.780, 0.431, 0.184, 0.375, 0.198, 0.890, 0.192, 0.463, 0.486, 0.521, 0.366, 0.486, 0.116, 0.511, 0.612, 0.117, 0.384, 0.326, 0.057, 0.412, 0.586, 0.517, 0.570, 0.588, 0.497, 0.246, 0.234, 0.228, 0.552, 0.893, 0.403, 0.458, 0.134, 0.338.

The second data set by Bader and Priest [3] consists of 63 samples representing the strength measured in GPA for single carbon fibers with gauge lengths of 10mm given by: 1.901, 2.132, 2.203, 2.228, 2.257, 2.350, 2.361, 2.396, 2.397, 2.445, 2.454, 2.474, 2.518, 2.522, 2.525, 2.532, 2.575, 2.614, 2.616, 2.618, 2.624, 2.659, 2.675, 2.738, 2.740, 2.856, 2.917, 2.928, 2.937, 2.937, 2.977, 2.996, 3.030, 3.125, 3.139, 3.145, 3.220, 3.223, 3.235, 3.243, 3.264, 3.272, 3.294, 3.332, 3.346, 3.377, 3.408, 3.435, 3.493, 3.501, 3.537, 3.554, 3.562, 3.628, 3.852, 3.871, 3.886, 3.971, 4.024, 4.027, 4.225, 4.395, 5.020.

The output results of the analyzed data sets include the estimated parameters, the information measures such as Akaike information criteria $\{ AIC = -2 \log L(x; \Theta) \}$, Bayesian information criteria $\{ BIC = -2 \log L(x; \Theta) + k \log L(x; \Theta) \}$, Hannan Quen information criteria $\{ HQIC = -2 \log L(x; \Theta) + 2k \log [\log(n)] \}$, where constants n and k represent the sample size and the number of parameters respectively. The goodness-of-fit statistics tests are the Anderson Darling test, Cramer Von Mises test, and the Kolmogorov Smirnov test statistic with a p -value. The existing distributions that are compared with LBI_vED are I_vED , exponential distribution (ED), weighted exponential distribution (WED), and length-biased exponential distribution ($LBED$). The $p.d.f$'s are given as,

$$ED = f(x; \Theta) = \frac{1}{\sigma} e^{-x/\sigma} \tag{43}$$

$$WED = f(x; \Theta) = \frac{(\sigma + 1)}{\sigma} \theta e^{-\theta x} (1 - e^{-\sigma \theta x}) \tag{44}$$

$$LBED = f(x; \Theta) = \frac{[\theta(\sigma + 1)]^2}{\sigma(\sigma + 2)} x e^{-\theta x} (1 - e^{-\sigma \theta x}) \tag{45}$$

Table 2: Various measures from the first data set

Models	Information Measures						
	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\rho}$	$\log L$	AIC	BIC	$HQIC$
LBI_vED	0.01975	0.13979	1.00007	7.60999	-9.21998	-3.48391	-7.03565
I_vED	0.04305	0.19339	1.00033	6.10016	-6.20032	-0.46426	-4.01600
$LBED$	1.749e+4	4.65200	-	4.11157	-3.96781	-0.39908	-2.76691
WED	1.278e-04	4.65203	-	4.11157	-4.22313	-0.39908	-2.76691
ED	0.42990	-	-	-7.78987	17.5797	19.4918	18.3078
Models	Goodness-of-fit Tests						
	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\rho}$	C_{VM}	AD	KS	p -value
LBI_vED	0.01975	0.13979	1.00007	0.14036	0.75075	0.12723	0.3932
I_vED	0.04305	0.19339	1.00033	0.17074	0.91949	0.15728	0.1685
$LBED$	1.749e+4	4.65200	-	0.13315	0.71331	0.16757	0.1206
WED	1.278e-04	4.65203	-	0.13315	0.71330	0.16761	0.1205
ED	0.42990	-	-	0.13395	0.71750	0.23317	0.0087

Table 3: Various measures from the second data set

Models	Information Measures						
	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\rho}$	$-\log L$	AIC	BIC	HQIC
LBI_vED	1.87153	0.50522	1.00011	59.884	125.768	132.197	128.297
I_vED	1.87408	0.59253	1.00026	60.663	127.326	133.756	129.855
$LBED$	1.58e-05	0.9806086	-	97.954	199.908	204.194	201.594
WED	1.717e-5	6.537e-01	-	110.35	224.696	228.983	226.382
ED	3.05930	-	-	133.45	268.892	271.035	269.734
Models	Goodness-of-fit Tests						
	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\rho}$	C_{VM}	AD	KS	<i>p-value</i>
LBI_vED	1.87153	0.50522	1.00011	0.06424	0.42637	0.10492	0.4919
I_vED	1.87408	0.59253	1.00026	0.06944	0.46604	0.11296	0.3975
$LBED$	1.58e-05	0.9806086	-	0.05891	0.36294	0.33488	1.4e-06
WED	1.717e-5	6.537e-01	-	0.05890	0.36281	0.39019	9.3e-09
ED	3.05930	-	-	0.05885	0.36229	0.48600	2.3e-13

From Table 2 and Table 3, it is visible that all the existing distributions lack superiority against the new LBI_vED , the information measures and the goodness-of-fit test results displayed are relatively lower for LBI_vED with a higher significant *p-value*. Hence, the LBI_vED is treated as the best-fitted model.

CONCLUSION

In this article, the WI_vED , a generalization of the I_vED that may be used to describe time-to-event data sets as a lifetime distribution is presented and examined. The length biased variant of the I_vED is derived as a special instance of the WI_vED and its model applicability features are thoroughly examined. We find that the LBI_vED may be used as a flexible model in place of the traditional lifetime models taken in comparison that are commonly used in the literature to describe real-life time-to-event data. As evidenced by its excellent distributional, reliability, and survival qualities, we anticipate that the proposed WI_vED and the length-biased variant of I_vED will function as a competitive model for representing data from reliability analysis, survival analysis, and other domains of the statistics.

ACKNOWLEDGEMENT(S)

The first author is very grateful to Pondicherry University for providing him with a research fellowship to carry out this work.

REFERENCES

- [1] Ahmed, M. A. (2020). On the Alpha Power Kumaraswamy Distribution: Properties, Simulation and Application. *Revista Colombiana de Estadística*, 43 (2): 285–313. <https://doi.org/10.15446/rce.v43n2.83598>
- [2] Bhat, V. A. and Pundir, S. (2022). Intervened exponential distribution: properties and applications. *Pakistan Journal of Statistics and Operation Research*, 18 (1): 71–84. <https://doi.org/10.18187/pjsor.v18i1.3829>
- [3] Bader, M. G., and Priest, A. M. (1982). Statistical aspects of fiber and bundle strength in hybrid composites. *Progress in science and engineering of composites*, 1129–1136.
- [4] Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics*, 7 (4): 579–588. <https://doi.org/10.1080/00401706.1965.10490300>

- [5] Dennis, B. and Patil, G. P. (1984). The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences*, 68 (2): 187–212. [https://doi.org/10.1016/0025-5564\(84\)90031-2](https://doi.org/10.1016/0025-5564(84)90031-2)
- [6] Finkelstein, M. (2008). Failure rate modelling for reliability and risk. London: *Springer Science & Business Media*. <https://doi.org/10.1007/978-1-84800-986-8>
- [7] Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *The Annals of Eugenics*, 6 (1): 13–25. <https://doi.org/10.1111/j.1469-1809.1934.tb02105.x>
- [8] Gupta, R. D. and Kundu, D. (2009). A new class of weighted exponential distribution. *Statistics*, 43 (6): 621–634. <https://doi.org/10.1080/02331880802605346>
- [9] Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions. *Australian & New Zealand Journal of Statistics*, 41 (2): 173–188. <https://doi.org/10.1111/1467-842X.00072>
- [10] Laird, N., Patil, G.P., and Taillie, C. (1988). Comment on S. Iyengar & J.B. Greenhouse, selection models and the file drawer problem, *Statistical Science*, 3, p. 126–128. <https://www.jstor.org/stable/2245928>
- [11] Lin, C. T., Duran, B. S., and Lewis, T. O. (1989). Inverted gamma as life distribution, *Microelectronics Reliability*, 29 (4): 619–626. [https://doi.org/10.1016/0026-2714\(89\)90352-1](https://doi.org/10.1016/0026-2714(89)90352-1)
- [12] Patil, G. P. and Taillie, C. (1989). Probing encountered data, meta analysis and weighted distribution methods, in *Statistical Data Analysis and Inference*, Y. Dodge, ed., Elsevier, Amsterdam, PP. 317–345. <https://doi.org/10.1016/B978-0-444-88029-1.50035-6>
- [13] Patil, G. P., Rao, C. R., and Zelen, M. (1988). Weighted Distributions, in *Encyclopedia of Statistical Sciences*, Vol. 9, S. Kotz & N.L. Johnson, eds, Wiley, New York, p. 565–571.
- [14] Patil, G. P. and Ord, J. K. (1976). On size-biased sampling and related form-invariant weighted distributions. *Sankhya*, 38 (1): 48–61. <https://www.jstor.org/stable/25051990>
- [15] Patil, G.P. & Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families *Biometrics*, 34: 179–189. <https://doi.org/10.2307/2530008>
- [16] Patil, G. P. (1997). Weighted distributions, in *Encyclopedia of Biostatistics*, Vol. 6, P. Armitage & T. Colton, eds, Wiley, Chichester, pp. 4735–4738.
- [17] Rényi A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Volume 1: Contributions to the Theory of Statistics. University of California Press; pp. 547–561.
- [18] Rao, C. R. (1965). On Discrete Distributions Arising Out of Methods of Ascertainment, in *Classical and Contagious Discrete Distributions*, G.P. Patil, ed., Pergamon Press and Statistical Publishing Society, Calcutta. 320–332.
- [19] Rao, C. R. (1985). Weighted Distributions Arising out of Methods of Ascertainment, in a *Celebration of Statistics*, A.C. Atkinson & S.E. Fienberg, eds, Springer-Verlag, New York, Chapter 24, p. 543–569. https://doi.org/10.1007/978-1-4613-8560-8_24
- [20] Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. New York: Springer. <https://doi.org/10.1007/978-0-387-34675-5>
- [21] Shanmugam, R. (1985). An intervened Poisson distribution and its medical application. *Biometrics*, 41 (4): 1025–1029. <https://doi.org/10.2307/2530973>
- [22] Shanmugam, R., Bartolucci, A. A., and Singh, K. P. (2002). The analysis of neurologic studies using an extended exponential model. *Mathematics and Computers in Simulation*, 59 (1-3): 81–85. [https://doi.org/10.1016/S0378-4754\(01\)00395-0](https://doi.org/10.1016/S0378-4754(01)00395-0)
- [23] Taillie, C., Patil, G. P., & Hennemuth, R. C. (1995). Modelling and analysis or recruitment distributions. *Environmental and Ecological Statistics*, 2: 315–329. <https://doi.org/10.1007/BF00569361>
- [24] Tsallis, C. (1988). Possible generalization of Boltzmann–Gibbs statistics. *Journal of Statistical Physics*, 52: 479–487. <https://doi.org/10.1007/BF01016429>

TRADE CREDIT FINANCING SCHEME ON RETAILER'S ORDERING QUANTITY FOR IMPERFECT QUALITY ITEM WITH LEARNING EFFECTS AND STOCKING STRATEGIES

A. R. NIGWAL¹, U. K. KHEDLEKAR², N. GUPTA³ and L. Sharma⁴

^{1,4}Department of Mathematics, Ujjain Engineering College, Ujjain, 456010, Madhya Pradesh India,

^{2,3}Department of Mathematics and Statistics, Dr. Harisingh Gour Vishwavidyalaya,
Sagar, 470003, Madhya Pradesh, India, (A Central University)

¹arnw@rediffmail.com

Abstract

Today's life is a age of modern life, and in the modern life for any kind of business setup, customer service, pricing, stocking strategies and trade credit financing schemes are effective, essential and survival parameters to grow the business. In this paper, we have developed, an economical order quantity model for imperfect quality product by considering retailer's stock sensitive demand of product under trade credit financing policy. Further in this paper we have studied the Learning effect on screening process on every batch of imperfect quality product. Under the trade credit financing scheme, we have considered that, the supplier proposes to the retailer, a fixed credit time period for payment and retailer also offers to his customers to a fixed credit time period of payment.

Finally an appropriate total profit function per unit time has been derived under the various trade credit financing periods of payment including various expenditure and other related parameters. A sensitivity analysis has been done to verify the optimum results and also a numerical example has been given to verify the model's outputs.

Keywords: Learning Effect, Stocking, Imperfect quality items, Trade credit policy. Screening process

1. INTRODUCTION

Generally, it has been analyzed, that a large number of consumer goods displayed in a shelf at shopping center or supermarket are connected with on sale items to induce more sales and profits. For any kind of an item, increment of shelf space induces more consumers to buy it. This possible because of its visibility, prominence or variety. Conversely, low stocks of certain paved goods might raise the feeling that they are not good and fresh. Therefore, demand is often based on inventory-level. In the last some decade, a considerable literature has been written in the operational research area on how inventory-level-dependent demand affects inventory control policies. Wu *et al.* [1] developed a inventory model for determining the optimal ordering quantity for non-instantaneous deteriorating products considering with stock-level dependent demand. They suggested that this may more beneficial for those situations in which backlogging parameter increases and decreases the order quantity. In (2005), Teng and Chang [2] Giri, and Bardhan, [3] formulated two layer supply chain coordination policy for deteriorating items with price stock level depended market demand of single product under revenue sharing contract. They concluded that the centralized system is the best strategy instead of decentralized system. developed EPQ models for deteriorating items considering with selling price and stock level dependent demand. They proposed appropriate decision for managerial activities. Ray and Chaudhuri,

[4], developed an EOQ model assuming a deterministic stock-dependent demand, incorporating with shortage, inflation and time discounting. They show through the numerical example that inventory backlogging is beneficial from both (retailer and supplier) organizational as well as economic viewpoints; Giri, and Chaudhuri, [5] designed a model considering with deterministic stock-dependent demand rate of perishable products incorporating time dependent nonlinear holding cost of the products. Parlar and Wang, [6] designed a quantity discounting decisions model for supplier and buyer relationship in which they start with Stackelberg equilibrium of the problem. They concluded quantity discount policy can be very effective in obtaining the maximum profit increase that the supplier and the retailer can possibly obtain together in certain cases. A deterministic inventory model is developed by Pal *et al.*, [7] assuming that the demand rate is stock-dependent and the items deterioration rate is constant. They highlighted the various problems related stocking of goods and further they optimized the profit function with respect to decision variable time and economical order quantity. Muth, and Spremann [8] provided a classical square root formula on the class of economical lot sizing problem considering with learning effects on the production process. Salameh *et al.* [9] developed a economical production inventory (EMQ) model which was formulated under the learning curve effect on the finite production rate. Cheng [10] Formulated an economical manufacturing quantity (EOM) under the influence of learning process. The order size is considered as to be large enough to allow the manufacturing learning phenomenon to manifest itself. The set-up cost is also assumed to reduce as a result of learning over the life of the product. Salameh and Jaber [11] developed a traditional (EOQ/EPQ) model for imperfect quality items using the (EOQ/EPQ) formulas. They also assumed that at the end of 100 percent screening work the poor-quality items are sold as a single batch with lower price.

Jaber and Guiffrida [12] Provided an (EPQ) model with rework for imperfect quality items using (WLC) wright learning curve. For this they proposed two different cases, first one is learning process adopted in production, no learning process in reworks and second one is learning process adopted in production and rework both. Eroglu and Ozdemir [13] developed an economical order quantity (EOQ) model in which they considered that each ordered lot contains some defective items incorporating with shortages at retailers end. They analyzed, how to affects optimal solution by increasing rate of percentage of defective items. They also assumed that, after 100 percent screening of each lot, the good and defective items are separated into two collection of imperfect quality and scrap items.

Jaber *et al.* [14] extended the work of Salameh and Jaber [11] by introducing the assumption that the percentage defective items per lot decreases under the learning curve (LC) effects, which was experimentally certified and validated by actual data of automotive industry. Jaber *et al.* [15] investigated the quality learning curve (QLC) for the assumption that the manufacturing process is interrupted due to maintain the quality to bring the process in control again. In this article they developed two various cases, first one is learning process is adopted in production, no learning process is adopted in reworks and second one is learning process is adopted in production and reworks both. Pan [16] analyzed the effect of learning curve on setup cost for their (CRI) model. They also assumed that the controllable lead time with the mixture of backorder and partial lost sales. Lin [17] investigated the market survey and manufacturing problem for a monopolist firm for quality and cumulative sales dependent demand. They also assumed that per unit production cost reduces with the cumulative manufacturing and learning effect.

Yoo *et al.* [18] focused on the problem that not only imperfect production process is possible but also inspection processes are always not perfect, due to generating defects and inspection errors. For this they developed a profit-maximizing economical manufacturing (EMQ) model by incorporating imperfect quality production and two-way imperfect inspection both. Sui *et al.* [19] provided a model for Vendor-Managed Inventory (VMI) system in place of traditional retailer-managed inventory applying with learning curve approach in which the supplier makes decisions of inventory management for the retailer.

Khan *et al.* [20] extended the paper of Salameh and Jaber's [11] model by introducing a new

case where there is learning in inspection. The model is more realistic than Salameh and Jaber's [11] model in the that they considered situations of lost sales and back-orders.

Wahab and. Jaber [21] developed a model for the optimal lot sizes of an item with imperfect quality which is extension of Salameh and Jaber [11] by incorporating different holding cost for good and defective items. Jaber and Khan [22] presented a model to develop a combination of performance of average processing time and process yield with respect to the number of equal batches. For the development of model, they changed the learning curve parameters in production system and rework both.

Das *et al.* [23] introduced a production-inventory model (EPQ) for deteriorating items in an indefinite conditions characterized by inflation and timed value of money by considering with static demand. They also considered that the planning interval of the business activity time is random in nature and follows exponential distribution function with a known mean. Khan *et al.* [24] extended the model of Salameh and Jaber [11] by introducing the inspection error in the time of the screening process and the probability of inspection errors is assumed to be known.

Konstantaras *et al.* [25] developed an economical order quantity (EOQ) model for imperfect quality items considering with shortages. They also assumed that the fraction of perfect quality in each shipment increases with respect to learning effect. A new and more advance inventory model for imperfect quality items has been developed by Jaggi *et al.* [26] under the situations of permissible delay in payments. Shortages are also allowed and fully backlogged, which are fulfilled during screening process. In this model It has been assumed that screening rate is enough greater than the demand rate.

Teng *et al.* [27] also proposed an economical order quantity (EPQ) model from the retailer's point of view to determine his/her optimal production lot size (EPQ) and trade credit financing period simultaneously. For this they assumed that (i) trade credit financing scheme encourage not only sales but also opportunity cost and default risk, and (ii) production cost reduces with respect to learning curve effect. Kumar *et al.* [28] proposed the effect of learning on the economical ordering policy (EPQ) for deteriorating items incorporating shortages and partially backlogging. They also assumed that due to impact learning process the ordering cost is partly constant and partly decreasing in each cycle. Further they also considered the two-level storage cost for replenishment inventor.

Givi *et al.* [29] introduced a Human Reliability Analysis (HRA) model that estimates the human error rate while performing a collectively job under the domination of learning-forgetting and fatigue-recovery. This model is enable to quantify the human error rate dynamically with time. Agi and Soni [30] proposed a deterministic demand inventory model for jointly pricing and inventory control of a perishable product considering both physical deterioration and freshness condition degradation. They considered the market demand of product as price and stock sensitive. They suggested that when the primary demand is high, then the retailer would be interested in greedy the benefit of this higher primary demand by increasing the retailing price and accelerating the inventory turnover by reducing a cycle time in place of pricing strategies. Sarkar, and Sumon [31] extended an inventory model for deteriorating items with stock-level dependent market demand. This model has been studied in that situations in which backlogging rate and deterioration rate are time varying with respect to time. Further a sensitivity analysis is analyzed of the model's outputs with respect to key parameters. Jayaswal *et al.* [32] introduced trade credit financing inventory model for imperfect quality items under the effects of learning on ordering policy. They derived average profit function per cycle time by incorporating various expenditure costs and related parameters for the retailers and the optimization process is also shown by a numerical example. Yadav *et al.* [33], developed two layer supply chain model to study the effect of imperfect quality items under the asymmetric information with market expenditure sensitive demand.

Soni, and Shah [34] developed an economical production quantity (EPQ) model for retailer's by considering partially constant and partially stock stock sensitive demand. Further they also consider a new progressive credit period. They concluded which credit period is more beneficial for business activity. Benyong and Feng [35] developed a two layer supply chain inventory model

with revenue sharing contract and service requirement under the unpredictability of supply and demand. They formulated the buyer's and supplier's optimal coordination and service requirement situations. They demonstrated how the service requirement's impacts the buyer's and supplier's decisions. Nigwal *et al.*[36] designed an EPQ model on retailer's order quantity using learning effect on screening process under trade credit financing scheme. Nigwal *et al.*[37] developed three stage price dependent trade credit policy for supplier, manufacturer and retailer.

Generally, in the traditional economical order quantity (EOQ) models, it is assumed that the retailer pay to supplier as soon as the product is received. But in the practice, supplier to stimulate sales of his products,he offers to the retailer a certain permissible delay period of payment and after end of this delayed period he charges the interest. In this chapter we have considered a two stage trade credit financing periods, in which firstly supplier offers to the retailer a permissible delay period of payment and the retailer also offers to his customers a permissible delay period of payment without interest. Further, we have also assumed that every manufacturing system may manufacture some defective and good items both. The defective items may be detected by the screening process after delivery of imperfect quality items' batches.To separate the good and defective items we apply screening process on each batches of imperfect quality items on retailer's end. Furthermore we have applied learning effects on screening process.

Learning curve (LC) or Experience curve (EC) was derived first by Wright [38] in 1936. It is a mathematical tool which relates the learning variables and cumulative quantity of units. In this chapter we study the impact of learning on screening process on imperfect quality items. Sigmoid function is the ideal shape of all other learning curves and in this paper we use Sigmoid function which is formulated as $\alpha(n) = \left(\frac{\alpha}{g+e^{\beta n}}\right)$, where $\alpha(n)$ represents defective percentage rate of item in the single batch and n represents number of batches. $\beta, g > 0$ and $a > 0$ are the learning curve parameters.

Table 1: Comparative table for contribution of different authors:

Authors	Learning Effects	Screening	Trade Credit Financing	Pricing	Stock level Strategies
Wright (1936)	✓	×	×	×	×
Muth, and Spremann (1983)	✓	×	×	×	×
Salameh <i>et al.</i> (1993)	✓	×	×	×	×
Pal <i>et al.</i> (1993)	×	×	×	✓	×
Parlar and Wang (1994)	×	×	×	✓	×
Cheng (1994)	✓	×	×	×	×
Ray, and Chaudhuri(1997)	×	×	×	✓	✓
Giri, and Chaudhuri. (1998)	×	×	×	✓	×
Salameh and Jaber (2000)	✓	✓	×	×	×
Jaber <i>et al.</i> (2004)	✓	✓	×	×	×
Teng and Chang (2005)	×	×	×	✓	✓
Wu <i>et al.</i> (2006)	×	×	×	✓	✓
Eroglu and Ozdemir(2007)	✓	✓	×	×	×
Jaber and Guiffrida (2008)	✓	✓	×	×	×
Pan (2008)	✓	×	×	×	×
Lin (2008)	✓	×	×	✓	×
Soni, and Shah (2008)	×	×	✓	✓	✓
Jaber <i>et al.</i> (2008)	✓	✓	×	×	×
Yoo <i>et al.</i>	✓	✓	×	×	×
Sui, <i>et al.</i> (2010)	✓	✓	×	×	×
Khan <i>et al.</i> (2010)	✓	✓	×	×	×
Wahab and. Jaber (2010)	✓	✓	×	×	×
Jaber and Khan (2010)	✓	×	×	×	×
Das <i>et al.</i> (2010)	✓	×	×	×	×
Khan <i>et al.</i> (2011)	×	✓	×	×	×
Giri, and Bardhan (2012)	×	×	×	✓	✓
Sarkar, and Sumon (2013)	×	×	×	✓	✓
Konstantaras <i>et al.</i> (2012)	✓	✓	×	×	×
Jaggi <i>et al.</i> (2013)	×	✓	✓	×	×

Authors	Learning Effects	Screening	Trade Credit Financing	Pricing	Stock level Strategies
Teng <i>et al.</i> (2013)	✓	×	×	×	×
Kumar <i>et al.</i> (2013)	✓	×	×	×	×
Givi <i>et al.</i> (2015)	✓	×	×	×	×
Benyong and Feng (2017)	×	×	×	✓	✓
Jayaswal <i>et al.</i> (2019)	✓	✓	✓	×	×
Maher and Soni (2020)	×	×	×	✓	✓
Nigwal <i>et al.</i> (2022)	×	×	✓	✓	✓
Nigwal <i>et al.</i> (2022)	✓	✓	✓	✓	×
This paper	✓	✓	✓	×	✓

2. THE MATHEMATICAL MODEL

I. Notations and Assumptions:

ϕ_n : Lot size of the n^{th} batch,

D : Demand rate of items in units per unit of time for perfect quality items, Where,
 $D = a + bI(t)$,

S_c : Setup cost per order,

ϕ_n : Initial stock of inventory,

C_p : Purchasing cost per unit of an item,

h : Holding cost of items per unit time,

p : Retailing price per unit of perfect quality items,

v : Retailing price (On discounted rate) per unit of defective items ($p > v$),

$\alpha(n)$: Percentage rate of defective items per batch,

T_n : Length of cycle for shipment per order,

χ : Screening rate of items per unit time ($D < \chi$),

C_s : Screening cost per unit items,

τ_n : Screening time of ϕ_n batch in planing time T_n , where, $\tau_n = \frac{\phi_n}{\chi} < T_n$,

I_e : Interest rate per unit \$ earned by retailer,

I_p : Interest rate per unit \$ paid by retailer,

TSR : Sells revenue,

TE : Total cost,

$\Pi(\phi_n)$: Retailer's total profit per unit time,

L : Length delay period of payment offered per cycle time by supplier to the retailer,

M : Length delay period of payment offered per cycle time by retailer to customers,

The following assumptions are assumed during the development of model:

- The supplier provides a fixed and predetermined credit period to settle the accounts to the supplier,
- For infinite supply rate, selling price p and optimal lot size ϕ_n , are decision variable,

- No scrape item will be obtain during the screening process,
- Screening procedure and demand of items occurs simultaneously ($D < \chi$),
- It has been assumed that each lot size contains perfect and imperfect items both,
- It has been assumed that the price of the perfect quality items is greater than the imperfect quality items,
- It has been assumed that the earned interest rate is less than the payable interest rate,
- It has been that the retailer offers a permissible delay period of payment to his customers without interest to stimulate the sales,
- We has been assumed that a limited but maximum amount of stock displayed in a super-market without leaving a negative impact on customers,
- It has been assumed that $L, M \in [0, T_n]$, only.
- During the formulation of profit, T_n is approximated by second term, because $b < 1$ and a is very large, therefore $\frac{b^2}{a^2} \approx 0$.

II. The Mathematical formulation of model

The inventory level of perfect quality items at any time t , is governed by the following differential equation:

$$\frac{dI(t)}{dt} = -(a + bI(t)), 0 \leq t \leq T_n, \quad (1)$$

with the boundary conditions: $I(0) = (1 - \alpha(n))\phi_n$, and $I(T_n) = 0$

solution of this equation : Where A is arbitrary constant, to remove the constant using the conditions $I(0) = (1 - \alpha(n))\phi_n$, then solutions becomes

$$I(t) = \frac{a}{b}(e^{-bt} - 1) + (1 - \alpha(n))\phi_n e^{-bt} \quad (2)$$

at time $t = T_n$, the T_n can be determined by the following formula

$$T_n = \frac{\log(1 - \alpha(n))\phi_n}{a} \quad (3)$$

and according to the assumptions the screening time τ_n is given by the following formula

$$\tau_n = \frac{\phi_n}{\chi} \quad (4)$$

The Sales revenue $SR = p(1 - \alpha(n))\phi_n + v\alpha(n)\phi_n$, Ordering Cost = O_c , Purchasing Cost = $P_c\phi_n$, Screening Cost = $S_c\phi_n$, Inventory Holding Costs = $h \left[\frac{1}{a}(1 - (e^{bT_n} T_n b)) + \frac{(1 - \alpha(n))\phi_n}{b}(1 - e^{-bT_n}) \right]$ and $h \left[\frac{\phi_n^2 \alpha(n)}{\chi} \right]$ Now the total expenditure per cycle is given by:

$$TE = S_c + p_c\phi_n + s_c\phi_n + = h \left[\frac{a}{b} \left(-\frac{e^{-bT_n}}{b} - T_n \right) + \frac{(1 - \alpha(n))\phi_n e^{-bT_n}}{b} + \frac{a}{b^2} + \frac{(1 - \alpha(n))\phi_n}{b} \right] \quad (5)$$

At a time of each replenishment a fixed and certain credit period of payment L is provided by supplier to the retailer and similarly a fixed and certain period of payment M is also provided by retailer to their customers. Where $M, L \in (0, T_n)$ and $\tau_n \neq T_n$ There are four different cases available for retailer and their customers.

$$(1) \tau_n \geq L \geq M \quad (2) \tau_n \geq M \geq L \quad (3) L \geq M \geq \tau_n \quad (4) M \geq L \geq \tau_n$$

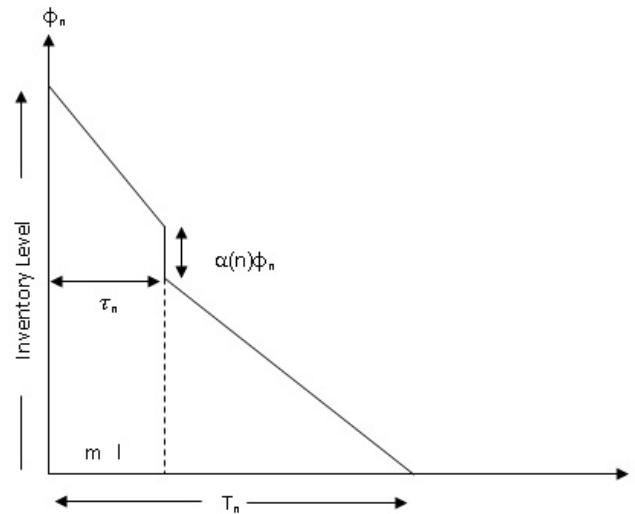


Figure 1: Inventory Level Chart

the retailer's whole profit $\Pi_i(\phi_n)$, $i=1,2,3, 4$ per unit of time can be defined as:

$$\Pi_i(\phi_n) = TSR_i - TE_i + (\text{Earned Interest}) - (\text{Paid Interest}), \text{ where } i=1, 2, 3, 4 \quad (6)$$

Case 1: $\tau_n \geq L \geq M$

As assumed credit periods, firstly we consider L is greater than M as per depicted in Figure 1, the earned interest and paid interest for this case is estimated as follows:

Earned Interest by retailer:

$$EI_r = I_e p [a + b(1 - \alpha(n))\phi(L - M)] \quad (7)$$

Paid Interest by retailer:

$$PI_r = I_p C_p [a + b(1 - \alpha(n))\phi_n b(L - \frac{(1 - \alpha(n))}{a}\phi_n)] + C_p I_p \alpha(n)\phi_n (\frac{\phi_n}{\chi} - L) \quad (8)$$

and total profit function per unit time may be defined as follow:

$$\Pi_1(\phi_n) = \frac{TSR_1 - TE_1 + (\text{Earned Interest}) - (\text{Paid Interest})}{T_n} \quad (9)$$

Hence the total profit function per unit time is:

$$\begin{aligned} \Pi_1(\phi_n) = & pa + \frac{v\alpha(n)a}{(1 - \alpha(n))} + \frac{I_e p a^2}{(1 - \alpha(n))\phi_n} + ba(L - M)I_e p \\ & - \frac{(C_s + C_p\phi_n + S_c\phi_n)a}{(1 - \alpha(n))\phi_n} - \frac{2hb}{a} - h(1 - \alpha(n))\phi_n - \frac{h\phi_n\alpha(n)}{(1 - \alpha(n))\chi} \\ & - I_p C_p \left[\frac{a}{(1 - \alpha(n))\phi_n} + ab \left(L - \frac{(1 - \alpha(n))}{a}\phi_n \right) \right] + \frac{C_p I_p \alpha(n)a}{(1 - \alpha(n))} \left(\frac{\phi_n}{\chi} - L \right). \quad (10) \end{aligned}$$

Theorem 2.1. Retailer's profit function is an optimum at retailer's ordering quantity ϕ_n^* , where ϕ_{n1}^* is given by the following equation:

$$\phi_{n1}^* = \sqrt{\frac{C_s a - I_e p a^2 + C_p I_p a^2}{h(1 - \alpha(n))^2 + \frac{h\alpha(n)}{\chi} - C_p I_p b(1 - \alpha(n))^2 + \frac{C_p I_p \alpha(n)a}{\chi}}}. \quad (11)$$

Proof. On differentiating the equations (10) with respect to ϕ_n , we get

$$\begin{aligned} \frac{d\Pi_1(\phi_n)}{d\phi_n} &= -\frac{I_e pa^2}{(1-\alpha(n))\phi_n^2} + \frac{C_s a}{(1-\alpha(n))\phi_n^2} - h(1-\alpha(n)) - \frac{h\alpha(n)}{\chi(1-\alpha(n))} \\ &+ \frac{C_p I_p a^2}{(1-\alpha(n))\phi_n^2} + C_p I_p b(1-\alpha(n)) - \frac{C_p I_p \alpha(n)a}{(1-\alpha(n))\chi} \end{aligned} \quad (12)$$

As per optimality condition, on equating to zero the above equation (12), yields

$$-I_e pa^2 + C_s a - h(1-\alpha(n))^2 \phi^2 - \frac{h\alpha(n)\phi^2}{\chi} + C_p I_p \left(a^2 + b(1-\alpha(n))^2 \phi^2 - \frac{\alpha(n)a\phi^2}{\chi} \right) = 0$$

On solving the above equation we obtain the equation (11) □

Theorem 2.2. *As per optimality condition, at the point of optimality, the second derivative $\frac{d^2\Pi_1}{d\phi_n^2}$ is always negative if*

$$\frac{2I_e pa^2}{(1-\alpha(n))\phi_n^3} - \frac{2C_s a}{(1-\alpha(n))\phi_n^3} - \frac{C_p I_p \alpha(n)a^2}{(1-\alpha(n))^3} < 0 \quad (13)$$

Proof. On differentiating again the equation (12) with respect to ϕ_n , we get the second order derivative is

$$\frac{d^2\Pi_1}{d\phi_n^2} = \frac{2I_e pa^2}{(1-\alpha(n))\phi_n^3} - \frac{2C_s a}{(1-\alpha(n))\phi_n^3} - \frac{C_p I_p \alpha(n)a^2}{(1-\alpha(n))^3} \quad (14)$$

As per assumptions all the terms $\frac{2I_e pa^2}{(1-\alpha(n))\phi_n^3}$, $\frac{2C_s a}{(1-\alpha(n))\phi_n^3}$ and $\frac{C_p I_p \alpha(n)a^2}{(1-\alpha(n))^3}$ are always positive, and therefore by numerical analysis

$$\frac{2I_e pa^2}{(1-\alpha(n))\phi_n^3} - \frac{2C_s a}{(1-\alpha(n))\phi_n^3} - \frac{C_p I_p \alpha(n)a^2}{(1-\alpha(n))^3} < 0. \quad \square$$

Case 2: $\tau_n \geq M \geq L$

As assumed credit periods, we consider M is greater than L as per depicted in the Figure 2, the earned interest and paid interest for this case is estimated as follows:

Earned Interest by retailer:

$$EI_r = 0 \quad (15)$$

Paid Interest by retailer:

$$PI_r = I_p C_p \left[a + b(1-\alpha(n))\phi_n b \left(L - \frac{(1-\alpha(n))}{a} \phi_n \right) \right] + C_p I_p \alpha(n) \phi_n \left(\frac{\phi_n}{\chi} - L \right) \quad (16)$$

and total profit function per unit time may be defined as follow:

$$\Pi_2(\phi_n) = \frac{TSR_2 - TE_2 + (\text{Earned Interest}) - (\text{Paid Interest})}{T_n} \quad (17)$$

Hence the total profit function per unit time is:

$$\begin{aligned} \Pi_2(\phi_n) &= pa + \frac{v\alpha(n)a}{(1-\alpha(n))} - \frac{(C_s + C_p \phi_n + S_c \phi_n)a}{(1-\alpha(n))\phi_n} - \frac{2hb}{a} - h(1-\alpha(n))\phi_n \\ &- \frac{h\phi_n \alpha(n)}{(1-\alpha(n))\chi} - I_p C_p \left[\frac{a^2}{(1-\alpha(n))\phi_n} + ab \left(\frac{(1-\alpha(n))\phi_n}{a} - M \right) \right] \\ &- \frac{C_p I_p \alpha(n)a}{(1-\alpha(n))} \left(\frac{\phi_n}{\chi} - L \right). \end{aligned} \quad (18)$$

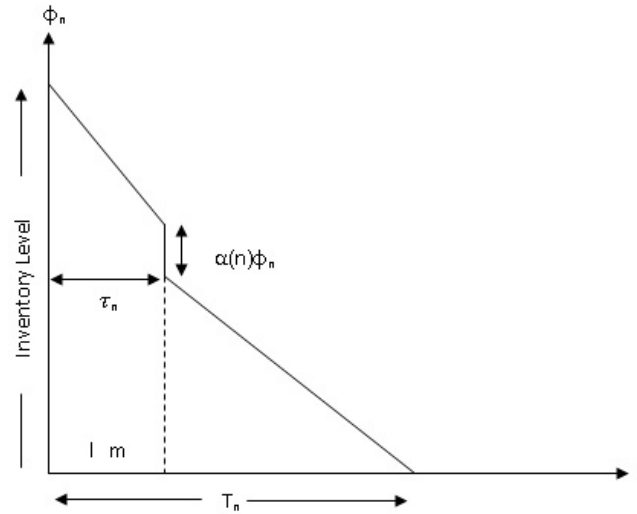


Figure 2: Inventory Level Chart

Theorem 2.3. Retailer's profit function is an optimum at retailer's ordering quantity ϕ_{n2}^* , where ϕ_n^* is given by the following equation:

$$\phi_{n2}^* = \sqrt{\frac{C_s a + C_p I_p a^2}{h(1 - \alpha(n))^2 + \frac{h\alpha(n)}{\chi} + C_p I_p b(1 - \alpha(n))^2 + \frac{C_p I_p \alpha(n)a}{\chi}}}. \quad (19)$$

Proof. On differentiating the equations (18) with respect to ϕ_n , we get

$$\begin{aligned} \frac{d\Pi_2(\phi_n)}{d\phi_n} &= \frac{C_s a}{(1 - \alpha(n))\phi_n^2} - h(1 - \alpha(n)) - \frac{h\alpha(n)}{\chi(1 - \alpha(n))} \\ &+ \frac{C_p I_p a^2}{(1 - \alpha(n))\phi_n^2} - C_p I_p b(1 - \alpha(n)) - \frac{C_p I_p \alpha(n)a}{(1 - \alpha(n))\chi} \end{aligned} \quad (20)$$

As per optimality condition, on equating to zero the above equation (20), yields

$$C_s a - h(1 - \alpha(n))^2 \phi^2 - \frac{h\alpha(n)\phi^2}{\chi} + C_p I_p \left(a^2 - b(1 - \alpha(n))^2 \phi^2 - \frac{\alpha(n)a\phi^2}{\chi} \right) = 0$$

On solving the above equation we obtain the value of ϕ_{n2}^* given in the equation (19) □

Theorem 2.4. As per optimality condition, at the point of optimality, the second derivative $\frac{d^2\Pi_2}{d\phi_n^2}$ is always negative if

$$\frac{2C_s a}{(1 - \alpha(n))\phi_n^3} + \frac{C_p I_p a^2}{(1 - \alpha(n))^3} > 0 \quad (21)$$

Proof. On differentiating again the equations (20) with respect to ϕ_n , we get the second order derivative is

$$\frac{d^2\Pi_2}{d\phi_n^2} = -\frac{2C_s a}{(1 - \alpha(n))\phi_n^3} - \frac{C_p I_p a^2}{(1 - \alpha(n))^3} \quad (22)$$

As per article's assumptions all the terms $\frac{2C_s a}{(1 - \alpha(n))\phi_n^3}$ and $\frac{C_p I_p a^2}{(1 - \alpha(n))^3}$ are always positive, and therefore

$$-\frac{2C_s a}{(1 - \alpha(n))\phi_n^3} - \frac{C_p I_p a^2}{(1 - \alpha(n))^3} < 0. \quad \square$$

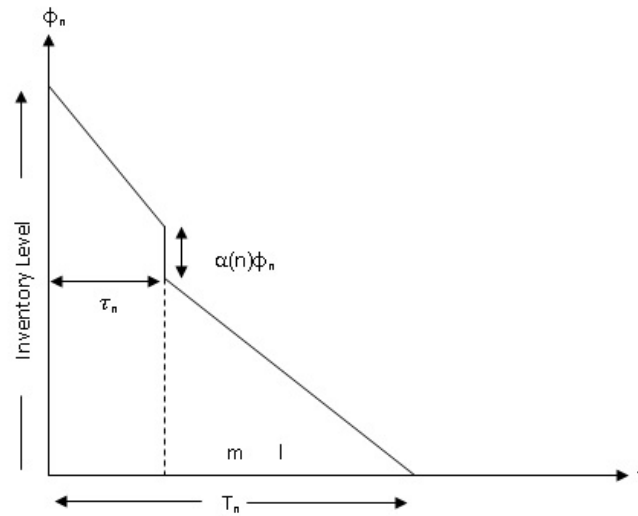


Figure 3: Inventory Level Chart

Case 3: $L \geq M \geq \tau_n$

As assumed credit periods, we consider L is greater than τ and M as pr depicted in the Figure 3, the earned interest and paid interest for this case is estimated as follows:

Earned Interest by retailer:

$$EI_r = I_e p [a + b(1 - \alpha(n))\phi(M - L)] + v I_e \alpha(n)\phi(L - \tau) \quad (23)$$

Paid Interest by retailer:

$$PI_r = I_p C_p \left[a + b(1 - \alpha(n))\phi_n \left(\frac{(1 - \alpha(n))\phi_n}{a} - L \right) \right] \quad (24)$$

and total profit function per unit time may be defined as follow:

$$\Pi_3(\phi_n) = \frac{TSR_3 - TE_3 + (\text{Earned Interest}) - (\text{Paid Interest})}{T_n} \quad (25)$$

Hence the total profit function per unit time is:

$$\begin{aligned} \Pi_3(\phi_n) &= pa + \frac{v\alpha(n)a}{(1 - \alpha(n))} + \frac{v I_e \alpha(n)a}{(1 - \alpha(n))} \left(L - \frac{\phi_n}{\chi} \right) - \frac{(C_s + C_p \phi_n + S_c \phi_n)a}{(1 - \alpha(n))\phi_n} - \frac{2hb}{a} \\ &- I_p C_p \left[a - ab \left(L + \frac{a^2 L}{1 - \alpha(n)\phi_n} \right) + b((1 - \alpha(n))\phi_n) \right] - h(1 - \alpha(n))\phi_n \\ &- \frac{h\phi_n \alpha(n)a}{(1 - \alpha(n))\chi} + \frac{C_p I_p \alpha(n)a}{(1 - \alpha(n))} \left(\frac{\phi_n}{\chi} - L \right). \end{aligned} \quad (26)$$

Theorem 2.5. Retailer's profit function is an optimum at retailer's ordering quantity ϕ_{n3}^* , where ϕ_{n3}^* is given by the following equation:

$$\phi_{n3}^* = \sqrt{\frac{C_s a + C_p I_p L a^2}{h(1 - \alpha(n))^2 + \frac{h\alpha(n)a}{\chi} + C_p I_p b(1 - \alpha(n))^2 + \frac{v\alpha(n)I_e a}{\chi}}}. \quad (27)$$

Proof. On differentiating the equations (26) with respect to ϕ_n , we get

$$\begin{aligned} \frac{d\Pi_3(\phi_n)}{d\phi_n} &= -\frac{I_e p a^2}{(1 - \alpha(n))\phi_n^2} + \frac{C_s a}{(1 - \alpha(n))\phi_n^2} - h(1 - \alpha(n)) - \frac{h\alpha(n)}{\chi(1 - \alpha(n))} \\ &+ \frac{C_p I_p a^2}{(1 - \alpha(n))\phi_n^2} + C_p I_p b(1 - \alpha(n)) - \frac{C_p I_p \alpha(n)a}{(1 - \alpha(n))\chi} \end{aligned} \quad (28)$$

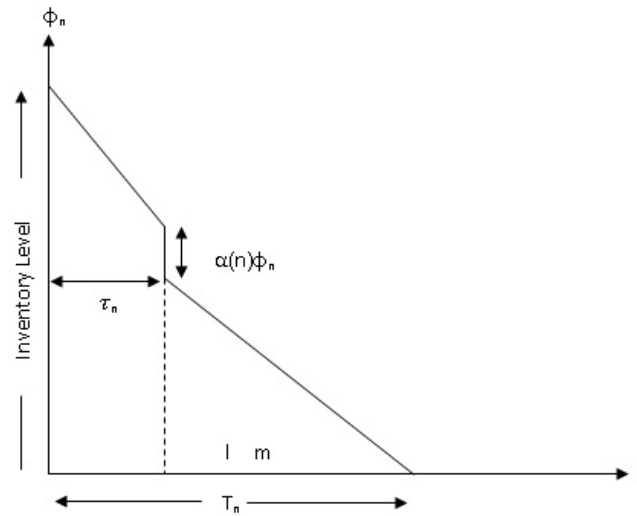


Figure 4: Inventory Level Chart

As per optimality condition, on equating to zero the above equation (28), yields

$$C_s a + a^2 C_p I_p L - h(1 - \alpha(n))^2 \phi_n^2 - \frac{h\alpha(n)\phi_n^2 a}{\chi} - C_p I_p b(1 - \alpha(n))^2 \phi_n^2 - \frac{v I_e \alpha a \phi^2}{\chi} = 0 \quad (29)$$

On solving the above equation we obtain the equation (27) □

Theorem 2.6. As per optimality condition, at the point of optimality, the second derivative $\frac{d^2 \Pi_3}{d\phi_n^2}$ is always positive if

$$\frac{2C_s a}{(1 - \alpha(n))\phi_n^3} + \frac{2C_p I_p a^2 L}{(1 - \alpha(n))\phi_n^3} > 0 \quad (30)$$

Proof. On differentiating again the equations (28) with respect to ϕ_n , we get the second order derivative is

$$\frac{d^2 \Pi_1}{d\phi_n^2} = -\frac{2C_s a}{(1 - \alpha(n))\phi_n^3} - \frac{2C_p I_p a^2 L}{(1 - \alpha(n))\phi_n^3} \quad (31)$$

As per assumptions all the terms $\frac{2C_s a}{(1 - \alpha(n))\phi_n^3}$ and $\frac{2C_p I_p a^2 L}{(1 - \alpha(n))\phi_n^3}$ are always positive, and therefore,

$$-\frac{2C_s a}{(1 - \alpha(n))\phi_n^3} - \frac{2C_p I_p a^2 L}{(1 - \alpha(n))\phi_n^3} < 0. \quad \square$$

Case 4: $M \geq L \geq \tau_n$

As assumed credit periods, we consider M is greater than L and τ as per depicted in the Figure 4, the earned interest and paid interest for this case is estimated as follows:

Earned Interest by retailer:

$$EI_r = I_e p [a + b(1 - \alpha(n))\phi_n] L + v I_e \alpha(n)\phi_n(L - \tau_n) \quad (32)$$

Paid Interest by retailer:

$$PI_r = I_p C_p \left[a + b(1 - \alpha(n))\phi_n \left(\frac{(1 - \alpha(n))\phi_n}{a} - L \right) \right] \quad (33)$$

and total profit function per unit time may be defined as follow:

$$\Pi_4(\phi_n) = \frac{TSR_4 - TE_4 + (\text{Earned Interest}) - (\text{Paid Interest})}{T_n} \quad (34)$$

Hence the total profit function per unit time is:

$$\begin{aligned} \Pi_4(\phi_n) &= pa + \frac{v\alpha(n)a}{(1-\alpha(n))} - C_p I_p (b(1-\alpha(n))\phi_n - aL) - \frac{(C_s + C_p\phi_n + S_c\phi_n)a}{(1-\alpha(n))\phi_n} \\ &- \frac{2hb}{a} - h(1-\alpha(n))\phi_n - \frac{h\phi_n\alpha(n)a}{(1-\alpha(n))\chi} - I_e pLa \left[\frac{a}{(1-\alpha(n))\phi_n} + b \right] \\ &+ \frac{vI_e\alpha(n)a}{(1-\alpha(n))} \left(L - \frac{\phi_n}{\chi} \right) - C_p I_p a \left(1 + \frac{aL}{(1-\alpha(n))\phi_n} \right). \end{aligned} \quad (35)$$

Theorem 2.7. Retailer's profit function is an optimum at retailer's ordering quantity ϕ_{n4}^* , where ϕ_{n4}^* is given by the following equation:

$$\phi_{n4}^* = \sqrt{\frac{C_s a - I_e p a^2 L - C_p I_p a^2 L}{h(1-\alpha(n))^2 + \frac{h\alpha(n)a}{\chi} + C_p I_p b(1-\alpha(n))^2 + \frac{v\alpha(n)}{\chi}}}. \quad (36)$$

Proof. On differentiating the equations (35) with respect to ϕ_n , we get

$$\begin{aligned} \frac{d\Pi_4(\phi_n)}{d\phi_n} &= \frac{C_s a}{(1-\alpha(n))\phi_n^2} - h(1-\alpha(n)) - \frac{h\alpha(n)a}{\chi(1-\alpha(n))} \\ &- \frac{I_e p a^2 L}{(1-\alpha(n))\phi_n^2} - C_p I_p b(1-\alpha(n)) - \frac{v\alpha(n)a}{(1-\alpha(n))\chi} - \frac{I_p C_p L}{(1-\alpha(n))\phi_n^2}. \end{aligned} \quad (37)$$

As per optimality condition, on equating to zero the above equation (37), yields

$$C_s a - h(1-\alpha(n))^2\phi^2 - I_e p a^2 L - \frac{\alpha(n)\phi^2}{\chi} (ah + I_e v) - C_p I_p (a^2 + b(1-\alpha(n))^2\phi^2) = 0 \quad (38)$$

On solving the above equation we obtain the equation (36) □

Theorem 2.8. As per optimality condition, at the point of optimality, the second derivative $\frac{d^2\Pi_4}{d\phi_n^2}$ is always negative if

$$\frac{2I_e p a^2 L}{(1-\alpha(n))\phi_n^3} - \frac{2C_s a}{(1-\alpha(n))\phi_n^3} - \frac{2C_p I_p a^2 L}{(1-\alpha(n))\phi_n^3} < 0 \quad (39)$$

Proof. On differentiating again the equation (37) with respect to ϕ_n , we get the second order derivative is

$$\frac{d^2\Pi_1}{d\phi_n^2} = \frac{2I_e p a^2 L}{(1-\alpha(n))\phi_n^3} - \frac{2C_s a}{(1-\alpha(n))\phi_n^3} - \frac{2C_p I_p a^2 L}{(1-\alpha(n))\phi_n^3} \quad (40)$$

As per assumptions all the terms $\frac{2I_e p a^2 L}{(1-\alpha(n))\phi_n^3}$, $\frac{2C_s a}{(1-\alpha(n))\phi_n^3}$ and $\frac{2C_p I_p a^2 L}{(1-\alpha(n))\phi_n^3}$ are always positive, and therefore by numerical analysis

$$\frac{2I_e p a^2 L}{(1-\alpha(n))\phi_n^3} - \frac{2C_s a}{(1-\alpha(n))\phi_n^3} - \frac{2C_p I_p a^2 L}{(1-\alpha(n))\phi_n^3} < 0.$$

□

3. NUMERICAL EXAMPLES

Case-1:

We have considered the following data set of input parameters is given as: $\alpha = 160$ units/unit time, $\beta = 1$, $S_c = \$0.5$, $h = \$0.8$ unit/ unit time, $C_p = \$150$ /unit, $v = \$45$ per/unit, $\chi = 5000$ units, $C_s = \$100$ /unit, $I_e = 0.003$ /unit time, $I_p = \$0.004$ /unit time, $\alpha(n) = 0.1599$, $n = 1$, $a = 455$, $b = 0.25$, $g = 999$ $L = 0.06$ /unit time, $M = 0.05$ /unit time.

Following the proposed restrictions for this case we may get the optimal ordering quantity (OOQ) $\phi_n = 339$ units per unit time, $p = 190$ and after substituting these optimum values ϕ_n , and p into the equation (10) we get the retailer's profit $\Pi(\phi_n) = 18266$, screening time $\tau_n = 0.06$ per unit time, and time interval is $T_n = 0.250$ in year.

Case-2:

We have considered the following data set of input parameters is given as: $\alpha = 160$ units/unit time, $\beta = 1$, $S_c = \$0.5$, $h = \$0.8$ unit/ unit time, $C_p = \$150$ /unit, $v = \$45$ per/unit, $\chi = 5000$ units, $C_s = \$100$ /unit, $I_e = 0.003$ /unit time, $I_p = \$0.004$ /unit time, $\alpha(n) = 0.1599$, $n = 1$, $a = 455$, $b = 0.25$, $g = 999$ $L = 0.06$ /unit time, $M = 0.07$ /unit time.

Following the proposed restrictions for this case we may get the optimal ordering quantity (OQ) $\phi_n = 423$ units per unit time, $p = 190$ and after substituting these optimum values ϕ_n , and p into the equation (18) we get the retailer's profit $\Pi(\phi_n) = 17174$, screening time $\tau_n = 0.08$ per unit time, and time interval $T_n = 0.363$ in year.

Case-3:

We have considered the following data set of input parameters is given as: $\alpha = 160$ units/unit time, $\beta = 1$, $S_c = \$0.5$, $h = \$0.8$ unit/ unit time, $C_p = \$190$ /unit, $v = \$45$ per/unit, $\chi = 5000$ units, $C_s = \$100$ /unit, $I_e = 0.003$ /unit time, $I_p = \$0.004$ /unit time, $\alpha(n) = 0.1599$, $\eta = 1$, $a = 455$, $b = 0.25$, $g = 999$ $L = 0.09$ /unit time, $M = 0.07$ /unit time.

Following the proposed restrictions for this case we may get the optimal ordering quantity (OQ) $\phi_n = 244$ units per unit time, $p = 190$ and after substituting these optimum values ϕ_n , and p into the equation (26) we get the retailer's profit $\Pi(\phi_n) = 39979$, screening time $\tau_n = 0.049$ per unit time, and time interval $T_n = 0.219$ in year.

Case-4:

We have considered the following data set of input parameters is given as: $\alpha = 160$ units/unit time, $\beta = 1$, $S_c = \$0.5$, $h = \$0.8$ unit/ unit time, $C_p = \$150$ /unit, $v = \$45$ per/unit, $\chi = 5000$ units, $C_s = \$100$ /unit, $I_e = 0.003$ /unit time, $I_p = \$0.004$ /unit time, $\alpha(n) = 0.1599$, $n = 1$, $a = 455$, $b = 0.25$, $g = 999$ $L = 0.06$ /unit time, $M = 0.09$ /unit time.

Following the proposed restrictions for this case we may get the optimal ordering quantity (OQ) $\phi_n = 181$ units per unit time, $p = 190$ and after substituting these optimum values ϕ_n , and p into the equation (35) we get the retailer's profit $\Pi(\phi_n) = 17364$, screening time $\tau_n = 0.036$ per unit time, and time interval $T_n = 0.164$ in year.

4. SENSITIVITY ANALYSIS

A perusal of Table 2, shows that if the learning ability of the workers is 1 then 16 shipments will be required for the workers to acquire the proficiency. And, if the learning ability of the workers is 1.2 then 16 shipments will be required for the workers to acquire proficiency of screening process. And again, if the learning ability of the workers is 1.4 then 13 shipments will be required for the workers to acquire proficiency of screening process. In addition, as we increase the learning

ability of the workers, the lot size, profit function, screening time and planning time increase, while the number of defective items and number of shipments decrease. The same situations are applying for table numbers 3, 4 and 5 as well. A learning efficiency of worker's not only reduces the number of shipments but also increases the profit per unit time. Table 6 shows the comparative study of various cases. Observation of tables 6, 7, 8 and 9 reveals that Case 3. gives better results per unit time and there is no considerable effect of Beta on outputs of various cases.

Table 2: Impact of Learning Rate and No. of Shipments on Outputs (Case 1)

Learning Rate $\beta = 1$						
No. of Shipment (n)	ϕ_n	% of good item	τ_n	T_n	Profit	No of Shipment Required for Perfection
1	339	84.03%	0.068	0.253	9140	
4	336	84.81%	0.067	0.252	9670	
7	307	92.37%	0.061	0.251	14298	
10	284	99.31%	0.057	0.250	17930	
13	282	99.96%	0.056	0.250	18249	$n=16$
16	282	100.00%	0.056	0.250	18265	
19	282	100.00%	0.056	0.250	18266	
22	282	100.00%	0.056	0.250	18266	
25	282	100.00%	0.056	0.250	18266	
Learning Rate $\beta = 1.2$						
1	339	84.04%	0.068	0.253	9147	
4	332	85.72%	0.066	0.252	10269	
7	291	97.06%	0.058	0.251	16813	
10	282	99.90%	0.056	0.250	18219	
13	282	100.00%	0.056	0.250	18265	$n=13$
16	282	100.00%	0.056	0.250	18266	
19	282	100.00%	0.056	0.250	18266	
22	282	100.00%	0.056	0.250	18266	
25	282	100.00%	0.056	0.250	18266	
Learning Rate $\beta = 1.4$						
1	282	99.99%	0.056	0.250	18260	
4	282	100.00%	0.056	0.250	18266	
7	282	100.00%	0.056	0.250	18266	
10	282	100.00%	0.056	0.250	18266	
13	282	100.00%	0.056	0.250	18266	$n=4$
16	282	100.00%	0.056	0.250	18266	
19	282	100.00%	0.056	0.250	18266	
22	282	100.00%	0.056	0.250	18266	
25	282	100.00%	0.056	0.250	18266	

Table 3: Impact of Learning Rate and No. of Shipments on Outputs (Case 2)

Learning Rate $\beta = 1$						
No. of Shipment (n)	ϕ_n	% of good item	τ_n	T_n	Profit	No of Shipment Required for Perfection
1	500	84.03%	0.100	0.361	8047	
4	495	84.81%	0.099	0.361	8577	
7	456	92.37%	0.091	0.362	13206	
10	426	99.31%	0.085	0.363	16838	
13	423	99.96%	0.085	0.363	17157	$n=16$
16	423	100.00%	0.085	0.363	17173	
19	423	100.00%	0.085	0.363	17174	

No. of Shipment (n)	ϕ_n	% of good item	τ_n	T_n	Profit	No of Shipment Required for Perfection
22	423	100.00%	0.085	0.363	17174	
25	423	100.00%	0.085	0.363	17174	
Learning Rate $\beta = 1.2$						
1	500	84.04%	0.100	0.361	8054	
4	490	85.72%	0.098	0.361	9176	
7	435	97.06%	0.087	0.362	15721	
10	423	99.90%	0.085	0.363	17127	
13	423	100.00%	0.085	0.363	17173	$n=13$
16	423	100.00%	0.085	0.363	17174	
19	423	100.00%	0.085	0.363	17174	
22	423	100.00%	0.085	0.363	17174	
25	423	100.00%	0.085	0.363	17174	
Learning Rate $\beta = 1.4$						
1	500	84.05%	0.100	0.361	8062	
4	481	87.40%	0.096	0.361	10250	
7	426	99.16%	0.085	0.363	16767	
10	423	99.99%	0.085	0.363	17168	
13	423	100.00%	0.085	0.363	17174	$n=13$
16	423	100.00%	0.085	0.363	17174	
19	423	100.00%	0.085	0.363	17174	
22	423	100.00%	0.085	0.363	17174	
25	423	100.00%	0.085	0.363	17174	

Table 4: Impact of Learning Rate and No. of Shipments on Outputs (Case 3)

Learning Rate $\beta = 1$						
No. of Shipment (n)	ϕ_n	% of good item	τ_n	T_n	Profit	No of Shipment Required for Perfection
1	288	84.03%	0.058	0.217	35175	
4	285	84.81%	0.057	0.217	35454	
7	263	92.37%	0.053	0.218	37890	
10	246	99.31%	0.049	0.219	39802	
13	244	99.96%	0.049	0.219	39970	$n=16$
16	244	100.00%	0.049	0.219	39979	
19	244	100.00%	0.049	0.219	39979	
22	244	100.00%	0.049	0.219	39979	
25	244	100.00%	0.049	0.219	39979	
Learning Rate $\beta = 1.2$						
1	288	84.04%	0.058	0.217	35179	
4	282	85.72%	0.056	0.217	35770	
7	251	97.06%	0.050	0.218	39214	
10	244	99.90%	0.049	0.219	39955	
13	244	100.00%	0.049	0.219	39979	$n=13$
16	244	100.00%	0.049	0.219	39979	
19	244	100.00%	0.049	0.219	39979	
22	244	100.00%	0.049	0.219	39979	
25	244	100.00%	0.049	0.219	39979	
Learning Rate $\beta = 1.4$						
1	288	84.05%	0.058	0.217	35183	
4	277	87.40%	0.055	0.217	36335	
7	246	99.16%	0.049	0.219	39765	
10	244	99.99%	0.049	0.219	39976	
13	244	100.00%	0.049	0.219	39979	$n=13$

No. of Shipment (n)	ϕ_n	% of good item	τ_n	T_n	Profit	No of Shipment Required for Perfection
16	244	100.00%	0.049	0.219	39979	
19	244	100.00%	0.049	0.219	39979	
22	244	100.00%	0.049	0.219	39979	
25	244	100.00%	0.049	0.219	39979	

Table 5: Impact of Learning Rate and No. of Shipments on Outputs (Case 4)

Learning Rate $\beta = 1$						
No. of Shipment (n)	ϕ_n	% of good item	τ_n	T_n	Profit	No of Shipment Required for Perfection
1	213	84.03%	0.043	0.163	8238	
4	211	84.81%	0.042	0.163	8767	
7	195	92.37%	0.039	0.164	13396	
10	182	99.31%	0.036	0.164	17028	
13	181	99.96%	0.036	0.164	17347	$n = 16$
16	181	100.00%	0.036	0.164	17364	
19	181	100.00%	0.036	0.164	17364	
22	181	100.00%	0.036	0.164	17364	
25	181	100.00%	0.036	0.164	17364	
Learning Rate $\beta = 1.2$						
1	213	84.04%	0.043	0.163	8244	
4	209	85.72%	0.042	0.163	9366	
7	186	97.06%	0.037	0.164	15911	
10	181	99.90%	0.036	0.164	17317	
13	181	100.00%	0.036	0.164	17363	$n=13$
16	181	100.00%	0.036	0.164	17364	
19	181	100.00%	0.036	0.164	17364	
22	181	100.00%	0.036	0.164	17364	
25	181	100.00%	0.036	0.164	17364	
Learning Rate $\beta = 1.4$						
1	213	84.05%	0.043	0.163	8252	
4	205	87.40%	0.041	0.163	10440	
7	182	99.16%	0.036	0.164	16957	
10	181	99.99%	0.036	0.164	17358	
13	181	100.00%	0.036	0.164	17364	$n=13$
16	181	100.00%	0.036	0.164	17364	
19	181	100.00%	0.036	0.164	17364	
22	181	100.00%	0.036	0.164	17364	
25	181	100.00%	0.036	0.164	17364	

Table 6: Comparative Table for Case 1

β	No. of ship. required for proficiency	τ_n	T_n	Profit
1	16	0.056	0.363	18265
1.2	16	0.056	0.363	18266
1.4	4	0.056	0.363	18266

Table 7: Comparative Table for Case 2

β	No. of ship. required for proficiency	τ_n	T_n	Profit
1	16	0.085	0.363	17173
1.2	13	0.085	0.363	17173
1.4	13	0.085	0.363	17174

Table 8: Comparative Table for Case 3

β	No. of ship. required for proficiency	τ_n	T_n	Profit
1	16	0.049	0.219	39979
1.2	13	0.049	0.219	39979
1.4	13	0.049	0.219	39979

Table 9: Comparative Table for Case 4

β	No. of ship. required for proficiency	τ_n	T_n	Profit
1	16	0.036	0.164	17364
1.2	13	0.036	0.164	17364
1.4	13	0.036	0.164	17364

5. CONCLUSION

In this article we have optimized the retailer's ordering quantity for imperfect quality items with learning effects on screening process under the trade credit financing scheme. The main focus of this study is that how affects the retailer's ordering quantity when stocking strategies is beneficial for market situations. The various interval of credit periods have been analyzed and verified through the different numerical examples. A comparative study has been done through the numerical examples. and we have concluded that Case 3. is more beneficial for this type of trade credit financing strategies. This article suggests that, those item which sale depends on stocking may earn more and more profit by increasing T_n , ϕ_n , and τ_n . Article also suggests that, in the financing policy keep always $\tau_n > M > L$ for better outputs. This article may be extended by incorporating the rework process on defective items. One can also extended this article by incorporating procurement cost on ordering size of items. One can also extended this article by incorporating expected quantity of defective items.

REFERENCES

- [1] Wu K. S., Liang Y. O., and Chih T. Y. (2006). An Optimal Replenishment Policy for Non-Instantaneous Deteriorating Items with Stock-Dependent Demand and Partial Backlogging. *International Journal Production Economics*, **101**(2): 69 - 84.
- [2] Teng, J. T. and Chang T. C. (2005). Economic Production Quantity Models for Deteriorating Items with Price- and Stock-Dependent Demand. *Computers & Operations Research*, **32**(2): 297- 308.
- [3] Giri, B. C., and Bardhan S. (2012). Supply Chain Coordination for a Deteriorating Item with Stock and Price-Dependent Demand under Revenue Sharing Contract. *International Transactions Operational Research*, **19**(5): 753 -768.
- [4] Ray, J., and Chaudhuri K. S. (1997). An EOQ Model with Stock-Dependent Demand, Shortage, Inflation and Time Discounting. *International Journal Production Economics*, **53**(2): 71 - 80.
- [5] Giri, B. C., and Chaudhuri K. S. (1998). Deterministic Models of Perishable Inventory with Stock-Dependent Demand Rate and Nonlinear Holding Cost. *European Journal Operational Research*, **105**(3): 67 - 74.
- [6] Parlar, M., and Qinan W. (1994). Discounting Decisions in a Supplier-Buyer Relationship with a Linear Buyer's Demand. *IIE Transactions*, (Institute of Industrial Engineers). **26**(2): 34 - 41.

- [7] Pal, S., Goswami A. , and Chaudhuri K. S. (1993). A Deterministic Inventory Model for Deteriorating Items with Stock-Dependent Demand Rate. *International Journal Production Economics*, **32**(3): 91 - 99.
- [8] Muth E. J. and Spremann K. (1983). Note—Learning Effects in Economic Lot Sizing. *Management Science*, **29**: 264-269.
- [9] Salameh M. K. Mohamed-Asem U. A. M. and Mohamad Y. J. (1993). Mathematical modelling of the effect of human learning in the finite production inventory model. *Applied Mathematical Modelling*, **17**: 613-615.
- [10] Cheng T. C. E.(1994). An economic manufacturing quantity model with learning effects. *International Journal Production Economics*, **33**: 257-264.
- [11] Salameh M. K. and Jaber M. Y. (2000). Economic production quantity model for items with imperfect quality, *International Journal Production Economics*, **64**: 59-64.
- [12] Jaber M. Y. and Guiffrid A. L. (2004). A Learning curves for processes generating defects requiring reworks, *European Journal Operations Research*, **159**: 663-672.
- [13] Eroglua A, and Ozdemirb G. (2007). An economic order quantity model with defective items and shortages. *International Journal Production Economics*, **106**: 544-549.
- [14] Jaber M. Y. and Guiffrid A. L. (2008). Learning curves for imperfect production processes with reworks and process restoration interruptions. *European Journal Operations Research*, **189**: 93-104.
- [15] Jaber M.Y., Goyal S. K. and Imran M. (2008). Economic production quantity model for items with imperfect quality subject to learning effects. *International Journal Production Economics*, **115**: 143-150.
- [16] Pan J. C. (2008). The Learning Effect on Setup Cost Reduction for Mixture Inventory Models with variable Lead Time, *Asia Pacific Journal Operations Research*, **25**: 513-529.
- [17] Lin P. C. (2008). Optimal pricing, production rate, and quality under learning effects. *Journal of Business Research*, **61**: 1152-1159.
- [18] Yoo H. S., Kim D. S. and Park M. S. (2009), Economic production quantity model with imperfect-quality items, two-way imperfect inspection and sales return, *International Journal Production Economics*, **121**: 255-265.
- [19] Sui Z., Gosavi A. and Lin L. (2010). A Reinforcement Learning Approach for Inventory Replenishment in Vendor-Managed Inventory Systems With Consignment Inventory. *Engineering Management Journal*, **22**: 44-53.
- [20] Khan M., Jaber M. Y., and Bonney M. (2011). An economic order quantity (EOQ) for items with imperfect quality and inspection errors. *International Journal Production Economics*, **133**: 113 -118.
- [21] Wahab M. I. M. and Jaber M.Y. (2010). Economic order quantity model for items with imperfect quality, different holding costs, and learning effects: A note. *Computers & Industrial Engineering*. **58**: 186-190.
- [22] Jaber M. Y. and Khan M.(2010), Managing yield by lot splitting in a serial production line with learning, rework and scrap. *International Journal Production Economics*, **124**: 32-39.
- [23] Das D. Roy A. and Kar S. (2010). A Production-Inventory Model for a Deteriorating Item Incorporating Learning Effect Using Genetic Algorithm. Hindawi Publishing Corporation *Advances in Operations Research.*, Article ID 146042: 26 pages.

- [24] Khan M., Jaber M.Y. and Wahab M.I.M (2010). Economic order quantity model for items with imperfect quality with learning in inspection. *International Journal Production Economics*, **124**: 87-96.
- [25] Konstantaras I., Skouri K. and Jaber M. Y. (2012). Inventory models for imperfect quality items with shortages and learning in inspection. *Applied Mathematical Modelling*, **36**: 5334-5343.
- [26] Jaggi C. K., Goyal S. K. and Mittal M. (2013). Credit financing in economic ordering policies for defective items with allowable shortages. *Applied Mathematics and Computation*, **219**: 5268-5282.
- [27] Teng J. T., Lou K. R. and Wang L. (2014), Optimal trade credit and lot size policies in economic production quantity models with learning curve production costs. *International Journal Production Economics*, **155**, 318-323.
- [28] Kumar N., Singh S.R. and Kumari R. (2013): Learning effect on an inventory model with two-level storage and partial backlogging under inflation. *International Journal of Services and Operations Management*, **16**: 105-122.
- [29] Givi Z. S., Jaber M. Y. and Neumann W. P., (2015). Modelling worker reliability with learning and fatigue. *Applied Mathematical Modelling*, **39**: 5186-5199.
- [30] Agi Maher. A. N. and Soni Hardik N. (2020). Joint Pricing and Inventory Decisions for Perishable Products with Age- Stock, and Price-Dependent Demand Rate. *Journal Operational Research Society*, **71**(1): 85 - 99.
- [31] Sarkar B. and Sumon S. (2013). An Improved Inventory Model with Partial Backlogging, Time Varying Deterioration and Stock-Dependent Demand. *Economic Modelling*, **30**(1): 24 - 32.
- [32] Jayaswal M. K. , Sangala I., Mittal M. and Malik S. (2019). Effects of learning on retailer ordering policy for imperfect quality items with trade credit financing. *Uncertain Supply Chain Management*, **7**: 49-62.
- [33] Yadav R., Prateek S,, Mittal M. and Mehta S.(2018). Effects of Imperfect Quality Items in the Asymmetric Information Structure in Supply Chain Model. *Uncertain Supply Chain Management*, **6**: 287-298.
- [34] Soni H. and Shah N. H. (2008). Optimal Ordering Policy for Stock-Dependent Demand under Progressive Payment Scheme. *European Journal of Operational Research*, **184**(1): 91-100.
- [35] Benyong H. and Feng Y. (2017). Optimization and Coordination of Supply Chain with Revenue Sharing Contracts and Service Requirement under Supply and Demand Uncertainty. *International Journal Production Economics*, **183**(October 2016): 185-193.
- [36] Nigwal A. R., Khedlekar U. K. and Gupta N.,(2022), Learning Effects On Retailer Ordering Policy For Imperfect Quality Items Under Trade Credit Financing With Pricing Strategies, *Jnanabha*, **52**(2): 6-22.
- [37] Nigwal A. R., Khedlekar U. K. Sharma L. and Gupta N.(2022). Trade Credit Policies for Supplier, Manufacturer, and Retailer: An Imperfect Production-Inventory System with Rework. *Journal of Mathematical & Fundamental Sciences*, **54**(1): 76-108.
- [38] Wright T. P. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences*, **3**: 122-128.

A NOVEL METHOD TO GENERATE A FAMILY OF BATHTUB-SHAPED FAILURE RATES FROM A FAMILY OF UPSIDE DOWN BATHTUB-SHAPED FAILURE RATES AND VICE-VERSA

R.L. GIRI¹, SUBARNA BHATTACHARJEE^{2*}, SUCHANDAN KAYAL³, S. K. MISRA³

^{1,2} Department of Mathematics, Ravenshaw University, Cuttack-753003, Odisha, India

³ Department of Mathematics, National Institute of Technology Rourkela, Odisha, India

⁴ Department of Mathematics, KIIT University, Bhubaneswar, Odisha, India

¹rajiblochan.giri@gmail.com, ²subarna.bhatt@gmail.com, ³suchandan.kayal@gmail.com,

⁴satyamisra05@gmail.com

Abstract

It is indeed a matter of great significance for system engineers and scientists to derive new classes of lifetime distributions for providing a better statistical model which will fit a given lifetime data set. It is known that many real time data have varied characteristics and can be modeled by distributions with bathtub and upside down bathtub failure rates viz., Weibull, Modified Weibull, Inverse Weibull. This paper proposes a method which generates a family of distributions having bathtub (BT)-shaped failure rate from a distribution having upside down bathtub (UBT)-shaped failure rate and vice-versa. The proposed method is validated with the help of a few statistical distributions. The closure properties of the proposed model under various reliability operations are studied.

Keywords: Aging phenomenon, hazard rate, bathtub-shaped failure rate, upside down bathtub-shaped failure rate.

AMS 2020 Subject Classification: Primary 60E15, Secondary 62N05, 60E05

1. INTRODUCTION

Lifetime distributions are usually categorized based on their failure pattern. Given that a device has survived till time $t > 0$, the hazard (failure) rate provides instantaneous failure rate in a very small (future) time interval. The shape of the hazard rate function can be strictly decreasing, strictly increasing, constant, BT and UBT. Increasing failure rate often occurs in the real life situations, where devices are more likely to fail with respect to age. Decreasing hazard rate appears when materials become harder with respect to time. The concept of bathtub (resp. upside bathtub) hazard rate distribution is discussed in the literature based on whether the corresponding hazard rate is decreasing (resp. increasing) in the region $(0, T_0]$, constant in $[T_0, T_1]$, and increasing (resp. decreasing) in $[T_1, \infty)$ where T_0 and T_1 are non-negative real numbers. In that case,

*Corresponding author : E-mail: subarna.bhatt@gmail.com

the random variable X is said to be BT (resp. UBT). Here, T_0 and T_1 are considered as change points of the hazard rate function. This concept holds even if $T_0 = T_1$. BT-shaped hazard rate is a combination of three different types of shapes, which usually appears in the study of life cycle of an industrial product or in the whole life span of a biological entity. Due to design error or installation problem, there is a high chance that a device has high likelihood of failure in first few weeks of operation. After initial period, the failure rate becomes relatively low, known as normal wear period. Then, the device reaches at the end of its life and the failure probability becomes very high with respect to time due to ageing. We refer to Rajarshi and Rajarshi [6] and Lai et al.[4] for some discussions on such kind of distributions. There are some other situations, related to the study of lifetime of a patient after major surgery, where models having unimodal hazard rate or having UBT-shaped hazard rate are useful. In biological science, it is observed in the course of a disease whose mortality reaches a peak after some finite period and then declines gradually. The commonly used distributions with UBT-shaped hazard rate are inverse Gaussian distribution, log-normal distribution, etc.

There are various transformations used by researchers to convert a baseline distribution into a new statistical distribution to get better flexibility. For example, inverted family of distributions can be obtained from a baseline distribution after using an inverse transformation. It has been shown by Keller et al.[3] that for pistons, crankshaft, main bearings failure data sets, the inverse Weibull distribution provides a better fit than the exponential and Weibull distributions. Akgül et al. [1] explored that the wind speed data can be modelled by inverse Weibull distribution, which gives a better output than Weibull distribution. This paper aims to provide a new method for the generation of a family of BT-shaped failure rates from a family of UBT-shaped failure rates and vice-versa. It is well-known that a series system formed with independent components each having BT-shaped failure rate with different change points has a BT-shaped failure rate with an arbitrary change point. In this article, we propose a new transformation so as to have a common (specific) change point of the resulting BT-shaped failure rate. Some of the resulting mathematical avenues are also explored for reverse model. Let X be a non-negative absolutely continuous random variable with probability density function (PDF) $f(\cdot)$ and cumulative distribution function (CDF) $F(\cdot)$. Then, the hazard rate of X is denoted by $r(t) = f(t)/\bar{F}(t)$, where $\bar{F}(t) = 1 - F(t)$, $t > 0$. Throughout the paper, we assume that the derivative exists whenever, it is implemented.

The rest of the paper is arranged as follows. Section 2 provides a transformation/method for the generation of the BT-shaped failure rate distribution from UBT-shaped failure rate distribution. In addition, various properties of the resulting BT-shaped failure rate distribution are explored. Section 3 discusses a method of generating UBT-shaped failure rates from BT-shaped failure rate with some notable consequences. Finally, Section 4 concludes the paper.

2. A METHOD TO GENERATE A BT-SHAPED FAILURE RATE USING UBT-SHAPED FAILURE RATE

Let U be a non-negative absolutely continuous lifetime random variable with CDF $F_U(\cdot)$ having UBT-shaped failure rate $r_U(t)$, for $t \in (l_U, u_U)$, where l_U and u_U denote respectively the lower and the upper bounds of the support of the random variable U . In this section, we introduce an interesting method to generate a distribution with corresponding lifetime random variable B with a BT-shaped failure rate $r_B(t)$, $t \in (l_B, u_B)$, where l_B and u_B denote respectively the lower and the upper bounds of the support of B using the distribution with UBT-shaped failure rate $r_U(t)$. Throughout the paper, we assume $l_U = 0$ and $u_U = +\infty$.

Theorem 1. Denote by $r_B(t)$ and $r_U(t)$ the BT-shaped failure rate and UBT-shaped failure rate, respectively. Then, the UBT-shaped failure rate can be obtained using an equation given by

$$r_B(t) = kM - r_U(t), \text{ for } t \geq 0, \tag{2.1}$$

where $k \geq 1$ is a real number and $M = \max_{t \geq 0} r_U(t)$.

Proof. The proof is clear from the following discussion. Note that the graphs of $r_B(t)$ and $r_U(t)$ are geometrically equivalent because one is obtained from other by reflection about horizontal axis and then by vertical translation of kM units. For a given UBT-shaped failure rate function $r_U(t)$, $-r_U(t)$ represents its vertical reflected image (or reflection about t -axis) lying in fourth quadrant, which is eventually a BT-shaped failure rate function. To shift up and to drag $-r_U(t)$, for all $t \geq 0$, back to first quadrant, we give a positive (up) shift by kM units, k being greater than or equal to one, the minimum required factor being M , where $M = \max_{t \geq 0} r_U(t)$. This completes the proof of the result. ■

Remark 1. Clearly, $\{r_U(\cdot), k\}$ completely describes the aforementioned model which satisfies the hypothesis of Theorem 1. This notation will be used throughout the article wherever required. The parameter M is derivable from $\{r_U(\cdot), k\}$.

The next theorem provides the survival function $\bar{F}_B(\cdot)$ and the density function $f_B(\cdot)$ corresponding to the newly generated distribution with BT-shaped failure rate, which is obtained from the UBT-shaped failure rate model by the method discussed in Theorem 1. The proof is omitted since it easily follows from Theorem 1 and the well-known relationship

$$\bar{F}_B(t) = e^{-\int_0^t r_B(u) du}. \tag{2.2}$$

Theorem 2. The survival and density functions of the random variable B are respectively given by

$$\bar{F}_B(t) = \frac{\exp(-kMt)}{\bar{F}_U(t)}, \quad t \geq 0 \tag{2.3}$$

and

$$f_B(t) = \frac{1}{\bar{F}_U(t)} (kM - r_U(t)) \exp(-kMt), \quad t \geq 0. \tag{2.4}$$

The method, discussed in Theorem 1 can be implemented to generate a family of BT-shaped failure rate models using a single UBT-shaped failure rate model as stated (without proof) in the next theorem.

Theorem 3. For $i = 1, \dots, n$, the random variables B_i 's have BT-shaped failure rates as given by

$$r_{B_i}(t) = k_i M - r_U(t), \text{ for } t \geq 0, \tag{2.5}$$

where $r_U(t)$ is the UBT-shaped failure rate, k_i 's are real constants satisfying $k_i \geq 1$, and $M = \max_{t \geq 0} r_U(t)$.

Proof. The proof follows using similar arguments as in the proof of Theorem 1, and thus it is omitted. ■

Next, we consider an example to illustrate the result in Theorem 3.

Example 2.1. Let a random variable U follow inverse Weibull distribution (see Jiang et al. [2]) with survival function $\bar{F}_U(t) = 1 - \exp\left(-\left(\frac{\beta}{t}\right)^\alpha\right)$, $t \geq 0$, $\alpha > 0$, $\beta > 0$. This distribution has UBT-shaped failure rate for $\alpha < 1$. Taking $\alpha = 0.5$, $\beta = 2$, the corresponding failure rate $r_U(t) = \frac{\alpha(\beta/t)^\alpha}{t(\exp(\beta/t)^\alpha - 1)}$ can be shown to be UBT. Further, $M = \max_{t \geq 0} r_U(t) = 0.35536$. Now, the corresponding $r_{B_i}(t)$'s are plotted in Figure 1(a), where $r_{B_i}(t) = k_i M - r_U(t)$, with $k_i = i$, for $i = 1, \dots, 5$.

Below, we compare two random variables B_i and B_j having UBT-shaped hazard rates in the sense of the hazard rate order. Let X and Y be two non-negative random variables with hazard rate functions $r_X(\cdot)$ and $r_Y(\cdot)$, respectively. Then, X is said to be smaller than Y in the sense of the hazard rate order, denoted by $X \leq_{hr} Y$, if $r_X(x) \geq r_Y(x)$, for all $x > 0$. For various other stochastic orders, we refer to Shaked and Shanthikumar [7]. From Theorem 3, we can write $r_{B_n}(t) = k_n M - r_U(t)$, for $n = i, j$.

Corollary 1. Let B_i and B_j be two random variables with $r_{B_n}(t) = k_n M - r_U(t)$, for $n = i, j$ and $t \geq 0$. Then, $B_i \leq_{hr} B_j$, if and only if $k_i \geq k_j$.

Proof. The proof is straightforward, and thus it is omitted. ■

2.1. Properties of the resulting BT-shaped failure models

In this subsection, we establish an interesting property of the resulting BT-shaped failure models. The following theorem shows that a series system formed by n number of independent components each having BT-shaped failure rate obtained from a common UBT-shaped failure rate model possesses BT-shaped failure rate model. In other words, a series system is closed under the specified BT transformation as given by (2.1) and (2.5).

Theorem 4. Consider a series system formed by n components with independent lifetimes denoted by B_i , $i = 1, \dots, n$. Further, let B_i have BT-shaped failure rate, say $r_{B_i}(t)$ generated from a single component with UBT-shaped failure rate $r_U(t)$ satisfying $r_{B_i}(t) = k_i M - r_U(t)$, $i = 1, \dots, n$, for all $t \geq 0$, $k_i \geq 1$, and $M = \max_{t \in (0, \infty)} r_U(t)$. Then, the system has BT-shaped failure rate, denoted by $r_{B_S}(t)$.

Proof. Note that

$$r_{B_S}(t) = \sum_{i=1}^n r_{B_i}(t) = \sum_{i=1}^n k_i M - n r_U(t) = n M \left(\frac{\sum_{i=1}^n k_i}{n} \right) - n r_U(t). \tag{2.6}$$

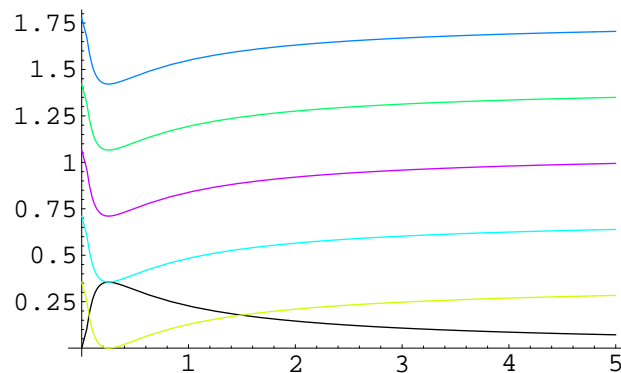
Further, $\max_{t \in (0, \infty)} nr_U(t) = nM$ and $\frac{\sum_{i=1}^n k_i}{n} \geq 1$. Thus, it follows from (2.1) and (2.6) that $r_{B_S}(t)$ represents a BT-shaped failure rate, implies the system has BT-shaped failure rate. This completes the proof. ■

The next remark gives an interesting fact about Theorem 4, and may be noted for independent interest.

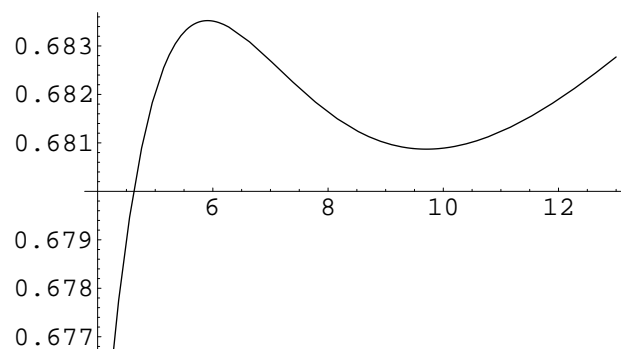
Remark 2. For $i = 1, \dots, n$, if $r_{B_i}(t) = k_i M - r_U(t)$, then $r_{B_i}(t)$ and $r_U(t)$ have the same change points as given by the roots of $\frac{d}{dt}(r_{B_i}(t)) = -\frac{d}{dt}(r_U(t)) = 0$, for all $t \in [0, +\infty)$. This leads to the fact that all of $r_{B_i}(t)$, for $i = 1, \dots, n$ have the same change point as a result of which $r_{B_S}(t) = \sum_{1 \leq i \leq n} r_{B_i}(t)$ has a change point equal to that of $r_{B_i}(t)$ (or $r_U(t)$) since from (2.6) we find

$$\frac{d}{dt}(r_{B_S}(t)) = -n \frac{d}{dt}(r_U(t)) = n \frac{d}{dt}(r_{B_i}(t)) = 0,$$

and hence B_S has a BT-shaped failure rate.



(a)



(b)

Figure 1: (a) Plots of $r_U(t)$ (the black curve) and $r_{B_i}(t)$, for $i = 1, \dots, 5$, respectively from bottom to top as in Example 2.1. (b) Plot of $r_X(t)$ versus t as in Counterexample 2.1

We now state a lemma with an outline of its proof, which will be used in proving upcoming theorem.

Lemma 1. Let S be a non-empty set and let $g_i(t)$ be defined on S such that $\max_{t \in S} g_i(t)$ exists for each $i = 1, \dots, n$. Then, we have

$$\max_{t \in S} \sum_{i=1}^n g_i(t) \leq \sum_{i=1}^n \max_{t \in S} g_i(t). \tag{2.7}$$

Proof. We know that $g_i(t) \leq \max_{t \in S} g_i(t)$, for $i = 1, \dots, n$ and $t \geq 0$. Thus,

$$\sum_{i=1}^n g_i(t) \leq \sum_{i=1}^n \max_{t \in S} g_i(t).$$

Further, since $\sum_{i=1}^n \max_{t \in S} g_i(t)$ is an upper bound of $\sum_{i=1}^n g_i(t)$, it follows that

$$\max_{t \in S} \sum_{i=1}^n g_i(t) \leq \sum_{i=1}^n \max_{t \in S} g_i(t).$$

Thus, the proof is completed. ■

Now, we present an example in the light of the above lemma, where strict inequality holds. It is quite easy to construct examples where equality holds.

Example 2.2. Let $g_1, g_2 : \mathbb{R} \rightarrow \mathbb{R}$ be given by

$$g_1(t) = \begin{cases} 1, & \text{for } t \in \mathbb{Q} \\ 0, & \text{for } t \in \mathbb{Q}^c \end{cases} \quad \text{and} \quad g_2(t) = \begin{cases} 0, & \text{for } t \in \mathbb{Q} \\ 1, & \text{for } t \in \mathbb{Q}^c \end{cases} \tag{2.8}$$

Clearly, $\max_{t \in [0, +\infty)} g_i(t) = 1$, for $i = 1, 2$ so that

$$\sum_{i=1}^2 \max_{t \in [0, +\infty)} g_i(t) = 2.$$

Furthermore, since $(g_1 + g_2)(t) = 1$, for all $t \in [0, +\infty)$, we have

$$\max_{t \in [0, +\infty)} \sum_{i=1}^2 g_i(t) = 1.$$

Thus,

$$\max_{t \in [0, +\infty)} \sum_{i=1}^2 g_i(t) < \sum_{i=1}^2 \max_{t \in [0, +\infty)} g_i(t)$$

is established.

In the upcoming theorem, we will observe that even though $r_{B_i}(t)$ possesses different change points yet

$$r_{B_{S^*}}(t) = \sum_{i=1}^2 r_{B_i}(t)$$

is BT-shaped. We pause for a while and read the next remark before going to Theorem 5.

Remark 3. For $i = 1, \dots, n$, if $r_{B_i}(t) = k_i M_i - r_{U_i}(t)$, then $r_{B_i}(t)$ possesses different change points given by the roots of $\frac{d}{dt}(r_{B_i}(t)) = -\frac{d}{dt}(r_{U_i}(t)) = 0$, for all $t \in [0, +\infty)$, provided that each $r_{U_i}(t)$ is differentiable. This leads to the fact that all of $r_{B_i}(t)$, for $i = 1, \dots, n$ have different change points.

Theorem 5. If B_{S^*} is a random variable denoting the lifetime of a series system formed by n independent components with lifetimes B_i , for $i = 1, \dots, n$ having BT-shaped failure rates $r_{B_i}(t)$ generated from independent components with UBT-shaped failure rates $r_{U_i}(t)$ satisfying $r_{B_i}(t) = k_i M_i - r_{U_i}(t)$, for all $t \geq 0$, $k_i \geq 1$, and $M_i = \max_{t \in (0, +\infty)} r_{U_i}(t)$, then $r_{B_{S^*}}(t)$ yields a distribution having BT-shaped failure rate.

Proof. Note that

$$r_{B_{S^*}}(t) = \sum_{i=1}^n r_{B_i}(t) = \sum_{i=1}^n (k_i M_i) - \sum_{i=1}^n r_{U_i}(t) = M \sum_{i=1}^n \left(\frac{k_i M_i}{M} \right) - \sum_{i=1}^n r_{U_i}(t), \tag{2.9}$$

where $M = \max_{t \in (0, \infty)} \sum_{i=1}^n r_{U_i}(t)$. Further, since each of $r_{U_i}(t)$, for $i \in \{1, \dots, n\}$ is concave, $\sum_{i=1}^n r_{U_i}(t)$ is also concave. Moreover, the local maximizer of a concave function defined over a convex set (here \mathbb{R}) is the global maximizer. Thus, $\sum_{i=1}^n r_{U_i}(t)$ possesses a UBT-shaped failure rate with unique maximizer. So, it suffices to show that $\sum_{i=1}^n \frac{k_i M_i}{M} \geq 1$ to establish our claim that $r_{B_{S^*}}(t)$ represents a BT failure rate as discussed in Theorem 1. Clearly,

$$\sum_{i=1}^n (k_i M_i) \geq \min_{1 \leq i \leq n} (k_i) \sum_{i=1}^n M_i \tag{2.10}$$

Again, using Lemma 1, one can show that

$$\left(\sum_{i=1}^n M_i \right) \sum_{i=1}^n \max_{t \in (0, +\infty)} r_{U_i}(t) \geq \max_{t \in (0, +\infty)} \sum_{i=1}^n r_{U_i}(t). \tag{2.11}$$

Thus, from (2.10) and (2.11), we conclude that

$$\sum_{i=1}^n (k_i M_i) \geq \min_{1 \leq i \leq n} (k_i) \sum_{i=1}^n \max_{t \in (0, \infty)} r_{U_i}(t) \geq \max_{t \in (0, \infty)} \sum_{i=1}^n r_{U_i}(t) = (M)$$

as $k_i \geq 1$, for all $i = 1, \dots, n$, that is $\sum_{i=1}^n \frac{k_i M_i}{M} \geq 1$. This completes the proof. ■

Since this special type of construction allows the BT-shaped failure rate system to be closed under the formation of series system, a natural question that arises is whether this result can be generalized to the formation of k -out-of- n system. We recall that k -out-of- n system works if atleast k components of n number of components work. In the following counterexample, we notice that the answer of this question in negative. It shows that the BT-shaped failure rate system is not closed under the formation of parallel system.

Counterexample 2.1. Consider a parallel system with lifetime X comprised of two components having failure rates, $r_{B_i}(t) = k_i M - r_U(t)$, $t \geq 0$ with $k_i = i + 1$, for $i = 1, 2$, and $r_U(t) = \frac{\beta(\alpha/t)^\beta}{t(\exp(\alpha/t)^\beta - 1)}$, $\alpha = 0.5$, $\beta = 2$, $M = \max_{t \geq 0} r_U(t) = 0.35536$. By Theorem 2, it follows that $\bar{F}_{B_i}(t) = \frac{\exp(-k_i M t)}{\bar{F}_U(t)}$, for $i = 1, 2$ so that $\bar{F}_X(t) = 1 - (1 - \bar{F}_{B_1}(t))(1 - \bar{F}_{B_2}(t))$, for all $t \geq 0$. The plot of $r_X(t)$ for $t \geq 0$ given in Figure 1(b) shows that it is roller coaster.

3. A METHOD TO GENERATE A UBT-SHAPED FAILURE RATE USING BT-SHAPED FAILURE RATE

Let B^* be a continuous non-negative random variable with CDF $F_{B^*}(\cdot)$ having BT-shaped failure rate $r_{B^*}(t)$ for $t \in [0, +\infty)$. On a similar line as discussed in the earlier section, we generate a distribution with corresponding random variable U^* having UBT-shaped failure rate as given in the next theorem. The proof is omitted for the sake of conciseness.

Theorem 6. A distribution with UBT-shaped failure rate denoted by $r_{U^*}(t)$ obtained from a distribution having BT-shaped failure rate $r_{B^*}(t)$ is generated by the following equation

$$r_{U^*}(t) = \begin{cases} 0 & \text{for } 0 < t \leq t_1 \\ km - r_{B^*}(t) & \text{for } t_1 \leq t \leq t_2 \\ 0 & \text{for } t \geq t_2, \end{cases} \quad (3.12)$$

where t_1 and t_2 are the positive roots of $km - r_{B^*}(t) = 0$ with $t_1 \leq t_2$ and $m = \min_{t \in [0, +\infty)} r_{B^*}(t)$ and k is a real number satisfying $k \geq 2$.

The next corollary is useful to obtain the survival function $\bar{F}_{U^*}(\cdot)$, and the density function $f_{U^*}(\cdot)$ of the newly generated UBT-shaped failure rate model obtained from BT-shaped failure rate model by the approach as discussed in Theorem 6.

Corollary 2. With reference to the hypothesis as in Theorem 6, it is easy to note that

(i) the survival function of the random variable U^* is

$$\bar{F}_{U^*}(t) = \begin{cases} 1 & \text{for } 0 < t \leq t_1 \\ \exp(-km(t - t_1)) \frac{\bar{F}_{B^*}(t_1)}{\bar{F}_{B^*}(t)} & \text{for } t_1 \leq t \leq t_2 \\ \exp(-km(t_2 - t_1)) \exp(-\int_{t_2}^t r_{B^*}(u) du) \frac{\bar{F}_{B^*}(t_1)}{\bar{F}_{B^*}(t)} & \text{for } t \geq t_2; \end{cases}$$

(ii) the density function of the random variable U^* can be obtained by simply differentiating $-\bar{F}_{U^*}(t)$ with respect to t .

The following proposition, which is useful to generate a family of UBT-shaped failure rate models using a single BT-shaped model, can be easily established from Theorem 3. The proof is omitted for the sake of brevity.

Proposition 3.1. A family of random variables U_i , for $i = 1, \dots, n$ each with UBT-shaped failure rate, given by

$$r_{U_i^*}(t) = \begin{cases} 0 & \text{for } 0 \leq t \leq t_1^{(i)} \\ k_i m - r_{B^*}(t) & \text{for } t_1^{(i)} \leq t \leq t_2^{(i)} \\ 0 & \text{for } t \geq t_2^{(i)}, \end{cases} \quad (3.13)$$

is generated from a random variable B^* with BT-shaped failure rate $r_{B^*}(t)$, where $t_1^{(i)}$ and $t_2^{(i)}$ are the positive roots of $k_i m - r_{B^*}(t) = 0$, with $t_1^{(i)} \leq t_2^{(i)}$, $m = \min_{t \in [0, +\infty)} r_{B^*}(t)$ and k_i is a real number satisfying $k_i \geq 2$.

The following corollary presents condition, under which the hazard rate order between U_i^* and U_j^* exists. We omit the proof since it is a consequence of Proposition 3.1.

Corollary 3. We have $U_i^* \geq_{hr} U_j^*$, if and only if $k_i \leq k_j$.

Let us use the notation $S_i = \{t \in \mathbb{R} \mid r_{U_i^*}(t) > 0\}$. Clearly, it follows from (3.13) that $S_i = (t_1^{(i)}, t_2^{(i)})$. Next, we state a strong result in the form of a lemma, which will be used later.

Lemma 2. If $k_i \leq k_j$, then $S_i \subseteq S_j$, for any $i, j \in \{1, \dots, n\}$.

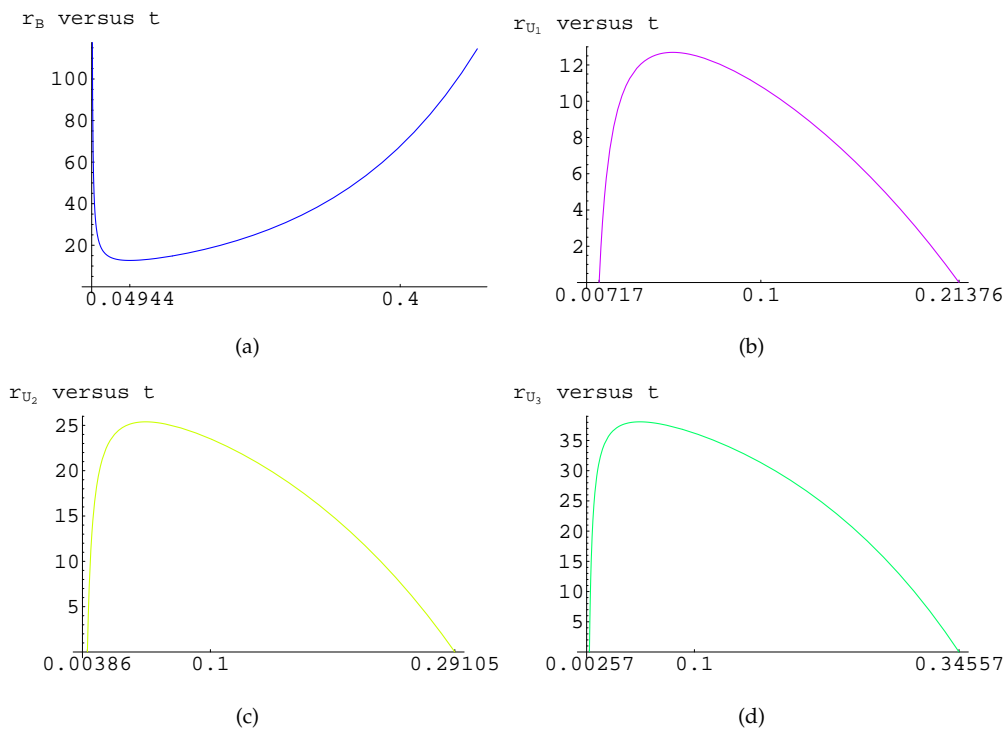


Figure 2: Plots of (a) $r_{B^*}(t)$, (b) $r_{U_1^*}(t)$, (c) $r_{U_2^*}(t)$, and (d) $r_{U_3^*}(t)$ versus t as in Example 3.1.

Proof. Since $k_i \leq k_j$, thus we have $r_{U_i^*}(t) \leq r_{U_j^*}(t)$, for $t > 0$. We claim that $S_i \subseteq S_j$ for any $i, j \in \{1, \dots, n\}$. If $x \in S_i$ but $x \notin S_j$, then $r_{U_i^*}(x) > 0$ and $r_{U_j^*}(x) = 0$, i.e., $r_{U_i^*}(x) > r_{U_j^*}(x)$, a contradiction. Hence the result follows. ■

Example 3.1. Let B^* have failure rate, given by $r_{B^*}(t) = a(\lambda t + b)e^{\lambda t}t^{b-1}$, $t \geq 0$. Taking $a = 2$, $b = 0.2$, $\lambda = 5$, it can be seen that B^* has bathtub-shaped failure rate (as $0 < b < 1$, (cf. Pham and Lai [5]). Here

$$m = \min_{t \in (0, +\infty)} r_{B^*}(t) = r_{B^*}(t_m) = 12.6948,$$

where $t_m = 0.0494427$. We construct $r_{U_i^*}(t) = k_i m - r_{B^*}(t)$, for $t \in [t_i^{(1)}, t_i^{(2)}]$ and $= 0$ otherwise, given that $k_i \in \{2, 3, 4\}$, for $i = 1, 2$ and 3 , respectively. Here, $[t_1^{(1)}, t_1^{(2)}] = [0.00717441, 0.21376]$, $[t_2^{(1)}, t_2^{(2)}] = [0.00386556, 0.291059]$, and $[t_3^{(1)}, t_3^{(2)}] = [0.00257858, 0.345577]$. This example has been implemented and plotted in Figure 2 and Figure 3(a).

3.1. Properties of the new UBT-shaped failure models

In this subsection, we show that the nature of the failure rate of a series system constituted by n independent components each having UBT-shaped failure rate, obtained from a single bathtub-shaped failure rate distribution can be derived using the concept in Proposition 3.1. If U_{S^*} is a random variable denoting the lifetime of a series system formed by n independent components with lifetimes U_i^* for $i = 1, \dots, n$ with corresponding UBT-shaped failure rate functions $r_{U_i^*}(t)$, all generated from a component with BT-shaped failure rate $r_{B^*}(t)$, then $r_{U_{S^*}}(t)$ can be derived as given in the following.

Let $r_{U_i^*}(t) = k_i m - r_{B^*}(t)$, for $t \in S_i$ and $r_{U_i^*}(t) = 0$ otherwise, where $S_i = \{t \mid r_{U_i^*}(t) \neq 0\}$, $i \in A = \{1, \dots, n\}$. Let $B = \{k_1, \dots, k_n\}$. Let us define k_i^* , for all $i \in A$ as

$$k_1^* = \min_{k_i \in B} k_i, \quad k_j^* = \min_{k_i \in B - \{k_1^*, k_2^*, \dots, k_{j-1}^*\}} k_i, \quad j \in A - \{1\}.$$

Further, consider the roots $t_1^{(p)}$ and $t_2^{(p)}$ of $r_{U_p^*}(t)$, for $p \in A$ with $t_1^{(p)} < t_2^{(p)}$, where

$$r_{U_p^*}(t) = k_p^* m - r_{B^*}(t), \quad p \in A, \quad t \in S_p,$$

with $S_p = \{t \mid r_{U_p^*}(t) \neq 0\}$. Now, one can prove that $r_{U_1^*}^*(t) < r_{U_2^*}^*(t) < \dots < r_{U_n^*}^*(t)$, for all $t \in \mathbb{R}$ as $k_j^* < k_{j+1}^*$, for all $j \in A$. Here, the two finite sets $\{r_{U_1^*}^*, \dots, r_{U_n^*}^*\}$ and $\{r_{U_n^*}^*, \dots, r_{U_1^*}^*\}$ are equal, i.e., its elements are rearrangement of each other. Each $r_{U_j^*}^*(t)$ is of UBT-shaped, for $t \in (t_1^{(j)}, t_2^{(j)})$.

From lemma 2, we know that $(t_1^{(j)}, t_2^{(j)}) \subseteq (t_1^{(j+1)}, t_2^{(j+1)})$, for all $j \in A$, as shown in Figure 3(b) giving

$$r_{U_{S^*}}(t) = \begin{cases} 0 & \text{for } t \in [0, t_1^{(n)}] \\ r_{U_n^*}^*(t) & \text{for } t \in [t_1^{(n)}, t_1^{(n-1)}] \\ r_{U_n^*}^*(t) + r_{U_{n-1}^*}^*(t) & \text{for } t \in [t_1^{(n-1)}, t_1^{(n-2)}] \\ \vdots & \\ \sum_{l=n-j}^n r_{U_l^*}^*(t) & \text{for } t \in [t_1^{(n-j)}, t_1^{(n-j-1)}] \\ \vdots & \\ \sum_{l=1}^n r_{U_l^*}^*(t) & \text{for } t \in [t_1^{(1)}, t_2^{(1)}] \\ \sum_{l=2}^n r_{U_l^*}^*(t) & \text{for } t \in [t_2^{(1)}, t_2^{(2)}] \\ \vdots & \\ \sum_{l=n-j}^n r_{U_l^*}^*(t) & \text{for } t \in [t_2^{(n-j-1)}, t_2^{(n-j)}] \\ \vdots & \\ r_{U_n^*}^*(t) & \text{for } t \in [t_2^{(n-1)}, t_2^{(n)}] \\ 0 & \text{for } t \in [t_2^{(n)}, +\infty), \end{cases}$$

so that

$$r_{U_{S^*}}(t) = \begin{cases} 0 & \text{for } t \in [0, t_1^{(n)}] \\ mk_n^* - r_{B^*}(t) & \text{for } t \in [t_1^{(n)}, t_1^{(n-1)}] \\ m(k_n^* + k_{n-1}^*) - 2r_{B^*}(t) & \text{for } t \in [t_1^{(n-1)}, t_1^{(n-2)}] \\ \vdots & \\ m \sum_{l=n-j}^n k_l^* - (j+1)r_{B^*}(t) & \text{for } t \in [t_1^{(n-j)}, t_1^{(n-j-1)}] \\ \vdots & \\ m \sum_{l=1}^n k_l^* - nr_{B^*}(t) & \text{for } t \in [t_1^{(1)}, t_2^{(1)}] \\ m \sum_{l=2}^n k_l^* - (n-1)r_{B^*}(t) & \text{for } t \in [t_2^{(1)}, t_2^{(2)}] \\ \vdots & \\ m \sum_{l=n-j}^n k_l^* - (j+1)r_{B^*}(t) & \text{for } t \in [t_2^{(n-j-1)}, t_2^{(n-j)}] \\ \vdots & \\ mk_n^* - r_{B^*}(t) & \text{for } t \in [t_2^{(n-1)}, t_2^{(n)}] \\ 0 & \text{for } t \in [t_2^{(n)}, +\infty). \end{cases}$$

Clearly, $(j + 1)r_{B^*}(t)$ represents failure rate of a bathtub distribution with

$$\min_{t \in [t_1^{(n-j)}, t_1^{(n-j-1)}]} (j + 1)r_{B^*}(t) = (j + 1)m,$$

where $m = \min_{t \in (0, +\infty)} r_{B^*}(t)$. One can note that $r_{U_{S^*}}(t)$ represents a UBT-shaped failure model as in $r_{U_s}(t) = m \sum_{l=n-j}^n k_l^* - (j + 1)r_{B^*}(t)$, for $t \in [t_1^{(n-j)}, t_1^{(n-j-1)}]$ and $r_{U_{S^*}}(t) = m \sum_{l=n-j}^n k_l^* - (j + 1)r_{B^*}(t)$, for $t \in [t_2^{(n-j-1)}, t_2^{(n-j)}]$. We find that $m \sum_{l=n-j}^n k_l^* \geq 2(j + 1)$, k_i being a real number satisfying $k_i \geq 2$ for all i .

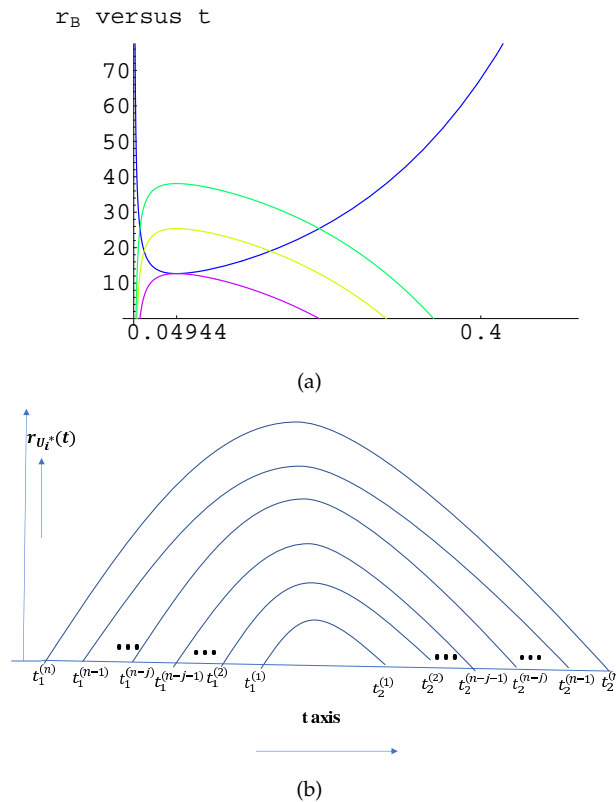


Figure 3: (a) Plot of $r_{B^*}(t)$ (blue color curve) and $r_{U_i^*}(t)$ versus t , for $i = 1, 2, 3$ (bottom to top) as in Example 3.1. (b) Plot of $r_{U_i^*}(t)$ for $i = 1, \dots, n$ versus t .

4. CONCLUDING REMARKS

In this paper, we have proposed a novel method which yields a family of distributions with BT shaped failure rate model from a distribution having UBT-shaped failure rate and vice-versa. Few examples have been presented for the validation of the newly proposed method. In addition, the closure properties of the proposed model have been studied under various reliability operations.

CONFLICT OF INTEREST STATEMENT

On behalf of all authors, the corresponding author states that there is no conflict of interest.

ACKNOWLEDGEMENTS

The corresponding author would like to thank Odisha State Higher Education Council for providing support to carry out the research project under OURIIP, Odisha, India (Grant No. 73/Seed Fund/2022/Mathematics).

REFERENCES

- [1] Akgül, F. G., Senoğlu, B. and Arslan, T. (2016). An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. *Energy Conversion and Management*, 114: 234—240.
- [2] Jiang, R., Murthy, D. and Ji, P. (2001). Models involving two inverse Weibull distributions. *Reliability Engineering and System Safety*, 73(1): 73—81.
- [3] Keller, A., Goblin, M. and Farnworth, N. (1985). Reliability analysis of commercial vehicle engines. *Reliability Engineering*, 10(1): 15–25.
- [4] Lai, C., Xie, M. and Murthy, D. (2001). Ch. 3 Bathtub-shaped failure rate life distributions. *Handbook of statistics*, 20: 69–104.
- [5] Pham, H. and Lai, C.-D. (2007). On recent generalizations of the Weibull distribution. *IEEE transactions on reliability*, 56(3): 454–458.
- [6] Rajarshi, S. and Rajarshi, M. (1988). Bathtub distributions: A review. *Communications in Statistics-Theory and Methods*, 17(8): 2597–2621.
- [7] Shaked, M. and Shanthikumar, J. G. *Stochastic Orders*, Springer, New York, 2007.

AN APPROACH FOR ASSESSMENT OF RELIABILITY INDICES CONSIDERING OMISSION OF FIXED REPAIR TIME FOR ELECTRIC TRACTION SYSTEM APPLYING MONTE CARLO SIMULATION

Aditya Tiwary

•

Department of Fire Technology & Safety Engineering, IPS Academy,
IES, Indore (M.P) India
raditya2002@gmail.com

Abstract

Assessment of numerous reliability indices is essential when availability and unavailability of supply in any electric power system is talked about. The reliability index which are very important for overall performance of any complex engineering system are mean up time, mean down time and unavailability. In this paper, assessment of various reliability indices for the electric traction system is done based on Monte Carlo simulation. If an engineering system fails then its repair is required to be performed at proper time interval. Omission of a threshold value of fixed repair time will not have so much impact on the overall reliability of the engineering system taken together. Electric traction system is very important as it is utilized for operation of passenger trains and freight trains across a large rail network throughout the continents. In view of above, modified values of mean up time, mean down time and unavailability have been obtained accounting fixed repair time omission for the electrical traction system taken under consideration.

Keywords: Monte Carlo simulation, Electric traction system, Mean up time, Repair time, Reliability indices.

1. Introduction

Reliability evaluation of a system or component or element is very important in order to predict its availability and other relevant indices. Reliability is the parameter which tells about the availability of the system under proper working conditions for a given period of time. A Markov cut-set composite approach to the reliability evaluation of transmission and distribution systems involving dependent failures was proposed by Singh et al. [1]. The reliability indices have been determined at any point of composite system by conditional probability approach by Billinton et al. [2]. Wojczynski et al. [3] discussed distribution system simulation studies which investigate the effect of interruption duration distributions and cost curve shapes on interruption cost estimates. New indices to reflect the integration of probabilistic models and fuzzy concepts was proposed by Verma et al. [4]. Zheng et al. [5] developed a model for a single unit and derived expression for availability of a component accounting tolerable repair time. Distributions of reliability indices resulting from two sampling techniques are presented and analyzed along with those from MCS

by Jirutitijaroen and Singh [6]. Dzobe et al. [7] investigated the use of probability distribution function in reliability worth analysis of electric power system. Bae and Kim [8] presented an analytical technique to evaluate the reliability of customers in a micro grid including distribution generations. Reliability network equivalent approach to distribution system reliability assessment is proposed by Billinton and Wang [9].

Customer and energy based indices consideration for reliability enhancement of distribution system using Improved Teaching Learning based optimization is discussed [10]. An Innovative Self-Adaptive Multi-Population Jaya Algorithm based Technique for Evaluation and Improvement of Reliability Indices of Electrical Power Distribution System, Tiwary et al. [11]. Determination of reliability indices for distribution system using a state transition sampling technique accounting random down time omission, Tiwary et al. [12]. Tiwary et al. [13] proposed a methodology based on Inspection-Repair-Based Availability Optimization of Distribution System Using Bare Bones Particle Swarm Optimization. Bootstrapping based technique for evaluating reliability indices of RBTS distribution system neglecting random down time was evaluated [14].

Volkanavski et al. [15] proposed application of fault tree analysis for assessment of the power system reliability. Li et al. [16] studies the impact of covered overhead conductors on distribution reliability and safety. Self-Adaptive Multi-Population Jaya Algorithm based Reactive Power Reserve Optimization Considering Voltage Stability Margin Constraints was obtained in Tiwary et al. [17]. A smooth bootstrapping based technique for evaluating distribution system reliability indices neglecting random interruption duration is developed [18]. Tiwary et al. [19] have developed an inspection maintenance based availability optimization methodology for feeder section using particle swarm optimization. The impact of covered overhead conductors on distribution reliability and safety is discussed [20]. Tiwary et al. [21] has discussed a methodology for reliability evaluation of an electrical power distribution system, which is radial in nature. Sarantakos et al. [22] introduced a method to include component condition and substation reliability into distribution system reconfiguration. Tiwary et al. [23] has discussed a methodology for evaluation of customer orientated indices and reliability of a meshed power distribution system. Reliability evaluation of engineering system is discussed [24]. Battu et al. [25] discussed a method for reliability compliant distribution system planning using Monte Carlo simulation. Application of non-parametric bootstrap technique for evaluating MTTF and reliability of a complex network with non-identical component failure laws is discussed [26]. Tiwary and Tiwary [27] have developed an innovative methodology for evaluation of customer orientated indices and reliability study of electrical feeder system. Tiwary and Tiwary [28] proposed the evaluation of reliability indices of Roy Billinton Test System (RBTS) Bus-2 Distribution System.

Tiwary and Tiwary [29] have proposed a methodology for reliability block diagram representation of electric traction system and identification of various reliability indices. In view of the above, in this paper reliability indices such as mean up time, mean down time and unavailability are obtained for the electrical traction system considering Monte Carlo simulation technique. Fixed repair time omission has been taken into account for obtaining the various reliability indices of importance. The result obtained is been shown in the result section and various analysis has also been done based on proposed method.

2. Reliability indices evaluation of electric traction system considering Omission of fixed Repair Time

Reliability block diagram modeling of the electric traction system is formulated and discussed by Tiwary and Tiwary [29]. The radial system which is taken into account, all the components is connected in series manner. System failure rate, repair time and unavailability are the parameters

which are considered as important reliability parameters.

The system failure rate is expressed as follows:

$$\lambda_{sys} = \Sigma \lambda_i \quad (1)$$

λ_i is the failure rate of each and every components.

The unavailability of a series system can be written as:

$$U_{sys} = \Sigma \lambda_i \cdot r_i \quad (2)$$

r_i is the repair time of each and every component.

The system repair time can be obtained by following relation:

$$r_{sys} = \frac{U_{sys}}{\lambda_{sys}} \quad (3)$$

3. Monte Carlo Simulation (MCS) Based Computational Algorithm

Certain value of the repair time which is fixed in nature can be omitted as it does not concern much to the system. A threshold value of the repair time can be omitted depending on the system under consideration. Evaluation of reliability indices with fixed repair time omission using Monte Carlo simulation (MCS) is discussed by Tiwary et al. [30].

Monte Carlo simulation is used to obtain the relevant reliability indices and the algorithm for the same is shown below:

Step 1: Obtain failure density function $f(t)$ of the components, repair density function $g(t)$ of the components and repair time omission (τ).

Step 2: Obtain random variates:

$$(T_{f,i}, T_{r,i}) \quad i=1, \dots, NS$$

$T_{f,i}$ = time to failure

$T_{r,i}$ = time to repair

NS = number of samples

Step 3: Obtain the modified value of the different reliability indices as follows:

If $T_{r,i} \leq \tau$

Then

$$\widetilde{T}_{f,i} = T_{r,i} + T_{f,i}$$

If $T_{r,i} > \tau$

Then

$$\widetilde{T}_{f,i} = T_{f,i}$$

If $T_{r,i} \leq \tau$

Then

$$\widetilde{T}_{r,i} = 0$$

If $T_{r,i} > \tau$

Then

$$\widetilde{T}_{r,i} = T_{r,i}$$

Where

$\widetilde{T}_{f,i}$ = modified time to failure

$\widetilde{T}_{r,i}$ = modified time to repair

Step 4: Repeat above steps for NS samples.

Step 5: Calculate the modified mean up time (\widetilde{MUT}) by using following relation:

$$\widetilde{MUT} = \frac{1}{NS} \Sigma \widetilde{T}_{f,i}$$

Step 6: Calculate the modified mean down time (\widetilde{MDT}) by using following relation:

$$\widetilde{MDT} = \frac{1}{NS} \sum \widetilde{T}_{r,i}$$

Step 7: Calculate the modified mean unavailability (\widetilde{U}) by using following relation:

$$\widetilde{U} = \frac{\widetilde{MDT}}{\widetilde{MUT}}$$

4. Result and Discussion

Table 1 shows the initial data such as failure rate per year and repair time in hours for the components of direct current (DC) electrification traction system. There are four components in the DC electrification traction system which are overhead wire, pantograph, motor control and motor and shown as components c1, c2, c3 and c4 respectively [29].

Table 1: Initial data for different components of the direct current electrification traction system [29].

component	c1	c2	c3	c4
Failure rate/year	0.04	0.03	0.005	0.004
Repair time (hrs.)	3	4	5	6

Table 2 shows the initial data such as failure rate per year and repair time in hours for the components of alternating current (AC) electrification traction system. There are six components in the AC electrification traction system which are overhead wire, pantograph, transformer, rectifier, motor control and motor and shown as components c1, c2, c3, c4, c5 and c6 respectively [29].

Table 2: Initial data for different components of the alternating current electrification traction system [29].

component	c1	c2	c3	c4	c5	c6
Failure rate/year	0.04	0.03	0.002	0.003	0.005	0.004
Repair time (hrs.)	3	4	6	4	5	6

Table 3 and Table 4 provide the component level evaluated modified failure rate, modified repair rate and modified unavailability for each and every component of the DC electrification traction system and AC electrification traction system respectively.

Table 3: Component level modified evaluated Reliability indices for each and every component of the DC electrification traction system.

Component Level	MUT, year	MDT, h	U, h/year	\widetilde{MUT} , year	\widetilde{MDT} , h	\widetilde{U} , h/year
C1	25.0000	3.0000	0.1200	25.3569	2.8856	0.1138
C2	14.2857	3.4286	0.2400	14.3569	3.1546	0.2197
C3	13.3333	3.5333	0.2650	13.8596	3.2548	0.2348
C4	12.6582	3.6582	0.2890	12.8896	3.4568	0.2682

Table 4: Component level modified evaluated Reliability indices for each and every component of the AC electrification traction system.

Component Level	MUT, year	MDT, h	U, h/year	\widetilde{MUT} , year	\widetilde{MDT} , h	\widetilde{U} , h/year
C1	25.0000	3.0000	0.1200	25.2365	2.8965	0.1148
C2	14.2857	3.4286	0.2400	14.3895	3.2245	0.2241
C3	13.8889	3.5000	0.2520	13.9623	3.3654	0.2410
C4	13.3333	3.5200	0.2640	13.5623	3.2361	0.2386
C5	12.5000	3.6125	0.2890	12.6985	3.4521	0.2718
C6	11.9048	3.7262	0.3130	11.9965	3.4523	0.2878

Fig. 1 shows the magnitude of modified mean up time at Component level C1, C2, C3 and C4 of DC electrification traction system. The magnitude of MUT has increased from 25.0000 to 25.3569 for the component C1. Components C2, C3 and C4 are having increase from 14.2857 to 14.3569, 13.3333 to 13.8596 and 12.6582 to 12.8896 respectively. Magnitude of modified mean down time and modified unavailability at Component level C1, C2, C3 and C4 of DC electrification traction system is shown in Fig. 2 and Fig. 3 respectively. Modified mean down time are 2.8856, 3.1546, 3.2548 and 3.4568 respectively. Improvement in the modified unavailability is 0.0062, 0.0203, 0.0302 and 0.0208 respectively.

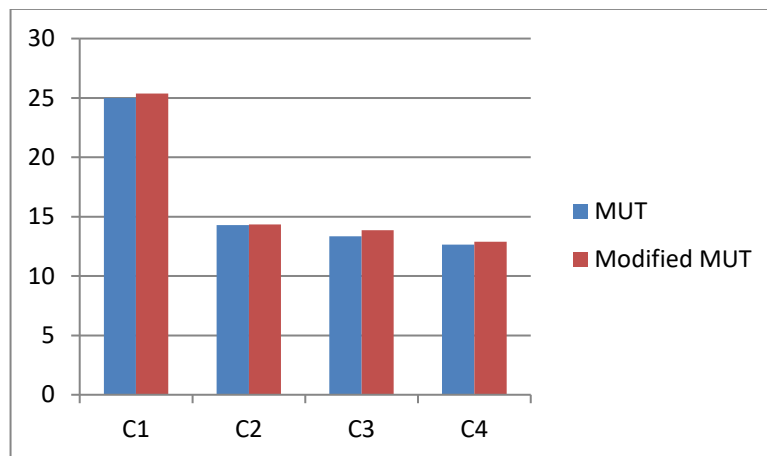


Figure 1: Magnitude of modified mean up time at Component level C1, C2, C3 and C4 of DC electrification traction system.

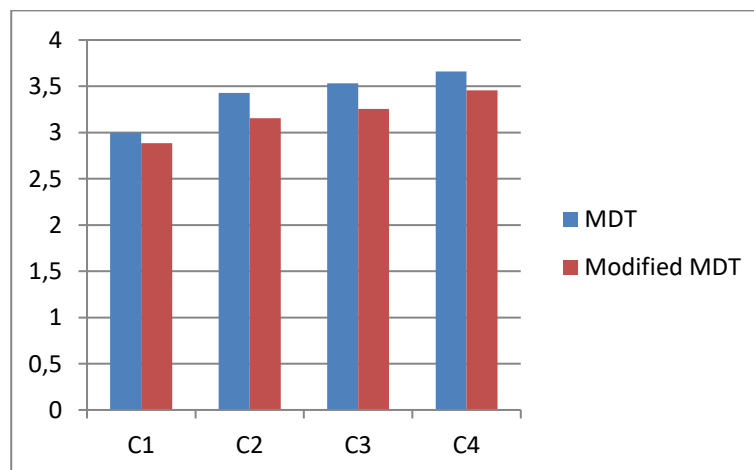


Figure 2: Magnitude of modified mean down time at Component level C1, C2, C3 and C4 of DC electrification traction system.

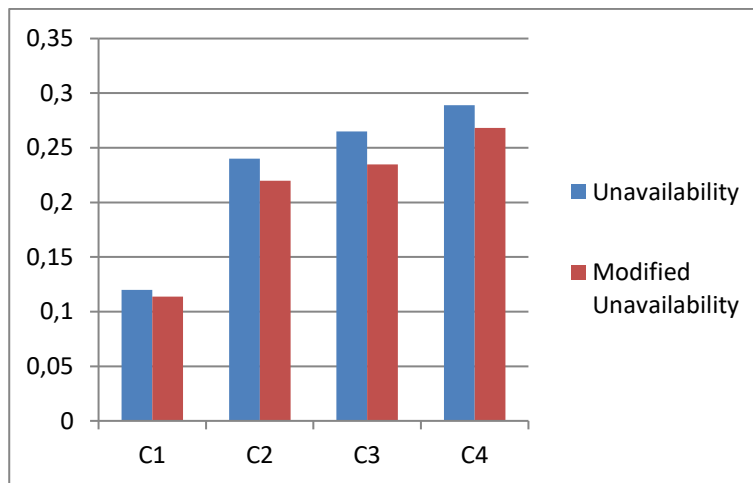


Figure 3: Magnitude of modified unavailability at Component level C1, C2, C3 and C4 of DC electrification traction system.

Fig. 4 shows the magnitude of modified mean up time at Component level C1, C2, C3, C4, C5 and C6 of AC electrification traction system. The magnitude of MUT has increased from 25.0000 to 25.2365 for the component C1. Components C2, C3, C4, C5 and C6 are having increase from 14.2857 to 14.3895, 13.8889 to 13.9623, 13.3333 to 13.5623, 12.5000 to 12.6985 and 11.9048 to 11.9965 respectively. Magnitude of modified mean down time and modified unavailability at Component level C1, C2, C3, C4, C5 and C6 of AC electrification traction system is shown in Fig. 5 and Fig. 6 respectively. Modified mean down time are 2.8965, 3.2245, 3.3654, 3.2361, 3.4521 and 3.4523 respectively. Improvement in the modified unavailability is 0.0052, 0.0159, 0.0110, 0.0254, 0.0172 and 0.0252 respectively.

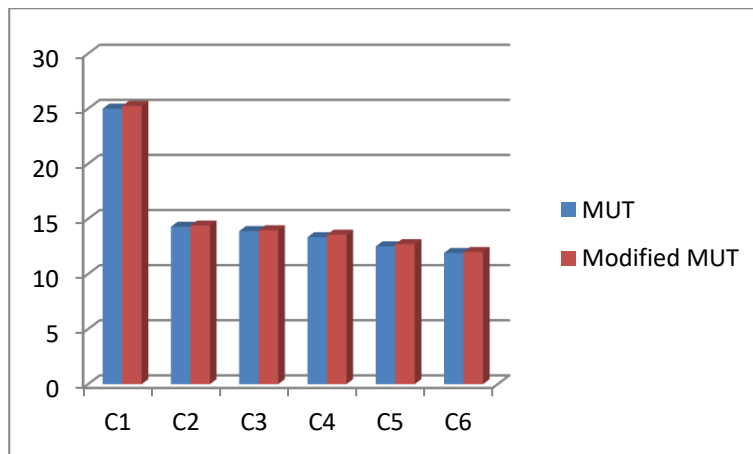


Figure 4: Magnitude of modified mean up time at Component level C1, C2, C3, C4, C5 and C6 of AC electrification traction system.

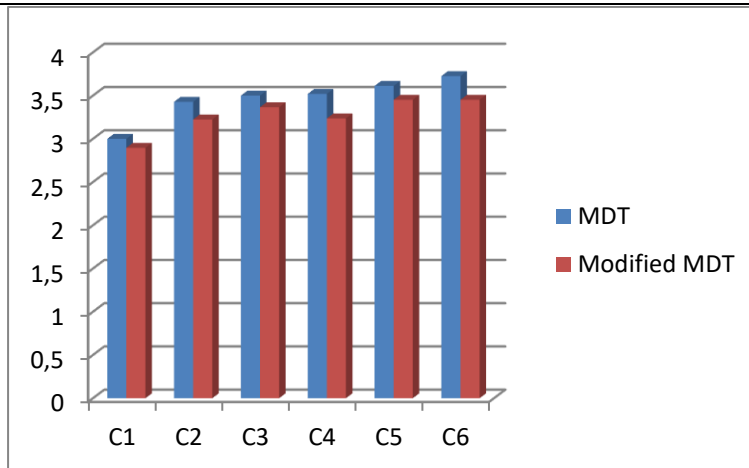


Figure 5: Magnitude of modified mean down time at Component level C1, C2, C3, C4, C5 and C6 of AC electrification traction system.

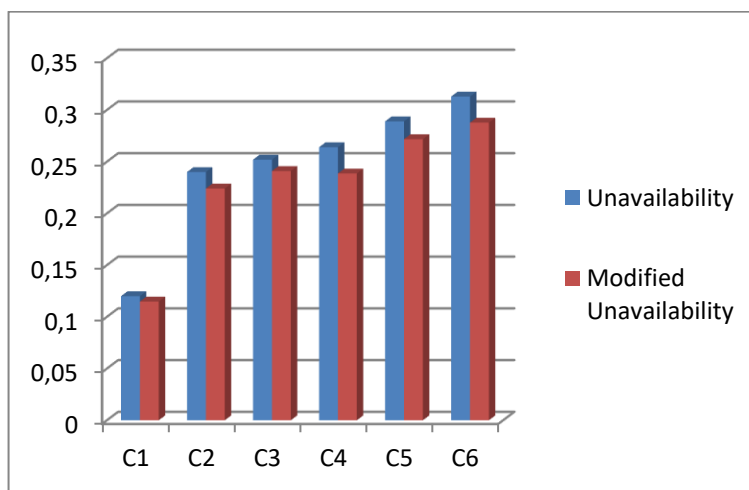


Figure 6: Magnitude of modified unavailability at Component level C1, C2, C3, C4, C5 and C6 of AC electrification traction system.

5. Conclusion

Evaluation of reliability of complex engineering system is necessary in order to obtain its overall availability. In view of the above, in this paper calculation of various reliability indices for the electric traction system is obtained based on Monte Carlo simulation methodology. Considering the omission of the threshold value of fixed repair time, the overall reliability indices of the engineering system taken into account is obtained. Based on the Monte Carlo simulation computational algorithm the component level modified evaluated reliability indices for each and every component of the DC and AC electrification traction system is obtained and shown. Modified mean up time, modified mean down time and modified mean unavailability are obtained and shown in result and discussion section for DC and AC electrification traction system respectively. The magnitude of modified mean up time at Component level of DC and AC electrification traction system is shown. It also provides the magnitude of modified mean down time at Component level of DC and AC electrification traction system. Magnitude of modified mean unavailability for both the electrification traction system is shown in result section.

References

- [1] Singh, C. (1981). Markov cut-set approach for the reliability evaluation of transmission and distribution systems. *IEEE Trans. on Power Apparatus and Systems*, 100: 2719-2725.
- [2] Billinton, R. (1969). Composite system reliability evaluation. *IEEE Trans. on Power Apparatus and Systems*, 88: 276-281.
- [3] Wojczynski, E. and Billinton, R. (1985). Effects of distribution system reliability index distributions upon interruption cost/reliability worth estimates. *IEEE Trans. on Power Apparatus and Systems*, 11: 3229-3235.
- [4] Verma, A. K., Srividya, A., Kumar, H. M. R. (2002). A framework using uncertainties in the composite power system reliability evaluation. *Electric Power Components and Systems*, 30: 679-691.
- [5] Zheng, Z., Cui, L., Hawkes, A. G. (2006). A study on a single-unit Markov repairable system with repair time omission. *IEEE Trans. on Reliability*, 55: 182-188.
- [6] Jirutitijaroen, P. and Singh, C. (2008). Comparison of simulation methods for power system reliability indexes and their distributions. *IEEE Trans. on Power Systems*, 23: 486-493.
- [7] Dzobe, O., Gaunt, C. T., Herman, R. (2012). Investigating the use of probability distribution functions in reliability-worth analysis of electric power systems. *Int. J. of Electrical Power and Energy Systems*, 37: 110-116.
- [8] Bae, I. S. and Kim, J. O. (2008). Reliability evaluation of customers in a microgrid. *IEEE Trans. on Power Systems*, 23: 1416-1422.
- [9] Billinton, R. and Wang, P. (1998). Reliability-network-equivalent approach to distribution-system-reliability evaluation. *IEE Proc. generation, transmission and distribution*, 145: 149-153.
- [10] Tiwary, A. (2017). Reliability enhancement of distribution system using Teaching Learning based optimization considering customer and energy based indices. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 3: 58-62.
- [11] Tiwary, A. (2018). An Innovative Self-Adaptive Multi-Population Jaya Algorithm based Technique for Evaluation and Improvement of Reliability Indices of Electrical Power Distribution System. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4: 299-302.
- [12] Tiwary, A., Arya, R., Choube, S. C., Arya, L. D. (2013). Determination of reliability indices for distribution system using a state transition sampling technique accounting random down time omission. *Journal of The Institution of Engineers (India): series B (Springer)*, 94: 71-83.
- [13] Tiwary, A. (2019). Inspection-Repair-Based Availability Optimization of Distribution System Using Bare Bones Particle Swarm Optimization. *Computational Intelligence: Theories, Applications and Future Directions – Volume II, Advances in Intelligent Systems and computing*, 799.
- [14] Tiwary, A., Arya, R., Arya, L. D., Choube, S. C. (2017). Bootstrapping based technique for evaluating reliability indices of RBTS distribution system neglecting random down time. *The IUP Journal of Electrical and Electronics Engineering*, X: 48-57.
- [15] Volkanavski, Cepin, M., Mavko, B. (2009). Application of fault tree analysis for assessment of the power system reliability. *Reliab Eng Syst Safety*, 94: 1116–27.
- [16] Li, B.M., Su, C.T., Shen, C.L. (2010). The impact of covered overhead conductors on distribution reliability and safety. *Int J Electr Power Energy Syst*, 32: 281–9.
- [17] Tiwary, A. (2018). Self-Adaptive Multi-Population Jaya Algorithm based Reactive Power Reserve Optimization Considering Voltage Stability Margin Constraints. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4: 341-345.
- [18] Arya, R., Tiwary, A., Choube, S. C., Arya, L. D. (2013). A smooth bootstrapping based technique for evaluating distribution system reliability indices neglecting random interruption duration. *Int. J. of Electrical Power and Energy System*, 51: 307-310.

- [19] Tiwary, A. (2018). Inspection–Maintenance-Based Availability Optimization of Feeder Section Using Particle Swarm optimization. *Soft Computing for Problem Solving-Advances in Intelligent Systems and Computing*, 816: 257-272.
- [20] BinLi, M., TzongSu, C., LungShen, C. (2010). The impact of covered overhead conductors on distribution reliability and safety. *Int. J. of Electrical Power and Energy System*, 32: 281-289.
- [21] Tiwary, A. (2019). Reliability evaluation of radial distribution system – A case study. *Int. J. of Reliability: Theory and Applications*, 14, 4(55): 9-13.
- [22] Sarantakos, I., Greenwood, D. M., Yi, J., Blake, S. R., Taylor, P. C. (2019). A method to include component condition and substation reliability into distribution system reconfiguration. *Int. J. of Electrical Power and Energy System*, 109: 122-138.
- [23] Tiwary, A. (2020). Customer orientated indices and reliability evaluation of meshed power distribution system. *Int. J. of Reliability: Theory and Applications*, 15, 1(56): 10-19.
- [24] Tiwary, A. and Patel, P. (2020). Reliability Evaluation of Hose Reel System - A Practical Approach. *Journal of Industrial Safety Engineering*, 7: 30-34.
- [25] Battu, N. R., Abhyankar, A. R., Senroy, N. (2019). Reliability Compliant Distribution System Planning Using Monte Carlo Simulation. *Electric power components and systems*, 47: 985-997.
- [26] Tiwary, A. (2020). Application of Non-Parametric Bootstrap Technique for evaluating MTTF and Reliability of a Complex Network with Non-Identical Component Failure Laws. *Reliability: Theory and Applications*, 15: 62-69.
- [27] Tiwary, A. and Tiwary, S. (2020). Evaluation of Customer Orientated Indices and Reliability Study of Electrical Feeder System. *Reliability: Theory and Applications*, 15: 36-43.
- [28] Tiwary, A. and Tiwary, S. (2021). Evaluation of Reliability Indices of Roy Billinton Test System (RBTS) Bus-2 Distribution System for Educational Purpose. *Reliability: Theory and Applications*, 16: 54-61.
- [29] Tiwary, A. and Tiwary, S. (2021). An Innovative Methodology for Evaluation of Reliability Indices of Electric Traction System. *Reliability: Theory and Applications*, 16: 13-21.
- [30] Tiwary, A., Arya, R., Choube, S. C., Arya, L. D. (2012). Reliability indices evaluation of a distributor with fixed repair time omission using Monte Carlo simulation. *17th National Power system conference- NPSC-2012*, Indian Institute of Technology- Banaras Hindu University, Varanasi.

BAYESIAN INTERVAL ESTIMATION FOR THE PARAMETERS OF POISSON TYPE LENGTH BIASED EXPONENTIAL CLASS MODEL

Rajesh Singh¹, Preeti A. Badge², Pritee Singh³

^{1,2}Department of Statistics, S. G. B. Amravati University, Amravati, India

³Department of Statistics, Institute of Science, Nagpur, India

rsinghamt@hotmail.com

preetibadge10@gmail.com

priteesingh25@gmail.com

Abstract

In this research paper, two sided Bayesian interval is proposed for Poisson type length biased exponential class software reliability growth model. The failure intensity function, mean time to failure function and likelihood function are derived. Bayesian interval estimation has been done for the parameters using non informative priors. The performance of proposed Bayesian interval is obtained by using Monte Carlo simulation technique. Average length and coverage probability of Bayesian interval for the parameters are calculated. From the obtained intervals it is concluded that Bayesian interval of parameters perform better for appropriate choice of execution time and certain values of parameters.

Keywords: Length biased exponential distribution, Non informative priors, Software reliability growth model (SRGM), Bayesian interval, average length, coverage probability.

1. Introduction

This paper considers Poisson type length biased exponential class model according to classification scheme of Musa and Okumoto [9]. Poisson execution time models are based on the premise that execution time is the best time domain for expressing reliability. Execution is the most practical measure of the failure inducing stress being placed on a program. Musa et al [8] have suggested that it is convenient to divide the program into number of runs. The run depends on the function executed by program. The time required for run is depends upon size of run. As the size of run varies the number of failure observed in single run may vary. Fisher [2] defined length biased and further formulated by Rao [12] Gupta and Keating [6] developed relationship between survival function, the failure rate and mean residual life function using length biased distribution. Patil and Rao [10] have given a table for some distribution and their size biased forms. Rao and Cunha [13] estimated credible interval and confidence interval through MLE for lognormal distribution and also compared average length and coverage probability of the calculated interval. Tamak [16] estimated reliability of web application using Goel-Okumoto Software Reliability Growth models (SRGM). Shreshtha and Kumar [14] computed MLE and Bayesian estimate for Rayleigh distribution using gamma prior. Singh et al [15] introduced length biased distribution as Software Reliability Growth models (SRGM). Fitrilia et al [3] estimate the failure rate by E-Bayesian estimation method. E-Bayesian estimation is an

expectation of Bayes estimation, in order to obtain Bayes estimation expectations is by calculating the mean of Bayes estimator. Rabie and Li [11] studied the Bayesian and E-Bayesian approaches under squared error and LINEX loss functions. Also construct Confidence intervals for maximum likelihood estimates, as well as credible intervals for the E-Bayesian and Bayesian estimates. Andure and Ade [1] proposed length biased quasi lindley distribution and discussed different properties of proposed distribution. Gupta et al [5] obtained Bayesian and non-Bayesian estimators under symmetric (squared error) and asymmetric (linex and precautionary) loss functions using a non-informative prior. And compared risk efficiencies of Bayes estimators with maximum likelihood estimators.

In this paper, it is considered that the time to failure of an individual fault following length biased exponential distribution and the failure experienced by time t is distributed as Poisson type. In this model it is assumed that the software failures are independent of each other but depend on length of the time interval which contains the software failures.

The structure of the paper is such that section 2 presents derivation of failure intensity and expected number of failures using length biased exponential distribution, derivation of likelihood function, selection of priors, and derives joint and marginal posterior distribution of model. Section 3 presents derivation of two sided Bayesian interval for the parameters θ_0 and θ_1 . Results and discussion is given in the section 4 while concluding remarks are provided in section 5.

2. Model Formulation

Consider that software is tested for its performance and observed the time of failure occurs during software system performance. Let the number of failures present in software be θ_0 , and t_e be the execution time i.e. time during which CPU is busy and m_e be the number of failures observed up to execution time t_e . Consider time between the failures t_i ($i=1,2,\dots,m_e$) follows the exponential distribution with parameter θ_1 . The length biased exponential distribution is given as

$$f^*(t) = \begin{cases} t\theta_1^2 e^{-\theta_1 t} & , t > 0, \theta_1 > 0, E[t] \neq 0 \\ 0 & otherwise \end{cases} \quad (1)$$

Where $f^*(t)$ denotes the length biased exponential distribution.

The failure intensity function is obtained by using equation (1)

$$\lambda(t) = \theta_0 t \theta_1^2 e^{-\theta_1 t} \quad , t > 0, \theta_0 > 0 \quad (2)$$

Where θ_0 express the number of failures and θ_1 express the for failure rate.

The mean failure function i.e. expected number of failures at time t_e can be obtained by using equations (2)

$$\mu(t_e) = \theta_0 \theta_1^2 I_1 \quad (3)$$

Where, $I_1 = \int_0^{t_e} t_i e^{-\theta_1 t_i} dt$ and by solving (see Gradshteyn and Ryzhik [4] p. 357) we get,

$$\mu(t_e) = \theta_0 [1 - (1 + \theta_1 t_e) e^{-\theta_1 t_e}] \quad , t > 0, \theta_0 > 0, \theta_1 > 0 \quad (4)$$

Behavior of failure intensity and expected number of failure of length biased exponential class model has been studied by Singh et al [15]. They have compared the maximum likelihood estimates i.e. MLE's and Bayesian estimators on the basis of risk efficiencies.

Now for a system, considering that m_e software failures are observed at times t_i , $i = 1, 2, \dots, m_e$ up to execution time is t_e ($\geq t_{me}$) and the likelihood function of parameters θ_0 and θ_1 with the help of failure intensity and mean failure function can be obtained as (cf. Singh et al [15])

$$L(\theta_0, \theta_1) = \theta_0^{m_e} \theta_1^{2m_e} \left[\prod_{i=1}^{m_e} t_i \right] e^{-T\theta_1} e^{-\theta_0 [1 - (1 + \theta_1 t_e) e^{-\theta_1 t_e}]} \quad (5)$$

3. Bayesian interval Estimation of parameters θ_0 and θ_1

Bayesian estimation is done by combining prior information with information obtained from sample data. While testing the software, the experimenter have very little knowledge relative to the total number of failures present in the software.. Here insufficient prior information is not available about parameters θ_0 and θ_1 , hence non- informative priors are considered. The following non informative prior distributions $g(\theta_0)$ and $g(\theta_1)$ are considered for parameters θ_0 and θ_1 as follows:

$$g(\theta_0) \propto \begin{cases} \theta_0^{-1} & , \theta_0 \in [0, \infty) \\ 0 & , otherwise \end{cases} \quad (6)$$

and

$$g(\theta_1) \propto \begin{cases} \theta_1^{-1} & , \theta_1 \in [0, \infty) \\ 0 & , otherwise \end{cases} \quad (7)$$

The joint posterior of θ_0 and θ_1 given \underline{t} ($=t_i, i = 1, 2, \dots, m_e$) is obtained by using equations (5),(6) and (7) is as follows:

$$\pi(\theta_0, \theta_1 | \underline{t}) = D^{-1} \theta_0^{m_e-1} \theta_1^{2m_e-1} e^{-T\theta_1} e^{-\theta_0} e^{[\theta_0(1+\theta_1 t_e) e^{-\theta_1 t_e}]} \quad m_e < \theta_0 < \infty, 0 < \theta_1 < \infty \quad (8)$$

Where, D is normalizing constant.

$$D = \sum_{k=0}^{\infty} \sum_{r=0}^k \binom{k}{r} \left[\frac{\Gamma(2m_e + r) \Gamma(m_e + k, m_e) t_e^r (T^*)^{-2m_e-r}}{k!} \right]$$

$$\text{Where, } T^* = T + k t_e, \quad T = \sum_{i=1}^{m_e} t_i$$

The marginal posterior distribution of θ_0 given \underline{t} is,

$$\pi(\theta_0 | \underline{t}) = D^{-1} \sum_{k=0}^{\infty} \sum_{r=0}^k \binom{k}{r} \left[\frac{t_e^r \Gamma(2m_e+r)(T^*)^{-2m_e-r}}{k!} \right] [\theta_0^{m_e+k-1} e^{-\theta_0}], \theta_0 > m_e \quad (9)$$

The marginal posterior of θ_1 , say $\pi(\theta_1 | \underline{t})$ can be obtained as

$$\pi(\theta_1 | \underline{t}) = D^{-1} \sum_{k=0}^{\infty} \left[\frac{\Gamma(m_e+k, m_e)}{k!} \right] [\theta_1^{2m_e-1} (1 + \theta_1 t_e)^k e^{-\theta_1 T^*}], \theta_1 > 0 \quad (10)$$

A symmetric 100(1- α) % two sided Bayes probability interval (θ_L, θ_U) is given as

$$\int_{-\infty}^{\theta_L} \pi(\theta | \underline{t}) dt = \alpha/2$$

$$\int_{\theta_U}^{\infty} \pi(\theta | \underline{t}) dt = \alpha/2$$

Where $\pi(\theta | \underline{t})$ is the marginal posterior distribution of θ for details see Martz and Waller [7]

Now, to obtain two sided Bayes interval for the parameter θ_0 and θ_1 by integrating equation (9) and (10) w.r.t. θ_0 and θ_1 respectively can be given as:

$$\tilde{\theta}_{0L} = D^{-1} \sum_{k=0}^{\infty} \sum_{r=0}^k \binom{k}{r} \left[\frac{t_e^r \Gamma(2m_e+r)(T^*)^{-2m_e-r}}{k!} \right] \Gamma(m_e + k, \theta_{0*})$$

$$\tilde{\theta}_{0U} = D^{-1} \sum_{k=0}^{\infty} \sum_{r=0}^k \binom{k}{r} \left[\frac{t_e^r \Gamma(2m_e+r)(T^*)^{-2m_e-r}}{k!} \right] \Gamma(m_e + k, \theta_0^*)$$

$$\tilde{\theta}_{1L} = D^{-1} \sum_{k=0}^{\infty} \sum_{r=0}^k \binom{k}{r} \left[\frac{\Gamma(m_e+k, m_e)}{k!} \right] t_e^r (T^*)^{-2m_e-r} \gamma(2m_e + r, T^* \theta_{1*})$$

$$\tilde{\theta}_{1U} = D^{-1} \sum_{k=0}^{\infty} \sum_{r=0}^k \binom{k}{r} \left[\frac{\Gamma(m_e+k, m_e)}{k!} \right] t_e^r (T^*)^{-2m_e-r} \Gamma(2m_e + r, T^* \theta_1^*)$$

Where, $\Gamma(m_e + k, \theta_0^*)$, $\Gamma(2m_e + r, T^* \theta_1^*)$ are incomplete gamma functions.

$$\int_u^\infty x^{v-1} e^{-\mu x} dx = \mu^{-v} \Gamma(v, \mu u) \quad [Re u > 0, Re \mu > 0, n > 0]$$

is known as incomplete gamma function. The details about the incomplete gamma function can be seen from Gradshteyn and Ryzhik [7]. Therefore, $(\tilde{\theta}_{0L}, \tilde{\theta}_{0U})$ and $(\tilde{\theta}_{1L}, \tilde{\theta}_{1U})$ are the Bayesian interval estimates for parameter θ_0 and θ_1 respectively.

4. Results and Discussion

Here, two sided Bayesian interval are obtained for the parameter total number of failures i.e. θ_0 and the parameter θ_1 . The interval estimate is the posterior probability distribution about the parameter. To study the performance of two sided Bayesian interval, sample of different sizes (say m_e) was generated from the length biased exponential distribution and it is repeated 1000 times. Monte Carlo simulation technique is used for simulation and the result is presented in the tables given below. The credible interval depends upon the values of execution time i.e. t_e and number of failures experienced at times t_i , $i = 1, 2, \dots, m_e$. Average length and coverage probability of the proposed Bayesian interval have been calculated for different values of parameters θ_0 and θ_1 and for certain execution time. Here, it is assumed that two sided Bayesian interval maintains the credible level if the estimated coverage probability is found in between the range of 0.940 to 0.960 i.e. $(1-\alpha) \pm 0.01$ where, $\alpha = 0.05$.

Here tables (1) to (4) give average length and coverage probability for Bayesian interval for parameter θ_0 i.e. total number of failures. Average length have been obtained by assuming parameter $\theta_0 (=1(1)5)$, $\theta_1 (= 0.6(0.1)1)$. From tables, we can see that Average length is computed for Bayesian interval for parameter θ_0 is shorter. Average length increases as the values of θ_0 increases for fixed execution time i.e. t_e . It is also observed that as the values of θ_1 increases average length decreases. From table, we can see that coverage probability increases as the θ_0 increases and coverage probability decreases as the θ_1 increases. From the table it also observes that as execution time increases average length decreases.

Table (5) to (8) gives average length and coverage probability for Bayesian interval for parameter the θ_1 . The Average length have been calculated by assuming values $\theta_0 (=1(1)5)$ and $\theta_1 (= 0.6(0.1)1)$. Average length increases as the values of θ_0 increases for fixed execution time i.e. t_e and it is decreases as θ_1 increases. From table, it can be seen that coverage probability increases as the θ_0 increases and coverage probability decreases as parameter θ_1 increases. From the table it also observes that average length decreases as execution time increases.

Table 1: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{0L}, \tilde{\theta}_{0U})$ of θ_0 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 5$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	2.12412 (0.994)	2.95254 (0.994)	5.07429 (0.994)	5.39784 (0.995)	6.45182 (0.995)
0.7	2.06795 (0.993)	2.18189 (0.994)	4.17136 (0.994)	5.07228 (0.995)	6.69461 (0.995)
0.8	1.90843 (0.993)	2.98570 (0.993)	3.93268 (0.994)	4.72848 (0.994)	6.42171 (0.995)
0.9	1.42513 (0.993)	2.26186 (0.993)	3.56985 (0.994)	4.40744 (0.994)	5.06234 (0.995)
1	1.40031 (0.993)	2.04029 (0.993)	2.28113 (0.993)	3.85033 (0.994)	4.29281 (0.994)

*The values in the parenthesis are coverage probability of true value of parameter.

Table 2: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{0L}, \tilde{\theta}_{0U})$ of θ_0 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 7$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.723187 (0.994)	1.544598 (0.994)	1.58816 (0.994)	3.681955 (0.995)	4.526929 (0.995)
0.7	0.476033 (0.993)	1.222565 (0.994)	1.236041 (0.994)	2.442999 (0.994)	3.921342 (0.995)
0.8	0.470561 (0.993)	0.847506 (0.993)	1.191867 (0.994)	1.889722 (0.994)	2.520656 (0.995)
0.9	0.463198 (0.993)	0.71856 (0.993)	1.092366 (0.994)	1.828802 (0.994)	2.504829 (0.994)
1	0.375691 (0.993)	0.570903 (0.993)	0.795695 (0.993)	1.412542 (0.994)	2.446384 (0.994)

*The values in the parenthesis are coverage probability of true value of parameter.

Table 3: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{0L}, \tilde{\theta}_{0U})$ of θ_0 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 10$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.267986 (0.993)	0.324152 (0.993)	0.465934 (0.993)	0.57678 (0.994)	1.019986 (0.994)
0.7	0.260624 (0.993)	0.322248 (0.993)	0.353603 (0.993)	0.431857 (0.993)	0.906414 (0.994)
0.8	0.260624 (0.993)	0.282712 (0.993)	0.34624 (0.993)	0.421758 (0.993)	0.868347 (0.994)
0.9	0.260624 (0.993)	0.275349 (0.993)	0.338877 (0.993)	0.412504 (0.993)	0.801439 (0.994)
1	0.257986 (0.993)	0.267986 (0.993)	0.316789 (0.993)	0.402406 (0.993)	0.718558 (0.993)

*The values in the parenthesis are coverage probability of true value of parameter.

Table 4: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{0L}, \tilde{\theta}_{0U})$ of θ_0 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 12$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.260624 (0.993)	0.260624 (0.993)	0.309427 (0.993)	0.368328 (0.993)	0.667052 (0.993)
0.7	0.258620 (0.993)	0.275349 (0.993)	0.280105 (0.993)	0.34624 (0.993)	0.591506 (0.993)
0.8	0.256715 (0.993)	0.267986 (0.993)	0.275349 (0.993)	0.331515 (0.993)	0.402406 (0.993)
0.9	0.253609 (0.993)	0.264624 (0.993)	0.267986 (0.993)	0.319525 (0.993)	0.380318 (0.993)
1	0.252286 (0.993)	0.260624 (0.993)	0.260624 (0.993)	0.309427 (0.993)	0.365592 (0.993)

*The values in the parenthesis are coverage probability of true value of parameter.

Table 5: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{1L}, \tilde{\theta}_{1U})$ of θ_1 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 5$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.00605 (0.994)	0.01269 (0.994)	0.02099 (0.994)	0.02412 (0.995)	0.02947 (0.995)
0.7	0.00543 (0.993)	0.01146 (0.994)	0.01713 (0.994)	0.02134 (0.994)	0.02700 (0.995)
0.8	0.00474 (0.993)	0.00859 (0.993)	0.01511 (0.994)	0.02037 (0.994)	0.02694 (0.995)
0.9	0.00387 (0.993)	0.00710 (0.993)	0.01386 (0.994)	0.01912 (0.994)	0.02613 (0.994)
1	0.00348 (0.993)	0.00616 (0.993)	0.01001 (0.993)	0.01895 (0.993)	0.02599 (0.993)

*The values in the parenthesis are coverage probability of true value of parameter

Table 6: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{1L}, \tilde{\theta}_{1U})$ of θ_1 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 7$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.00049 (0.994)	0.00158 (0.994)	0.00252 (0.994)	0.00572 (0.994)	0.00908 (0.994)
0.7	0.00046 (0.993)	0.00106 (0.994)	0.00229 (0.994)	0.00498 (0.994)	0.00890 (0.994)
0.8	0.00041 (0.993)	0.00093 (0.993)	0.00149 (0.993)	0.00401 (0.994)	0.00771 (0.994)
0.9	0.0007 (0.993)	0.00082 (0.993)	0.00139 (0.993)	0.00248 (0.994)	0.00358 (0.994)
1	0.0002 (0.992)	0.00067 (0.993)	0.00109 (0.993)	0.00143 (0.993)	0.00245 (0.994)

*The values in the parenthesis are coverage probability of true value of parameter.

Table 7: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{1L}, \tilde{\theta}_{1U})$ of θ_1 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 10$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.000066 (0.992)	0.000102 (0.992)	0.001163 (0.993)	0.00227 (0.992)	0.00265 (0.993)
0.7	0.000066 (0.992)	0.000071 (0.992)	0.00025 (0.992)	0.000219 (0.992)	0.00184 (0.993)
0.8	0.000065 (0.992)	0.000067 (0.992)	0.00022 (0.992)	0.000207 (0.992)	0.00159 (0.993)
0.9	0.000063 (0.992)	0.000060 (0.992)	0.00021 (0.992)	0.000130 (0.992)	0.000249 (0.993)
1	0.000062 (0.992)	0.000052 (0.992)	0.000041 (0.992)	0.000062 (0.992)	0.000109 (0.993)

*The values in the parenthesis are coverage probability of true value of parameter.

Table 8: Average length and coverage probability of Bayesian interval $(\tilde{\theta}_{1L}, \tilde{\theta}_{1U})$ of θ_1 calculated for different values of parameters θ_0 and θ_1 when execution time $t_e = 12$

$\theta_0 \backslash \theta_1$	1	2	3	4	5
0.6	0.000063 (0.992)	0.000594 (0.992)	0.000603 (0.992)	0.00064 (0.992)	0.00151 (0.993)
0.7	0.0000532 (0.992)	0.0000599 (0.992)	0.0000608 (0.992)	0.000541 (0.992)	0.000179 (0.992)
0.8	0.0000330 (0.992)	0.0000466 (0.992)	0.0000607 (0.992)	0.000116 (0.992)	0.000462 (0.992)
0.9	0.0000204 (0.992)	0.0000229 (0.992)	0.0000385 (0.992)	0.0000602 (0.992)	0.000220 (0.992)
1	0.0000128 (0.991)	0.0000193 (0.992)	0.0000359 (0.992)	0.0000596 (0.992)	0.000124 (0.992)

*The values in the parenthesis are coverage probability of true value of parameter.

5. Conclusion

In this paper, two-sided Bayesian interval has been obtained for length biased exponential class model with parameters i.e. total number of failures θ_0 and scale parameter θ_1 . Bayesian interval is obtained and studied on the basis of average length and coverage probability. It is found that Bayesian interval has shorter average length and high coverage probability. As execution time increases average length decreases and coverage probability increases. From results it is concluded that the proposed Bayesian interval preferred for parameter total number of failures i.e. θ_0 and θ_1 .

References:

- [1] Andure, N.W. and Ade R.B. (2021). The new Length biased quasi Lindley distribution and its applications, *Reliability theory and Applications*, 16:331-345.
- [2] Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies, *Ann. Eugenics*, 6:13-25
- [3] Fitrilia A., Fithriani I. and Nurrohmah1 S. (2018). Parameter estimation for the Lomax distribution using the E Bayesian method, *Journal of Physics*.
- [4] Gradshteyn, I. S and Ryzhik, I. M., Table of Integrals, Series, and Products, Alan Jeffrey (editor) 5th Ed., Academic Press, New York, (1994).
- [5] Gupta, I. and Gupta, R. (2018). Bayesian and Non Bayesian Method of Estimation of Scale Parameter of Gamma Distribution under Symmetric and Asymmetric Loss Functions, *World Scientific News*, 101:172-191.
- [6] Gupta, R. C. and Keating, J. P. (1986). Relations for reliability measures under length biased sampling; *Scandinavian Journal of Statistics*, 13: 49-56.
- [7] Martz, H.F and Waller, R.A. Bayesian Reliability Analysis, New York Wiley (1982).
- [8] Musa, J. D. Iannino, A. and Okumoto K., Software Reliability: Measurement, Prediction, Application, New York, McGraw-Hill (1987).
- [9] Musa, J.D. and Okumoto, K. (1984). A logarithmic Poisson execution time model for software reliability measurement, *Proc. 7th International Conference on Software Engineering, Orlando, Florida*, 230–238.
- [10] Patil, G.P. and Rao, C.R. (1978). Weighted distributions and Size biased Sampling with Applications to Wildlife Populations and Human Families, *Biometrics* 34:179-34189.

- [11] Rabie Abdalla and Junping Li (2019) "E-Bayesian Estimation Based on Burr-X Generalized Type-II Hybrid Censored Data" *Symmetry* 2019, 11, 626;
- [12] Rao, C.R. (1965). On discrete distributions arising out of method of ascertainment in classical and contagious discrete, Pergamum Press and Statistical publishing society, Calcutta. 320 -332.
- [13] Rao. K. A. and. D'cunha, J.G. (2014). Bayesian Inference for Mean of the Lognormal Distribution, *International Journal of Scientific and Research Publications*, 4: 195-203.
- [14] Shreshtha, S.K. and Kumar, V. (2014). Bayesian Analysis for the Generalized Rayleigh Distribution, *International Journal of Statistika and Matematika*, 9: 118-131.
- [15] Singh, R., Singh, P. and Kale, K. (2016). Bayes Estimators for the parameters of Poisson type length biased exponential class model using non- informative priors, *Journal of Reliability and Statistical Studies*, 9: 21-28.
- [16] Tamak, J. (2013). Use of software reliability growth model to estimates the reliability of web applications, *International Journal of Advanced Research in Computer and Software Engineering*, 3(6) 53-59.

RESILIENCE OF A TELECOMMUNICATIONS NETWORK SUBJECTED TO CORRELATED GEOGRAPHICAL FAILURES

DORA JIMÉNEZ¹ AND ABIGAIL MEDINA²



¹Universidad Adolfo Ibáñez. ²Universidad de Carabobo
dora.jimenez.a@edu.uai.cl, acolina6@uc.edu.ve

Abstract

Due to the COVID-19 pandemic, the way in which routine activities are carried out changed, taking a leading role telecommunications networks, it is important to evaluate their operation, service interruption and progressive deterioration, especially those generated by natural disasters, we have focused on earthquakes. Venezuela is a seismic country, being vulnerable to economic and human losses caused by this disaster. Many of the infrastructures that are used by both public and private institutions may not follow the current laws on earthquake-resistant structures. We seek to evaluate the damage by correlated geographic faults produced by earthquakes in a telecommunications network using probabilistic seismic risk analysis.

Keywords: reliability, seismic, simulation

1. INTRODUCTION

In the last three years due to the social isolation caused by the COVID-19 pandemic, the daily tasks, work and academic, have taken a turn driving more and more the need to access and have reliable telecommunications networks, since the activities that were traditionally executed in person, have been replaced by tools that allow to perform these tasks from home via the Internet, in addition to this, the trade could continue thanks to this. In a research carried out in 2021 by Valeria Castro in [1], it points out that the use of Internet increased globally by 70%, in fact it also shows a projection for 2025, and it is expected that this percentage will continue to increase. In [2] the methods currently used to assess individual seismic risk are based on many years of statistical data, they propose an end-to-end calculation-experimental approach to estimate possible losses and individual risk based on actual data on hazard, seismicity and earthquake resistance.

Due to the importance that telecommunication networks have taken, it is relevant to study the fragility of this, in order to know the possibility of partial or complete failure of the network, caused by typical or geographical failures, such as those generated by seismic events. Despite scientific and technological advances, human beings continue to be exposed to natural disasters, leaving in evidence the vulnerability of society to such events. It should be emphasized that these events cannot be avoided; however, thanks to globalization, it is possible to obtain information in real time about the predictions of these disasters (in case it is possible to predict them), or the strata caused by them, and this has been achieved thanks to the progress in telecommunications.

Venezuela throughout history has been affected by high intensity earthquakes, additionally most of the country's population is concentrated in the capital and northwest coast, being these some of the areas of greatest seismic threat [3], since along this region are the fault systems Boconó, San Sebastián and La Victoria, making the buildings that are located there more vulnerable to damage by such events.

Once the motivations for this research have been stated, the study of the vulnerability of the telecommunications network to earthquakes is focused on using as a case study a representation of the Venezuelan state communications network (Compañía Anónima Nacional Teléfonos de Venezuela, CANTV). A phased model is proposed through Monte Carlo simulation to determine the operability of the network components.

2. METHODOLOGY

The seismic hazard of Venezuela was identified in order to select the model that adapts to the geographic faults of the Venezuelan territory. John Douglas, in [4], proposes different functions for the study of ground motion in an earthquake, classifying them according to the zone and type of earthquake. Taking this into account, it was decided to use the equation to calculate the *SA* (spectral acceleration) described by Norman Abrahamson et al. (2016) and BC Hydro (2012) [5]. We follow the approach proposed by [6], where, using probabilistic seismic risk analysis, they evaluate the reliability of a telecommunications network in the event of earthquake failure in Chile, one of the most seismic countries in the world.

In the diagram 1 the processes executed in the simulation are visualized, applying a three-stage model, in the first one the characteristics of the earthquake itself necessary to generate the following stages are generated, some parameters of interest are the moment magnitude, depth and epicenter, subsequently an intensity measure describing the ground motion is generated, In this opportunity we work with the spectral acceleration (*SA*) and in the last one we calculate the fragility curves to obtain the probability that the components of the network suffer a specific level of damage (in this opportunity the levels of mild and extensive damage), which occurred given the spectral acceleration, as established by the HAZUS manual [7], this is achieved using Monte Carlo simulation, these steps are repeated *n* times, to finally calculate the marginal probability of failure of each component of the network.

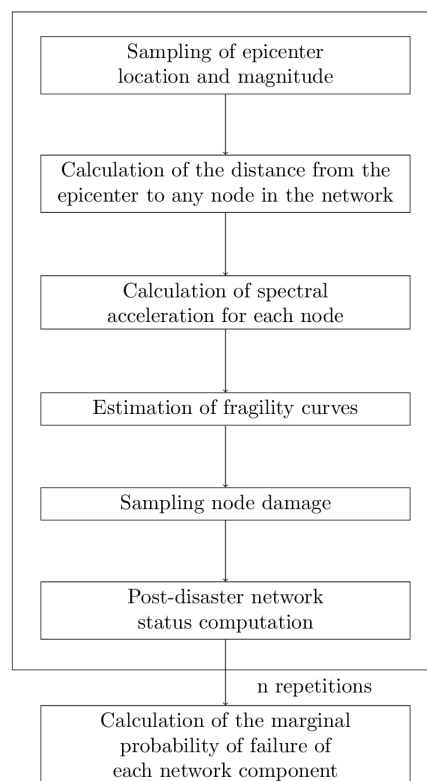


Figure 1: Flowchart of a generic Monte Carlo simulation run

The processes shown in the 1 diagram are presented in more detail below, starting with the model used to simulate ground motion.

2.1. Equation to predict ground motion

I know use the following equation of ground motion from Carlos Arteta et al. in [8], where SA_{ij} is the spectral acceleration at a given period; $f(M_i, r_{ij}, \theta)$ is the attenuation equation as a function of magnitude M , distance r and the model parameter vector Θ ; ϵ_{ij} is the error term for log j of event i and η_i is the random effect for event i .

$$\ln(SA_{ij}) = f(M_i, r_{ij}, \theta) + \eta_i + \epsilon_{ij} \quad (1)$$

The term θ_i represents between-event (or interevent) variations, while ϵ_{ij} represents within-event (or intra-event) variations. The residuals of η , ϵ are uncorrelated and are normally distributed with variances t^2 and f^2 , respectively. The total standard deviation of the ground motion model σ can be expressed as.

$$\sigma = \sqrt{\tau^2 + \sigma^2} \quad (2)$$

2.1.1 Functional form of the ground motion model

The mean of the model posited by Abrahamson (2016) and BC Hydro (2012) in [5], in future reference the abbreviation GMPE-Am2016 will be used for simplification of writing, is described in the following equation for interface and intraplate earthquakes:

$$\begin{aligned} \ln(SA) = & \theta_1 + \theta_4 \delta C_1 + (\theta_2 + \theta_{14} F_{event} + \theta_3 (M - 7.8)) \ln(R_{rup, hypo}) + (M - 6)) \\ & + C_4 \exp(\theta_9 + \theta_6 R_{rup, hypo} + \theta_{10} F_{event} + f_{mag}(M) + f_{depth}(R_{hypo}) F_{event} + \\ & + f_{FABA}(R_{rup, hypo}) + f_{site}(PGA_{1000}, V_{S30}) \end{aligned} \quad (3)$$

- SA is the spectral acceleration in gravity.
- M is the magnitude of the momentum.
- R is the event-dependent distance, e.g., R_{rup} the rupture distance for the interface and R_{hypo} the hypocentral distance for intraplate events.
- θ_j is the dependent event.
- $F_{event} = \begin{cases} 0 & \text{for interface events} \\ 1 & \text{for intra-board events} \end{cases}$
- $F_{FABA} = \begin{cases} 0 & \text{foreground or unknown sites} \\ 1 & \text{posterior arc sites} \end{cases}$

Model for magnitude scaling

$$f_{mag}(M) = \begin{cases} \theta_4 (M - (C_1 + \delta C_1)) + \theta_{13} (10 - M)^2 & \text{para } M \leq C_1 + \delta C_1 \\ \theta_5 (M - (C_1 + \delta C_1)) + \theta_{13} (10 - M)^2 & \text{otros casos} \end{cases} \quad (4)$$

Where $C_1 = 7.8$. The values of C_1 capture the epistemic uncertainty of magnitudes.

Model for scaling depth

$$f_{depth}(R_{hypo}) = \theta_{11} (\min(R_{hypo}, 120)) F_{event} \quad (5)$$

Model for scaling the fore/back arc:

$$f_{FABA}(R) = \begin{cases} \theta_7 + \theta_8 \ln\left(\frac{\max(R_{hypo}, 85)}{40}\right) F_{FABA} & \text{para } F_{event} = 1 \\ \theta_{15} + \theta_{16} \ln\left(\frac{\max(R_{rup}, 100)}{40}\right) F_{FABA} & \text{para } F_{event} = 0 \end{cases} \quad (6)$$

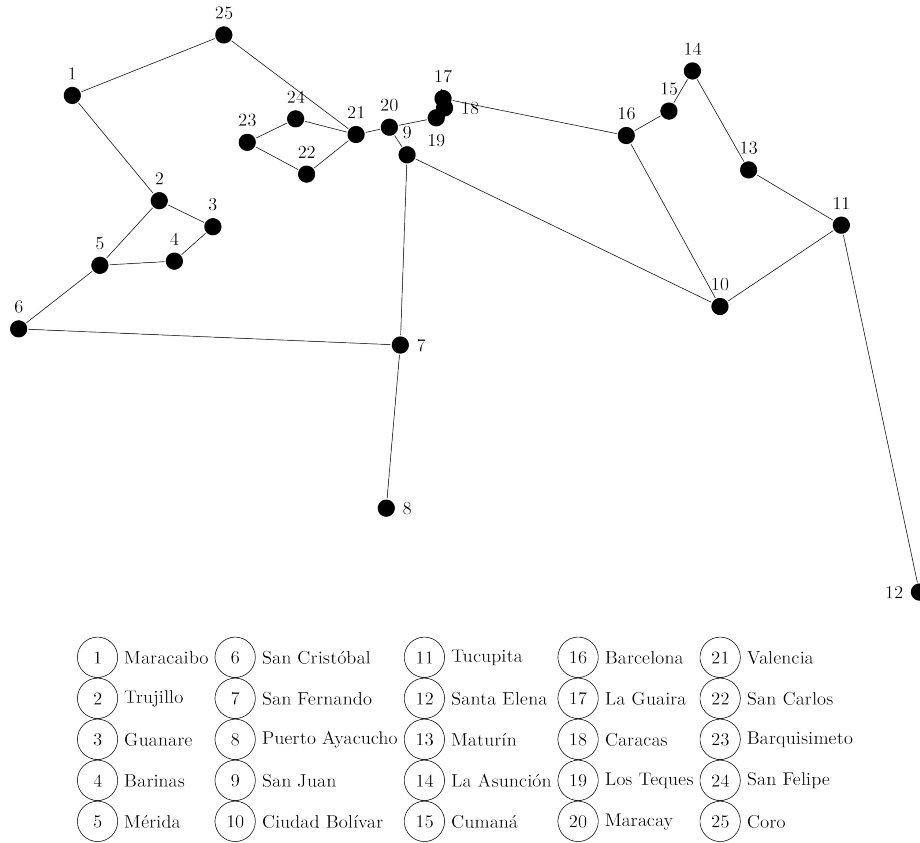


Figure 2: Venezuela's simplified telecommunications network graph

Model for scaling site response,

$$f_{site}(PGA_{1000}, V_{S30}) = \begin{cases} \theta_{12} \ln \frac{V_S^*}{V_{lin}} - b \ln PGA_{1000} + c + & \text{para } V_{S30} < V_{lin} \\ +b \ln PGA_{1000} + c \left(\frac{V_S^*}{V_{lin}} \right)^n & \\ \theta_{12} \ln \frac{V_S^*}{V_{lin}} - b \ln \frac{V_S^*}{V_{lin}} & \text{para } V_{S30} \geq V_{lin} \end{cases} \quad (7)$$

where PGA_{1000} = is the PGA median f $V_{S30} = 1000$ m/s and

$$V_S^* = \begin{cases} 1000 & \text{para } V_{S30} > 1000 \\ V_{S30} & \text{para } V_{S30} \leq 1000 \end{cases} \quad (8)$$

In this research for the calculation of the SA, the equation for the calculation of interface type earthquakes was used. Likewise, a model was proposed in which the different components of the Venezuelan telecommunications network are evaluated by damage levels. With the objective of calculating the marginal probability of failure for the nodes (cities of the network).

To calculate the probability of operability of each component it is necessary to use fragility curves, this corresponds to the graphical representation of the cumulative distribution function, also seen as the probability of reaching or exceeding a previously defined state of damage. In the case of earthquakes, such fragility curves have been defined in the HAZUS manual [7].

A telecommunication network can be viewed as a graph composed of nodes and arcs, the arcs connecting nodes to each other. The capitals of each state that make up the country and an additional city (Santa Elena de Uairén) were considered as the nodes of the network 2.

To perform the simulation of ground motion within an acceptable margin, with the GMPEAm-2016 model, a historical seismic base is necessary, for which it is essential to use a database that has a seismic research institution either national or international.

We use the data provided by the USGS (United States Geological Survey) Earth Explorer. Considering seismic events with magnitudes between 6 and 8, from 1900 to 2019, as shown in Figure 3.

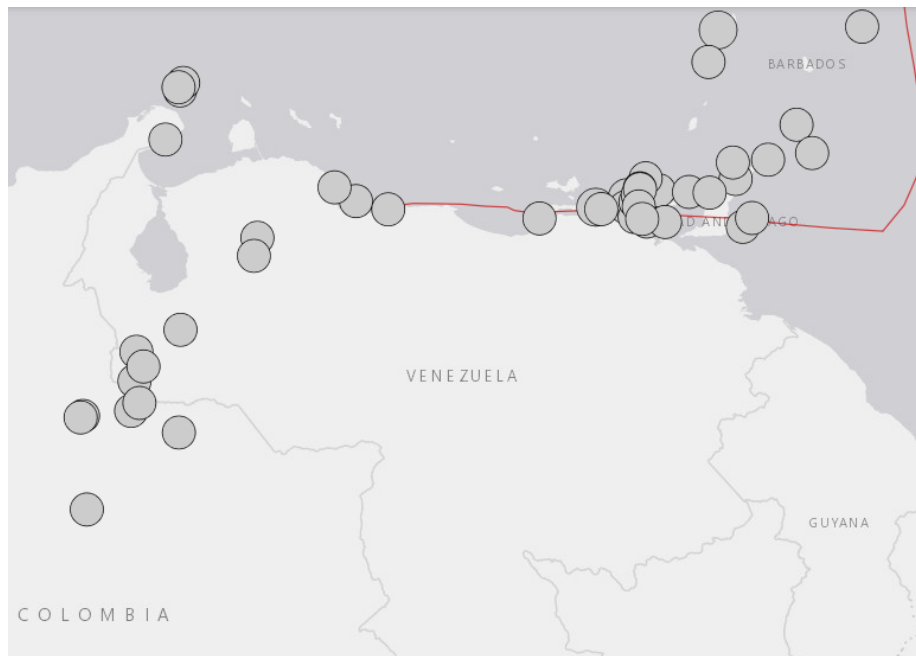


Figure 3: Geolocation of earthquakes between magnitude 6 and 8, image taken from [9]

Once the data was obtained, the ground motion simulation was performed, then the conditional probability was defined to evaluate the possibility of the component being operative after the earthquake. The probability of damage occurring, given that a displacement occurred in the component under study, is calculated using the equation for estimating fragility given by the HAZUS manual [7]. After obtaining the probability of failure, the state of the component was evaluated, for which a stochastic binary system is defined. Finally, the structure function for the Venezuelan telecommunications network was performed to calculate the marginal probability of each component. Finally, by quantifying the expected damage at each node, the corresponding analyses were performed to better understand the state of the network.

3. RESULTS

For this research, the RStudio program was used as a tool for the simulation, the calculations of the probabilities, as well as the graphs that are presented. The calculations and graphs were obtained by means of codes implemented in the aforementioned program using an Intel Core i3 – 3110M 1.4G laptop computer with 4GB of RAM memory. One thousand (1000) simulations were performed, a process that took 1.3399 seconds to execute, indicating that the associated computational cost is low.

For the mapping of the earthquakes, the data presented in Table 1, taken from [9], were used.

The locations used for the nodes are shown in the table below 2

For ground motion simulation, the GMPE-Am2016 model was implemented for interface type earthquakes, in the following Tables the parameters used for this model are presented.

Table 3, which was taken from [5], shows the coefficients to be used for the GMPE-Am2016 model, regardless of the period being worked with.

Table 1: Earthquakes (6 – 8 Mw), from 1900 to 2019, data obtained from [9]

Latitude	Longitude	Depth (Km)	Moment Magnitude
10.7731	-62.9019	146.82	7.3
6.7757	-72.9875	155	6.2
10.9048	-62.3150	63	6.0
10.7090	-67.9270	14	6.4
10.8760	-61.7560	53	6.1
11.1240	-62.5590	110.3	6.2
10.5980	-63.4860	19.9	7.0
11.1120	-60.8920	5	6.7
11.4120	-60.9420	45	6.1
5.0500	-72.9160	17.3	6.5
7.4140	-72.0330	11.6	6.0
10.2410	-60.7580	36.0	6.2
6.4700	-71.2100	20.3	6.1
10.9600	-68.3250	20.3	6.0
10.4020	-60.5870	56.2	6.7
10.5970	-62.9280	18.8	6.3
10.4190	-62.7640	40.0	6.1
10.0400	-69.7550	33.0	6.1
10.5630	-63.3820	34.0	6.1
9.4770	-72.5880	175.7	6.1
10.6970	-62.7480	103.4	6.5
10.5590	-67.3300	25.0	6.6
6.7470	-73.0320	161.2	6.8
10.3360	-62.5400	15.0	6.1
10.9420	-62.6680	15	6.0
10.8240	-62.7060	23.3	6.4
6.8680	-72.0950	25.0	6.6
8.3570	-71.1810	15.0	6.1
10.8630	-61.3740	49.0	6.1
7.0130	-71.9410	28.2	6.5
9.7100	-69.8190	15.0	6.4
10.5430	-64.4440	10.0	6.7
10.3620	-62.8040	20.0	6.3
11.8290	-71.4610	15	6.3
11.0000	-66.0000	0.0	7.7

Table 2: Geographic location of Venezuelan telecommunications network nodes

Number	City	Latitude	Longitude
1	Coro	11.4046	-69.6563
2	Valencia	10.1725	-67.9941
3	Maracay	10.2669	-67.6052
4	Los Teques	10.3486	-67.0344
5	Caracas	10.5002	-66.9191
6	La Guaira	10.6025	-66.9308
7	San Felipe	10.3400	-68.7452
8	Barquisimeto	10.0680	-69.3475
9	San Carlos	9.6600	-68.5811
10	San Juan de los Morros	9.9127	-67.3613
11	Ciudad Bolívar	8.0886	-63.5536
12	Barcelona	10.1450	-64.6783
13	Cumana	10.4322	-64.1833
14	La Asunción	11.0278	-63.8583
15	Maturín	9.7358	-63.1919
16	Tucupita	9.0697	-62.0456
17	San Fernando de Apure	7.8808	-67.4692
18	San Cristóbal	7.77194	-72.2264
19	Mérida	8.5700	-71.1808
20	Barinas	8.6231	-70.2372
21	Guanare	9.0378	-69.7289
22	Trujillo	9.3691	-70.4396
23	Maracaibo	10.645	-71.6131
24	Puerto Ayacucho	5.6625	-67.5828
25	Santa Elena de Uairén	4.6081	-61.1072

Table 3: Period-independent subduction model coefficients used in the regression analysis, Table taken from [5]

Coefficients	Value in all the periods
n	1.18
θ_3	0.1
θ_4	0.9
θ_5	0.0
θ_9	0.4
C_1	7.8
dC_1	0.2
C_4	10

Table 4: Regression coefficients for the median (units of g) of the ground motion model for interface earthquakes, taken from [5]

Coefficients	Value in all the PGA_{1000}	Value for the SA
Period	0.0000	0.2500
V_{lim}	865.1	654.3
b	-1.186	-2.381
θ_1	4.2203	5.0594
θ_2	-1.3500	-0.9940
θ_6	-0.0012	-1.3000
θ_7	1.988	2.8000
θ_6	-1.4200	0.0129
θ_8	0.9996	-1.3000
θ_{10}	3.1200	2.800
θ_{11}	0.0130	0.0129
θ_{12}	0.9800	2.4800
θ_{13}	-1.4200	-0.0172
θ_{15}	0.9996	1.1600
θ_{16}	-1.0000	-1.1700

In Table 4, the coefficients that depend on the period are shown, being equal to 0 used for the calculation of the PGA_{1000} and equal to 0.25 for the calculation of the SA.

Since the spectral acceleration was calculated for interface earthquakes, the coefficient θ_{14} has been omitted, since it is not necessary for this type of telluric movement. The value of V_{S30} used corresponds to [200,850], taking into consideration the works of Escobar et al. in [10] and [11], that of Morales and several authors in [12], and that of Acosta together with other authors in [13].

For the fragility curves, the following parameters presented in Table 5, selected from the HAZUS Manual [7] for reinforced concrete buildings (C2L), since the main CANTV telecommunications network exchanges are of this type of structures, two levels of damage were taken into consideration, slight and extensive, the first refers to cracks in the surfaces of the walls or small detachments of the concrete in some places, while the second is when most of the walls have exceeded their creep capacity, thus indicating cracks that go through the wall, extensive spalling around the cracks and wall reinforcement visibly bent or rotation of narrow walls with inadequate foundations, and partial collapse of the walls, the units are given in inches, so the corresponding conversion must be made, taking into account that the spectral acceleration is given in g .

Table 5: Structural parameters of the fragility curve

	Displacement Spectral (Inches)	
	Slight	Extended
$\bar{S}_{d,ds}$	0.72	1.04
β_{ds}	3.55	0.99

An Unif [0,1] was sampled to evaluate the state of the network components, i.e., whether there was damage or not, for which a dynamic binary system was established, taking the number one (1) to indicate that damage actually occurred in the node, and zero (0) for the opposite scenario. Once the thousands (1000) repetitions had been carried out, the marginal probability of each node receiving a slight or extensive level of damage was calculated. The following results were

obtained:

In Figures 4(a) and 4(b), the horizontal axis corresponds to the numbering assigned in the network for each city (see Figure 2 or Table 2), and the vertical axis to the number of times a failure occurred.

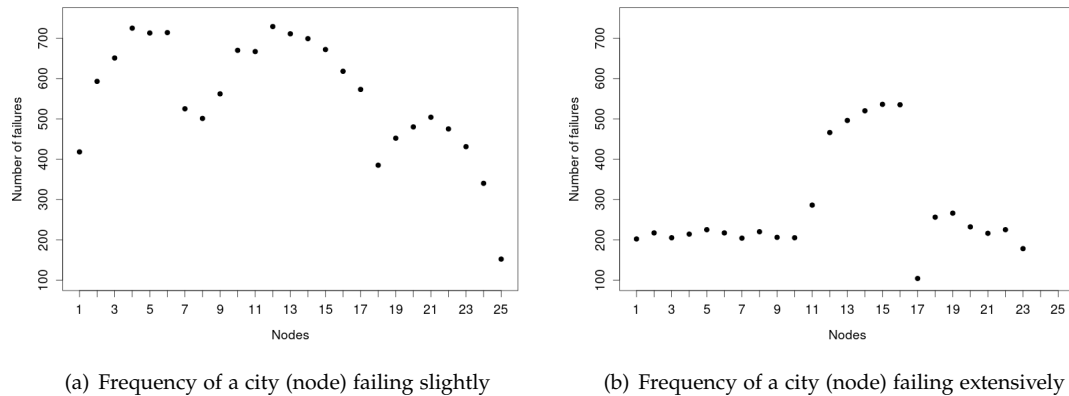


Figure 4: Frequency of occurrence of failure per damage level at each node

In Figures 4(a) and 4(b) three facts are observed, the first is that it is more frequent that a node receives a slight damage, as opposed to one of extensive level, second is that in Figure 4(a) the nodes that were damaged more than 600 times, were those corresponding to the following cities: Valencia (2), Maracay (3), Los Teques (4), Caracas (5), La Guaira (6), San Felipe (7), San Juan de los Morros (10), Ciudad Bolivar (11), Barcelona (12), Cumana (13), La Asunción (14) and Maturin (15), Los Teques and Barcelona being the two most recurrent nodes to receive slight damage, and close to 600 are San Carlos (9), Tucupita (16) and San Fernando de Apure (17), finally, in Figure 4(b) shows that La Asunción, Maturín and Tucupita received more than 500 times extensive damage, Barcelona and Cumaná between 400 and 500 times. Thus obtaining the probabilities presented in Figure 5, again the horizontal axis corresponds to the nodes, while the vertical axis in these graphs corresponds to the probability of failure.

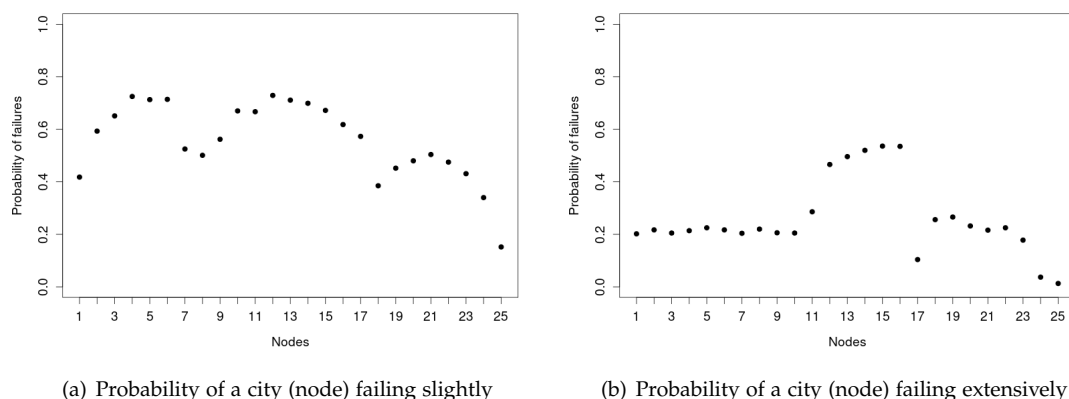


Figure 5: Probabilities of failure per damage level at each node

When comparing 5(a) and 5(b), with the geolocation of the earthquakes shown in the map 3, a geographical correlation is evidenced in which the nodes that have greater probability of

suffering damage, whether slight or extensive, are those that are closer to the areas with greater seismic activity; since the nodes that are observed in the graph 5(a) with a probability greater than 0.6 are mostly belonging to the capital, central and eastern regions of the country, and it is precisely these regions or near them that a greater number of earthquakes occur due to the system of active faults Boconó (Los Andes), San Sebastián (north-central Venezuela), and El Pilar (northeast of the country), their locations can be seen in the image 6.

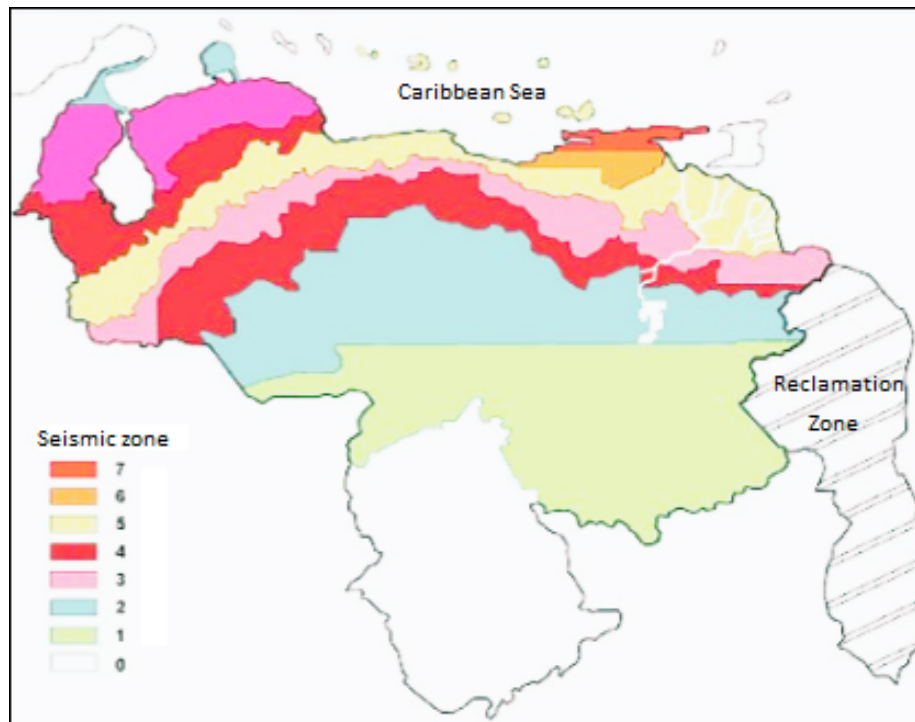


Figure 6: Seismic zonation map of Venezuela

The nodes with a probability of extensive damage greater than 0.5, shown in Figure 5(b), are those corresponding to La Asunción Maturín and Tucupita, which are located in one of the high seismic hazard zones (indicated by the orange color), as shown in the map 6. This correlation is also observed when analyzing the nodes with lower probability of damage in the two different levels, being the most notorious the Santa Elena de Uairén node (25) which has a slight probability of damage equal to 0.152, located in a zone of low seismic hazard (green color); and a probability of extensive damage to 0.013, another node with a low probability of extensive damage is Puerto Ayacucho (24) with an occurrence equal to 0.037, and also located in a zone with seismic hazard similar to the previous one.

In the graph 7 the fragility curves for C2L buildings are shown, where the blue curve represents the probability that given a spectral acceleration the mild damage state is exceeded or reached, and the green one for extensive damage, on the horizontal axis the SA is expressed and on the horizontal axis the conditional probability that a damage state is exceeded given the intensity measure, in this case the SA. It is clear from this graph that slight damage has a higher probability of occurrence than extensive damage, which is the expected result.

4. CONCLUSIONS

A study was carried out on the operability of the components of the Venezuelan telecommunications network when affected by earthquakes. For this purpose, a model was implemented to calculate the spectral acceleration in order to perform a simulation of ground motion, taking into consideration the characteristics of the infrastructure and geography, and thus be able to evaluate

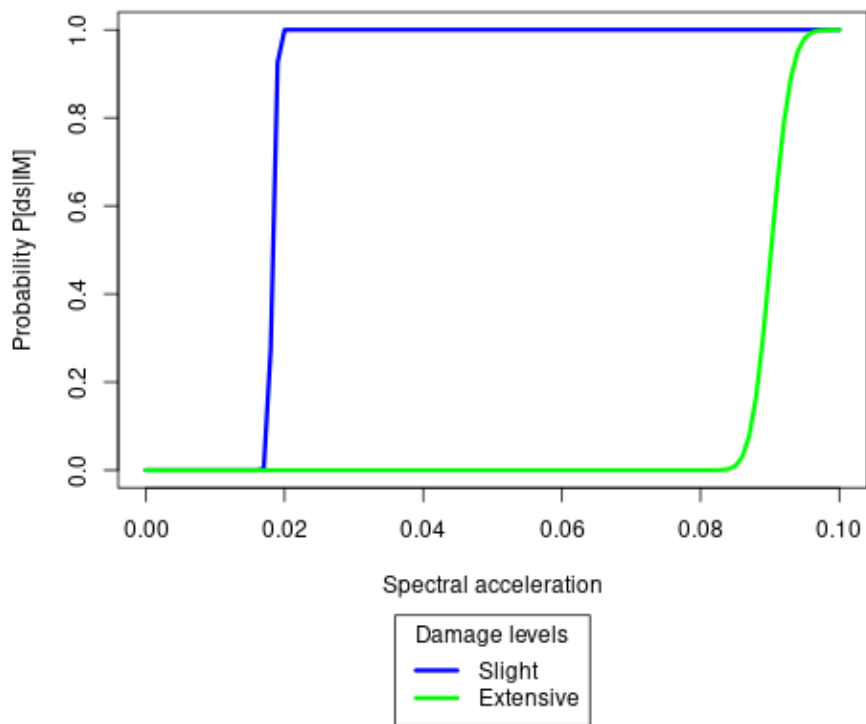


Figure 7: Fragility curves for mild and extensive damage levels

the probability of exceeding a previously defined damage state (in this opportunity, light and extended damage were established), and thus obtain the marginal probability of each component receiving a level of damage.

The network was tested with specific earthquakes (6 – 8 Mw), obtaining as a result that the probability that the nodes of the network suffer a slight or extensive level of damage is geographically related to the location of the earthquakes, so they have a higher probability of slight damage compared to extensive damage.

It is suggested to consider possible damage to the network arcs, since it is known that earthquakes can cause damage to these components. This study represents a novel proof of concept for the Venezuelan case and serves as a starting point for further studies.

ACKNOWLEDGMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] Valeria Castro Obando. Regulación, infraestructura y telecomunicaciones en la crisis del covid-19.
- [2] Gennadiy Nigmatov, Andrey Savinov, and Temir Nigmatov. Assessment of individual seismic risk for the population, taking into account the actual seismic resistance of buildings and the seismicity of soils. *Reliability: Theory & Applications*, 17(SI 4 (70)):172–179, 2022.
- [3] Omar Pérez. Sismicidad y tectónica en Venezuela y áreas vecinas, 1998.
- [4] Jhon Douglas. Ground motion prediction equations 1964–2021, May 2021.
- [5] Norman Abrahamson, Nicholas Gregor, and Kofi Addo. BC Hydro ground motion prediction equations for subduction earthquakes, 2016.
- [6] Dora Jiménez, Javiera Barrera, and Héctor Cancela. Communication network reliability under geographically correlated failures using probabilistic seismic hazard analysis. *IEEE Access*, 2023. Accepted.
- [7] MH HAZUS. Multi-hazard loss estimation methodology: earthquake model, 2003.
- [8] Carlos Arteta, Cesar Pajaro, Vicente Mercado, Julián Montejo, Mónica Arcila, and Norman Abrahamson. Ground-motion model for subduction earthquakes in northern South America. *Earthquake Spectra*, 37(4):2419–2452, 2021.
- [9] Usgs.Gov. Latest earthquakes. <https://earthquake.usgs.gov/>.
- [10] Marysol Mijares, Michael Schmitz, Javier Sanchez, and Freddy Rondón. Estudio del parametro VS30 mediante iMASW en la ciudad de Valencia.
- [11] Víctor Escobar, Michael Schmitz, Javier Sánchez, and Freddy Rondón. Estudio de interferometría del análisis multicanal de ondas superficiales (iMASW) para VS30 en Maracay study of interferometric analysis of surface waves (iMASW) for VS30 in Maracay.
- [12] Cecilio Morales, Julio Hernández, Michael Schmitz, Víctor Cano, and Mauricio Tagliaferro. Velocidades promedio de ondas de corte en los primeros 30 m de profundidad (V_{s30}), inferidas a partir del relieve en el área metropolitana de Caracas. *Revista de la Facultad de Ingeniería Universidad Central de Venezuela*, 26(2):161–168, 2011.
- [13] Rafael Acosta, Freddy Rondón, and Michael Schmitz. Mapa de VS30 de Carúpano a partir de la topografía v_{s30} map of Carúpano based on the analysis of the topography.

A NEW ZERO-TRUNCATED DISTRIBUTION AND ITS APPLICATIONS TO COUNT DATA

Na Elah¹ Peer Bilal Ahmad*² Muneeb Ahmad Wani³

•
Department of Mathematical Sciences,
Islamic University of Science & Technology-192122, Kashmir^{1,2,3}
naaelashah@gmail.com
bilalahmadpz@gmail.com*
wanimuneeb163@gmail.com

Abstract

Numerous disciplines, including engineering, public health, sociology, psychology, and epidemiology, are particularly interested in the analysis and modelling of zero truncated count data. As a result, we suggest a novel and straightforward structural model in this study called zero truncated new discrete distribution. We examine its statistical properties including probability mass function, cumulative function, and moments. The parametric estimation of the zero-truncated new discrete distribution is explained by Maximum Likelihood Estimation method and, to investigate its performance, a simulation study is proposed. The importance of the distribution is evaluated using two real-world data sets as well as one simulated data set and the model comparison is made on the basis of AIC and BIC criterions.

Keywords: truncation, zero-truncated distribution, simulation maximum likelihood estimation, goodness-of-fit

1. INTRODUCTION

The truncation of probability distributions is a significant statistical phenomenon that is employed in numerous domains, including medicine, reliability theory, industry, queueing systems, and many others. When a range of probability values for the variables is either ignored or unobservable, probability models are truncated. For instance, if we want to see how many journal articles across different disciplines have been published, how many tickets were given out to teenagers based on their academic performance, how many children have ever been born to a sample of mothers who are over 40, etc. Typically, the Zero Truncated Poisson(ZTP), Zero Truncated Negative Binomial(ZTNB), and Zero Truncated Poisson Lindley(ZTPL) distributions are used to model the aforementioned scenarios. For the examination of gall-cell counts and amount of eggs in flower heads, Finney and Varley [7] used the ZTP distribution. Brass [2] used the ZTNB distribution to simulate the number of children ever born to a sample of mothers above the age of 40. The ZTNB distribution was used in a regression model by Lee *et al.* [15] to examine the over-dispersed data of ischemic stroke hospitalizations. ZTNB distribution was also used by Phange and Loh [18] to analyze the prevalence of rare species and hospital stay. In 1990, Creel and Loomis [5] used the ZTP distribution and applied it to deer hunting data set which was collected in California. Also, Lindsay [14] analyzed the postal survey data using ZTP distribution. Further, ZTPL distribution was introduced by Ghitany *et al.* [9] in 2008 and a real count data set was analyzed.

The zero truncated models outlined above are seen to have neither a high peak nor a heavy tail. These models do not perform well when over and under-dispersion problems are present in the data. Some other researchers who have worked in this field are Coleman and James [4], Mathews and Appleton [17], they have discussed various applications of zero-truncated distributions particularly ZTP distribution. Further, a detailed discussion was given by Best *et al.* [3] in 2007 on the applications including goodness-of-fit of ZTP distribution. Hassan *et al.*[10] in 2008 obtained the Bayes estimator and reliability function of the ZTP and also derived its recurrence relations. In 2008, Ghitany *et al.* [9] introduced the ZTPL distribution and also derived the method of moments (MoM) and maximum likelihood estimators (MLE) of the parameter including their large sample properties and simulation procedure. In addition to this, the effectiveness of the MoM and ML estimators have been compared, it has been found that the ML estimators are more effective than the MoM estimators. In 2017, Shanker & Shukla [21] introduced a Zero Truncated Two Parameter Poisson Lindley(ZTTPPL) distribution using compounding technique by mixing Size Biased Poisson(SBPD) with a continuous distribution. They have showed that ZTTPPL model gives better fitting than ZTPD in case of biological science data. Simon and Shanker [24] in 2018, obtained a Zero Truncated Discrete Lindley(ZTDL) distribution and analyzed that both the methods i.e., MoM and MLE gave the same estimates of the distributions parameter. In 2020, Kiani [13] proposed a new model, named as Zero Truncated Two Parameter Discrete Lindley(ZTTPDL) distribution. The distribution has some embedded models and represents a two-component mixture of a Zero Truncated Geometric(ZTG) distribution and a ZTNBD with certain parameters. Another distribution named as Poisson Ishita distribution(PID) was obtained by Shukla & Shanker [25] which was further studied by Kamlesh *et al.* [27] in 2020. They obtained the zero truncation of the PID model. Similarly, Shanker [22] in 2017 obtained a zero truncation for Poisson-Akash model of Shanker [23]. Sium in 2020 [26] obtained a zero-truncated model of discrete Akash distribution and obtained its different structural properties. Now some of the aforementioned shortcomings served as the motivation for the proposal of a Zero Truncated New Discrete (ZTND) distribution with a basic structure and only one parameter. The proposed model is equi-dispersed, over-dispersed as well as under-dispersed. Additionally, the model's adaptability is examined by looking at how well it agrees with some real data sets as shown by checking its p-value, and some criterions like AIC and BIC. For this purpose, three data sets including one simulated data are examined , and it has been found that the proposed model, which displays minimum values for certain statistics, is the most appropriate distribution among the others.

2. ZERO TRUNCATED NEW DISCRETE DISTRIBUTION

Sanjay *et al.*[12] in 2021, obtained a new discrete version of the failure model which was given by Siddiqui in 2016 [20] by using the cdf and reliability function of the Continuous Distribution (CD).

Let y be a random variable following new discrete distribution [12] then pmf is given as:

$$P_0(y; \theta) = \frac{\theta - 1}{\theta^{(y+1)}} \quad ; y = 0, 1, 2, 3, \dots; \quad ; \theta > 1 \quad (1)$$

Now, the formula given below can be used to determine the pmf of ZTND distribution.

$$P(y; \theta) = \frac{P_0(y; \theta)}{1 - P_0(0; \theta)} \quad (2)$$

Using equation (1) and (2) we get the pmf of ZTND as:

$$P(y; \theta) = \frac{\theta - 1}{\theta^y} \quad ; y = 1, 2, 3, \dots \quad \theta > 1 \quad (3)$$

The cdf of ZTNDD is provided as:

$$F(y; \theta) = \frac{\theta^y - 1}{\theta^y}$$

$$F(y; \theta) = \frac{\theta^y - 1}{\theta^y}$$

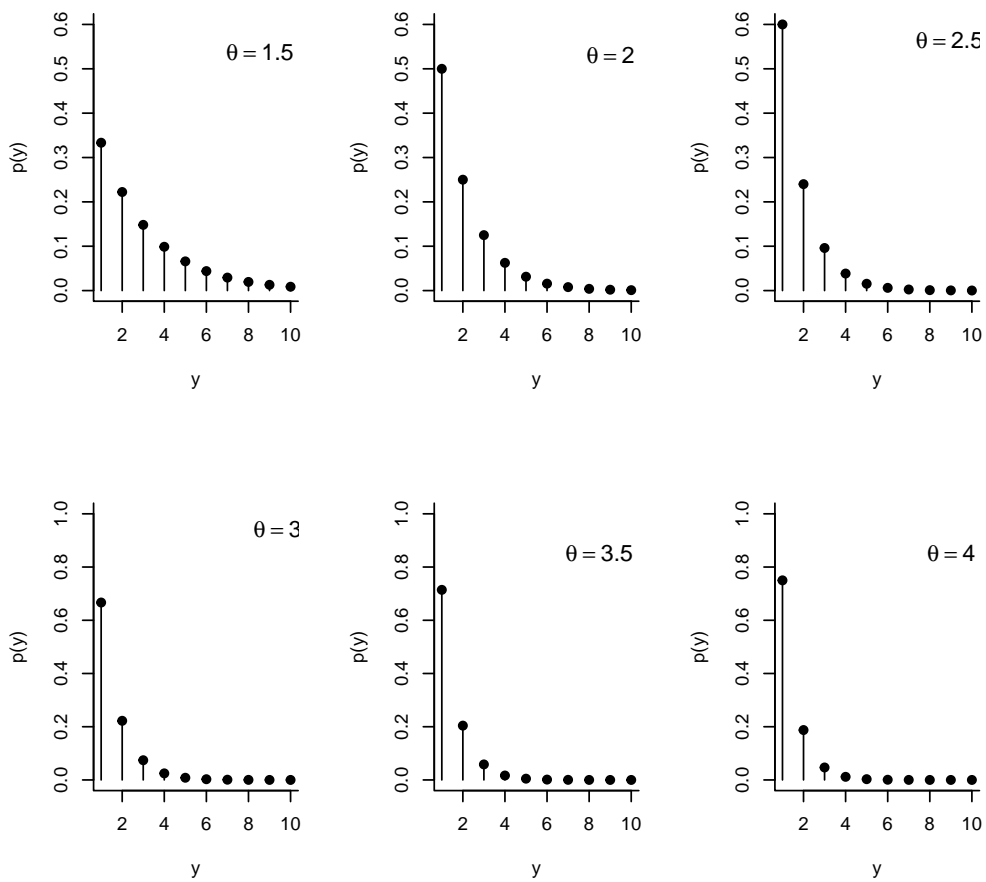


Figure 1: PMF plots of ZTNDD

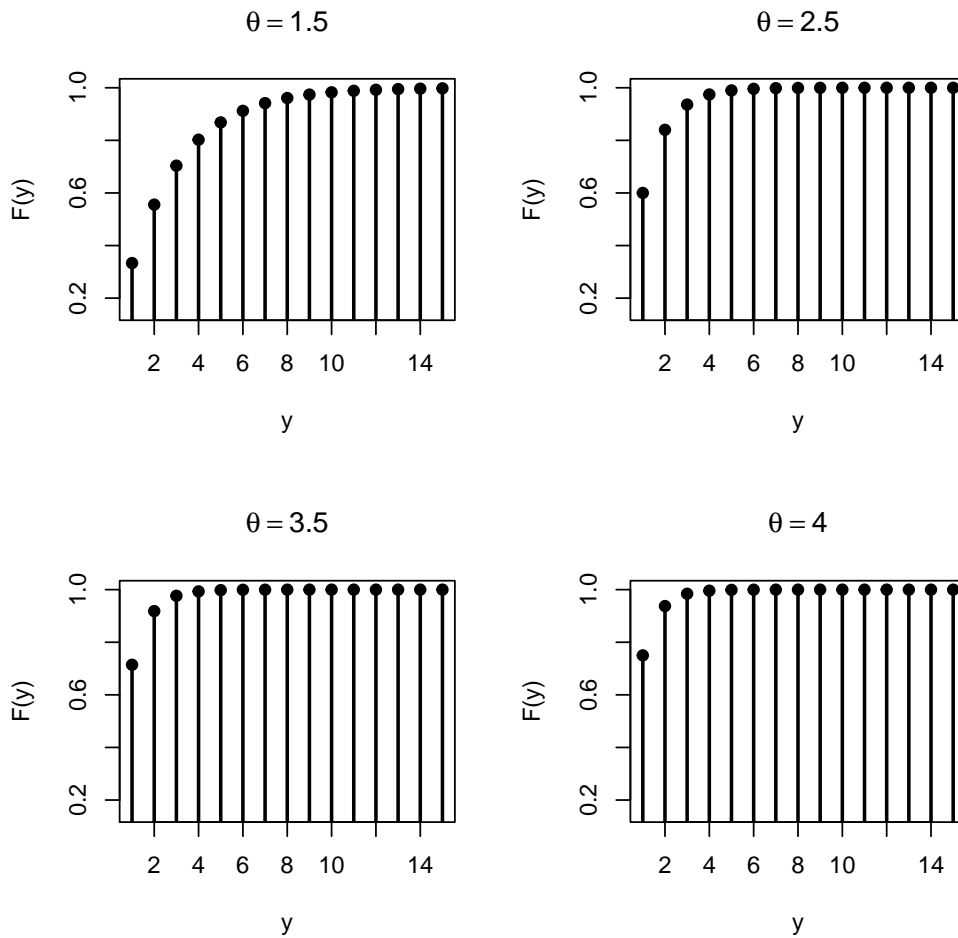


Figure 2: CDF plots of ZTNDD

3. STRUCTURAL PROPERTIES AND GENERATING FUNCTIONS

In this part, we acquired various structural properties including factorial moments and raw moments. Also, MGF is obtained and is given in next subsection.

3.1. Raw Moments/Moments about Origin

The first four raw moments of the proposed model are:

$$m'_1 = \frac{\theta}{\theta - 1}$$

$$m'_2 = \frac{\theta^2 + \theta}{(\theta - 1)^2}$$

$$m'_3 = \frac{\theta(\theta^2 + 4\theta + 1)}{(\theta - 1)^3}$$

$$m'_4 = \frac{\theta(\theta^3 + 11\theta^2 + 11\theta + 1)}{(\theta - 1)^4}$$

3.2. Factorial Moments

The four factorial moments are:

$$m^{(1)} = \frac{\theta}{\theta - 1}$$

$$m^{(2)} = \frac{2\theta}{(\theta - 1)^2}$$

$$m^{(3)} = \frac{6\theta}{(\theta - 1)^3}$$

$$m^{(4)} = \frac{24\theta}{(\theta - 1)^4}$$

3.3. Moment Generating Function(M.G.F)

From the formula given below, we can get the M.G.F of proposed model as:

$$M_y(t) = E(e^{ty}) = \sum_{y=1}^{\infty} e^{ty} P(Y = y)$$

$$\Rightarrow M_y(t) = \frac{(\theta - 1)}{\theta} \sum_{y=1}^{\infty} \frac{e^{ty}}{\theta^y}$$

$$\Rightarrow M_y(t) = \frac{(\theta - 1)e^t}{(\theta - e^t)}$$

From the above calculations, the mean and variance of the proposed model are:

$$E(y) = m'_1 = \frac{\theta}{\theta - 1} \tag{4}$$

$$V(y) = \sigma^2 = \frac{\theta}{\theta^2 + 1 - 2\theta}$$

Index of Dispersion(IoD) is given as

$$IoD = \frac{1}{\theta - 1}$$

Skewness and Kurtosis are given as:

$$\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3} = \frac{(1 + \theta)^2}{\theta}$$

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{(\theta^2 + 7\theta + 1)}{\theta}$$

Coefficient of Variation for the proposed model is:

$$C.V = \frac{1}{\sqrt{\theta}}$$

Figure 3: Index of Dispersion plot for ZTNDD

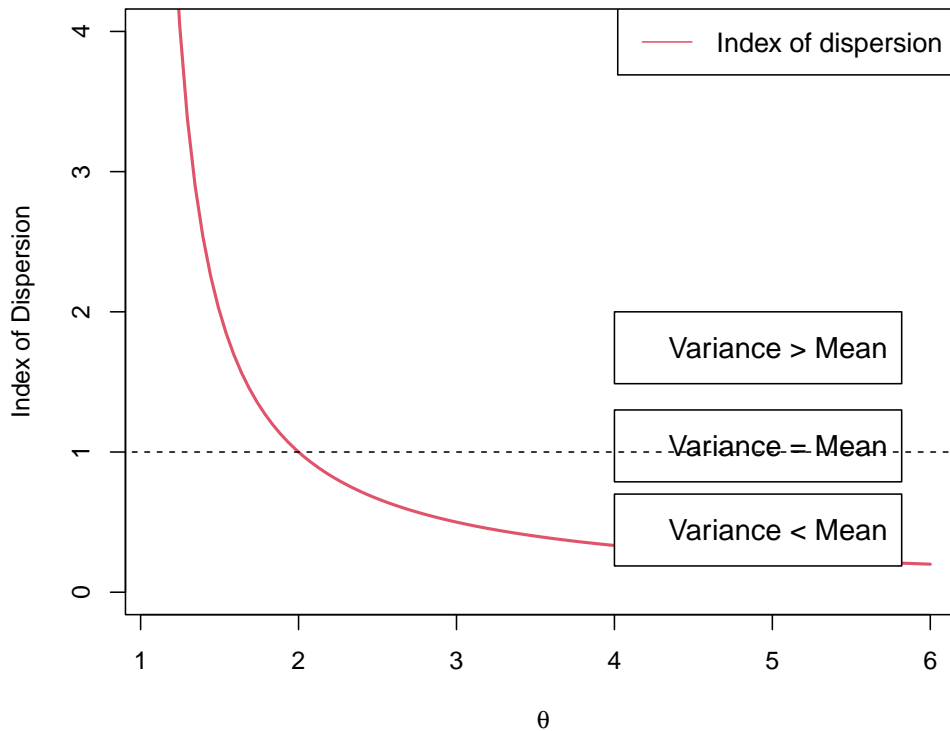


Table 1: Behaviour of the model's descriptive statistics for various parameter values.

θ	Mean	Variance	C.V	Skewness	Kurtosis	IoD
1.5	3.0000	6.0000	0.8165	4.1667	9.1667	2.0000
2.0	2.0000	2.0000	0.7071	4.5000	9.5000	1.0000
2.5	1.6667	1.1111	0.6324	4.9000	9.9000	0.6667
3.0	1.5000	0.7500	0.5774	5.3333	10.3333	0.5000
3.5	1.4000	0.5600	0.5345	5.7857	10.7857	0.4000
4.0	1.3333	0.4444	0.5000	6.2500	11.2500	0.3333
4.5	1.2857	0.3673	0.4714	6.7222	11.7222	0.2857
5.0	1.2500	0.3125	0.4472	7.2000	12.2000	0.2500

From the table, it is clear that our model is positively skewed and leptokurtic. Also, the model is equi-dispersed for ($\theta=2$), underdispersed for ($\theta>2$) and over-dispersed for ($1<\theta<2$).

4. PARAMETRIC ESTIMATION

One of the key problems in mathematical statistics is the parameter estimation. In this segment, we will look at the parametric estimation of the ZTNDD using the maximum likelihood estimation method as well as the method of moments.

4.1. Maximum Likelihood Method of Estimation

Let's take a sample $y_1, y_2, y_3, \dots, y_n$ from ZTNDD of size n with parameter θ . The likelihood function is given by

$$L = \frac{(\theta - 1)^n}{\theta^{\sum y_i}}$$

$$\Rightarrow \log L = n \log(\theta - 1) - \log \theta \sum_{i=1}^{\infty} y_i$$

Differentiating above equation w.r.t. θ and equating to zero we get

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta - 1} - \frac{\sum_{i=1}^{\infty} y_i}{\theta} = 0$$

$$\Rightarrow \hat{\theta} = \frac{\bar{y}}{\bar{y} - 1}$$

4.2. Method of Moments

We must compare the first sample moment with the corresponding population moment in order to estimate the unknown parameter of the ZTND model using the technique/method of moments. Now, replacing μ by \bar{y} in equation (4) we get the moment estimate of the parameter θ as:

$$\hat{\theta} = \frac{\bar{y}}{\bar{y} - 1}$$

5. RELIABILITY ANALYSIS

This section deals with Reliability and hazard rate function of the proposed model.

5.1. Reliability Function

The likelihood that a system will continue to work after a certain amount of time is the reliability function. For the proposed model, it is given as:

$$R(y; \theta) = P(Y > y) = \frac{1}{\theta^y} \quad ; y = 1, 2, 3, \dots$$

5.2. Hazard Rate

The hazard rate for ZTNDD is given by:

$$h(y; \theta) = \frac{P(y; \theta)}{R(y; \theta)}$$

$$\Rightarrow h(y; \theta) = \frac{(\theta - 1)}{\theta^y} \theta^y = (\theta - 1)$$

Table 2: Values of Reliability function $R(y;\theta)$ for the proposed model

y	$\theta = 1.5$	$\theta = 2.0$	$\theta = 2.5$	$\theta = 3.0$	$\theta = 3.5$	$\theta = 4.0$
1	0.667	0.500	0.400	0.333	0.286	0.250
2	0.444	0.250	0.160	0.111	0.082	0.0625
3	0.296	0.125	0.064	0.037	0.0233	0.016
4	0.198	0.063	0.026	0.012	0.007	0.004
5	0.132	0.031	0.010	0.004	0.002	-
6	0.088	0.016	0.004	0.001	-	-
7	0.059	0.008	0.002	-	-	-
8	0.039	0.004	-	-	-	-
9	0.026	0.002	-	-	-	-
10	0.017	-	-	-	-	-

5.3. Simulation Study

This part is based on extensive simulation studies to compare the effectiveness of the created estimator. Six parameter settings ($\theta=1.5,2.0,2.5,2.7,3.0,3.5$) are taken into account from ZTNDD with sizes $n=10, 25, 60, 350,$ and 500 . The simulation process, which is described below, is based on 1000 iterations of the suggested model.

Table 3: Simulation study of MLEs for proposed model

Sample Size(n)	$\theta = 1.5$				$\theta = 2.0$			
	Bias	Variance	MSE	Coverage Probability (95%)	Bias	Variance	MSE	Coverage Probability (95%)
10	0.11539	0.07045	0.08376	1.00	0.57527	2.55556	2.88649	0.94
25	0.02430	0.01274	0.01332	0.98	0.10359	0.06458	0.07531	1.00
60	0.01663	0.00710	0.00738	0.96	0.00713	0.02679	0.02684	0.96
350	-0.00388	0.00099	0.00101	0.96	0.01830	0.00680	0.00713	0.96
500	0.00652	0.00083	0.00087	0.96	0.00118	0.00473	0.00473	0.90
Sample Size(n)	$\theta = 2.5$				$\theta = 2.7$			
	Bias	Variance	MSE	Coverage Probability (95%)	Bias	Variance	MSE	Coverage Probability (95%)
10	0.45452	0.96431	1.17090	0.94	0.74182	4.06066	4.61096	0.94
25	0.13630	0.12885	0.14743	0.96	0.05464	0.21413	0.21712	0.98
60	0.03798	0.04603	0.04748	0.98	0.02543	0.07962	0.08027	0.92
350	0.03642	0.01411	0.01543	0.92	-0.02979	0.02648	0.02737	0.90
500	0.02586	0.00647	0.00714	0.88	-0.01670	0.01099	0.01127	0.94
Sample Size(n)	$\theta = 3.0$				$\theta = 3.5$			
	Bias	Variance	MSE	Coverage Probability (95%)	Bias	Variance	MSE	Coverage Probability (95%)
10	0.68743	2.96483	3.4374	0.98	0.54364	3.79498	4.09053	0.96
25	0.25330	0.51149	0.57565	0.98	0.32301	0.65368	0.75802	0.98
60	0.16092	0.17636	0.20225	0.96	0.11374	0.24541	0.25835	0.92
350	0.03626	0.03187	0.03319	0.94	0.06039	0.04573	0.04938	0.96
500	-0.00804	0.00909	0.00916	0.96	-0.01819	0.02315	0.02349	0.96

It is clear from the simulation table (3) that the Bias and MSE decreases significantly as sample size increases. Additionally, when sample size grows, coverage probability gets closer to 0.95. We

can therefore claim that MLE exhibits the consistency property. Therefore, we draw the conclusion that the MLE does a good job of forecasting the distribution’s parameter.

6. APPLICATIONS

In this part, we show that, in comparison to other competing models, our suggested model gives a better fitting when analysing two real-world datasets and one simulated data set. The first two data sets reflect European red mites, are taken from Garman [8] and Mcguire [16] respectively. These data sets are shown in Tables (4) and (7) respectively. The parameters of each of these distributions are determined using the maximum likelihood approach. The expected frequencies for fitting ZTNDD, ZTPD, ZTNBD, SBPD, SBNBD(Size Biased Negative Binomial), SBPLD(Size Biased Poisson Lindley Distribution)[1], and SBDLD(Size Biased Discrete Lindley)[24] are obtained using R studio, and the goodness of fit of the model is evaluated using Pearson’s chi-square test. For each fitted model, the expected counts, chi square, and p-value are shown in Table (5 and 8). Additionally, we take into account the AIC (Akaike information criterion) and BIC criteria in order to evaluate our proposed distribution to the other competing models. The fact is lesser the AIC and BIC values, better is the distribution.

$$AIC = 2k - 2 \log L \quad \text{and} \quad BIC = k \log n - 2 \log L$$

where n is the sample size, k is the number of parameters in the model, and $\log L$ is the log-likelihood function’s maximum value for the model under consideration. Here Table (6 and 12) show that, in comparison to other competing models, the ZTND distribution has lower AIC and BIC values.

6.1. Data Set I

This data set is related to European red mites taken by Garman [8]. This data set was recently used by Rama & Simon in 2018 [13] in modelling zero truncaed discrete distribution.

Table 4: *European Red Mites[8]*

Count	1	2	3	4	5	6	7
Frequency	38	17	10	9	3	2	1

Table 5: *Expected frequencies and p-values of ZTND model and other competing models for data set I*

Count	Frequency	ZTNDD	ZTPD	ZTNBD	SBPD	SBNBD	SBPLD	SBDLD
1	38.00	37.00	29.00	36.00	25.00	37.00	32.00	30.00
2	17.00	20.00	26.00	21.00	29.00	20.00	24.00	25.00
3	10.00	11.00	15.00	11.00	17.00	11.00	13.00	14.00
4	9.00	6.00	7.00	6.00	6.00	6.00	6.00	6.00
5	3.00	3.00	2.00	3.00	2.00	3.00	3.00	3.00
6	2.00	2.00	1.00	2.00	0.00	2.00	1.00	1.00
7	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
Degrees of Freedom		3.00	2.00	2.00	2.00	2.00	2.00	2.00
χ^2		2.07	10.08	2.46	20.73	2.07	6.36	8.34
p-value		0.56	0.01	0.29	<0.01	0.36	0.04	0.02

Table 6: Different criterions for ZTND model and other competing models for data set I

Criteria	ZTNDD	ZTPD	ZTNBD	SBPD	SBNBD	SBPLD	SBDLD
$-l$	118.80	122.79	119.73	127.89	119.78	120.07	121.08
AIC	239.60	247.59	241.46	257.78	241.56	242.13	244.17
BIC	241.99	249.97	246.22	260.16	246.33	244.51	246.55

z

6.2. Data Set II

This data set is also related to European red mites taken from Mcguire [16]. This data set was recently used by Hassan & Mir[11] in analysing the Bayesian estimation of size biased generalized geometric series distribution.

Table 7: European red mites [11]

Count	1	2	3	4	5
Frequency	128	37	18	3	1

Table 8: Expected frequencies and p-values of ZTND model and other competing models for Data set II

Count	Frequency	ZTNDD	ZTPD	ZTNBD	SBPD	SBNBD	SBPLD	SBDLD
1	128.00	128.00	121.00	126.00	118.00	127.00	123.00	122.00
2	37.00	40.00	49.00	43.00	54.00	42.00	46.00	49.00
3	18.00	13.00	13.00	13.00	12.00	13.00	13.00	13.00
4	3.00	4.00	3.00	4.00	2.00	4.00	3.00	3.00
5	1.00	1.00	0.00	1.00	0.00	1.00	1.00	1.00
Degrees of Freedom		2.00	1.00	1.00	1.00	1.00	1.00	1.00
χ^2		2.35	5.59	2.99	10.77	2.73	3.43	4.70
p-value		0.31	0.02	0.08	0.00	0.10	0.06	0.03

Table 9: Different criterions for ZTND model and other competing models for data set II

Criteria	ZTNDD	ZTPD	ZTNBD	SBPD	SBNBD	SBPLD	SBDLD
$-l$	170.09	171.16	171.96	173.83	170.92	170.48	171.04
AIC	342.19	344.32	343.92	349.66	344.05	342.95	344.09
BIC	345.42	347.55	350.38	352.89	350.51	346.19	347.32

6.3. Simulated Data Set

For further testing our claim that our model is suitable for fitting purposes, we ran a simulation experiment. The simulated data is present in Table(10).

Table 10: Simulated Data Set

Count	1	2	3	4	5	6	7	8	9	10
Frequency	523	250	104	61	31	15	8	6	1	1

Table 11: Expected frequencies and p-values of ZTND model and other competing models for Simulated data

Count	Frequency	ZTNDD	ZTPD	ZTNBD	SBPD	SBNBD	SBPLD	SBDLD
1	523.00	514.00	425.00	510.00	388.00	514.00	457.00	439.00
2	250.00	250.00	323.00	253.00	367.00	250.00	300.00	316.00
3	104.00	122.00	164.00	123.00	174.00	122.00	145.00	152.00
4	61.00	59.00	62.00	60.00	55.00	59.00	61.00	61.00
5	31.00	29.00	19.00	29.00	13.00	29.00	24.00	22.00
6	15.00	14.00	5.00	14.00	2.00	14.00	9.00	7.00
7	8.00	7.00	1.00	7.00	0.00	7.00	3.00	2.00
8	6.00	3.00	0.00	3.00	0.00	3.00	1.00	1.00
9	1.00	2.00	0.00	1.00	0.00	2.00	0.00	0.00
10	1.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00
Degrees of Freedom		6	4	4	3	5	4	4
χ^2		3.90	172.81	4.86	260.35	3.90	56.42	92.80
p - value		0.69	<0.01	0.30	<0.01	0.56	<0.01	<0.01

Table 12: Loglikelihood, AIC and BIC values of proposed model and other competing models for Simulated

Criteria	ZTNDD	ZTPD	ZTNBD	SBPD	SBNBD	SBPLD	SBDLD
-l	1348.836	1418.286	1349.320	1477.478	1348.845	1373.345	1387.746
AIC	2699.672	2838.572	2702.641	2956.956	2701.690	2748.690	2777.492
BIC	2704.580	2843.480	2712.456	2961.864	2711.506	2753.598	2782.399

7. DISCUSSION AND CONCLUSION

From Table(1), it is evident that our model is positively skewed and leptokurtic in nature. Also, from the index of dispersion plot(3), it is obvious that our model can accommodate equi-dispersed, under-dispersed as well as over-dispersed data. From Table(3), it can be observed that the Bias and MSE decreases with increase in sample size and also as sample size increases the coverage probability approaches to 0.95.

From table(5,8, and 11), it can be observed that our model has highest p-value among all the competing models in all the three data sets. Also, the information criterion is used to validate our model, and it can be seen from table(6,9, and 12) that our model has lowest values of AIC, BIC criterions which agrees with the fact that lesser the values of AIC and BIC, better is the model. In this paper we have introduced a zero truncated model namely zero truncated new discrete distribution(ZTNDD). We have derived the structural properties and generating functions of the truncated model including Moments about origin, Factorial moments, Moment generating function. Also, MLE is calculated, and for checking the behaviour of MLE, we have carried out a simulation study. We have fitted two real life data sets and one simulated data set, we compared our model with several other discrete models and it has been found that our model performs better than all models.

Author contributions All the authors have contributed equally.

Data Availability The data are given in the manuscript.

Declarations

Ethical statements Three datasets are used in the application section and two of them are taken from the literature.

Conflict of interest The authors have no conflict of interest.

Funding The author received no specific funding for this study.

REFERENCES

- [1] Adhikari, T. R., & Srivastava, R. S. (2014). Poisson-size-biased Lindley distribution. *International Journal of Scientific and Research Publications*, 4(1), 1-6.
- [2] Brass, W.(1959). Simplified methods of fitting the truncated negative binomial distribution. *Biometrika* 45,59-68.
- [3] Best, D. J., Rayner, J. C. W., & Thas, O. (2007). Goodness of fit for the zero-truncated Poisson distribution. *Journal of Statistical Computation and Simulation*, 77(7), 585-591.
- [4] Coleman JS and James J. The equilibrium size distribution of freely forming groups, *Sociometry* 1961; 24: 36-45.
- [5] Creel, M. D. and Loomis, J. B. (1990). Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California. *American Journal of Agricultural Economics*, 72, 434-441.
- [6] Davis, D. J. F. (1952). An analysis of some failure data. *Journal of the American Statistical Association*, 47(258), 113-150.
- [7] Finney DJ and Varley GC. An example of the truncated Poisson distribution, *Biometrics* 1955; 11: 387-94.
- [8] Garman, P. (1923). The European red mites in Connecticut apple orchards. *Connecticut Agri. Exper. Station Bull.*, 252,103-125
- [9] Ghitany, M. E., Al-Mutairi, D. K., & Nadarajah, S. (2008). Zero-truncated Poisson-Lindley distribution and its application. *Mathematics and Computers in Simulation*, 79(3), 279-287.
- [10] Hassan, A., Ahmad, P. B., & Bhatti, M. I. (2008). On the bayes estimator of parameter and reliability function of the zero-truncated poisson distribution. *Journal of the Korean Society for Industrial and Applied Mathematics*, 12(2), 97-108.
- [11] Hassan, A., & Mir, K. A. (2008). On Bayesian estimation of size-biased generalized geometric series distribution and its applications. *Journal of Statistical Research*, 42(1), 117-126.
- [12] Jain, S., Siddiqui, S. A., Dwivedi, S., Siddiqui, I., & Kamal, M. (2021). A NEW DISCRETE DISTRIBUTION WITH ITS MATHEMATICAL PROPERTIES. *Int. J. Agricult. Stat. Sci.* Vol, 17(2), 693-697.
- [13] Kiani, T. H. (2020). A zero truncated discrete distribution: Theory and applications to count data. *Pakistan Journal of Statistics and Operation Research*, 167-190.
- [14] Lindsay, J.K.(1995). *Modeling Frequency and Count Data*, Clarendon Press, Oxford University New York.
- [15] Lee, A. H., Wang, K., Yau, K. K. W and Somerford, P.G. (2003). Truncated Negative Binomial Mixed Regression Modelling of Ischaemic Stroke Hospitalizations. *Statistics in Medicine*, 22, 1129-1139.
- [16] Mcguire, J.U., Brindley, T.A.,& Bancroft, T.A. (1957). The Distribution of European Corn Borer Larvae *Pyrausta nubilalis* (Hbn.), in Field Corn. *Biometrics*, 13, 65.
- [17] Mathews JNS and Appleton DR. An application of the truncated Poisson distribution to immunogold assay, *Biometrics* 1993; 49: 617-621.
- [18] Phang, A. L. and Loh, E.F. (2013).Zero Truncated Strict Arcsine Model. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 7(7), 989-991.
- [19] Rama, S., & Simon, S. (2018). A zero-truncated discrete Lindley distribution with applications. *Int. J. of Statistics in Medical and Biological Research*, 2(1), 8-15.

- [20] Siddiqui, S. A., Jain, S., Siddiqui, I., Khan, K., & Alam, M. (2016). Characterization and development of a new failure model. *Journal of Theoretical and Applied Information Technology*, 86(1), 87.
- [21] Shanker, R., & Shukla, K. K. (2017). A zero-truncated two-parameter Poisson-Lindley distribution with an application to Biological Science. *Turkiye Klinikleri Biyoistatistik*, 9(2), 85-95.
- [22] Shanker, R. (2017). Zero-truncated Poisson–Akash distribution and its applications. *American Journal of Mathematics and Statistics*. 2017c, 7(6), 227-236.
- [23] Shanker, R. (2017): The discrete Poisson-Akash distribution, *International Journal of probability and Statistics*, 6(1), 1-10
- [24] Sium, S., & Shanker, R. (2018). Size-biased discrete-lindley distribution and its applications to model distribution of freely-forming small group size. *Biom Biostat Int J*, 7(2), 131-136.
- [25] Shukla, K.K. and Shanker,R. (2019) : A discrete Poisson-Ishita distribution and its applications, *International Journal of Statistics and Economics* 20(2),109-122.
- [26] Sium, S., & Shanker, R. (2020). A zero-truncated discrete Akash distribution with properties and applications. *Hungarian Statistical Review*, 3(2), 12-25.
- [27] Shukla, K. K., Shanker, R., & Tiwari, M. K. (2020). Zero-truncated poisson-ishita distribution and its applications. *Journal of Scientific Research*, 64(2).

Use of tribodiagnostics in practice

Pavol Mikuš¹ Alena Breznická²



Faculty of Special Technology, Alexander Dubček University of Trenčín,
Ku kyselke 469, 911 06, Trenčín, Slovakia^{1,2}
pavol.mikus@tnuni.sk
alena.breznicka@tnuni.sk

Abstract

Tribodiagnostics deals with the problems of lubrication, friction and analysis of oils in technical fluids. Based on the results of parameter monitoring and chemical analysis of the oil, it is possible to determine the impending failure of the entire system very accurately. Today, this relatively young field of technical diagnostics is gradually becoming very viable and its results are fully in line with classical vibroacoustic diagnostics or thermodynamics. It is used in all mechanical systems containing oil systems. This is one of the methods of non-disassembly technical diagnostics, which is based on the knowledge that the lubricant after a certain period of use in the lubrication system reflects the condition of the equipment and the conditions in which this equipment was operated.

Keywords: Tribodiagnostics, lubrication fluid, oil, friction

1. Introduction

The growing demand for vehicles forces us to think about ensuring a high level of operational reliability, which should be close to the inherent reliability, which is ensured by optimal use, maintenance, repairs, etc. For the maintenance to be technically and economically optimal, it is also necessary to optimize the technical diagnostics, resp. also a significant part of it - tribodiagnostics. The term tribology (from the Greek TRIBOS - friction and logos - science) is historically very old and has probably existed since the beginnings of written history. Examples of the development of wheels, bearings, friction surfaces, etc. are documented. Already in the early civilizations (Archimedes) and also the targeted scientific development of tribology has a relatively long history (15th century), when the foundations of the law come from Leonardo da Vinci. Important scientists who dealt with tribology were e.g. Lavoisier, Leibnitz, Tower, Reynolds, Stribeck and others [2].

2. Tribotechnical systems

From a narrower point of view, tribology is a science and practice that deals with the behavior of contacting surfaces in motion, or in an attempt to move relative to each other (sliding, rolling, rotating, impact, oscillating, flow of gases, liquids, etc.). When the surfaces interact with each other, there is resistance to movement and friction. At the same time, the engineering observation of friction has a predominantly phenomenological character, since it uses in particular its external

manifestations, effects in the field of contact and effects on the environment. Generally speaking, there are generally two basic areas of research and application of tribology:

- the field of man-made artificial technical tribological systems,
- the area of natural tribological systems (e.g. human locomotor system - joints, plant roots, etc.),

An approximation of the sizes and dimensions in which tri-diagnostics is performed in the area of vehicle groups is shown in Fig. 1 and Fig. 2.

Both artificial and natural tribological systems include at least two, but usually four, system elements, namely a basic friction body, a counter-acting friction body, an Interfacial Medium and an Ambient Medium.

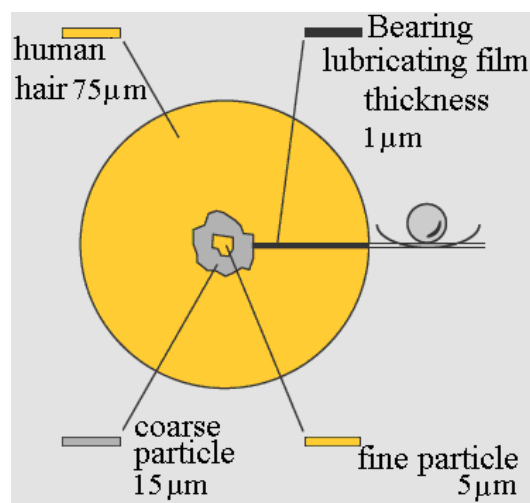


Figure 1: An example of the size of the lubricating film and wear particles in the field of tribodiagnosics

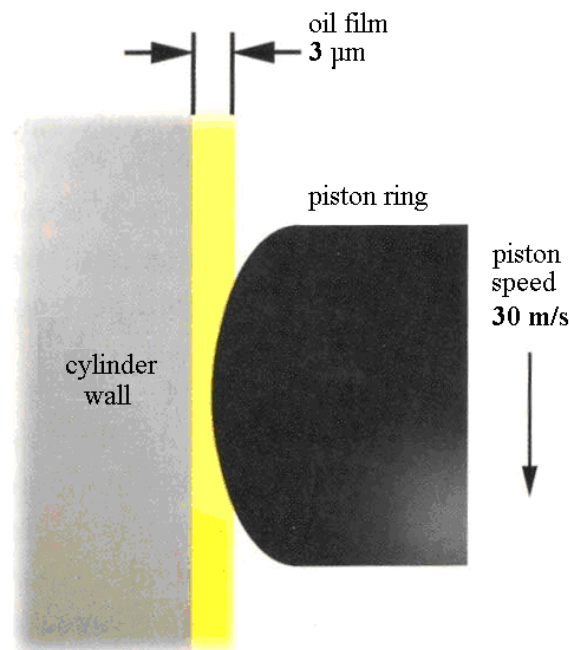


Figure 2: Example of an oil film size for an internal combustion engine

The Tribotechnical System (TTS) generally includes relationships between the following variables:

- Input Variables:
 - desirable,
 - undesirable (disturbing).
- Output Variables:
 - useful variables,
 - loss variables.

Undesirable - interfering input variables negatively affect the values of usable and lossy output variables. The role of TTS in practice is the conversion of input variables such as input torque, input speed, type of input movement, resp. in the case of several different movements, their sequence also includes, for technically usable output variables e.g. output torque, output speed and output movement. While the two contacting surfaces are part of each TTS, the interfacial medium and the surrounding medium may be absent from the system if it is a process taking place e.g. in a vacuum. Open systems are those where the base body is in contact in time with a different type of contact body or with several different bodies, e.g. when transporting materials or machining. Closed TTSs are those where the contacting bodies meet repeatedly. In addition to the other features of the tribological process mentioned above, when comparing open and closed systems, the ability of the system to function properly depends on:

- in the case of open TTS only for wear of the base body,
- in the case of closed TTS for wear of the base and opposing body [4].

3. Tribological load and interaction

The tribological load in tribotechnical systems is caused by the already mentioned input and disturbing variables, more precisely by their influence on the structure of the tribotechnical system. Tribological loading includes contact, kinematic, dynamic and thermal processes. A tribological load is a contact load of a base surface in the solid phase by another surface, which may be in the solid, liquid or gaseous phase, with relative relative movement of the two surfaces. This is done with the help of real contact surfaces. Plastic deformation and wear can cause a change in the size of individual real contact surfaces during the tribological process. Mechanical energy dissipates by friction, i.e. they dissipate and are converted into thermal energy, which affects the rate of wear. The nominal or apparent surface is decisive for lubrication when the two contact surfaces are not in direct contact and there is a sufficient amount of lubricant between them. With mixed lubrication, the lubrication parameter Λ is equal:

$$\Lambda = \frac{h_{\min}}{\left(R_{q1}^2 + R_{q2}^2\right)^{1/2}} \quad (1)$$

where h_{\min} is minimum thickness of lubricating layer (μm), R_{q1} root mean square deviation of the base body surface profile (μm), R_{q2} mean square deviation of the surface profile of the opposing body (μm).

In the range $1 \leq \Lambda < 5$, in the case of limit friction (lubrication) $\Lambda < 1$ and in the case of dry friction, when both bodies are in direct contact, whether partial or complete, the boundary or real contact surfaces have a decisive influence. If there is direct contact between the friction surfaces, there will also be interactions between atoms / molecules and mechanical interactions at the locations of the

real contact surfaces and in the affected area close below the surface of both bodies. This gradually leads to elastic and finally to plastic contact deformations and the creation of real contact surfaces. The type of interaction that occurs depends mainly on the state of lubrication. If it is a sufficiently lubricated contact, then atomic / molecular interactions are insignificant compared to mechanical interactions.

4. Wear

Wear can be characterized as an increasing loss of material from the surface of the solid phase upon interaction and relative motion with the solid phase body, liquid or gas. Wear of two rigid bodies occurs in direct contact, ie. in case of insufficient thickness of lubricating film, or in case of absence of lubricant. Wear is manifested by the release of particles from the surface of the material of one or both bodies in frictional contact. Wear can be caused by several mechanisms, the following four being the most important of them:

- surface fatigue,
- abrasion,
- adhesion,
- tribochemical reaction or erosion [1].

5. Tribotechnical diagnostics

Tribotechnical diagnostics is a set of methods and means of checking the technical condition (diagnosis, localization, prognosis, or genesis) of usually complex, closed friction moving joints of mechanical systems using lubricating media (oils, greases, greases, etc.) hydraulic liquids and. i. It organically combines the measurement, evaluation and forecasting of parameters and characteristics of processes taking place in a given facility. The results of the analyzes are used to perform the following tasks:

1. Monitoring the condition, trend and mode of wear of machinery based on, e.g. determination of the content of abrasions, resp. abrasion metals in the lubricant, while the decisive factor is mainly the trend of measured values.
2. Determination of the service life of the lubricant by determining the degree of its degradation by chemical reactions, products of thermal-oxidation processes, internal contamination, external impurities, etc. Increased number of impurities, e.g. in oil it not only means greater wear of the lubricated parts, but the contained deposits can clog the lubrication holes and grooves. The service life of the lubricant is expressed by a set of relatively objectively determined indicators.
3. Determination of optimal times for changing individual lubricants. The importance of this task is currently increasing with the rising price of lubricants and cost-saving measures.

By fulfilling the above tasks, we can get an overview of the technical condition of the relevant mechanical system, the aging and deterioration of the lubricant, wear of functional parts of the machine, or. about the location of excessive wear, which is usually the cause of failures and sometimes system crashes. Analytical data on the lubricant provide, in addition to diagnostic information, also prognostic information and make it possible to predict and also prevent accident situations. Lubricant analysis makes it possible to very sensitively determine the wear rate of the system as a function of time, resp. in real time, provides additional control options, e.g. filtration systems, tightness of cooling systems, etc. In addition to the requirement of complexity,

tribotechnical diagnostics must meet the condition of correct selection of the necessary tribodiagnostic methods, their simplicity, speed and unambiguous responses to the state (mode) of system wear and further usability of the lubricant. In terms of use, depending on the complexity of the technique, the organizational level, the traffic intensity, the instrumentation and the personnel possibilities, the methods of tribotechnical diagnostics can be divided as follows:

- Simple methods and tests - express methods (speed methods).
- Standard methods and tests - according to STN EN.
- Special methods and tests - tribodiagnostic methods.

From the point of view of the essence and physico-chemical principles, the methods of tribodiagnosics can be divided into:

- Methods for determining the concentration of abrasive metals.
- Methods for evaluating the morphology and distribution of abrasive particles.
- Methods for determining the physico - chemical properties of a lubricant [5,6].

5. Ferrography

The detection of wear of oil-lubricated mechanical systems is based on the knowledge that the oil after a certain period of operation reflects the technical condition of the mechanical system and the operating conditions. This multidimensional information is carried by metal abrasion, which is dispersed in the oil and which, after quantification by a suitable method, allows indirect monitoring of the wear regime and mechanical changes in the system in which the oil is used. From the detected amount of metal abrasion, the intensity of the increase in the number of particles, the shape, morphology, size and material composition of particles and wear fragments, certain conclusions can be drawn - if the increase in abrasion and other parameters are systematic and compared with the nominal values determined for a given mechanical system (determined by calculation, long-term monitoring, etc.), it can be relatively reliably judged to be a normal course of wear without an increased risk of system failure. A sudden increase in the number of metal particles and the finding of particles of shapes characteristic of abnormal wear mechanisms signal an extraordinary event. From the size and shape of the particles, the growth rate, their number, morphology and other parameters, the severity of the disorder and the urgency of corrective action can be inferred. An important diagnostic circumstance is the ability to locate the site of increased abrasion and incipient disorders. According to the material composition of the metal abrasion, it is possible to determine the friction pair in which there is a sharp increase in wear. For these purposes, a suitable method is ferrography, based on the separation of solid metallic and non-metallic particles contained in the oil filling of lubrication systems of machines and equipment from the actual oil. Describes trapped particles (especially ferromagnetic) and assigns them to individual wear mechanisms; allows you to detect an impending machine failure. Abrasive particles can be divided according to their composition, size and other characteristics using this method. The separation takes place in a ferrograph, Fig. 3 - a sample of the examined lubricant flows down an inclined pad, which is placed in a magnetic field. The largest ferromagnetic particles settle at the beginning of the substrate and then the particles settle according to their magnetic properties, composition, size and shape. With this method it is possible to distinguish the shape of particles, their origin, place of origin (location of wear), morphology, etc. Ferrography is focused on the analysis of ferromagnetic abrasives in a lubricant using a magnetic field. It is a technique for separating metallic (and non-metallic) substances from used oil. In the ferrographic analysis, a diluted sample of oil is drained over an inclined transparent substrate (foil), under which a strong magnet is placed. The inclination of the substrate causes a particle size distribution along the transparent substrate due to the gradient (variable force) of the magnetic field. At the beginning, larger particles are captured (>15

micrometers) and the closer the film is to the magnet, the smaller particles are captured (<5 micrometers, or at the end up to 1-2 micrometers). After passing the oil sample, the oil is washed away with a suitable solvent (technical gasoline) and the particles are fixed on a transparent support with a transparent varnish, thus obtaining the so-called Ferrogram. Ferrogram allows to assess particle size, ratio of large particles (10-100 micrometers) to small particles, morphological (shape) characteristics of particles, etc. Based on the observation of particles on a special bichromatic microscope (combination of metallographic and biological microscopes - reflected as well as transmitted light is used), the wear regime of the mechanical system can be determined.

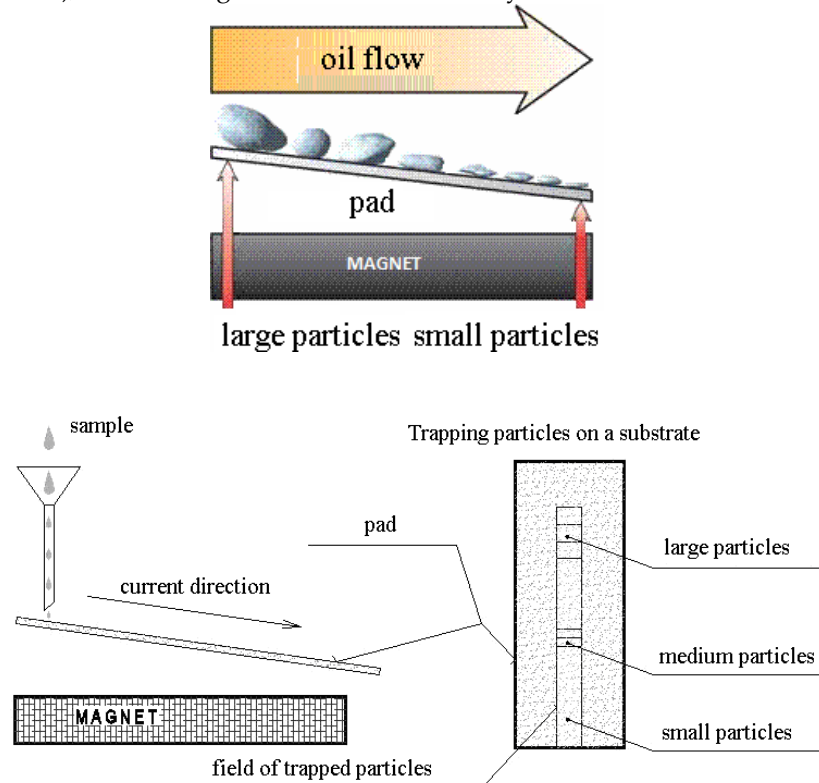


Figure 3: Schematic of a ferrograph and the principle of ferrographic analysis

The operating conditions, in particular the efficiency of the air filter, the presence of water in the oil and the overall care of the technical staff for the equipment, shall be clearly indicated at the end of the sedimentation trace on the ferrogram. Image analysis can be used for quantitative evaluation of ferrograms. In Fig. 4 and Fig. 5 are particles isolated from oil filters.

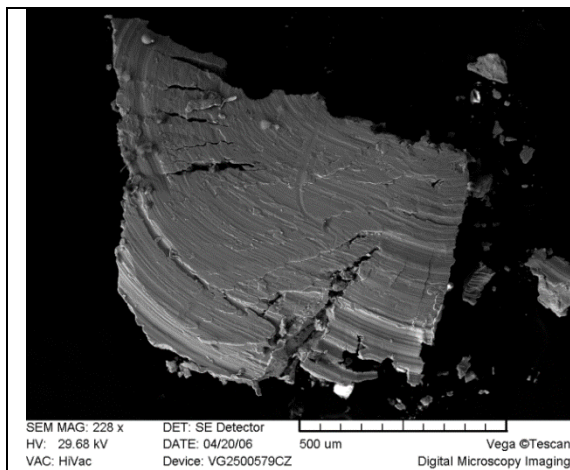


Figure 4: Particles from oil filters

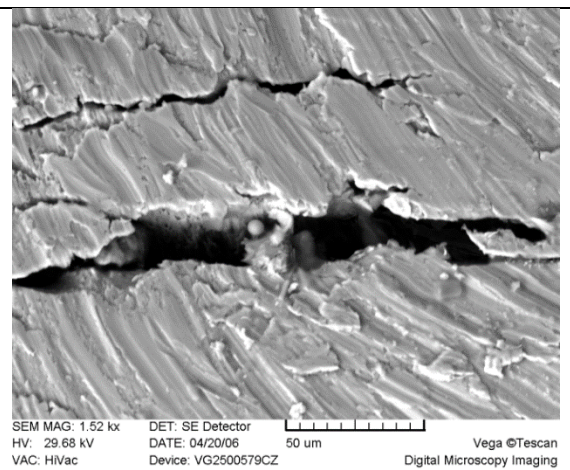


Figure 5: Particles from oil filters

Fig. 4 shows laminar particles where traces of abrasive wear due to high pressures at the contact of the friction surfaces are visible and incipient cracks are visible in the edge portions. In Fig. 5 is a spheroid artifact typical of fatigue wear. The ball is formed by the slow growth of a fatigue crack extending into the oil-soaked surface [7].

6. Results and discussion

The maintenance program is usually based on vibration monitoring, selected operating parameters and tribodiagnosics, which allows you to assess the specific condition of the equipment in real time. It is important that maintenance only applies to those parts or machines that really need it. The fault can be detected at the stage of occurrence and thus prevent more extensive damage, there are no unexpected outages and at the same time no unnecessary work is performed. Tribodiagnosics is based on regular sampling of lubricants (oils) from monitored machines and their analysis. With the help of tribodiagnostic analysis, we can determine both the condition of the oil itself and especially the condition of the monitored machine. The lubricating oil serves as a medium containing wear particle of the lubricated parts of the monitored machine. By analyzing these particles, we obtain important information about the mode of wear and events in the machine. It is very important to monitor the machine systematically and continuously from the beginning of the operational deployment and to obtain trends in the content of particles in the oil, resp. other tribodiagnostic parameters in the oil, as this information provides a reliable indication of changes in the wear regime and the actual technical condition. The second part of tribodiagnosics is the analysis of the oil itself, which we find out by changing its physico-chemical properties, as well as its pollution by foreign substances, e.g. water, mechanical pollution, chemical compounds. The basis of the success of tribodiagnosics is a correctly taken oil sample. The sample must be truly representative, ie. it must contain all substances in the proportion in which they occur in the lubrication system of the monitored machine. The optimal place for sampling is the return line, where the oil returns from the lubricated places to the oil tank. Some manufacturers already equip the machine (engine) with a sampling tap located just on the oil return line. Sampling must be performed while the machine is running, if possible, or shortly after it has stopped. Sampling containers (sample boxes) must be clean and dry. An important aspect of tribodiagnosics is the speed of response and the accuracy of the results. Regular samples should be analyzed quickly, based on the results of the diagnosis, the results are sent to the machine operator. Only the results of the analysis of the oil sample or the evaluation of the analysis with a recommendation for further action may be given in the relevant report. It depends on the system that suits the knowledge and experience of the workers.

7. Conclusion

This article briefly analyzes the crucial problems of friction and wear that occur during the operation of vehicles (especially vehicle combustion engine, transmissions, hydraulic systems, etc.), defined tribological unit as the smallest element where friction and wear take place. The tribological unit includes interactions of min. two friction surfaces, lubricant and environment. A separate part is devoted to the tribodiagnostic method of ferrography. This focuses on the detection of wear of oil-lubricated mechanical systems is based on the knowledge that the oil after a certain period of operation reflects the technical condition of the mechanical system and the operating conditions. There are still few such maintenance personnel who practically use and apply the current state of the art to identify the condition of the equipment based on the condition of the oil system. The introduction of new approaches to the care of means of production, technology, and processes enables the rational use of the results of analyzes for reliable and trouble-free operation. Proper treatment e.g. by filtering or adding the right additives will significantly affect the economy of

operation. If the oil is still clean and properly maintained, it does not need to be changed. This will significantly reduce the environmental risk as well as increased environmental protection.

References

- [1] Deters, L. *Springer Handbook of Mechanical Engineering*, Part B. [s.l.]: [s.n.], Tribology, 2009. ISBN 978-80-540-491.
- [2] Stodola, J.: *Diagnostika bojových a speciálních vozidel*. Vysokoškolská učebnice U- 3086: Univerzita obrany Brno, 2005. ISBN 80-7231-017-8.
- [3] Macian, V., Tormos, B., Olmeda, P., Montoro, L.: *Analytical Approach to Wear Rate Determination for Internal Combustion Engine Condition Monitoring Based on Oil Analysis*. Tribology International 36 (2003), p. 771-776.
- [4] Krtička, F.: *Obrazová analýza částic v provozních hmotách a konstrukčních materiálech pro dopravní prostředky*. Dopravní fakulta Jana Pernera. Univerzita Pardubice, 2007.
- [5] Leugner, L.: *The Practical Handbook of Machinery Lubrication*. Maintenance Technology International, 2005.
- [6] Michael A. Taylor.: *Quantitative Measures for Shape and Size of Particles*, Powder Technology, 124 (2001), p. 94–100.
- [7] Totten, E., G.: *Handbook of Lubrication and Tribology*. Volume I. Taylor&Francis Group, New York, 2006, ISBN 0-8493-2095-X.
- [8] Leitner, B. (2020). The procedure of operational risks management in railway companies, In. Transport Means - Proceedings of the International Conference, Pages 57 - 622020 24th International Scientific Conference on Transport Means, Kaunas, ISSN 1822296X.

MOMENTS PROPERTIES OF CONCOMITANTS OF GENERALIZED ORDER STATISTICS FROM FGMTBM EXPONENTIAL DISTRIBUTION

Mustafa Kamal¹, Nayabuddin², Intekhab Alam³, Ahmadur
Rahman⁴, Abdul Salam⁵, Shazia Zarrin⁶

¹Department of Basic Sciences, College of Science and Theoretical Studies, Saudi
Electronic University, KSA

e-mail: m.kamal@seu.edu.sa

²Faculty of Public Health and Tropical Medicine, Jazan University, KSA

e-mail: nhanif@jazanu.edu.sa

³Department of Management St. Andrews Institute of Technology &
Management Gurugram, Haryana, India

⁴Department of Statistics and O.R., Aligarh Muslim University, India

e-mail: ahmadur.st@gmail.com (Corresponding Author)

⁵Department of Epidemiology, Faculty of Public Health and Tropical Medicine,
Jazan University, KSA

⁶Uttaranchal Unani Medical College and Hospital, Haridwar, India.

e-mail: shaziazzarrin@gmail.com

Abstract

Concurrent or induced order statistics are produced when individuals in a random sample are ordered in compliance with the corresponding values of some other random sample. Concomitants are most helpful when $k(n)$ individuals are to be chosen using a selection technique based on their X -values. The relevant Y -values are then used to reflect how well a characteristic has performed. In this paper, concomitants of generalized order statistics (GOS) from Farlie Gumbel Morgenstern type bivariate moment (FGMTBM) exponential distribution are obtained. Additionally, distribution function (df) and probability density function (pdf) r -th generalized order statistics and a joint pdf of r -th and s -th GOS were also obtained. Furthermore, we provide the minimum variance linear unbiased estimator (MVLUE) of the position and scale parameters of the concomitants of the k -th upper record values and order statistics for the distribution under consideration. Finally, an implementation of the suggested methodology has been taken into account.

Keywords: Generalized order statistics, Concomitants, Record Values, FGMTBM exponential distribution, Single and Product Moments, Minimum variance linear unbiased estimator.

AMS Subject Classification: 62G30, 62E10

1. Introduction

In a wide range of statistical applications, order statistics (OS) and record values are often used in statistical modeling and inference. In both models, random variables are listed in descending order of magnitude. GOS offers an integrated approach to a wide variety of ordered random variable models with various interpretations. The theory of GOS is pioneered by [1]. Since then, a number of

writers have incorporated the idea of GOS into their works, including [2-3], and others.

Let $n \in N, n \geq 2, k > 0, \tilde{m} = (m_1, m_2, \dots, m_{n-1}) \in \mathfrak{R}^{n-1}, M_r = \sum_{j=r}^{n-1} m_j$, such that $\gamma_r = k + n - r + M_r > 0$ for all $r \in \{1, 2, \dots, n-1\}$. Then $X(r, n, \tilde{m}, k), r \in \{1, 2, \dots, n\}$ are called *gos* if their joint pdf is given by

$$k \left(\prod_{j=1}^{n-1} \gamma_j \right) \left(\prod_{i=1}^{n-1} [1 - F(x_i)]^{m_i} f(x_i) \right) [1 - F(x_n)]^{k-1} f(x_n) \quad (1.1)$$

on the cone $F^{-1}(0) < x_1 \leq x_2 \leq \dots \leq x_n < F^{-1}(1)$ of \mathfrak{R}^n .

Now, by selecting the proper values for the parameters, models such as ordinary OS ($\gamma_i = n - i + 1; i = 1, 2, \dots, n$ i.e. $m_1 = m_2 = \dots = m_{n-1} = 0, k = 1$), k^{th} record values ($\gamma_i = k$ i.e. $m_1 = m_2 \dots = m_{n-1} = -1, k \in N$), sequential OS ($\gamma_i = (n - i + 1)\alpha_i$;) ($\alpha_1, \alpha_2, \dots, \alpha_n > 0$), OS with non-integral sample size ($\gamma_i = (\alpha - i + 1); \alpha > 0$), Pfeifer's record values ($\gamma_i = \beta_i; \beta_1, \beta_2, \dots, \beta_n > 0$) and progressive type-II censored OS ($m_i \in N_0, k \in N$) are obtained.

The pdf of r -th GOS, $X(r, n, m, k)$ is

$$f_{X(r,n,m,k)} = \frac{C_{r-1}}{(r-1)!} [\bar{F}(x)]^{\gamma_{r-1}} f(x) [g_m(F(x))]^{r-1} \quad (1.2)$$

and joint pdf of $X(s, n, m, k)$ and $X(r, n, m, k)$, $1 \leq r < s \leq n$, is

$$f_{X(r,s,n,m,k)}(x, y) = \frac{C_{s-1}}{(r-1)!(s-r-1)!} [\bar{F}(x)]^m f(x) [g_m(F(x))]^{r-1} \\ \times [h_m(F(y)) - h_m(F(x))]^{s-r-1} [\bar{F}(y)]^{\gamma_{s-1}} f(y), \alpha \leq x < y \leq \beta \quad (1.3)$$

where,

$$C_{r-1} = \prod_{i=1}^r \gamma_i, \gamma_i = k + (n - i)(m + 1), g_m(x) = h_m(x) - h_m(0), x \in (0, 1)$$

$$\text{and } h_m(x) = \begin{cases} -\frac{1}{m+1} (1-x)^{m+1} & , m \neq -1 \\ -\log(1-x) & , m = -1 \end{cases}$$

Let $F(x, y)$ be the df of some arbitrary bivariate population. Assume also that $(X_i, Y_i), i = 1, 2, \dots, n$, are n pairs of independent random variables from the population having distribution $F(x, y)$. Let the ascending order of X variates is $X(1, n, m, k) \leq X(2, n, m, k) \leq \dots \leq X(n, n, m, k)$, now, if we arrange the Y variates pairwise (necessarily not in increasing order) with these generalized ordered statistics, then Y variates are called the concomitants of GOS and are generally denoted by $Y_{[1,n,m,k]}, Y_{[2,n,m,k]}, \dots, Y_{[n,n,m,k]}$. Now, the pdf of $Y_{[r,n,m,k]}$, the r^{th} concomitant of GOS is given as

$$g_{[r,n,m,k]}(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_{X(r,n,m,k)}(x) dx \quad (1.4)$$

and the joint pdf of $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$ is

$$g_{[r,s,n,m,k]}(y_1, y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} f_{Y|X}(y_1|x_1) f_{Y|X}(y_2|x_2) f_{X(r,s,n,m,k)}(x_1, x_2) dx_1 dx_2 \quad (1.5)$$

The FGM family of bivariate distributions is frequently employed in practice. This family is defined by the given marginal dfs $F_X(x)$ and $F_Y(y)$ of random variables X and Y and a parameter α . As a result, the FGM bivariate distribution function is given as follows:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)[1 + \alpha(1 - F_X(x))(1 - F_Y(y))], \quad (1.6)$$

together with associated pdf

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)[1 + \alpha(1 - 2F_X(x))(1 - 2F_Y(y))]. \quad (1.7)$$

The marginals of $f_{X,Y}(x, y)$ in this case are $f_X(x)$ and $f_Y(y)$. The two random variables X and Y are independent when the parameter α , also known as the association parameter, is zero. For exponential marginals, such a model was first presented by [5] and examined by [6]. The general form in Eq. (1.6) is credited to [7] and [8]. The acceptable range of the association parameter α is $-1 \leq \alpha \leq 1$, and the maximum value of the Pearson correlation coefficient ρ between X and Y must not exceed the value $1/3$. Now, for given X , the conditional df and pdf of Y are:

$$F_{Y|X}(y|x) = F_Y(y)[1 + \alpha(1 - F_X(x))(1 - F_Y(y))] \quad (1.8)$$

and

$$f_{Y|X}(y|x) = f_Y(y)[1 + \alpha(1 - 2F_X(x))(1 - 2F_Y(y))] \quad (1.9)$$

There have been various studies that discuss the concomitants of OS in the research. Weighted inverse Gaussian distribution used to conduct a comparative study on the concomitant of OS and record values by [9]. Shannon's entropy is calculated by [10] after taking into account the concomitants of distribution of FGM family. Readers might check [11, 12] for some outstanding reviews on the concomitants of GOS. In [13], parameter estimation is discussed. They estimated the parameters of the FGM-type bivariate exponential distribution using the concomitants of GOS. The concomitants of GOS for the bivariate Lomax distribution is investigated by [14]. In addition, [15] obtained the concomitants of GOS for bivariate Pareto distribution. Also, [16] explored the concomitants of m-GOS from the generalized FGM distribution family. For a few examples of recent studies based on the notion of concomitants of OS, readers may refer to [15, 17]. For some applications of OS in reliability and accelerated life testing, readers may refer to [18-22].

In this study, we examined the FGMTBM exponential distribution. For $\theta > 0, 0 < x, y < \infty, -1 \leq \alpha \leq 1$, the pdf, df and the conditional pdf of Y given X of the FGMTBM exponential distribution is described by the following equation [23].

$$f(x, y) = \theta^4 xy e^{-\theta(x+y)} \{1 + \alpha[1 - 2\{1 - (1 + \theta x)e^{-\theta x}\}] [1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}]\} \quad (1.10)$$

$$F(x, y) = \{1 - (1 + \theta x)e^{-\theta x}\} \{1 - (1 + \theta y)e^{-\theta y}\} [1 + \alpha(1 + \theta x)(1 + \theta y)e^{-\theta(x+y)}] \quad (1.11)$$

$$f(y|x) = \theta^2 y e^{-\theta y} \{1 + \alpha[1 - 2\{1 - (1 + \theta x)e^{-\theta x}\}] [1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}]\} \quad (1.12)$$

The marginal pdf and df of X are respectively

$$f(x) = \theta^2 x e^{-\theta x}, \quad 0 < x < \infty, \quad \theta > 0, \quad (1.13)$$

$$F(x) = 1 - (1 + \theta x)e^{-\theta x}, \quad 0 < x < \infty, \quad \theta > 0, \quad (1.14)$$

2. PDF of Concomitants

For the FGMTBM exponential distribution as given in Eq. (1.10), using Eq. (1.12) and Eq. (1.2) in Eq. (1.4), the pdf of r -th concomitants $Y_{[r,n,m,k]}$ of GOS for $m \neq -1$ is given as:

$$g_{[r,n,m,k]}(y) = \frac{c_{r-1}}{(r-1)!(m+1)^{r-1}} \theta^2 y e^{-\theta y}$$

$$\begin{aligned} & \times \int_0^\infty \{1 + \alpha[1 - 2\{1 - (1 + \theta x)e^{-\theta x}\}][1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}] \\ & \times [(1 + \theta x)e^{-\theta x}]^{\gamma_{r-1}-1} [1 - \{(1 + \theta x)e^{-\theta x}\}^{m+1}]^{r-1} \theta^2 x e^{-\theta x} dx. \end{aligned} \quad (2.1)$$

Setting $z = (1 + \theta x)e^{-\theta x}$ in Eq. (2.1), we get

$$= \frac{C_{r-1}}{(r-1)!(m+1)^{r-1}} \theta^2 y e^{-\theta y} \int_0^1 \{1 + \alpha(1 - 2z)[1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}]\} z^{\gamma_{r-1}-1} [1 - z^{m+1}]^{r-1} dz. \quad (2.2)$$

Making transformation $t = 1 - z^{m+1}$, we get

$$= \frac{C_{r-1}}{(r-1)!(m+1)^{r-1}} \theta^2 y e^{-\theta y} \int_0^1 t^{r-1} (1 - t)^{\frac{\gamma_r}{m+1}-1} \{1 + \alpha[1 - 2(1 - t)^{\frac{1}{m+1}}][1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}]\} dt \quad (2.3)$$

$$= \frac{C_{r-1}}{(r-1)!(m+1)^{r-1}} \theta^2 y e^{-\theta y} \{B(r, \frac{\gamma_r}{m+1}) + \alpha[B(r, \frac{\gamma_r}{m+1}) - 2B(r, \frac{\gamma_r+1}{m+1})][1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}]\}. \quad (2.4)$$

Which after simplification yields

$$g_{[r,n,m,k]}(y) = \theta^2 y e^{-\theta y} [1 + \alpha\{1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}\}]\{1 - 2\prod_{i=1}^r (1 + \frac{1}{\gamma_i})^{-1}\}. \quad (2.5)$$

It may be verified that $\int_0^\infty g_{[r,n,m,k]}(y) dy = 1$

Remark 2.1: Set $m = 0, k = 1$ in Eq. (2.5), to get the *pdf* of $r - th$ concomitants of OS from FGMTBM exponential distribution as

$$g_{[r:n]}(y) = \theta^2 y e^{-\theta y} [1 + \alpha\{1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}\}]\{1 - 2(\frac{n-r+1}{n+1})\}.$$

Remark 2.2: At $m = -1$ in Eq. (2.5), to get the *pdf* of $r - th$ concomitants of $k - th$ upper record values from FGMTBM exponential distribution as

$$g_{[r,n,-1,k]}(y) = \theta^2 y e^{-\theta y} [1 + \alpha\{1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}\}]\{1 - 2(\frac{k}{k+1})^r\}.$$

3. Single Moments

Now, by using the results from the previous section, we obtain the moments of $Y_{[r,n,m,k]}$ for the FGMTBM exponential distribution in this section. As a result, in the light of Eq. (2.5), the moments of $Y_{[r,n,m,k]}$ are given as:

$$E[Y_{[r,n,m,k]}^{k_1}] = \theta^2 \int_0^\infty y^{k_1+1} e^{-\theta y} [1 + \alpha\{1 - 2\{1 - (1 + \theta y)e^{-\theta y}\}\}]\{1 - 2\prod_{i=1}^r (1 + \frac{1}{\gamma_i})^{-1}\} dy. \quad (3.1)$$

Note that

$$\int_0^\infty x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}. \quad (3.2)$$

Using Eq. (3.2) in Eq. (3.1), we get after simplification

$$E[Y_{[r,n,m,k]}^{k_1}] = \frac{\Gamma_{k_1+2}}{\theta^{k_1}} [1 + \alpha \{ \frac{k_1+4}{2^{k_1+2}} - 1 \} \{ 1 - 2 \prod_{i=1}^r (1 + \frac{1}{y_i})^{-1} \}]. \quad (3.3)$$

Remark 3.1: Insert $m = 0, k = 1$ in Eq. (3.3) to retrieve the moments of OS concomitants from the FGMTBM exponential distribution as:

$$E[Y_{[r:n]}^{k_1}] = \frac{\Gamma_{k_1+2}}{\theta^{k_1}} [1 + \alpha \{ \frac{k_1+4}{2^{k_1+2}} - 1 \} \{ 1 - \frac{2(n-r+1)}{n+1} \}].$$

Remark 3.2: The moments of concomitants of the k-th upper record statistics from the FGMTBM exponential distribution can be obtained as follows by setting $m = -1$ in Eq. (3.3):

$$E[Y_{[r,n,-1,k]}^{k_1}] = \frac{\Gamma_{k_1+2}}{\theta^{k_1}} [1 + \alpha \{ \frac{k_1+4}{2^{k_1+2}} - 1 \} \{ 1 - 2(\frac{k}{k+1})^r \}]$$

Table 3.1: Mean of the concomitant of OS for FGMTBM exponential distribution with $\theta = 0.4$

n	r	$\alpha = -0.9$	$\alpha = -0.5$	$\alpha = -0.1$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
1	1	5	5	5	5	5	5
2	1	4.437	4.687	4.937	5.062	5.312	5.562
	2	5.562	5.312	5.062	4.937	4.687	4.437
3	1	4.156	4.531	4.906	5.093	5.468	5.843
	2	5	5	5	5	5	5
	3	5.843	5.468	5.093	4.906	4.531	4.156
4	1	3.987	4.437	4.887	5.112	5.562	6.012
	2	4.662	4.812	4.962	5.037	5.187	5.337
	3	5.337	5.187	5.037	4.962	4.812	4.662
	4	6.012	5.562	5.112	4.887	4.437	3.987

Table 3.2: Mean of the concomitant of record value for FGMTBM exponential distribution with $\theta = 0.4$ and $k = 2$.

r	$\alpha = -0.9$	$\alpha = -0.5$	$\alpha = -0.1$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
1	4.437	4.687	4.937	5.062	5.312	5.562
2	5.187	5.104	5.02	4.979	4.895	4.812
3	5.687	5.381	5.076	4.923	4.618	4.312
4	6.02	5.567	5.113	4.886	4.432	3.979

4. Joint PDF of Two Concomitants

Now, using Eqs. (1.3) and (1.12) in Eq. (1.5), the joint pdf of the r-th and s-th concomitants of GOS $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$ for $m \neq -1$ for the FGMTBM exponential distribution can be derived as:

$$g_{[r,s,n,m,k]}(y_1, y_2) = \frac{c_{s-1}}{(r-1)!(s-r-1)!} \theta^4 y_1 y_2 e^{-\theta y_1} e^{-\theta y_2} \\
\times \int_0^\infty \int_0^{x_2} \{ 1 + \alpha [1 - 2 \{ 1 - (1 + \theta x_1) e^{-\theta x_1} \}] [1 - 2 \{ 1 - (1 + \theta y_1) e^{-\theta y_1} \}] \}$$

$$\begin{aligned} & \times \{1 + \alpha[1 - 2\{1 - (1 + \theta x_2)e^{-\theta x_2}\}] [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]\} \\ & \times (\theta^4 \theta e^{-\theta x_1} e^{-\theta x_2}) \left\{ \frac{1}{m+1} (1 - [(1 + \theta x_1)e^{-\theta x_1}]^{m+1}) \right\}^{r-1} \{(1 + \theta x_2)e^{-\theta x_2}\}^{y_2 s-1} \\ & \times \left\{ -\frac{1}{m+1} [(1 + \theta x_2)e^{-\theta x_2}]^{m+1} + \frac{1}{m+1} [(1 + \theta x_1)e^{-\theta x_1}]^{m+1} \right\}^{s-r-1} \\ & \times [(1 + \theta x_1)e^{-\theta x_1}]^m dx_1 dx_2 \end{aligned} \quad (4.1)$$

And utilizing the transformations $U = 1 - (1 + \theta x_1)e^{-\theta x_1}$ and $V = 1 - (1 + \theta x_2)e^{-\theta x_2}$ in Eq. (4.1), we get:

$$\begin{aligned} & = \frac{c_{s-1}}{(r-1)!(s-r-1)!(m+1)^{s-2}} \theta^4 y_1 y_2 e^{-\theta y_1} e^{-\theta y_2} \int_0^1 \int_0^{V=U} \{1 + \alpha(1 - 2U) [1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}]\} \\ & \times \{1 + \alpha(1 - 2V) [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]\} U^m [1 - U^{m+1}]^{r-1} [U^{m+1} - V^{m+1}]^{s-r-1} V^{y_2 s-1} du dv \end{aligned} \quad (4.2)$$

Subsequently, using the transformation $V = Ut$ in the preceding equation, we obtained the following results:

$$\begin{aligned} & = \frac{c_{s-1}}{(r-1)!(s-r-1)!(m+1)^{s-2}} \theta^4 y_1 y_2 e^{-\theta y_1} e^{-\theta y_2} \int_0^1 \int_0^1 \{1 + \alpha(1 - 2U) [1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}]\} \\ & \times \{1 + \alpha(1 - 2Ut) [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]\} U^{y_2 r-1} [1 - U^{m+1}]^{r-1} t^{y_2 s-1} [1 - t^{m+1}]^{s-r-1} du dt \end{aligned} \quad (4.3)$$

By setting $p = U^{m+1}$ and $q = t^{m+1}$ in Eq. (4.3), we get after simplification

$$\begin{aligned} & = \frac{c_{s-1}}{(r-1)!(s-r-1)!(m+1)^s} \theta^4 y_1 y_2 e^{-\theta y_1} e^{-\theta y_2} \int_0^1 \int_0^1 \{1 + \alpha(1 - 2p^{\frac{1}{m+1}}) [1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}]\} \\ & \times \{1 + \alpha(1 - 2p^{\frac{1}{m+1}} q^{\frac{1}{m+1}}) [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]\} p^{\frac{y_2 r}{m+1}-1} (1 - p)^{r-1} q^{\frac{y_2 s}{m+1}-1} (1 - q)^{s-r-1} dp dq \end{aligned} \quad (4.4)$$

Upon further simplification of the preceding result, we obtain

$$\begin{aligned} g_{[r,s,n,m,k]}(y_1, y_2) & = \theta^4 y_1 y_2 e^{-\theta y_1} e^{-\theta y_2} [1 + \alpha^2 [1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}] [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]] \\ & \times \{1 - 2 \prod_{i=1}^r (1 + \frac{1}{\gamma_i})^{-1} - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1} + 4 \prod_{i=1}^r (1 + \frac{2}{\gamma_i})^{-1}\} \\ & + \alpha \{ [1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}] + [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}] \} [1 - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1}] \end{aligned} \quad (4.5)$$

It may be verified that $\int_0^\infty \int_0^\infty g_{[r,s,n,m,k]}(y_1, y_2) dy_1 dy_2 = 1$.

Remark 4.1: We can obtain the joint pdf of concomitants of OS by putting $m = 0, k = 1$ in Eq. (4.5), as well as the joint pdf of concomitants of the k-th upper record values for the FGMTBM exponential distribution by setting $m = -1$.

5. Product Moments of Two Concomitants

The product moments of two concomitants $Y_{[r,n,m,k]}$ and $Y_{[s,n,m,k]}$ are given as:

$$E(Y_{[r,n,m,k]}^l Y_{[s,n,m,k]}^p) = \int_0^\infty \int_0^\infty y_1^l y_2^p g_{[r,s,n,m,k]}(y_1, y_2) dy_1 dy_2 \tag{5.1}$$

In view of Eq. (4.5) and Eq. (5.1), we have

$$\begin{aligned} E(Y_{[r,n,m,k]}^l Y_{[s,n,m,k]}^p) &= \theta^4 \int_0^\infty y_2^{p+1} e^{-\theta y_2} \left\{ \int_0^\infty y_1^{l+1} e^{-\theta y_1} [1 + \alpha^2 [1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}]] \right. \\ &\quad \times [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]\{1 - 2 \prod_{i=1}^r (1 + \frac{1}{\gamma_i})^{-1} - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1} + 4 \prod_{i=1}^r (1 + \frac{2}{\gamma_i})^{-1}\} \\ &\quad \left. + \alpha\{[1 - 2\{1 - (1 + \theta y_1)e^{-\theta y_1}\}] + [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]\{1 - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1}\}\} dy_1 \right\} dy_2 \end{aligned} \tag{5.2}$$

$$\begin{aligned} E(Y_{[r,n,m,k]}^l Y_{[s,n,m,k]}^p) &= \frac{\theta^2 \Gamma(l+2)}{\theta^{l+2}} \int_0^\infty y_2^{p+1} e^{-\theta y_2} \left\{ [1 + \alpha^2 (\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}})] [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}] \right. \\ &\quad \times [1 - 2 \prod_{i=1}^r (1 + \frac{1}{\gamma_i})^{-1} - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1} + 4 \prod_{i=1}^r (1 + \frac{2}{\gamma_i})^{-1}] + \alpha[(\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}) \\ &\quad \left. + [1 - 2\{1 - (1 + \theta y_2)e^{-\theta y_2}\}]] [1 - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1}] \right\} dy_2 \end{aligned} \tag{5.3}$$

Solving Eq. (5.2), we get after simplification

$$\begin{aligned} E(Y_{[r,n,m,k]}^l Y_{[s,n,m,k]}^p) &= \frac{\Gamma(l+2) \Gamma(p+2)}{\theta^{l+p}} \left\{ [1 + \alpha^2 (\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}) (\frac{2^{p+2} - 2^{p+3} + p + 4}{2^{p+2}})] \right. \\ &\quad \times [1 - 2 \prod_{i=1}^r (1 + \frac{1}{\gamma_i})^{-1} - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1} + 4 \prod_{i=1}^r (1 + \frac{2}{\gamma_i})^{-1}] \\ &\quad \left. + \alpha[(\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}) + (\frac{2^{p+2} - 2^{p+3} + p + 4}{2^{p+2}})] [1 - 2 \prod_{i=1}^s (1 + \frac{1}{\gamma_i})^{-1}] \right\} \end{aligned} \tag{5.4}$$

Remark 5.1: The product moments of the concomitants of OS from the FGMTBM exponential distribution can be obtained by setting $m = 0, k = 1$ in Eq. (5.4) as follows:

$$\begin{aligned} E(Y_{[r:n]}^l Y_{[s:n]}^p) &= \frac{\Gamma(l+2) \Gamma(p+2)}{\theta^{l+p}} \left\{ [1 + \alpha^2 (\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}) (\frac{2^{p+2} - 2^{p+3} + p + 4}{2^{p+2}})] \right. \\ &\quad \times [1 + 4(\frac{(n-s+1)(n-s+2)}{(n+1)(n+2)}) - 2(\frac{n-s+1}{n+1}) - 2(\frac{n-r+1}{n+1})] \\ &\quad \left. + \alpha[(\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}) + (\frac{2^{p+2} - 2^{p+3} + p + 4}{2^{p+2}})] [1 - 2(\frac{n-r+1}{n+1})] \right\} \end{aligned}$$

Remark 5.2: The product moments of the concomitants of $k - th$ upper record value from the FGMTBM exponential distribution can be obtained by setting $m = -1$ in Eq. (5.4) as follows:

$$E(Y_{[r,n,-1,k]}^l Y_{[s,n,-1,k]}^{j-p}) = \frac{\Gamma(l+2) \Gamma(p+2)}{\theta^{l+p}} \left\{ [1 + \alpha^2 (\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}) (\frac{2^{p+2} - 2^{p+3} + p + 4}{2^{p+2}})] \right\}$$

$$\times [1 + 4\left(\frac{k}{k+2}\right)^r - 2\left(\frac{k}{k+1}\right)^r - 2\left(\frac{k}{k+1}\right)^s]$$

$$+ \alpha \left[\left(\frac{2^{l+2} - 2^{l+3} + l + 4}{2^{l+2}}\right) + \left(\frac{2^{p+2} - 2^{p+3} + p + 4}{2^{p+2}}\right) \right] \left[1 - 2\left(\frac{k}{k+1}\right)^s \right]$$

Table 5.1: Covariance of the concomitant of OS for FGMTBM exponential distribution with $\theta = 0.4$

n	s	r	$\alpha = -0.9$	$\alpha = -0.5$	$\alpha = -0.1$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
4	1	1	30.981	19.839	12.493	10.243	8.589	10.731
		2	13.102	13.488	11.939	10.439	5.988	-0.397
	2	1	11.567	6.996	4.179	3.429	3.246	4.817
		2	1.297	9.011	11.46	10.71	5.261	-5.452
		3	-0.692	2.378	3.695	3.695	2.378	-0.692
	3	1	-2.682	-4.253	-4.07	-3.32	-0.503	4.067
		2	-4.431	6.41	11.056	11.056	6.41	-4.431
		3	-6.877	-0.363	3.285	4.035	3.386	-0.127
		4	-9.322	-7.136	-4.485	-2.985	0.363	4.177
	4	1	-11.768	-13.91	-12.256	-10.006	-2.66	8.481
		2						
		3						
4								

Table 5.2: Covariance of the concomitant of record value for FGMTBM exponential distribution with $\theta=0.4$, and $k = 2$.

s	r	$\alpha = -0.9$	$\alpha = -0.5$	$\alpha = -0.1$	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
1	1	7.352	2.287	0.098	0.081	2.204	7.202
2	1	21.549	7.57	0.633	-0.193	3.431	14.099
	2	5.002	1.537	0.059	0.064	1.565	5.052
3	1	31.013	11.092	0.991	-0.377	4.25	18.697
	2	14.091	4.944	0.411	-0.123	2.268	9.275
	3	6.607	2.016	0.072	0.092	2.118	6.79
4	1	37.323	13.44	1.229	-0.499	4.795	21.762
	2	20.151	7.214	0.646	-0.248	4	12.09
	3	12.5	4.236	0.305	-0.034	2.535	9.439
	4	9.298	2.836	0.101	0.131	2.987	9.57

6. Application

Finding the MVLUE of the location and scale parameters is an intriguing application of this work. Although the [24] methodology could be utilized to obtain MVLUE but it is quite challenging to acquire the inverse of the variance-covariance matrix in closed form for this methodology.

Therefore, these estimates are computed using numerical techniques.

Assume that the distribution of the random variables have μ and σ as the location and scale parameters. Now, based on the [24] technique, the MVLUE of θ is given as $\hat{\theta} = (A'V^{-1}A)^{-1}(A'V^{-1}y)$, where, $V = (V_{i,j})$ represents the variance of the i -th and j -th concomitants, V^{-1} represents the inverse of the matrix V , and y' is the observed value of the vector $Y' = Y_{[1,n,m,k]}, Y_{[2,n,m,k]}, \dots, Y_{[n,n,m,k]}$. For $\mu = 0$ and $\sigma = 1$, A and θ can be obtained as:

$$A = \begin{bmatrix} 1, & 1 & \dots & 1 \\ \mu_{d[1,n,m,k]} & \mu_{d[2,n,m,k]} & \dots & \mu_{d[n,n,m,k]} \end{bmatrix}$$

$$\theta' = [\mu \ \sigma]$$

Table 6.1: Coefficients of MVLUE of μ and σ for FGMTBM exponential distribution based on OS with $n = 4$, $\theta = 0.4$

α	Estimate	Coefficients			
-0.9	$\hat{\mu}$	-0.0828	-0.1638	1.26766	-0.5523
	$\hat{\sigma}$	-0.5017	-0.9922	7.67812	-3.3452
-0.5	$\hat{\mu}$	5.14576	3.93097	-8.4672	0.39049
	$\hat{\sigma}$	-0.9503	-0.825	1.83408	-0.0588
-0.1	$\hat{\mu}$	14.4297	28.8015	-32.732	-9.4992
	$\hat{\sigma}$	-2.7511	-5.8872	6.69428	1.944
0.1	$\hat{\mu}$	-9.1516	-35.805	33.9053	12.0516
	$\hat{\sigma}$	1.98121	7.01331	-6.6369	-2.3576
0.5	$\hat{\mu}$	-5.0134	-3.5421	10.2922	-0.7368
	$\hat{\sigma}$	1.07724	0.74098	-2.047	0.2288
0.9	$\hat{\mu}$	-1.9202	-0.6639	1.18162	2.40244
	$\hat{\sigma}$	0.40411	0.19224	-0.1153	-0.481

Table 6.2: Coefficients of MVLUE of μ and σ for FGMTBM exponential distribution based on record values with $n = 4$, $\theta = 0.4$ and $k = 2$

α	Estimate	Coefficients			
-0.9	$\hat{\mu}$	-9.0704	37.5315	-32.689	5.22765
	$\hat{\sigma}$	2.27034	-8.4608	7.36915	-1.1787
-0.5	$\hat{\mu}$	-9.6243	57.7496	-68.289	21.1632
	$\hat{\sigma}$	2.26517	-12.314	14.5604	-4.5111
-0.1	$\hat{\mu}$	26.7814	-5.3193	24.1669	-44.629
	$\hat{\sigma}$	-5.2229	1.07976	-4.8968	9.04001
0.1	$\hat{\mu}$	-32.758	4.67707	12.0107	17.0699
	$\hat{\sigma}$	6.66864	-0.9234	-2.3731	-3.3721
0.5	$\hat{\mu}$	-19.659	26.1102	4.18588	-9.6375
	$\hat{\sigma}$	3.88893	-4.915	-0.7883	1.81434
0.9	$\hat{\mu}$	-7.5351	10.0486	-1.2653	-0.2481
	$\hat{\sigma}$	1.53447	-1.8065	0.22748	0.04455

7. Conclusions

In this study, FGMTBM exponential distributions was considered and concomitants of GOS were obtained. Additionally, pdf pdf r -th GOS and a joint pdf of r -th and s -th GOS were also obtained. Furthermore, the MVLUE of the location and scale parameters of the concomitants of the k -th upper record values, as well as OS, were obtained for the distribution under consideration. Finally, an example was considered for the purpose of putting the suggested methodology into exercise. When $k(<n)$ individuals are to be picked using a selection method based on their X -values, concomitants are most useful. The performance of a characteristic is then depicted using the pertinent Y -values. When individuals in a random sample are arranged in accordance with the corresponding values of another random sample, concurrent or induced OS are created.

References

- [1] Kamps, U. (1995). A concept of generalized order statistics, B.G. Teubner Stuttgart, Germany (Ph. D. Thesis).
- [2] Khan, A.H., Khan R.U. and Yaqub, M. (2006). Characterization of continuous distributions through conditional expectation of generalized order statistics, J. Appl. Prob. Statist 1, 15-131.
- [3] Tavangar M. and Asadi M. (2008) On a characterization of generalized Pareto distribution based on generalized order statistics, Commun. Statist. Theory Meth. 37, 1347-1352.
- [4] Athar H. and Nayabuddin (2013). Recurrence relations for single and product moments of generalized order statistics from Marshall- Olkin extended general class of distributions, J. Stat. Appl. Prob. 2(2), 63-72.
- [5] Morgenstern, D. (1956). Einfache Beispiele Zweidimensionaler Verteilungen, Mitteilungsblatt fur Mathematische Statistik, 8, 234-235.
- [6] Gumbel, E.J. (1960). Bivariate exponential distributions, J. Amer. Statist. Assoc. 55, 698-707.
- [7] Farlie, D.J.G. (1960). The performance of some correlation coefficients for a general bivariate distribution, Biometrika , 47, 307-323.
- [8] Johnson, N.L. and Kotz, S. (1975). On some generalized Farlie-Gumbel-Morgenstern distributions, Commun. statist. Theor. Meth. 4, 415-427.
- [9] Das, K. K., Das, B. and Baruah, B. K. (2012). A comparative study on concomitant of order statistics and record values for weighted inverse Gaussian distribution. International Journal of Scientific and Research Publication 2, 1-7.
- [10] Tahmasebi, S. and Behboodian, J. (2012). Shannon information for concomitants of generalized order statistics in Farlie-Gumbel-Morgenstern (FGM) family. Bulletin of the Malaysian Mathematical Science Society 34, 975-981.
- [11] Ahsanullah, M. and Beg, M. I. (2006). Concomitant of generalized order statistics in Gumbel's bivariate exponential distribution. Journal of Statistical Theory and Applications 6, 118-132.
- [12] Beg, M. I. and Ahsanullah, M. (2007). Concomitants of generalized order statistics from Farlie Gumbel Morgenstern type bivariate Gumbel distribution, Statistical Methodology. 1-20.
- [13] Chacko, M. and Thomas, P.Y. (2011) . Estimation of a parameter of Morgenstern type bivariate exponential distribution using concomitants of order statistics, Statistical Methodology, 8, 363-376.
- [14] Nayabuddin, (2013). Concomitants of generalized order statistics from bivariate Lomax distribution, ProbStat Forum 6, 73-88.
- [15] Nayabuddin, Athar H. and Al- Helan, M. (2018). Concomitants of generalized order statistics from bivariate Burr XII distribution, *Aligarh Journal of statistics*, 18, 1-18.

- [16] Domma, F. and Giordano, S. (2016). Concomitants of m -generalized order statistics from generalized Farlie-Gumbel-Morgenstern distribution family. *Journal of Computational and Applied Mathematics*, 294, 413-435
- [17] Alawady, Barakat, H. M., Shengwu Xiong and Abd Elgawad, M. A.(2020). Concomitants of generalized order statistics from iterated Farlie–Gumbel–Morgenstern type bivariate distribution, *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2020.1842452
- [18] Saxena, S., Zarrin, S., & Kamal, M. (2012). Computation of reliability and Bayesian analysis of system reliability for Mukherjee Islam failure model. *American Journal of Mathematics and Statistics*, 2(2), 1-4.
- [19] Kamal, M. (2013). Application of geometric process in accelerated life testing analysis with type-I censored Weibull failure data. *Reliability: Theory & Applications*, 8(3 (30)), 87-96.
- [20] Kamal, M. (2021). Parameter estimation for progressive censored data under accelerated life test with k levels of constant stress. *Reliability: Theory & Applications*, 16(3 (63)), 149-159.
- [21] Kamal, M., Rahman, A., Zarrin, S., & Kausar, H. (2021). Statistical inference under step stress partially accelerated life testing for adaptive type-II progressive hybrid censored data. *Journal of Reliability and Statistical Studies*, 585-614.
- [22] Kamal, M. (2022). Parameter estimation based on censored data under partially accelerated life testing for hybrid systems due to unknown failure causes. *CMES-Computer Modeling in Engineering & Sciences*, 130(3), 1239-1269.
- [23] Hasnain, S.A. (2013). Exponentiated moment Exponential Distribution, National College of Business Administration and Economics, Lahore, Pakistan. (Ph.D. Thesis).
- [24] Llyod, E.H.(1952) Least squares estimation of location and scale parameter using order statistics, *Biometrika*, 39, 88-95.

PROJECT CHARACTERISTICS WITH TRIANGULAR FUZZY NUMBER

Adilakshmi Siripurapu¹, Ravi Shankar Nowpada²

•

Dept. of Basic Science and Humanities, Vignan's Institute of Information Technology (A),
Duvvada, Visakhapatnam, AP, India¹
laxmimaths2008@gmail.com

Dept. of Mathematics, Institute of Science, GITAM
(Deemed to be University), Visakhapatnam, AP, India²
drravi68@gmail.com

Abstract

The Critical Path Method (CPM) is required to plan, organize, and arrange for major project networks. A clear estimate of the time duration will help in the successful execution of the CPM. However, the time duration cannot be precisely specified in real life. As a result, there is always uncertainty about the duration of activities, leading to the invention of the fuzzy critical path method. This study proposes a simple method for critical path analysis and project characteristics in a project network with a triangular fuzzy number. Furthermore, this study defines the most critical path and the relative path degree of criticality, both theoretically valid and practical. The suggested method can determine the critical path, project characteristics and critical degree of an activity as shown by an example discussed in some earlier studies. The proposed approach is very simple to apply and does not require knowing the explicit form of the membership functions of the fuzzy activity times.

Keywords: Activity duration, critical degree, critical path method, project characteristics, triangular fuzzy number,

1. Introduction

The project manager needs to complete the required work in time, but recognition of project obstruction plays the main task. Projects may be complicated to handle in various scenarios. To meet this challenge, the CPM has been proven to be a practical implementation in efficiently controlling projects when the activity durations are predictable and known. CPM's objective is to recognize essential activities in the project if resources can be focused on to shorten project duration.

The effective outcome of CPM needs the provision of clearly defined time frames for an individual task. Anyhow, this condition is sometimes difficult to meet in practice since many tasks will be performed initially. To meet such realistic conditions, Zadeh, 1965 [21] familiarized the theory of a fuzzy set. Because of ambiguity regarding the activities period in a network design, the Fuzzy Critical Path Method (FCPM) has been presented in the 1970s. Several techniques have been developed in past years for determining the fuzzy critical path.

Dubois, et al., 1978 [6] calculated the earliest, latest and slack times and criticality levels of activities using fuzzy numbers to demonstrate the network topology and series-parallel graphs. Gazdik [7] initially determined the activity durations and critical paths using fuzzy arithmetic operations. Atanassov et al. 1989 [3] defined membership value, non-membership value, and another element called hesitation; it has been used in various fields. Rommelfanger et al., 1994[16] calculated the activity's earliest, latest and slack times in a project network using fuzzy intervals and expanded all concepts of classical network procedure using the suggested methodology. In order to allocate renewable resources, the FPS (Fuzzy Project Scheduling) method was used. Using time parameters, understand durations, beginning timings, and due dates of certain activities. L-R fuzzy numbers are utilized to model the uncertainty of these parameters. Constructing optimistic and pessimistic schedules based on α -cut levels that have been specified are possible. A fuzzy project was a planning decision support system that applies L-R fuzzy numbers to evaluate activity times. Hapke et al., 1994 [8] proposed fuzzy project scheduling technique is employed in a software project scheduling context to allocate resources across related tasks. Scheduling is made more accurate by using a novel resource allocation approach that accounts for and includes uncertainty in the scheduling process. Loterapong et al., 1994 [12] proposed the fuzzy set theory approach to distribute resources to minimize unexpected project delays, use available resources, and minimize resource shortages. Three criteria are used to measure performance compared to the Minimum Late Finish (MINLFT) and the Minimum Slack (MINSLK) rules project duration, resource utilization, and resource interruption. Chang et al., 1995 [5] introduced the fuzzy Delphi method effectively determines the project length and the critical degree for each direction in a project. Attempted to substitute probabilistic or deterministic assumptions in the project network review with probabilistic ones and decrease the complexity caused by inexact and incomplete knowledge about activity times by representing activity times as fuzzy numbers. Chanas et al., 2002 [4] introduced the idea of criticality in the network having fuzzy activity times. It computed the path degree of criticality using two alternative approaches. Stefan et al., 2002 [20] evaluated the criticality of a network in terms of path and activity duration periods, where intervals or fuzzy intervals describe activity duration times. Sireesha et al., 2010 [19] adapted fuzzy triangular numbers to solve project scheduling problems. This technique computes every activity's total float and the network's critical path beyond performing any forward or backward computations. Oladeinde et al. 2013 [14] utilized an altered fuzzy backward pass technique with a recursive relation to obtaining the most recent start and finish times, which enabled them to overcome negative fuzzy numbers. Khalaf 2013 [11] suggested a new mechanism for dealing with fuzzy project scheduling problems by applying a ranking function. They also suggested a novel technique for estimating every activity's fuzzy cumulative, free, and independent slack. Jaya Gowri et al. [9] proposed an algorithm to tackle the problem in an intuitionistic fuzzy environment. The Triangular intuitionistic fuzzy number is defuzzified using graded mean integration representation. Now the intuitionistic fuzzy number is converted to a crisp number. Then applied, the proposed algorithm to find the critical path. Adilakshmi et al. 2021 [17] find the critical path and project duration with the fuzzy hexagonal number. In their paper, the activity times are represented as a hexagonal fuzzy number and obtained a new ranking function in hexagonal fuzzy number by the centroid of centroid method. By using the ranking function, hexagonal fuzzy numbers are transformed into crisp values. Priyadharshini et al. 2022 [15] developed a different algorithm, namely the maximum edge distance method, to find the optimal path in an intuitionistic fuzzy weighted directed graph with its edge weights as an intuitionistic triangular fuzzy number. The method is an alternative way to identify the critical path in the fuzzy environment. Adilakshmi et al. 2022 [1] find the critical path and project duration with the pentagonal fuzzy number. In their paper, the activity times are represented as a pentagonal fuzzy

number and obtained a new ranking function in pentagonal fuzzy number by the centroid of centroid method. Then applying a ranking function to PFNs is transformed into crisp values. Mahesh et al. [13] developed a new Fuzzy linear programming-based method using a single MF, called modified logistics MF. The modified MF logistics and its modifications taking into account the characteristics of the parameter are from the analysis process. This MF was tested for useful performance by modeling using FLP. Adilakshmi et al. 2022 [2] improved Dijkstra's algorithm for finding the critical path and project duration with the triangular fuzzy number. Their paper represents the activity times as a triangular fuzzy number. Then applied an algorithm to find the critical path.

This paper suggested measuring a fuzzy network's Critical Degree and project Characteristics utilizing tabular representation with Triangular fuzzy activity times and a new subtraction arithmetic operation.

2. Preliminaries

In this section, we will look at a few key definitions.

2.1. Fuzzy Set [16]

As stated in Zadeh's paper, the formalization of a fuzzy set is:

Let X be a space of points (objects), with a generic element of X denoted by x . Thus, $X = \{x\}$. A fuzzy set (class) A in X is characterized by a membership (characteristic function) function $\mu_A(x)$, which associates with each point in X a real number in the interval $[0,1]$, with the value of $\mu_A(x)$ at x representing the "grade of membership" of x in A . When A set in the ordinary sense of the term, its membership function can take on only two values, 0 and 1, $\mu_A(x) = 1$ or 0 according to x does or does not belong to A .

2.2. Fuzzy Number [8]

It is a Fuzzy set of the following conditions:

- Convex fuzzy set
- Normalized fuzzy set.
- Its membership function is piece-wise continuous.
- It is defined in the real number.

Fuzzy numbers should be normalized and convex. Here the condition of normalization implies that the maximum membership value is 1.

2.3. Triangular Fuzzy Number (TFN) [8]

The TFN indicated by $\tilde{A} = (a, b, c)$ and its membership function is given by;

$$\mu_{\tilde{A}}(x) = \begin{cases} 0, & -\infty < x \leq a \\ \frac{x-a}{b-a}, & a \leq x < b \\ \frac{c-x}{c-b}, & b \leq x < c \\ 0, & c \leq x < \infty \end{cases}$$

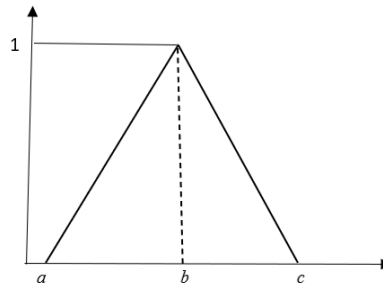


Figure 1: Graphical representation of Triangular fuzzy number

2.4. Arithmetic Operations of Triangular fuzzy number [4]

Let $\tilde{A} = (a_1, a_2, a_3)$ and $\tilde{B} = (b_1, b_2, b_3)$ be two Triangular fuzzy numbers then;

$$k\tilde{A} = k(a_1, a_2, a_3) = (ka_1, ka_2, ka_3)$$

$$\tilde{A} \oplus \tilde{B} = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$$

$$\tilde{A} \ominus \tilde{B} = (a_1 - b_1, a_2 - b_2, a_3 - b_3)$$

$$\tilde{A} \otimes \tilde{B} = (a_1 * b_1, a_2 * b_2, a_3 * b_3)$$

$$\tilde{A} \oslash \tilde{B} = \left(\frac{a_1}{a_2}, \frac{a_2}{b_2}, \frac{a_3}{b_3}\right)$$

Example:

Let $\tilde{A} = (3,6,9)$ and $\tilde{B} = (2,4,6)$ then

$$\tilde{A} \oplus \tilde{B} = (5,10,15)$$

$$\tilde{A} \ominus \tilde{B} = (1,2,3)$$

$$\tilde{A} \otimes \tilde{B} = (6,24,54)$$

$$\tilde{A} \oslash \tilde{B} = (1.5, 1.5, 1.5)$$

Remark: Some authors defined $\tilde{A} \ominus \tilde{B} = (a_1 - b_3, a_2 - b_2, a_3 - b_1)$.

How is it possible?

Here I consider one example.

Let $\tilde{A} = (2,4,6)$ & $\tilde{B} = (2,4,6)$, Here both \tilde{A} and \tilde{B} are same TFNs.

Now $\tilde{A} \ominus \tilde{B} = (2 - 6, 4 - 4, 6 - 2)$

$$= (-4, 0, 4).$$

It is a completely wrong output since both \tilde{A} and \tilde{B} are the same.

According to my definition;

$$\tilde{A} \ominus \tilde{B} = (2 - 2, 4 - 4, 6 - 6) = (0, 0, 0)$$

3. Proposal Method

Fuzzy numbers are acceptable representations in project networks since activity periods are frequently challenging to predict or define exactly. In this chapter, the length of every activity is expressed as a positive TFN.

The below steps to follow to find project critical path and characteristics.

Step1: Construct a network diagram $G(V, E, T)$ with the set of given vertices and edges, where $V = \{v_1, v_2, \dots, v_n\}$ be set of n vertices, E be set of edges which represent activities, and T be set of fuzzy triangular numbers as activity times.

Step2: Construct a square table with $(n - 1) \times (n - 1)$ order. "i" indicate the number of rows, and the column number is indicated by j , where $i = 1, 2, 3, \dots, n - 1; j = 2, 3, \dots, n$. Two numbers (i, j) describes each entry in the table, where i indicates the row number and j indicates the column number.

Step3: Set the entries \tilde{T}_{ij} , the highest fuzzy time needed from source node to tail node $i \rightarrow j$. Insert the times \tilde{t}_{ij} in the table's first row. Then add the preceding time of the second node and t_{2j} to complete the elements in the second row $i = 2$. Then proceed by adding the preceding time of the third node and t_{3j} to the elements of the third row $i = 3$, and so on. If we get more than one path, then we take the maximum value. Continuing this procedure on down the rows up to the ending row is attained.

Step4: To determine the critical path, go backward from the known endpoint. Begin by picking the element from the end column $j = n$ with the highest fuzzy value. This value represents the project completion time. The row number corresponding to the highest value in the end column provides the preceding activity's end node index; hence, the next activity head node on the critical path is noted. If column- n has the greatest number in row k , column k is checked, and the greatest number in that column is needed. This step is continued until the initial node is reached when a list of all actions along the fuzzy critical path is obtained.

Step5: Create a new $(n - 1) \times (n - 1)$ order table to compute free slack times. Choose the entry with the largest value $\tilde{T}_{end} = \tilde{T}_{kj}$ where $k = 2, 3, \dots, n - 1; j = n$.

The free slack of activity (i, j) is calculated using $F\tilde{S}_{ij} = \max\{\tilde{T}_{kj}\} - \tilde{T}_{ij}$.

Where $\tilde{T}_{ij} = t_{ij} +$ the maximum fuzzy completion time in column j .

Step6: Form another $(n - 1) \times (n - 1)$ order table for computing total slack time using $T\tilde{S}_{ij} = F\tilde{S}_{ij} + \min\{T\tilde{S}_{jk}\}$ for all $k > j$.

In the last row $T\tilde{S}_{ij} = F\tilde{S}_{ij}$ is taken to begin the computations.

Step7: Compute the earliest and latest activity periods by using;

Earliest start time $(E\tilde{S}_{ij}) = \tilde{E}_i$

Earliest finish time $(E\tilde{F}_{ij}) = \tilde{E}_i + \tilde{t}_{ij}$

Latest finish time $(L\tilde{F}_{ij}) = \tilde{L}_j$

Latest start time $(L\tilde{S}_{ij}) = \tilde{L}_j - \tilde{t}_{ij}$; Total Float $(T\tilde{F}_{ij}) = L\tilde{S}_{ij} - E\tilde{S}_{ij}$ or $L\tilde{F}_{ij} - E\tilde{F}_{ij}$

4. Application

Consider a site preparation and concrete slab foundation that includes nine different activities. The network table is shown in Table 1, and the network diagram is presented in Figure 1, where every activity duration is presented as a TFN.

Table1: Project Activity duration with TFN

Activity	Activity description	TFN
1→2	Removal tress	(6,12,18)
1→3	Site clearing	(13,25,33)
2→3	General excavation	(15,20,29)
2→4	Placing Formwork and reinforcement for concrete	(2,11,20)
2→5	Excavation for utility frame work	(7,16,25)
3→5	Grading general area	(17,25,33)
4→5	Installing sewer line	(8,9,10)
4→6	Pouring concrete	(15,21,31)
5→6	Installing other utilities	(4,11,14)

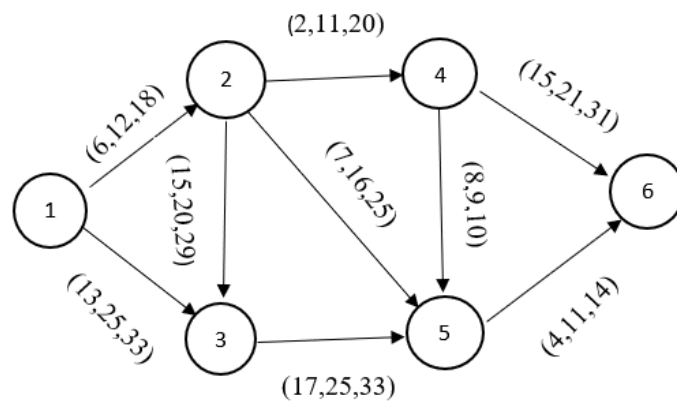


Figure2: Fuzzy project network with activity time as TFN

The tabular representation of Fuzzy activity times is presented in Table 2.

Table 2: Tabular description of Fuzzy activity times with TFN

(i, j)	2	3	4	5	6
1	(6,12,18)	(13,25,33)			
2		(15,20,29)	(2,11,20)	(7,16,25)	
3				(17,25,33)	
4				(8,9,10)	(15,21,31)
5					(4,11,14)

5. Results

5.1. Project critical path

In section 3 explained the procedure. From this procedure, calculated fuzzy critical path. Table3 represents maximum fuzzy time durations.

Table 3: Tabular representation of Maximum Fuzzy activity duration (\tilde{T}_{ij})

(i, j)	2	3	4	5	6
1	(6,12,18)	(13,25,33)			
2		(21,32,47)	(8,23,38)	(13,28,43)	
3				(38,57,80)	
4				(16,32,48)	(23,44,69)
5					(42,68,94)

Using the following ranking formula, calculated rank of maximum fuzzy time duration (\tilde{T}_{ij}) and represented in Table 4.

Let (a, b, c) be a TFN, then $\mathcal{R}(a, b, c) = \frac{a+2b+c}{4}$.

Table 4: Tabular representation of Rank of \tilde{T}_{ij}

(i, j)	2	3	4	5	6
1	12	24			
2		33	23	28	
3				58	
4				32	45
5					68

From Table 4,

In the 6th column; $\mathcal{R}\{\tilde{T}_{56}\} > \mathcal{R}\{\tilde{T}_{46}\} \Rightarrow \tilde{T}_{56} > \tilde{T}_{46}$

As a result, the sixth row has the greatest fuzzy value in the last column.

Therefore, the last activity's head node index is 5, and the last activity's tail node index is 6.

In the 5th column; $\mathcal{R}\{\tilde{T}_{35}\} > \mathcal{R}\{\tilde{T}_{45}\} > \mathcal{R}\{\tilde{T}_{25}\} \Rightarrow \tilde{T}_{35} > \tilde{T}_{45} > \tilde{T}_{25}$

Therefore, the 3rd row is the highest fuzzy value in the 5th column.

Thus, the last activity's head node index is 3, and the last activity's tail node index is 5.

In the 4th column;

\tilde{T}_{24} is the only fuzzy number in the column.

Therefore, the 2nd row is the highest fuzzy value in the 4th column.

Thus, the last activity's head node index is 2, and the last activity's tail node index is 4.

In the 3rd column;

In the 3rd column, $\mathcal{R}\{\tilde{T}_{23}\} > \mathcal{R}\{\tilde{T}_{13}\} \Rightarrow \tilde{T}_{23} > \tilde{T}_{13}$

Therefore, the 3rd row is the highest fuzzy value in the 3rd column.

Thus, the last activity's head node index is 2, and the last activity's tail node index is 3.

In the 2nd column, \tilde{T}_{12} is the only fuzzy number.

Therefore, 1st row is the highest fuzzy value in 2nd column.

Thus, the last activity's head node index is 1, and the last activity's tail node index is 2.

The head node on the fuzzy critical path of the next to last activity is identified if the column's highest fuzzy number is in row k , then column k is checked, and the highest fuzzy number in that column is needed. Then, the last activity's head node index is 1, and the last activity's tail node index is 2.

The tabular representation of identified critical path is presented in Table 5.

Table 5: Maximum Fuzzy duration of Critical path

(i, j)	2	3	4	5	6
1	(6,12,18)				
2		(21,32,47)			
3				(38,57,80)	
4					
5					(42,68,94)

Therefore, the project critical path is 1→2→3→5→6 and the project end period is 68 days.

5.2. Project Characteristics

In section 3, explained earliest, latest and total floats formulas. Adapting those formulas, computed earliest, latest and total float of every activity.

5.2.1. Fuzzy Earliest and Latest Times of events

Table 6 shows the fuzzy events earliest and latest times.

Table 6: Earliest and Latest times of Fuzzy Network events with TFN

Event (<i>i</i>)	\tilde{E}_i	\tilde{L}_i
1	(0,0,0)	(0,0,0)
2	(6,12,18)	(6,12,18)
3	(21,32,47)	(21,32,47)
4	(8,23,38)	(27,47,63)
5	(38,57,80)	(38,57,80)
6	(42,68,94)	(42,68,94)

Table 7 represents activities earliest; latest and total float time of every activity.

Table 7: Earliest, latest and total float times of every activity with TFN

$i \rightarrow j$	$E\tilde{S}_{ij}$	$E\tilde{F}_{ij}$	$L\tilde{S}_{ij}$	$L\tilde{F}_{ij}$	Total float($T\tilde{F}_{ij}$)
1→2	(0,0,0)	(6,12,18)	(0,0,0)	(6,12,18)	(0,0,0) *
1→3	(0,0,0)	(13,25,33)	(8,7,14)	(21,32,47)	(8,7,14)
2→3	(6,12,18)	(21,32,47)	(6,12,18)	(21,32,47)	(0,0,0) *
2→4	(6,12,18)	(8,23,38)	(25,36,43)	(27,47,63)	(19,24,25)
2→5	(6,12,18)	(13,28,43)	(31,41,55)	(38,57,80)	(25,29,37)
3→5	(21,32,47)	(38,57,80)	(21,32,47)	(38,57,80)	(0,0,0) *
4→5	(8,23,38)	(16,32,48)	(30,48,70)	(38,57,80)	(22,25,32)
4→6	(8,23,38)	(23,44,69)	(27,47,63)	(42,68,94)	(19,24,25)
5→6	(38,57,80)	(42,68,94)	(38,57,80)	(42,68,94)	(0,0,0) *

5.3. Fuzzy slack times

Slack time is an important factor in project management and planning. There are two kinds of slacks for each activity, i.e., fuzzy free slack ($F\tilde{S}_{ij}$) and fuzzy total slack($T\tilde{S}_{ij}$).

5.3.1. Fuzzy free slack time

An activity's fuzzy free slack time ($F\tilde{S}_{ij}$) is the duration of activity that could be prolonged, or its beginning can be delayed without impacting the beginning period of the immediately following activity. By default, every value in the \tilde{T}_{ij} table reflects the maximal duration needed to achieve and complete the task(i, j); thus, the gap within the maximum column and particular element in that column indicates how much free slack time is feasible for that particular activity.

In section 3, the method to find Fuzzy free slack times using a tabular representation is explained.

Fuzzy free slack time calculations

$$F\tilde{S}_{12} = (6,12,18) - (6,12,18) = (0,0,0)$$

$$F\tilde{S}_{13} = (21,32,47) - (13,25,33) = (8,7,14)$$

$$F\tilde{S}_{23} = (21,32,47) - (21,32,47) = (0,0,0)$$

$$F\tilde{S}_{24} = (8,3,28) - (8,23,38) = (0,0,0)$$

$$F\tilde{S}_{25} = (38,57,80) - (13,28,43) = (25,29,37)$$

$$F\tilde{S}_{35} = (38,57,80) - (38,57,80) = (0,0,0)$$

$$F\tilde{S}_{45} = (38,57,80) - (16,32,48) = (22,25,32)$$

$$F\tilde{S}_{46} = (42,68,94) - (23,44,69) = (19,24,25)$$

$$F\tilde{S}_{56} = (42,68,94) - (42,68,94) = (0,0,0)$$

The calculated free slack times of every activity are represented in Table 8.

Table 8: Tabular representation of free slack times

$i \rightarrow j$	2	3	4	5	6
1	(0,0,0)	(8,7,14)			
2		(0,0,0)	(0,0,0)	(25,29,37)	
3				(0,0,0)	
4				(22,25,32)	(19,24,25)
5					(0,0,0)

5.3.2. Total Slack Fuzzy Times

The total slack time is the delay or an activity extension(i, j)period that can occur without affecting the project's completion.

In section 3, explained how to calculate total slack time for each activity.

Total slack time calculations for fuzzy project network

$$T\tilde{S}_{56} = (0,0,0) + (0,0,0) = (0,0,0)$$

$$T\tilde{S}_{46} = (19,24,25) + (0,0,0) = (19,24,25)$$

$$T\tilde{S}_{45} = (22,25,32) + (0,0,0) = (22,25,32)$$

$$T\tilde{S}_{35} = (0,0,0) + (0,0,0) = (0,0,0)$$

$$T\tilde{S}_{25} = (25,29,37) + (0,0,0) = (25,29,37)$$

$$T\tilde{S}_{24} = (0,0,0) + (22,25,32) = (22,25,32)$$

$$T\tilde{S}_{23} = (0,0,0) + (0,0,0) = (0,0,0)$$

$$T\tilde{S}_{13} = (8,7,14) + (0,0,0) = (8,7,14)$$

$$T\tilde{S}_{12} = (0,0,0) + (0,0,0) = (0,0,0)$$

Table 9 displays the total fuzzy slack periods of every activity in the fuzzy project network.

Table 9: Tabular representation of total float of each activity

$i \rightarrow j$	2	3	4	5	6
1	(0,0,0)	(8,7,14)			
2		(0,0,0)	(19,24,25)	(25,29,37)	
3				(0,0,0)	
4				(22,25,32)	
5					(0,0,0)

The above table shows that activities 1→2, 2→3, 3→5, 5→6 have zero slack time. Therefore, the critical path is 1→2→3→5→6.

5.4. Critical degree

In the PERT, an activity considers a critical activity; its total float is zero. That is, the total float reduces, and then criticality increases. In the Fuzzy methodology, if $C\tilde{D}_{ij} = 1$, the activity (i, j) is considered as a critical activity.

Let $T\tilde{F}_{ij} = (a_i, b_i, c_i)$ be the total float of fuzzy activity (i, j) , then the Critical Degree of activity is described as follows:

$$C\tilde{D}_{ij} = \begin{cases} 1, & b_i \leq 0 \\ \frac{-a_i}{b_i - a_i}, & a_i < 0 < b_i \\ 0, & a_i \geq 0 \end{cases}$$

Table 10 provides the critical degree calculations for each activity.

Table 10: Critical Degree of Each Activity

Activity	Total float	Critical degree
1→2	(0,0,0)	1*
1→3	(8,7,14)	0
2→3	(0,0,0)	1*
2→4	(19,24,25)	0
2→5	(25,29,37)	0
3→5	(0,0,0)	1*
4→5	(22,25,32)	0
4→6	(19,24,25)	0
5→6	(0,0,0)	1*

We noticed from the above table that activities 1→2, 2→3, 3→5, 5→6 have critical degree 1. According to the critical degree definition, activities 1→2, 2→3, 3→5, and 5→6 is considered as critical activities.

Therefore, the critical path is 1→2→3→5→6.

6. Conclusion

This paper presented various methods to determine the critical path with a triangular fuzzy number. Earlier work on network planning employing fuzzy sets theory has developed approaches for project scheduling. However, these techniques do not directly enable backward-pass computations in the same way they do forward-pass calculations. Project parameters like earliest, latest times, total float, slack times and critical degree are computed with TFNs. One significant set of this technique uses primary arithmetic fuzzy operations to get relevant determinable outputs. We presented a novel strategy in this paper: the Fuzzy tabular representation.

References

- [1] Adilakshmi Siripurapu, and Ravi Shankar Nowpada. "Fuzzy Project Planning and Scheduling with Pentagonal Fuzzy Number" Reliability: Theory & Applications, vol. 17, no. 3 (69), 2022, pp. 131-138. doi:10.24412/1932-2321-2022-369-131-138
- [2] Adilakshmi Siripurapu, Ravi Shankar Nowpada , and K. Srinivasa Rao. "Improving Dijkstra's algorithm for Estimating Project Characteristics and Critical Path" Reliability: Theory & Applications, vol. 17, no. 4 (71), 2022, pp. 65-73. doi:10.24412/1932-2321-2022-471-65-73

- [3] Atanassov. K, G. Gargov (1989), Interval valued intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, Volume 31, Issue 3, 1989, Pages 343-349, ISSN 0165-0114.
- [4] Chanas S, Dubois D, Zielinski P. (2002), On the sure criticality of tasks in activity networks with imprecise durations. *IEEE Trans Syst Man Cybern B Cybern.* 32(4):393-407.
- [5] Chang Seong, Yasuhiro Tsujimura, Mitsuo Gen, Tatsumi Tozawa (1995), An efficient approach for large scale project planning based on fuzzy Delphi method, *Fuzzy Sets and Systems*, Volume 76, Issue 3, Pages 277-288, ISSN 0165-0114.
- [6] Dubois. D, Henri Prade (1987), The mean value of a fuzzy number, *Fuzzy Sets and Systems*, Volume 24, Issue 3, Pages 279-300, ISSN 0165-0114.
- [7] Gazdik (1983), Fuzzy-Network Planning - FNET, *IEEE Transactions on Reliability*, vol. R-32, no. 3, pp. 304-313.
- [8] Hapke Maciej & Roman Slowinski (1993), A DSS for Resource-Constrained Project Scheduling under Uncertainty, *Journal of Decision Systems*, volume 2, Issue 2, PP 111-128.
- [9] Jayagowri P. & Geetharamani G. (2014). A Critical Path Problem Using Intuitionistic Trapezoidal Fuzzy Number. *Applied Mathematical Sciences*, 8, 2555-2562.
- [10] Kaufmann, A. and Madan M. Gupta (1986). Introduction to fuzzy arithmetic: theory and applications. *Van Nostrand Reinhold*, New York.
- [11] Khalaf, Wakas. (2013). Solving the Fuzzy Project Scheduling Problem Based on a Ranking Function. *Australian Journal of Basic and Applied Sciences*. 7. 806-811.
- [12] Lorterapong, P. (1994), A fuzzy heuristic method for resource-constrained project scheduling. *Project Management Journal*, 25(4), 12-18.
- [13] Mahesh M. Janolkar, Kirankumar L. Bondar, and Pandit U. Chopade. "Fuzzy Linear Programming Approach for Solving Production Planning Problem" *Reliability: Theory & Applications*, vol. 17, no. 4 (71), 2022, pp. 176-185. doi:10.24412/1932-2321-2022-471-176-185
- [14] Oladeinde, M., & Itsisor, D. (2013), Application of Fuzzy theory to project scheduling with critical path method. *Journal of Applied Sciences and Environmental Management*, 17, 161-166.
- [15] Priyadarshini. S and G. Deepa. "CRITICAL PATH INTERMS OF INTUITIONISTIC TRIANGULAR FUZZY NUMBERS USING MAXIMUM EDGE DISTANCE METHOD" *Reliability: Theory & Applications*, vol. 17, no. 1 (67), 2022, pp. 382-390.
- [16] Rommelfanger Heinrich J. (1994), Network analysis and information flow in fuzzy environment, *Fuzzy Sets and Systems*, Volume 67, Issue 1, Pages 119-128, ISSN 0165-0114.
- [17] S. Adilakshmi and Dr. N. Ravi Shankar. "A New Ranking in Hexagonal Fuzzy number by Centroid of Centroids and Application in Fuzzy Critical Path" *Reliability: Theory & Applications*, vol. 16, no. 2 (62), 2021, pp. 124-135. doi:10.24412/1932-2321-2021-262-124-135
- [18] Sireesha, V., & Shankar, N.R. (2010), A New Approach to find Total Float time and Critical Path in a fuzzy Project Network. *Society (NAFIPS 2007)* pp. 277-282.
- [19] Sireesha, V., & Shankar, N.R. (2010), A New Approach to find Total Float time and Critical Path in a fuzzy Project Network. *Society (NAFIPS 2007)* pp. 277-282.
- [20] Stefan Chanas & Paweł Zieliński (2001), Critical path analysis in the network with fuzzy activity times, *Fuzzy Sets and Systems*, Volume 122, Issue 2, Pages 195-204, ISSN 0165-0114.
- [21] Zadeh L.A. (1965), Fuzzy sets, *Information and Control*, Volume 8, Issue 3, Pages 338-353, ISSN 0019-9958.

RELIABILITY FACTORS OF SOFTWARE FOR MICROPROCESSOR PROTECTIONS

M. I. Uspensky



Institute for Socio-Economic & Energy Problems of the North of FRC Komi Science Center,
Ural Branch, Russian Academy of Sciences, Syktyvkar, Russia
uspensky@energy.komisc.ru

Abstract

The features of the program functioning reliability that are typical for critical applications operating in real mode for microprocessor protections are considered. Among the main characteristics of the relay protection functioning are reliability indicators. With the transition to the execution of such protections on a microprocessor-based basis, in addition to hardware reliability, it became necessary to characterize its operation by software reliability. The importance of the solve tasks in the operation process refers it to programs used in safety-critical software and operating in real mode. This, in turn, tightens the requirements for their reliability evaluation. Comparative analysis made it possible to assess the generality of the considered reliability factors for the functioning of the application software under consideration. At the same time, we were able to highlight some of the features that are typical for the implementation of relay protection on microprocessors. An example of such an assessment is given, showing that despite all the difficulties of complete testing for microprocessor protection programs the software erroneous contribution still amounts to 2.5% of the total contribution. It is shown that microprocessor protection programs are related to programs of critical applications operating in real mode, which makes it possible to use the experience and characteristics of such software in solving relay protection challenges. Nevertheless, some features are given that are specific for relay protection based on a microprocessor. Further tasks are defined.

Key words: safety-critical software, microprocessor relay protection, software reliability.

1. INTRODUCTION

Recently, programmers have been paying a lot of attention to assessing the reliable operation of programs used in mission-critical applications and operating in real mode. Among a whole range of such programs that perform important functions in nuclear power plants, for military equipment, etc., a large share of them is involved in control and protection of electric power system (EPS) equipment. Such software is required to perform its functions with high reliability.

In the engineering community, software systems have a reputation for being unstable. There have been cases where programmers' mistakes have led to major accidents with large financial losses. So on 14.08.2013 on the east coast of the USA, 55 million people were without electricity, because the operator of the power plant turned off the alarm about the violation, so as not to interfere with the work, corrected the mode and forgot to turn on the alarm, and the program did not remind about it, and the next violation of the mode led to a major accident, because the alarm about it was not received. Because the designers forgot to cross the date line, 12 fighters on their way from the United States to Okinawa lost access to fuel data, speed and altitude sensors, and communications were partially disrupted. For hours, America's most advanced F-22 Raptor fighters flew across the ocean completely helpless. In the end they were only able to land thanks to the skill of their pilots. And there are hundreds of similar examples.

2. SPECIFICS OF SAFETY-CRITICAL SOFTWARE

So why don't engineers avoid software if it is not trustworthy? There are usually three main advantages of replacing hardware with software:

1. Software technology allows more logic to be built into the system. Software-controlled computer systems can distinguish between a large number of situations and produce outputs corresponding to each situation. Hard logic systems could not achieve this behavior without an exorbitant amount of hardware. Programmable hardware is less expensive than equivalent hardwired logic because it is regularly structured and mass produced. The economics of the situation also allow program-controlled systems to perform more checks; reliability can be increased by periodically running programs that check the hardware.
2. On the other hand, the cost of writing and certifying truly reliable software is very high; in addition, maintenance costs must be taken into account – again, one that does not undermine reliability and security. An illustrative example: only maintenance of relatively simple and not very big in volume (about 400 thousand words) software for the onboard computer installed on the American spacecraft, type Shuttle, has cost NASA 100 million dollars per year [1].
3. The logic implemented in the software is theoretically easier to change than the logic implemented in the hardware. Many changes can be made without adding new components. If the system is replicated or in a hard to reach physical location, it is much easier to make changes to the software than to the hardware.
4. On the other hand, changes to software modules are easy to make technically, but difficult to do without introducing new bugs. The verification and certification required for security guarantees mean new high costs. In addition, the longer the lifetime of a program, the greater the danger of introducing bugs along with changes – for example, because some developers cease to be such over time, and documentation is rarely exhaustive. Meanwhile, the scale of changes in software can be very large. For example, the Shuttle software underwent 14 modifications over 10 years of maintenance since 1980, which resulted in 152,000 lines of code changes (the total software volume is 400,000 lines). The need for software upgrades was dictated by periodic hardware upgrades, adding functionality, and by the need to correct identified defects.
5. Computer technology and software flexibility allow more information to be made available to operators and provided in a more useful form. The operator of a modern software-controlled system can be provided with information that would be unthinkable in a purely hardware system. All of this can be accomplished with less space and power than was used in noncomputerized systems [2].

In considering reliability relative to software, it should be remembered that:

- The most obvious difference between software and hardware technologies is their complexity. Thus, accurate documentation in fairly general notation for small software systems can fill a bookcase, whereas for hardware a few diagrams and a brief description are sufficient. Another indicator of complexity is the time it takes the programmer to become intimately familiar with the system. Even with small software systems, it is often the case that it takes a programmer a year to work with a program before he or she can be trusted to make improvements on his or her own.
- The next notable property of software is its sensitivity to small errors. In conventional mechanical engineering, every design and manufacturing measurement can be characterized by

a tolerance. One is not required to be an exact match; it is sufficient to be within the specified tolerance of the desired value. The use of tolerances is justified by the assumption that small errors have small consequences. It is well known that in software, trivial clerical errors can have the most serious consequences. For software, there is no known useful interpretation of a tolerance. A single punctuation error can be catastrophic, even though fundamental missteps sometimes have minor consequences.

- Software is notoriously difficult to adequately test. It often happens that a piece of software, subjected to careful and disciplined testing, is seriously flawed. Analog device testing is based on interpolation. It is assumed that devices that perform well at two close points will also perform well at intermediate points. In software, this assumption is wrong. The number of cases that must be tested in order to generate confidence in the software is usually very large.
- Many of the assumptions that are commonly made when designing high-reliability equipment are invalid for software. Developers of high-reliability equipment are concerned about manufacturing failures and wear and tear phenomena. They may conduct their analysis under the assumption that failures are not highly correlated and simultaneous failures are unlikely. Those evaluating the reliability of hardware systems would have to be concerned about design errors and correlated failures; however, in many situations, the effects of other types of errors are dominant. Software has few errors introduced at the production (compilation) stage; when such errors do exist, they are systematic rather than random. Software does not wear out. The errors that software reliability specialists must deal with are design errors. These errors cannot be considered statistically independent. There is ample evidence that even when programs for a given task are written by people who do not know each other, they have closely related errors [3, 4].

These properties are a fundamental consequence of the fact that the mathematical functions implemented by the software are not continuous functions, but functions with an arbitrary number of discontinuities. The lack of continuity constraints on functions describing software effects makes it difficult to find compact descriptions of the software. The lack of such constraints gives software flexibility, but also complexity. Similarly, sensitivity to small errors and difficulty in testing can be attributed to fundamental mathematical properties; a miracle cure is unlikely to be found. Reliability-critical software systems will always require a great deal of discipline and scrutiny.

In contrast to the situation with hardware systems, you cannot achieve higher reliability by duplicating software components. Errors are simply duplicated. Even if programs are written independently, errors made by one programmer are often shared by others. As a result, one cannot expect to improve the reliability of software systems only by having three computers where one is enough, although even here everything is not so simple.

It is appropriate to recall that in the practice of developing and using software, great importance is attached to solving the problem of its safety by executing a standard series of the IEC 61508-N-2010 "«Functional safety of electrical/electronic/programmable electronic safety-related systems", where $N = 1, \dots, 4$. This series details the goals and requirements for managing the safety of functioning, including programmable systems. But the actual filling of these requirements remains with the software developers.

3. FEATURES OF RELAY PROTECTION PROGRAMS ON MICROPROCESSOR

Let us consider application of the above properties and peculiarities to relay protection programs on microprocessor which also refer to reliability-critical software systems.

Applied program modules of relay protection are characterized by relatively small volume of programs (in average 3-5 thousand lines in assembler equivalent), that is determined, first of all, by requirements to speed of protection. On the other hand, such volumes create a good visibility of the program text, which reduces the appearance of design errors. This also includes the fact that in some cases such application programs are prepared in the languages of programmable logic controllers (PLCs) [5], which also reduces the likelihood of program errors. However, the operating environment is written in more traditional software languages, such as C, Java, etc.

Another characteristic of these modules is related to the fact that this or that violation of the mode of EPS components can be detected in different ways on the basis of the same input data and often on different microprocessors. In traditional protection, this approach is known as redundancy. In microprocessor protection, it allows duplicating software when processing the input data, because the algorithms of their processing are not repeated. As a result, at such organization the restriction on duplication of programs is to some extent weakened, because such organization of their work is rather a redundancy than a duplication, although in reliability modeling they are included in parallel in terms of task performance (detection of an unacceptable mode). Of course, here too, each of the programs may have its own errors, which are not detected and are not similar to each other, but this is closer to the duplication of the hardware part of protection.

And of course it is important that the protection system is not loaded with some auxiliary background tasks, which can be a source of errors and crashes. It should be as autonomous as possible if we consider it a critical application. The tendency to combine several protection tasks on a single processor leads to an increase in the size of the software, hence to a worse visibility of the programs, and hence to an increase in the probability of software errors.

Thus, various probabilistic assessments of reliability indicators of relay protection software functioning, similar to hardware assessments, give only a qualitative picture in a rather wide range. Quantitative assessment can be judged only from statistics, which appears after years of operation of a significant number of devices.

So according to statistical data, the number of microprocessor-based relay protection and automatic devices (RPAD) in operation in 2013 were 274062 devices, and in 2014 - 319912 devices ([6], Table 4). From the data "Distribution of cases of malfunction of RPAD devices by types of technical reasons and types of RPAD devices for the period from 01.01.2020 to 30.06.2020" [7] we know that from 727 cases of RPAD failure 29 cases are related to failure of microprocessor protections hardware, and 18 cases - to failure or failure of their software. Then the forecast number of relay protection devices for 2020 in relation to 2013 (in 7 years) can be made from the formula

$$a_n = a_1(1 + r)^n$$

with $r = (a_2 - a_1)/a_1$, where $a_1 = 274062$ is the number of devices of the first year (2013), $a_2 = 319912$ (2014) is the number of devices of the next year, a_n is the number of devices for year n (2020), r is the average annual growth of devices, then

$$a_7 = 274062 \left(1 + \frac{319912 - 274062}{274062} \right)^7 = 809321 \text{ devices.}$$

Let's take Rosseti's share of RPAD to be 70% of all devices in [7]. Then a rough estimation of intensity of failures $\lambda = \frac{m}{0.7 \cdot N \cdot t} = \frac{18 \cdot 2}{0.7 \cdot 809321} = 6.35 \cdot 10^{-5}$ years⁻¹. Here m - the number of device failures due to software, for half a year (2 is multiplier up to a year), $0.7 \cdot N$ - the number of all Rosseti's microprocessor protections, t - design period (year). When the recovery time $t_r = 2$ h

$$(\mu = \frac{1}{t_r} = \frac{8760}{2} = 4380 \text{ years}^{-1}),$$

software availability factor

$$K_{av.sw} = \frac{\mu}{\lambda + \mu} = \frac{4380}{6.35 \cdot 10^{-5} + 4380} = 0.99999985$$

And the average operating time before the error

$$t_e = 1/\lambda = 15748 \text{ years}$$

The proportion of errors due to software from the number of all protection failures is

$$Pr_{sw} = \frac{m_{sw}}{m_{sw} + m_{oth}} \cdot 100\% = \frac{18}{727} \cdot 100\% \approx 2.5\%$$

where $m_{sw} = 18$ is failures of devices due to software, $m_{sw} + m_{oth} = 727$ all cases of incorrect operation of microprocessor protections. 29 of them are related to hardware. Note that the proportion of failures due to software errors was 2.5%.

According to the data of work [8] at small sampling a error share of software in the total number of failures can be estimated, as $(3+1+3+4) / (11+15+18+17) = 11/61 \cdot 100\% = 18\%$, where in numerator - failures of microprocessor devices because of software, and in denominator - total failures. Of course, the small sample does not allow to judge confidently about the representativeness of the figures, but nevertheless some idea of the ratio is given.

4. CONCLUSIONS

Microprocessor-based protection programs refer to the programs of critical applications operating in real mode, and therefore they are required to perform their functions with high reliability. Given the development and widespread use of computer technology and the flexibility of microprocessor-based protection software, we must not forget the complexity of their description, sensitivity to even small deviations from the basic algorithm, testing problems associated with the lack of continuity limitations of the functions of programs describing the algorithms of solutions, which does not contribute to the design with errors that have little effect on the performance of the required functions.

There are some properties of these programs that improve their reliability performance, such as brevity, programmable logic controller languages, but this is not enough to solve the reliability problems of microprocessor protection programs as programs, critical applications. On the other hand, sufficient autonomy of individual protection devices is necessary, because the desire for comprehensiveness often reduces the reliability parameters of protection. Here it is necessary to look for optimal solutions, taking into account both the functionality of programmable protections, and the achievement of the required characteristics of the reliability of the execution of these functions.

The work was done within the framework of the topic AAA-A20-120051590026-3 "Models and methods of adaptation of power engineering systems in modern conditions".

References

- [1] An Assessment of Space Shuttle Flight Software Development processes. - Committee for Review of Oversight Mechanisms for Space Shuttle Flight Software Development Processes, National Research Council, 1993. 206 p. <https://ntrs.nasa.gov/api/citations/19930019745/downloads/19930019745.pdf>
- [2] Parnas D.L., Schouwen A.J., Kwan S.P. Evaluation of Safety-Critical Software // Communications of the ACM. 1990, Vol. 33, No. 6. Pp. 636-648
- [3] Bishop P.G., Pullen F.D. Error Masking: A Source of Failure Dependency in Multi-Version Programs / 1991, January. 17 p. DOI: 10.1007/978-3-7091-9123-1.3.
- [4] Eckhardt D.E. et al. An experimental evaluation of software redundancy as a strategy for improving reliability / D.E. Eckhardt A.K. Caglayan J.C. Knight L.D. Lee D.F. McAllister J.P.J. Kelly // IEEE Transactions on Software Engineering. Vol.17, Is.7, 1991. Pp.692-702. DOI: 10.1109/32.83905
- [5] Livshits, Yu. E. Programmable Logic Controllers for Process Control / Minsk: BNTU, 2014. - Part 1. - 206 p. (In Russian).
- [6] The concept of development of relay protection and automation of the electric grid complex // Appendix No. 1 to the Protocol of the Board of JSC "Rosseti" dated 06/22/2015 No. 356pr. M., 2015. - 49 p. Available in: <https://mig-energo.ru/wp-content/uploads/2015/12/rza-fsk.pdf> (In Russian).
- [7] Distribution of cases of incorrect operation of relay protection devices by types of technical reasons and types of relay protection devices for the period from 01.01.2020 to 30.06.2020// Available in: https://www.so-ups.ru/fileadmin/files/company/rza/rza_rez_info/rza_rez_vid_teh_1-2k2020.xl
- [8] Zakharov O. G. Reliability of digital relay protection devices. Indicators. Requirements. Estimates. - Moscow: Infra-Engineering, 2018. - 128 p. (In Russian).

ANALYSIS OF $M^{[X_1]}, M^{[X_2]}/G_1, G_2/1$ RETRIAL QUEUE WITH PRIORITY SERVICES, DIFFERENTIATE BREAKDOWN, REPAIR, SYNCHRONIZED RENEGING AND OPTIONAL VACATION

G. AYYAPPAN, S. NITHYA

•

Department of Mathematics
Puducherry Technological University
Puducherry-605014, India
ayyappan@ptuniv.edu.in, nithyamouttouraman@gmail.com

Abstract

This study deals with the steady-state analysis of single-server retrial non-preemptive priority queue with differentiate breakdown, repair, synchronized renegeing and optional vacation. For this purpose, two categories of customers are considered, priority and ordinary customers, who arrive as per Poisson arrival process. The server consistently affords single service for these customers based on general distribution. The server randomly fails while providing service to the customer. Hard failure and soft failure are the two kinds of system failure. Hard failure is defined as an equipment failure that requires a repairman with specialized knowledge to be physically present, which is a time-consuming process. Whereas soft failure is defined as failure caused by events rather than physical condition and is usually resolved rebooting the system. Ordinary customers may renege the orbit if the server is engaged or unavailable. Furthermore, once the service of all priority customers is completed by the server, the server goes for a vacation or becomes idle. In this study, we used probability generating function and supplementary variable technique to solve the Laplace transforms of time-dependent probabilities of system states. Finally, we evaluated performance measures and expressed the results in numerical values.

Keywords: Batch arrivals; Priority queues; Differentiate Breakdown; Optional Vacation; Synchronized Reneging

AMS Subject Classification (2010): 60K25, 68M30, 90B22.

1. INTRODUCTION

A significant feature of a queuing situation (e.g., telecommunication system) is that when all servers are busy, an arriving customer is forced to exit the area of service and return to the retrial group after a particular period. These scenarios can be overcome by using retrial queues; for example if a server is found unavailable by the customers who arrives, then they will join the orbit to try their requests in random orders and at random moments. In real world, retrial queues are widely utilized as models for stochastic phenomena such as telecommunication networks, telephone switching systems, and computer systems so as to gain service from central processing unit. Recently, the literature on retrial queues has grown rapidly. Many researchers have investigated single-server retrial queues with two classes of customers. Wua and Lian [16] analyzed an $M^{[1]}, M^{[2]}/G/1$ G-queueing system with retrial customers and a server subject to breakdown and repair. Choudhury et al. [10] extensively analyzed an $M^{[X]}/G/1$ retrial queue with service interruption and optional service. Ammar and Rajadurai [2] studied a preemptive priority queueing system with disaster and the server working at the lower speed. Ayyappan and Udayageetha [4] discussed a priority retrial queueing system with collisions, working

breakdown, reneging, two-way communication and immediate feedback . Arivudainambi and Godhandaraman [3] investigated single-server retrial queueing system with second optional service, balking and single vacation.

In the history of queueing analysis, priority queueing system has gained crucial attention. Preemptive and non-preemptive disciplines are the two types of priority disciplines. Priority customers, in case of non-preemptive discipline, have to wait until the service to ordinary customers completed. However, in case of preemptive discipline, priority customers will always interrupt the service provided to ordinary customers. Dhas et al. [11] described a preemptive priority queue with general bulk service and heterogeneous arrivals. Brandwajn and Begin [8] examined preemptive priority system with general inter arrival and service times. Kim et al. [17] explained a non-preemptive priority queue with two classes of customers and multiple vacations. Krishnamoorthy et al. [18] analyzed non-preemptive priority queue with priorities and service time that are self generated and follows PH-distribution. Dimitriou [12] investigated a retrial queue with mixed priorities, unreliable server, negative arrivals and multiple vacations.

A server in a queueing system would be unavailable for some time due to different reasons, for example, being under maintenance work, busy at another queue, or simply taking a break. Baruah et al. [7] explained a two-stage of service with reneging during breakdown and vacation periods. Choudhury and Kalita [9] analyzed the steady-state behaviour of a $M/G/1$ queue with optional repeated service and two types of general heterogeneous service subject to server breakdown that randomly occurs at any point in time while serving the customers and during delayed repair. Maragathasundari and Srinivasan [20] investigated a non-Markovian queueing model with multistage of services, in which the authors considered reneging to prevail in case of unavailability of the server during system breakdown or vacation periods. Jain et al. [14] studied a general retrial $M^{[X]}/G/1$ queue with Bernoulli vacation and second optional service. In this model, breakdowns are observed at random intervals at any point in time during the provision of service to the customers. Janani [15] described transient analysis of single-server queueing model with differentiated breakdown.

The server is assumed to take vacations at random intervals. However, even if the system has a single priority customer, no vacation is allowed. Therefore, the server may take an optional random length vacation after the last service of the last priority customer is served. Gupta et al. [22] generalized impatient customers in queueing system with optional vacation policies. Madan and Rawwash [19] determined $M^{[X]}/G/1$ queueing system with feedback and optional server vacations based on single vacation. Laxmi et al. [24] investigated $M/M/1$ queueing system with second optional service, correlated reneging and working vacations. Ayyappan et al. [5] analysed $M^{[X]}/G/1$ queueing system with optional server vacation and two phases of service. Ordinary customers become impatient if the server is busy or unavailable. However, they execute synchronized abandonments motivated by remote systems. Adan et al. [1] have performed a detailed analysis of queueing models with impatient customers and vacations. Economou and Kapodistria [13] explained single-server queueing system with synchronized reneging customers. A single server queueing model with reneging, feedback and balking was examined by Rakesh Kumar and Soodam [23]

Two different types of customers, priority and ordinary customers are to be considered in this work with differentiated breakdown, repair, synchronized reneging optional vacation and followed by non preemptive priority discipline. The server provide service to the customer but randomly fails. Hard failure is defined as an equipment failure that usually necessitates the physical presence of a repairman with specialized knowledge, which is a time-consuming process. Soft failure, on the other hand, is defined as failure caused by events rather than physical condition and usually resolved rebooting the system. Ordinary customers may go back from the orbit if the server is unavailable. Furthermore, the server becomes idle or goes for a vacation on completing the service of all priority customers.

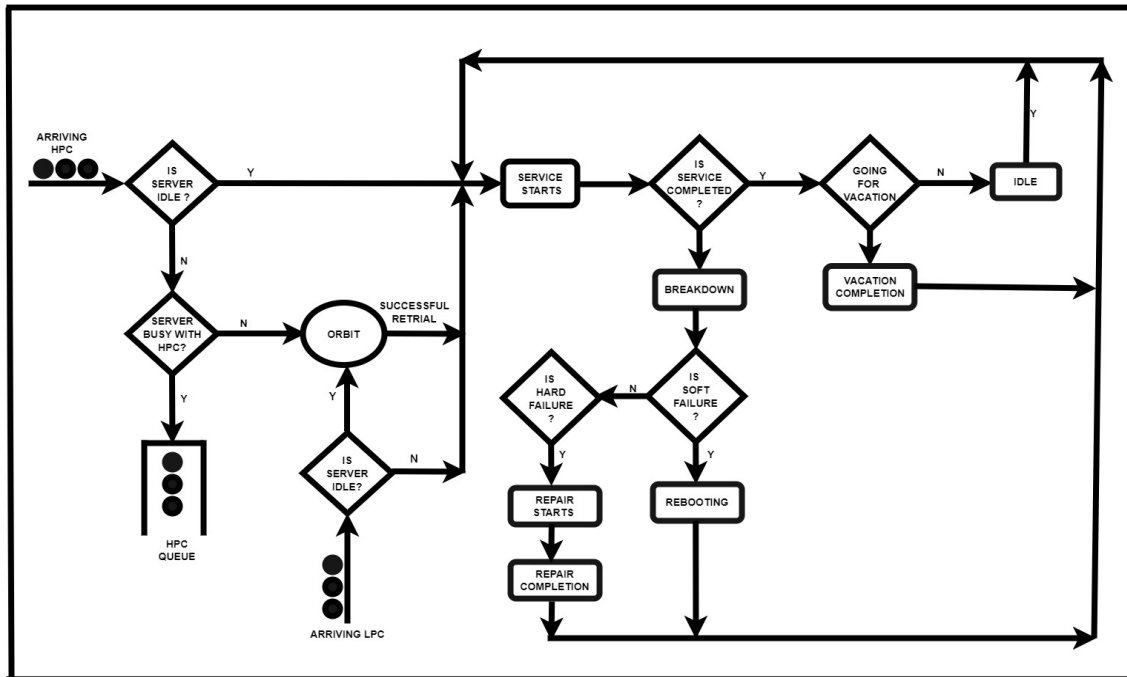


Figure 1: Schematic representation

This article is organized as follows. Mathematical model is described in Section 2 and queue size distribution is analyzed in Section 3. An explicit expression for governing equation is enlisted in Section 4. Steady state analysis is discussed in Section 5. Particular cases are obtained in Section 6. The effect of system performance measures is illustrated in Section 7. Numerical and graphical results are derived and conclusion is obtained in Section 8 and 9.

2. DESCRIPTION OF THE MODEL

- **Arrival Process :** Two different types of units arrive in batches with independent Poisson compound process. Let $\lambda_1, \lambda_2 > 0$ be the arrival rate for priority and ordinary customers, respectively. Assume that the first order probabilities for priority and ordinary units $\lambda_1 c_i dt$ ($i = 1, 2, 3, \dots$) and $\lambda_2 c_j dt$ ($j = 1, 2, 3, \dots$) with batch size i and j units arrive at the system during a short interval of time $(t, t + dt)$. Here, $0 \leq c_i \leq 1, \sum_{i=1}^{\infty} c_i = 1, 0 \leq c_j \leq 1, \sum_{j=1}^{\infty} c_j = 1$.
- **Retrial Service Process :** Ordinary customers are known as retrial customers. These customers will go back to the orbit and will request repeatedly for their service after some time if the server is busy or unavailable.
- **Regular Service Process :** Ordinary and priority customers ordinate in batches with distinct queues. Service rate follows general distribution and server renders single service for priority customers and ordinary customers with service rate $\mu_i(v), i = 1, 2$ respectively. The service for ordinary customers starts when the priority queue is empty.
- **Differentiate Breakdown and repair :** The rates of hard and soft failure are exponentially distributed with rate α_1 & α_2 respectively. For soft failure, the repair time follows exponential distribution with rate η_1 and for hard failure, the repair time follows general distribution with rate $\eta_2(v)$.
- **synchronized Reneging :** If the server is not available in the system, ordinary customer either exit the orbit with probability ζ or join the orbit with probability $1 - \zeta$.
- **Vacation:** The server may take a vacation with probability θ or it may remain idle with probability $1 - \theta$ after serving all priority customers. The random variable for vacation time

V , with rate $\gamma(v)$ follows general distribution.

3. ANALYSIS OF QUEUE SIZE DISTRIBUTION

This section deals with the derivation of governing equations. On account of non-Markovian queueing system, probability generating function and supplementary variable have been used to solve this model.

Let,

$N_1(t)$ = Number of priority customers in the queue at time t ,

$N_2(t)$ = Number of ordinary customers in the queue at time t ,

$Y(t)$ = State of the server at time t .

Here $M^0(t)$, $B_i^0(t)$ for $i = 1, 2$, $V^0(t)$, $R^0(t)$, indicates elapsed service time for retrial, service for priority and ordinary customers, vacation and repair at time t .

To obtain a bivariate Markov process $\{N_1(t), N_2(t), Y(t), t > 0\}$, $Y(t)$ denotes the server state. Here $Y(t) = (0, 1, 2, 3, 4, 5)$, which mean as follows: 0, the server is idle; 1, server is in retrial state; 2, busy with priority customers; 3, busy with ordinary customers; 4, on vacation and 5, repair.

Let us assume that, $M(0) = 0$, $M(\infty) = 1$, $B_i(0) = 0$, $B_i(\infty) = 1$, $V(0) = 0$, $V(\infty) = 1$ and $R^{(2)}(0) = 0$, $R^{(2)}(\infty) = 1$ be continuous at $v = 0$ for $i = 1, 2$.

The functions $\beta(v)$, $\mu_1(v)$, $\mu_2(v)$, $\gamma(v)$ and $\eta_2(v)$ be the hazard rate for retrial, priority and ordinary customers service rate, vacation and repair.

$$\beta(v) = \frac{dM(v)}{1 - M(v)}; \quad \mu_i(v) = \frac{dB_i(v)}{1 - B_i(v)}, i = 1, 2 \quad \gamma(v) = \frac{dV(v)}{1 - V(v)}; \quad \eta_2(v) = \frac{dR^{(2)}(v)}{1 - R^{(2)}(v)}.$$

The probability $I_{0,n}(v, t) = Pr\{N_1(t) = 0, N_2(t) = n, Y(t) = 0\}$ and probability densities are as follows:

$$\begin{aligned} I_{0,n}(v, t)dv &= Pr\{N_1(t) = 0, N_2(t) = n, Y(t) = 1; v \leq I^0(t) \leq v + dv\}, n \geq 1 \\ P_{m,n}(v, t)dv &= Pr\{N_1(t) = m, N_2(t) = n, Y(t) = 2; v \leq B_1^0(t) \leq v + dv\}, \\ Q_{m,n}(v, t)dv &= Pr\{N_1(t) = m, N_2(t) = n, Y(t) = 3; v \leq B_2^0(t) \leq v + dv\}, \\ V_{m,n}(v, t)dv &= Pr\{N_1(t) = m, N_2(t) = n, Y(t) = 4; v \leq V^0(t) \leq v + dv\}, \\ R_{m,n}(v, t)dv &= Pr\{N_1(t) = m, N_2(t) = n, Y(t) = 5; v \leq R^0(t) \leq v + dv\}, \end{aligned}$$

for $v \geq 0, t \geq 0, m \geq 0$ and $n \geq 0$.

4. EQUATION GOVERNING THE SYSTEM

$$\begin{aligned} \frac{d}{dt} I_{0,0}(t) &= -(\lambda_1 + \lambda_2)I_{0,0}(t) + (1 - \theta) \int_0^\infty P_{0,0}(v, t)\mu_1(v)dv \\ &+ (1 - \theta) \int_0^\infty Q_{0,0}(v, t)\mu_2(v)dv + \int_0^\infty R_{0,0}^{(2)}(v, t)\eta_2(v)dv \\ &+ R_{0,0}^{(1)}(t)\eta_1 + \int_0^\infty V_{0,0}(v, t)\gamma(v)dv. \end{aligned} \tag{1}$$

$$\frac{\partial}{\partial t} I_{0,n}(v, t) + \frac{\partial}{\partial v} I_{0,n}(v, t) = -(\lambda_1 + \lambda_2 + \beta(v))I_{0,n}(v, t) \quad \text{for } n \geq 1. \tag{2}$$

$$\begin{aligned} \frac{\partial}{\partial t} P_{m,n}(v, t) + \frac{\partial}{\partial v} P_{m,n}(v, t) = & -(\lambda_1 + \lambda_2 + \alpha_1 + \alpha_2 + \mu_1(v))P_{m,n}(v, t) \\ & + \lambda_1(1 - \delta_{0m}) \sum_{i=1}^m c_i P_{m-i,n}(v, t) \\ & + \lambda_2(1 - \delta_{0n}) \sum_{j=1}^n c_j P_{m,n-j}(v, t) \quad \text{for } m, n \geq 1. \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\partial}{\partial t} Q_{m,n}(v, t) + \frac{\partial}{\partial v} Q_{m,n}(v, t) = & -(\lambda_1 + \lambda_2 + \alpha_1 + \alpha_2 + \mu_2(v))Q_{m,n}(v, t) \\ & + \lambda_1(1 - \delta_{0m}) \sum_{i=1}^m c_i Q_{m-i,n}(v, t) \\ & + \lambda_2(1 - \delta_{0n}) \sum_{j=1}^n c_j Q_{m,n-i}(v, t) \quad \text{for } m, n \geq 1. \end{aligned} \quad (4)$$

$$\begin{aligned} \frac{\partial}{\partial t} V_{m,n}(v, t) + \frac{\partial}{\partial v} V_{m,n}(v, t) = & -(\lambda_1 + \lambda_2 + \zeta + \gamma(v))V_{m,n}(v, t) + \zeta V_{m,n+1}(v, t) \\ & + \lambda_1(1 - \delta_{0m}) \sum_{i=1}^m c_i V_{m-i,n}(v, t) \\ & + \lambda_2(1 - \delta_{0n}) \sum_{j=1}^n c_j V_{m,n-i}(v, t) \quad \text{for } m, n \geq 1. \end{aligned} \quad (5)$$

$$\begin{aligned} \frac{d}{dt} R_{m,n}^{(1)}(v, t) + \frac{d}{dv} R_{m,n}^{(1)}(v, t) = & -(\lambda_1 + \lambda_2 + \zeta + \eta_1)R_{m,n}^{(1)}(t) + \zeta R_{m,n+1}^{(1)}(t) \\ & + \alpha_1 \int_0^\infty (P_{m,n}(v, t) + Q_{m,n}(v, t)) dv \\ & + \lambda_1(1 - \delta_{0m}) \sum_{i=1}^m c_i R_{m-i,n}^{(1)}(t) \\ & + \lambda_2(1 - \delta_{0n}) \sum_{j=1}^n c_j R_{m,n-i}^{(1)}(t) \quad \text{for } m, n \geq 1. \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{\partial}{\partial t} R_{m,n}^{(2)}(v, t) + \frac{\partial}{\partial v} R_{m,n}^{(2)}(v, t) = & -(\lambda_1 + \lambda_2 + \zeta + \eta_2(v))R_{m,n}^{(2)}(v, t) + \zeta R_{m,n+1}^{(2)}(v, t) \\ & + \lambda_1(1 - \delta_{0m}) \sum_{i=1}^m c_i R_{m-i,n}^{(2)}(v, t) \\ & + \lambda_2(1 - \delta_{0n}) \sum_{j=1}^n c_j R_{m,n-i}^{(2)}(v, t) \quad \text{for } m, n \geq 1. \end{aligned} \quad (7)$$

Define, the boundary conditions at $v = 0$

$$\begin{aligned} P_{m,n}(0, t) = & \int_0^\infty P_{m+1,n}(v, t) \mu_1(v) dv + \int_0^\infty Q_{m+1,n}(v, t) \mu_2(v) dv \\ & + \int_0^\infty R_{m+1,n}^{(2)}(v, t) \eta(v) dv + R_{m+1,n}^{(1)}(t) \eta_1 + \lambda_1 c_{m+1} I_{0,n}(t), \end{aligned} \quad (8)$$

$$Q_{0,0}(0, t) = \lambda_2 c_1 I_{0,0}(t) + \int_0^\infty I_{0,1}(v, t) \beta(v) dv \quad (9)$$

$$\begin{aligned} Q_{0,n}(0, t) = & \lambda_2 c_{n+1} I_{0,0}(t) + \int_0^\infty I_{0,n+1}(v, t) \beta(v) dv + \sum_{i=1}^n \lambda_2 C_i(v, t) \\ & + \int_0^\infty I_{0,n+1-i}(v, t) dv \quad \text{for } n \geq 1. \end{aligned} \quad (10)$$

$$V_{0,n}(0, t) = \theta \int_0^\infty P_{0,n}(v, t) \mu_1(v) dv + \theta \int_0^\infty Q_{0,n}(v, t) \mu_2(v) dv, \quad \text{for } n \geq 0. \quad (11)$$

$$R_{m,n}^{(2)}(0, t) = \alpha_2 \int_0^\infty P_{m-1,n}(v, t) \mu_1(v) dv + \alpha_2 \int_0^\infty Q_{m,n}(v, t) \mu_2(v) dv, \quad \text{for } m, n \geq 0. \quad (12)$$

$$I_{0,n}(0, t) = (1 - \theta) \int_0^\infty P_{0,n}(v, t) \mu_1(v) dv + \int_0^\infty R_{0,n}^{(2)}(v, t) \eta_2(v) dv + R_{0,n}^{(1)}(t) \eta_1 \\ + (1 - \theta) \int_0^\infty Q_{0,n}(v, t) \mu_2(v) dv + \int_0^\infty V_{0,n}(v, t) \gamma(v) dv, \quad (13)$$

$$P_{m,n}(0) = Q_{m,n}(0) = R_{m,n}^{(1)}(0) = R_{m,n}^{(2)}(0) = V_{m,n}(0) = 0, \quad \text{for } m, n \geq 0 \text{ and } I_{0,0} = 1, \\ I_{0,n}(0) = 0, \quad \text{for } n \geq 1 \text{ are the initial conditions.} \quad (14)$$

Now, we define the Probability Generating Function (PGF),

$$I(v, t, z_0) = \sum_{n=1}^\infty z_0^n I_{0,n}(v, t); \quad A(v, t, z_1, z_2) = \sum_{m=0}^\infty \sum_{n=0}^\infty z_1^m z_2^n A_{m,n}(v, t); \\ A(v, t, z_1) = \sum_{m=0}^\infty z_1^m A_m(v, t); \quad A(v, t, z_2) = \sum_{n=0}^\infty z_2^n A_n(v, t); \quad (15)$$

here $A = P, Q, V, R^{(1)}, R^{(2)}$.

By applying Laplace transforms to equations (1) to (13) and by using (14) and (15), we obtain the following equations:

$$\bar{I}_0(v, s, z_2) = \bar{I}_0(0, s, z_2) e^{-(s+\lambda_1+\lambda_2)v - \int_0^v \beta(t) dt}, \quad (16)$$

$$\bar{P}(v, s, z_1, z_2) = \bar{P}(0, s, z_1, z_2) e^{-\phi_1(s, z) v - \int_0^v \mu_1(t) dt}, \quad (17)$$

$$\bar{Q}(v, s, z_1, z_2) = \bar{Q}(0, s, z_1, z_2) e^{-\phi_1(s, z) v - \int_0^v \mu_2(t) dt}, \quad (18)$$

$$\bar{V}(v, s, z_1, z_2) = \bar{V}(0, s, z_1, z_2) e^{-\phi_2(s, z) v - \int_0^v \gamma(t) dt}, \quad (19)$$

$$\bar{R}^{(2)}(v, s, z_1, z_2) = \bar{R}^{(2)}(0, s, z_1, z_2) e^{-\phi_2(s, z) v - \int_0^v \eta_2(t) dt}. \quad (20)$$

where,

$$\phi_1(s, z) = s + \lambda_1(1 - C(z_1)) + \lambda_2(1 - C(z_2)) + \alpha_1 + \alpha_2, \quad (21)$$

$$\phi_2(s, z) = s + \lambda_1(1 - C(z_1)) + \lambda_2(1 - C(z_2)). \quad (22)$$

$$\bar{Q}(0, s, z_2) = \frac{\left\{ \begin{aligned} &1 - (s + \lambda_1 + \lambda_2) \bar{I}_{0,0}(s) E_1(s, z) + C(z_2) \lambda_2 \bar{I}_{0,0}(s) \\ &\frac{1}{z_2} [A(s, z) D(s, z) + B(s, z) E_1(s, z)] F(s, z) + C(z_2) \lambda_2 \bar{I}_{0,0}(s) \end{aligned} \right\}}{\left\{ \begin{aligned} &z_2 [E_1(s, z) - A(s, z) C(g(z_2)) \lambda_1 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right]] \\ &- [A(s, z) D(s, z) + B(s, z) E_1(s, z)] F(s, z) \end{aligned} \right\}} \quad (23)$$

$$\bar{I}_0(0, s, z_2) = \frac{\left\{ \begin{aligned} &1 - (s + \lambda_1 + \lambda_2) \bar{I}_{0,0}(s) E_1(s, z) + C(z_2) \lambda_2 \bar{I}_{0,0}(s) \\ &\frac{1}{z_2} [A(s, z) D(s, z) + B(s, z) E_1(s, z)] \end{aligned} \right\}}{\left\{ \begin{aligned} &z_2 [E_1(s, z) - A(s, z) C(g(z_2)) \lambda_1 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right]] \\ &- [A(s, z) D(s, z) + B(s, z) E_1(s, z)] \end{aligned} \right\}} \quad (24)$$

$$\bar{P}(0, s, z_1, z_2) = \frac{\left\{ \begin{aligned} &\lambda_1 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right] \bar{I}_0(0, s, z_2) [E_1(s, z) C(z_1) - C(g(z_2)) E(s, z)] \\ &\bar{Q}(0, s, z_2) [E_1(s, z) (d_1(s, z) - d_2(s, z)) - E(s, z) (d_1^*(s, z) - d_2^*(s, z))] \end{aligned} \right\}}{\left\{ (z_1 - C(s, z)) E_1(s, z) \right\}} \quad (25)$$

$$\bar{V}_0(0, s, z_2) = \frac{\left\{ \begin{aligned} &\theta \bar{B}_1 \psi_1(s, z) C(g(z_2)) \lambda_1 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right] \bar{I}_0(0, s, z_2) \\ &+ \theta [E_1(s, z) \bar{B}_2 \psi_1(s, z) + (d_1^*(s, z) - d_2^*(s, z)) \bar{B}_1 \psi_1(s, z)] \bar{Q}(0, s, z_2) \end{aligned} \right\}}{\{E_1(s, z)\}} \quad (26)$$

$$\bar{R}^{(1)}(0, s, z_1, z_2) = \frac{\alpha_1}{\phi(s, z)} \left[\bar{P}(0, s, z_1, z_2) \left[\frac{1 - \bar{B}_1(\phi_1(s, z))}{\phi_1(s, z)} \right] + \bar{Q}(0, s, z_2) \left[\frac{1 - \bar{B}_2(\phi_1(s, z))}{\phi_1(s, z)} \right] \right] \quad (27)$$

$$\bar{R}^{(2)}(0, s, z_1, z_2) = \alpha_2 z_1 \bar{P}(0, s, z_1, z_2) \left[\frac{1 - \bar{B}_1(\phi_1(s, z))}{\phi_1(s, z)} \right] + \alpha_2 \bar{Q}(0, s, z_2) \left[\frac{1 - \bar{B}_2(\phi_1(s, z))}{\phi_1(s, z)} \right]. \quad (28)$$

Theorem.1 When the system is in regular service, breakdown, repair and vacation by using the Laplace transforms the probability generating function of the number of customers in the respective queue is given by

$$\bar{I}_0(s, z_2) = \bar{I}_0(0, s, z_2) \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right], \quad (29)$$

$$\bar{P}(s, z_1, z_2) = \bar{P}(0, s, z_1, z_2) \left[\frac{1 - \bar{B}_1(\phi_1(s, z))}{\phi_1(s, z)} \right], \quad (30)$$

$$\bar{Q}(s, z_1, z_2) = \bar{Q}(0, s, z_2) \left[\frac{1 - \bar{B}_2(\phi_1(s, z))}{\phi_1(s, z)} \right], \quad (31)$$

$$\bar{V}(s, z_1, z_2) = \bar{V}(0, s, z_1, z_2) \left[\frac{1 - \bar{V}(\phi_2(s, z))}{\phi_2(s, z)} \right], \quad (32)$$

$$\bar{R}^{(2)}(s, z_1, z_2) = \bar{R}^{(2)}(0, s, z_1, z_2) \left[\frac{1 - \bar{R}^{(2)}(\phi_2(s, z))}{\phi_2(s, z)} \right]. \quad (33)$$

Proof: Integrating the preceding equations (29) to (33) with respect to v and applying the solution of renewal theory we obtain the following

$$\int_0^\infty [1 - H(v)] e^{-sv} dv = \frac{1 - \bar{h}(s)}{s}. \quad (34)$$

Here, the LST of the distribution function of a random variable $H(v)$ is denoted as $\bar{h}(s)$. The absolute outcomes of the probability generating functions for the successive states, $\bar{P}(s, z_1, z_2)$, $\bar{Q}(s, z_1, z_2)$, $\bar{V}(s, z_1, z_2)$, and $\bar{R}^{(2)}(s, z_1, z_2)$ are obtained by using equation (29) to (33).

5. STEADY STATE ANALYSIS

According to Tauberian property,

$$\lim_{s \rightarrow 0} s \bar{f}(s) = \lim_{t \rightarrow \infty} f(t).$$

Despite of the state of the system, the probability generating function of the queue size is as follows:

$$W_q(z_1, z_2) = \frac{Nr(z_1, z_2)}{Dr(z_1, z_2)}, \quad (35)$$

where

$$\begin{aligned} Nr(z_1, z_2) = & N_2(z)D_3(z)\phi_1(z) \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right] \left[\phi_2 + \frac{\theta \bar{B}_1 \psi_1(z) C(g(z_2)) \lambda_1 (1 - \bar{V} \phi_2)}{E_1} \right] \\ & + N_4 D_2 (1 - \bar{B}_1 \phi_1(z)) \left[\left(1 + \frac{\alpha_1}{\phi(z)} \right) \phi_2 + \alpha_2 z_1 (1 - \bar{R}^{(2)} \phi_2(z)) \right] \\ & + N_3 D_3 \left[\left(1 + \frac{\alpha_1}{\phi(z)} \right) \phi_2 + \alpha_2 (1 - \bar{R}^2 \phi_2(z)) \right] (1 - \bar{B}_2 \phi_1(z)) \\ & + (1 - \bar{V} \phi_2) \left[\frac{\theta E_1(z) \bar{B}_2 \psi_1(z) + \bar{B}_1 \psi_1(z) (d_1^*(z) - d_2^*(z))}{E_1} \right] \end{aligned}$$

$$Dr(z_1, z_2) = D_2(z)D_3(z)\phi_1(z)\phi_2(z),$$

where,

$$N_2 = z_2(-\lambda_1 + \lambda_2) \bar{I}_{0,0} E_1(s, z) + C(z_2) \lambda_2 \bar{I}_{0,0} [A(s, z)(d_1^*(s, z) - d_2^*(s, z)) + B(s, z)E_1(s, z)]$$

$$\begin{aligned} N_3 = & [(-\lambda_1 + \lambda_2) \bar{I}_{0,0} E_1(s, z) + \frac{1}{z_2} C(z_2) \lambda_2 \bar{I}_{0,0} [A(s, z)(d_1^*(s, z) - d_2^*(s, z)) \\ & + B(s, z)E_1(s, z)]] F(s, z) + \frac{1}{z_2} C(z_2) \lambda_2 \bar{I}_{0,0} \end{aligned}$$

$$\begin{aligned} N_4 = & N_2 \lambda_1 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right] [E_1(s, z)C(z_1) - C(g(z_2))E(s, z)] \\ & + N_3 [E_1(s, z)(d_1(s, z) - d_2(s, z)) - E(s, z)(d_1^*(s, z) - d_2^*(s, z))] \end{aligned}$$

$$\begin{aligned} D_2 = & z_2 [E_1(s, z) - A(s, z)C(g(z_2))\lambda_1 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right] \\ & - [A(s, z)(d_1^*(s, z) - d_2^*(s, z)) + B(s, z)E_1(s, z)] F(s, z)] \end{aligned}$$

$$D_3 = (z_1 - C(s, z))E_1(s, z)D_2(s, z)$$

$$A(s, z) = (1 - \theta + \theta \bar{V} \psi_2(z)) \bar{B}_1 \psi_1(z) + \left[\frac{\alpha_1 \eta_1}{\psi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \psi_2(z) \right] \left[\frac{1 - \bar{B}_1 \psi_1(z)}{\psi_1(z)} \right]$$

$$B(s, z) = (1 - \theta + \theta \bar{V} \psi_2(z)) \bar{B}_2 \psi_1(z) + \left[\frac{\alpha_1 \eta_1}{\psi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \psi_2(z) \right] \left[\frac{1 - \bar{B}_2 \psi_1(z)}{\psi_1(z)} \right]$$

$$C(s, z) = \bar{B}_1 \phi_1(z) + \left[\frac{\alpha_1 \eta_1}{\phi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \phi_2(z) \right] \left[\frac{1 - \bar{B}_1 \phi_1(z)}{\phi_1(z)} \right]$$

$$d_1(s, z) = \bar{B}_2 \phi_1(z) + \left[\frac{\alpha_1 \eta_1}{\phi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \phi_2(z) \right] \left[\frac{1 - \bar{B}_2 \phi_1(z)}{\phi_1(z)} \right]$$

$$d_2(s, z) = (1 + \theta \bar{V} \psi_2(z) - \theta \bar{V} \phi_2(z)) \bar{B}_2 \psi_1(z) + \left[\frac{\alpha_1 \eta_1}{\psi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \psi_2(z) \right] \left[\frac{1 - \bar{B}_2 \psi_1(z)}{\psi_1(z)} \right]$$

$$d_1^*(s, z) = \bar{B}_2 \sigma_1(z) + \left[\frac{\alpha_1 \eta_1}{\sigma(z)} + \alpha_2 z_1 \bar{R}^{(2)} \sigma_2(z) \right] \left[\frac{1 - \bar{B}_2 \sigma_1(z)}{\sigma_1(z)} \right]$$

$$d_2^*(s, z) = (1 + \theta \bar{V} \psi_2(z) - \theta \bar{V} \sigma_2(z)) \bar{B}_2 \psi_1(z) + \left[\frac{\alpha_1 \eta_1}{\psi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \psi_2(z) \right] \left[\frac{1 - \bar{B}_2 \psi_1(z)}{\psi_1(z)} \right]$$

$$E(s, z) = (1 + \theta \bar{V} \psi_2(z) - \theta \bar{V} \phi_2(z)) \bar{B}_1 \psi_1(z) + \left[\frac{\alpha_1 \eta_1}{\psi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \psi_2(z) \right] \left[\frac{1 - \bar{B}_1 \psi_1(z)}{\psi_1(z)} \right]$$

$$E^*(s, z) = (1 + \theta \bar{V} \psi_2(z) - \theta \bar{V} \sigma_2(z)) \bar{B}_1 \psi_1(z) + \left[\frac{\alpha_1 \eta_1}{\psi(z)} + \alpha_2 z_1 \bar{R}^{(2)} \psi_2(z) \right] \left[\frac{1 - \bar{B}_1 \psi_1(z)}{\psi_1(z)} \right]$$

$$F(z) = \bar{M}(\lambda_1 + \lambda_2) + c(z_2)\lambda_2 \left[\frac{1 - \bar{M}(s + \lambda_1 + \lambda_2)}{s + \lambda_1 + \lambda_2} \right]$$

$$f_1(z_1, z_2) = \phi_2(z) + \frac{\theta \bar{B}_1 \psi_1(z) C(g(z_2)) \lambda_1 (1 - \bar{V} \phi_2(z))}{E^*(z_1, z_2)}$$

$$f_2(z_1, z_2) = \left(1 + \frac{\alpha_1}{\phi(z)} \phi_2(z) + \alpha_2 z_1 (1 - \bar{R}^{(2)} \phi_2) \right)$$

$$f_3(z_1, z_2) = \left[\left(1 + \frac{\alpha_1}{\phi(z)} \phi_2(z) + \alpha_2 z_1 (1 - \bar{R}^{(2)} \phi_2) \right) (1 - \bar{B}_2 \phi_1(z)) \right. \\ \left. + \left[\frac{\theta (E^*(z) \bar{B}_2 \psi_1(z) + \bar{B}_1 \psi_1(z) (d_1^*(z) - d_2^*(z)))}{E^*(z)} \right] (1 - \bar{V} \phi_2(z)) \right]$$

$$\phi(z) = \lambda_1 (1 - C(z_1)) + \lambda_2 (1 - C(z_2)) + \eta_1 + \xi \left(1 - \frac{1}{z_2} \right)$$

$$\sigma(z) = \lambda_1 (1 - C(g(z_2))) + \lambda_2 (1 - C(z_2)) + \eta_1 + \xi \left(1 - \frac{1}{z_2} \right)$$

$$\sigma_1(z) = \lambda_1 (1 - C(g(z_2))) + \lambda_2 (1 - C(z_2)) + \alpha_1 + \alpha_2$$

$$\sigma_2(z) = \lambda_1 (1 - C(g(z_2))) + \lambda_2 (1 - C(z_2)) + \xi \left(1 - \frac{1}{z_2} \right)$$

$$\psi(z) = \lambda_1 + \lambda_2 (1 - C(z_2)) + \eta_1 + \xi \left(1 - \frac{1}{z_2} \right)$$

$$\psi_1(z) = \lambda_1 + \lambda_2 (1 - C(z_2)) + \alpha_1 + \alpha_2$$

$$\psi_2(z) = \lambda_1 + \lambda_2 (1 - C(z_2)) + \xi \left(1 - \frac{1}{z_2} \right).$$

Using normalization condition $W_q(1, 1) + I_{0,0} = 1$, we get

$$I_{0,0} = \frac{\left\{ \begin{aligned} & [D_2(1, 1) D_3'(1, 1) (\alpha_1 + \alpha_2) (\xi - (\lambda_1 + \lambda_2))] \\ & - [N_3(1, 1) D_3'(1, 1) f_3'(1, 1) + N_2'(1, 1) D_3'(1, 1) f_1'(1, 1) (\alpha_1 + \alpha_2) \end{aligned} \right\}}{\left\{ D_2(1, 1) D_3'(1, 1) (\alpha_1 + \alpha_2) (\xi - (\lambda_1 + \lambda_2)) \right\}} \quad (36)$$

and the utilization factor is given by

$$\rho = \frac{\left\{ \begin{aligned} & [N_3(1, 1) D_3'(1, 1) f_3'(1, 1) + N_2'(1, 1) D_3'(1, 1) f_1'(1, 1) (\alpha_1 + \alpha_2) \\ & \left[\frac{1 - \bar{M}(\lambda_1 + \lambda_2)}{\lambda_1 + \lambda_2} \right] + N_4'(1, 1) D_2(1, 1) f_2'(1, 1) (1 - B_1 (\alpha_1 + \alpha_2)) \end{aligned} \right\}}{\left\{ D_2(1, 1) D_3'(1, 1) (\alpha_1 + \alpha_2) (\xi - (\lambda_1 + \lambda_2)) \right\}}. \quad (37)$$

The stability condition for the model under which steady state exists is $\rho < 1$.

6. PERFORMANCE MEASURES

The expected queue size for priority customer is as follows:

$$L_{q1} = \frac{d}{dz_1} W_q(z_1, 1) |_{z_1=1} \quad (38)$$

The expected orbit size for ordinary customer is as follows:

$$L_{q2} = \frac{d}{dz_2} W_q(1, z_2) |_{z_2=1} \quad (39)$$

then

$$L_{q_1} = \frac{Dr''(1)Nr'''(1) - Dr'''(1)Nr''(1)}{3(Dr''(1))^2}, \quad (40)$$

$$L_{q_2} = \frac{Dr'''(1)Nr^{(iv)}(1) - Dr^{(iv)}(1)Nr'''(1)}{4(Dr'''(1))^2}. \quad (41)$$

The expected waiting time for priority queue is as follows:

$$W_{q_1} = \frac{L_{q_1}}{\lambda_1} \quad (42)$$

The expected waiting time for orbit is as follows:

$$W_{q_2} = \frac{L_{q_2}}{\lambda_2}. \quad (43)$$

7. PARTICULAR CASES

Case 1:

In the absence of priority queue, when there is no breakdown, no renegeing, no vacation and no retrial then the above model becomes

$$W_q(z) = \frac{-(1 - \bar{B}_2)(\lambda - \lambda C(z_2))I_{0,0}}{z_2 - \bar{B}_2(\lambda_2 - \lambda_2 C(z_2))}$$

which is the PGF of Medhi [21].

Case 2:

In the absence of priority queue, when there is no breakdown, no renegeing and no retrial then the above model becomes

$$W_q(z) = \frac{\left\{ \begin{aligned} &[-\phi_1(z, s)I_{0,0}(1 - \theta\bar{V}(\phi_1(z, s)) + \theta\bar{V}(\psi_1(z, s)))(1 - \bar{B}_1(\phi(z, s))) \\ &+ \theta\lambda_1 C(g(z_2))I_{0,0}1 - \theta\bar{V}(\phi_1(z, s))z_1 - \bar{B}_1(\phi(z, s))] \\ &+ [Q_0(0, z_2)[\theta(\bar{B}_2(\sigma(z, s)) - \bar{B}_2(\psi(z, s)))(1 - \bar{V}(\phi(z, s))) \\ &(z_1 - \bar{B}_1(\phi(z, s))) - (z_2 - \bar{B}_2(\phi_1(z, s)))(1 - \theta\bar{V}(\phi(z, s))) \\ &+ \theta\bar{V}(\psi_1(z, s)))(1 - \bar{B}_1(\phi(z, s))) + (z_1 - \bar{B}_1(\phi(z, s))) \\ &(1 - \bar{B}_2(\phi(z, s)))(1 - \bar{V}(\sigma(z, s))) + \theta\bar{V}(\psi(z, s))] \end{aligned} \right\}}{\{ [z_1 - \bar{B}_1(\phi(z, s))(1 - \theta\bar{V}(\sigma(z, s)) + \theta\bar{V}(\psi(z, s)))\phi(z, s)] \}}$$

which is the PGF of Ayyappan and Thamizhselvi [6].

8. NUMERICAL RESULTS

This section deals with the numerical and graphical studies of this model. We assume that the service time, breakdown, repair and vacation time are distributed exponentially.

Table 1: Effect of priority arrival rate (λ_1)

λ_1	$I_{0,0}$	ρ	L_{q_1}	W_{q_1}	L_{q_2}	W_{q_2}
0.1	0.2806	0.7194	0.2082	2.0816	0.2463	0.1231
0.2	0.2776	0.7224	0.4758	2.3789	0.3580	0.1790
0.3	0.2754	0.7246	0.7650	2.5501	0.4246	0.2123
0.4	0.2743	0.7257	1.0720	2.6801	0.4554	0.2277
0.5	0.2742	0.7258	1.3930	2.7859	0.4616	0.2308

Table 1 shows that when the arrival rate (λ_1) of priority customers for priority queue increases, then the probability of the server being idle decreases. However, average queue lengths busy period and average waiting time for customers in the queues all increase: we assume the values as $\lambda_2 = 2, \alpha_1 = 0.1, \alpha_2 = 0.1, \mu = 5, \eta_1 = 10, \eta_2 = 15, \gamma = 10, \theta = 0.9, \beta = 20, \xi = 0.9$ and $\lambda_1 = 0.1$ to 0.5.

Table 2: Effect of service rate (μ)

μ	$I_{0,0}$	ρ	L_{q_1}	W_{q_1}	L_{q_2}	W_{q_2}
5	0.3058	0.6942	0.4275	1.4250	0.6219	0.3110
6	0.4823	0.5177	0.4270	1.4233	0.5472	0.2736
7	0.5950	0.4050	0.4200	1.4001	0.4856	0.2428
8	0.6728	0.3272	0.4109	1.3695	0.4348	0.2174
9	0.7296	0.2704	0.4012	1.3374	0.3926	0.1963
10	0.7727	0.2273	0.3918	1.3060	0.3576	0.1788

Table 2 shows that when the service rate (μ) increases then the probability of the server being idle increases. However, average queue lengths, busy period, and average waiting time for customers in the queues all decrease: we assume the values as $\lambda_1 = 0.3, \lambda_2 = 2, \alpha_1 = 0.2, \alpha_2 = 0.5, \eta_2 = 10, \eta_1 = 15, \gamma = 10, \theta = 0.8, \beta = 20, \xi = 0.9$ and $\mu = 5$ to 10.

Table 3: Effect of ordinary customers arrival rate (λ_2)

λ_2	$I_{0,0}$	ρ	L_{q_1}	W_{q_1}	L_{q_2}	W_{q_2}
1.5	0.6275	0.3275	0.4029	0.8058	0.0409	0.0273
1.6	0.6406	0.3594	0.4347	0.8694	0.2939	0.1837
1.7	0.6073	0.3927	0.4663	0.9326	0.4609	0.2711
1.8	0.5725	0.4275	0.4974	0.9949	0.5752	0.3196
1.9	0.5361	0.4639	0.5279	1.0557	0.6556	0.3451
2.0	0.4981	0.5019	0.5573	1.1145	0.7133	0.3567

Table 3 shows that when the arrival rate (λ_2) of ordinary customers increases, then the probability server being idle decreases. Busy period, average queue lengths, and average waiting time for customers in the queues all increase: we assume the values as $\lambda_1 = 0.5, \mu = 6, \alpha_1 = 0.3, \alpha_2 = 0.5, \mu = 5, \eta_1 = 6, \eta_2 = 15, \gamma = 10, \theta = 0.8, \beta = 18, \xi = 0.9$ and $\lambda_2 = 1.5$ to 2.

In graphical representations, we assume that the service time, breakdown, repair, and vacation time are follows Erlang-2 distribution. The two-dimensional graphs are shown in Figure 2 - 4. Figure 2 exhibits that the expected length of the queue (L_{q_1}, L_{q_2}) rises, the expected length of the queue extends together with the priority arrival rate (λ_1). The behaviour of the queue sizes (L_{q_1}, L_{q_2}), which depends on the service rate (α), is shown in Figure 3, the length of the queue as the service rate rises. Figure 4 shows the expected queue length (L_{q_1}, L_{q_2}), which depends on the ordinary customers arrival rate (λ_2), the expected queue length grows together with the repair rate.

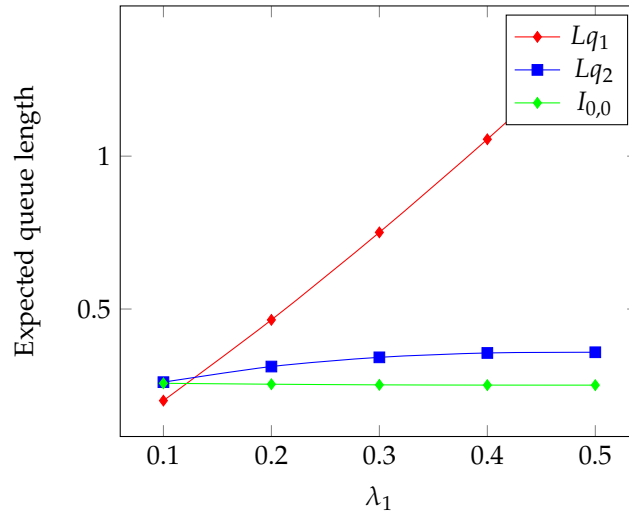


Figure 2: Expected queue length vs priority arrival rate λ_1

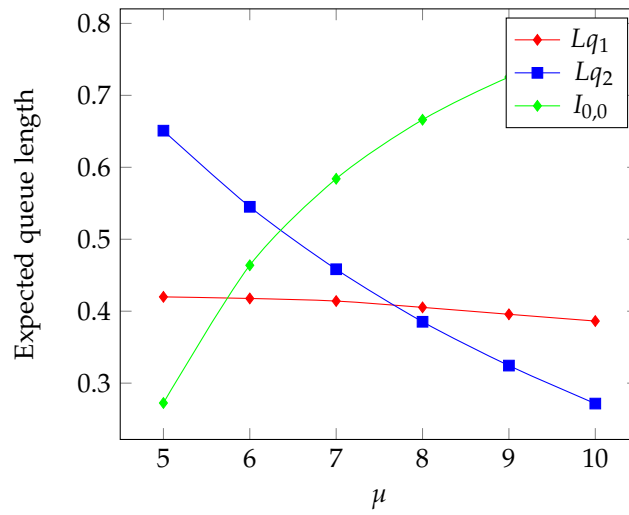


Figure 3: Expected queue length vs service rate μ

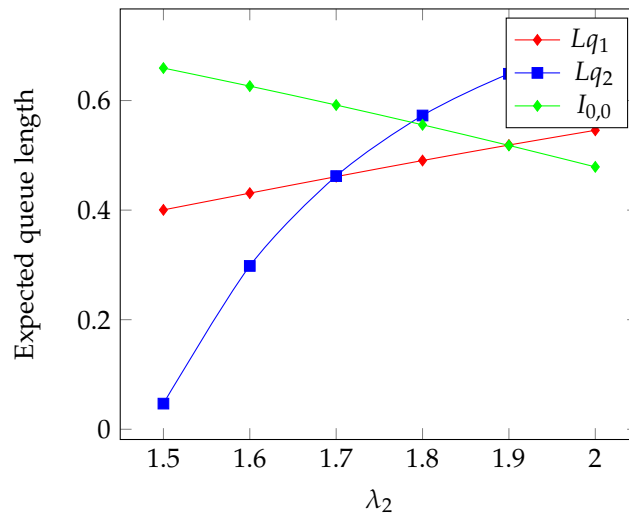


Figure 4: Expected queue length vs ordinary arrival rate λ_2

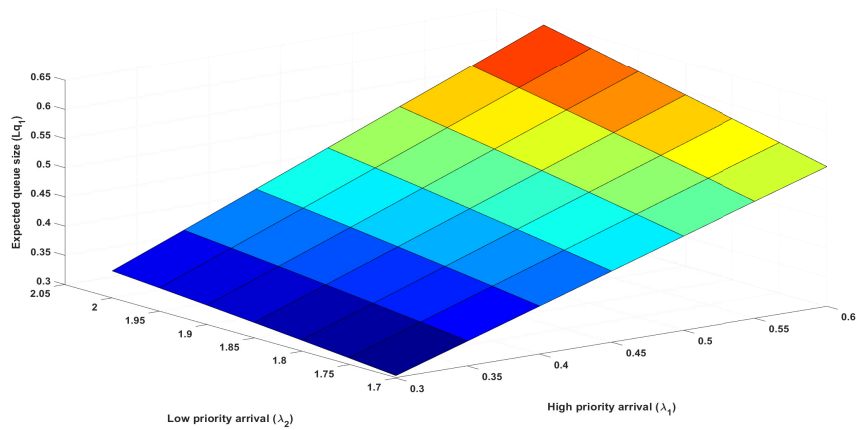


Figure 5: L_{q_1} Vs λ_1 and λ_2

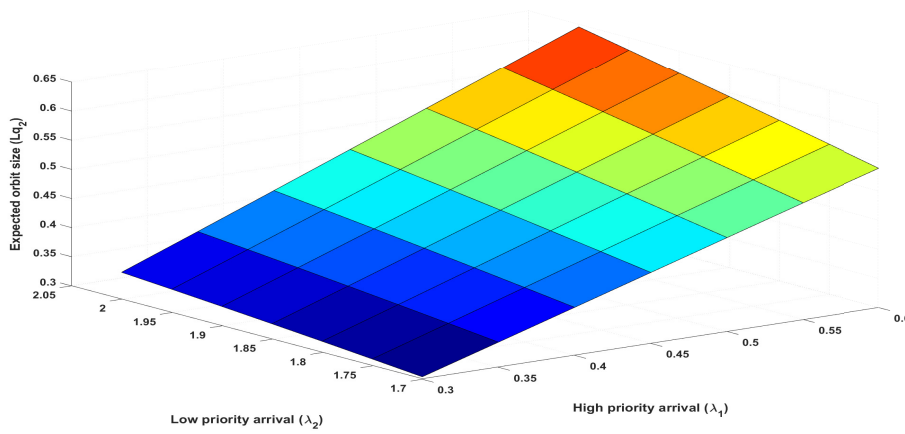


Figure 6: L_{q_2} Vs λ_1 and λ_2

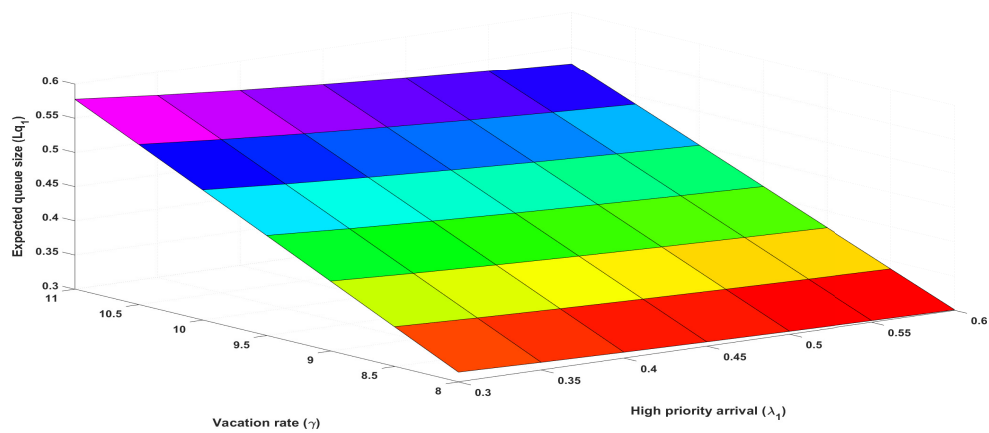


Figure 7: L_{q_1} Vs γ and λ_1

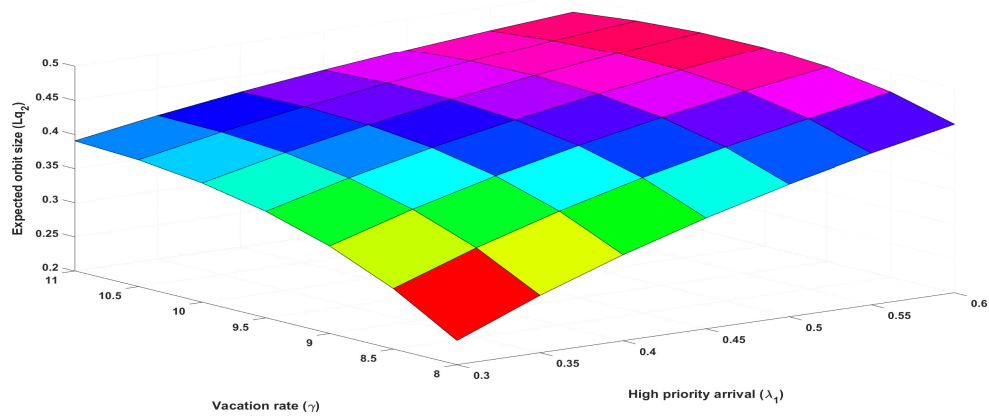


Figure 8: L_{q_2} Vs γ and λ_1

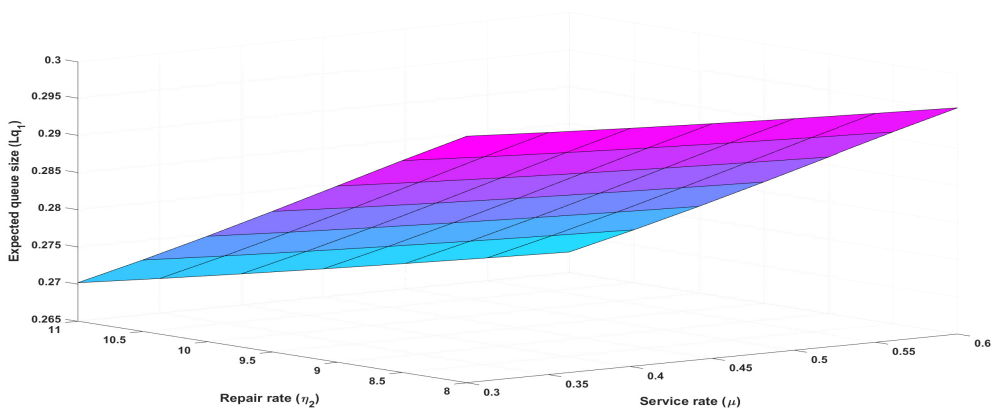


Figure 9: L_{q_2} Vs η_2 and μ

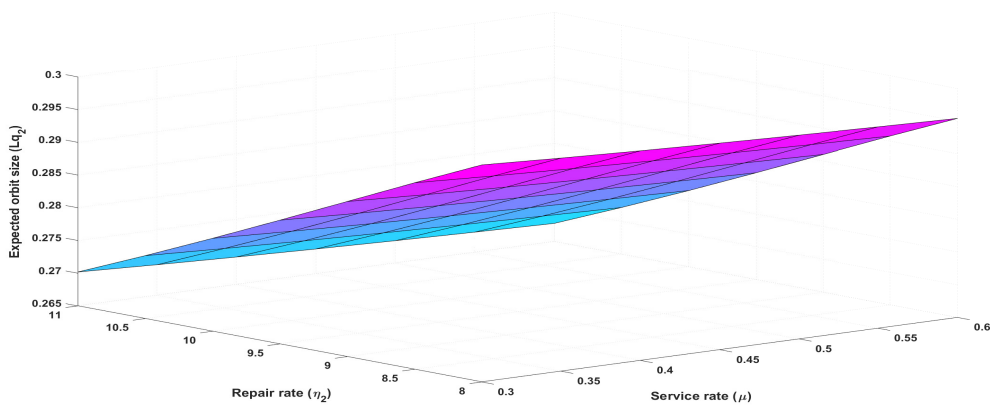


Figure 10: L_{q_2} Vs η_2 and μ

Graphs in three dimensions can be found in Figures 5 - 10. Figures 5 and 6 in the reference indicate that as the priority arrival rate (λ_1) and ordinary arrival rate (λ_2) increase, the expected queue size (Lq_1) and orbit size (Lq_2) increase as well. The behaviour of queue size (Lq_1) and orbit size (Lq_2) rises for increasing priority arrival rate (λ_1) and vacation rate (γ) as shown in Figure 7 and 8. The behaviour of queue size (Lq_1) and orbit size (Lq_2) increases with lowering repair rate (η_2) and service rate (μ) in Figure 9 and 10 .

9. CONCLUSION

In this study, we examined a non-preemptive priority retrial queue on a single server with distinct breakdown, repair, synchronised reneging, and optional server vacation. Random failures of servers are common. The breakdown of the model proposed in this study differs from traditional breakdown, which can be either a long breakdown (hard failure) or a short breakdown (soft failure). For example, if a computer system fails, a simple rebooting of the system (soft failure) or repair by a skilled personnel (hard failure) can fix the problem. Ordinary customers may exit the orbit if the server is either down or busy with priority queue. Furthermore, on completing the priority service, the server may become idle or may go for a vacation. Probability generating functions for the system size and its orbit were found by using supplementary variable technique. System characteristics such as steady-state probabilities and mean system size were also obtained. The results obtained analytically were confirmed with numerical illustrations. The model proposed in this study finds significant practical applications in computer processing systems.

REFERENCES

- [1] Adan, I., Economou, A. and Kapodistria, S. (2009). Synchronized reneging in queueing systems with vacations, *Queueing Syst, Queueing Syst*, 62:1-33.
- [2] Ammar, S.I. and Rajadurai, P. (2019). Performance Analysis of Preemptive Priority Retrial Queueing System with Disaster under Working Breakdown Services, *Symmetry*, 419(11):1-15.
- [3] Arivudainambi, D. and Godhandaraman, P. (2015). Retrial queueing system with balking, optional service and vacation, *Symmetry*, 229(5):67-84.
- [4] Ayyappan, G. and Udayageetha, J. (2018). Transient Analysis of $M^{[X_1]}, M^{[X_2]}/G_1, G_2/1$ retrial queueing system with priority services, Collisions, Orbital Search, Working Breakdown, Startup/Close down time, Feedback, modified Bernoulli vacation and Balking, *Symmetry*, 13(11):8783-8795.
- [5] Ayyappan, G., Muthu Ganapathi Subramanian, A. and Sathiya, K. (2013). $M^{[X]}/G/1$ Queue with Two Phase of Service and Optional Server Vacation, *International Journal of Computer Applications*, 66(6):4-10.
- [6] Ayyappan, G. and Thamizhselvi, P. (2016). riority queueing System with a Single Server Serving Two Queues $M^{[X_1]}, M^{[X_2]}/G_1, G_2/1$ with and Optional Server Vacation, *Applications of Applied Mathematics*, 11(1):61-82.
- [7] Baruah, M., Madan, K.C. and Eldabi, T. (2013). A Two Stage Batch Arrival Queue with Reneging during Vacation and Breakdown Periods, *American Journal of Operations Research*, 3(6):570-580.
- [8] Brandwajn, A. and Begin, T. (2017). Multi-server preemptive priority queue with general arrivals and service times, *European Journal of Operational Research*, 115(1):150-164.
- [9] Choudhury, G. and Kalita, C.R. (2017). An $M/G/1$ queue with two types of general heterogeneous service and optional repeated service subject to server breakdown and delayed repair, *Quality Technology & Quantitative Management*, 15(5):622-654.
- [10] Choudhury, G., Tadj, L. and Deka, K. (2010). A batch arrival retrial queueing system with two phases of service and service interruption, *Computers and Mathematics with Applications*, 126(59):437-450.

- [11] Dhas, D., Nadarajan, R. and Lee, H.W. (1992). A preemptive priority queue with general bulk service and heterogeneous arrivals, *Performance Evaluation*, 18(3):327-333.
- [12] Dimitriou, I. (2013). A mixed priority retrial queue with negative arrivals, unreliable server and multiple vacations, *Applied Mathematical Modelling*, 37(3):1295-1309.
- [13] Economou, A. and Kapodistria, S. (2010). Synchronized abandonments in a single server unreliable queue, *European Journal of Operational Research*, 203:143-155.
- [14] Jain, M., Sharma, G.C. and Sharma, R. (2012). A batch arrival retrial queuing system for essential and optional services with server breakdown and Bernoulli vacation, *International Journal of Internet and Enterprise Management*, 8(1):16-45.
- [15] Janani, B. (2022). Transient Analysis of Differentiated Breakdown Model, *Applied Mathematics and Computation*, 417(3):131-140 .
- [16] Wua, J. and Lian, Z. (2013). Analysis of the $M^{[1]}, M^{[2]}/G/1$ G-queueing system with retrial customers, *Nonlinear Analysis Real World Applications Journal*, 69(14):365-382.
- [17] Kim, B., Kim, J. and Bueker, O. (2021). Non-preemptive priority $M/M/m$ queue with servers' vacations, *Computers & Industrial Engineering*, 160:171-179.
- [18] Krishnamoorthy, A., Babu, S. and Narayanan, V.C. (2009). The $MAP/(PH/PH)/1$ queue with self-generation of priorities and non-preemptive service, *European Journal of Operational Research*, 195(1):174-185.
- [19] Madan, K.C. and Rawwash, M. (2005). On the $M^{[X]}/G/1$ queue with feedback and optional server vacations based on a single vacation policy, *Applied Mathematics and Computation*, 160(3):909-919.
- [20] Maragathasundari, S. and Srinivasan, S. (2014). A Non-Markovian multistage batch arrival queue with breakdown and reneging, *Hindawi Publishing Corporation*, 14:1-16.
- [21] Medhi, J. (2002). A single server Poisson input queue with second optional channel, *Queueing Systems, Theory Applications*, 42:239-242.
- [22] Gupta, P., Gupta, R. and Malik, S. (2022). Impatient Customers in Queueing System with Optional Vacation Policies and Power Saving Mode, *Applications and Applied Mathematics*, 17(1):1-17.
- [23] Rakesh Kumar and Soodam, B.S. (2019). Transient numerical analysis of a queueing model with correlated reneging, balking and feedback, *Reliability Theory and Applications*, 14(4):46-54.
- [24] Laxmi, P.V. Bhavani, E.G. and Kumar, R. (2022). Correlated reneging in an optional service Markovian queue with working vacations, *Reliability Theory and Applications*, 17(2):507-518.

MULTI-COMPONENT CONDITIONAL STRESS-STRENGTH PARAMETER

KAVOOS KHORSHIDIAN¹, MORTEZA TAHERI SAIF ABAD²

¹Department of Statistics, Faculty of Science, Shiraz University, Shiraz, Iran
khorshidian@shirazu.ac.ir

²Department of Statistics, Faculty of Science, Shiraz University, Shiraz, Iran
taherisaifmorteza@gmail.com

Abstract

There are situations in which the experimenter has some information about the components of the operating system and he/she wants to use this information for better assessment or operating of the underlying system. In such cases the notion of conditional probability may help the operator to use that information and improve his/her task. In the present study this notion has been examined, and some conditional stress-strength parameters have been introduced for s of k systems. The multi-component conditional stress-strength parameter (MCCSSP) and its maximum likelihood estimator have been calculated when the strength and stress random variables are exponentially distributed. In the case of having extra information about the parameters, a closed form has been derived for the Bayes estimator of MCCSSP and has been calculated by using an algorithm together with Monte Carlo method. For the case of non-exponential stress or strengths, the nonparametric estimator of the defined parameter has also been derived. Finally, some simulation study on the MLE and Bayes estimator, as well as real data analysis for nonparametric estimators have been done to verify the analytic results.

Keywords: Conditional Reliability, Exponential Distribution, Maximum Likelihood Estimator, Multi-Component Systems, Stress-Strength Parameter

1. INTRODUCTION

The effects of resistance and shocks which enter to a system are usually studied via a stress-strength model. The term stress-strength was first introduced by [1]. Since then the stress-strength models have been inspected by many researchers due to their applicability in different fields, such as engineering, economics, psychology, medicine and so on. In such models, when the stress that experienced by the system have been represented by a random variable (RV) X and the strength of system by a RV Y , the stress-strength parameter is denoted by $R = P(X > Y)$, it measures the chance that the system fails. It should be mentioned that $1 - R$ is the chance that the considered system operates well and is known as the reliability function or parameter of the system. For the majority of the well-known distributions, including Normal, Exponential, Pareto, Uniform, Weibull, Gamma, Beta, logistic, and Laplace, R has been studied by [2]. Some of the recent studies about R can be seen in [3], [4], [5], [6] and [7]

There are situations that one have some information about the stress and strength RV's and knows that they are greater than some pre-specified values, or one wants to know how much a system can be reliable when stress and strength increase or decrease. Considering conditions like these, the conditional stress-strength parameter was introduced by [8] as:

$$R^{a,b} = P(X > Y \mid X > a, Y > b). \quad (1)$$

Nowadays in the real life and industries most of the operating systems have become complex with more than one active component, i.e., a lot of working systems are multi-component rather than simple and uni-component. The reliability of a multi-component stress-strength model was first developed by [9]. Afterwards, applications and studies on different characteristics of multi-component stress-strength models grow up rapidly. Some of the recent studies can be seen in [10],[11], [12], [13], [14] and [15].

By developments in most technologies, in many situations there are a lot of information about the working mechanisms which will be precise and helpful, if they have been employed corrected, e.g. in the case of second hand and used devices. For example, consider a large drilling machine in a mine. This machine uses several gears or drills simultaneously for drilling, which are the most important parts of this machine and are often iteratively replaced by another one. Therefore, a lot of information about the amount of stress and strength experienced by this part of the machine can be collected . In this article, we have focussed on the notion of conditional stress-strength parameter to extend, generalize and employ such information in multi-component systems. In order to prepare a complete pack about MCCSSP, it has been calculated and estimated by using different methods for employing it in different real situations of practice. For exponential distribution as the first and most exploited candidate of the lifetimes of components in operating systems, the MCCSSP has been calculated, its MLE has been estimated through samples and its asymptotic behaviors has been studied, as well. For the circumstances that we have extra information about the varying structure of exponentially distributed stress and strengths random variables, the Bayes estimators of MCCSSP has been also derived based on the information included in samples of stress and strength. For the case of non-exponential or unknown life time distributions the non-parametric estimators have been also derived.

The structure of this article is as follows: A general formula for computing MCCSSP will have been provided in Section 2. In Section 3, the MCCSSP has been computed in the case of exponential distributions as well as its maximum likelihood estimator and asymptotic distribution of the later. The Bayes estimator of this parameter has been obtained in Section 4, by adopting an algorithm and using the Monte Carlo method. The corresponding nonparametric estimator of this parameter has been obtained in Section 5. Section 6 is devoted to the presentation of some simulation studies on the MLE, Bayesian and nonparametric estimators and their comparison. Some numerical results for a real data-set have been presented in Section 7. Finally in Section 8, some concluding remarks have been given.

2. THE MCCSSP

In this section, the MCCSSP will have been introduced and a general formula have been presented to compute it.

Definition 1. Consider the independent RV's X_1, \dots, X_k with common continuous distribution function $F(\cdot)$, independent of continuous RV Y with distribution function $G(\cdot)$. The MCCSSP is defined as:

$$R_{s,k}^{a,b} = P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y \mid X_1 > a, \dots, X_k > a, Y > b). \quad (2)$$

The particular cases $s = 1$ and $s = k$ correspond to parallel and series systems, respectively. Note that a special case of this quantity for $a = b = -\infty$ is

$$R_{s,k} = P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y) = \sum_{i=s}^k \binom{k}{i} \int_{-\infty}^{\infty} (1 - F(y))^i (F(y))^{k-i} dG(y) \quad (3)$$

which is introduced by [9] as the multi-component stress-strength parameter.

Suppose that there a lot of information about one of the stress RV's X_z , some specified z , $1 \leq z \leq k$. For example, in some systems, one of the parts wears out more and is replaced more often, such as drilling machines, where the drill bit is very important and is replaced a lot, and

the other parts are replaced less often. Therefore, there are more information about the lifetime of a specified part than the other parts. For this case $R_{s,k}^{(z),a,b}$ as the MCCSSP when $X_z > a$ is defined as:

Definition 2.

$$R_{s,k}^{(z),a,b} = P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y \mid X_z > a, Y > b) \quad (4)$$

Note that (3) is again a special case of (4). A formula for computing (2) has been presented in the following theorem.

Theorem 1. If $R_{a,b}^{s,k}$ is defined by (2), then

$$R_{s,k}^{a,b} = \begin{cases} \frac{\sum_{i=s}^k \binom{k}{i} \int_b^\infty [1-F(y)]^i [F(y)-F(b)]^{k-i} dG(y)}{[1-F(a)]^k [1-G(b)]} & a \leq b \\ \frac{\sum_{i=s}^k \binom{k}{i} (\int_a^\infty [G(x)-G(b)] dF(x))^i (\int_a^\infty [1-G(x)] dF(x))^{k-i}}{[1-F(a)]^k [1-G(b)]} & a > b \end{cases} \quad (5)$$

Proof. First, we write (2) as follows:

$$R_{s,k}^{a,b} = \frac{P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y, X_1 > a, \dots, X_k > a, Y > b)}{P(X_1 > a, \dots, X_k > a, Y > b)}.$$

Since X_1, \dots, X_k and Y are independent, the dominator is $(1 - F(a))^k(1 - G(b))$. To compute the numerator, first we write it as follows:

$$\begin{aligned} P(\text{at least } s \text{ of } X_i \text{ exceed } Y, X_1 > a, \dots, X_k > a, Y > b) &= P((X_1, \dots, X_k, Y) \in A) \\ &= \int \dots \int_A dF(x_1) \dots dF(x_k) dG(y), \end{aligned}$$

where $A = \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, x_1 > a, \dots, x_k > a, y > b\}$. To compute this integral, partition A into two regions A_1 and A_2 for the cases $a \leq b$ and $a > b$, where:

$$\begin{aligned} A_1 &= \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, x_1 > a, \dots, x_k > a, y > b, a \leq b\} \\ &= \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, a < x_1 < b, \dots, a < x_k < b, y > b, a \leq b\} \\ &\quad \cup \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, x_1 > b, \dots, x_k > b, y > b, a \leq b\} \\ &= B_1 \cup B_2, \end{aligned}$$

and

$$\begin{aligned} A_2 &= \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, y > b, x_1 > a, \dots, x_k > a, a > b\} \\ &= \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } (b, x_1), \dots, (b, x_k) \text{ contain } y, y > b, x_1 > a, \dots, x_k > a, a > b\}. \end{aligned}$$

where

$$\begin{aligned} B_1 &= \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, a < x_1 < b, \dots, a < x_k < b, y > b, a \leq b\}, \\ B_2 &= \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, x_1 > b, \dots, x_k > b, y > b, a \leq b\}. \end{aligned}$$

Let

$$R_1 = \int_{A_1} dF(x_1) \dots dF(x_k) dG(y), \quad (6)$$

and

$$R_2 = \int_{A_2} dG(y) dF(x_1) \dots dF(x_k) \quad (7)$$

then

$$\begin{aligned} R_1 &= \int_{B_1} dF(x_1) \dots dF(x_k) dG(y) + \int_{B_2} dF(x_1) \dots dF(x_k) dG(y) \\ &= \int_{B_2} dF(x_1) \dots dF(x_k) dG(y) \\ &= \sum_{i=s}^k \binom{k}{i} \int_b^\infty [1-F(y)]^i [F(y)-F(b)]^{k-i} dG(y) \end{aligned}$$

The first integral becomes zero because $P(X_i > Y, a < X_i < b, Y > b) = 0$ for $i = 1, \dots, k$, and

$$\begin{aligned} R_2 &= P(\text{at least } s \text{ of } (b, X_1), \dots, (b, X_k) \text{ contain } Y, Y > b, X_1 > a, \dots, X_k > a, a > b) \\ &= \sum_{i=s}^k \binom{k}{i} \left[\int_a^\infty [G(x) - G(b)] dF(x) \right]^i \left[\int_a^\infty [1 - G(x)] dF(x) \right]^{k-i}, \end{aligned}$$

This completes the proof. ■

Remark 1. Consider an s of k multi-component system, which their strengths are denoted by iid RV's X_1, X_2, \dots, X_k with common continuous distribution function $F(\cdot)$. Also suppose that each component experiences a random stress Y with continuous distribution function $G(\cdot)$, independent of the strengths. Note that the system stays alive only if at least s of k strengths be greater than the stress. Then the conditional reliability of the multi-component system has the following form:

$$R_{s,k}^{a,b} = P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y \mid X_1 > a, \dots, X_k > a, Y > b). \quad (8)$$

In this model, the conditional reliability of the system is represented by (5).

Remark 2. In practice the information in hand and given condition may not have exactly the form $\{x_1 > a, \dots, X_k > a, Y > b\}$, but be as $\{X_1 \in \mathcal{A}_1, \dots, X_k \in \mathcal{A}_k, Y \in \mathcal{B}\}$ where $\mathcal{A}_1, \dots, \mathcal{A}_k$ and \mathcal{B} are linear Borel sets on $(0, \infty)$. In this case, by applying some procedure similar to the approach of Theorem 1, one can compute this generalized MCCSSP. Based on the structure of $\mathcal{A}_1, \dots, \mathcal{A}_k$ and \mathcal{B} , it is expected that the analytic derivations may be complicated. In this situation and more general case some non-parametric method similar to that given in section 5 as well as Monte Carlo simulation may be applied.

Remark 3. By formula (5), one may show that for the case $a \leq b$, the MCCSSP $R_{s,k}^{a,b}$ is an increasing function of a , which is expected trivially. Note that in this case:

$$\begin{aligned} \frac{\partial R_{s,k}^{a,b}}{\partial a} &= \frac{kf(a)(1 - F(a))^{k-1} \sum_{i=s}^k \binom{k}{i} \int_b^\infty [1 - F(y)]^i [F(y) - F(b)]^{k-i} dG(y)}{[1 - F(a)]^{2k} [1 - G(b)]} \\ &= kR_{s,k}^{a,b} \frac{f(a)}{1 - F(a)} \geq 0. \end{aligned}$$

According to the calculations resulting in the formula (5), it can be seen that if X_1, \dots, X_k have different distributions, it is not easy to calculate the analogous of this formula. In what follows, the formula (5) has been calculated when X_1, \dots, X_k and Y have the same distributions.

Corollary 1. Suppose that the continuous RV's X_1, \dots, X_k and Y are independent and identically distributed with probability density function(pdf) $f(\cdot)$ and cumulative distribution function(cdf) $F(\cdot)$. Then,

$$R_{s,k}^{a,b} = \begin{cases} \frac{\sum_{i=s}^k \binom{k}{i} \int_{F(b)}^1 [1-y]^i [y-F(b)]^{k-i} dy}{[1-F(a)]^k [1-F(b)]}, & a \leq b \\ \left(\frac{1}{2}\right)^k \frac{\sum_{i=s}^k \binom{k}{i} [1-2F(b)+F(a)]^i [1-F(a)]^{k-i}}{[1-F(b)]}, & a > b. \end{cases} \quad (9)$$

Remark 4. Put $R_{s,k}^{(z)1,a,b} = P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y, X_z \geq Y \mid X_z > a, Y > b)$ and $R_{s,k}^{(z)2,a,b} = P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y, X_z < Y \mid X_z > a, Y > b)$ for some $z, 1 \leq z \leq k$. According to the approach of the proof for Theorem 1, after some computation, we have:

$$R_{s,k}^{(z)1,a,b} = \begin{cases} \frac{\sum_{i=s}^k \binom{k}{i} \int_b^\infty [1-F(y)]^i F(y)^{k-i} dG(y)}{[1-F(a)]^k [1-G(b)]} & a \leq b \\ \frac{\sum_{i=s}^k \binom{k}{i-k+i} \left[\int_0^\infty [G(x)-G(b)] dF(x) \right]^{i-1} \left[\int_a^\infty [G(x)-G(b)] dF(x) \right] \left[\int_0^\infty [1-G(x)] dF(x) \right]^{k-i}}{[1-F(a)]^k [1-G(b)]} & a > b, \end{cases} \quad (10)$$

$$R_{s,k}^{(z)2,a,b} = \begin{cases} \frac{\sum_{i=s}^{k-1} \binom{k}{i,k-i-1} \int_b^\infty [1-F(y)]^i [F(y)-F(b)] F(y)^{k-(i+1)} dG(y)}{[1-F(a)]^k [1-G(b)]} & a \leq b \\ \frac{\sum_{i=s}^{k-1} \binom{k}{i,k-i-1} \left[\int_0^\infty [G(x)-G(b)] dF(x) \right]^i \left[\int_a^\infty [1-G(x)] dF(x) \right] \left[\int_0^\infty [1-G(x)] dF(x) \right]^{k-(i+1)}}{[1-F(a)]^k [1-G(b)]} & a > b. \end{cases} \quad (11)$$

Therefore

$$R_{s,k}^{(z),a,b} = \begin{cases} R_{s,k}^{(z)1,a,b} & X_z \geq Y \\ R_{s,k}^{(z)2,a,b} & X_z < Y. \end{cases} \quad (12)$$

3. ESTIMATION FOR EXPONENTIAL DISTRIBUTION

In this section, the measure (5) has been evaluated for the Exponentially distributed stresses and strength RV's with different parameters. The probability density and cumulative distribution functions of a random variable $X \sim E(\alpha)$ are denoted by: $f(x) = \alpha e^{-\alpha x}$, and $F(x) = 1 - e^{-\alpha x}$ where $x \geq 0, \alpha > 0$. Suppose that $X_i \sim E(\lambda_1)$ for $i = 1, \dots, k$ and $Y \sim E(\lambda_2)$ are independent, we have:

$$R_1 = \lambda_2 e^{-b(\lambda_1 k + \lambda_2)} \sum_{i=s}^k \sum_{j=0}^{k-i} \binom{k}{i,j} \frac{(-1)^j}{\lambda_1(i+j) + \lambda_2},$$

and

$$R_2 = e^{-ak(\lambda_1 + \lambda_2)} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right]^k \sum_{i=s}^k \binom{k}{i} \left[\frac{\lambda_1 + \lambda_2}{\lambda_1} e^{-\lambda_2(b-a)} - 1 \right]^i,$$

by dividing the above equations by $[1 - F(a)]^k [1 - G(b)] = e^{-(ak\lambda_1 + b\lambda_2)}$ we have:

$$R_{s,k}^{a,b} = \begin{cases} \lambda_2 e^{-\lambda_1 k(b-a)} \sum_{i=s}^k \sum_{j=0}^{k-i} \binom{k}{i,j} \frac{(-1)^j}{\lambda_1(i+j) + \lambda_2} & a \leq b \\ e^{-\lambda_2(ak-b)} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right]^k \sum_{i=s}^k \binom{k}{i} \left[\frac{\lambda_1 + \lambda_2}{\lambda_1} e^{-\lambda_2(b-a)} - 1 \right]^i & a > b. \end{cases} \quad (13)$$

Remark 5. From (13), we conclude that $R_{s,k}^{a,b}$ for $a \leq b$ in Exponential distribution depends only on the difference between a and b . In other words, if $b_1 - a_1 = b_2 - a_2$ then $R_{s,k}^{a_1,b_1} = R_{s,k}^{a_2,b_2}$ for $a_1 \leq b_1$ and $a_2 \leq b_2$.

Figure 1 show the effect of changes in the values a and b in (13). These figures show what happens when the values a and b increase or decrease, in all Figures $(s, k) = (1, 3)$.

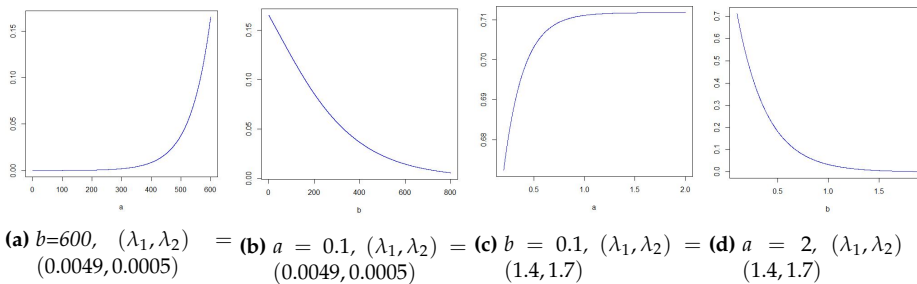


Figure 1: MCCSSP

By assuming the Exponential distributions for stresses and strength, from (10) and (12), after some calculation it follows that for the case $X_z \geq Y$, we have:

$$R_{s,k}^{(z),a,b} = \begin{cases} \sum_{i=s}^k \sum_{j=0}^{k-i} \binom{k}{i,j} \frac{\lambda_2 (-1)^j e^{-\lambda_1(b(k-j)-a)}}{\lambda_1(k-j) + \lambda_2} & a \leq b \\ e^{-\lambda_2(a-b)} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right]^k \sum_{i=s}^k \binom{k}{i} \left[\frac{\lambda_1 + \lambda_2}{\lambda_1} e^{-\lambda_2 b} - 1 \right]^{i-1} \left[\frac{\lambda_1 + \lambda_2}{\lambda_1} e^{-\lambda_2(b-a)} - 1 \right] & a > b, \end{cases} \quad (14)$$

and for the case $X_z < Y$:

$$R_{s,k}^{(z),a,b} = \begin{cases} \lambda_2 e^{-\lambda_1(b-a)} \left[\sum_{i=s}^{k-1} \sum_{j=0}^{k-i} \binom{k}{i} \binom{k-i}{j} (-1)^j e^{-\lambda_1(b(i+j))} \left[\frac{1}{\lambda_1(i+j) + \lambda_2} - \frac{e^{-\lambda_1(b-a)}}{\lambda_1(i+j+1) + \lambda_2} \right] \right] & a \leq b \\ e^{-\lambda_2(a-b)} \left[\frac{\lambda_1}{\lambda_1 + \lambda_2} \right]^k \left[\sum_{i=s}^{k-1} \binom{k}{i} \left[\frac{\lambda_1 + \lambda_2}{\lambda_1} e^{-\lambda_2 b} - 1 \right]^i \right] & a > b. \end{cases} \quad (15)$$

Indeed, as in (13) to (15), the stress-strength parameters $R_{s,k}^{a,b}$ and $R_{s,k}^{(z),a,b}$ are functions of λ_1 and λ_2 . Therefore, it is rational that for evaluating the maximum likelihood estimators of MCCSSP, the first step to be calculating the MLE's of λ_1 and λ_2 .

3.1. Maximum Likelihood Estimation

Suppose that X_1, \dots, X_n and Y_1, \dots, Y_m are two independent random samples from $E(\lambda_1)$ and $E(\lambda_2)$. Then the likelihood function is

$$L(\lambda_1, \lambda_2) = \lambda_1^n \lambda_2^m e^{-\lambda_1 \sum_1^n x_i} e^{-\lambda_2 \sum_1^m y_j} \tag{16}$$

and the MLE's of the parameters λ_1 and λ_2 are $\hat{\lambda}_1 = \frac{1}{\bar{X}}$ and $\hat{\lambda}_2 = \frac{1}{\bar{Y}}$, respectively. Therefore, by using the invariance property for MLE's, and substituting $\hat{\lambda}_1$ and $\hat{\lambda}_2$ instead of λ_1 and λ_2 in (14) and (15), one may write the MLE of (13) by:

$$\hat{R}_{s,k}^{a,b} = \begin{cases} \hat{\lambda}_2 e^{-\hat{\lambda}_1 k(b-a)} \sum_{i=s}^k \sum_{j=0}^{k-i} \binom{k}{i,j} \frac{(-1)^j}{\hat{\lambda}_1^{i+j} + \hat{\lambda}_2} & a \leq b \\ e^{-\hat{\lambda}_2(ak-b)} \left[\frac{\hat{\lambda}_1}{\hat{\lambda}_1 + \hat{\lambda}_2} \right]^k \sum_{i=s}^k \binom{k}{i} \left[\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\hat{\lambda}_1} e^{-\hat{\lambda}_2(b-a)} - 1 \right]^i & a > b. \end{cases} \tag{17}$$

3.2. Asymptotic Distribution

In this subsection the asymptotic distribution of $\hat{R}_{s,k}^{a,b}$ will have been obtained by using the asymptotic normality of the MLE's and the multivariate delta method. By the fact that $\hat{\lambda} \rightarrow N_2(\lambda, \Sigma)$ as n, m tend to infinity, $\frac{n}{m} \rightarrow d$ for some $0 < d < \infty$, where $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2)^T$, $\lambda = (\lambda_1, \lambda_2)^T$ and Σ is the inverse of Fisher's information matrix $I(\lambda)$, it is easy to see that

$$I(\lambda) = \begin{bmatrix} \frac{n}{\lambda_1^2} & 0 \\ 0 & \frac{m}{\lambda_2^2} \end{bmatrix}, \quad \text{and so} \quad \Sigma = \begin{bmatrix} \frac{\lambda_1^2}{n} & 0 \\ 0 & \frac{\lambda_2^2}{m} \end{bmatrix}.$$

The well-known delta method enables us to derive the asymptotic behaviour of functions of an estimator, whenever the estimator is itself asymptotically normal. The delta method have been present and applied in different forms, we have used the following presentation.

Proposition 1. Let $g(\cdot)$ be a mapping $g(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$, such that $g(\cdot)$ is continuous in a neighborhood of $\mu \in \mathbb{R}^d$. If \mathbf{X}_n is a sequence of d -dimensional random vectors such that $\mathbf{X}_n \rightarrow N_d(\mu, \Sigma)$ in distribution, then $\frac{g(\mathbf{X}_n) - g(\mu)}{\tau} \rightarrow N(0, 1)$ in distribution, where $\tau^2 = \nabla^T \Sigma \nabla > 0$ and $\nabla = \frac{\partial g(\mu)}{\partial \mu}$.

We will apply Proposition 1 to $\mathbf{X}_n = \hat{\lambda}$ and

$$g(x_1, x_2) = \begin{cases} x_2 e^{-x_1 k(b-a)} \sum_{i=s}^k \sum_{j=0}^{k-i} \binom{k}{i,j} \frac{(-1)^j}{x_1^{i+j} + x_2} & a \leq b \\ e^{-x_2(ak-b)} \left[\frac{x_1}{x_1 + x_2} \right]^k \sum_{i=s}^k \binom{k}{i} \left[\frac{x_1 + x_2}{x_1} e^{-x_2(b-a)} - 1 \right]^i & a > b. \end{cases}$$

The asymptotic distribution of $\hat{R}_{s,k}^{a,b}$ may be obtained as below:

$$(\hat{R}_{s,k}^{a,b} - R_{s,k}^{a,b}) \rightarrow N(0, \nabla^T \Sigma \nabla), \tag{18}$$

where

$$\begin{aligned} \nabla &= \left(\frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_1}, \frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_2} \right)^T, \\ \nabla^T \Sigma \nabla &= \left[\frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_1} \right]^2 \frac{\lambda_1^2}{n} + \left[\frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_2} \right]^2 \frac{\lambda_2^2}{m}, \end{aligned} \tag{19}$$

For the cases $a \leq b$ and $a > b$, denote (19) by σ_1^2 and σ_2^2 respectively. Put $\psi = \frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_1} \Big|_{a \leq b}$, $\nu = \frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_1} \Big|_{a > b}$, $\theta = \frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_2} \Big|_{a \leq b}$, $\kappa = \frac{\partial g(\lambda_1, \lambda_2)}{\partial \lambda_2} \Big|_{a > b}$, we arrive at:

$$\sigma_1^2 = \psi^2 \frac{\lambda_1^2}{n} + \theta^2 \frac{\lambda_2^2}{m}, \tag{20}$$

$$\sigma_2^2 = \nu^2 \frac{\lambda_1^2}{n} + \kappa^2 \frac{\lambda_2^2}{m}. \tag{21}$$

Therefore, the asymptotic normalized distribution of $\hat{R}_{s,k}^{[a,b]}$ for different values of a and b are as follow:

$$\frac{\hat{R}_{s,k}^{[a,b]} - R_{s,k}^{[a,b]}}{\sigma_i} \rightarrow N(0,1) \quad i = 1,2, \tag{22}$$

where σ_1 and σ_2 stands for the cases $a \leq b$ and $a > b$ respectively. The above statistics can be used for constructing confidence intervals for $R_{s,k}^{[a,b]}$. By employing a similar approach and performing some steps like the above, using lemma 1 and asymptotic normality of $\hat{R}_{s,k}^{[(z),a,b]}$, one may arrive at the asymptotic distribution of $R_{s,k}^{[(z),a,b]}$.

4. BAYES ESTIMATION

In this section, the Bayesian estimation of the reliability parameter (13) has been considered. Suppose that the parameters λ_1 and λ_2 are RV's, and have independent Gamma prior distributions with parameters $(\alpha_i, \beta_i), i = 1,2$ respectively. The pdf of a random variable $X \sim Gamma(\alpha_i, \beta_i)$ is denoted by

$$\pi(x) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i-1} e^{-\beta_i x} \quad x > 0, \alpha_i > 0, \beta_i > 0. \tag{23}$$

The joint posterior density function of the parameters based on this prior density and the likelihood function can be written as follows:

$$\pi^*(\lambda_1, \lambda_2 | \mathbf{x}, \mathbf{y}) = \frac{\pi(\lambda_1, \lambda_2, \mathbf{x}, \mathbf{y})}{\int_0^\infty \int_0^\infty \pi(\lambda_1, \lambda_2, \mathbf{x}, \mathbf{y}) d\lambda_1 d\lambda_2} \tag{24}$$

where

$$\pi(\lambda_1, \lambda_2, \mathbf{x}, \mathbf{y}) = \pi(\lambda_1)\pi(\lambda_2)L(\lambda_1, \lambda_2) \propto \lambda_1^{\alpha_1+n-1} e^{-\lambda_1(\beta_1+\sum_{i=1}^n x_i)} \lambda_2^{\alpha_2+m-1} e^{-\lambda_2(\beta_2+\sum_{j=1}^m y_j)}.$$

It is easily seen that the posterior density functions of λ_1 and λ_2 are respectively

$$\pi^*(\lambda_1 | \lambda_2, \mathbf{x}, \mathbf{y}) \propto \Gamma(\alpha_1 + n, \beta_1 + \sum_{i=1}^n x_i), \tag{25}$$

$$\pi^*(\lambda_2 | \lambda_1, \mathbf{x}, \mathbf{y}) \propto \Gamma(\alpha_2 + m, \beta_2 + \sum_{j=1}^m y_j). \tag{26}$$

The Bayes estimator of $R_{s,k}^{[a,b]}$ under the squared error loss (SEL) is obtained as

$$\tilde{R}_{s,k}^{[a,b]} = E(R_{s,k}^{[a,b]} | \mathbf{x}, \mathbf{y}) = \int_0^\infty \int_0^\infty R_{s,k}^{[a,b]} \pi^*(\lambda_1, \lambda_2 | \mathbf{x}, \mathbf{y}) d\lambda_1 d\lambda_2. \tag{27}$$

It is not possible to calculate equation (27) analytically. Therefore, to compute the Bayes estimate of reliability parameter $R_{s,k}^{[a,b]}$, a Monte Carlo (MC) method has been adopted as follows:

- Step 1: Set $l=1$.
- Step 2: Generate X_1, \dots, X_n from $Exp(\lambda_1)$.
- Step 3: Generate Y_1, \dots, Y_m from $Exp(\lambda_2)$
- Step 4: Generate λ_1^l from $Gamma(\alpha_1 + n, \beta_1 + \sum_{i=1}^n x_i)$.
- Step 5: Generate λ_2^l from $Gamma(\alpha_2 + m, \beta_2 + \sum_{j=1}^m y_j)$.
- Step 6: Compute $R_{s,k}^{[a,b]}$ at $(\lambda_1^l, \lambda_2^l)$.
- Step 7: $l=l+1$.

Step 8: Repeat Steps 2 to 7, M times and obtain the posterior sample $R_{s,k}^{l|a,b}$ for $l = 1, \dots, M$.
 Now the Bayes estimate of $R_{s,k}^{a,b}$ with respect to SEL will be obtained as follows:

$$\tilde{R}_{s,k}^{a,b} = \frac{1}{M} \sum_{l=1}^M R_{s,k}^{l|a,b}. \tag{28}$$

5. NONPARAMETRIC ESTIMATION

In this section a nonparametric method for estimating $R_{s,k}^{a,b}$ has been presented. In many situations, we may have no information about the distribution of data or computing $R_{s,k}^{a,b}$ via 1 may require complex computations, or even may not have a definite answer. Therefore, employing the nonparametric method, in which the structure of the model may have been determined from data, can lead us to better results or at least be more applicable. Let $n(\cdot)$ be the counting measure. For the sample space \mathbf{S} and the event \mathbf{D} as a subset of \mathbf{S} the nonparametric estimator of $P(\mathbf{D})$ is defined as $\hat{P}(\mathbf{D}) = \frac{n(\mathbf{D})}{n(\mathbf{S})}$. To obtain the nonparametric estimator of MCCSSP, one may write (2) in the form:

$$R_{s,k}^{a,b} = \frac{P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y, X_1 > a, \dots, X_k > a, Y > b)}{P(X_1 > a, \dots, X_k > a, Y > b)} \tag{29}$$

where $P(X_1 > a, \dots, X_k > a, Y > b) > 0$. Since X_1, \dots, X_k and Y are independent, equation (29) can be written as follows:

$$R_{s,k}^{a,b} = \frac{P(\text{at least } s \text{ of } X_1, \dots, X_k \text{ exceed } Y, X_1 > a, \dots, X_k > a, Y > b)}{P(X_1 > a, \dots, X_k > a)P(Y > b)}, \tag{30}$$

where $P(X_1 > a, \dots, X_k > a)P(Y > b) > 0$.

Let $\mathbf{A} = \{(x_1, \dots, x_k, y) \mid \text{at least } s \text{ of } x_1, \dots, x_k \text{ exceed } y, x_1 > a, \dots, x_k > a, y > b\}$, $\mathbf{B} = \{(x_1, \dots, x_k) \mid x_1 > a, \dots, x_k > a\}$ and $\mathbf{C} = \{y \mid y > b\}$. The nonparametric estimator of (30) can be written as follows:

$$R_{s,k}^{NP|a,b} = \frac{n(\mathbf{A})}{n(\mathbf{B})n(\mathbf{C})}. \tag{31}$$

Let $X_{1i}, \dots, X_{ki} \sim X$ for $i = 1, \dots, n$ and $Y_1, \dots, Y_m \sim Y$ be independent random samples. Also, let $I(E)$ be the indicator function of the event E , that is a RV that takes value 1 when the event E happens and 0 when the event does not happen. By assuming $n(\cdot)$ as the counting measure, we have:

$$n(\mathbf{B}) = \sum_{i=1}^n I(X_{1i} > a, \dots, X_{ki} > a), \tag{32}$$

$$n(\mathbf{C}) = \sum_{j=1}^m I(Y_j > b), \tag{33}$$

and by the properties of the indicator function:

$$\begin{aligned} n(\mathbf{A}) &= \sum_{i=1}^n \sum_{j=1}^m I(s \text{ of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j) I(X_{1i} > a, \dots, X_{ki} > a) I(Y_j > b) + \dots \\ &+ \sum_{i=1}^n \sum_{j=1}^m I(k \text{ of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j) I(X_{1i} > a, \dots, X_{ki} > a) I(Y_j > b). \end{aligned} \tag{34}$$

Let $\mathbf{X}_i = (X_{1i}, \dots, X_{ki})$ for $i = 1, \dots, n$. Those observations \mathbf{X}_i and Y_j for them both $\mathbf{X}_i \leq a$ and $Y_j \leq b$ simultaneously, have been removed in calculating $n(\mathbf{A})$, since in details of calculating $P(\mathbf{A})$ or $R_{s,k}^{NP|a,b} = \frac{n(\mathbf{A})}{n(\mathbf{B})n(\mathbf{C})}$, the numerator is an strict subset of denominator. Note that in this

case the values of the second and third indicators will automatically equal one in $n(\mathbf{A})$, (34). It is worth noting that the number of reminded samples of X_i and Y_j are $n(\mathbf{B})$ and $n(\mathbf{C})$, so $n(\mathbf{A})$ can be written as follows:

$$n(\mathbf{A}) = \sum_{i=1}^{n(\mathbf{B})} \sum_{j=1}^{n(\mathbf{C})} I(\text{s of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j) + \dots + \sum_{i=1}^{n(\mathbf{B})} \sum_{j=1}^{n(\mathbf{C})} I(\text{k of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j)$$

In the case of $n = m$, the formula (31) may have simpler form and computations, since we only keep those $(X_{1i}, \dots, X_{ki}, Y_j)$ $i = 1, \dots, n$ which for them $(X_{1i} > a, \dots, X_{ki} > a, Y_j > b)$ and remove the rest and also, $n(\mathbf{B}) = n(\mathbf{C})$. In what follows, we introduce a definition and representation for non-parametric estimator of multi-component stress-strength parameter. To the best of our knowledge, interestingly this estimator has not been defined till now.

Definition 3. The nonparametric estimator of $R_{s,k}$ is defined as follows:

$$R_{s,k}^{NP} = \frac{n(\mathbf{A})}{n(\mathbf{B})n(\mathbf{C})} \tag{35}$$

where $n(\mathbf{B}) = n$, $n(\mathbf{C}) = m$ and

$$n(\mathbf{A}) = \sum_{i=1}^n \sum_{j=1}^m I(\text{s of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j) + \dots + \sum_{i=1}^n \sum_{j=1}^m I(\text{k of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j).$$

Note that (35) can be obtained from (31) by assuming $a = b = 0$.

Remark 6. (i): By (31), and according to the definitions of $n(\mathbf{A})$, $n(\mathbf{B})$ and $n(\mathbf{C})$, it can be concluded that for fixed values of a , $a \leq b$, the estimator $R_{s,k}^{NP|a,b}$ is a decreasing function of b .

(ii): By (31), and according to the definitions of $n(\mathbf{A})$, $n(\mathbf{B})$ and $n(\mathbf{C})$, it can be concluded that for fixed values of b , $a > b$, the estimator $R_{s,k}^{NP|a,b}$ is an increasing function of a .

In applications, the data observed for different stresses may differ greatly in their values. Therefore, selecting a minimum value of a , w.r.t. it all stresses in MCCSSP through definition 1, satisfy the corresponding condition $X_i > a$, may be not useful. So, in what follows, the MCCSSP has been defined in some general way to be more realistic and applicable.

Definition 4. The generalized conditional multi-component stress-strength parameter is defined as follows:

$$R_{s,k}^{a_1, \dots, a_k, b} = P(\text{at least s of } X_1, \dots, X_k \text{ exceed } Y \mid X_1 > a_1, \dots, X_k > a_k, Y > b) \tag{36}$$

where the RV's Y, X_1, \dots, X_k are independent, $G(\cdot)$ is the continuous distribution function of Y and $F(\cdot)$ is the common continuous distribution function of X_1, \dots, X_k .

Theorem 2. If $X_{ri} > \max(a_1, \dots, a_k)$ for $r = 1, \dots, k$; $i = 1, \dots, n$ and $Y_j > b$ for $j = 1, \dots, m$ then $R_{s,k}^{NP|a_1, \dots, a_k, b} = R_{s,k}^{NP}$.

Proof. Replace $I(X_{1i} > a_1, \dots, X_{ki} > a_k)$ with $I(X_{1i} > a, \dots, X_{ki} > a)$ in (32) and (34). Since $I(X_{1i} > a_1, \dots, X_{ki} > a_k) = 1$ and $I(Y_j > b) = 1$ we have $n(\mathbf{B}) = n$, $n(\mathbf{C}) = m$ and $n(\mathbf{A}) = \sum_{i=1}^n \sum_{j=1}^m I(\text{s of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j) + \dots + \sum_{i=1}^n \sum_{j=1}^m I(\text{k of } X_{1i}, \dots, X_{ki} \text{ exceed } Y_j)$. ■ Of course, a special case of (36) is (2). In parametric case (MLE method) when a_1, \dots, a_k are closed in values, a can be considered as the minimum or maximum of a_1, \dots, a_k and approximate (36) through (4). In some situations, a_1, \dots, a_k are very different, and using (36) is not very helpful or may not be accurate. In these cases, the non-parametric method is more practical and it is enough to consider $\mathbf{A} = \{(x_1, \dots, x_k, y) \mid \text{at least s of } x_1, \dots, x_k \text{ exceed } y, x_1 > a_1, \dots, x_k > a_k, y > b\}$, and $\mathbf{B} = \{(x_1, \dots, x_k) \mid x_1 > a_1, \dots, x_k > a_k\}$ in (31). It is easy to see that the results of nonparametric estimation of (29) can also be used for nonparametric estimation of (36), where a_i is substituted instead of a for $i = 1, \dots, k$. Note that in this case, one advantage of the nonparametric method is that the assumption of common distribution for stress RV's may be relaxed. The later makes this method much more practical. If $\mathbf{B} = \{(x_1, \dots, x_k, y) \mid x_1 > a_1, \dots, x_k > a_k, y > b\}$, then the nonparametric estimator of the generalized MCCSSP where stresses and strength RV's are not independent, can also be easily computed through the same method.

6. SIMULATION

In this section, a simulation study has been done to assess the quality and the efficiency of performance of $R_{s,k}^{ab}$, its MLE, Bayes and nonparametric estimators. The performances of the MLE, Bayes and nonparametric estimators have been studied by using their biases. The performances of the confidence intervals for MLE are studied by using average confidence lengths (ACL's) and coverage probabilities (CP's). It would be mentioned that, the proportion of the times that the intervals contain the true value of interest is called the coverage probability of a confidence interval. The simulations have been only done for $a \neq b$ since for $a = b$ the conditional and unconditional cases have the same results.

The results for $R_{s,k}^{ab}$, MLE's, Biases, MSE's, ACL's and CP's and different values of m and n where the other parameters are fixed, have been shown in the Tables 1 for $\lambda_1 = 1, \lambda_2 = 2$ and shown in 2 for $\lambda_1 = 1.5, \lambda_2 = 0.7$. According to these tables larger sample sizes have more reliable results. A comparison among MLE, $R_{1,3}^{NP|a,b}$ and $\hat{R}_{1,3}^{a,b}$ assuming $\alpha_1 = 2, \beta_1 = 3, \alpha_2 = 5, \beta_2 = 4$ for different values of a and b, $n = m = 100, \lambda_1 = 0.0003$ and $\lambda_2 = 0.0005$ has been done and the results presented in Tables 3 and 4. A comparison among $\hat{R}_{2,4}^{a,b}$ and $\tilde{R}_{2,4}^{a,b}$ assuming $\lambda_1 = 3, \lambda_2 = 2, \alpha_1 = 5, \beta_1 = 0.8, \alpha_2 = 4, \beta_2 = 0.2$ for different sample sizes has been done and the results presented in Table 5. A nonparametric simulation for different values of $a_1, a_2, a_3, \lambda_1 = 0.004$ and $\lambda_2 = 0.002$ has been done and the results are presented in Table 6.

Table 1: Comparison of estimators, $R_{1,3}^{0.7,1} = 0.3659, R_{2,4}^{0.7,1} = 0.2409$

(s,k)	n	15	20	35	50	85	100
	m	15	25	35	50	75	100
(1,3)	$\hat{R}_{1,3}^{0.7,1}$	0.3556	0.3577	0.3612	0.3628	0.3632	0.3646
	MSE	0.0439	0.0256	0.0180	0.0125	0.0083	0.0062
	Bias	-0.0102	-0.0082	-0.0046	-0.0030	-0.0026	-0.0013
	ACL	0.6889	0.5202	0.4425	0.3674	0.2990	0.2588
	CP	0.9384	0.9498	0.9736	0.9802	0.9818	0.9844
(2,4)	$\hat{R}_{2,4}^{0.7,1}$	0.2372	0.2357	0.2390	0.2393	0.2400	0.2400
	MSE	0.0243	0.0139	0.0100	0.0069	0.0046	0.0034
	Bias	-0.0036	-0.0052	-0.0018	-0.0015	-0.0008	-0.0008
	ACL	0.5145	0.3879	0.3296	0.2741	0.2228	0.1921
	CP	0.8988	0.9012	0.938	0.9518	0.9640	0.9690

Table 2: Comparison of estimators, $R_{1,3}^{1.2,0.5} = 0.4603, R_{2,4}^{1.2,0.5} = 0.2798$

(s,k)	n	15	20	35	50	75	100
	m	15	25	35	50	75	100
(1,3)	$\hat{R}_{1,3}^{1.2,0.5}$	0.4404	0.4481	0.4551	0.4537	0.4562	0.4567
	MSE	0.0391	0.0225	0.0097	0.0096	0.0060	0.0046
	Bias	-0.0199	-0.0122	-0.0052	-0.0065	-0.0041	-0.0035
	ACL	0.5655	0.4581	0.3110	0.3085	0.2481	0.2183
	CP	0.9374	0.9696	0.9826	0.9838	0.9876	0.9932
(2,4)	$\hat{R}_{2,4}^{1.2,0.5}$	0.2608	0.2679	0.2718	0.2742	0.2758	0.2771
	MSE	0.0060	0.0045	0.0025	0.0017	0.0010	0.0008
	Bias	-0.0197	-0.0118	-0.0079	-0.0056	-0.0040	-0.0027
	ACL	0.2349	0.2079	0.1600	0.1351	0.0104	0.0964
	CP	0.9156	0.9548	0.9620	0.9690	0.9712	0.9837

Table 3: Comparison of $\hat{R}_{1,3}^{a,b}$, $R_{1,3}^{NP|a,b}$, $\tilde{R}_{1,3}^{a,b}$ for $a \leq b$

a	10	25	70	78	170	215	300
b	20	40	74	120	190	260	310
$R_{1,3}^{a,b}$	0.8607	0.8568	0.8653	0.8362	0.8530	0.8340	0.8607
$\hat{R}_{1,3}^{a,b}$	0.8555	0.8516	0.8602	0.8311	0.8478	0.8327	0.8555
$\tilde{R}_{1,3}^{a,b}$	0.8598	0.8559	0.8646	0.8349	0.8519	0.8321	0.8598
$R_{1,3}^{NP a,b}$	0.8657	0.8655	0.8697	0.8681	0.8691	0.8646	0.8668
Bias($\hat{R}_{1,3}^{a,b}$)	-0.0051	-0.0051	-0.0051	-0.0051	-0.0051	-0.0013	-0.0051
Bias($\tilde{R}_{1,3}^{a,b}$)	-0.0008	-0.0009	-0.0007	-0.0013	-0.0010	-0.0019	-0.0008
Bias($R_{1,3}^{NP a,b}$)	0.0049	0.0087	0.0043	0.0618	0.0161	0.0305	0.0061

Table 4: Comparison of $\hat{R}_{1,3}^{a,b}$, $R_{1,3}^{NP|a,b}$, $\tilde{R}_{1,3}^{a,b}$ for $a > b$

a	7	22	45	67	100	120	240
b	4	11	38	65	90	70	230
$R_{1,3}^{a,b}$	0.9437	0.9377	0.9124	0.8877	0.8664	0.8867	0.7532
$\hat{R}_{1,3}^{a,b}$	0.9397	0.9338	0.9088	0.8844	0.8633	0.8836	0.7514
$\tilde{R}_{1,3}^{a,b}$	0.9372	0.9313	0.90559	0.8812	0.8598	0.8805	0.7466
$R_{1,3}^{NP a,b}$	0.8654	0.8658	0.8658	0.8653	0.8674	0.8686	0.8680
Bias($\hat{R}_{1,3}^{a,b}$)	-0.0039	-0.0038	-0.0036	-0.0033	-0.0030	-0.0030	-0.0017
Bias($\tilde{R}_{1,3}^{a,b}$)	-0.0064	-0.0063	-0.0064	-0.0065	-0.0065	-0.0062	-0.0065
Bias($R_{1,3}^{NP a,b}$)	-0.0782	-0.0719	-0.0466	-0.0224	0.0097	-0.0181	0.1147

Table 5: Comparison of $\hat{R}_{2,4}^{a,b}$, $\tilde{R}_{2,4}^{a,b}$, exact values $R_{2,4}^{0.6,0.8} = 0.0430$, $R_{2,4}^{0.9,0.6} = 0.0243$

n	10	20	30	60	95	100	150
m	10	22	30	58	100	120	150
$\hat{R}_{2,4}^{0.6,0.8}$	0.0505	0.0475	0.0457	0.0443	0.0437	0.0436	0.0434
$\tilde{R}_{2,4}^{0.6,0.8}$	0.0360	0.0407	0.0405	0.0418	0.0419	0.0421	0.0424
Bias($\hat{R}_{2,4}^{0.6,0.8}$)	-0.0075	-0.0045	-0.0027	-0.0013	-0.0007	-0.0006	-0.0004
Bias($\tilde{R}_{2,4}^{0.6,0.8}$)	0.0069	0.0022	0.0024	0.0011	0.0010	0.0008	0.0005
$\hat{R}_{2,4}^{0.9,0.6}$	0.0269	0.0260	0.0254	0.0249	0.0246	0.0246	0.0245
$\tilde{R}_{2,4}^{0.9,0.6}$	0.0342	0.0301	0.0290	0.0271	0.0261	0.0257	0.0252
Bias($\hat{R}_{2,4}^{0.9,0.6}$)	-0.0026	-0.0016	-0.0010	-0.0006	-0.0003	-0.0002	-0.0001
Bias($\tilde{R}_{2,4}^{0.9,0.6}$)	-0.0098	-0.0057	-0.0046	-0.0027	-0.0017	-0.0013	-0.0008

Table 6: Values of $R_{1,3}^{NP|a_1,a_2,a_3,b}$ for $\lambda_1 = 0.004$ and $\lambda_2 = 0.002$

a_1	1	1	8	27	40	40	95	100	100
a_2	3	5	14	60	40	42	98	100	100
a_3	7	7	28	90	40	47	100	100	100
b	5	3	19	43	30	30	110	110	180
$R_{1,3}^{NP a_1,a_2,a_3,b}$	0.532	0.535	0.532	0.540	0.546	0.549	0.510	0.509	0.468

7. REAL DATA ANALYSIS

In this section the numerical results of the parameters estimation for a real data set with Exponential distribution have been presented. This data set was used for the first time by [16] and can be find in it. Also, it have been used by many other authors, e.g., [17], [18] and [19]. These data present the tensile properties of the jute fibres at different gauge lengths 5, 10, 15 and 20 mm which measured in MPa. The data sets corresponding to the breaking strength of jute fibres with 10mm and 15mm gauge lengths have been considered as the stresses measurement and 20mm in gauge lengths, which represents the strength measurement.

Each data has been separately fitted to the some Exponential distribution and examined by using the Kolmogorov-Smirnov goodness-of-fit test, the results have been reported in Table 7. The Kolmogorov-Smirnov statistics and the corresponding P-values indicate that the Exponential distribution fits the data sets. The estimation of MCCSSP for different values of a and b by MLE, nonparametric methods and Bayesian approach assuming $\alpha_1 = 2, \beta_1 = 3, \alpha_2 = 5, \beta_2 = 4$ for parameters of prior distributions have been presented in Table 8. The estimation of MCCSSP for different values of a_1, a_2 and b by nonparametric methods have been presented in Table 9. The estimation of (4) for $a_1 = 0$ or $a_2 = 0$ by nonparametric methods have been presented in Table 10. The data set consisting of the breaking strength of jute fiber 5 mm in gauge length have been fitted with the Normal distribution with mean 384.37 and standard deviation 188.77 using the Kolmogorov-Smirnov goodness-of-fit test. For this data, the Lilliforce significance correction criteria (modified Kolmogorov-Smirnov test to check the normality of the data) and the P-value are 0.143 and 0.122. Note that by adding this length to the model, the assumption of exponentially for all stresses fails and the MLE method may not be employed. The nonparametric estimators of MCCSSP for real data and different values of a_1, a_2, a_3 and b have been presented in Table 11 where X_1 has Normal distribution, X_2 and X_3 have Exponential distribution.

Table 7: Estimate of parameters, K-S test for strength of jute fiber data

data	Mean	$\hat{\lambda}$	K-S	p-value
10 mm	365.72	0.0027	0.958	0.317
15 mm	367.87	0.0027	0.999	0.271
20 mm	340.74	0.0029	0.727	0.666

Table 8: Values of estimates of MCCSSP for real data

a	30	45	45	78	85	100	220
b	25	50	40	90	75	80	245
$\hat{R}_{1,2}^{a,b}$	0.7200	0.6680	0.6893	0.6432	0.6280	0.6288	0.5996
$\tilde{R}_{1,2}^{a,b}$	0.7431	0.6737	0.6458	0.6207	0.6477	0.5970	0.6151
$R_{1,2}^{NP a,b}$	0.6744	0.6760	0.6886	0.6462	0.6485	0.6485	0.6944

Table 9: Values of $R_{1,2}^{NP|a_1,a_2,b}$ for real data

a_1	20	30	90	180	202	200	300
a_2	40	100	70	170	200	250	280
b	30	70	80	175	201	225	290
$R_{1,2}^{NP a_1,a_2,b}$	0.674	0.640	0.654	0.755	0.755	0.694	0.760

Table 10: Values of $R_{1,2}^{NP|(z),a,b}$ for real data

a_1	0	0	90	160	0	190	0
a_2	30	50	0	0	150	0	280
b	45	35	40	145	160	255	290
$R_{1,2}^{NP a_1,a_2,b}$	0.663	0.670	0.679	0.608	0.663	0.617	0.640

Table 11: Values of $R_{1,3}^{NP|a_1,a_2,a_3,b}$ for real data

a_1	10	42	80	111	150	215	300
a_2	30	58	90	121	160	221	400
a_3	60	71	100	171	170	240	100
b	34	54	85	154	165	220	340
$R_{1,3}^{NP a_1,a_2,a_3,b}$	0.730	0.736	0.705	0.750	0.859	0.625	0.750

8. CONCLUSION

The MCCSSP ($R_{s,k}^{a,b}$) as an appropriate extension of multi-component stress-strength parameter has been introduced. A general formula for computing $R_{s,k}^{a,b}$ in the case of continuous RV's has been presented. The maximum likelihood estimator of $R_{s,k}^{a,b}$ for Exponential distribution has been estimated. The asymptotic distribution of maximum likelihood estimator has been obtained and been used to obtain asymptotic confidence intervals of $R_{s,k}^{a,b}$. A Formula for estimating the MCCSSP by nonparametric method has also been presented. Some numerical computation and simulation studies have been done for illustrating the inferential procedures.

In the past decades, a lot of researches have been done for studying the behavior of reliability function in multi-component stress-strength models, many of similar works can be done for the conditional case. As an specific idea, $R_{s,k}^{a,b}$ can be obtained and estimated for other distributions. As another idea, one may interested in the amounts of information which are measurable, lost, unpredictable, etc.

DECLARATIONS

- **Authors Contribution:** All parts of this study has been done jointly by both authors, unless the simulation and graphical study which has been compiled by M.Taheri and the final edition which has been done by K.Khorshidian.
- **Competing Interest:** There is no conflict of interest between authors.
- **Availability of Data and Materials:** Please contact M.Taheri, taherisaifmorteza@gmail.com in order to request any data corresponding to the simulation or graphical subsection.

- **Funding:** No funding was obtained for this study.

REFERENCES

- [1] Church, J. D. and Harris, B. (1970). The estimation of reliability from stress-strength relationships. *Technometrics*, 12(1): 49–54.
- [2] Kotz, S., Lumelski, Y. and Pensky, M. *The Stress-Strength Model and its Generalizations: Theory and Applications*, World Scientific, Singapore, 2003.
- [3] Jose, J. K. (2022). Estimation of stress-strength reliability using discrete phase type distribution. *Communications in Statistics-Theory and Methods*, 51(2): 368–386.
- [4] Khan, A. H., Jan, T. R. (2020). Estimation of stress-strength reliability model using finite mixture of M-transformed Exponential distributions. *Reliability: Theory and Applications*, 15(2): 90–103.
- [5] Varghese, A. K., Chacko, V. M. (2022). Estimation of stress-strength reliability for Akash distribution. *Reliability: Theory and Applications*, 17(3): 52–58.
- [6] Chauhan, K. S. (2022). Estimation and testing procedures of P ($Y < X$) for the inverse distributions family under type-II censoring. *Reliability: Theory and Applications*, 17(3): 328–339.
- [7] Xavier, T., Jose, J. K. (2021). A study of stress-strength reliability using a generalization of power transformed half-logistic distribution. *Communications in Statistics-Theory and Methods*, 50(18): 4335–4351.
- [8] Saber, M. M. and Khorshidian, K. (2021). Introduction to reliability for conditional stress-strength parameter. *Journal of sciences, Islamic Republic of Iran*, 32(4):349–357.
- [9] Bhattacharya, G. k. and Johnson, R. A. (1974). Estimation of reliability in a multicomponent stress-strngth model. *Journal of American statistical association*, 69(348): 966–970.
- [10] Cetinkaya C. (2021). Reliability estimation of the stress-strength model with non-identical jointly type-II censored Weibull component strengths. *Statistical Computation and Simulations*, 91(14): 2917–2936.
- [11] Goswami, A., Seal, B. (2022). Stress-strength Reliability for Equi-correlated Multivariate Normal and its estimation. *Reliability: Theory and Applications*, 17(4): 249–267.
- [12] Awodutire, P. O., Xavier, T., Jose, J. K. (2022). Inferences on stress-strength reliability in multi-component system for type I generalized half-Logistic distribution. *Reliability: Theory and Applications*, 17(1): 18–32.
- [13] Jana, N. and Bera, N. (2022). Interval estimation of multi-component stress-strength reliability based on inverse Weibull distribution. *Mathematics and Computers in Simulation*, 92(4):667–704.
- [14] Kotb, M. S. and Raqab, M. Z. (2021). Estimation of multi-component stress-strength model based on modified Weibull distribution. *Statistical Papers*, 62(6): 2763–2797.
- [15] Saini, S., Tomer, S. and Garg, R. (2022). On the reliability estimation of multi-component stress-strength model for Burr XII distribution using progressively first-failure censored samples. *Statistical Computation and Simulation*, 191: 95–119.
- [16] Xia, Z. P., Yu, J. Y., Cheng, L. D., Liu, L. F. and Wang, W. M. (2009). Study on the breaking strength of Jute Fibers using modified Weibull distribution. *Compus-A*, 40:54–59.
- [17] Chaturvedi, A., Kumar, N. and Kumar, K. (2018). Statistical inference for the reliability functions of a family of lifetime distributions based on progressive type II right censoring. *Statistica*, 78(1):81–101.
- [18] Hassan, A. S., Naqy, H. F., Muhammad, H. Z. and Saad, M. S. (2020). Estimation of multi-component stress-strengt reliability following weibull distribution based on upper record values. *Journal of Taibah University for science*, 1(14):244–253.
- [19] Rao, G. S., Mbwambo, S. and Pak, A. (2018). Estimation of multi-component stress-strength reliability from exponentiated inverse Rayleigh distribution. *Journal of Statistics and Management Systems*, 1–20.

Bayesian Analysis of ARMA and BSTS models for COVID-19 data using R and Stan

MUHAMMED NAVAS T AND ATHAR ALI KHAN



Department of Statistics and Operations Research

Aligarh Muslim University, Aligarh-202002, India

navasmku@gmail.com

atharkhan1962@gmail.com

Abstract

This study compares the performance of Bayesian ARIMA and BSTS models for COVID-19 data using Bayesian approach. Many studies in the literature have compared the BSTS model and classical ARIMA models for infectious disease modelling, and the BSTS model performs well. Apart from the literature, this study is trying to prove the Bayesian ARIMA model gives a better result than the BSTS model. This study uses a different modelling and model comparison method to compare widely used autoregressive integrated moving average (ARIMA) models with their Bayesian structural time series (BSTS) models for COVID-19 data using the Bayesian approach. It is essential to find the order of the ARIMA model before doing bayesian analysis. We find the order of the ARIMA model using measurement LOO information criteria , using the Hamiltonian Montecarlo algorithm and rstan estimate the parameters of ARIMA and BSTS models for COVID-19 data. Furthermore, compare both models using Loaic and Waic values; Bayesian ARIMA models outperform in this study.

Keywords: BSTS, COVID-19, ARIMA, MCMC, LOOIC, WAIC, Stan

1. INTRODUCTION

The literature contains a variety of traditional studies that use ARIMA models. The statistical models and techniques for evaluating discrete time series are discussed in [1], along with some of the methodology's most significant applications. The class of autoregressive integrated moving average (ARIMA) models and various extensions of these models are among the models taken into consideration. [2] used the ARIMA model, a version of ARMA, and fitted the same to non-seasonal data by identifying autoregressive and moving average terms with the help of PACF and ACF. [3] describes how to find ARIMA models using the extended autocorrelation function. Over the past few decades, the Bayesian method's significance in econometrics has grown significantly. In this sense, significant references include [4], [5], [6], [7], and [8], etc. [9] clarifies the autoregressive and moving average parameters are implicitly constrained to the stationary and invertible region using a straightforward reparameterisation. [10] offers a few new transformation concepts and looks at how they fit into an effective numerical integration strategy for ARIMA models. The ARMA model can be used to model a variety of data sets due to its universal structure. Typically, the theory does not specify which model should be chosen, so it must be chosen from among a variety of competing models. The choice of an ARIMA model is vital for both statistical inference and prediction.

The bayesian structural time series (BSTS) model, which is based on [11] technique, is another model we used in this work. Create numerous layers utilising this model, such as trends, seasonality, and regression components. In addition, unlike with traditional ARIMA models,

stationary time series are not required for BSTS models, since the BSTS model can deal with structural changes in time series. In contrast to classical ARIMA and machine learning models, a notable advantage of the BSTS model is its distinct interpretable structure in both observable and unobservable dynamic components.

2. METHODOLOGY

In many studies comparison between classical ARIMA and BSTS models for COVID-19 data, the BSTS model performs well in these studies [12, 13, 14]. Apart from the literature, this study is trying to prove the Bayesian ARIMA model gives a better result than the BSTS model for COVID-19 data. The study's goal is to provide a Bayesian approach to assess ARIMA models and the BSTS model, utilizing the Hamiltonian Monte Carlo (HMC) algorithm and the programming language Rstan. This is being done for the COVID-19 cumulative number of cases and the cumulative number of deaths from the date 01 March 2020, to 30 June 2021. Performing this study in two steps; first, for the analysis of Bayesian ARIMA models, it is important to find the order of the ARIMA model. We find the optimized order using the LOOIC value, and the model with the lowest LOOIC value is considered the best model [15]. In the second step, using various priors, we estimate the parameters of the BSTS model and the Bayesian ARIMA model. In the following phase, we compare the BSTS model and the Bayesian ARIMA model using various measures for the corresponding data, such as the looic and waic values [16]. The model with the lowest looic and waic values is considered the best model [17].

3. ARIMA MODEL

ARMA model with AR coefficient ϕ_i 's and MA coefficient θ_j 's is defined as

$$y_t = \mu_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \tag{1}$$

in equation 1, p and q are order of ARMA model, y_t 's are the data points collected over time, μ_0 is the intercept, and ϵ 's are the error terms distributed according to IID normal variate with mean zero and variance σ^2 . Equation 1 is denoted by ARIMA (p,q) model. Invertibility enables us to estimate the noise (ϵ_t) recursively by

$$\hat{\epsilon}_t = y_t - \mu_0 - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j}, \quad t = 1, 2, \dots, T \tag{2}$$

As an approximation, we consider the starting values $y_t = \epsilon_t = 0$ for $t \leq 0$. Choosing $p + q \leq 2$ greatly simplifies the model, and it is frequently observed that higher order ARMA models are more difficult to justify in practise. (see, for example, [6]). A careful review of the literature also suggests that the problems of stationarity and invertibility can be easily tackled for small p and q , say $p + q \leq 2$.

Advantages of Bayesian ARIMA models:

- Flexibility: The Bayesian ARIMA model can handle a wide range of time series patterns, such as non-stationary, multi-seasonal, and multi-trend data.
- Better forecasting: The Bayesian approach to ARIMA models allows for more accurate predictions by taking into account uncertainties in the model parameters.
- Model selection: The Bayesian framework enables the use of model selection methods, such as LOOIC, which helps to determine the optimal number of AR and MA terms in the model.
- Prior information: The Bayesian ARIMA model allows for incorporating prior information or domain knowledge into the model, making it more robust and reliable.

- Model uncertainty: The Bayesian ARIMA model provides a quantification of model uncertainty, which can be useful in decision-making processes.
- Model comparison: The Bayesian ARIMA model allows for comparison with other models, and selection of the best model based on a range of criteria.

3.1. Bayesian Model Formulation of ARMA model

Let $y : y_1, y_2, \dots, y_T$ be the time series observations from a strictly stationary and invertible ARMA model 1. To write the approximate likelihood function for an ARMA(p, q) model with y_t 's are the time series observed data is

$y_t \sim N((\mu_0 + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}), \sigma^2)$ under the assumption of $y_t = \epsilon_t = 0$ for $t \leq 0$. The conditional density of y_t given $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ can then be written as,

$$f(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}; \mu_0, \Phi, \Theta) \propto \left(\frac{1}{\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2} \left(y_t - \mu_0 - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j}\right)^2\right). \tag{3}$$

The likelihood function is defined as:

$$f(\underline{y} | \mu_0, \Phi, \Theta) \propto \prod_{t=p+1}^T f(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}; \mu_0, \Phi, \Theta) \tag{4}$$

which reduces to:

$$f(\underline{y} | \mu_0, \Phi, \Theta) \propto \left(\frac{1}{\sigma^2}\right)^{(T-p)/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=p+1}^T \left(y_t - \mu_0 - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j}\right)^2\right) \tag{5}$$

4. BSTS MODEL

The BSTS model is defined by a pair of equations,

$$\begin{aligned} y_t &= Z_t^T \alpha_t + \epsilon_t, \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, \end{aligned} \tag{6}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and $\eta_t \sim \mathcal{N}(0, Q_t)$, both error terms are independent of all other unknowns [11]. In this study, we are using a local linear trend model (a BSTS model without regression components) that assumes level and slope as random walk components.

4.1. Local linear trend model

The local linear trend model is a popular option for trend modelling because it responds rapidly to local variation, which is important when making short-term forecasts. The equation as follows:

$$y_t = \mu_t + \epsilon_t \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{7}$$

the equation of the level component is:

$$\mu_{t+1} = \mu_t + \delta_t + \eta_t \quad \eta_t \sim N(0, \sigma_\eta^2) \tag{8}$$

and the equation of the slope is:

$$\delta_{t+1} = \delta_t + \zeta_t \quad \zeta_t \sim N(0, \sigma_\zeta^2) \tag{9}$$

The local linear trend is based on the supposition that both the mean and slope components follow random walks.

5. BAYESIAN ESTIMATION USING R AND STAN

5.1. Bayesian estimation of ARIMA models

In this session, use the corresponding priors to express the following stan code for the ARIMA(1,0,1) model in R:

$$\begin{aligned}\mu_0 &\sim \text{studentt}(0, 1, 6) \\ \phi_1 &\sim \text{normal}(0, 0.5) \\ \theta_1 &\sim \text{normal}(0, 0.5) \\ \sigma &\sim \text{studentt}(0, 1, 7)\end{aligned}$$

For all ARIMA models, we use the same priors for μ and σ ; for all AR components, we use prior as $\text{normal}(0, 0.5)$, and for all MA components, we use prior as $\text{normal}(0, 0.5)$. Stan code for the ARIMA(1,0,1) model is given below:

```
armamodel="data {
int<lower=1> N;
real y[N];
}
parameters {
real mu0;
real phi1;
real theta1;
real<lower=0> sigma;
}
model {
vector[N] nu;
vector[N] err;
psi[1] = mu0 + phi1 * mu0;
err[1] = y[1] - psi[1];
for ( n in 2:N) {
psi[n] = mu0 + phi1 * y[n-1] + theta1 * err[n-1];
err[n] = y[n] - psi[n];
}
mu0 ~ studentt(0, 1, 6);
phi1 ~ normal(0, 0.5);
theta1 ~ normal(0, 0.5);
sigma ~ studentt(0, 1, 7);
err ~ normal(0, sigma);
}
"
```

Only the ARIMA(1,0,1) model code is expressed here; other types of ARIMA models were used in this study but were not displayed owing to space limitations.

5.2. BSTS models

In this session, the Stan formulation of the Bayesian structural time series model is discussed. The appropriate priors for the parameters are discussed below:

$$\begin{aligned}u_{err} &\sim N(0, 1) \\ v_{err} &\sim N(0, 1) \\ \sigma_{slope} &\sim N(0, 0.5)\end{aligned}$$

$$\sigma_{level} \sim N(0, 0.5)$$

$$\sigma_{obs} \sim N(5, 10)$$

Stan code for BSTS model is:

```

\label{stan:2}
bstsmodel="
data {
  int <lower=0> N;
  vector[N] y;
}
parameters {
  vector[N] a_err;
  vector[N] b_err;
  real beta;
  real <lower=0> sigma_obs;
  real <lower=0> sigma_slope;
  real <lower=0> sigma_level;
}

transformed parameters {
  vector[N] a;
  vector[N] b;
  a[1] = a_err[1];
  b[1] = b_err[1];
  for (n in 2:N) {
    a[n] = a[n-1] + b[n-1] + s_level*u_err[n] ;
    b[n] = b[n-1] + s_slope*v_err[n] ;
  }
}

model {
  a_err ~ normal(0,1);
  b_err ~ normal(0,1);
  sigma_slope~normal(0,0.5);
  sigma_level~normal(0,0.5);
  sigma_obs~normal(5,10);
  for(n in 1:N){
    y[n] ~ normal (a[n] ,sigma_obs);
  }
}

generated quantities{
  vector[N] yrepg;
  real log_lik[N];
  for(n in 1:N){log_lik[n]=normal_lpdf(y[n] |a[n] ,sigma_obs);
}
}
"

```

In this chapter for the BSTS model, we apply the same priors and stan code for the remaining data sets; when we use the same priors for all data sets, we can more precisely compare our looic and waic values.

6. RESULTS AND DISCUSSIONS

During this session, we reviewed the estimated parameters obtained after executing Stan code for both Bayesian ARIMA and BSTS models.

6.1. ARIMA model Bayesian estimation

6.1.1 ARIMA model selection

Finding the AR component, MA component, and order of differencing for the model is the first step before using an ARIMA model. Table 1 lists various p and q order ARIMA models and their associated LOOIC values. We assess alternative order ARIMA models with LOOIC values for the cumulative number of cases and the cumulative number of deaths for each of the five countries. We evaluate the model with the lowest LOOIC value as the ideal model, and we employ that model in further bayesian analysis. For instance, from table 1, take the cumulative number of cases in the USA, the minimum LOOIC value is 10720.14 for the order of ARIMA (1, 2, 1), so we find our ARIMA model using the minimum LOOIC value. Similarly, for the cumulative number of deaths in the USA, the minimum LOOIC value is 7000.51 for the model ARIMA (3, 2, 2), so we use this model in further Bayesian analysis. Similarly, we ordered ARIMA for further Bayesian analysis for all other countries' cumulative cases and deaths.

Table 1: ARIMA model selection

Countries	Cases model	AIC value	Deaths model	AIC value
USA	ARIMA(2,,2,2)	10721.76	ARIMA(2,2,2)	7007.206
	ARIMA(0,2,1)	10730.94	ARIMA(1,2,2)	7176.245
	ARIMA(1,2,2)	10721.83	ARIMA(2,2,1)	7131.75
	ARIMA(2,2,1)	10722.18	ARIMA(3,2,2)	7000.51
	ARIMA(1,2,1)	10720.14	ARIMA(3,2,1)	7115.77
	ARIMA(2,2,0)	10751.45	ARIMA(3,2,3)	7012.184
UK	ARIMA(2,2,2)	8921.798	ARIMA(2,2,2)	6181.108
	ARIMA(0,2,0)	8935.118	ARIMA(1,2,2)	6204.96
	ARIMA(1,2,0)	8919.275	ARIMA(2,2,1)	6237.94
	ARIMA(0,2,1)	8917.05	ARIMA(5,2,2)	5801.33
	ARIMA(1,2,1)	8918.23	ARIMA(5,2,1)	5968.22
	ARIMA(1,2,2)	8920.127	ARIMA(4,2,1)	6114.78
UAE	ARIMA(0,2,0)	6906.30	ARIMA(2,2,2)	1987.16
	ARIMA(1,2,0)	6788.24	ARIMA(0,2,1)	1984.84
	ARIMA(0,2,1)	6729.45	ARIMA(1,2,1)	1983.08
	ARIMA(1,2,1)	6732.08	ARIMA(2,2,1)	1985
	ARIMA(0,2,2)	6731.07	ARIMA(1,2,2)	1973.254
Bahrain	ARIMA(2,2,2)	5911.402	ARIMA(2,2,2)	2096.992
	ARIMA(1,2,2)	5908.39	ARIMA(0,2,1)	2096.94
	ARIMA(1,2,1)	5929.98	ARIMA(1,2,1)	2099.94
	ARIMA(2,2,1)	5931.204	ARIMA(1,2,2)	2101.53
India	ARIMA(2,2,2)	10000.67	ARIMA(2,2,2)	6895.87
	ARIMA(2,2,1)	10033.83	ARIMA(1,2,1)	6895.35
	ARIMA(3,2,2)	9941.79	ARIMA(0,2,1)	6892.32
	ARIMA(3,2,1)	9990.04	ARIMA(1,2,2)	6895.8

Using the above mentioned stan code in 5.1 and using different priors to estimate unknown parameters shown in table 2. Figure 1a and 1b show the posterior density plot and trace plot of the ARIMA model for the cumulative number of UK cases and the cumulative number of Bahrain

deaths, respectively. Due to space restrictions, here is the convergence plot of two countries: the UK and Bahrain are only displayed.

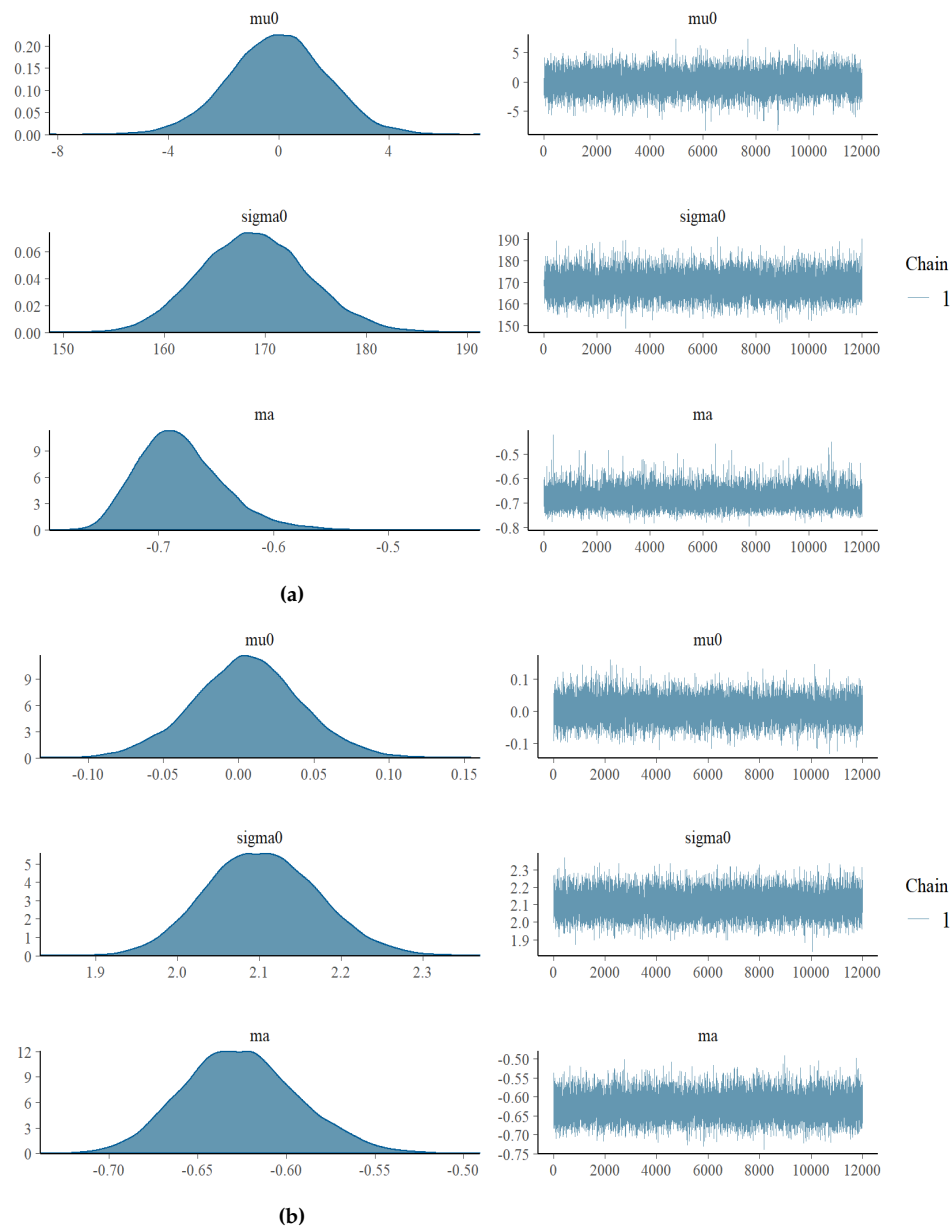


Figure 1: (a) ARIMA model for UK cases Posterior density plot, and Traceplot (b) ARIMA model for bahrain deaths Posterior density plot, and Traceplot

Table 2 displays the anticipated outcomes for the ARIMA models. For all countries, it is simple to verify that the majority of estimated parameters across the ARIMA models are significantly different from zero. Let's use the results for deaths in the UK as an example. Except for μ_0 , all other parameters are statistically significant in the ARIMA(5,2,2) model. This model's parameters have a 95% credible interval that excludes zero and demonstrates the statistical significance of the estimate. Additionally, the model's ar.1 and ma.2 components contribute favourably, but ar.2, ar.3, ar.4, and ar.5 components have statistically significant negative contributions to the ARIMA(5, 2, 2) model. For the USA's cumulative number of cases, the recommended model is ARIMA (1, 2, 2). There is one AR component, two MA components, and two other parameters, μ_0 and σ_0 . All five parameters are statistically significant, which means their 95% credible interval does not contain

zero, and the first MA component contributes to the model negatively. Similarly, we can interpret other countries' parameters using table 2.

Countries	Items (Models)	μ_0	σ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	θ_1	θ_2
USA	Cases (ARIMA(1,2,1))	0.0545 (0.0286)	15711.35 (4.6476)	0.2375 (0.0006)					-0.6858 (0.0004)	
	Deaths (ARIMA(3,2,2))	0.0367 (0.0209)	360.2125 (0.1044)	0.7407 (0.0004)	-0.4641 (0.0005)	-0.2742 (0.0004)			-0.9896 (0.0001)	0.3402 (0.0004)
UK	Cases (ARIMA(0,2,1))	0.0717 (0.0277)	2433.14 (0.7048)						-0.2125 (0.0004)	
	Deaths (ARIMA(5,2,2))	0.0263 (0.0211)	95.9465 (0.0277)	0.1916 (0.0004)	-0.8767 (0.0003)	-0.1517 (0.0005)	-0.4477 (0.0003)	-0.5309 (0.0004)	-0.6963 (0.0004)	0.6722 (0.0003)
UAE	Cases (ARIMA(0,2,1))	1.2845 (0.0371)	252.93 (0.1261)						-0.6439 (0.0005)	
	Deaths (ARIMA(1,2,2))	0.0135 (0.0003)	1.8694 (0.0005)	-0.2635 (0.0020)					-0.5254 (0.0021)	-0.1067 (0.0016)
Bahrain	Cases (ARIMA(1,2,2))	0.0278 (0.0279)	108.622 (0.0564)	0.4209 (0.0027)					-0.8067 (0.0026)	0.3154 (0.0010)
	Deaths (ARIMA(0,2,1))	0.0054 (0.0003)	2.1048 (0.0006)						-0.6267 (0.0003)	

Table 2: ARIMA model Bayesian estimation: estimation of the posterior means.

6.2. BSTS model Bayesian estimation

From table 3, the majority of the posterior estimates for the BSTS model are statistically significant, which is consistent with the results from the BSTS models. For UK deaths, for instance, a posterior parameter estimate of σ_{obs} with a range of 142.04 to 155.34 implies that zero excludes from a credible interval of 95 %, indicating that the parameter is statistically significant. Similarly, for parameter σ_{slope} and σ_{level} , UK deaths have statistical significance. Similarly, for parameter σ_{slope} and σ_{level} , UK deaths have statistical significance. The parameter value of σ_{slope} is 0.02, and σ_{level} is 0.38 for the UK cumulative number of deaths. We can provide a similar justification for the parameter estimate for the other four countries. Figure 2 to 9 show the autocorrelation plot, caterpillar plot, posterior density plot, and trace plot of the BSTS model for the UK cumulative number of cases and deaths, respectively. From the trace plot, we can interpret the convergence of the Markov chain, due to space restrictions not displaying the plots of all countries.

Table 3: BSTS model Bayesian estimation: estimation of the posterior means.

Countries	Items	σ_{obs}	σ_{slope}	σ_{level}
USA	Cases	1948.43 (0.04)	3.25 (0.001)	0.40 (0.001)
	Deaths	284.66 (0.07)	0.03 (0.001)	0.41 (0.02)
UK	Cases	1239.84 (0.10)	11.51 (0.01)	0.39 (0.01)
	Deaths	148.45 (0.04)	0.02 (0.001)	0.38 (0.001)
UAE	Cases	218.46 (0.07)	0.03 (0.001)	0.40 (0.01)
	Deaths	2.41 (0.08)	0.02 (0.001)	0.02 (0.001)
Bahrain	Cases	107.37 (0.08)	0.03 (0.001)	0.41 (0.02)
	Deaths	2.50 (0.08)	0.01 (0.001)	0.02 (0.001)
India	Cases	1240.02 (0.10)	11.53 (0.01)	0.40 (0.01)
	Deaths	233.30 (0.09)	0.04 (0.001)	0.39 (0.001)

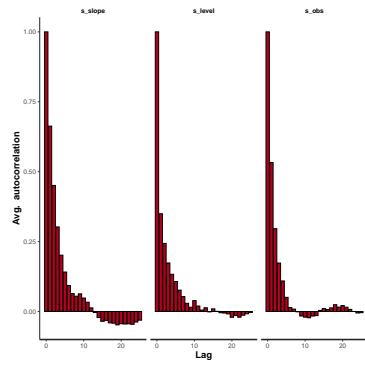


Figure 2: BSTS model for UK cases autocorrelation plot

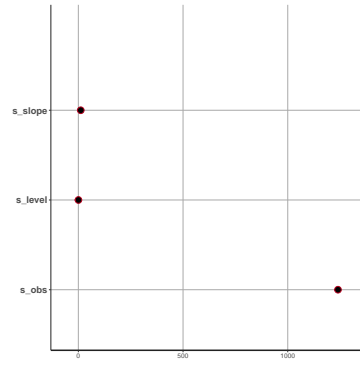


Figure 3: BSTS model for UK cases caterpillar plot

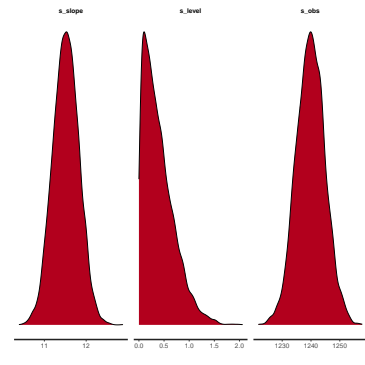


Figure 4: BSTS model for UK cases posterior density plot

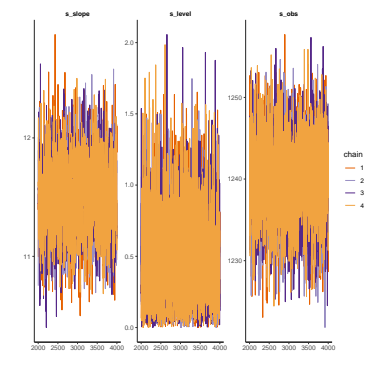


Figure 5: BSTS model for UK cases trace plot

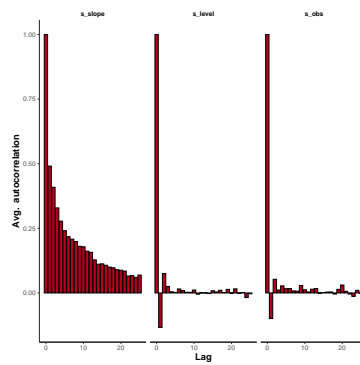


Figure 6: BSTS model for UK deaths autocorrelation plot

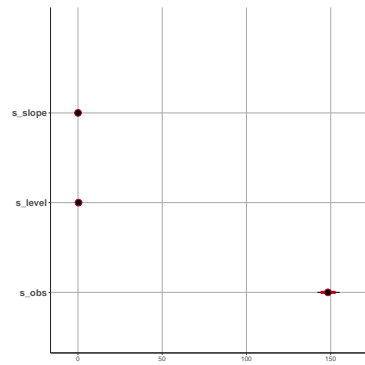


Figure 7: BSTS model for UK deaths caterpillar plot

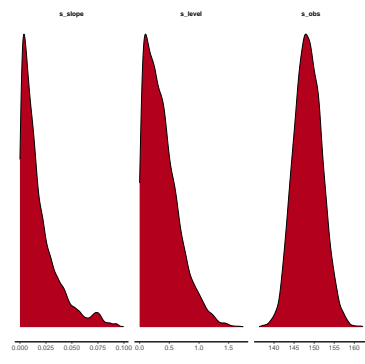


Figure 8: BSTS model for UK deaths posterior density plot

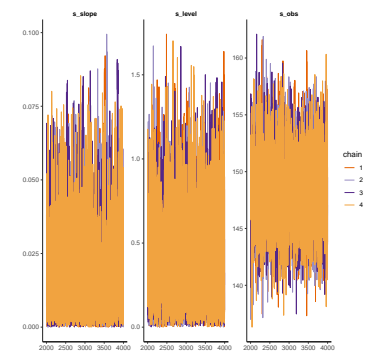


Figure 9: BSTS model for UK deaths trace plot

7. CONCLUSION

In this study, we used a Bayesian estimation to evaluate the widely used ARIMA and BSTS models while modelling COVID-19 cumulative number of cases and the cumulative number of deaths in five countries. From table 4, we discover that the ARIMA model LOOIC value is lower than that of their corresponding BSTS models. The model with the lowest LOOIC value is considered to be the best model. For example, Bahrain’s cumulative number of cases LOOIC value for the Bayesian ARIMA model is 5945.3 and for the BSTS model is 6028.5. Here Looic value is the minimum for the ARIMA model and which is the best model. Similarly, for Bahrain cumulative number of deaths LOOIC value for the Bayesian ARIMA model is 2110.7, and for the BSTS model is 2273.2, therefore the minimum LOOIC value is for the ARIMA model. Therefore

almost all countries' data set LOOIC value is minimum for the Bayesian ARIMA model. Hence, when compared to the corresponding BSTS model, the Bayesian ARIMA model performs the best.

Table 4: ARIMA and BSTS model validation using LOOIC and WAIC.

		Loaic	Waic	Loaic	Waic
USA	Cases	10771.5	10771.6	47343.5	47887.2
	Deaths	7221.3	7260	7665.8	7665
UK	Cases	8958.4	8958.3	24255.4	24362.8
	Deaths	5831.3	5831.2	6448.6	6448.6
UAE	Cases	6786	6800.9	7131.2	7180.2
	Deaths	1993.4	1993.5	2236	2236
Bahrain	Cases	5945.3	5945.3	6028.5	6028.5
	Deaths	2110.7	2110.9	2273.2	2273.2
India	Cases	10095.5	10096.1	24219.5	24326.3
	Deaths	6981.2	6983.5	7279.4	7361.7

REFERENCES

- [1] BOX, G. (1976). Time series analysis, forecasting and control.
- [2] Tsay, R. S. and Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary arma models. *Journal of the American Statistical Association*, 79(385):84–96.
- [3] Pankratz, A. (2012). *Forecasting with dynamic regression models*. John Wiley & Sons.
- [4] Zellner, A. (1971). An introduction to bayesian inference in econometrics. Technical report.
- [5] Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2):183–206.
- [6] Marriott, J., Ravishanker, N., Gelfand, A. E., and Pai, J. (1996). Bayesian analysis of arma processes: Complete sampling-based inference under exact likelihoods. In *Bayesian analysis in statistics and econometrics: essays in honor of Arnold Zellner*, pages 243–256.
- [7] Kleibergen, F. R. and Hoek, H. (2000). Bayesian analysis of arma models.
- [8] Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*, volume 20. Springer.
- [9] Monahan, J. F. (1984). A note on enforcing stationarity in autoregressive-moving average models. *Biometrika*, 71(2):403–404.
- [10] Marriott, J. and Smith, A. (1992). Reparametrization aspects of numerical bayesian methodology for autoregressive moving-average models. *Journal of Time Series Analysis*, 13(4):327–343.
- [11] Scott, S. L. and Varian, H. R. (2014). Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2):4–23.
- [12] Navas Thorakkattle, M., Farhin, S., et al. (2022). Forecasting the trends of covid-19 and causal impact of vaccines using bayesian structural time series and arima. *Annals of Data Science*, 9(5):1025–1047.

- [13] Feroze, N. (2020). Forecasting the patterns of covid-19 and causal impacts of lockdown in top five affected countries using bayesian structural time series models. *Chaos, Solitons & Fractals*, 140:110196.
- [14] Zhang, Y. and Fricker, J. D. (2021). Quantifying the impact of covid-19 on non-motorized transportation: A bayesian structural time series model. *Transport policy*, 103:11–20.
- [15] Voss, M. S. and Feng, X. (2002). Arma model selection using particle swarm optimization and aic criteria. *IFAC Proceedings Volumes*, 35(1):349–354.
- [16] Vehtari, A., Gelman, A., and Gabry, J. (2015). Efficient implementation of leave-one-out cross-validation and waic for evaluating fitted bayesian models. *arXiv preprint arXiv:1507.04544*.
- [17] Khan, Athar Ali and others (2022). Bayesian Analysis of Type II Generalized Topp–Leone Accelerated Failure Time Models Using R and Stan. *Reliability: Theory & Applications*, 17(4 (71)), 477-493.

Analysis of projected profit in an M/M/K Encouraged Arrival Queueing model using Chi-Square Test

V. Narmadha¹, P. Rajendran²

^{1,2} Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology,
Vellore, Tamil Nadu 632014, India.

¹narmadha.v@vit.ac.in, ²prajendran@vit.ac.in

* Correspondence: prajendran@vit.ac.in

Abstract

Nowadays, queues be seen in fast food restaurants and in all service-based businesses. This study is a mathematical analysis of such business firms with the help of Queueing Theory . The discounts and promotions entice customers to the firm and in this study such attracted customers are referred to as Encouraged Arrivals. The Chi-square test is used to determine the kind of encouraged arrival pattern that adheres to the data observed from a fastfood outlet. We introduce the encouraged arrivals in an M/M/k queueing model for the analysis of performance metrics. The performance metrics of the various encouraged arrival patterns are compared and the ideal one is chosen for the firm. The economic analysis shows that with encouraged arrivals, the cost associated with the time lost due to waiting is reduced gradually with increasing number of servers. Thus the firm increases its projected profit with encouraged arrivals. This study helps the entrepreneurs to decide the kind of discounts that would attract the customers simultaneously improving the firm's profit. Little's law is also verified.

Keywords: Encouraged Arrivals (ea), Promotions, Expenses, Chi-square test, Frequency, Degrees of freedom (Dof), Little's formula (LF).

1. Introduction

Queueing theory is one of the earliest and most used quantitative analytic methodology. It involves the study of waiting queues. Everyday activities such as grocery shopping, gas purchases, bank deposits, and many organisations are affected by waiting lines. The Latin word cauda, which means "tail," is where the word queue first originated. Any service system will inevitably have customers wait in line to get services, making queue management a significant task. It is commonly referred to as the theory of congestion, which comes under the branch of operational research examines the connection between the level of demand for a system of services and the delays experienced by its users. The goal of the research of queues is to quantify the phenomenon of standing in queues by employing benchmark performance indicators like average length of the queue, queue's average waiting time and utilization.

Poor service patterns, queue management issues, unhelpful service staff attitudes, subpar

amenities and delivery are widespread in most service-based businesses including restaurants. All of these elements have an impact on customer relationships and overall efficiency. This study aims to analyse the data to increase the efficacy and economy of the organization. The analysis is carried out based on the primary data collected from a fastfood outlet [1].

The objective of this study is to reduce the amount of time consumers spend waiting in lines by using the M/M/k model in a fast food outlet. First in, first out (FIFO) queue discipline is utilised and we have taken into account ear.

A comparison of network tools was performed to figure out the ideal values of performance metrics for flexible manufacturing systems in [2]. Tsarouhas [3] presented a theory to estimate the overall waiting times for processing each pizza at a workstation in a food manufacturing outlet. A model describing the psychological mechanisms that influence the connection between satisfaction and perceived wait time was suggested and tested by McGuire [4]. In [5] according to Mahmoud and Lu, numerous branches of research and engineering use queuing theory and Markov chains as common analysis, evaluation, and decision-making methods. For analysis of the steady-state and transient behaviour, real-world systems could be modelled. According to stochastic replacement intervals, special discounts are taken into account for a specific item from the supplier in [6]. A demand-satisfaction dilemma involving two products that are interchangeable was examined in [7]. The goal was to obtain the order quantity for each product that optimises the combined profit function. Location-inventory models were also taken into account in [8]. They employed a bi-level Markov process(MP) to create a stochastic inventory model. Som et al. [9] studied a multi- server queuing model with limited capacity for any organization encountering ear and reverse reneing. Customers who are drawn into an organisation as a result of special offers (known as Encouraged Arrivals (ear), a term coined by Som and Seth [10]. Jain et al. [11] described the idea of customer mobilisation and stated that a system attracts a new consumer by taking a look at its sizable customer base. ear deals with the percentage change in clients as a result of promotions and discounts. Som et al. [12] studied a multi- server queuing model with limited capacity for any organization encountering ear and reverse reneing. The Banking sector has become the most inevitable part of public units. Most banks make use of common queuing models. An M/M/1 queueing model is used to analyse the ATM's performance in [13]. The performance metrics of a toll plaza is analysed to find the traffic flow and to set up the system in an efficient way in [14]. A review on bulk arrivals [15] helps the researchers to model problems without congestion.

The introduction of the paper is given in Section 1. Section 2 provides the mathematical notation. The proposed mathematical model is given in section 3. In section 4, chi-square test is performed to check the fit of various encouraged arrivals. Analysis of performance metrics and economic analysis is given in section 5. Little's law is verified in section 6. Section 7 wraps up the paper with remarks and conclusion.

2. Mathematical Notation

The proposed queueing model uses the following notations.

- Encouraged arrivals are denoted by 'ear' and frequency of arrivals is denoted by 'fa'. The arrivals happen sequentially according to a Poisson process with the parameter $\lambda(1+ear)$, where "ear" denotes the change of percentage in the total count of clients estimated from observed data. For instance, if a firm previously offered discounts and a percentage change in the total count of clients was noticed of +10%, +30% or +50%, then $ear=0.1, 0.3$ or $ear=0.5$, respectively.
- The model follows an exponentially distributed service times with parameter μ .
- Customers are served in the order of their arrival i.e., FCFS.
- The system has k parallel servers and there is no limit placed on the waiting space in the

system at any time 't'.

- Probability that there is no clients in the system is given by P_{b0} .

3. Mathematical Model

The following diagram depicts the proposed model. We construct an M/M/k model to analyse the performance measures of the encouraged arrivals. The arrival rate is denoted by the parameter λ_n ,

$$\lambda_n = \lambda(1 + ea\eta), \text{ for all } n$$

Where n is the number of customers in the system.

The model follows an exponentially distributed service times with parameter μ and there are k servers in the model. If there are k or more clients, then it is understood that all the k servers are busy.

Hence, the service rate is given by $\mu_n = \begin{cases} n\mu & (1 \leq n \leq k) \\ k\mu & (n \geq k) \end{cases}$

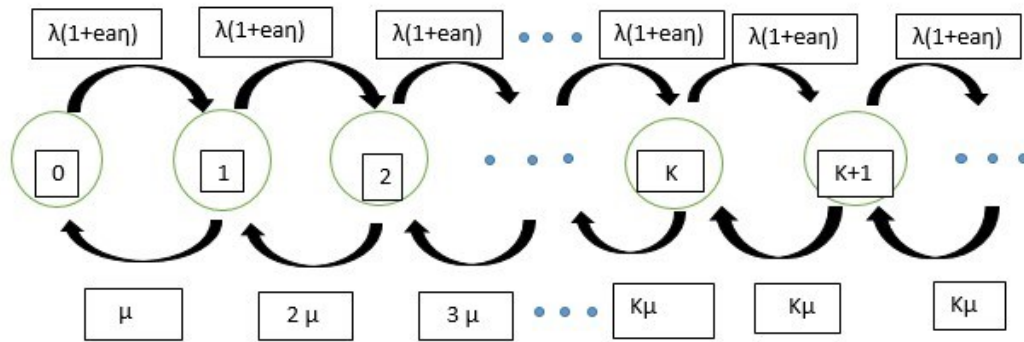


Figure 1. Rate transition diagram of the proposed model

4. Chi-square test to check the goodness of fit

In this study, Chi-square test is employed to determine the encouraged arrival pattern that adheres to the data observed from a fastfood outlet. This particular statistical test was created to examine the consistency between a set of actual frequencies and expected frequencies under the presumption of a hypothesis for the phenomenon under study. The test is employed to determine if two classification attributes are dependent on one another or not. The chi-square formula provides a measurement for the gap between reported and expected frequency which is given by

$$\chi^2 = \sum \frac{(fa - ea)^2}{ea} \tag{1}$$

Here 'fa' denotes the frequency of arrivals and 'ea' denotes the expected arrivals, in our case we consider ea as follows:

i.e., $ea = \lambda(1 + ea\eta)$ when 20% , 30% or 40% $ea\eta$ are offered .

Case 1:

Consider the following hypothesis to check whether the 20% $ea\eta$ fits the observed data [1]

Null hypothesis: The 20% $ea\eta$ fits the distribution

Alternative hypothesis: The 20% $ea\eta$ does not fits the distribution.

Table 1 : Chi-square test to check the goodness of fit of 20% eaη

Number of arrivals	Frequency of arrivals (fa = λ)	ea =λ (1+ 0.2)	$\frac{(fa - ea)^2}{ea}$
0	6	7.2	0.2
1	14	16.8	0.467
2	26	31.2	0.867
3	18	21.6	0.6
4	6	7.2	0.2
5	2	2.4	0.067

The ea should be greater than or equal to 5 for good approximation in a chi-square test. If any of the ea does not satisfy this condition then it has to be combined with some other ea until the condition is satisfied.

Table 1 thus becomes

Table : 1.1

Number of arrivals	Frequency of arrivals (fa = λ)	ea =λ (1+ 0.2)	$\frac{(fa - ea)^2}{ea}$
0	6	7.2	0.2
1	14	16.8	0.467
2	26	31.2	0.867
3	18	21.6	0.6
4 or 5	8	9.6	0.267

$$\sum \frac{(fa - ea)^2}{ea} = 2.401$$

$$\chi^2 = \sum \frac{(fa - ea)^2}{ea} = 2.401$$

The Dof is given by (number of observed data) – (number of parameters to be estimated) -1

Hence, Dof = 3

The value of chi-square at 5% level of significance for 3 Dof from the chi-square distribution table = 7.8147

The calculated value is lesser than table value therefore null hypothesis is accepted.

(i.e.,) The 20% eaη fits the distribution.

Table 2 : To find arrival rate from the 20% eaη

Number of arrivals	Frequency of arrivals (fa = λ)	ea =λ (1+ 0.2)	Number of arrivals × ea
0	6	7.2	0
1	14	16.8	16.8
2	26	31.2	62.4
3	18	21.6	64.8
4	6	7.2	28.8
5	2	2.4	12
			Total = 184.8

Hence the system has the capacity $N= 184.8$

The interarrival time for the N clients is observed as 714 minutes [1]

The time taken to serve the N clients is observed as 1012 minutes [1]

$$\text{Then } \lambda = \frac{N}{\text{inter arrival Time for } N \text{ clients}} = \frac{184.8}{714} = 0.258$$

$$\mu = \frac{N}{\text{Time taken for } N \text{ clients to be served}} = \frac{184.8}{1012} = 0.183$$

Case 2:

Consider the following hypothesis to check whether the 30% eaη fits the observed data [1]

Null hypothesis: The 30% eaη fits the distribution

Alternative hypothesis: The 30% eaη does not fits the distribution

Table 3 : Chi-square test to check the goodness of fit of 30% eaη

Number of arrivals	Frequency of arrivals (fa = λ)	ea = λ (1+ 0.3)	$\frac{(fa - ea)^2}{ea}$
0	6	7.8	0.4153
1	14	18.2	0.9692
2	26	33.8	1.8
3	18	23.4	1.246
4 or 5	8	10.4	0.553
			$\sum \frac{(fa - ea)^2}{ea} = 4.9835$

$$\chi^2 = \sum \frac{(fa - ea)^2}{ea} = 4.9835$$

The Dof is given by (number of observed data) – (number of parameters to be estimated) -1

Hence, Dof = 3

The value of chi-square at 5% level of significance for 3 Dof from the chi-square distribution table= 7.8147

The calculated value is lesser than table value therefore null hypothesis is accepted.

(i.e.,) The 30% eaη fits the distribution.

Table 4 : To find arrival rate from the 30% eaη

Number of arrivals	Frequency of arrivals (fa = λ)	ea =λ (1+ 0.3)	Number of arrivals × ea
0	6	7.8	0
1	14	18.2	18.2
2	26	33.8	67.6
3	18	23.4	70.2
4	6	7.8	31.2
5	2	2.6	13
			Total = 200.2

Let the system has the capacity $N= 200.2$

The interarrival time for the N clients is observed as 714 minutes [1]

The time taken to serve the N clients is observed as 1012 minutes [1]

$$\text{Then } \lambda = \frac{N}{\text{inter arrival Time for } N \text{ clients}} = \frac{200.2}{714} = 0.28$$

$$\mu = \frac{N}{\text{Time taken for } N \text{ clients to be served}} = \frac{200.2}{1012} = 0.197$$

Case 3:

Consider the following hypothesis to check whether the 40% eaη fits the observed data [1]

Null hypothesis: The 40% eaη fits the distribution

Alternative hypothesis: The 40% eaη does not fits the distribution.

Table 5: Chi-square test to check the goodness of fit of 40% eaη

Number of arrivals	Frequency of arrivals (fa = λ)	ea = λ (1+ 0.4)	$\frac{(fa - ea)^2}{ea}$
0	6	8.4	0.685
1	14	19.6	1.6
2	26	36.4	2.97
3	18	25.2	2.057
4 or 5	8	11.2	0.9142
			$\sum \frac{(fa - ea)^2}{ea} = 8.2262$

$$\chi^2 = \sum \frac{(fa - ea)^2}{ea} = 8.2262$$

The Dof is given by (number of observed data) – (number of parameters to be estimated) -1

Hence, Dof = 3

The value of chi-square at 5% level of significance for 3 Dof from the chi-square distribution table= 7.8147

The calculated value is gretae than table value therefore null hypothesis is not accepted.

(i.e.,) The 40% eaη does not fits the distribution.

Remarks :

- Using the chi-square test it is very easy to find the kind of eaη that fits the observed data[1].
- Thus we can infer from the chi-square test that 20% and 30% eaη fits the observed data with which we can further investigate the performance metrics to analyse which is more effective for the firm to increse the projected profit.

5. Analysis of performance metrics

Probability that there are no clients in the system is given by Pb₀

$$\sum_{j=0}^{\infty} Pb_j = 1$$

$$\sum_{j=0}^{k-1} Pb_n + \sum_{j=k}^{\infty} Pb_n = 1$$

$$\sum_{j=0}^{k-1} \frac{1}{j!} \left(\frac{\lambda(1+ea\eta)}{\mu} \right)^j Pb_0 + \frac{\left(\frac{\lambda(1+ea\eta)}{\mu} \right)^k}{k! \left(1 - \frac{\lambda(1+ea\eta)}{k\mu} \right)} Pb_0 = 1$$

$$Pb_0 = \left(\sum_{j=0}^{k-1} \frac{1}{j!} \left(\frac{\lambda(1+ea\eta)}{\mu} \right)^j + \frac{\left(\frac{\lambda(1+ea\eta)}{\mu} \right)^k}{k! \left(1 - \frac{\lambda(1+ea\eta)}{k\mu} \right)} \right)^{-1}$$

$$Pb_0 = \left(\frac{m^k}{k!(1-r)} + \sum_{n=0}^{k-1} \frac{m^n}{n!} \right)^{-1} \text{ where } m = \frac{\lambda(1+ea\eta)}{\mu}, r = \frac{m}{k} \quad (2)$$

Provided that $\frac{m}{k} < 1$

The expected number of clients in the queue is given by L_q

$$L_q = \sum_{j=k+1}^{\infty} (j-k)Pb_j$$

$$L_q = \left(\frac{m^k r}{k!(1-r)^2} \right) Pb_0 \quad (3)$$

Using LF, $L_q = \lambda W_q$

Where W_q is the expected waiting time in queue

$$w_q = \frac{\lambda}{L_q}$$

$$W_q = \frac{\lambda(1+ea\eta)}{\left(\frac{m^k r}{k!(1-r)^2} \right) Pb_0} \quad (4)$$

The expected number of clients in the system is given by $L_s = m + L_q$

The expected waiting time in the system is given by $W_s = \frac{1}{\mu} + W_q$

Expected time lost per day due to waiting = $\lambda(1+ea\eta) \times W_q \times 8$ hours
(The working hours per day is taken as 8 hours) [1]

Expected Cost associated with lost time = $W_q \times \text{Rs. } 50$

(The cost associated with time lost by waiting is taken as Rs. 50) [1]

Table 6: Comparison of performance measures between 20% and 30% eaη

Number of servers	2	3	4
L_q for 20% eaη	1.97	0.3862	0.0948
L_q for 30% eaη	2.033	0.3968	0.0978
W_q for 20% eaη	7.635	1.496	0.367
W_q for 30% eaη	7.26	1.417	0.3492
L_s for 20% eaη	3.379	1.7952	1.503
L_s for 30% eaη	3.454	1.817	1.5188
W_s for 20% eaη	13.099	6.96	5.831
W_s for 30% eaη	12.33	6.493	5.425

Comparing all the performance metrics of eaη with respect to number of servers in Table 6.

Remarks:

- From comparing all the performance metrics of 20% and 30% eaη we observe the following
- In case of expected total count of clients both in queue as well as system (i.e.,) L_q and L_s respectively, 30% eaη increases the count when compared with 20% eaη with varying number of servers.
- In case of waiting time in queue as well as system (i.e.,) W_q and W_s 30% eaη reduces the time spent in waiting when compared with 20% eaη with varying number of servers.

- Therefore we come to a conclusion that in case of performance metrics (i.e.,) expected count of clients and expected waiting time 30% eaη increases the size and at the same time reduces the waiting time.

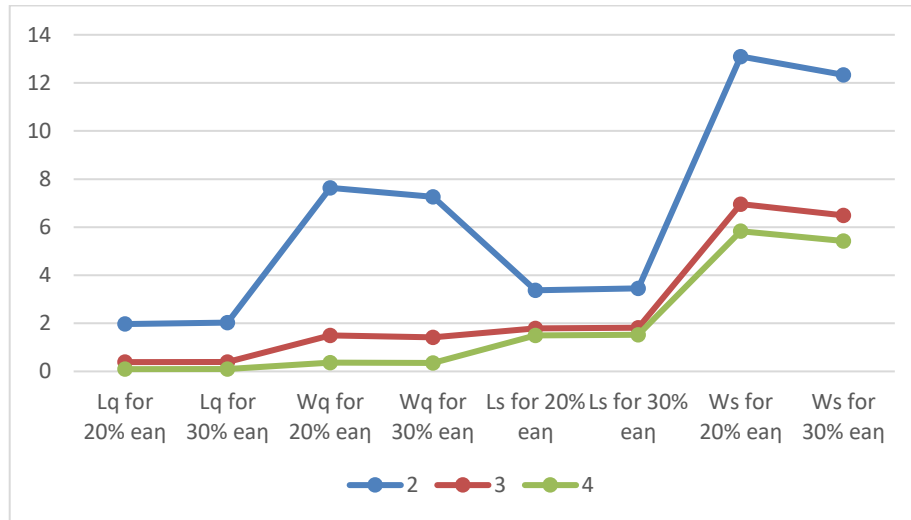


Figure 2. Comparing performance metrics between 20% and 30% eaη with respect to number of servers

5.1. Comparison of expected time lost per day with respect to number of servers

Now, the expected time lost per day for poisson, 20% eaη and 30% eaη with respect to number of servers are compared.

Expected time lost per day due to waiting = $\lambda (1 + ea\eta) \times W_q \times 8$ hours
(The working hours per day is taken as 8 hours) [1]

Table 7 : Calculating Expected lost time for poisson arrival

Number of servers	Expected lost time for poisson arrival	Expected lost time for 20% eaη	Expected lost time for 30% eaη
2	18.7719	15.758	16.2624
3	3.1612	3.0877	3.174
4	0.775	0.7574	0.7822

Calculating the Expected lost time for poisson arrival with respect to number of servers in Table 7.

Remarks:

We infer from the Table 7 that the lost time for poisson is more than that of 20% and 30% eaη. Any firm's aim is to reduce the waiting time thus administering eaη helps us to reduce the amount of time lost in waiting for service.

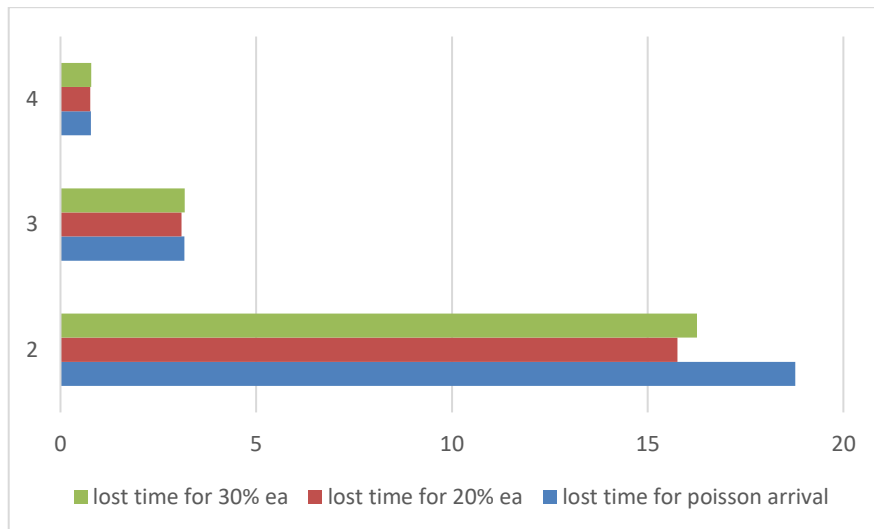


Figure 3. Variation in Expected lost time per day with respect to number of servers

5.2. Comparison of cost associated with lost time per day

Comparing the cost associated with lost time per day for poisson, 20% eaη and 30% eaη with respect to number of servers.

$$\text{Expected Cost associated with lost time} = W_q \times \text{Rs. } 50$$

(The cost associated with time lost by waiting is taken as Rs. 50) [1]

Table 8 : Calculating cost associated with lost time for 20% eaη

Number of servers	$\lambda (1+ 0.2)$	Cost associated with lost time for poisson arrival (in Rs.)	Cost associated with lost time for 20% eaη (in Rs.)	Savings with 20% eaη (in Rs.)
2	0.258	938.597	787.932	150.665
3	0.258	158.06	154.387	3.63
4	0.258	38.875	37.87	1.005

Calculating the cost associated with lost time for 20% eaη and also the cost saved with 20% eaη when compared with λ [1] with respect to number of servers in Table 8.

Table 9 : Calculating cost associated with lost time for 30% eaη

Number of servers	$\lambda (1+ 0.3)$	Cost associated with lost time for poisson arrival (in Rs.)	Cost associated with lost time for 30% eaη (in Rs.)	Savings with 30% eaη (in Rs.)
2	0.28	938.597	813.12	125.477
3	0.28	158.06	158.704	-0.644
4	0.28	38.875	39.11	-0.235

Calculating the cost associated with lost time for 30% eaη and also the cost saved with 30% eaη when compared with λ [1] with respect to number of servers in Table 9.

Remarks:

We infer from the Tables 7,8 and 9 that the cost associated with lost time due to waiting is more for λ [1] than 20% and 30% eaη. While comparing the cost saved with eaη and λ, we observe that among the eaη 20% yields better gain than 30% for the firm which is our primary goal. Thus we conclude that 20% eaη is the best to be offered by the firm as it would increase the organisation’s profit.

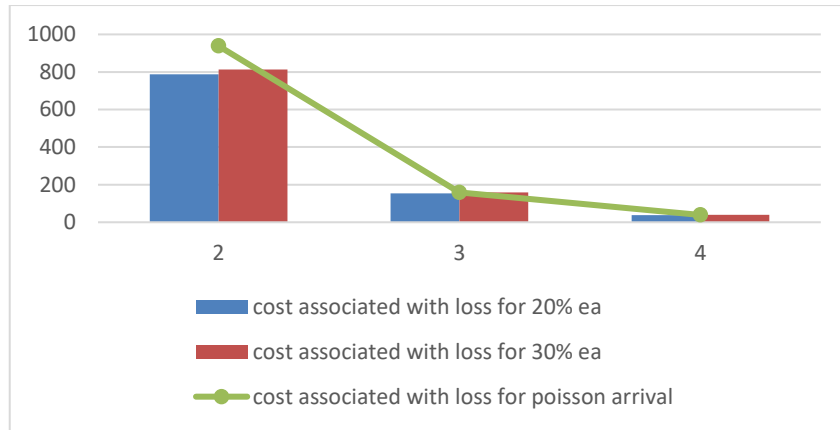


Figure 4. Variation in cost associated with lost time with respect to number of servers

6. Verification of Little Formula(LF)

Using LF, $L_s = \lambda W_s$

Table 10: Verification of LF for 20% eaη with respect to system

Number of servers	$\lambda(1+ ea\eta)$	L_s	W_s	$L_s = \lambda W_s$
2	0.258	3.379	13.099	3.379 = (0.258)(13.099) verified
3	0.258	1.7952	6.96	1.7952 = (0.258)(6.96) verified
4	0.258	1.503	5.831	1.503 = (0.258)(5.831) verified

Table 11: Verification of LF for 30% eaη with respect to system

Number of servers	$\lambda(1+ ea\eta)$	L_s	W_s	$L_s = \lambda W_s$
2	0.28	3.454	12.33	3.454 = (0.28)(12.33) Verified
3	0.28	1.817	6.493	1.817 = (0.28)(6.493) Verified
4	0.28	1.5188	5.425	1.5188 = (0.28)(5.425) verified

The LF is verified for both 20% and 30% ea η in terms of queue as well as system size which shows that the system is well balanced.

7. Conclusion

- Using the chi-square test, it is very easy to find the kind of ea η that fits the observed data[1].
- We infer from the chi-square test that, 20% and 30% ea η fits the observed data with which we can further investigate the performance metrics to analyse and is found to be more effective for the firm to increase the profit.
- While comparing the expenses with ea η and λ , we observe that among the ea η 20% yields better gain than 30% for the firm which is our primary goal.
- We conclude that 20% ea η is the best to be offered by the firm as it would increase the organisation's projected profit.

We infer from the model created that adding one additional server will assist shorten the time clients wait in line and lower the cost associated with it. In order to decrease the time customers must wait to receive services and to lower the expense associated with waiting, we advise the organisation to raise the number of servers to at least three.

With encouraged arrivals the results show that the cost associated with lost time reduces gradually with increasing number of servers than poisson arrivals [1] and we see that 20% ea η is ideal for the proposed model as 20% ea η yields better projected profit. Therefore by using the chi-square test we analysed the kind of encouraged arrival pattern that adheres to the firm simultaneously increasing the firm's projected profit. Thus this study helps the entrepreneurs to decide the kind of discounts that would attract the customers simultaneously improving the firm's profit.

Declaration of conflicting interest: The authors declare that there is no conflict of interest.

References

- [1] Igbinoba, J.O., Sule, O.K., Ugboya, A.P. and Akinwunmi, M. (2019). Application Of Queueing Theory In Optimization Of Service Process, A Case Study Of Gt Plaza Fast Food, *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, 6(1):2458-9403.
- [2] Ullah, H. (2011). Petri net versus queueing theory for evaluation of FMS, *Assembly Automation*, 31(1):29-37.
- [3] Tsarouhas, P.H. (2011). A comparative study of performance evaluation based on field failure data for food production lines, *Journal of Quality in Maintenance Engineering*, 17(1):26-39.
- [4] McGuire, K.A., Kimes, S.E., Lynn, M., Pullman, M.E. and Lloyd, R.C. (2010). A framework for evaluating the customer wait experience, *Journal of Service Management*, 21(3):269-290.
- [5] Kunfeng Lu., Jiqing Qiu. and Mahmoud, S. (2011). Robust passive filter design for uncertain singular stochastic Markov jump systems with mode-dependent time delays, 2nd International Conference on Intelligent Control and Information Processing, *IEEE*.
- [6] Karimi-Nasab, M. and Konstantaras, I. (2013). An inventory control model with stochastic replenishment interval and special sale offer, *European Journal of Operational Research*, 227(1):81-87.
- [7] Krommyda, I.P., Skouril, K. and Konstantaras, I. (2015). Optimal ordering quantities for substitutable products with stock-dependent demand, *Applied Mathematical Modelling*, 39(1):147-164.
- [8] Rashid, R., Hoseini, S.F. and Gholamian, M.R. (2015). Application of queueing theory in production-inventory optimization. *Journal of Industrial Engineering International*, 11:485-494
- [9] Som, B. K. and Seth, S. (2017). An M/M/1/N Queueing system with Encouraged Arrivals, *Global Journal of Pure and Applied Mathematics*, 17:3443-3453.

- [10] Som, B.K. (2020). Multi-server Finite Waiting-space Encouraged Arrival Queuing System with Reverse Reneging, *Jagannath University Research Journal*, 1(1):2582-6263.
- [11] Jain, N.K., Kumar, R. and Som, B.K. (2014). An M/M/1/N Queuing system with reverse balking, *American Journal of Operational Research*, 4(2):17-20.
- [12] Patel, B. and Bhathawala, P. (2012). Case study for bank ATM queuing model. *International Journal of Engineering Research and Applications (IJERA)*, 2(5).
- [13] Kim, B.J.(2011). Conceptualization of traffic flow for designing Toll Plaza configuration: A case study using simulation with estimated traffic, *International Journal of Industrial Engineering*, 18(1):51-57.
- [14] Niranjana, S.P. and Indhira, K. (2016). A review on classical bulk arrival and batch service queueing model, *International Journal of Pure and Applied Mathematics*, 106(8):45-51.
- [15] Donald Gross, John F. Shortle, James M. Thompson. *Fundamentals of Queueing Theory*, Wiley Series in Probability and Statistics, 2018.
- [16] Medhi J. *Stochastic models in Queueing theory*, 2003.
- [17] Veerarajan T. *Probability, statistics and random processes with Queueing Theory and Queueing Networks*, 2009.
- [18] Taha A. *Operations research an introduction*, 2002.
- [19] Ajay kumar Sharma . *Queueing theory approach with queueing model: A study*, 2013.
- [20] Daniel A. – *Introduction to Queueing theory*, 1995.

MOMENTS OF GENERALIZED RECORD VALUES FROM MODIFIED FRÉCHET DISTRIBUTION AND ITS CHARACTERIZATION

Zaki Anwar¹, Abdul Nasir Khan² and Rafiqullah Khan³

•

^{1,3}Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh-202002,
India

²School of Mathematics and Statistics, MIT World Peace University, Pune-411038, India

zakistats@gmail.com¹

nasirgd4931@gmail.com²

aruke@rediffmail.com³

Abstract

The aim of this paper is to introduce the relations for moments and characterizing results for the newly introduced modified Fréchet distribution based on generalized record values. Here, we used an ordered random variable approach like generalized record values for generating the results. We have established the recurrence relations for single and product moments of generalized record values from modified Fréchet distribution. These relations are also deduced for the lower record values and some specific distributions, which are the special cases of modified Fréchet distribution. Further, the characterization results for this distribution have been established by using recurrence relations for single and product moments and conditional expectation of a function of generalized record values and truncated moments.

Keywords: Order statistics, generalized record values, modified Fréchet distribution, single moments, product moments, recurrence relations and characterization.

1. Introduction

The modified Fréchet distribution is an extension of the Fréchet distribution which was introduced by Tablada and Cordeiro [23] and pointed out that this distribution is quite effective to provide the best fits for real data sets. Since the results on real life data compared with other known distributions such as Fréchet, exponentiated Fréchet, Marshall–Olkin Fréchet, exponentiated Weibull, revealed that modified Fréchet distribution provides a better fit for modeling real life data.

A random variable X follows modified Fréchet distribution, if it's probability density function pdf is of the form

$$f(x) = \frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x}\right)^\lambda \exp\left[-\alpha x - \left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}\right], \quad x \geq 0, \quad \alpha, \beta, \lambda > 0 \quad (1)$$

with the distribution function (*df*)

$$F(x) = \exp\left[-\left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}\right], \quad x \geq 0, \quad \alpha, \beta, \lambda > 0. \quad (2)$$

Where β and λ are shape parameters.

Note that $f(x)$ and $F(x)$ satisfy the relation.

$$f(x) = \frac{(\lambda + \alpha x)}{x} [-\ln F(x)] F(x). \quad (3)$$

The Fréchet and standard Gumbel distributions are the special cases of the modified Fréchet distribution, when $\lambda = 0$ and $\lambda = 0, \alpha = 1$ respectively.

Initially, Chandler [8] was the first who laid down the concept of record values inspired by the extreme weather conditions. As a result, he designed the model for successive extremes values in a sequence of identically independently distributed (*iid*) continuous random variables. Dziubdziela and Kopociński [9] have generalized the concept of record values by choosing random variables of more generalized nature and these random variables are called the k -th record values. Later, the record values defined by Dziubdziela and Kopociński [9] have been called as generalized record values by Minimol and Thomas [15], since the r -th member of the sequence of the ordinary record values is also known as the r -th record value. Setting $k = 1$, we obtain ordinary record statistics.

Generally, the record values means the values which are not acquired before, e.g., fastest century in the one day cricket match, the longest winning streak in basketball, the world record in high jumping, the lowest time to cover a fixed distance in freestyle swimming and so on. The observation which is greater (or less) than the previous all observations is known as the record value. Record values arise naturally in many real life applications involving data relating to weather, sports, economics and life-tests.

For more details on the applications of record values, see Ahsanullah [1], Ahsanullah and Nevzorov [2], Arnold et. al. [5].

Let $\{X_n, n \geq 1\}$ be a sequence of independently identically distributed (*iid*) random variables with *df* $F(x)$ and *pdf* $f(x)$. The r -th order statistics of a random sample X_1, X_2, \dots, X_n is denoted by $X_{r:n}$. For fixed $k \geq 1$, we define the sequence $\{L_k(n), n \geq 1\}$ of k -th record times of $\{X_n, n \geq 1\}$ as follows:

$$L_k(1) = 1$$

$$L_k(n+1) = \min\{j > L_k(n) : X_{k:L_k(n)+k-1} > X_{k:j+k-1}\}.$$

The sequence $\{Z_n^{(k)}, n \geq 1\}$ with $Z_n^{(k)} = X_{k:L_k(n)+k-1}, n = 1, 2, \dots$ is called the sequence of k -th lower record values of $\{X_n, n \geq 1\}$. For convenience, we shall also take and $Z_0^{(k)} = 0$. Note that for $k=1$ we have $Z_n^{(1)} = X_{L(n)}, n \geq 1$. Then *pdf* of $Z_n^{(k)}$ and the joint *pdf* $Z_m^{(k)}$ and $Z_n^{(k)}$ are as follows:

$$f_{Z_n^{(k)}}(x) = \frac{k^n}{(n-1)!} [-\ln F(x)]^{n-1} [F(x)]^{k-1} f(x), \quad n \geq 1 \quad (4)$$

$$f_{Z_m^{(k)}, Z_n^{(k)}}(x, y) = \frac{k^n}{(m-1)!(n-m-1)!} [-\ln F(x)]^{n-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-1} \times [F(y)]^{k-1} f(y), \quad x > y, \quad 1 \leq m < n, \quad n \geq 2. \quad (5)$$

The conditional *pdf* of $Z_n^{(k)}$ given $Z_m^{(k)} = x$ as given

$$f_{Z_n^{(k)}|Z_m^{(k)}}(y|x) = \frac{k^{n-m}}{(n-m-1)!} [\ln F(x) - \ln F(y)]^{n-m-1} \left(\frac{F(y)}{F(x)} \right)^{k-1} \frac{f(y)}{F(x)}, \quad y < x. \quad (6)$$

For some recent developments on generalized record values with special reference to those arising from NH, exponentiated Rayleigh, Kappa distribution, additive-Weibull lifetime, Power function, extended Erlang-truncated exponential, Kumaraswamy-log-logistic, Weibull-Rayleigh, Weibull-power function, Fréchet distributions see, Alam et al. [4], Khan et al. ([12], [13]), Khan et al. [14], MirMostafae et al. [16], Paul [17], Singh and Khan [19], Singh et al. ([20], [21], [22]) Thomas and Paul [24], etc. In this paper we mainly studied the generalized lower record values arising from the modified Fréchet distribution.

The plots represent the shapes of the *pdf* of lower record values, arises from the modified Fréchet distribution.

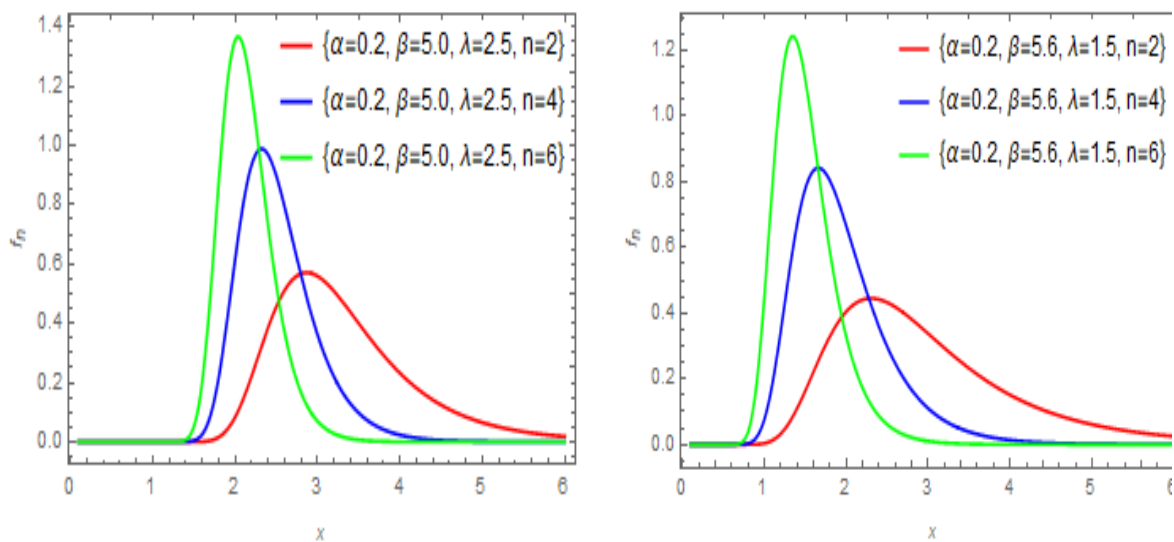


Figure. Plots of the *pdf* of lower record values from modified Fréchet distribution for selected values of parameters.

2. Relations for single moments

Theorem 2.1. For the modified Fréchet distribution given in (2) and $1 \leq k \leq n$, $j = 0, 1, \dots$

$$E(Z_{n+1}^{(k)})^j - E(Z_n^{(k)})^j = \frac{j\alpha}{(j+1)\lambda} \{E(Z_n^{(k)})^{j+1} - E(Z_{n+1}^{(k)})^{j+1}\} - \frac{j}{n\lambda} E(Z_n^{(k)})^{j+1} \quad (7)$$

Consequently for $n \geq 1$, $1 \leq k \leq n$ and $j = 0, 1, \dots$

$$jE(Z_{n+1}^{(k)})^{j+1} = jE(Z_1^{(k)})^{j+1} + \frac{j+1}{\alpha} \sum_{p=1}^n \frac{(p\lambda - j)}{p} E(Z_p^{(k)})^j - \frac{\lambda(j+1)}{\alpha} \sum_{p=2}^{n+1} E(Z_p^{(k)})^j. \quad (8)$$

Proof. From (3) and (4), we get

$$E(Z_n^{(k)})^j = \frac{\lambda k^n}{(n-1)!} \int_0^\infty x^{j-1} [-\ln F(x)]^n [F(x)]^k dx + \frac{\alpha k^n}{(n-1)!} \int_0^\infty x^j [-\ln F(x)]^n [F(x)]^k dx. \quad (9)$$

In view of Bieniek and Szynal [7], note that

$$E(Z_n^{(k)})^j - E(Z_{n-1}^{(k)})^j = -j \frac{k^{n-1}}{(n-1)!} \int_0^\infty x^{j-1} [-\ln F(x)]^{n-1} [F(x)]^k dx.$$

Therefore,

$$E(Z_{n+1}^{(k)})^j - E(Z_n^{(k)})^j = -j \frac{k^n}{n!} \int_0^\infty x^{j-1} [-\ln F(x)]^n [F(x)]^k dx.$$

On substituting in (9), we get

$$E(Z_n^{(k)})^j = \frac{n\lambda}{j} \{E(Z_n^{(k)})^j - E(Z_{n+1}^{(k)})^j\} + \frac{n\alpha}{(j+1)} \{E(Z_n^{(k)})^{j+1} - E(Z_{n+1}^{(k)})^{j+1}\}.$$

On rewriting above expression, we derive the recurrence relation in (7). Then, by repeatedly applying the recurrence relation in (7), we simply derive the recurrence relation in (8).

Remark 2.1. For $k=1$ in (7), the recurrence relation for single moments of lower record values from the modified Fréchet distribution given as

$$E(X_{L(n+1)})^j - E(X_{L(n)})^j = \frac{j\alpha}{(j+1)\lambda} \{E(X_{L(n)})^j - E(X_{L(n+1)})^j\} - \frac{j}{n\lambda} E(X_{L(n)})^j.$$

Remark 2.2. Putting $k=1$, $\lambda=0$, $\alpha=1$ in (7), we deduced the recurrence relation for single moments of lower record values from the standard Gumbel distribution as obtained by Balakrishnan et al. [6].

Remark 2.3. Setting $\alpha=0$ in (7), we deduced the recurrence relation for generalized record values from inverse Weibull distribution as established by Pawlas and Szynal [18] for replacing n by $n-1$.

Table I Moments of lower record values

n	$\alpha = 1.5, \beta = 4$			$\alpha = 2, \beta = 5$		
	$\lambda = 1.5$	$\lambda = 2.5$	$\lambda = 3.5$	$\lambda = 1.5$	$\lambda = 2.5$	$\lambda = 3.5$
1	1.45394	1.78177	2.02126	1.32077	1.68029	1.95621
2	1.06968	1.44242	1.71475	1.00841	1.39679	1.69422
3	0.90020	1.28891	1.57416	0.86681	1.26561	1.57166
4	0.79618	1.19248	1.48489	0.77828	1.18208	1.49298
5	0.72301	1.12323	1.42020	0.71510	1.12151	1.43553

Table II Variances of lower record values

n	$\alpha = 1.5, \beta = 4$			$\alpha = 2, \beta = 5$		
	$\lambda = 1.5$	$\lambda = 2.5$	$\lambda = 3.5$	$\lambda = 1.5$	$\lambda = 2.5$	$\lambda = 3.5$
1	0.28211	0.21581	0.17375	0.18045	0.14615	0.12349
2	0.08176	0.06541	0.05419	0.05577	0.04695	0.04059
3	0.04147	0.03480	0.02952	0.02950	0.02579	0.02271
4	0.02588	0.02267	0.01960	0.01900	0.01718	0.01535
5	0.01798	0.01638	0.01440	0.01355	0.01263	0.01142

3. Relations for product moments

Theorem 3.1. For the modified Fréchet distribution given in (2) and $n \geq 1$, $m \geq k$, $i, j = 0, 1, \dots$

$$E[(Z_{m+1}^{(k)})^{i+j}] - E(Z_m^{(k)})^i (Z_{m+1}^{(k)})^j = \frac{i\alpha}{(i+1)\lambda} \{E[(Z_m^{(k)})^{i+1} (Z_{m+1}^{(k)})^j] - E(Z_{m+1}^{(k)})^{i+j+1}\} - \frac{i}{m\lambda} E[(Z_m^{(k)})^{i+j}]. \quad (10)$$

and for $1 \leq m < n$, $n \geq 2$, $i, j = 0, 1, \dots$

$$E[(Z_{m+1}^{(k)})^i (Z_n^{(k)})^j] - E(Z_m^{(k)})^i (Z_n^{(k)})^j$$

$$= \frac{i\alpha}{(i+1)\lambda} \{E[(Z_m^{(k)})^{i+1} (Z_n^{(k)})^j] - E(Z_{m+1}^{(k)})^{i+1} (Z_n^{(k)})^j\} - \frac{i}{m\lambda} E[(Z_m^{(k)})^i (Z_n^{(k)})^j]. \quad (11)$$

Proof. From (3) and (5), we have

$$E[(Z_m^{(k)})^i (Z_n^{(k)})^j] = \frac{\lambda k^n}{(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^{i-1} y^j [-\ln F(x)]^m [\ln F(x) - \ln F(y)]^{n-m-1}$$

$$\times [F(y)]^{k-1} f(y) dx dy + \frac{\alpha k^n}{(m-1)!(n-m-1)!} \times \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^m$$

$$\times [\ln F(x) - \ln F(y)]^{n-m-1} [F(y)]^{k-1} f(y) dx dy,$$

$$= \frac{\lambda k^n}{(m-1)!(n-m-1)!} \int_0^\infty y^j [F(y)]^{k-1} f(y) I_{i-1}(x, y) dy$$

$$+ \frac{\alpha k^n}{(m-1)!(n-m-1)!} \int_0^\infty y^j [F(y)]^{k-1} f(y) I_i(x, y) dy, \quad (12)$$

where

$$I_b(x, y) = \int_y^\infty x^b [-\ln F(x)]^m [\ln F(x) - \ln F(y)]^{n-m-1} dx.$$

Integrating by part, taking x^b for integration and rest of the part for differentiation, we get

$$I_b(x, y) = \frac{m}{b+1} \int_y^\infty x^{b+1} [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-1} dx$$

$$- \frac{(n-m-1)}{b+1} \int_y^\infty x^{b+1} [-\ln F(x)]^m \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-2} dx.$$

Substituting in (12), we get

$$E[(Z_m^{(k)})^i (Z_n^{(k)})^j] = \frac{m\lambda k^n}{i(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)}$$

$$\times [\ln F(x) - \ln F(y)]^{n-m-1} [F(y)]^{k-1} f(y) dx dy - \frac{\lambda k^n}{i(m-1)!(n-m-2)!}$$

$$\times \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^m \frac{f(x)}{F(y)} [\ln F(x) - \ln F(y)]^{n-m-2} [F(y)]^{k-1} f(y) dx dy$$

$$+ \frac{m\alpha k^n}{(i+1)(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^{i+1} y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)}$$

$$\times [\ln F(x) - \ln F(y)]^{n-m-1} [F(y)]^{k-1} f(y) dx dy$$

$$- \frac{\alpha k^n}{(i+1)(m-1)!(n-m-2)!} \int_0^\infty \int_y^\infty x^{i+1} y^j [-\ln F(x)]^m \frac{f(x)}{F(x)}$$

$$\times [\ln F(x) - \ln F(y)]^{n-m-2} [F(y)]^{k-1} f(y) dx dy.$$

After simplification, we obtain the required result as given in (11).

Proceeding in a similar manner for the case $n = m + 1$, the recurrence relation given in (10) can easily be established.

On can also note that Theorem 2.1. can be deduced from Theorem 3.1. by putting $j = 0$.

Remark 3.1. Putting $k = 1$ in (11), the recurrence relations for product moments of lower record values is deduced for the modified Fréchet distribution in the form

$$E[(X_{L(m+1)})^i (X_{L(n)})^j] - E[(X_{L(m)})^i (X_{L(n)})^j] = -\frac{i}{m\lambda} E[(X_{L(m)})^i (X_{L(n)})^j] \\ + \frac{i\alpha}{(i+1)\lambda} \{E[(X_{L(m)})^{i+1} (X_{L(n)})^j] - E[(X_{L(m+1)})^{i+1} (X_{L(n)})^j]\}.$$

Remark 3.2. Setting $k = 1$, $\lambda = 0$, $\alpha = 1$ in (11), we get the recurrence relations for product moments of lower record values from the standard Gumbel distribution as obtained by Balakrishnan et al. [6].

Remark 3.3. Assuming $\alpha = 0$ in (11), the recurrence relations for product moments of generalized record values is deduced for inverse Weibull distribution as established by Pawlas and Szynal [18].

4. Characterizations

Theorem 4.1. If k and j be are positive integers. A necessary and sufficient condition for a random variable X to be distributed with *pdf* given by (2) is that

$$E(Z_{n+1}^{(k)})^j - E(Z_n^{(k)})^j = \frac{j\alpha}{(j+1)\lambda} \{E(Z_n^{(k)})^{j+1} - E(Z_{n+1}^{(k)})^{j+1}\} - \frac{j}{n\lambda} E(Z_n^{(k)})^j. \quad (13)$$

Proof. The necessary part follows from Theorem 2.1. On the other hand if the recurrence relation (13) is satisfied, then on using Bieniek and Szynal [7], we have

$$\frac{k^n}{(n-1)!} \int_0^\infty x^j [-\ln F(x)]^{n-1} [F(x)]^{k-1} f(x) dx = \frac{\lambda k^n}{(n-1)!} \int_0^\infty x^{j-1} [-\ln F(x)]^n [F(x)]^k dx \\ + \frac{\alpha k^n}{(n-1)!} \int_0^\infty x^j [-\ln F(x)]^n [F(x)]^k dx,$$

which implies

$$\int_0^\infty x^j [-\ln F(x)]^{n-1} [F(x)]^k \left\{ \frac{f(x)}{F(x)} - \frac{(\lambda + \alpha x)}{x} [-\ln F(x)] \right\} dx = 0.$$

Now applying a generalization of the Müntz-Szász theorem (see for example Hwang and Lin [11]) to above expression, we get

$$f(x) = \frac{(\lambda + \alpha x)}{x} [-\ln F(x)] F(x),$$

which proves the sufficiency part.

Theorem 4.2. For a positive integer $k \geq 1$ and let i, j are non-negative integers, a necessary and sufficient condition for a random variable X to be distributed with *pdf* given by (1) is that

$$E[(Z_{m+1}^{(k)})^i (Z_n^{(k)})^j] - E[(Z_m^{(k)})^i (Z_n^{(k)})^j] = -\frac{i}{m\lambda} E[(Z_m^{(k)})^i (Z_n^{(k)})^j] \\ + \frac{i\alpha}{(i+1)\lambda} \{E[(Z_m^{(k)})^{i+1} (Z_n^{(k)})^j] - E[(Z_{m+1}^{(k)})^{i+1} (Z_n^{(k)})^j]\}. \quad (14)$$

Proof. The necessary part follows from Theorem 3.1. On the other hand if the relation in (14) is satisfied, then (14) can be written as

$$\begin{aligned} & \frac{k^n}{(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-1} \\ & \times [F(y)]^{k-1} f(y) dx dy = \frac{\lambda m k^n}{i(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} \\ & \times [\ln F(x) - \ln F(y)]^{n-m-2} [F(y)]^{k-1} f(y) \left\{ \ln F(x) - \ln F(y) - \frac{(n-m-1)[-\ln F(x)]}{m} \right\} dx dy \\ & + \frac{\alpha m k^n}{(i+1)(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^{i+1} y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-2} \\ & \times [F(y)]^{k-1} f(y) \left\{ \ln F(x) - \ln F(y) - \frac{(n-m-1)[-\ln F(x)]}{m} \right\} dx dy. \end{aligned} \quad (15)$$

Let

$$h(x, y) = -\frac{1}{m} [-\ln F(x)]^m [\ln F(x) - \ln F(y)]^{n-m-1}$$

Differentiating both sides with respect to x , we get

$$\begin{aligned} \frac{\partial}{\partial x} h(x, y) &= [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-2} \\ & \times \left\{ \ln F(x) - \ln F(y) - \frac{(n-m-1)[-\ln F(x)]}{m} \right\}. \end{aligned}$$

Thus, (15) can be written as

$$\begin{aligned} & \frac{k^n}{(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-1} [F(y)]^{k-1} f(y) dx dy \\ & = \frac{\lambda m k^n}{i(m-1)!(n-m-1)!} \int_0^\infty y^j [F(y)]^{k-1} f(y) I_i(x, y) dy \\ & + \frac{\alpha m k^n}{(i+1)(m-1)!(n-m-1)!} \int_0^\infty y^j [F(y)]^{k-1} f(y) I_{i-1}(x, y) dy, \end{aligned} \quad (16)$$

where

$$I_b(x, y) = \int_y^\infty x^b \frac{\partial}{\partial x} h(x, y) dx.$$

Integrating by parts, treating x^b for differentiation and rest of the part for integration, we get

$$I_b(x, y) = -b \int_y^\infty x^{b-1} h(x, y) dx,$$

On substituting $I_b(x, y)$ in (16), yields

$$\begin{aligned} & \frac{k^n}{(m-1)!(n-m-1)!} \int_0^\infty \int_y^\infty x^i y^j [-\ln F(x)]^{m-1} \frac{f(x)}{F(x)} [\ln F(x) - \ln F(y)]^{n-m-1} \\ & \times [F(y)]^{k-1} f(y) \left\{ \frac{f(x)}{F(x)} - \frac{(\lambda + \alpha x)}{x} [-\ln F(x)] \right\} dx dy = 0. \end{aligned}$$

Now applying a generalization of the Müntz-Szász theorem (see for example Hwang and Lin [11]), we get

$$f(x) = \frac{(\lambda + \alpha x)}{x} [-\ln F(x)] F(x).$$

Hence the sufficiency part proved.

Theorem 4.3. Let X be an absolutely continuous non-negative random variable having $df = F(x)$,

with $F(0) = 0$ and $0 \leq F(x) \leq 1 \quad \forall \quad 0 < x < \infty$, then

$$E[\xi(Z_n^{(k)}) | Z_m^{(k)} = x] = -\ln F(x) + \frac{n-m}{k} \quad (17)$$

if and only if

$$F(x) = \exp\left[-\left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}\right], \quad x > 0, \quad \alpha, \beta, \lambda > 0,$$

where

$$\xi(y) = \left(\frac{\beta}{y}\right)^\lambda e^{-\alpha y}.$$

Proof. From (6), we have

$$E[\xi(Z_n^{(k)}) | (Z_m^{(k)}) = x] = \frac{k^{n-m}}{(n-m-1)!} \int_0^x \left(\frac{\beta}{y}\right)^\lambda e^{-\alpha y} [\ln F(x) - \ln F(y)]^{n-m-1} \left(\frac{F(y)}{F(x)}\right) \frac{f(y)}{F(x)} dy.$$

By setting $z = \frac{F(y)}{F(x)}$, we have

$$E[\xi(Z_n^{(k)}) | (Z_m^{(k)}) = x] = \frac{k^{n-m}}{(n-m-1)!} \int_0^1 [-\ln F(x) - \ln z] [-\ln z]^{n-m-1} z^{k-1} dz. \quad (18)$$

Now (17) can be seen in view of (Gradshteyn and Ryzhik [10], p-551)

$$\int_0^1 (-\ln x)^{\mu-1} x^{\nu-1} dx = \frac{\Gamma(\mu)}{\nu^\mu}, \quad \mu > 0, \quad \nu > 0.$$

To prove the sufficient part, we have

$$\frac{k^{n-m}}{(n-m-1)!} \int_0^x \left(\frac{\beta}{y}\right)^\lambda e^{-\alpha y} [\ln F(x) - \ln F(y)]^{n-m-1} [F(y)]^{k-1} f(y) dy = [F(x)]^k g_{n|m}(x), \quad (19)$$

where

$$g_{n|m}(x) = -\ln F(x) + \frac{n-m}{k}.$$

Differentiating both sides of (19) with respect to x , we get

$$\begin{aligned} \frac{k^{n-m} f(x)}{(n-m-2)! F(x)} \int_0^x \left(\frac{\beta}{y}\right)^\lambda e^{-\alpha y} [\ln F(x) - \ln F(y)]^{n-m-2} \\ \times [F(y)]^{k-1} f(y) dy = g'_{n|m}(x) [F(x)]^k + k g_{n|m}(x) [F(x)]^{k-1} f(x) \end{aligned}$$

or

$$k g_{n|m+1}(x) [F(x)]^{k-1} f(x) = g'_{n|m}(x) [F(x)]^k + k g_{n|m}(x) [F(x)]^{k-1} f(x).$$

Therefore,

$$\frac{f(x)}{F(x)} = \frac{g'_{n|m}(x)}{k[g_{n|m+1}(x) - g_{n|m}(x)]} = \frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x},$$

where

$$g'_{n|m}(x) = -\frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}$$

$$g_{n|m+1}(x) - g_{n|m}(x) = -\frac{1}{k}.$$

Now integrating both the sides with respect to x , we get

$$\ln F(x) = -\left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x} + \log C$$

which implies

$$F(x) = C \exp\left[-\left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}\right].$$

Since $F(x) = 0$ as $x \rightarrow 0$ and $F(x) = C$ as $x \rightarrow \infty$.

Thus, by definition of $\lim_{x \rightarrow \infty} F(x) = 1$, this implies that $C = 1$.

Hence the sufficiency part proved.

Remark 4.3. If $k=1$ in (17), we get the following characterization of lower record values for modified Fréchet distribution

$$E[\xi(X_{L(n)} | X_{L(m)} = x)] = -\ln F(x) + n - m.$$

Theorem 4.4. Suppose X be an absolutely continuous (with respect to Lebesgue measure) random variable with the $\text{cdf } F(x)$ and $\text{pdf } f(x) \forall 0 < x < \infty$, such that $f'(x)$ and $E(X | X \leq x)$, exist for all $x, 0 < x < \infty$, then

$$E(X | X \leq x) = g(x)\eta(x), \tag{20}$$

where

$$\eta(x) = \frac{f(x)}{F(x)}$$

and

$$g(x) = \frac{x^2 e^{\alpha x}}{(\lambda + \alpha x) \left(\frac{\beta}{x}\right)^\lambda} - \frac{x e^{\alpha x} \exp\left[\left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}\right]}{(\lambda + \alpha x) \left(\frac{\beta}{x}\right)^\lambda} \int_0^x \exp\left[-\left(\frac{\beta}{u}\right)^\lambda e^{-\alpha u}\right] du$$

if and only if

$$f(x) = \frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x}\right)^\lambda \exp\left[-\alpha x - \left(\frac{\beta}{x}\right)^\lambda e^{-\alpha x}\right], \quad x \geq 0, \quad \alpha, \beta, \lambda > 0.$$

Proof. From (1), we have

$$E(X | X \leq x) = \frac{1}{F(x)} \int_0^x u \frac{(\lambda + \alpha u)}{u} \left(\frac{\beta}{u}\right)^\lambda e^{-\alpha u} \exp\left[-\left(\frac{\beta}{u}\right)^\lambda e^{-\alpha u}\right] du.$$

Integrating by parts, taking $\frac{(\lambda + \alpha u)}{u} \left(\frac{\beta}{u}\right)^\lambda e^{-\alpha u} \exp\left[-\left(\frac{\beta}{u}\right)^\lambda e^{-\alpha u}\right]$ for integration and rest of the integrand for differentiation, we get

$$E(X | X \leq x)$$

$$= \frac{1}{F(x)} \left\{ x \exp \left[- \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} \right] - \int_0^x \exp \left[- \left(\frac{\beta}{u} \right)^\lambda e^{-\alpha u} \right] du \right\}$$

Now dividing and multiplying by $f(x)$, we obtain the result as given in (20).

For proving sufficient part, we have from (20)

$$\int_0^x u f(u) du = g(x) f(x).$$

Differentiating on both sides with respect to x , we find that

$$x f(x) = g'(x) f(x) + g(x) f'(x).$$

Therefore,

$$\frac{f'(x)}{f(x)} = \frac{x - g'(x)}{g(x)} \quad \text{Ahsanullah et. al [24]}$$

$$\frac{f'(x)}{f(x)} = -\frac{\lambda}{x} - \frac{\lambda}{x(\lambda + \alpha x)} + \frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} - \alpha, \quad (21)$$

where

$$g'(x) = x - g(x) \left(-\frac{\lambda}{x} - \frac{\lambda}{x(\lambda + \alpha x)} + \frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} - \alpha \right).$$

On integrating (21) both sides with respect to x , we get

$$f(x) = C \frac{(\lambda + \alpha x)}{x} \left(\frac{1}{x} \right)^\lambda e^{-\alpha x} \exp \left[- \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} \right].$$

Further, To obtain the value of C (constant of integration), we have used the property of *pdf* that is

$$\int_0^\infty f(x) dx = 1.$$

Thus,

$$\frac{1}{C} = \int_0^\infty \frac{(\lambda + \alpha x)}{x} \left(\frac{1}{x} \right)^\lambda e^{-\alpha x} \exp \left[- \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} \right] dx = \frac{1}{\beta^\lambda},$$

which proves that

$$f(x) = \frac{(\lambda + \alpha x)}{x} \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} \exp \left[- \left(\frac{\beta}{x} \right)^\lambda e^{-\alpha x} \right], \quad x > 0, \quad \alpha, \beta, \lambda > 0.$$

Remark 4.5. Setting $\alpha = 1$ and $\lambda = 0$, Theorem 4.4 gives characterizing result for standard Gumbel distribution and for $\alpha = 0$, it gives the characterizing result for inverse Weibull distribution.

Conclusion: In this paper, we have presented the new results for the single and product moments of modified Fréchet distribution based on generalized lower record values. These results include some well-known results for standard Gumbel and inverse Weibull distributions as obtained by

Balakrishnan et al. [6] and Pawlas and Szynal [18]. Later, we established the characterizing results for this distribution by utilizing the relations for single and product moments and conditional expectation of a function of generalized lower record value, and using truncated moments.

References

- [1] Ahsanullah, M. Record Statistics, Nova Science Publishers, New York, (1995).
- [2] Ahsanullah, M. and Nevzorov, V. B. Record via Probability Theory, Atlantis Press, Paris, (2015).
- [3] Ahsanullah, M., Shakil, M. and Golam Kibria, B.M. (2016). Characterization of continuous distribution by truncated moment. Journal of Modern Applied Statistical Method, 15: 316-331.
- [4] Alam, M., Khan, M. A. and Khan, R. U (2020). Characterization Of NH distribution through generalized record values. Applied Mathematics E-Notes, 20: 406-414.
- [5] Arnold B. C., Balakrishnan, N. and Nagaraja, H. N. Records, Wiley, New York, (1998).
- [6] Balakrishnan, N., Ahsanullah, M. and Chan, P. S. (1992). Relations for single and product moments of record values from Gumbel distribution. Statistics and Probability Letters, 15: 223-227.
- [7] Bieniek, M. and Szynal, D. (2002). Recurrence relations for distribution functions and moments of k – th record values. Journal of Mathematical Sciences, 111: 3511-3519.
- [8] Chandler, K.N. (1952). The distribution and frequency of record values. Journal of the Royal Statistical Society: Series B, 14: 220-228.
- [9] Dziubdziela, W. and Kopociński, B. (1976). Limiting properties of the k – th record value. Mathematica Applicanda, 15: 187-190.
- [10] Gradshteyn, I.S. and Ryzhik, I.M. Tables of Integrals, Series of Products, Academic Press, New York, (2007).
- [11] Hwang, J.S. and Lin, G.D. (1984). On a generalized moments problem II. Proceedings of the American Mathematical Society, 91: 577-580.
- [12] Khan, M. A. R., Khan, R. U. and Singh, B. (2019a). Relations for moments of dual generalized order statistics from exponentiated Rayleigh distribution and associated Inference. Journal Statistical Theory and Applications, 18: 402-415.
- [13] Khan, M. A. R., Khan, R. U. and Singh, B. (2019b). Moments of dual generalized order statistics from two parameter Kappa distribution and characterization. Journal of Applied Probability and Statistics, 14: 85-101.
- [14] Khan, R. U., Khan, M. A. and Khan, M. A. R. (2017). Relations for moments of generalized record values from additive-Weibull lifetime distribution and associated inference. Statistics Optimization and Information Computing, 5: 127-136.
- [15] Minimol, S. and Thomas, P.Y (2013). On some properties of Makeham distribution using generalized record values and its characterization. Brazilian Journal of Probability and Statistics, 27: 487-501.
- [16] MirMostafae, S. K., Asgharzadeh, A. and Fallah, A. (2016). Record values from NH distribution and associated inference. Metron, 74: 37-59.
- [17] Paul, J. (2014). On generalized lower (k) record values arising from power function distribution. Journal of the Kerala Statistical Association, 25: 49-64.
- [18] Pawlas, P. and Szynal, D. (2000). Characterizations of the inverse Weibull distribution and generalized extreme value distributions by moments of k – th record values. Mathematica Applicanda, 27: 197-202.
- [19] Singh, B. and Khan, R. U. (2018). Moments of extended Erlang-truncated exponential distribution based on k – th lower record values and characterizations. International Journal of Mathematics and Statistics Invention, 6: 65-74.

[20] Singh, B., Khan, R. U and Khan, M. A. R (2019a). Moments of generalized record values from Kumaraswamy-log-logistic distribution and related inferences. *Thailand Statistician*, 17: 93-103.

[21] Singh, B., Khan, R. U. and Khan, M. A. (2019b). Exact moments and characterizations of the Weibull-Rayleigh distribution based on generalized upper record statistics. *Applied Mathematics E-Notes*, 19: 675-688.

[22] Singh, B., Khan, R. U. and Zarrin, S. (2020). Moments of generalized upper record values from Weibull-power function distribution and characterization. *Journal of Statistics Applications and Probability*, 9: 309-318.

[23] Tablada, C. J. and Cordeiro, G. M. (2017). The modified Fréchet distribution and its properties. *Communications in Statistics-Theory and Methods*, 46: 10617-10639.

[24] Thomas, P. Y. and Paul, J. (2014). On generalized lower (k) record values from the Fréchet distribution. *Journal of The Japan Statistical Society*, 44: 157-178.

PROFITABILITY ANALYSIS OF A FOOD INDUSTRIAL SYSTEM HAVING MAKE-AND-PACK PRODUCTION STRATEGY WITH PRIORITY BASIS REPAIR

Monika¹ and Garima Chopra^{2*}

^{1,2}Department of Mathematics, University Institute of Engineering & Technology, Maharshi Dayanand University, Rohtak, Haryana, India

[1monika.dhull8@gmail.com](mailto:monika.dhull8@gmail.com)

[2*garima.chopra@gmail.com](mailto:garima.chopra@gmail.com)

*corresponding author

Abstract

The present investigation is concerned with the profitability analysis of a food industrial system where production is based on the make-and-pack strategy. The system is assumed to have two subsystems: first subsystem is for making while the second is for packing the product so formed. As per the gathered information about the production procedure in the food industrial plant, the priority of repair is given to the making subsystem over the packing subsystem. Here, failure of either subsystem leads to a complete breakdown of the system. Also, two types of failures are considered in the packing subsystem i.e. minor failures and major failures. Two kinds of repairers (operator and fitter) are appointed to tackle the failures in the subsystems. For minor and major failures in the packing subsystem, the operator and fitter respectively are responsible for repairs. However, any failure in making subsystem is repaired by the combined efforts of the operator and fitter. Reliability characteristics such as mean time to system failure (MTSF), system availability, and expected busy period of the repair persons are studied by employing the semi-Markov process and regenerative point technique. The system profitability is graphically analyzed concerning to failure rates of both subsystems.

Keywords: make-and-pack production process, priority basis repair, regenerative point technique, semi-Markov process

1. Introduction

Globalization has opened up new opportunities and challenges to the manufacturing industries which force producers to seek out more efficient ways to manufacture their products. The producers adopt various strategies to obtain the most reliable product and as a result, the production process is becoming more multifarious. So, for the smooth handling of the production process, it is divided into some stages i.e. known as two-stage production or three-stage production, etc. From the existing literature, it is visible that Johnson [1] was the first one who used the term two-stage production in which he considered one machine in each stage. Afterward, a number of significant studies came into existence regarding two-stage production with consideration of more than one machine in the

second stage as taken by Gupta and Tunc [2] and Honkomp *et al.* [3]. From the general point of view of the production, this two-stage production was specifically named the make-and-pack process. Usually, a make-and-pack production process is coupled with two stages: the first stage is responsible for the making or formation of the product while the second stage is responsible for the packaging or packing of the product so formed. This mode of production makes workability more flexible by managing the processing rates of the packaging lines with the respect that of formulation lines. It is frequently confronted in many production industries such as the paper industry, pharmaceutical industry, food industry, chemical industry, etc.

Make-and-pack production process is a well-known concept in the food processing industry. Akkerman *et al.* [4] examined the effect of some capacity and time constraints on the performance of a two-stage food production system. Sel *et al.* [5] discussed the planning and scheduling of make-and-pack production under lifetime uncertainty. Klanke *et al.* [6] analyzed the short-term scheduling of the make-and-pack process for minimizing the total production makespan of the schedule. It is apparent from the extensive literature review that most of the research on make-and-pack process is limited to the scheduling of this production process. But there is another side rather than the scheduling of this process which can be equally responsible for the financial loss in manufacturing industries and that is the failure in the systems used for production. As if the system will fail and did not come into operation timely, then the company can face a great loss in terms of money and reputation. So, the present paper is an effort to develop the reliability model for a food industrial system considering two subsystems one for making and another for the packing stage. In reliability analysis, a modeling approach is usually adopted to understand and predict the system's behavior in given situations with the help of probabilistic concepts. Also, the priority of repair and two types of repairmen are taken from the point of making the system more available so that profit can be maximized.

As far as reliability analysis is concerned, it has been a very compelling topic for a long time. It has been dealing with various types of industrial systems since the 1960s. A lot of research has been done and appreciated on the reliability analysis of various systems as seen in the literature. The various types of systems viz. single-unit systems or two or more unit industrial systems have been studied in the literature considering various factors affecting the system performance such as preventive maintenance of the units, perfect/imperfect switchover of the units, priority basis repair, two or more types of repairmen, standby units, inspections, replacement of the failed unit or its repair in the online or offline mode and plenty of factors are also there. These all factors were sized up to see the feasibility of profit maximization. Gaver [7] made derivations regarding MTSF and the availability of the type of systems that are composed of two paralleled subsystems. He considered situations for when there is waiting time to repair for another subsystem or not when both fail simultaneously. Pandey *et al.* [8] discussed the stochastic modeling of a powerloom plant consisting of two units having mechanical failures along with the concept of two additional failures due to poorly trained weavers; common cause failure and human error. Mathew *et al.* [9] studied a two-unit system of continuous casting plant with three types of failures such as repairable, replaceable, or requiring reconditioning/ reinstallation. Taj *et al.* [10] have examined a single-machine subsystem involved in a cable plant by considering its three types of maintenance strategies. Kumari *et al.* [11] investigated the profit of the butter-oil (ghee) manufacturing system through the supplementary variable technique. Singh *et al.* [12] evaluated the reliability metrics of a complex repairable system having two subsystems connected in series with imperfect switching. Saini *et al.* [13] analyzed a redundant system with non-identical units; one original and another duplicate cold standby unit where priority was given to the original unit over the repair of the duplicate unit. Bashir *et al.* [14] proposed a model considering two units along with their controlled and uncontrolled failures in terms of repairing and replacing respectively. Andalib and Sarkar [15] discussed a repairable system

with two spare units along with their repair/service by two repair persons. Sharma and Drishti [16] studied the seasonal effect on the workability of an ice-cream plant. Monika and Chopra [17] have developed the reliability model for the food industrial system by considering demand-based seasons. Some other recent reliability studies on realistic systems can be explored in Rizwan *et al.* [18], Sachdeva *et al.* [19], and Yusuf and Sanusi [20].

It is apparent from the literature that several aspects have been taken regarding the reliability modeling of industrial systems. So, on studying these aspects and by the visit of the industrial system under consideration, the concept of priority of repair of the making subsystem over the packing subsystem and two types of repairmen are considered. In all, the present study helps to fill the gap between the scheduling of the make-and-pack process of the food production systems and failures that cause a delay in food production following the financial loss. We develop a model with consideration of the two subsystems (one for making and another for packing) of the food industrial system based on the production strategy. Also, it is assumed that the packing subsystem has more than one unit working in parallel and this subsystem can have failures of two types i.e. minor and major failures. Accordingly, various measures of system effectiveness have been evaluated with the help of the regenerative point technique to see the behavior of the profit function, and also profit maximizing parameters are deduced.

2. Notations

Table 1: Notations used throughout the paper

Notations	Description
λ_1	Constant failure rate of the making subsystem
λ_2	Constant failure rate of the packing subsystem
$g_1(t)/G_1(t)$	pdf/cdf of repair time of making subsystem
$g_{21}(t)/G_{21}(t)$	pdf/cdf of time to complete repair of minor failures of packing subsystem
$g_{22}(t)/G_{22}(t)$	pdf/cdf of time to complete repair of major failures of packing subsystem
$q_{ij}(t)/Q_{ij}(t)$	pdf/cdf of transition time from a state 'i' to a state 'j'
$\phi_i(t)$	cdf of first passage time from a regenerative state 'i' to a failed state
A_0	Steady state availability of the system
B_0^o/B_0^f	Busy period of the operator/fitter for repair
$M_i(t)$	Probability that the system is up initially in regenerative state 'i' and is up at instant t without going through any other regenerative state
$W_i(t)$	Probability that the repair person is busy in repair of the subsystem (making/packing) initially in regenerative state 'i' and is engaged at time t without visiting to any other regenerative state
m_{ij}	The unconditional mean time taken by the system to visit any regenerative state 'j' when the time is measured from the time of entrance into state 'i'
μ_i	Mean sojourn time, i.e., the expected spent time in a regenerative state 'i' before visiting any other state
*/@	Symbol for Laplace transform / Laplace convolution
**/@	Symbol for Laplace Stieltjes transform/ Laplace Stieltjes convolution

Table 2: Notations regarding the states of the system

State	Symbol	Meaning
State 0	(O_m, O_p)	Operative state of the system where both subsystems viz. making and packing subsystems are operative
State 1	(O_m, F_{urp_1})	Operative state of the system where packing subsystem is under minor repair
State 2	(F_{urm}, F_{wrp_1})	Failed state of the system where making subsystem is under repair and the packing subsystem is waiting for minor repair
State 3	(F_{urm}, D_p)	Failed state of the system where making subsystem is under repair and the packing subsystem is in down state
State 4	(D_m, F_{urp_2})	Failed state of the system where making subsystem is in down state and the packing subsystem is under major repair

This section is devoted to notations used in the present study. All notations are specified in Table 1 and Table 2.

3. Description of the system and Assumptions

3.1. Description of the system

This paper deals with a food industrial system in which production is based on the make-and-pack strategy. This system has two subsystems; one for making and another for packing the product so formed. The making subsystem has four units working in series and the packing subsystem has two units working in series which further have five and three subunits working in parallel respectively (see Figure 1). When there are failures in the making subsystem then the system will completely fail and the packing subsystem will be kept in a down state to manage the production because the packing subsystem will not have material for packing, while the packing subsystem can behave in two manners when it has failures depending on the type of failures i.e. minor or major failures. When minor failures are there, then the system will work in the same manner (as it is assumed that the failures will be repaired within negligible time or online we can say) but in case of major failures, the system will completely fail and in that case also the making subsystem will be in the down state as if not kept in the down state then the surplus product will form which will not be worked upon further due to failure of the packing subsystem. Also, when there will be a simultaneous failure in both subsystems, then the priority of repair will be given to making subsystem. It is due to the reason that the failures in the packing subsystem, at the same time, can be handled by shifting the material to other subunits working in parallel; as per the information from the officials of the food processing plant. Five states are there depicting the possibilities taken with the system (see Figure 2) which are described in Table 2.

3.2. Assumptions

The following assumptions have been taken throughout the paper discussion:

- Initially, both the making and packing subsystems are in operative condition.
- The system will be in the state of complete failure only in two cases; either there is a failure in the making subsystem or there is a major failure in the packing subsystem.

- The operator and fitter both work together for repair only when there is a failure in making subsystem as it has priority of repair.
- The failure times and repair times follow exponential and general distributions respectively.
- The repaired system works as a new one.

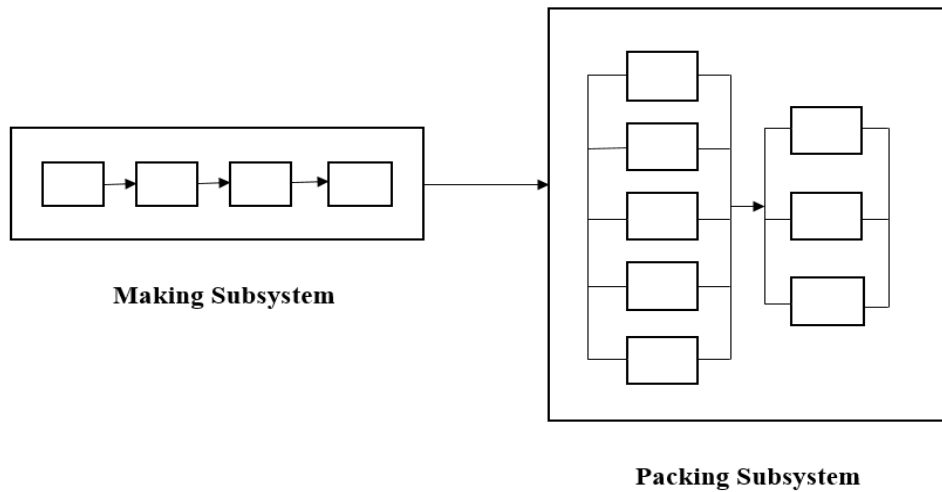


Figure 1: System Block Diagram

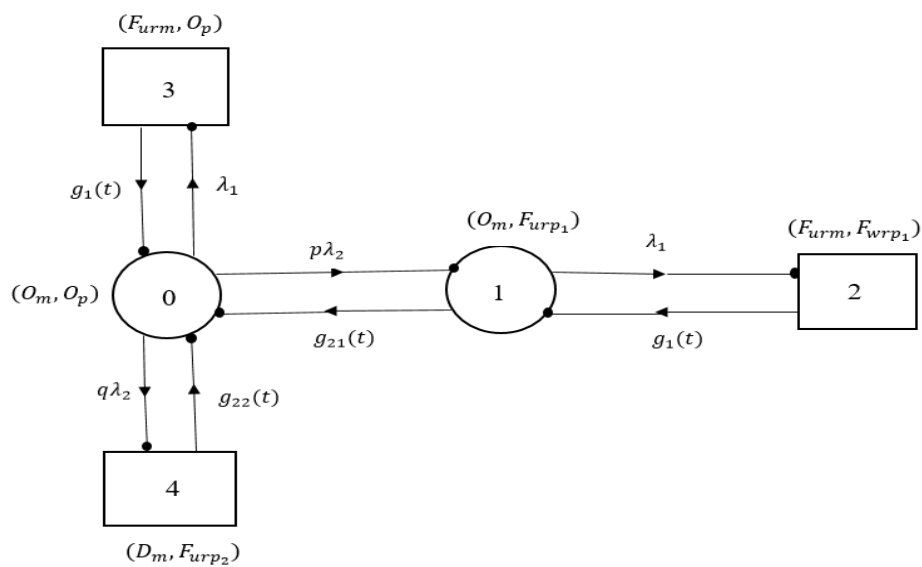


Figure 2: State Transition Diagram

4. Solution of Mathematical Model

The time points of entry into 0 to 4 states are the regeneration points and consequently, these states are termed as regenerative states. The states 0 and 1 are up states of the system whereas the states 2, 3, and 4 are failed states of the system. The transition probabilities for the states are mentioned as under:

$$q_{01}(t) = p\lambda_2 \cdot e^{-(\lambda_1+\lambda_2)t}, \quad q_{03}(t) = \lambda_1 \cdot e^{-(\lambda_1+\lambda_2)t}, \quad q_{04}(t) = q\lambda_2 \cdot e^{-(\lambda_1+\lambda_2)t}, \quad q_{10}(t) = g_{21}(t) \cdot e^{-\lambda_1 t},$$

$$q_{12}(t) = \lambda_1 e^{-\lambda_1 t} \cdot \overline{G_{21}(t)}, \quad q_{21}(t) = g_1(t), \quad q_{30}(t) = g_1(t), \quad q_{40}(t) = g_{22}(t).$$

Further, the non-zero elements p_{ij} , where $p_{ij} = \lim_{s \rightarrow 0} q_{ij}^*(s)$, are

$$p_{01} = \frac{p\lambda_2}{\lambda_1+\lambda_2}, \quad p_{03} = \frac{\lambda_1}{\lambda_1+\lambda_2}, \quad p_{04} = \frac{q\lambda_2}{\lambda_1+\lambda_2}, \quad p_{10} = g_{21}^*(\lambda_1), \quad p_{12} = 1 - g_{21}^*(\lambda_1), \quad p_{21} = p_{30} = p_{40} = 1.$$

From these, we can conclude that

$$p_{01} + p_{03} + p_{04} = 1, \quad p_{10} + p_{12} = 1.$$

Corresponding to the states 0 to 4, the mean sojourn times, μ_i are evaluated as:

$$\mu_0 = \frac{1}{\lambda_1+\lambda_2}, \quad \mu_1 = \frac{1-g_{21}^*(\lambda_1)}{\lambda_1}, \quad \mu_2 = -g_1^*(0), \quad \mu_3 = -g_1^*(0), \quad \mu_4 = -g_{22}^*(0).$$

Further, the unconditional mean times, m_{ij} used in the solution of the model are as given below:

$$m_{01} = \frac{p\lambda_2}{(\lambda_1+\lambda_2)^2}, \quad m_{03} = \frac{\lambda_1}{(\lambda_1+\lambda_2)^2}, \quad m_{04} = \frac{q\lambda_2}{(\lambda_1+\lambda_2)^2}, \quad m_{10} = -g_{21}^*(\lambda_1), \quad m_{12} = g_{21}^*(\lambda_1) + \frac{1-g_{21}^*(\lambda_1)}{\lambda_1},$$

$$m_{21} = -g_1^*(0), \quad m_{30} = -g_1^*(0), \quad m_{40} = -g_{22}^*(0).$$

Furthermore, we get

$$m_{01} + m_{03} + m_{04} = \mu_0, \quad m_{10} + m_{12} = \mu_1, \quad m_{21} = \mu_2, \quad m_{30} = \mu_3, \quad m_{40} = \mu_4.$$

The expressions for reliability measures and profit have been appraised in the upcoming subsections.

4.1. MTSF

The MTSF is the metric related to the failure of the system. It provides the expected time to which the system is operational before the complete failure. For determining the MTSF of the system, the failed states have been considered as the absorbing states. The resulting iterative relations for $\phi_i(t)$ are given below

$$\phi_0(t) = Q_{01}(t) \otimes \phi_1(t) + Q_{03}(t) + Q_{04}(t) \tag{1}$$

$$\phi_1(t) = Q_{10}(t) \otimes \phi_0(t) + Q_{12}(t) \tag{2}$$

Taking Laplace Stieltjes transform of equations (1-2) and solving them, we have

$$\phi_0^{**}(s) = \frac{N_1(s)}{D_1(s)}$$

where,

$$N_1(s) = \begin{vmatrix} q_{03}^*(s) + q_{04}^*(s) & -q_{01}^*(s) \\ q_{12}^*(s) & 1 \end{vmatrix} \text{ and } D_1(s) = \begin{vmatrix} 1 & -q_{01}^*(s) \\ -q_{10}^*(s) & 1 \end{vmatrix}$$

Assuming that the considered system initiates its journey from the state '0', the MTSF is calculated as:

$$\text{MTSF} = \lim_{s \rightarrow 0} \frac{(1 - \phi_0^{**}(s))}{s} = \frac{N_1}{D_1}$$

where,

$$N_1 = p_{01} \cdot \mu_1 + \mu_0 \text{ and } D_1 = 1 - p_{01} \cdot p_{10}.$$

4.2. Availability

Availability is a key performance indicator of a production system. This metric is used in the production process as it keeps an eye on the continuous operations of the system which results in terms of maximum profit for the manufacturers.

To determine the system availability, let us define $A_i(t)$ as the probability that starting from the regenerative state 'i', the system is in the up state at the instant 't'. Thus, the iterative relations will be:

$$A_0(t) = M_0(t) + q_{01}(t) \odot A_1(t) + q_{03}(t) \odot A_3(t) + q_{04}(t) \odot A_4(t) \quad (3)$$

$$A_1(t) = M_1(t) + q_{10}(t) \odot A_0(t) + q_{12}(t) \odot A_2(t) \quad (4)$$

$$A_2(t) = q_{21}(t) \odot A_1(t) \quad (5)$$

$$A_3(t) = q_{30}(t) \odot A_0(t) \quad (6)$$

$$A_4(t) = q_{40}(t) \odot A_0(t) \quad (7)$$

where,

$$M_0(t) = e^{-(\lambda_1 + \lambda_2)t} \text{ and } M_1(t) = e^{-\lambda_1 t} \cdot \overline{G_{21}(t)}.$$

Taking Laplace transform of equations (3-7) and then solving them further, the steady-state availability is evaluated as

$$A_0 = \lim_{s \rightarrow 0} \left[s \cdot \frac{N_2(s)}{D_2(s)} \right] = \frac{N_2}{D_2}$$

where,

$$N_2(s) = \begin{vmatrix} M_0^*(s) & -q_{01}^*(s) & 0 & -q_{03}^*(s) & -q_{04}^*(s) \\ M_1^*(s) & 1 & -q_{12}^*(s) & 0 & 0 \\ 0 & -q_{21}^*(s) & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix} \text{ and}$$

$$D_2(s) = \begin{vmatrix} 1 & -q_{01}^*(s) & 0 & -q_{03}^*(s) & -q_{04}^*(s) \\ -q_{10}^*(s) & 1 & -q_{12}^*(s) & 0 & 0 \\ 0 & -q_{21}^*(s) & 1 & 0 & 0 \\ -q_{30}^*(s) & 0 & 0 & 1 & 0 \\ -q_{40}^*(s) & 0 & 0 & 0 & 1 \end{vmatrix}$$

which gives

$$N_2 = \mu_0 \cdot p_{10} + \mu_1 \cdot p_{01} \text{ and } D_2 = p_{10} \cdot \mu_0 + p_{01} \cdot \mu_1 + (p_{12} \cdot p_{01} + p_{03} \cdot p_{10}) \cdot \mu_2 + p_{10} \cdot p_{04} \cdot \mu_4.$$

4.3. Busy period

In our study, we have two repairers i.e., operator and fitter. They have been assigned separate repair duties based on the failures in the subsystems. Their busy periods' expressions are evaluated in the next two subsections.

4.3.1. Operator's busy period for repair

To calculate the operator's busy period with the system, we have the following iterative relations for $B_i^o(t)$, where $B_i^o(t)$ is defined as the probability when the operator is fully engaged in repair at an instant 't', provided that the system moved in regenerative state 'i' at $t = 0$.

$$B_0^o(t) = q_{01}(t) \odot B_1^o(t) + q_{03}(t) \odot B_3^o(t) + q_{04}(t) \odot B_4^o(t) \quad (8)$$

$$B_1^o(t) = W_1(t) + q_{10}(t) \odot B_0^o(t) + q_{12}(t) \odot B_2^o(t) \quad (9)$$

$$B_2^o(t) = W_2(t) + q_{21}(t) \odot B_1^o(t) \quad (10)$$

$$B_3^o(t) = q_{30}(t) \odot B_0^o(t) \quad (11)$$

$$B_4^o(t) = q_{40}(t) \odot B_0^o(t) \quad (12)$$

where,

$$W_1(t) = e^{-\lambda_1 t} \cdot \overline{G_{21}(t)} \text{ and } W_2(t) = \overline{G_1(t)}.$$

Further, we have taken the Laplace transform of the aforementioned equations (8-12). The obtained busy period of the operator in the steady state is

$$B_0^o = \lim_{s \rightarrow 0} \left[s \cdot \frac{N_{31}(s)}{D_{31}(s)} \right] = \frac{N_{31}}{D_{31}}$$

where,

$$N_{31}(s) = \begin{vmatrix} 0 & -q_{01}^*(s) & 0 & -q_{03}^*(s) & -q_{04}^*(s) \\ W_1^*(s) & 1 & -q_{12}^*(s) & 0 & 0 \\ W_2^*(s) & -q_{21}^*(s) & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix} \text{ and}$$

$$D_{31}(s) = \begin{vmatrix} 1 & -q_{01}^*(s) & 0 & -q_{03}^*(s) & -q_{04}^*(s) \\ -q_{10}^*(s) & 1 & -q_{12}^*(s) & 0 & 0 \\ 0 & -q_{21}^*(s) & 1 & 0 & 0 \\ -q_{30}^*(s) & 0 & 0 & 1 & 0 \\ -q_{40}^*(s) & 0 & 0 & 0 & 1 \end{vmatrix}$$

which gives

$$N_{31} = p_{01} \cdot (\mu_1 + \mu_2 \cdot p_{12}) \text{ and } D_{31} = p_{10} \cdot \mu_0 + p_{01} \cdot \mu_1 + (p_{12} \cdot p_{01} + p_{03} \cdot p_{10}) \cdot \mu_2 + p_{10} \cdot p_{04} \cdot \mu_4.$$

4.3.2. Fitter's busy period for repair

For the fitter's busy period with the system, the following recursive relations can be obtained where $B_i^f(t)$ is defined as the probability when the fitter is fully engaged at an instant 't', provided that the system moved in regenerative state 'i' at $t = 0$.

$$B_0^f(t) = q_{01}(t) \odot B_1^f(t) + q_{03}(t) \odot B_3^f(t) + q_{04}(t) \odot B_4^f(t) \tag{13}$$

$$B_1^f(t) = q_{10}(t) \odot B_0^f(t) + q_{12}(t) \odot B_2^f(t) \tag{14}$$

$$B_2^f(t) = W_2(t) + q_{21}(t) \odot B_1^f(t) \tag{15}$$

$$B_3^f(t) = W_3(t) + q_{30}(t) \odot B_0^f(t) \tag{16}$$

$$B_4^f(t) = W_4(t) + q_{40}(t) \odot B_0^f(t) \tag{17}$$

where,

$$W_2(t) = \overline{G_1(t)}, W_3(t) = \overline{G_1(t)} \text{ and } W_4(t) = \overline{G_{22}(t)}.$$

The Laplace transform of the equations (13-17) is considered, and we obtain the busy period of the fitter as:

$$B_0^f = \lim_{s \rightarrow 0} \left[s \cdot \frac{N_{32}(s)}{D_{32}(s)} \right] = \frac{N_{32}}{D_{32}}$$

where,

$$N_{32}(s) = \begin{vmatrix} 0 & -q_{01}^*(s) & 0 & -q_{03}^*(s) & -q_{04}^*(s) \\ 0 & 1 & -q_{12}^*(s) & 0 & 0 \\ W_2^*(s) & -q_{21}^*(s) & 1 & 0 & 0 \\ W_3^*(s) & 0 & 0 & 1 & 0 \\ W_4^*(s) & 0 & 0 & 0 & 1 \end{vmatrix} \text{ and}$$

$$D_{32}(s) = \begin{vmatrix} 1 & -q_{01}^*(s) & 0 & -q_{03}^*(s) & -q_{04}^*(s) \\ -q_{10}^*(s) & 1 & -q_{12}^*(s) & 0 & 0 \\ 0 & -q_{21}^*(s) & 1 & 0 & 0 \\ -q_{30}^*(s) & 0 & 0 & 1 & 0 \\ -q_{40}^*(s) & 0 & 0 & 0 & 1 \end{vmatrix}$$

which gives

$$N_{32} = p_{10} \cdot p_{04} \cdot \mu_4 + p_{03} \cdot p_{10} \cdot \mu_2 + p_{12} \cdot p_{01} \cdot \mu_2 \quad \text{and}$$

$$D_{32} = p_{10} \cdot \mu_0 + p_{01} \cdot \mu_1 + (p_{12} \cdot p_{01} + p_{03} \cdot p_{10}) \cdot \mu_2 + p_{10} \cdot p_{04} \cdot \mu_4.$$

4.4. Profit

The main purpose of every organization is to gain profit to survive in this competitive era. Reliability analysis helps us to provide profitable strategies with the aid of measures of system effectiveness. The profit earned by the system can be evaluated by the following expression

$$\text{Profit} = C_0 \cdot A_0 - (C_{ro} \cdot B_0^o + C_{rf} \cdot B_0^f)$$

where, C_0 is revenue per unit up time of the system and C_{ro}, C_{rf} refer to the respective service costs per unit time for which the operator and fitter are busy.

5. Numerical Calculation

Based on gathered data from the food industrial plant, we have estimated the rates and probabilities involved in the model. The estimated rates and probabilities are given in Table 3.

Table 3: Rates and probabilities

Failure rate of making subsystem	0.08
Failure rate of packing subsystem	0.04
Repair rate of making subsystem	1.64
Repair rate of packing subsystem due to minor failures	5.37
Repair rate of packing subsystem due to major failures	1.76
Probability that there are minor failures in packing subsystem	0.22
Probability that there are major failures in packing subsystem	0.78

The considered system has been assessed by assuming the repair time for making subsystem as well as packing subsystem for both minor and major failures as exponentially distributed with parameters β_1, β_{21} , and β_{22} respectively, so that $g_1(t) = \beta_1 e^{-\beta_1 t}$, $g_{21}(t) = \beta_{21} e^{-\beta_{21} t}$, $g_{22}(t) = \beta_{22} e^{-\beta_{22} t}$.

Correspondingly, we have $p_{10} = \frac{\beta_{21}}{\lambda_1 + \beta_{21}}$, $p_{12} = \frac{\lambda_1}{\lambda_1 + \beta_{21}}$, $\mu_1 = \frac{1}{\lambda_1 + \beta_{21}}$, $\mu_2 = \frac{1}{\beta_1}$ and $\mu_4 = \frac{1}{\beta_{22}}$. The costs involved are as; $C_0 = 100000$, $C_{ro} = 1600$, and $C_{rf} = 3000$. Using the rates mentioned in Table 3, we have obtained measures of system effectiveness as $MTSF = 8.996875124$, $A_0 = 0.937665183$, $B_0^o = 0.001552568$, and $B_0^f = 0.062334817$.

6. Result and Discussion

We have noticed that although the system performance is governed by the various parameters involved, yet the aspect of failure rates of both subsystems is the base of our study. The impact of both subsystems' failure rates on the MTSF, system availability, and overall profit is examined graphically. From Figure 3, it is observed that MTSF decreases with the increase in the failure rate of the making as well as the packing subsystem. The availability of the system is effected by the failure rates of both subsystems which is evident from Figure 4. Here, from the trend of the availability of the system with the failure rates of the subsystems, it is clear that availability declines with the rise in failure rates of the making subsystem and the packing subsystem.

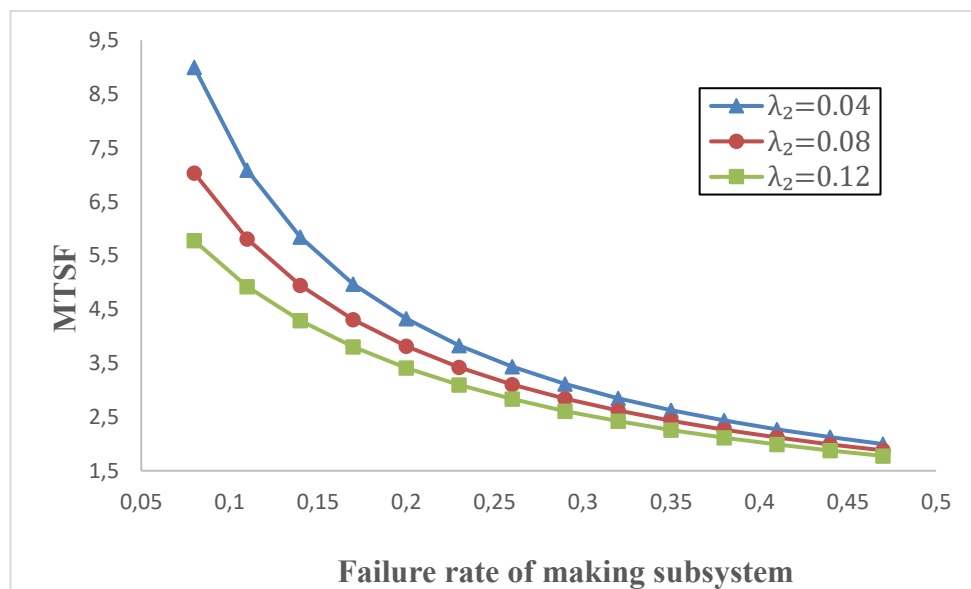


Figure 3: MTSF v/s Failure rate of making subsystem for various values of λ_2

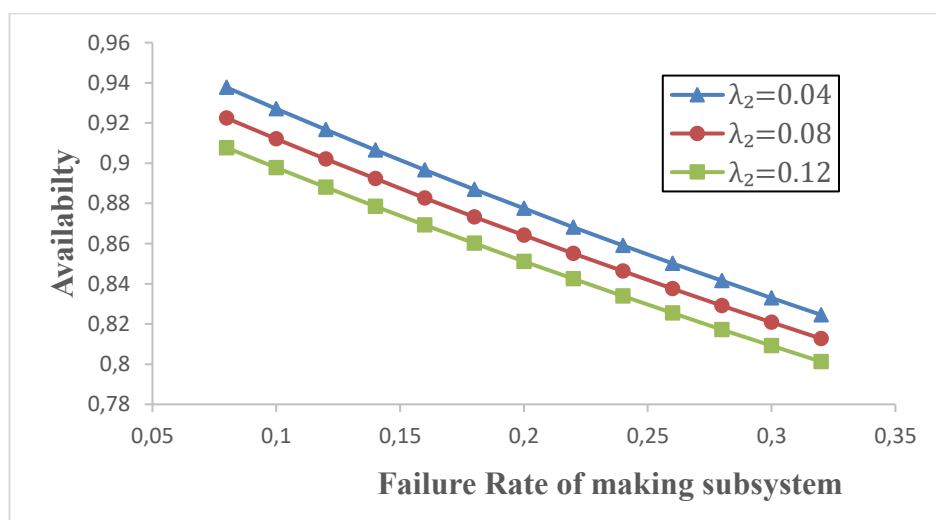


Figure 4: Availability v/s Failure rate of making subsystem with various values of λ_2

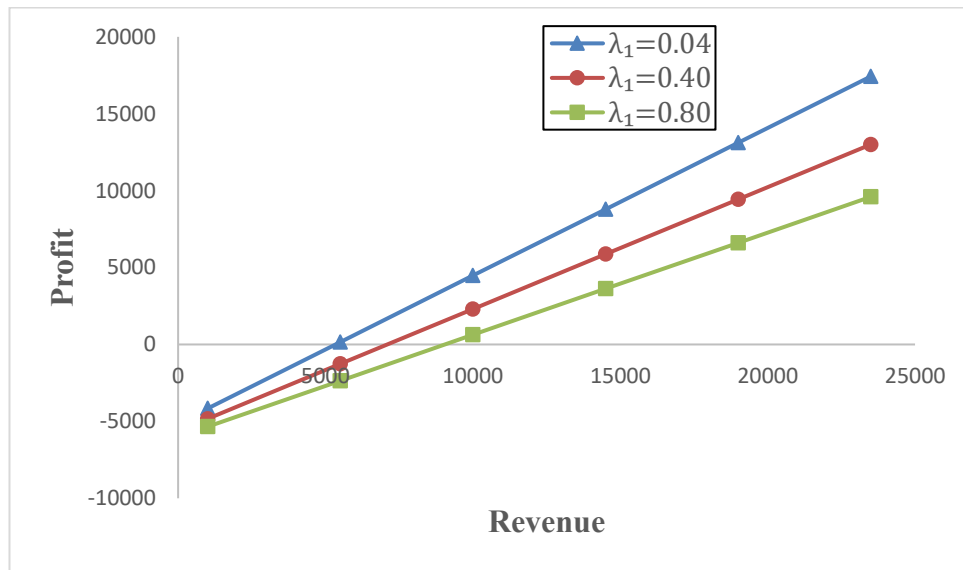


Figure 5: Profit v/s system revenue for various values of λ_1

For studying the effect of failure rates on system profit, we have fixed some parameters such as $\beta_1 = 1.64$, $\beta_{21} = 5.37$, $\beta_{22} = 1.76$, $p = 0.22$, $q = 0.78$, $C_{ro} = 1600$ and $C_{rf} = 3000$. To examine the effect of λ_1 on system profit, we have varied the parameter λ_1 as 0.04, 0.40, and 0.80. Figure 5 indicates that for the varying λ_1 as 0.04, 0.40, and 0.80, the profit is positive or zero or negative when the system revenue i.e., C_0 is greater or equal or lesser than 5339.423, 7096.091, and 9047.954 respectively. Further, we have explored the effect of λ_2 on profit, by varying the parameter λ_2 as 0.04, 0.40, and 0.80. It can be seen in Figure 6 that for λ_2 as 0.04, 0.40, and 0.80, the profit is positive or zero or negative when the system revenue i.e., C_0 is greater or equal or lesser than 5534.609, 7096.091, and 9772.342 respectively.

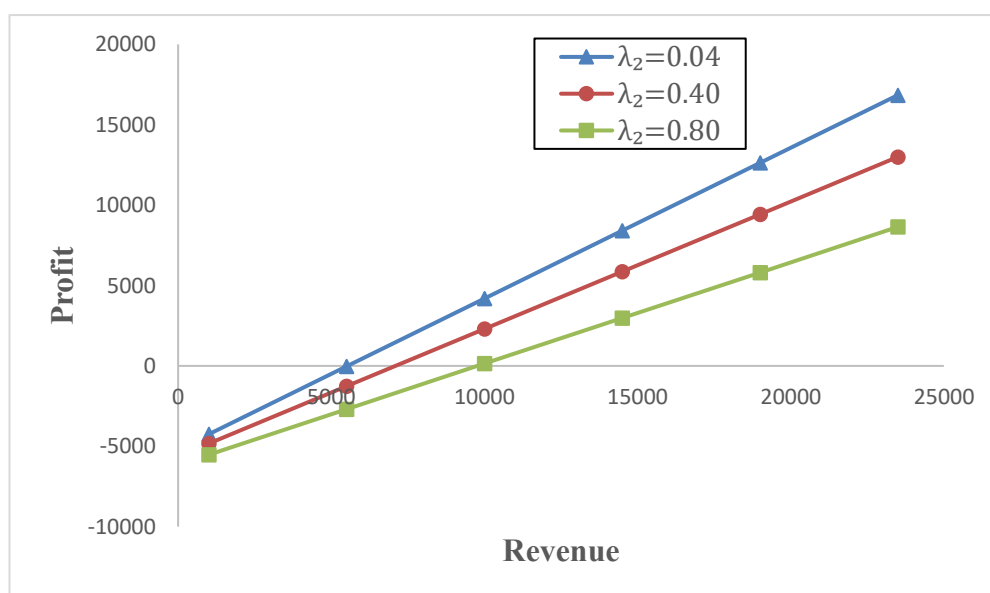


Figure 6: Profit v/s system revenue for various values of λ_2

7. Conclusion

In the present paper, we have figured out the MTSE, availability, and profitability of the food industrial system that is based on the make-and-pack production process in the aspect of the failure rates of the subsystems. The graphical interpretation confirms that both reliability metrics, MTSE, and system availability reduce with the surge in time and failure rates of both making and packing subsystems. It is also found true that the profit of the studied system increases with the increase in the system revenue and decreases when the failure rates of the subsystems increase. So, it is concluded that for lower values of failure rates, the expected profit is high as compared to higher values of failure rates. As far as the subsystems are concerned, it is graphically observed that for the same value of failure rate i.e., $\lambda_1, \lambda_2 = 0.04$, more revenue is needed for the packing subsystem as compared to the making subsystem for profitability. In other words, we can say that for the same number of failures in making as well as packing subsystems, the packing subsystem for some values of revenue results in loss to the system and at the same time making subsystem provides profit with those values. Therefore, priority control of the failures of packing subsystem is more advisable for the system's profitability as when failures will be controlled then revenue will be more. Such valuations can help to assess the failure times of the subsystems that can be afforded by the manufacturers. The developed model can be compared to the case when there is no priority basis repair to see the difference in earned profits.

Conflict of Interest Declaration: The authors have no conflicts of interest to declare.

References

- [1] Johnson, S. M. (1954). Optimal two- and three-stage production schedules with setup times included. *Naval Research Logistics Quarterly*, 1: 61-68.
- [2] Gupta, J. N. D. and Tunc, E. A. (1991). Schedules for a two-stage hybrid flowshop with parallel machines at the second stage. *INT. J. PROD. RES.*, 29(7):1489-1502.
- [3] Honkomp, S. J., Lombardo, S., Rosen, O. and Pekny, J. F. (2000). The curse of reality – why process scheduling optimization problems are difficult in practice. *Computers and Chemical Engineering*, 24: 323-328.
- [4] Akkerman, R., Donk, D. P. V. and Gaalman, G. (2007). Influence of capacity-and time-constrained intermediate storage in two-stage food production systems. *International Journal of Production Research*, 45(13): 2955-2973.
- [5] Sel, C., Bilgen, B. and Ruwaard, J. B. (2017). Planning and scheduling of the make-and-pack dairy production under lifetime uncertainty. *Applied Mathematical Modelling*, 51:129-144.
- [6] Klanke, C., Yfantis, V., Corominas, F. and Engell, S. (2021). Short-term scheduling of make-and-pack processes in the consumer goods industry using discrete-time and precedence-based MILP models. *Computers and Chemical engineering*, 154: 107453.
- [7] Gaver, D. P. (1963). Time to failure and availability of paralleled systems with repair. *IEEE Transactions on Reliability*, 12(2): 30-38.
- [8] Pandey, D., Jacob, M. and Tyagi, S. K. (1996) Stochastic modelling of a powerloom plant with common cause failure, human error and overloading effect. *International Journal of Systems Science*, 27(3): 309-313.
- [9] Mathew, A. G., Rizwan, S. M., Majumder, M. C. and Ramachandran, K. P. (2011). Reliability modelling and analysis of a two unit continuous casting plant. *Journal of the Franklin Institute*, 348(7): 1488-1505.
- [10] Taj, S. Z., Rizwan, S. M., Alkali, B. M., Harrison, D. K. and Taneja, G. (2017). Reliability modelling and analysis of a single machine subsystem of a cable plant. *7th International Conference on Modeling, Simulation, and Applied Optimization (ICMSAO)*, 2017: 1-4.
- [11] Kumari, P., Kadyan, M. S., and Kumar, J. (2019). Profit analysis of butter-oil (ghee) producing system of milk plant using supplementary variable technique. *International Journal of System Assurance Engineering and Management*, 10(6): 1627-1638.
- [12] Singh, V. V., Poonia, P. K. and Adbullahi, A. H. (2020). Performance analysis of a complex repairable system with two subsystems in series configuration with an imperfect switch, *Journal of Mathematical and Computational Science*, 10(2): 359-383.
- [13] Saini, M., Devi, K., and Kumar, A. (2021). Stochastic modeling of a non-identical redundant System with priority in repair activities. *Thailand Statistician*, 19(1): 154–161.
- [14] Bashir, N., Joorel, J.P.S., and Jan, T. R. (2021). Reliability analysis of two unit standby model with controlled and uncontrolled failure of unit and replacement facility available in the system. *Pakistan Journal of Statistics and Operation Research*, 17(3): 625-632.
- [15] Andalib, V. and Sarkar, J. (2021). A repairable system supported by two spare units and serviced by two types of repairers. *Journal of Statistical Theory and Applications*, 20(2): 180-192.
- [16] Sharma, U. and Drishti (2022). The seasonal effect of working conditions of an ice-cream plant. *Reliability: Theory & Applications*, 17(4): 192-203.

- [17] Monika and Chopra, G. (2022). Reliability modeling of a food industrial system with two types of repair persons wherein demand is season dependent. *International Journal of Mathematical, Engineering and Management Sciences*, 7(6): 918-937.
- [18] Rizwan, S. M., Nabhani, H. A., Rahbi, Y. A. and Alagiriswamy, S. (2022). Reliability analysis of a three-unit pumping system. *International Journal of Engineering Trends and Technology*, 70(6): 31.
- [19] Sachdeva, K., Taneja, G. and Manocha, A. (2022). Sensitivity and economic analysis of an insured system with extended conditional warranty. *Reliability: Theory & Applications*, 17(3): 315-327.
- [20] Yusuf, I. and Sanusi, A. (2022). Reliability assessment and profit analysis of automated teller machine system under copular repair policy. *Predictive Analytics in System Reliability (Springer Series in Reliability Engineering Book)*: 97-117.

HYBRID AND BLIND WATERMARKING FRAMEWORK FOR PRIVACY PROTECTION AND CONTENT AUTHENTICATION OF DIGITAL MULTIMEDIA

Swati J. Patel¹

Dr. Mehul C. Parikh²

•

¹Ph.D. Scholar, Information Technology Department, Gujarat Technological University, Ahmedabad, India. swatijpatel@ldce.ac.in

²Associate Professor, Information Technology Department, L.D.College of Engineering, Ahmedabad, India. mehulparikh@ldce.ac.in

Abstract

Nowadays, due to inexpensive and conveniently available internet access at the fingertip, the illegitimate sharing of digital multimedia i.e. image, audio, and video is becoming a universal and significant threat. Illegal transmission of digital multimedia through the internet creates an issue of authentication and copyright protection; hence, piracy protection is a vital need for protecting digital media. Digital watermarking is a method of preventing digital theft in which additional information, known as a watermark, is inserted into digital multimedia. This technology was originally designed for still photos, but it has subsequently been expanded to include additional multimedia artifacts such as audio and video, due to its countless use in today's era. Digital watermarking is an effective method of limiting piracy and providing authenticity and copyright ownership to digital content. Watermarking can be performed either in the spatial or in the transform domain. In this paper, a hybrid digital video watermarking technique for copyright protection, data security, and content authentication of multimedia, based on Discrete Wavelet Transform, Discrete Cosine Transform, and Singular Value Decomposition is presented. The authenticity of the content has been ensured by embedding a watermark in the transform domain, while copyright protection has been provided by a strong watermark. The experimental results show that the proposed schemes achieve a PSNR greater than 51 dB on average, which illustrates that the proposed method gives excellent performance for robustness, authentication, and security. A comparison of the proposed framework to various cutting-edge techniques illustrates its effectiveness and superiority.

Keywords: Digital Video Watermarking, Robustness, Imperceptible, Authentication, Copyright Protection, Discrete Wavelet Transform, Discrete Cosine Transform, Singular Value Decomposition

1. Introduction

Almost everyone's smartphone is connected to one of the several widely available and fairly priced high-speed internet networks in today's world. People's conceptions of information transmission have been radically altered by the power of digitalization, which now allows them to send and receive text, data, images, and even films using their mobile devices. They are inadvertently exposing themselves to a serious technological risk of having their personal information misused. Anyone can easily copy and distort someone else's work, resulting in a slew of piracy-related concerns as well as a detrimental influence on the owner's financial benefits and intellectual property rights [1][2]. The

challenges of video watermarking are greater than those of image watermarking. The data in the video is inherently more scattered between frames. Uneven distribution of moving and stationary areas. The complete video watermarking process can be divided into three stages, as indicated in figure 1: watermark attachment or embedding, watermark sending or distribution over the channel, and watermark extraction or detection [3].

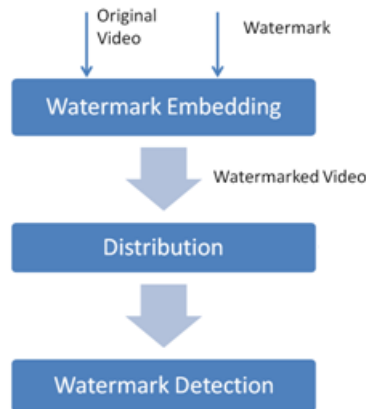


Figure 1: Watermarking Process

Section 2 represents the literature review. Section 3 introduces the three main concepts of the paper i.e., Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Singular Value Decomposition (SVD). Section 4 introduces the proposed method of video watermarking. Section 5 represents experimental results and discussion. Section 6 gives the concluding remarks.

2. Literature Review

Digital watermarking has three parameters: data payload, fidelity, and robustness. Digital watermarking has long been used to protect still photos, but it is now employed to protect audio and video files[4]. A comparison of different works is illustrated in Table 1.

Table 1: Summary of earlier work done

Ref.	Technique used	Purpose	Scheme Type	Results	Attacks	Remark
[5]	PCA, RDWT	Robustness, Imperceptibility	Blind	PSNR = 67.87, MSE = 0.0106, NCC = 0.995, SSIM = 0.9997	GN, Rot, Sc, Compression, RS	It can be improved to work for color images.
[6]	DWT, SVD, Entropy, Pixel position shuffling	Robustness, Imperceptibility, Security	Non-Blind	PSNR= 42.6369, NC = 1	JPEG, Rot, GN, S&P	It is not time Efficient
[7]	DCT, DWT, Arnold	Robustness, Imperceptibility	Non-Blind	PSNR = 47.1836, NC = 0.1936	Rot, Noise, RS, JPEG	No balance between PSNR and NC value is achieved
[8]	2D -DWT,	Robustness,	Non-	PSNR = 54.96,	S&P, Rot,	Not good for

	Encryption	Imperceptibility, Security	Blind	MSE = 0.2047, CC = 0.9749	JPEG	geometric attacks and have low capacity
[9]	DWT, SVD, Block selection scheme	Robustness, Security	Non- Blind	PSNR = 61.7524, SSIM = 0.9999, Cor = 1	GN, S&P, Rot, JPEG	It is not good for active attacks
[10]	DCT, DWT, SVD, Arnold	Robustness, Imperceptibility, Security	Non- Blind	PSNR = 43.88, NC = 0.9888, BER = 0.2174	JPEG, S&P, GN, Rot, Cr, RS	For the high gain factor, the quality of the watermarked image is poor

Abbreviations used for the various attack:

GN: Gaussian Noise

Rot: Rotation

Sc: Scaling

RS: Resize

Cr: Cropping

S&P: Salt and Pepper

3. Transform Domain Methods

The video frame of a host video sequence is translated into the new domain, and then the inverse frequency domain is applied after embedding a watermark. Data is converted using transforms like the Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Singular Value Decomposition (SVD). Each transform has its properties for video frame representation. This approach offers the advantages of being more resistant to malicious attacks, more stable, and more imperceptible than a spatial domain [11] [12].

3.1 Discrete Wavelet Transform (DWT):

In discrete wavelet transform, video frame pixels are converted into wavelets, which can subsequently be used for wavelet-based compression and coding [13]. A video frame is divided into four bands using a mathematical technique called lower resolution approximation (LL), vertical (LH), horizontal (HL), and diagonal (D) (HH). The low-frequency component is designated as LL, while the high-frequency sections are designated as LH, HH, and HL. Multiple-scale wavelet decompositions can be obtained by repeating the process as shown in figure2(b). The low-frequency district information is a frame that is quite similar to the original frame. This frequency district contains the majority of the original frame's signal information. The level detail, upright detail, and diagonal detail of the original image are represented by the frequency districts LH, HL, and HH, respectively. The watermark can be inserted in these three subbands to retain higher image quality because human eyes are significantly more sensitive to the low-frequency component (the LL subband). They are sensitive to changes in the smooth district of a frame, but not too small changes in the edge, profile, or streak, according to the HVS character [14]. The watermark's robustness is increased by embedding it in higher-level sub-bands. However, the visual integrity may be compromised, as evaluated by PSNR. In the high-frequency region, the edges and texture can be easily distinguished using the DWT. As a result, it's difficult to see that inserting the watermarking signal into the frame's DWT-converted high-

frequency band has a large amplitude coefficient.

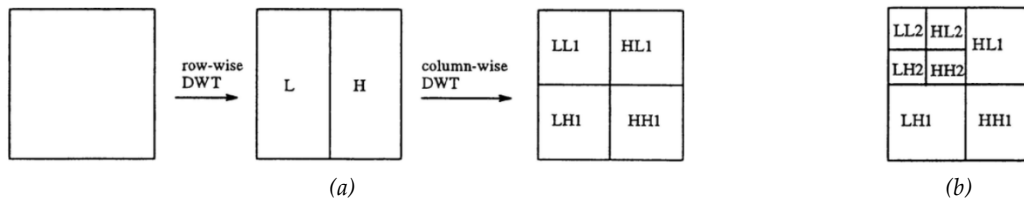


Figure 2: (a) First level of decomposition (b) Second level of decomposition

3.2 Discrete Cosine Transform (DCT):

Watermarking based on DCT can be divided into two types: global and block-based. A full DCT is conducted on the entire frame in the former, whereas a frame is divided into non-overlapping blocks of a particular size and the DCT is performed on each block [1] [15]. The DCT technique separates the coefficients into three frequency bands: low, middle, and high, with the mid-frequency band being the most commonly used for watermark embedding. Because the low-frequency band and DC component contain the greatest signal energy per frame, they are the most critical portions of a video signal, and any change to this band has a significant impact on how the human eye perceives distortion both are vulnerable to watermarking attacks [16]. Watermarking with DCT is based on two facts. The first is that much of the signal energy is concentrated in the low-frequency sub-band, which contains the frame's most significant visual elements. The second fact is that high-frequency frame components are frequently deleted due to compression and noise attacks [14]. DCT can be conducted in a variety of dimensions, including 1D, 2D, and 3D. In the context of images or frames, 2D DCT is used to convert them into a cosine series [15].

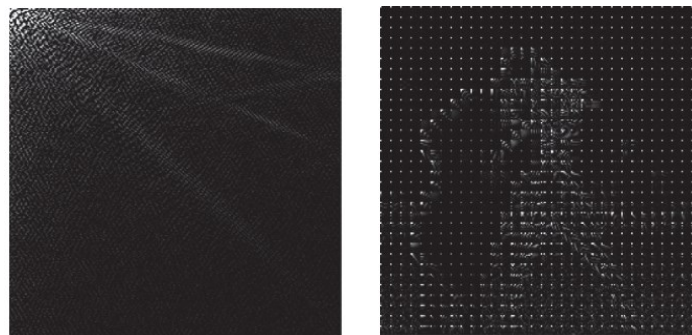


Figure 3: (a) The whole frame (Global) (b) Block-wise process

1D-DCT is given as:

$$f(i) = F(u) \tag{1}$$

2D-DCT is given as:

$$F(u, v) = \alpha(u)\alpha(v) \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} f(i, j) \cos \left[\frac{(2i + 1)u\pi}{2M} \right] \cos \left[\frac{(2j + 1)v\pi}{2M} \right]$$

Where $u, v = 0, 1, 2, \dots, M-1$, M is size of sequence; $f(i, j)$ is image in spatial domain and $F(u, v)$ is in frequency domain

$$\alpha(u) = \left(\frac{1}{\sqrt{M}} \text{ if } u = 0 \right) \text{ or } \left(\sqrt{\frac{2}{M}} \text{ if } u \neq 0 \right) \tag{2}$$

4. Singular Value Decomposition (SVD):

Singular Value Decomposition is a technique used in a variety of applications, including image compression, image concealment, and image watermarking. A frame is an array of nonnegative scalar entries that can be treated as a matrix from the viewpoint of linear algebra. A matrix can be decomposed into three matrices of the same size as the original matrix using the SVD transformation [14]. Considering a frame of the video sequence is a square matrix of size $M \times M$, the formula for SVD is defined as,

$$f = USV^T \quad (3)$$

Where U and V are orthogonal (or unitary) matrices and S is a diagonal matrix, with the diagonal elements in the descending order of S being called the singular values of the frame (f) and V^T is the transpose of a $M \times M$ matrix containing the orthonormal eigenvectors.

Watermarking strategies based on SVD embed the watermark by changing U , V , or S . Due to the strong stability of the singular values, SVD techniques are commonly used in video watermarking. This means that when a little perturbation is applied to a frame, these values do not fluctuate much. Although this characteristic of the SVD provides robustness to attacks, a limitation is that performing it on an image is computationally expensive[17].

5. Proposed Method

We proposed a DWT, DCT, and SVD-based hybrid watermarking technique using the salient properties of DWT, DCT, and the SVD in our suggested watermarking scheme. The algorithm for watermark embedding and extraction processes are described in algorithm 1 and algorithm 2 respectively, which shows how the watermark is embedded with the host frame and how the embedded watermark is extracted from the attacked watermarked frame. First of all Host video is taken and converted into a sequence of frames. Among them all video frames, the keyframes are selected for further processing. The proposed approach for watermarking is depicted in figure 4.

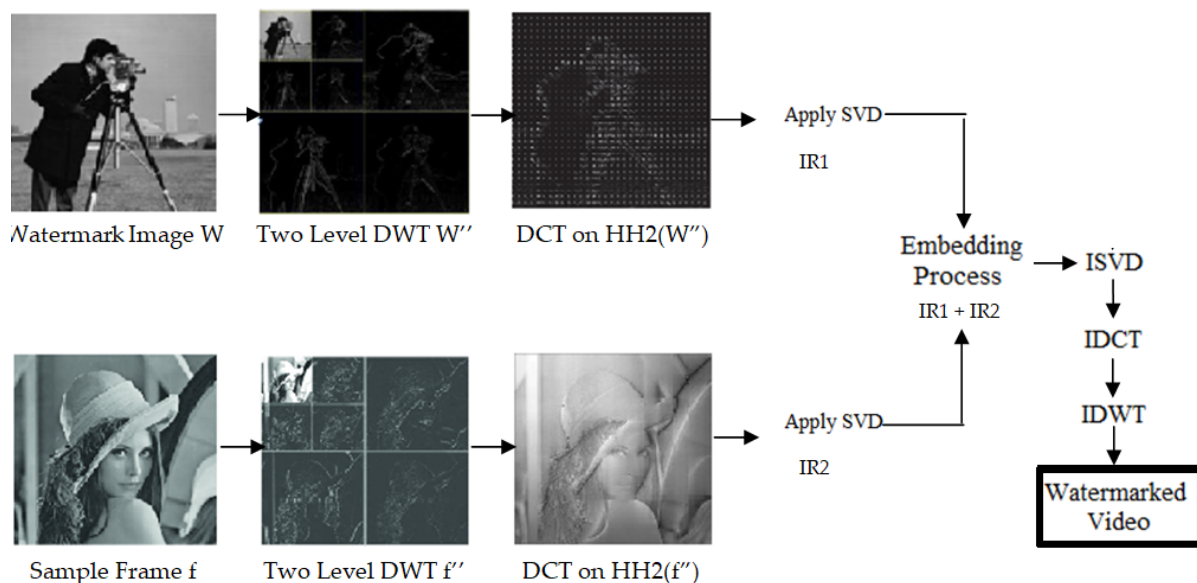


Figure 4: Schematic Representation of the Proposed System

Algorithm 1: Watermark Embedding Process

```

Host Video V, f = Frames of (V)
Watermark Image W
W' = DWT (W)
W'' = DWT (W') // Two Level DWT
Result1 = DCT on HH2(W'') as per eqn. (2)
IR1 = SVD (Result1) as per eqn. (3) // Intermediate Result1
fRow = All frames of Video V
for f = 1 to n where f = All keyframes of Video V
{
f' = DWT(f)
f'' = DWT(f')
Result 2 = DCT on HH2(f'') as per eqn. (2)
IR2 = SVD (Result 2) as per eqn. (3) // Intermediate Result 2
IR3 = IR1 + IR2 // Embedding Process
RR1 = ISVD (IR3)
RR2 = IDCT (RR1)
Ri = IDWT (RR2)
Final Result FR =  $\sum_{i=1}^n Ri$ 
}

```

Algorithm 2: Watermark Extraction Process

```

Watermarked Video FR
Rawf = frames (FR)
F = keyframe (Rawf)
For f=1 to n where f = All keyframes of Video FR
{
f' = DWT(f)
f'' = DWT(f')
R1 = DCT on HH2(f'') as per equ (2)
IR1 = SVD (R1) as per equ (3)
R2 = IDCT (S (IR1)) on singular values only.
R = IDWT (R2)
Final Extracted Result FER =  $\sum_{i=1}^n Ri$ 
}

```

6. Experimental Results and Discussion

In this proposed watermarking algorithm various mp4 videos of different backgrounds and different lengths are used. Videos are converted into a sequence of frames, frame size is 1280×720. Figure 5 shows the cameraman's image taken as a watermark of size 256×256. Figure 6 shows a frame of videos taken as input videos. Imperceptibility and robustness are the attributes that are assessed for the proposed scheme based on PSNR and MSE values. The term imperceptibility refers to the fact that the video's quality should not be compromised once the watermark is applied. After embedding the watermark, Peak Signal to Noise Ratio (PSNR) is used to calculate the imperceptibility, the degradation in the watermarked video, compared to the host video. PSNR is expressed as a decibel scale. Higher the value of PSNR higher the quality of the video. PSNR is represented as shown in equation (4). Watermark robustness refers to the fact that the watermark is not destroyed after intentional or unintentional attacks and may still be utilized to offer certification. It is calculated "after the attack." Mean Square Error (MSE) measures the mean of the square of the original watermark and

the extracted watermark from the attacked frame for robust capabilities. The lower the value of MSE lower will be the error. MSE is represented as shown in equation (5).

$$PSNR = 10 \log_{10} \left[\frac{P^2}{MSE} \right] \tag{4}$$

Where P is the peak signal value. P is equal to 255 for frames having a channel depth of 8-bit.

$$MSE = \frac{1}{ab} \sum_{x=1}^a \sum_{y=1}^b [c(x,y) - e(x,y)]^2 \tag{5}$$

Where a and b are the height of the original frame and distorted frame, respectively. $c(x,y)$ is the pixel value of the host video frame and $e(x,y)$ is the pixel value of the embedded video frame.



Figure 5: Watermark



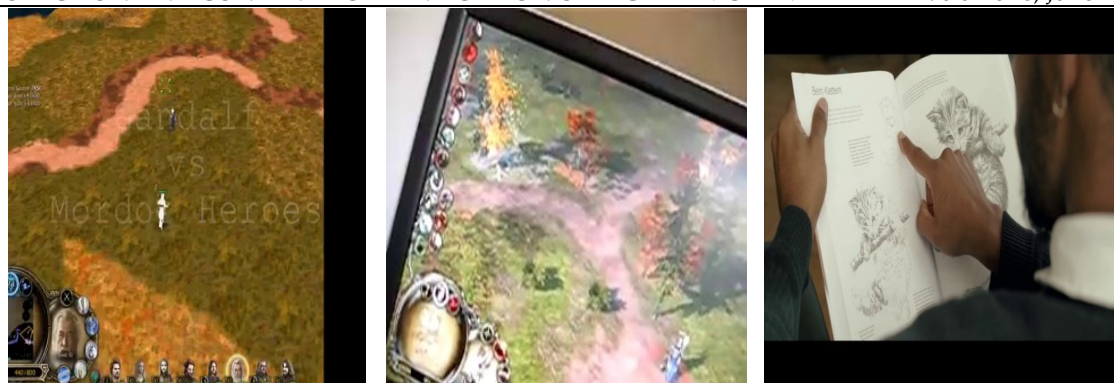
Test Video 1
(HonourableVCSirMessage)



Test Video 2
(Battle for Middle Earth 2 Goblins Fighting)



Test Video 3
(Filmstro & Film Riot One Minute Short Film)



Test Video 4
 (Gandalf vs Mordor Heroes)

Test Video 5
 (Lord Of The Rings Battle For
 Middle Earth 2)

Test Video 6
 (Perfection - 1 Minute Short Film)

Figure 6: Frame of various host videos

The PSNR values of the watermarked videos are shown in Table 2. These values demonstrate the scheme's undetectable property, as the PSNR values are high, implying that there is very little quality distortion after embedding the watermark. The graphical representations of the same are shown in figure 7.

Table 2: Performance results in terms of PSNR (in dB)

Test Data	DCT+SVD	DCT+DWT+ SVD	DWT+DCT+SVD
Test Video 1	54.9353	51.9685	55.1118
Test Video 2	53.8887	51.4956	54.1171
Test Video 3	53.9193	53.6994	54.0628
Test Video 4	51.8861	51.2743	52.5427
Test Video 5	50.8698	50.0266	51.5588
Test Video 6	53.8909	52.9037	54.5608

The MSE values of the watermark are shown in Table 3. These MSE values show the average term difference between the original video and the output, watermarked video. The lower the MSE values show the lower the level of degradation. The graphical representations of the same are shown in figure 8.

Table 3: Values of MSE of the watermark embedded in various host videos

Test Data	DCT+SVD	DCT+DWT+ SVD	DWT+DCT+SVD
Test Video 1	3.2097287e-06	6.3554344e-06	3.0818992e-06
Test Video 2	3.2443077e-06	4.086472e-06	3.0875129e-06
Test Video 3	3.2215482e-06	4.266353e-06	3.0920397e-06
Test Video 4	7.2462806e-06	7.457069e-06	7.0825214e-06
Test Video 5	7.2585012e-06	9.938821e-06	6.984163e-06
Test Video 6	4.2426826e-06	5.1242314e-06	3.498743e-06

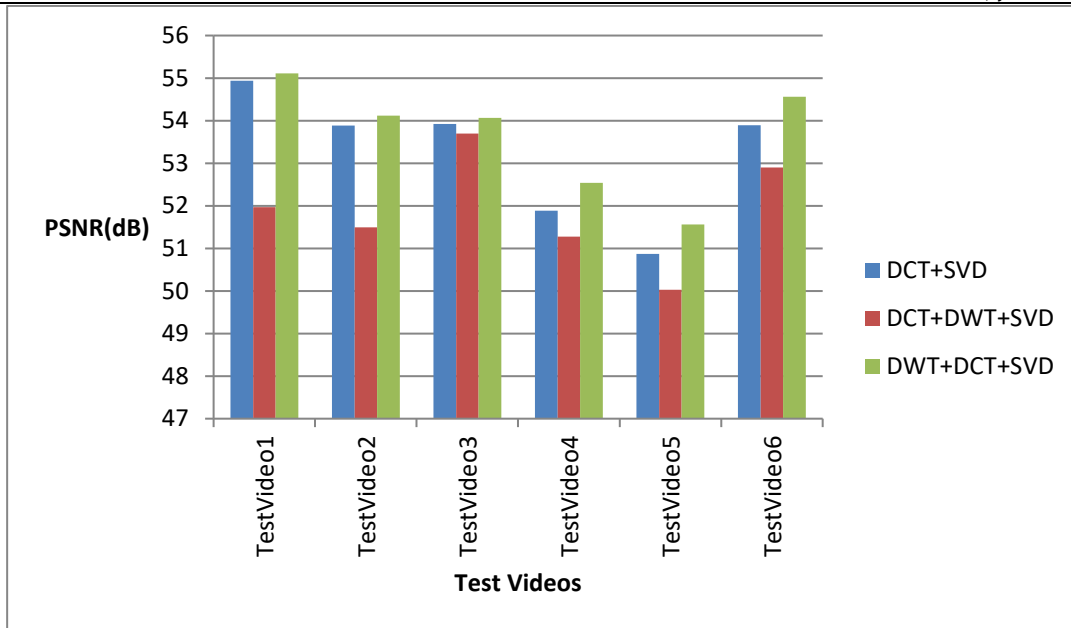


Figure 7: Comparison between various combinations for imperceptibility

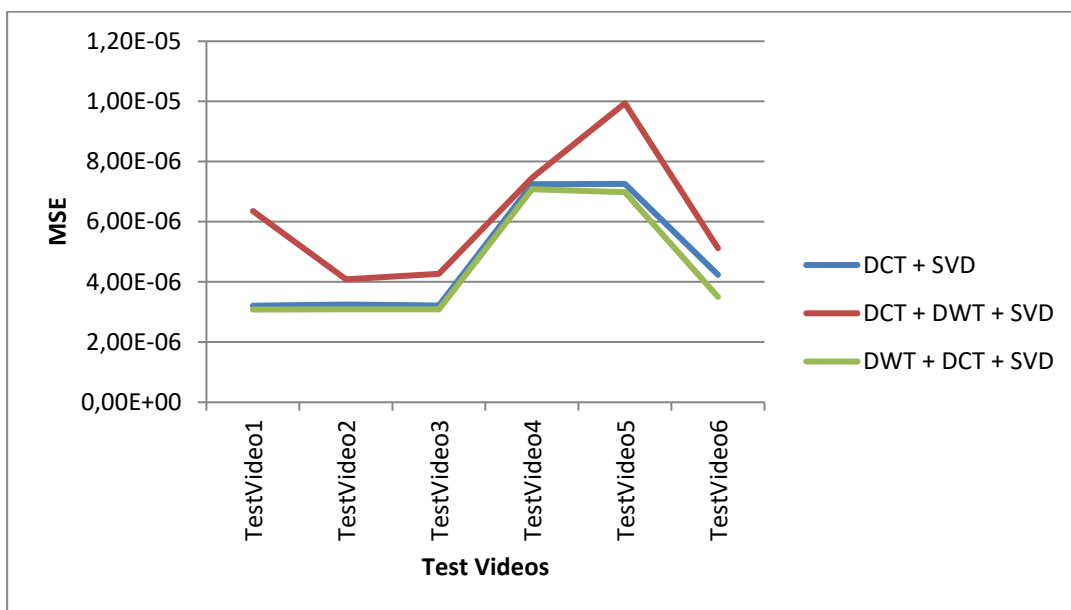


Figure 8: Comparison between various combinations for robustness

The current state of the art as discussed in the literature review section mainly focuses on the few numbers of bits or blocks within an image or a few numbers of images which make them more prone to various attacks such as geometric attacks, active attacks, compression attack, JPEG attack and many more. While the proposed approach is designed for watermarking the entire video with a special focus on keyframes thereby safeguarding the system from various attacks.

7. Conclusion and future work

The proposed DWT+DCT+SVD digital video watermarking technique outperforms the state-of-the-art methods while being more robust and safe. The DWT takes care of the finer variations of the images at various scales and it also helps in the localization of information. The DCT plays an important role

in capturing a finite sequence of data points and the SVD is pivotal for preserving important geometrical insights. In the pursuit of the added advantage of DWT, DCT, and SVD combination the computational time of the proposed approach can be considered as a small trade-off against a significant improvement in watermarking performance. As a part of future work, the same approach may be deployed and analyzed for analog systems such as audio signals. Also, the experimental analysis to discretize the scale parameter will be an interesting study to pursue.

References

- [1] M. Asikuzzaman and M. R. Pickering, "An Overview of Digital Video Watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2131–2153, 2018.
- [2] G. Kaur, S. S. Kasana, and M. K. Sharma, "An efficient watermarking scheme for enhanced high-efficiency video coding/h.265," *Multimed. Tools Appl.*, vol. 78, no. 9, pp. 12537–12559, 2019.
- [3] M. R. Keyvanpour, N. Khanbani, and M. Boreiry, "A secure method in digital video watermarking with transform domain algorithms," *Multimed. Tools Appl.*, vol. 80, no. 13, pp. 20449–20476, 2021.
- [4] G. Doërr and J. L. Dugelay, "A guide tour of video watermarking," *Signal Process. Image Commun.*, vol. 18, no. 4, pp. 263–282, 2003.
- [5] D. Rajani and P. R. Kumar, "An optimized blind watermarking scheme based on principal component analysis in redundant discrete wavelet domain," *Signal Processing*, vol. 172, p. 107556, 2020.
- [6] P. Garg and R. Rama Kishore, "Secured and multi optimized image watermarking using SVD and entropy and prearranged embedding locations in transform domain," *J. Discret. Math. Sci. Cryptogr.*, vol. 23, no. 1, pp. 73–82, 2020.
- [7] A. K. Abdulrahman and S. Ozturk, "A novel hybrid DCT and DWT based robust watermarking algorithm for color images," *Multimed. Tools Appl.*, vol. 78, no. 12, pp. 17027–17049, 2019.
- [8] S. P. Ambadekar, J. Jain, and J. Khanapuri, *Digital image watermarking through encryption and DWT for copyright protection*, vol. 727. Springer Singapore, 2019.
- [9] D. G. Savakar and A. Ghuli, "Robust Invisible Digital Image Watermarking Using Hybrid Scheme," *Arab. J. Sci. Eng.*, vol. 44, no. 4, pp. 3995–4008, 2019.
- [10] W. Zheng *et al.*, "Robust and high capacity watermarking for image based on DWT-SVD and CNN," *Proc. 13th IEEE Conf. Ind. Electron. Appl. ICIEA 2018*, pp. 1233–1237, 2018.
- [11] F. Arab, M. Zamani, S. Poger, C. Manigault, and S. Yu, "A framework to evaluate the performance of video watermarking techniques," *2019 IEEE 2nd Int. Conf. Inf. Comput. Technol. ICICT 2019*, pp. 114–117, 2019.
- [12] S. Das, M. Banerjee, and A. Chaudhuri, "An improved video key-frame extraction algorithm leads to video watermarking," *Int. J. Inf. Technol.*, vol. 10, no. 1, pp. 21–34, 2018.
- [13] A. M. U. Wagdarikar, R. K. Senapati, and S. Ekkeli, *A secure video watermarking approach using CRT theorem in DCT domain*, vol. 521. Springer Singapore, 2019.
- [14] M. I. Khan, M. Rahman, and I. H. Sarker, "Digital Watermarking for Image Authentication Based on Combined DCT, DWT and SVD Transformation."
- [15] P. Garg and R. R. Kishore, "Performance comparison of various watermarking techniques," *Multimed. Tools Appl.*, 2020.
- [16] H. Fang, H. Zhou, Z. Ma, W. Zhang, and N. Yu, "A robust image watermarking scheme in DCT domain based on adaptive texture direction quantization," *Multimed. Tools Appl.*, vol. 78, no. 7, pp. 8075–8089, 2019.
- [17] G. Singh and N. Goel, "Entropy based image watermarking using discrete wavelet transform and singular value decomposition," *Proc. 10th INDIACom; 2016 3rd Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2016*, vol. 59, no. 11, pp. 2700–2704, 2016.

EMPIRICAL STUDY ON ROBUST REGRESSION ESTIMATORS AND THEIR PERFORMANCE

LAKSHMI R¹, DR.SAJESH T A¹



Department of Statistics, St. Thomas College (Autonomous), Thrissur 680001, University of Calicut¹
lakshmi.nss19@gmail.com, sajesh.t.abraham@gmail.com

Abstract

Regression Analysis is statistical technique to model data. But the presence of outliers and influential points affect data modelling and its interpretation. Robust regression analysis is an alternative choice to this. Here we made an attempt to study different robust estimators and propose a new robust reweighted S_n covariance based regression estimator. We have evaluated the performance empirically and the simulation study shows our proposed estimator is preferable to OLS and other robust regression estimators in terms of the MSE criteria. Also, proposed robust S_n covariance regression estimator produce outperforming results for regression equivariance and breakdown criterion. Robustness of the proposed estimator is proved empirically. The proposed method is innovatively used to model fluid data. R software is used for simulation and study.

Keywords: robust S_n regression, influential observations, modelling, data analysis,

1. INTRODUCTION

One of the most essential statistical methods in data modelling is regression analysis. It helps in the prediction of a link between the predictors and the response variable. All academic disciplines, including social science, health science, engineering, physical science, and others, frequently use it. Regression Analysis mainly rely on ordinary least squares method, which is very vulnerable in the midst of the outliers. Informally outlier can be defined as those observations which lie out of the place with respect to other observations in the data set. Thus, when there are polluted points in the data set, robust regression was created as an improved and effective alternative to least squares. There are numerous robust regression techniques; among them some are resistant too. In this paper, we discuss about some of the mainstream and efficient robust regression techniques for contaminated data in multiple linear regression models. Apart from that, multiple regression can achieve efficiently through expressing classical normal equations in covariance matrix form. In this paper, apart from discussing robust regression estimators, we propose a robust reweighted regression based on S_n covariance matrix. The main inherent idea is to compare the techniques using simulated data set, and determine the properties of the proposed estimator through vast empirical simulations alone. Simulation is done and evaluated by using Monte Carlo technique. Section 2 briefly describes about the OLS method, necessity of robust techniques and important robust estimators developed over years and propose a new reweighted regression estimator based on robust covariance matrix technique. Section 3 of this paper presents different simulation methods for comparing proposed estimator and other robust regression estimators, along with that, properties of the proposed estimator studied through wide simulation. Section 4 provides the real life data application and conclusion of the paper.

2. ORDINARY LEAST SQUARES

Linear regression model is about estimating the parameter $\beta \in R^p$

$$y_i = x_i\beta + \epsilon_i \quad i = 1, 2, 3, \dots, n. \quad (1)$$

where $(x_i, y_i) \in (R^p, R)$ comprise the data and β is the p - dimensional unknown vector and ϵ_i are unknown errors. The best-known estimator of β is the least square estimators obtained by:

$$\min \sum_{i=1}^n (y_i - x_i\beta)^2$$

The least square estimators are very popular because of Gauss Markov theorem and very easy to use. These classical estimators are the best when their assumptions are met by the data. Whenever there are outliers in the data, OLS results in unstable estimate prediction and are renowned for misbehaving. The data may contain outliers for a number of reasons, including incorrect data entry, incorrect scoring, and unusual sample data. In regression, outliers can be classified according to their location and effect. Observations would be unusual with respect to y values or x values. They are categorised as outliers, leverages and influential points based on how they affect the model. The impact of these observations depends on the location where they occur. Extreme values in the predicted variables are called as leverages. Leverages measure how far an independent variable deviates from its mean. The direction of the distance between the remaining data points is not taken into account by leverages. Leverages do not affect the estimates of the regression coefficients. It affects the model summary statistics, standard errors of regression coefficient etc. Influential points are those points with unusual x coordinate and the unusual y value. The regression coefficients are noticeably affected by influential points. Influential points pull the regression model in its direction. Outliers in either the x or y directions constitute a significant hazard to least square estimators. Statistical or graphical methods can be used to identify outliers. Mahalanobis distance is a statistical procedure used to locate the outliers in the x direction. We cannot say Mahalanobis distance as a perfect method, as it fails to detect the outliers in y direction. Other statistical outlier diagnostics works on the idea of erasing one observation at a time and recalculates the regression coefficients; they are called as regression diagnostics, in which diagnostic quantities are obtained using the data with aim of identifying influential points. Following the identification, they are either eliminated or corrected, and then the least squares analysis is performed. As a result, such statistics estimate the change in regression coefficients that would occur if a single observation were removed following analysis. These statistics are also known as deletion statistics, useful for pinpointing influential points. Cook distance, Studentised residuals, DFFITS, DFBETAS and Jackknife residuals are some of such deletion statistics. Calculation of these diagnostic statistics become complicated when there are multiple unusual observations. Robust regression estimation is alternative strategy for handling outliers. Robust methods aim to create estimators that are immune to outliers. Diagnostic tools remove outliers before fitting the data using the least square approach, whereas, Robust regression, on the other hand, fits a regression model to the great majority of the data before identifying outliers as regions with substantial residuals. The breakdown point, concept of bounded influence and relative efficiency are ideas that are pertinent to the study of robust regression. The presence of single outlier can completely invalidate the OLS estimator. Contrast to it; we will see estimators that can handle certain percentage of outliers. This particular concept is called as breakdown point and [3] provided the first explanation of a breakdown point. It'd only evaluate location in a single dimension. Also, [5] provided broad description of braekdown, but it was highly mathematical in nature and asymptotic. It was [4], suggested a limited sample version of breakdown point.

For a sample Z of n observations,

$$Z = (x_{11}, \dots, x_{1p}, y_1), \dots, (x_{n1}, \dots, x_{np}, y_n)$$

Let T represents a regression estimator. When T is applied to such a sample, the result is a regression coefficient vector as $T(Z) = \hat{\beta}$. Let j of the sample data points swapped by arbitrary values and call them as corrupted sample Z' . The maximal bias generated by such contamination is then calculated as:

$$bias(j; T, Z) = \sup_{Z'} \|T(Z') - T(Z)\|,$$

Where the supremum is over all possible Z' . If the bias is infinite, j outliers have a significant impact on the estimator. Thus, breakdown of the estimator T at the sample Z is defined as

$$\epsilon_n^*(T, Z) = \min\left(\frac{j}{n}; bias(j; T, Z) \text{ is infinite}\right).$$

Or the least amount of contamination that an estimator can tolerate is known as the breakdown point. The breakdown point of ordinary least estimator is $\epsilon_n^*(T, Z) = \frac{1}{n}$. That is, even the presence of single outlier in the data set can affect least square estimators.

2.1. Proposed Method

OLS estimator can express as solution to 2 proposed by [8] in the following way. Let $z = (x, y)$ be the joint variable of independent and dependent variables. Let μ be the location and Σ be the scatter matrix of z . Partitioning μ and Σ yields the notation

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix} \quad (2)$$

Generally the estimates $\hat{\mu}$ and $\hat{\Sigma}$ are estimated in empirical way. The least square estimates of β and α can be written as function of $\hat{\mu}$ and $\hat{\Sigma}$, namely,

$$\hat{\beta} = \hat{\Sigma}_{xx}^{-1} \hat{\Sigma}_{xy}, \quad \hat{\alpha} = \hat{\mu}_y - \hat{\beta}^T \hat{\mu}_x. \quad (3)$$

Major drawback of the above mentioned estimators is, classical estimators of location and scatter are sensitive to the presence of the outliers. Robustification of the classical estimators of scatter and location improve the performance of the estimators and [9] in their paper proposed a robust method for detecting multiple outliers and thus robust covariance matrix estimation in multidimensional data set denoted as S_n method. Another objective of our paper is to propose a robust reweighted regression estimator based on S_n covariance estimator in equation (4). In this paper the performance of the proposed robust reweighted S_n regression estimator of the joint variable z is evaluated and studied through empirical simulations. The performance and properties of the estimator is investigated through wide range of simulations and Mean Squared error is used to compare the performance of proposed estimator with other estimators in different scenarios.

Let $X = X_1, X_2, X_3, \dots, X_p$ be a $n \times p$ matrix of size n and p being the number of variables. The robust covariance matrix based on S_n method of the matrix X is defined as:

$$S_n(X_i, X_j) = med_i med_{j \neq i} [(x_i - x_j)(y_i - y_j)] \quad , i, j = 1, 2, 3, \dots, p, \quad (4)$$

where med is an abbreviation for low median ($[\frac{n+1}{2}]^{th}$ order statistic). Inner median will be taken up by $[n/2]^{th}$ order statistic for odd value of n . The corresponding correlation matrix of equation 4 is defined as:

$$\delta_{S_n}(X) = DCOV_{S_n}(X)D^t \quad (5)$$

where D is the diagonal matrix with diagonals $1/S_n(x_i), i = 1, 2, 3, \dots, p$. Here $S_n(x_i)$ is nothing but robust scale estimator of univariate random variable X and is defined as,

$$S_n(X) = 1.1926 med_i med_j |x_i - x_j|$$

The covariance matrix mentioned in equation 4 is non-positive semi definite and [7] in their paper describe procedure to solve non positive semi definite and obtain positive semi definite and approximately affine equivariant estimators. The following steps provide us positive semi definite dispersion matrix and robust estimates:

- Let e_j be the eigen vector corresponding to the eigen value λ_j of correlation matrix δ_{S_n} . Let E be $p \times p$ matrix with columns e_j for $j = 1, 2, \dots, p$.
- Let $R = D^{-1}E$ and $z_i = R^{-1}X_i$ and Z be an orthogonalised matrix with rows z_i^T ($i = 1, 2, 3 \dots n$) and columns Z_j ($j = 1, 2, \dots, p$).

The resulting robust S_n estimate of location and scatter is defined as:

$$\hat{\mu}_{S_n} = Rv \quad \text{and} \quad \hat{\Sigma}_{S_n} = R\Gamma R^T \quad (6)$$

where $v = (\text{med}(Z_1), \text{med}(Z_2), \dots, \text{med}(Z_p))^T$ and $\Gamma = \text{diag}(S_n(Z_1)^2, \dots, S_n(Z_p)^2)$. Here med stands for median and S_n is the univariate robust scale estimate. The process can be iterate to enhance the estimates by replacing covariance estimator used in equation 5 with above $\hat{\Sigma}_{S_n}$. Let us call the robust estimator 6 of location and scatter as initial S_n estimator of z . The associated robust squared Mahalanobis distance of each observation z_i is defined as The resulting robust S_n estimate of location and scatter is defined as:

$$RD(z_i) = (z_i - \hat{\mu}_{S_n})^t \hat{\Sigma}_{S_n}^{-1} (z_i - \hat{\mu}_{S_n})$$

Robust Mahalanobis distance is an efficient outlier detection method. Let w_i be a weighted function based on the above Mahalanobis distance defined as $w_i = w(RD(z_i))$. The reweighted estimators take over the robustness properties of initial estimators with increasing their efficiency ([6]). Therefore the reweighted S_n location and scatter matrix be obtained as:

$$\hat{\mu}_{wS_n} = \frac{\sum_{i=1}^n w_i z_i}{\sum_{i=1}^n w_i} \quad \text{and} \quad \hat{\Sigma}_{wS_n} = \frac{\sum_{i=1}^n w_i (z_i - \mu_{wS_n})(z_i - \mu_{wS_n})^T}{\sum_{i=1}^n w_i} \quad (7)$$

The weights above are computed as $w_i = w(RD(z_i)) = I(RD(z_i) \leq c)$, which assign weight 1 to the z_i for $i = 1, 2, \dots, n$, where

$$c = \begin{cases} \chi_{0.95,p}^2 & \text{if } p < 15 \\ \frac{\chi_{0.95,p}^2 \text{med}(rd_1, \dots, rd_n)}{\chi_{0.5,p}^2} & \text{if } p \geq 15 \end{cases} \quad (8)$$

Based on $\hat{\mu}_{wS_n}$ and $\hat{\Sigma}_{wS_n}$ we obtain $\hat{\beta}_{wS_n}$ and $\hat{\alpha}_{wS_n}$ the robust reweighted S_n regression estimator defined as:

$$\hat{\beta}_{S_n} = (\hat{\Sigma}_{wS_n})_{xx}^{-1} (\hat{\Sigma}_{wS_n})_{xy} \quad \text{and} \quad \hat{\alpha}_{S_n} = (\hat{\mu}_{wS_n})_y - (\hat{\beta}_{S_n})^T (\hat{\mu}_{wS_n})_x \quad (9)$$

The efficiency, breakdown and affine equivariant property of the proposed estimator 9 is evaluated.

3. SIMULATION STUDY

Simulation study is done to evaluate the performance of the proposed robust reweighted S_n regression estimator. And the results are compared with ordinary least squares and some of the other robust regression estimators like: LTS, LMS, S, and MM estimator. The simulations are done in R and all the values are reported in tables at the end of the paper. Consider the linear regression model form:

$$y = \alpha + X\beta + \epsilon$$

where X is $n \times p$ matrix, $\beta = (\beta_1, \dots, \beta_p)^T$ is the unknown regression coefficient vector of size $p \times 1$, α is the unknown intercept of the model and ϵ is the i.i.d error term and are independent from X . The X variables are distributed as $N(\mathbf{0}_p, I_p)$, where I_p is the p - dimensional identity matrix. Following sets of dimensions and sample sizes are considered in this study respectively: $p=5, 10$, and 20 with $n=50, 100, 500$. The simulations are repeated 1000 times and each time parameter estimates are noted.

Mainly three simulation scenarios, as that found in the literature [2] are considered here.

- The dependent variable is generated from standard normal distribution with corresponding regression coefficients including intercept equals zero, and standard normal errors are considered [NES].
- The dependent variable is simulated from t distribution with 3 degrees of freedom with corresponding regression coefficients including intercept equals zero and heavy tailed errors (t distribution with 3 degrees of freedom) [HTS].
- Regression with normal error, some percentage(δ) of randomly selected observation in independent variable replaced as $N(\lambda\sqrt{\chi^2_{(0.99,p)}}, 1)$ and the dependent variables were replaced as $N(k\sqrt{\chi^2_{(0.99,1)}}, 1)$ where $\lambda, k=0.5, 1, 1.5, 2, 3, 5, 7, 8, 10$. The percentage of contamination considered in this scenario is 10% and 20%.

3.1. Efficiency

It is a well-known fact that ordinary least squares have maximum efficiency under normal errors. Thus under normal error case, the efficiency of each robust method is calculated relative to OLS. Let $\Phi = (\beta^T, \alpha)^T$ be the joint vector of regression parameters, intercept and slope. Dimension of Φ is $(p + 1) \times 1$. The finite efficiency for the joint estimator $\hat{\Phi}_{Re}$ of a robust method (Re) is defined as:

$$Eff = \frac{1/1000 \sum_{i=1}^{1000} \|\hat{\Phi}_{OLS}^i - \Phi\|_2^2}{1/1000 \sum_{i=1}^{1000} \|\hat{\Phi}_{Re}^i - \Phi\|_2^2}$$

Table1 exhibit the simulated efficiency of $\hat{\Phi}$, of proposed robust reweighted S_n estimator and other robust regression estimators, with respect to the classical least square estimator, under normal error scenario described above. In the table, bold letters in each row represents the highest efficiency and italic letters represent the lowest efficiency. Estimators with higher efficiencies are represented in bold and lowest efficiencies are represented in italics. Among the estimators under consideration, proposed reweighted S_n estimator exhibits highest efficiency throughout all the randomly chosen dimensions and sample size considered. MM estimator also possesses efficiency greater than 90% in some of the cases under consideration. Among the estimators LMS perform poorly. The proposed estimator has highest efficiency for all randomly chosen sample sizes and dimensions considered.

In the second scenario, we are considering heavy tailed error distribution. Thus least square estimators cannot be maximum efficient estimator. Hence we consider the Mean Squared Error of the estimators instead of Eff . The table 2 results shows that proposed estimator out perform in all scenarios with mean squared error lower than other estimators. Also, as the sample size increases the mean squared error decreases for all the robust methods.

3.2. Robustness

To study the robustness, simulations accordingly in third scenario [CS] defined above are carried out. Here we have randomly considered different dimensions ($p=5$ and 10) with sample size (50 and 500). The criteria used here to compare the different estimators is mean squared error of the estimated joint parameter vector $\hat{\Phi}$, averaged over 1000 simulation runs, similar criteria considered in [2]. Tables 3 to 9 below shows the maximum (across λ and k) MSE for both estimated intercept and slope for different combination of dimension and sample size. We are considering the maximum value of mean squared error obtained over all considered k values, for each value of lambda.

$$i.e. MSE_{\lambda}(\cdot) = \max_{k \in [0.5, 1, 1.5, 2, 3, 5, 7, 8, 10]} MSE_{\lambda, k}(\cdot)$$

Lowest value of MSE in each row is notated in bold letter in the tables. Tables 3 to 9 gives $MSE_{\lambda}(\cdot)$ of different robust estimators and proposed estimator. Among the values, proposed estimator shows minimum MSE in all cases considered. For higher dimensions, proposed estimator possess very low MSE than other estimators, even when percentage of contamination increases, proposed estimator shows low MSE and consistently maintain low error throughout different level of

contamination. Here we can see for same dimension, when number of observations increases, mean squared error of reweighted S_n estimator decreases. Thus the performance of proposed estimator increases with increase in the number of observations. Proposed estimator out perform in all the scenarios constantly. MM and S estimators mainly collapse for $\lambda, k = 0.5, 1, 1.5, 2, 3$ of contamination. And for other values of (λ, k) , MM and S estimators perform moderately with MSE values mostly greater than proposed estimator. LTS and LMS estimators performs consistently for all values of λ, k , but the mean squared error obtained is higher than proposed estimator in all scenarios. Among all the scenarios, OLS possess highest mean squared error than other estimators.

3.3. Breakdown Property

The breakdown point evaluates the maximum percentage of outliers an estimator can tolerate. 50% is the highest breakdown value an estimator can attain. Even though high contamination level occurs rarest of rare in general, here we propose to study the performance of the estimators in extreme contamination and evaluate the consistency in their performance. It has shown that repeated median regression estimator has 50% asymptotic breakdown point through simple mathematical induction by [11]. The same lemma quoted in [11] is applicable for reweighted S_n estimator, since the estimator is nothing but nested median of observations. For this, a criterion [CS] is used with percentage of contamination 30%, 35%, 40%. Dimensions considered here are 5 and 30. And we consider the maximum MSE across all combinations of λ, k . i.e. $MMSE(\cdot) = \max_{\lambda \in \{0.5, 1, 1.5, 2, 3, 5, 7, 8, 10\}} MSE_{\lambda}(\cdot)$. The results are shown in table 10.

In general, robust estimators like LTS, S, and MM have high breakdown point, but their computations are challenging. In all these mentioned methods regression estimators are obtained by resampling algorithm. Resampling algorithms are used to obtain number of subsamples, and then robust regression estimators are obtained by making use of an initial high breakdown estimator. Thus all these established methods depend purely on number of subsamples and initial estimates. Reweighted S_n estimator proposed here is not dependant on resampling and initial estimates. Also, proposed estimator possess high empirical breakdown even to large contamination and higher dimension.

Based on all simulation results, we can observe that all robust estimators, except our suggested estimator, have a constant increase in mean square error value, which reaches a maximum as the fraction of outliers in the vertical direction reaches a maximum for a certain lambda value. Although both MM and S estimators are stated to have a high breakdown, S estimator performs poorly as the dimension of the variable grows. MM estimator could be consider to be as reasonably good robust estimator, shows low MSE among other established robust estimators. However, the MSE of MM estimator is much higher than that of proposed reweighted S_n estimator. As the percentage of outliers in the data increases, LTS perform poorly. Even though LTS has a 50% breakdown point, the performance of LTS estimator depends merely on the correct choice of tuning constant h, here we used default value $h=0.5$. Throughout various levels of contamination, our suggested reweighted S_n estimator consistently maintains a low MSE. Also, even in higher dimensions, the proposed estimator has a lower MSE than other well-known robust estimators, indicating that proposed estimator is more resistant to large numbers of outliers, which can be termed as high empirical breakdown point.

3.4. Equivariance Property

Rather than theoretical goodness, practical usefulness of an estimator is determined by equivariance, breakdown and robustness properties. These three qualities are considerable properties of a regression estimator and discussed in our paper. Breakdown property is described in the above section. For regression estimators three types of equivariance are considered:

1. Regression equivariance is defined as: if we transform the dependent variable by adding a linear function of independent variables, is equal to adding the coefficients of this linear

function to the estimators.

2. y- equivariance is defined as, if the dependent variable is transformed linearly, then the estimators get transformed in the same manner.

Let $\hat{\Phi}(X, Y) = (\hat{\beta}^T, \hat{\alpha})^T$, where X is $n \times p$ matrix and Y is $n \times 1$ matrix. Then (1) and (2) can be combined to form: $\hat{\Phi}(X, Yb + Xg + u) = \hat{\Phi}(X, Y)b + (g^T, u)^T$, where $b \in R$, a non-zero constant, g is $p \times 1$ vector and $u \in R$ is any constant. Keeping X as same and transforming the dependent variable as $Yb + Xg + u$, then the resulting estimator would be: $\hat{\beta}_{new} = b(\hat{\beta}) + g$ and $\hat{\alpha}_{new} = b\hat{\alpha} + u$.

3. x- equivariance is defined as, if the independent variables are linearly transformed, then the equivalent transformed estimator is: $\hat{\Phi}(XA, Y) = (\hat{\beta}^T(A^{-1})^T, \hat{\alpha})^T$.

That is, if the independent variables are transformed as XA , with a non-singular $p \times p$ matrix A , the resulting new estimators are $\hat{\beta}_{new} = A^{-1}\hat{\beta}$ and $\hat{\alpha}_{new} = \hat{\alpha}$. It is not possible to explore available transformations, so, [7] and [10] proposed in their papers, to generate A matrices randomly for the purpose of checking x-equivariance as $A = TD$, where T a random orthogonal matrix and D is a $p \times p$ diagonal matrix with diagonal entries are independently and uniformly distributed. The methods outlined above are employed in our paper to investigate the suggested estimator's equivariance property. When the above mentioned transformations are performed on simulated data sets, the MSE of the suggested estimator is examined. Here we consider two dimensions, $p=5$ and $p=30$. Also, we consider contaminated scenario [CS] with contamination of 10% and 20%. First the estimator is applied to untransformed data and the estimator obtained $\hat{\Phi}_{S_n}$ is recorded. The above mentioned transformed data is then used to estimate x-equivariance, y-equivariance, regression equivariance, and the resulting new estimator $\hat{\Phi}_{S_n, new}$ is stored. The MSE is calculated between $\hat{\Phi}_{S_n, new}$ and estimator value which has to be obtained if the above properties hold. The estimator consistently performs and maintains low MSE. Even when the contamination increases with increase in dimension, the estimator exhibits low MSE. Low MSE indicates that the model can be predicted more accurately. The suggested estimator is essentially affine equivariant since the model's error is managed and kept low.

Table 11 shows MMSE results for x- equivariance. From the table, we can see that the error remains controlled and low for varied proportion of vertical outliers and leverage points. The error value increases as the dimension increases, but in a regulated manner. The suggested estimator is nearly x-equivariant since the errors are controlled.

Table 12 shows the MMSE for y-equivariance and regression equivariance. The mean square error remains very low throughout for different proportion of vertical outliers and leverage points. Also, we can see mean square error shows decreasing pattern as dimension increases. Though mean square error increases with increase in the percentage of contamination level, the increments are very small and close to zero. As a result, we can state that the mean square error is well-controlled and kept to a minimum in all scenarios evaluated. Thus the proposed estimator is approximately y-equivariant and regression equivariant. We have empirically demonstrated three equivariance features using simulated samples with contamination at various degrees and dimensions.

4. REAL LIFE DATA APPLICATION

4.1. Fluid Dynamics

A substance capable of flowing is termed as a fluid. Fluids are of two types, namely liquids and gases. The study of fluid's behaviour at rest (termed fluid statics) and in motion (termed fluid dynamics) is jointly known as fluid mechanics. Many real-world applications, including cancer treatment, car radiators, air conditioning, refrigeration, microwave ovens, blow moulding, and petrochemical processing, heavily rely on heat and mass transmission. Through extensive studies scientists have been successful in improving these transmission qualities. Choi and Eastman were the first to notice that the introduction of nano sized particles in a conventional fluid was able to bring a significant improvement in these transmission qualities. Internal heat sources play an

important role in various heat transfer applications.

For its vital role in photo thermal and photodynamic therapy, the dynamics of water-based TiO_2 nano liquid over an elongated nonlinear surface was elucidated by [1]. The flow problem was modeled using partial differential equations which were solved using finite-difference based bvp5c technique with the help of apposite similarity transformations. Further, the authors employed response surface methodology and sensitivity analysis to elucidate the heat transfer rate for the consequence of magnetic field ($0.5 \leq M \leq 1.5$), thermal radiation ($0.5 \leq R_d \leq 1.5$) and exponential heat source ($0.2 \leq Q_E \leq 0.4$). The optimal heat transfer rate was observed when $M=0.5, R_d =1.5$, and $Q_E =0.2$.

Recently, researchers have examined the influence of effectual parameters on the engineering quantities using statistical methods like regression analysis, and response surface methodology. By establishing a quantitative relationship between the independent (relevant characteristics) and dependent (physical process of interest) variables, the inclusion of these statistical techniques tends to broaden perception. Typographical errors while handling such data is a possibility, owing to those outliers could occur in such datasets, which should be tackled scientifically.

In this paper, the fluid data for conducting multiple linear regression analysis has been derived from Areekara et al. The derived data has been analyzed using proposed regression estimator along with other robust regression estimators. The data consist of 20 observations with three independent variables and one independent variable. Initially we analyze the performance in original data without outlier and then we conduct the same procedure by replacing 10th and 19th observation into outliers. The results are reported in tables 13 and 14 below. Proposed estimator performs in normal scenario and outlier injected data as that of other existing robust estimator like MM, S estimators. We have also reported the OLS estimated values, from which it is clear that classical method fails in the performance of outliers. Also, LTS, LMS regression methods fail to perform in these data sets due to their computational complexity. Thus we have shown in our paper that, for conducting linear regression analysis in such fluid data sets, it is always better to use robust regression methods; even there are no outliers in the data set. Also, proposed estimator works well even without outliers in the datasets. In this paper, the fluid data for conducting multiple linear regression analysis has been derived from Areekara et al. The derived data has been analyzed using proposed regression estimator along with other robust regression estimators. The data consist of 20 observations with three independent variables and one independent variable. Initially we analyze the performance in original data without outlier and then we conduct the same procedure by replacing 10th and 19th observation into outliers. The results are reported in tables 13 and 14 below. Proposed estimator performs in normal scenario and outlier injected data as that of other existing robust estimator like MM, S estimators. We have also reported the OLS estimated values, from which it is clear that classical method fails in the performance of outliers. Also, LTS, LMS regression methods fail to perform in these data sets due to their computational complexity. Thus we have shown in paper that, for conducting linear regression analysis in such fluid data sets, it is always better to use robust regression methods; even there are no outliers in the data set. Also, proposed estimator works well even without outliers in the datasets.

4.2. Belgian Phone Call Data

Belgian phone calls data was published by Belgium Statistical Survey and [[6]] used the data in their work. The data consists of annual count of international calls from Belgium during the period 1950 to 1973. The data comprise of two variables, the year (X) and the number of call received (Y). The data contains six outliers in Y direction. From the table 9 below, it is clear that, OLS provides highly misleading estimates in the presence of anomalies. Also, M and LMS performances are not remarkable. MM, S and LTS don't exhibits remarkable performance in outlier detection and in providing estimates. Among them, S_n estimator detects the outlying observations as (15th, 16th, 17th, 18th, 19th, 20th, and 24th observation) and MM estimator detects outlying observations as (15th, 16th, 17th, 18th, 19th, 20th, and 21st observation). Thus, proposed

estimator performs well and detects outliers correctly as that of MM estimator and it can be considered as an add-on to the covariance based regression method.

5. CONCLUSION

Classical regression analysis is very sensitive to the presence of contaminated observations. Several robust alternative methods are available in the literature. In this paper, we propose an improved robust reweighted S_n regression estimator. Here we are proposing a new robust regression technique using alternative form of OLS. We propose a new robust reweighted robust S_n regression estimator. The properties and performance of our proposed estimator are inferred through wide range of empirical simulation methods. Also, the performance of proposed estimator in fluid dynamics data and Belgian phone call data is evaluated. Proposed estimator exhibit a consistent performance in all the cases considered. The robustness property, affine equivariance and breakdown property of the proposed estimator is compared with OLS, MMS, LTS, LMS, S estimators using simulation study. And in all scenarios considered, proposed estimator outperforms other existing robust estimators. The results are tabulated below. Although many robust regression estimators have already been proposed in the literature, we could add proposed estimator to the list of available regression estimators, since proposed estimator exhibit excellent performance than other estimators. A thorough comparison has done and we can conclude that proposed estimator possess high breakdown, robustness equivariance property. Also, the proposed estimator is suitable for multiple regression estimation and is a good alternative to the classical estimator. Developing theoretical properties of the proposed estimator is the future aim of our work.

Table 1: Table showing efficiency in case of Normal error scenario

p	n	S_n	MM	S	LTS	LMS
5	30	0.9684	0.9309	0.2813	0.6708	0.0873
	50	0.9286	0.9228	0.2814	0.4639	0.1797
	100	0.9235	0.8951	0.2312	0.5166	0.1299
	500	0.9857	0.9309	0.2728	0.7414	0.0509
10	50	0.9466	0.9832	0.2520	0.6917	0.0292
	100	0.9429	0.8461	0.2888	0.4819	0.1211
	500	0.9852	0.8706	0.1992	0.6314	0.0315
	1000	0.9283	0.9321	0.2123	0.7119	0.0163
30	100	0.8751	0.7409	0.2617	0.7157	0.0033
	500	0.9628	0.8706	0.1992	0.6314	0.0315
	1000	0.9335	0.9321	0.2123	0.7119	0.0163
	5000	0.9817	0.7409	0.2617	0.7157	0.0033

Table 2: Table showing the MSE in case of t tailed error distribution

p	n	S_n	MM	S	LTS	LMS
5	30	0.4530	0.5326	1.1059	0.9081	2.0202
	50	0.2429	0.2501	0.5465	0.3708	0.8954
	100	0.1083	0.1192	0.2059	0.1391	0.4232
	500	0.0196	0.0203	0.0349	0.0224	0.1449
10	50	0.5387	0.5307	2.3917	0.8783	2.4023
	80	0.2996	0.3984	0.6684	0.3959	1.3699
	100	0.2111	0.3667	0.4762	0.2979	0.9966
	500	0.0396	0.1467	0.0629	0.0421	0.5392
30	100	0.3448	0.8933	2.2357	1.2430	6.2807
	150	0.4834	0.5339	0.8678	0.6666	4.2174
	500	0.0254	0.0971	0.2479	0.1269	3.2732

Table 3: Table 3(1) showing $MSE_{\lambda}(\cdot)$ for $n=500, p=5, \delta=10\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	0.1595	2.7088	1.7421	0.4396	0.8785	91.915
1	0.1158	2.2949	1.8572	1.7479	1.0189	37.916
1.5	0.1166	1.7354	1.8309	1.7042	1.0428	18.686
2	0.1158	2.6151	2.6478	1.8624	1.0528	10.881
3	0.1114	2.7446	2.5763	1.8506	1.0556	4.9069
5	0.1139	1.7846	1.8011	2.6581	1.0500	1.7826
7	0.1122	0.9074	0.9328	2.4172	1.0560	0.9135
8	0.1142	0.6917	0.7117	2.6076	1.0563	0.6882
10	0.1162	0.4473	0.4600	1.8004	1.0519	0.4438

Table 4: Table 3(2) showing $MSE_{\lambda}(\cdot)$ for $n=500, p=5, \delta=20\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	1.6434	7.3536	7.4798	7.5241	1.0062	134.02
1	0.1488	10.669	9.9103	9.7256	1.0749	42.848
1.5	0.1216	10.811	9.7874	9.7707	1.0882	19.762
2	0.1217	7.2620	10.175	10.198	1.0854	11.196
3	0.1209	4.9965	5.0651	5.0491	1.0857	4.9839
5	0.1169	1.7835	1.8156	1.8243	1.0707	1.7938
7	0.1195	0.9116	0.9263	0.9398	1.0579	0.9109
8	0.1207	0.4474	0.7124	0.7150	1.0662	0.6961
10	0.1153	0.4455	0.3032	0.3072	1.0812	0.4463

Table 5: Table 3(3) showing $MSE_{\lambda}(\cdot)$ for $n=50, p=5, \delta=10\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	0.6232	3.0231	2.8642	3.2983	1.1107	99.802
1	0.4038	2.6021	2.8603	3.0759	1.1709	38.799
1.5	0.3998	2.6271	2.7735	3.4003	1.1968	18.827
2	0.3867	2.6629	2.8132	3.6539	1.1950	11.005
3	0.3904	2.8386	2.7222	3.3677	1.2375	4.9652
5	0.3834	1.6583	1.7868	1.6814	1.2058	1.8273
7	0.3945	0.9234	0.6863	0.9836	1.2035	0.9278
8	0.3804	0.7253	0.8329	0.7483	1.1877	0.7016
10	0.3995	0.4704	0.5515	0.5209	1.1998	0.4827

Table 6: Table 3(4) showing $MSE_\lambda(\cdot)$ for $n=50, p=5, \delta=20\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	1.1457	11.154	12.784	18.499	99.802	139.16
1	0.4178	12.066	12.407	18.179	38.799	42.948
1.5	0.4215	12.643	12.710	20.576	18.827	19.915
2	0.4124	10.342	10.441	11.408	11.005	11.268
3	0.4169	5.0657	5.2375	5.2001	4.9652	5.0114
5	0.4113	1.1892	1.8959	1.8915	1.8273	1.8074
7	0.4014	0.9862	1.0382	0.9977	1.5036	0.9311
8	0.4133	0.7476	0.8552	0.8458	1.1879	0.7362
10	0.4158	0.3324	0.6009	0.5703	1.1999	0.4736

Table 7: Table 3(5) showing $MSE_\lambda(\cdot)$ for $n=50, p=10, \delta=10\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	0.4056	6.8755	7.5017	8.4836	1.1244	1.0269
1	0.4073	7.8391	7.6248	9.9929	1.1958	1.1821
1.5	0.4216	7.8541	7.1342	9.2994	1.2212	12.714
2	0.3726	6.2529	6.1745	6.6639	1.2174	7.1902
3	0.3356	3.6121	3.2540	3.2603	1.2399	3.2836
5	0.3475	1.2000	1.2886	1.2639	1.2546	1.1877
7	0.3509	0.6345	0.4943	0.6841	1.2516	0.6207
8	0.3512	0.4756	0.5748	0.5516	1.2623	0.4852
10	0.3568	0.3197	0.2922	0.3760	1.2248	0.4229

Table 8: Table 3(6) showing $MSE_\lambda(\cdot)$ for $n=50, p=10, \delta=20\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	0.7139	42.630	37.898	7.5892	1.1537	108.34
1	0.4418	28.875	28.031	7.1634	1.1952	28.931
1.5	0.4239	13.086	13.160	7.2613	1.2349	12.923
2	0.4079	7.4023	7.5233	6.2072	1.2334	7.2643
3	0.4008	3.3492	2.2409	3.1659	1.2526	1.5981
5	0.4797	1.2284	1.3222	1.2963	1.2698	1.1756
7	0.4125	0.6363	0.7376	0.7444	1.2424	0.6244
8	0.4533	0.5037	0.4316	0.5726	1.2968	0.4769
10	0.4229	0.2499	0.4273	0.3948	1.2854	0.3298

Table 9: Table 3(7) showing $MSE_\lambda(\cdot)$ for $n=500, p=10, \delta=10\%$

λ	S_n	MM	S	LTS	LMS	OLS
0.5	0.1655	4.0859	4.5578	0.2901	1.0055	90.827
1	0.1626	5.8239	4.5036	1.1616	1.0425	27.440
1.5	0.1598	5.9110	5.3589	2.6196	1.0638	12.582
2	0.1603	5.8077	4.6791	4.5252	1.0524	7.1467
3	0.1618	3.1899	3.2266	2.6401	1.0529	3.1912
5	0.1622	1.1576	1.1758	4.6335	1.0474	1.1535
7	0.1612	0.5905	0.4584	5.3122	1.0526	0.5885
8	0.1558	0.4490	0.3022	4.5747	1.0559	0.4499
10	0.1589	0.2865	0.3001	4.6989	1.0422	0.2879

Table 10: Table.4 showing the MMSE(.)for checking breakdown property

Method	5			30		
	$\delta=30\%$	$\delta=35\%$	$\delta=40\%$	$\delta=30\%$	$\delta=35\%$	$\delta=40\%$
OLS	12.438	12.721	12.934	8.6068	10.133	11.681
S_n	0.5108	1.5456	1.7754	0.9073	1.0807	2.1518
MM	6.3341	9.2566	6.1267	7.4952	9.2222	12.584
S	6.8982	8.6637	13.662	36.451	34.157	30.675
LTS	47.399	48.661	206.54	886.55	1055.6	1033.1

Table 11: Table 5 showing the $MMSE_\lambda(\hat{\Phi}_{S_n, new})$ for checking x-equivariance

λ	p=5		p=30	
	$\delta=10\%$	$\delta=20\%$	$\delta=10\%$	$\delta=20\%$
0.5	0.01953	0.03739	0.10566	0.13301
1	0.05469	0.03865	0.29562	0.32564
1.5	0.03992	0.03566	0.11913	0.17027
2	0.03909	0.01530	0.13167	0.25173
3	0.03842	0.01900	0.16799	0.28351
5	0.03488	0.03279	0.16283	0.18638
7	0.03771	0.03911	0.19997	0.25546
8	0.01948	0.03119	0.10239	0.10734
10	0.02028	0.02552	0.12871	0.27826

Table 12: Table.6 showing the $MMSE_\lambda(\hat{\Phi}_{S_n, new})$ for checking y-equivariance and regression equivariance

λ	p=5		p=30	
	$\delta=10\%$	$\delta=20\%$	$\delta=10\%$	$\delta=20\%$
0.5	0.00319	0.00216	0.00067	0.00276
1	0.00145	0.00173	0.00013	0.01238
1.5	0.00102	0.00814	0.00039	0.01591
2	0.00056	0.00139	0.00039	0.00005
3	0.00109	0.00135	0.00050	0.00073
5	0.00178	0.00106	0.00141	0.00077
7	0.00021	0.00020	0.00024	0.00038
8	0.00012	0.00062	0.00064	0.00082
10	0.00011	0.00000	0.00438	0.00034

Table 13: Table 7 showing output of fluid data without outlier

	S_n	OLS	MM	S	LMS
β_0	1.1056	1.9304	1.8018	1.7995	2.0497
β_1	-0.1617	-0.1614	-0.1036	-0.0899	-0.2098
β_2	0.5713	0.5711	0.6274	0.6297	0.5077
β_3	-2.1968	-2.1954	-2.2016	-2.2424	-2.2676

Table 14: Table 8 showing output of fluid data with outliers

	S_n	OLS	MM	S	LMS
β_0	1.1109	0.6438	1.95204	2.0321	1.7494
β_1	-0.1547	0.0015	-0.1748	-0.2001	-0.1914
β_2	0.5688	0.9332	0.5556	0.5163	0.6315
β_3	-2.1839	0.2489	-2.2495	-2.2708	-1.7389

Table 15: Table 9 showing results on Belgian Phone Call data

Method	Intercept	Coefficient of X	MSE
OLS	58.566	0.587	78.1372
S_n	2.986	6.062	5001.829
MM	47.931	8.831	10041.74
S	48.060	8.894	10228.44
LTS	47.769	9.094	10717.14
LMS	48.439	8.658	9674.762
M	57.412	0.626	92.58984

REFERENCES

- [1] Areekara, S., Mackolil, J., Mahanthesh, B., Mathew, A., and Rana, P. (2022). A study on nanoliquid flow with irregular heat source and realistic boundary conditions: A modified Buongiorno model for biomedical applications. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 102, e202100167.
- [2] Cabana Garceran del Vall, E., Lillo Rodriguez, R. E., and Laniado Rodas, H. (2019). Shrinkage reweighted regression.
- [3] Hodges Jr, J. L. (1967, January). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 163–186).
- [4] Donoho, D. L., and Huber, P. J. (1983). The Notion of BreakdownPoint. *A festschrift for Erich L. Lehmann* 157184.
- [5] Hampel, F. R. (1971). A general qualitative definition of robustness. *The annals of mathematical statistics*, 42(6), 1887–1896.
- [6] Rousseeuw, P. J., and Leroy, A. M. (2005). *Robust regression and outlier detection*. John wiley & sons.
- [7] Maronna, R. A., and Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44(4), 307–317.
- [8] Maronna, R. and Morgenthaler, S. (1986). Robust regression through robust covariances. *Communications in Statistics-Theory and Methods*, 15(4), 1347–1365.
- [9] Kunjunni, S. O., and Abraham, S. T. (2022). Multidimensional outlier detection and robust estimation using S_n covariance. *Communications in Statistics-Simulation and Computation*, 51(7), 3912–3922.
- [10] Sajesh, T. A., and Srinivasan, M. R. (2012). Outlier detection for high dimensional data using the Comedian approach. *Journal of Statistical Computation and Simulation*, 82(5), 745–757.
- [11] Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika*, 69(1), 242–244.
- [12] Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, 642–656.

RATIO ESTIMATOR OF POPULATION MEAN USING A NEW LINEAR COMBINATION UNDER RANKED SET SAMPLING

Saba Riyaz¹, Khalid Ul Islam Rather^{2*}, Showkat Maqbool³ and T. R. Jan¹

•

*Division of Statistics and Computer Science, SKUAST-Jammu, India.²
Department of Statistics, University of Kashmir, Srinagar – 190006, J&K, India^{1,4}
Division of AGB, FVSC & AH, SKUAST-K, Shuhama, J&K, India³*

Saba.riyaz.syed@gmail.com

khalidstat34@gmail.com

showkatmaq@gmail.com

drtrjan@gmail.com

Abstract

Ranked set sampling is an approach to data collection originally combines simple random sampling with the field investigator's professional knowledge and judgment to pick places to collect samples. Alternatively, field screening measurements can replace professional judgment when appropriate and analysis that continues to stimulate substantial methodological research. The use of ranked set sampling increases the chance that the collected samples will yield representative measurements. This results in better estimates of the mean as well as improved performance of many statistical procedures. Moreover, ranked set sampling can be more cost-efficient than simple random sampling because fewer samples need to be collected and measured. The use of professional judgment in the process of selecting sampling locations is a powerful incentive to use ranked set sampling. This paper is devoted to the study, we introduce an approach to the mean estimators in ranked set sampling. The amount of information carried by the auxiliary variable is measured with the on populations and samples and to use this information in the estimator, the basic ratio and the generalized exponential ratio estimators are as an improved form of a difference cum exponential ratio type estimator under the ranked set sampling in order to estimate the population mean \bar{Y} of study variate Y using single auxiliary variable X . The expressions for the mean squared error of propose estimator under ranked set sampling is derived and theoretical comparisons are made with competing estimators. We show that the proposed estimator has a lower mean square error than the existing estimators. In addition, these theoretical results are supported with the aid of some real data sets using R studio. Therefore, Under RSS architecture, a better difference cum exponential ratio type estimator has been suggested. The estimator's mathematical form has been developed, and its efficiency requirements have been developed in relation to various already-existing estimators from the literature. By imputing various values for the constants used in the creation of our proposed estimator, we also provide several specific situations of our estimator.

Keywords: mean squared error, auxiliary variable, median, coefficient of variation, kurtosis

1. Introduction

It is well known that the information of the auxiliary variable is commonly used in order to increase efficiency and precision in sample surveys. It has also a role in the related methods of estimation, such as ratio, product, and regression. If the correlation between the study variable (Y) and the auxiliary variable (X) is highly positive, the ratio method of estimation is used. If not, the product method of estimation is employed effectively provided that this correlation is highly

negative. In recent years, there have been many articles on estimators for the population mean in the Sampling Theory Literature, such as unbiased estimators in general form for estimating the finite population mean in stratified random sampling [1], a generalized ratio estimator is proposed by using some robust measures with single auxiliary variable [2 and 3], an efficient families of ratio-type estimators to estimate finite population mean using known correlation coefficient between study variable and auxiliary variable by [5 and 6], Estimation of rare and clustered population mean using stratified adaptive cluster sampling and using auxiliary character in stratified random sampling [7 and 8]. The estimation of population mean using auxiliary attribute under ranked set sampling (RSS) [9, 10 and 11]. The problem of exponential estimator for estimating the population mean considered under RSS using attribute, two phase sampling by [12, 13, 14, and 15].

In addition to the Simple Random Sampling (SRS) method, RSS, which may be considered as a controlled random sampling design, was first introduced to estimate the pasture yield by [16]. The RSS procedure involves randomly drawing n sets of n units each from the population for which the mean is to be estimated. It is assumed that the units in each set can be ranked visually. From the first set of n units, the lowest unit ranked is measured. From the second set of n units, the second lowest unit ranked is measured. This process continues until the n^{th} ranked unit is measured. The gain in efficiency by a computation involving five distributions illustrated by [16]. As a simple introduction to the concept of RSS, when X is a random variable with a density function $F(x)$ and (x_1, x_2, \dots, x_n) are the unobserved values from n units, we may then rank them by visual inspection or based on a concomitant variable. RSS involves selecting one unit among every ranked set consisting of m units for quantification.

The RSS method can be briefly described step by step as follows:

Step 1: Randomly select m^2 units from the target population.

Step 2: Allocate the m^2 selected units as randomly as possible into m sets, each of size m .

Step 3: Without knowing any values of the variable of interest, rank the units within each set with respect to variable of interest. This may be based on personal professional judgment or done with concomitant variable correlated with the variable of interest.

Step 4: Choose a sample for actual quantification by including the smallest ranked unit in the first set, the second smallest ranked unit in the second set and this process continues in this way until the largest ranked unit is selected from the last set.

Step 5: Repeat Steps 1 through 4 for n cycles to obtain a sample of size mn for actual quantification. [17]

When it is ranked on the auxiliary variable, let $y_{(i)}, x_{(i)}$ denote an i^{th} judgment ordering in the i^{th} set for the study variable and the i^{th} order statistic in the i^{th} set for the auxiliary variable, respectively.

In the remaining part of this article, the estimators for the population mean under RSS are mentioned in Section 2, the adapted estimator from the SRS to RSS is given in Section 3, theoretical and numerical comparisons of the adapted estimator are performed with the existing adapted estimators in literature in Sections 4 and 5, respectively.

2. Estimators in literature

The estimator of the population ratio using the RSS as defined by [19].

$$t_{RSS} = \frac{\bar{y}_{[n]}}{\bar{x}_{[n]}} \tag{2.1}$$

Where $\bar{y}_{[n]} = \frac{1}{n} \sum_{i=1}^n y_{(i)}$ and $\bar{x}_{[n]} = \frac{1}{n} \sum_{i=1}^n x_{(i)}$. Note that the estimator in (2.1) can also be used for the population total and mean. Then, the estimator for the population mean can be written as follows:

$$\bar{y}_{rRSS} = \frac{\bar{y}_{[n]}}{\bar{x}_{[n]}} \bar{X} \tag{2.2}$$

Where it is assumed that the population mean \bar{X} of the auxiliary variable x is known and the MSE equation of the estimator in (2.2) can be given by

$$MSE(\bar{y}_{rRSS}) \cong \frac{1}{mr} (S_y^2 - 2RS_{yx} + R^2S_x^2) - \frac{1}{m^2r} \left(\sum_{i=1}^m \tau_{y(i)}^2 - 2R \sum_{i=1}^m \tau_{yx(i)} + R^2 \sum_{i=1}^m \tau_{x(i)}^2 \right) \quad (2.3)$$

Where, $R = \frac{\bar{Y}}{\bar{X}}$, S_x^2 is the population variance of the auxiliary variable, S_y^2 is the population variance of the study variable, S_{yx} is the population covariance between the auxiliary and study variables, $\tau_{x(i)} = (\mu_{x(i)} - \bar{X})$, $\tau_{y(i)} = (\mu_{y(i)} - \bar{Y})$, and $\tau_{yx(i)} = (\mu_{y(i)} - \bar{Y})(\mu_{x(i)} - \bar{X})$. Here, \bar{Y} is the population mean of the study variable. Note that the values of $\mu_{x(i)}$ and $\mu_{y(i)}$ depend on the order statistics from some specific distributions and these values can be found in [19]. We would like to remind that the values of $\mu_{x(i)}$ and $\mu_{y(i)}$ can be taken to be same in the absence of judgment error if the variables have the same distribution (see the appendix of [20])

The following estimator by adapting [21] to the RSS proposed by [22]:

$$\bar{y}_{kRSS} = \frac{k\bar{y}_{[n]}}{\bar{x}_{[n]}} \bar{X} \quad (2.4)$$

Where k is a constant.

The MSE of the estimator in (2.4) is given by

$$MSE(\bar{y}_{rRSS}) \cong \frac{1}{mr} (k^2 S_y^2 - 2Rk S_{yx} + R^2 S_x^2) + \bar{Y}^2 (k^* - 1)^2 - \frac{1}{m^2r} \left(\sum_{i=1}^m k^2 \tau_{y(i)}^2 - 2Rk \sum_{i=1}^m \tau_{yx(i)} + R^2 \sum_{i=1}^m \tau_{x(i)}^2 \right) \quad (2.5)$$

where $k^* = \frac{1+\gamma\rho C_x C_y - W_{yx(i)}}{1+\gamma C_y^2 - W_{y(i)}^2}$, Here, $W_{yx(i)} = \frac{1}{m^2 r \bar{X} \bar{Y}} \sum_{i=1}^m \tau_{yx(i)}$ and $W_{y(i)}^2 = \frac{1}{m^2 r \bar{Y}^2} \sum_{i=1}^m \tau_{y(i)}^2$, $\gamma = \frac{1}{mr}$, C_x and C_y are the population coefficients of variation of the auxiliary and study variables, respectively, ρ is the population correlation between the auxiliary and the study variables.

3. Proposed estimator

An improved difference cum-exponential ratio type is defined for estimating \bar{Y} as following [18 and 21]

$$\bar{Y}_{RK} = \{t_1 \bar{y}_{[n]} + t_2 (\bar{X} - \bar{x}_{[n]})\} \left\{ \exp \left[\frac{\bar{X} - \bar{x}_{[n]}}{\bar{X} + \bar{x}_{[n]}} \right] \right\} \quad (3.1)$$

To obtain the MSE of \bar{Y}_{RK} , write

$$\bar{y}_{(n)} = \bar{Y}(1 + \epsilon_0), \text{ and } \bar{x}_{(n)} = \bar{X}(1 + \epsilon_1),$$

Such that $E(\epsilon_0) = E(\epsilon_1) = 0$,

$$\text{and } E(\epsilon_0)^2 = V\left(\frac{\bar{y}_{(n)}}{\bar{Y}}\right) = \frac{1}{mr} \frac{1}{\bar{Y}^2} \left[S_y^2 - \frac{1}{m} \sum t_{y(i)}^2 \right] = [\theta C_y^2 - w_{y(i)}^2],$$

$$E(\epsilon_1)^2 = V\left(\frac{\bar{x}_{(n)}}{\bar{X}}\right) = \frac{1}{mr} \frac{1}{\bar{X}^2} \left[S_x^2 - \frac{1}{m} \sum t_{x(i)}^2 \right] = [\theta C_x^2 - w_{x(i)}^2],$$

$$E(\epsilon_0 \epsilon_1) = \frac{1}{mr} \frac{1}{\bar{Y} \bar{X}} \left[S_{yx} - \frac{1}{m} \sum t_{y(i)}^2 \right] = [\theta C_{yx} - w_{yx(i)}],$$

Where $W_{x(i)}^2 = \frac{1}{m^2 r \bar{X}^2} \sum_{i=1}^m \tau_{x(i)}^2$

Expressing (1.1) in terms of ϵ 's,

$$\begin{aligned} \bar{Y}_{RK} &= \{t_1 \bar{Y}(1 + \epsilon_0) + t_2 (\bar{X} - \bar{X}(1 + \epsilon_1))\} \left\{ \exp \left[\frac{\bar{X} - \bar{X}(1 + \epsilon_1)}{\bar{X} + \bar{X}(1 + \epsilon_1)} \right] \right\} \\ \bar{Y}_{RK} &= \{t_1 \bar{Y} + t_1 \bar{Y} \epsilon_0 - t_2 \bar{X} \epsilon_1\} \left\{ \exp \left(-\frac{\epsilon_1}{2} \right) \left[1 + \frac{\epsilon_1}{2} \right]^{-1} \right\} \end{aligned} \quad (3.2)$$

Expanding the right hand side of (1.2) and retaining terms up to the second power of ϵ 's,

$$\begin{aligned} \bar{Y}_{RK} &= \{t_1 \bar{Y} + t_1 \bar{Y} \epsilon_0 - t_2 \bar{X} \epsilon_1\} \left\{ \exp\left(-\frac{\epsilon_1}{2}\right) \left[1 + \frac{\epsilon_1}{2} + \frac{\epsilon_1^2}{4}\right] \right\} \\ \bar{Y}_{RK} &= \{t_1 \bar{Y} + t_1 \bar{Y} \epsilon_0 - t_2 \bar{X} \epsilon_1\} \left\{ 1 - \frac{\epsilon_1}{2} + \frac{\epsilon_1^2}{4} + \frac{\epsilon_1^2}{8} \right\} \end{aligned} \quad (3.3)$$

From (3.3),

$$\bar{Y}_{RK} - \bar{Y} = \bar{Y} \left\{ (t_1 - 1) + t_1 \epsilon_0 - \frac{t_1 \epsilon_1}{2} - t_2 R \epsilon_1 - \frac{t_1 \epsilon_0 \epsilon_1}{2} + \frac{t_2 R \epsilon_1^2}{2} + \frac{3t_1 \epsilon_1^2}{8} \right\} \quad (3.4)$$

Squaring (3.4) and then taking expectation of both sides, the MSE of the estimator \bar{Y}_{RK} is

$$MSE(\bar{Y}_{RK}) = \bar{Y}^2 \{t_1^2 \varphi_1 - t_1 \varphi_2 + t_2^2 R^2 \varphi_3 - t_1 t_2 R \varphi_4\} \quad (3.5)$$

Where,

$$\varphi_1 = \gamma \{C_y^2 + C_x^2 - 2C_{yx}\} - \{w_{y[i]}^2 + w_{x[i]}^2 - 2w_{yx[i]}\}$$

$$\varphi_2 = \gamma \{C_x^2\} - \{w_{x[i]}^2\}$$

$$\varphi_3 = \gamma \left\{ \frac{3}{4} C_x^2 - C_{yx} \right\} + \left\{ \frac{3}{4} w_{x[i]}^2 - w_{yx[i]} \right\}$$

$$\varphi_4 = \gamma \{C_x^2 - C_{yx}\} + \{w_{x[i]}^2 - w_{yx[i]}\}$$

Obtain the optimum t_1 and t_2 to minimize $MSE(\bar{Y}_{RK})$. Differentiate $MSE(\bar{Y}_{RK})$ with respect to t_1 and t_2 and equating the derivatives to zero, optimum values of t_1 and t_2 is given by

$$t_{1opt} = \frac{2\varphi_2\varphi_3}{4\varphi_1\varphi_3 - \varphi_4^2}$$

$$t_{2opt} = \frac{\varphi_2\varphi_4}{4\varphi_1\varphi_3 - \varphi_4^2}$$

Substituting the value of t_{1opt} and t_{2opt} in (3.5), we get the minimum value of $MSE(\bar{Y}_{RK})$ as

$$MSE_{min}(\bar{Y}_{RK}) = \bar{Y}^2 \{t_1^2 \varphi_1 - t_1 \varphi_2 - t_2^2 R^2 \varphi_3\} \quad (3.6)$$

4. Efficiency

In this section, the performances of the proposed estimator have been demonstrated over the traditional ratio estimator in the RSS and the estimator of [23] respectively, as follows:

$$\begin{aligned} MSE(\bar{y}_{rRSS}) - MSE_{min}(\bar{Y}_{RK}) &> 0 \\ \{(1 - t_1^2)\varphi_1 + t_1\varphi_2 + t_2^2 R^2 \varphi_3\} &> 0 \end{aligned} \quad (4.1)$$

$$\begin{aligned} MSE(\bar{y}_{kRSS}) - MSE_{min}(\bar{Y}_{RK}) &> 0 \\ \{(k^* - 1)^2 + (1 - t_1^2)\varphi_1 + t_1\varphi_2 + t_2^2 R^2 \varphi_3\} &> 0 \end{aligned} \quad (4.2)$$

Table 1: Some members of exponential ratio type estimator in ranked set sampling

Estimator	t_1	t_2
$\bar{Y}_{RK1} = \{\bar{y}_{[n]} + (\bar{X} - \bar{x}_{[n]})\} \left\{ \exp\left[\frac{\bar{X} - \bar{x}_{[n]}}{\bar{X} + \bar{x}_{[n]}}\right] \right\}$	1	1
$\bar{Y}_{RK2} = \{(\bar{X} - \bar{x}_{[n]})\} \left\{ \exp\left[\frac{\bar{X} - \bar{x}_{[n]}}{\bar{X} + \bar{x}_{[n]}}\right] \right\}$	0	1
$\bar{Y}_{RK3} = \{\bar{y}_{[n]}\} \left\{ \exp\left[\frac{\bar{X} - \bar{x}_{[n]}}{\bar{X} + \bar{x}_{[n]}}\right] \right\}$	1	0

5. Numerical example

To observe performances of the estimators, we use some real-life populations. The descriptions of these populations are given below:

Population I {source: [24]}

Y: Acceleration of automobiles

X: Engine horsepower of automobiles

Objective: To estimate population mean of Acceleration of automobiles.

The summary statistics are given below:

$$N = 392, n = 30, m = 10, r = 3, \mu_x = 104.4694, \mu_y = 15.5413, S_y = 2.7589, S_x = 38.4912, \\ C_x = 0.3684, C_y = 0.1775, C_{xy} = -0.0451, \beta_{2(x)} = 0.6541, \beta_{1(x)} = 1.079, \rho_{xy} = 0.9091$$

Population II {source: [25]}

Y: Body Mass Index (BMI) of Crohn's disease patients

X: Weight of Crohn's disease patients

Objective: To estimate population mean of Body Mass Index (BMI) of Crohn's disease patients.

The summary statistics are given below:

$$N = 117, n = 20, m = 5, r = 4, \mu_x = 69.0256, \mu_y = 26.0624, S_y = 4.9888, S_x = 14.2438, \\ C_x = 0.2063, C_y = 0.1914, C_{xy} = 0.0325, \beta_{2(x)} = 0.7746, \beta_{1(x)} = 0.6571, \rho_{xy} = 0.8222$$

Population III {source: [26]}

Y: Body Mass Index (BMI)

X: Thigh Circumference

Objective: To estimate population mean of Body Mass Index (BMI).

The summary statistics are given below:

$$N = 36, n = 8, m = 4, r = 2, \mu_x = 49.3806, \mu_y = 25.678, S_y = 3.8198, S_x = 3.7599, \\ C_x = 0.0761, C_y = 0.1488, C_{xy} = 0.0066, \beta_{2(x)} = -0.6159, \beta_{1(x)} = -0.0607, \rho_{xy} = 0.9848$$

Percent Relative Efficiencies (PREs) of our proposed estimators along with competitor estimators from literature have been presented in Table 2, 3 and 4 for different real-life populations.

Table 2: PRE of Estimators for Population I

	\bar{y}_{rRSS}	$\bar{y}_{\kappa RSS}$	\bar{Y}_{RK1}	\bar{Y}_{RK2}	\bar{Y}_{RK3}	\bar{Y}_{RK}
\bar{y}_{rRSS}	100					
$\bar{y}_{\kappa RSS}$	212.19	100				
\bar{Y}_{RK1}	245.19	231.72	100			
\bar{Y}_{RK2}	241.45	210.47	98.81	100		
\bar{Y}_{RK3}	238.97	189.37	90.84	93.08	100	
\bar{Y}_{RK}	361.74	275.18	249.18	245.15	213.49	100

Table 2, revealed the percent relative efficiencies (PRE) of estimators for population I. It is observed that the proposed difference cum exponential ratio type estimator in ranked set sampling \bar{Y}_{RK} proved to be the best estimator in the sense of having highest percent relative efficiency than usual unbiased estimators \bar{y}_{rRSS} , $\bar{y}_{\kappa RSS}$ for the population I. The generalized form of proposed difference cum exponential ratio type estimator \bar{Y}_{RK} is 361.74% more efficient than the existing estimator \bar{y}_{rRSS} and 275.18% more efficient than $\bar{y}_{\kappa RSS}$. Moreover, the special cases of our proposed generalized estimator \bar{Y}_{RK1} , \bar{Y}_{RK2} and \bar{Y}_{RK3} are also proved to be more efficient than existing estimators. These results suggest using proposed difference cum exponential ratio type estimator to estimate population mean of Acceleration of automobiles.

Table 3: PRE of Estimators for Population II

	\bar{y}_{rRSS}	$\bar{y}_{\kappa RSS}$	\bar{Y}_{RK1}	\bar{Y}_{RK2}	\bar{Y}_{RK3}	\bar{Y}_{RK}
\bar{y}_{rRSS}	100					
$\bar{y}_{\kappa RSS}$	204.74	100				
\bar{Y}_{RK1}	238.48	238.29	100			

\bar{Y}_{RK2}	237.37	204.28	93.92	100		
\bar{Y}_{RK3}	221.49	174.28	89.32	82.74	100	
\bar{Y}_{RK}	352.86	252.48	248.82	229.23	190.48	100

Table 3, showed the percent relative efficiencies (PRE) of estimators for population II. It is observed that the proposed difference cum exponential ratio type estimator in ranked set sampling \bar{Y}_{RK} also proved to be the best estimator in the sense of having highest percent relative efficiency than usual unbiased estimators \bar{Y}_{rRSS} , \bar{Y}_{kRSS} for the population II. The generalized form of proposed difference cum exponential ratio type estimator \bar{Y}_{RK} is 352.86% more efficient than the existing estimator \bar{Y}_{rRSS} and 252.48% more efficient than \bar{Y}_{kRSS} . Moreover, the special cases of our proposed generalized estimator \bar{Y}_{RK1} , \bar{Y}_{RK2} and \bar{Y}_{RK3} are also proved to be more efficient than existing estimators. These results suggest using proposed difference cum exponential ratio type estimator to estimate population mean of Body Mass Index (BMI) of Crohn's disease patients.

Table 4: PRE of Estimators for Population III

	\bar{Y}_{rRSS}	\bar{Y}_{kRSS}	\bar{Y}_{RK1}	\bar{Y}_{RK2}	\bar{Y}_{RK3}	\bar{Y}_{RK}
\bar{Y}_{rRSS}	100					
\bar{Y}_{kRSS}	238.48	100				
\bar{Y}_{RK1}	275.28	264.82	100			
\bar{Y}_{RK2}	263.82	249.27	98.47	100		
\bar{Y}_{RK3}	239.83	237.42	97.38	98.37	100	
\bar{Y}_{RK}	384.27	283.38	259.37	278.38	239.57	100

Table 4, showed the percent relative efficiencies (PRE) of estimators for population III. It is observed that the proposed difference cum exponential ratio type estimator in ranked set sampling \bar{Y}_{RK} also proved to be the best estimator in the sense of having highest percent relative efficiency than usual unbiased estimators \bar{Y}_{rRSS} , \bar{Y}_{kRSS} for the population III. The generalized form of proposed difference cum exponential ratio type estimator \bar{Y}_{RK} is 384.27% more efficient than the existing estimator \bar{Y}_{rRSS} and 283.38% more efficient than \bar{Y}_{kRSS} . Moreover, the special cases of our proposed generalized estimator \bar{Y}_{RK1} , \bar{Y}_{RK2} and \bar{Y}_{RK3} are also proved to be more efficient than existing estimators. These results suggest using proposed difference cum exponential ratio type estimator to estimate population mean of Body Mass Index (BMI).

6. Conclusion

In this paper, by using coefficient of variation and median of the auxiliary variable, the modified ratio estimators of population mean have been proposed under ranked set sampling (RSS). Large sample approximations to the mean square errors of these estimators have been derived and compared with the MSE of the usual ratio estimator and other existing estimators of same class. Through numerical illustration is concluded that the proposed estimator performs much better than the classic ratio estimator as well as the estimators given by Mehta and Mandowara (2016) based on RSS. Thus, the proposed new estimator can be used instead, in order to increase the efficiency of parameter estimates.

References

- [1] Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. The Journal of Agricultural Science, 30(2): 262-275.
- [2] Javaid, S., and Maqbool, S. (2016). Modified variance estimation using mid-range of an auxiliary variable. International Journal of Agricultural and Statistical Sciences, 12(2): 347-350.
- [3] Jeelani, M. I., Bouza, C. N., and Sharma, M. (2018). Modified ratio estimator under rank set sampling. Investigación Operacional, 38(1): 103-106.

- [4] Kadilar, C., Unyazici, Y., and Cingi, H. (2009). Ratio estimator for the population mean using ranked set sampling. *Statistical Papers*, 50(2): 301-309.
- [5] Maqbool, S., Raja, T. A., and Javaid, S. (2016). Generalized modified ratio estimator using non-conventional location parameter. *International Journal of Agricultural and Statistical Sciences*, Vol, 12(1): 2016.
- [6] McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*. 3(4): 385-390.
- [7] Mehta, N., and Mandowara, V. L. (2016). A modified ratio-cum-product estimator of finite population mean using ranked set sampling. *Communications in Statistics-Theory and Methods*, 45(2): 267-276.
- [8] Murthy, M. N. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A*, 69-74.
- [9] Raja, T. A., and Maqbool, S. (2021). On modified ratio estimator using a new linear combination. *International Journal of Agricultural and Statistical Sciences*. Vol, 17(1): 209-211.
- [10] Rather, K.U.I. and Kadilar, C. (2021). Exponential Type Estimator for the population Mean under Ranked Set Sampling. *Journal of Statistics: Advances in Theory and Applications*, 25(1): 1-12.
- [11] Rather, K.U.I., Eda, K. G. and Unal, C. (2022). New exponential ratio estimator in Ranked set sampling. *Pakistan Journal of Statistics and operation research*, 18(2), 403-409.
- [12] Robson, D. S. (1957). Applications of multivariate polykeys to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 52(280): 511-522.
- [13] Samawi, H. M., and Muttlak, H. A. (1996). Estimation of ratio using rank set sampling. *Biometrical Journal*, 38(6): 753-764.
- [14] Singh, H. P., and Kakran, M. S. (1993). A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. unpublished paper.
- [15] Sisodia, B. V. S., and Dwivedi, V. K. (1981). Modified ratio estimator using coefficient of variation of auxiliary variable. *Journal-Indian Society of Agricultural Statistics*.
- [16] Tailor, R., and Sharma, B. (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis. *Population*, 10: 1.
- [17] Takahasi, K., and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20(1): 1-31.
- [18] Upadhyaya, L. N., and Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 41(5): 627-636.
- [19] Wang, T., Y. Li and H. CUI. (2007). On weighted randomly trimmed means. *Journal of Systems Science and Complexity*, 20(1): 47-65.
- [20] Wolter, K. M. (1985). *Introduction to Variance Estimation*, Springer –Verlag.

Combined Redundancy Optimization for a System Comprising Operative, Cold Standby and Warm Standby Units

LALJI MUNDA

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
laljimaths0505@gmail.com

GULSHAN TANEJA*

•

Department of Mathematics, Maharshi Dayanand University, Rohtak, Haryana, India
drgtaneja@gmail.com

*Corresponding Author

Abstract

The company or industry can increase system reliability and provide stress-free operation by adding redundant equivalent subsystems to the active unit. The warm standby system is accessible if the operational unit malfunctions, and the cold standby system can take over. This paper aims to analyze a system comprised of one operative unit, cold standby unit, and one warm standby unit. Cold standby is activated to become warm standby when the operative unit fails, and warm standby becomes operational immediately. A minor defect causes the warm standby unit to fail, whereas a major fault causes the operational unit to fail. Such systems are used by many businesses, sectors, and facilities to prevent operational and reputational losses. Cut-off values for the failure rate, activation rate, revenue cost, and cost per repairman visit have been calculated to determine when the system is profitable. Various system performance measures have been defined by using the Markov process and regeneration point method.

Keywords: Cold Standby, Profit Analysis, Redundancy, Warm Standby

1. INTRODUCTION

The major focus of modern technology is to improve system reliability so that it can meet consumer demand. This results in the development of more expensive, sophisticated systems with significant repair costs in the event of failure. Redundancy is a desirable choice as a result. Standby redundancy is one of several strategies for implementing redundancy. It is a redundancy that has a secondary unit in case the active unit fails. Countless researchers have studied the coupling of redundant systems in the optimization of system reliability. Goel and Gupta [1] assumed three types of failures for a hot standby system. Tuteja et al.[2] investigated a redundant two-unit cold standby system with regular and expert maintenance personnel. When a system malfunctions or an ordinary repairman is unable to fix the problem, an expert repairman is contacted. Kumar et al.[3] examined a two-unit cold standby system that was repaired by an expert repairman when it failed. When the expert is working on the first unit failure and the second unit fails at the same time, the assistant repairman can fix the problem while following the expert's instructions. El-Said and EL-Sherbery [4] discussed regarding two units on cold standby with post-repair inspection. Parashar and Taneja[5] analyzed a two-unit PLC hot standby system,

one as the master unit and the other as the slave unit, with minor repairs being performed by regular repairmen and major repairs being handled by specialist repairmen. Mahmoud and Moshref [6] considered standby system that experienced both hardware and human error failure. For a two unit cold standby system, Manocha and Taneja[7] modified the outcome provided by EL-Said and EL-Sherbery. By taking into account the non-regenerative state and all possible general distributions, a stochastic model has been developed by them. Malhotra and Taneja [8] identified two different cable manufacturing industry reliability models. Model 1 uses a single unit while Model 2 has a cold standby system with two units. Manocha et al. [9] created a stochastic model for a hot standby database system with two units. Database administrators handle issues with the primary or backup database systems. Adlakha et al.[10] took into consideration a two-unit cold standby communication system that was first packed and assembled as necessary. One is in use, and the other is on cold standby. Batra and Taneja[11]-[14] optimized the standby units with one or two operational units in the system. A comparative analysis is also performed to determine which of these models is preferable for system profitability. Levitin et al. [15] considered a system with the potential for shocks during data transfer and operation. Every shock shortens an element’s lifespan and causes deterioration; as a result, the operating unit is immediately switched out for a cold standby unit. Malhotra et al. [16] examined the expected rest period and expected maintenance period for a two-unit cold standby system in a pharmaceutical company. Jia et al. [17] developed a stochastic model for a demand-based warm standby system with capacity storage.

It is noticeable that the combination of hot, warm, and cold standby units has not yet been investigated, despite the fact that many industrial systems have these unit backups to manage emergencies. In light of this, the present study aims to create a stochastic model with a single operational unit, warm and cold standby units. The rest of the article is organized as follows. In Section 2, system assumptions are listed. Sections 3 and 4 include notations to be applied in the study and state descriptions. Section 5 covers the transition probabilities and mean sojourn time. Sections 6, 7 and 8 describe various system performance metrics, profit functions, and their corresponding numerical assessments. In section 9, the paper is concluded with some insightful real-world applications.

2. ASSUMPTIONS

Initially, we have one operating, one warm standby, and one cold standby unit. The warm standby unit becomes active immediately in the event of the operative unit malfunctioning, and the cold standby unit is activated to become the warm standby unit. Before any event is over, the activation process must be finished. The operative unit fails due to a serious problem, but warm standby fails due to a small problem. The system is still functioning if at least one unit is up and running. Each and every parameter is exponential and statistically independent.

3. NOTATIONS

The various notations for rates/probabilities are as follows:

λ/λ_1	Failure rate of operative/warm standby unit
β	Activation rate of cold standby unit
α_1/α_2	Repair rate on major/minor fault
$B_{j_0}(t)$	Probability that a repairman is working on a major repair to the system at time t without transitioning to another regenerative state at time t=0
$B_i(t)$	Probability that system will remain in state i while operating rather than switching to other state.
$BM_0(t)$	Probability that a repairman working on a minor repair to the system in regenerative state i at time t without switching into

another regenerative state at time $t = 0$

For other notations, one may refer to [11]

4. STATES DESCRIPTION

The state transition diagram (Figure 1) shows the following states at a given time point as:

- | | | |
|--|--|--|
| State 0: (O, ws, cs); | State 1: (F_{wmjr}, O, CS_a); | State 2: (F_{wmnr}, O, CS_a); |
| State 3: (F_{mjr}, O, ws); | State 4: (F_{mnr}, O, ws); | State 5: (F_{mjr}, F_{wmjr}, O); |
| State 6: (F_{mjr}, O, F_{wmnr}); | State 7: (F_{mnr}, F_{wmjr}, O); | State 8: (F_{mnr}, O, F_{wmnr}); |
| State 9: ($F_{mjr}, F_{wmjr}, F_{wmjr}$); | State 10: ($F_{mjr}, F_{wmjr}, F_{wmnr}$); | State 11: ($F_{mnr}, F_{wmjr}, F_{wmjr}$); |
| State 12: ($F_{mnr}, F_{wmjr}, F_{wmnr}$); | | |

where,

- | | |
|--------------|--------------------------------|
| O : | operative unit. |
| ws : | warm standby unit. |
| cs : | cold standby unit. |
| CS_a : | cold standby under activation. |
| F_{mjr} : | under major repair. |
| F_{mnr} : | under minor repair. |
| F_{wmjr} : | waiting for major repair. |
| F_{wmnr} : | waiting for minor repair. |

5. MODEL DEVELOPMENT

Figure 1 represents the transition between states of system. The operative states are $\{0,1,2,3,4,5,6,7,8\}$ and failed states are $\{9,10,11,12\}$. A regenerative process is a stochastic process having time points at which the process probabilistically restarts itself, and the associated state of the system is known as the regenerative state, otherwise non-regenerative. Our analysis consistently considers the exponential distribution those results in memorylessness. All the states in the present model are, therefore, regenerative states. Hence, the defined system's possible transitions between different states and entry points into a specific state follow Markov and regenerative processes.

The densities $q_{ij}(t)$, transition from state i to j is as:

$$\begin{aligned}
 q_{01}(t) &= \lambda e^{-(\lambda+\lambda_1)t}, & q_{02}(t) &= \lambda_1 e^{-(\lambda+\lambda_1)t}, & q_{13}(t) &= \beta e^{-\beta t} \\
 q_{24}(t) &= \beta e^{-\beta t}, & q_{30}(t) &= \alpha_1 e^{-(\lambda+\lambda_1+\alpha_1)t}, & q_{35}(t) &= \lambda e^{-(\lambda+\lambda_1+\alpha_1)t} \\
 q_{36}(t) &= \lambda_1 e^{-(\lambda+\lambda_1+\alpha_1)t}, & q_{40}(t) &= \alpha_2 e^{-(\lambda+\lambda_1+\alpha_2)t}, & q_{47}(t) &= \lambda e^{-(\lambda+\lambda_1+\alpha_2)t} \\
 q_{48}(t) &= \lambda_1 e^{-(\lambda+\lambda_1+\alpha_2)t}, & q_{53}(t) &= \alpha_1 e^{-(\lambda+\alpha_1)t}, & q_{59}(t) &= \lambda e^{-(\lambda+\alpha_1)t} \\
 q_{64}(t) &= \alpha_1 e^{-(\lambda+\alpha_1)t}, & q_{6,10}(t) &= \lambda e^{-(\lambda+\alpha_1)t}, & q_{73}(t) &= \alpha_2 e^{-(\lambda+\alpha_2)t} \\
 q_{7,11}(t) &= \lambda e^{-(\lambda+\alpha_2)t}, & q_{84}(t) &= \alpha_2 e^{-(\lambda+\alpha_2)t}, & q_{8,12}(t) &= \lambda e^{-(\lambda+\alpha_2)t} \\
 q_{95}(t) &= \alpha_1 e^{-\alpha_1 t}, & q_{10,7}(t) &= \alpha_1 e^{-\alpha_1 t}, & q_{11,5}(t) &= \alpha_2 e^{-\alpha_2 t} \\
 q_{12,5}(t) &= \alpha_2 e^{-\alpha_2 t}
 \end{aligned}
 \tag{1}$$

Now defining steady state probability,

$$p_{ij} = \lim_{s \rightarrow 0} q_{ij}^*(s) = \lim_{s \rightarrow 0} L[q_{ij}^*(t)] = \lim_{s \rightarrow 0} \int_0^\infty e^{-st} q_{ij}(t) dt
 \tag{2}$$

$$\begin{aligned}
 p_{01} &= \frac{\lambda}{\lambda+\lambda_1} & p_{02} &= \frac{\lambda_1}{\lambda+\lambda_1}; & p_{13} &= 1; & p_{24} &= 1 \\
 p_{30} &= \frac{\alpha_1}{\lambda+\lambda_1+\alpha_1} & p_{35} &= \frac{\lambda}{\lambda+\lambda_1+\alpha_1} & p_{36} &= \frac{\lambda_1}{\lambda+\lambda_1+\alpha_1} & p_{40} &= \frac{\alpha_2}{\lambda+\lambda_1+\alpha_2} \\
 p_{47} &= \frac{\lambda}{\lambda+\lambda_1+\alpha_2} & p_{48} &= \frac{\lambda_1}{\lambda+\lambda_1+\alpha_2} & p_{53} &= \frac{\alpha_1}{\lambda+\alpha_1} & p_{59} &= \frac{\lambda}{\lambda+\alpha_1} \\
 p_{64} &= \frac{\alpha_1}{\lambda+\alpha_1} & p_{6,10} &= \frac{\lambda}{\lambda+\alpha_1} & p_{73} &= \frac{\alpha_2}{\lambda+\alpha_2} & p_{7,11} &= \frac{\lambda}{\lambda+\alpha_1}
 \end{aligned}$$

$$\begin{aligned}
 p_{84} &= \frac{\alpha_2}{\lambda + \alpha_2} & p_{8,12} &= \frac{\lambda}{\lambda + \alpha_2} & p_{95} &= 1 & p_{10,7} &= 1 \\
 p_{11,5} &= 1 & p_{12,7} &= 1 & & & &
 \end{aligned}
 \tag{3}$$

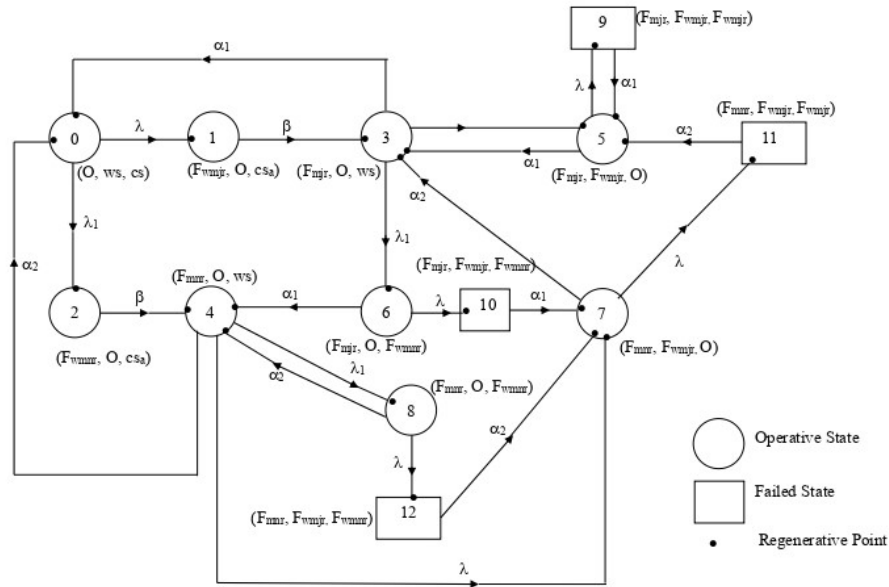


Figure 1: State Transition Diagram

The Mean Sojourn Time (μ_i), i.e., the stay time in particular state i , in state S_i as:

$$\mu_i = E(t) = \int_0^{\infty} t \cdot (\text{corresponding p.d.f. for moving from } i^{\text{th}} \text{ state}) dt. \tag{4}$$

$$\begin{aligned}
 \mu_0 &= \int_0^{\infty} e^{-(\lambda + \lambda_1)t} dt = \frac{1}{\lambda + \lambda_1}, & \mu_1 &= \int_0^{\infty} e^{-(\beta)t} dt = \frac{1}{\beta} \\
 \mu_2 &= \int_0^{\infty} e^{-(\beta)t} dt = \frac{1}{\beta}, & \mu_3 &= \int_0^{\infty} e^{-(\lambda + \lambda_1 + \alpha_1)t} dt = \frac{1}{\lambda + \lambda_1 + \alpha_1} \\
 \mu_4 &= \int_0^{\infty} e^{-(\lambda + \lambda_1 + \alpha_2)t} dt = \frac{1}{\lambda + \lambda_1 + \alpha_2}, & \mu_5 &= \int_0^{\infty} e^{-(\lambda + \alpha_1)t} dt = \frac{1}{\lambda + \alpha_1} \\
 \mu_6 &= \int_0^{\infty} e^{-(\lambda + \alpha_1)t} dt = \frac{1}{\lambda + \alpha_1}, & \mu_7 &= \int_0^{\infty} e^{-(\lambda + \alpha_2)t} dt = \frac{1}{\lambda + \alpha_2} \\
 \mu_8 &= \int_0^{\infty} e^{-(\lambda + \alpha_2)t} dt = \frac{1}{\lambda + \alpha_2}, & \mu_9 &= \int_0^{\infty} e^{-(\alpha_1)t} dt = \frac{1}{\alpha_1} \\
 \mu_{10} &= \int_0^{\infty} e^{-(\alpha_1)t} dt = \frac{1}{\alpha_1}, & \mu_{11} &= \int_0^{\infty} e^{-(\alpha_2)t} dt = \frac{1}{\alpha_2} \\
 \mu_{12} &= \int_0^{\infty} e^{-(\alpha_2)t} dt = \frac{1}{\alpha_2}, & & & &
 \end{aligned}
 \tag{5}$$

When counting from the epoch of entry into state i , the unconditional mean time that the system needs to transit any regenerative state j is calculated mathematically as

$$m_{ij} = \int_0^{\infty} tq_{ij}(t) dt = q'_{ij}^*(0) \tag{6}$$

$$\begin{aligned}
 m_{01} + m_{02} &= \mu_0 & m_{13} &= \mu_1 & m_{24} &= \mu_2 & m_{30} + m_{35} + m_{36} &= \mu_4 \\
 m_{40} + m_{47} + m_{48} &= \mu_4 & m_{53} + m_{59} &= \mu_5 & m_{64} + m_{6,10} &= \mu_6 & m_{73} + m_{7,11} &= \mu_7
 \end{aligned}$$

$$\begin{aligned}
m_{84} + m_{8,12} &= \mu_8 & m_{95} &= \mu_9 & m_{10,7} &= \mu_{10} & m_{11,5} &= \mu_{11} \\
m_{12,7} &= \mu_{12}
\end{aligned}
\tag{7}$$

6. SYSTEM PERFORMANCE MEASURES

6.1. Mean Time to System Failure (MTSF):

By using the definition of $\phi_i(t)$ and $Q_{ij}(t)$ define in section 3 and regarding the failed state as absorbing state, we have the recurrence relation for $\phi_i(t)$ from Figure 1 as

$$\begin{aligned}
\phi_0(t) &= Q_{01}(t) \otimes \phi_1(t) + Q_{02}(t) \otimes \phi_2(t) \\
\phi_1(t) &= Q_{13}(t) \otimes \phi_3(t) \\
\phi_2(t) &= Q_{24}(t) \otimes \phi_4(t) \\
\phi_3(t) &= Q_{30}(t) \otimes \phi_0(t) + Q_{35}(t) \otimes \phi_5(t) + Q_{36}(t) \otimes \phi_6(t) \\
\phi_4(t) &= Q_{40}(t) \otimes \phi_0(t) + Q_{47}(t) \otimes \phi_7(t) + Q_{48}(t) \otimes \phi_8(t) \\
\phi_5(t) &= Q_{53}(t) \otimes \phi_3(t) + Q_{59}(t) \otimes \phi_9(t) \\
\phi_6(t) &= Q_{64}(t) \otimes \phi_4(t) + Q_{610}(t) \otimes \phi_{10}(t) \\
\phi_7(t) &= Q_{73}(t) \otimes \phi_4(t) + Q_{711}(t) \otimes \phi_{11}(t) \\
\phi_8(t) &= Q_{84}(t) \otimes \phi_4(t) + Q_{812}(t) \otimes \phi_{12}(t)
\end{aligned}$$

Taking Laplace Stieljes transform of above equation and solving for $\phi_0^{**}(s)$, by crammer rule, we get

$$\phi_0^{**}(s) = \frac{M_1(s)}{T_1(s)}$$

where,

$$M_1(s) = \begin{vmatrix} 1 & -Q_{01}^{**}(s) & -Q_{02}^{**}(s) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -Q_{13}^{**}(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -Q_{24}^{**}(s) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -Q_{35}^{**}(s) & -Q_{36}^{**}(s) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -Q_{47}^{**}(s) & -Q_{48}^{**}(s) \\ -Q_{59}^{**}(s) & 0 & 0 & -Q_{53}^{**}(s) & 0 & 1 & 0 & 0 & 0 \\ -Q_{610}^{**}(s) & 0 & 0 & 0 & -Q_{64}^{**}(s) & 0 & 1 & 0 & 0 \\ -Q_{711}^{**}(s) & 0 & 0 & -Q_{73}^{**}(s) & 0 & 0 & 0 & 1 & 0 \\ -Q_{812}^{**}(s) & 0 & 0 & 0 & -Q_{84}^{**}(s) & 0 & 0 & 0 & 1 \end{vmatrix}$$

and

$$T_1(s) = \begin{vmatrix} 1 & -Q_{01}^{**}(s) & -Q_{02}^{**}(s) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -Q_{13}^{**}(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -Q_{24}^{**}(s) & 0 & 0 & 0 & 0 \\ -Q_{30}^{**}(s) & 0 & 0 & 1 & 0 & -Q_{35}^{**}(s) & -Q_{36}^{**}(s) & 0 & 0 \\ -Q_{40}^{**}(s) & 0 & 0 & 0 & 1 & 0 & 0 & -Q_{47}^{**}(s) & -Q_{48}^{**}(s) \\ 0 & 0 & 0 & -Q_{53}^{**}(s) & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -Q_{64}^{**}(s) & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -Q_{73}^{**}(s) & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -Q_{84}^{**}(s) & 0 & 0 & 0 & 1 \end{vmatrix}$$

The mean time to system failure (MTSF) is given by:

$$\begin{aligned}
MTSF &= \lim_{s \rightarrow 0} \frac{1 - \phi_0^{**}(s)}{s} \\
&= \lim_{s \rightarrow 0} \frac{1 - \frac{M_1(s)}{T_1(s)}}{s} \\
&= \lim_{s \rightarrow 0} \frac{T_1(s) - M_1(s)}{sT_1(s)}
\end{aligned}$$

Apply L'Hospital Rule , We get:

$$MTSF = \lim_{s \rightarrow 0} \frac{T_1'(s) - M_1'(s)}{sT_1'(s) + T_1(s)} = \frac{T_1'(0) - M_1'(0)}{T_1(0)}$$

which is evaluated by solving determinants in $M_1(s), T_1(s)$ and then further required steps using MATLAB.

6.2. Availability Analysis

We can obtain the following expression for the availability by proceeding the same way as in earlier section 6.1,

$$A_0^*(s) = \frac{M_2(s)}{T_2(s)}$$

where,

$$M_2(s) = \begin{vmatrix} B_0^*(s) - q_{01}^*(s) - q_{02}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B_1^*(s) & 1 & 0 & -q_{13}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B_2^*(s) & 0 & 1 & 0 & -q_{24}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B_3^*(s) & 0 & 0 & 1 & 0 & -q_{35}^*(s) - q_{36}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 \\ B_4^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & -q_{47}^*(s) - q_{48}^*(s) & 0 & 0 & 0 & 0 \\ B_5^*(s) & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{59}^*(s) & 0 & 0 \\ B_6^*(s) & 0 & 0 & 0 & -q_{64}^*(s) & 0 & 1 & 0 & 0 & 0 & -q_{6,10}^*(s) & 0 \\ B_7^*(s) & 0 & 0 & -q_{73}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) \\ B_8^*(s) & 0 & 0 & 0 & -q_{84}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{8,12}^*(s) \\ 0 & 0 & 0 & 0 & 0 & -q_{95}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{10,7}^*(s) & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -q_{11,5}^*(s) & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{12,7}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

$$B_i^*(s) = L(B_i(t)) = \lim_{s \rightarrow 0} \int_0^\infty e^{-st} B_i(t) dt$$

$$T_2(s) = \begin{vmatrix} 1 & -q_{01}^*(s) - q_{02}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -q_{13}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -q_{24}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -q_{30}^*(s) & 0 & 0 & 1 & 0 & -q_{35}^*(s) - q_{36}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -q_{40}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & -q_{47}^*(s) - q_{48}^*(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{59}^*(s) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -q_{64}^*(s) & 0 & 1 & 0 & 0 & 0 & -q_{6,10}^*(s) & 0 & 0 \\ 0 & 0 & 0 & -q_{73}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) & 0 \\ 0 & 0 & 0 & 0 & -q_{84}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{8,12}^*(s) \\ 0 & 0 & 0 & 0 & 0 & -q_{95}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{10,7}^*(s) & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -q_{11,5}^*(s) & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{12,7}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

The steady-state availability is given by

$$A_0 = \lim_{s \rightarrow 0} sA_0^*(s) = \lim_{s \rightarrow 0} \frac{sM_2(s)}{T_2(s)}$$

After being solved, this takes on an indeterminate form. Applying the L'Hospital Rule, we thus obtain:

$$A_0 = \lim_{s \rightarrow 0} \frac{sM_2'(s) + M_2(s)}{T_2'(s)}$$

$$= \lim_{s \rightarrow 0} \frac{M_2(s)}{T_2'(s)} = \frac{M_2(0)}{T_2'(0)}$$

Further calculations have been performed using MATLAB since there is a huge determinant and it's derivative to solve.

6.3. Busy Period Analysis for Major Repair

The average time for which repairman is busy for major repair of the system is given by

$$BJ_0^*(s) = \frac{M_3(s)}{T_2(s)}$$

where

$$M_3(s) = \begin{vmatrix} 1 & -q_{01}^*(s)-q_{02}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -q_{13}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -q_{24}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ W_3^*(s) & 0 & 0 & 1 & 0 & -q_{35}^*(s)-q_{36}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -q_{47}^*(s)-q_{48}^*(s) & 0 & 0 & 0 & 0 & 0 \\ W_5^*(s) & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{59}^*(s) & 0 & 0 & 0 \\ W_6^*(s) & 0 & 0 & 0 & -q_{64}^*(s) & 0 & 1 & 0 & 0 & 0 & -q_{6,10}^*(s) & 0 & 0 \\ 0 & 0 & 0 & -q_{73}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) & 0 \\ 0 & 0 & 0 & 0 & -q_{84}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{8,12}^*(s) \\ 0 & 0 & 0 & 0 & 0 & -q_{95}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ W_{10}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & -q_{10,7}^*(s) & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -q_{11,5}^*(s) & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{12,7}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

and $T_2(s)$ is same as mention in section 6.2

In steady state, the average time for which repairman is busy for major repair of the system is given by:

$$BJ_0 = \lim_{s \rightarrow 0} sBJ_0^*(s) = \lim_{s \rightarrow 0} \frac{s.M_3(s)}{T_2(s)}$$

After being solved, this takes on an indeterminant form. Applying the L'Hospital Rule, we thus obtain:

$$BJ_0 = \lim_{s \rightarrow 0} \frac{sM_3'(s) + M_3(s)}{T_2'(s)}$$

$$= \lim_{s \rightarrow 0} \frac{M_3(s)}{T_2'(s)} = \frac{M_3(0)}{T_2'(0)}$$

Further calculations have been performed using MATLAB since there is a huge determinant and it's derivative to solve.

6.4. Busy Period Analysis for Minor Repair

The average time for which repairman is busy for minor repair of the system is given by

$$BM_0^*(s) = \frac{M_4(s)}{T_2(s)}$$

where,

$$M_4(s) = \begin{vmatrix} 1 & -q_{01}^*(s)-q_{02}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -q_{13}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -q_{24}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -q_{35}^*(s)-q_{36}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -q_{47}^*(s)-q_{48}^*(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -q_{73}^*(s) & 0 & 1 & 0 & 0 & 0 & -q_{59}^*(s) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -q_{64}^*(s) & 0 & 1 & 0 & 0 & 0 & -q_{6,10}^*(s) & 0 & 0 \\ W_7^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) & 0 \\ W_8^*(s) & 0 & 0 & 0 & -q_{84}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{8,12}^*(s) \\ 0 & 0 & 0 & 0 & 0 & -q_{95}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{10,7}^*(s) & 0 & 0 & 1 & 0 & 0 \\ W_{11}^*(s) & 0 & 0 & 0 & 0 & -q_{11,5}^*(s) & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{12,7}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

where, $T_2(s)$ is same as mention in section 6.2

In steady state, the average time for which repairman is busy for minor repair of the system is given by

$$BM_0 = \lim_{s \rightarrow 0} sBM_0^*(s) = \lim_{s \rightarrow 0} \frac{sM_4(s)}{T_2(s)}$$

After being solved, this takes on an indeterminant form. Applying the L'Hospital Rule, we thus obtain:

$$\begin{aligned} BM_0 &= \lim_{s \rightarrow 0} \frac{sM_4'(s) + M_4(s)}{T_2'(s)} \\ &= \lim_{s \rightarrow 0} \frac{M_4(s)}{T_2'(s)} = \frac{M_4(0)}{T_2'(0)} \end{aligned}$$

Further calculations have been performed using MATLAB since there is a huge determinant and it's derivative to solve.

6.5. Expected Number of the Visit of the Repairman

The expected number of server's visits obtained as:

$$V_0^*(s) = \frac{M_5(s)}{T_2(s)}$$

where,

$$M_5(s) = \begin{vmatrix} -q_{01}^*(s)-q_{01}^*(s)-q_{02}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -q_{02}^*(s) & 1 & 0 & -q_{13}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -q_{24}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -q_{35}^*(s)-q_{36}^*(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -q_{47}^*(s)-q_{48}^*(s) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{59}^*(s) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -q_{64}^*(s) & 0 & 1 & 0 & 0 & 0 & -q_{6,10}^*(s) & 0 & 0 \\ 0 & 0 & 0 & -q_{73}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{7,11}^*(s) & 0 \\ 0 & 0 & 0 & 0 & -q_{84}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -q_{8,12}^*(s) \\ 0 & 0 & 0 & 0 & 0 & -q_{95}^*(s) & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{10,7}^*(s) & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -q_{11,5}^*(s) & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -q_{12,7}^*(s) & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

and $T_2(s)$ is same as mention in section 6.2

In steady state, the expected number of server's visits obtained as

$$V_0 = \lim_{s \rightarrow 0} sV_0^*(s) = \lim_{s \rightarrow 0} \frac{s.M_5(s)}{T_2(s)}$$

After being solved, this takes on an indeterminant form. Applying the L'Hospital Rule, we thus obtain:

$$\begin{aligned} V_0 &= \lim_{s \rightarrow 0} \frac{sM_5'(s) + M_5(s)}{T_2'(s)} \\ &= \lim_{s \rightarrow 0} \frac{M_5(s)}{T_2'(s)} = \frac{M_5(0)}{T_2'(0)} \end{aligned}$$

Further calculations have been performed using MATLAB since there is a huge determinant and it's derivative to solve.

7. PROFIT ANALYSIS:

Profit is the difference between total value generated and total expenditure. Thus, in steady state, the expected profit is

$$\text{Profit (P)} = C_0(A_0) - C_1(BJ_0) - C_2(BM_0) - C_3(V_0)$$

where

C_0 = Revenue per unit up time

C_1 = Cost per unit up time for which the repairman is busy for major repair.

C_2 = Cost per unit up time for which the repairman is busy for minor repair.

C_3 = Cost per visit of the repairman

8. NUMERICAL ANALYSIS

Various profit function graphs have been generated to determine the effect of various parameters, as illustrated. Assuming the hypothetical value of parameters as

$$\lambda = 0.1, \lambda_2 = 0.2, \beta = 0.3, \alpha_1 = 0.4, \alpha_2 = 0.05, C_0 = 200, C_1 = 500, C_2 = 350, C_3 = 50$$

The change in profit function (P) for different value of failure rate (λ) and revenue (C_0) shown in Figure 2.

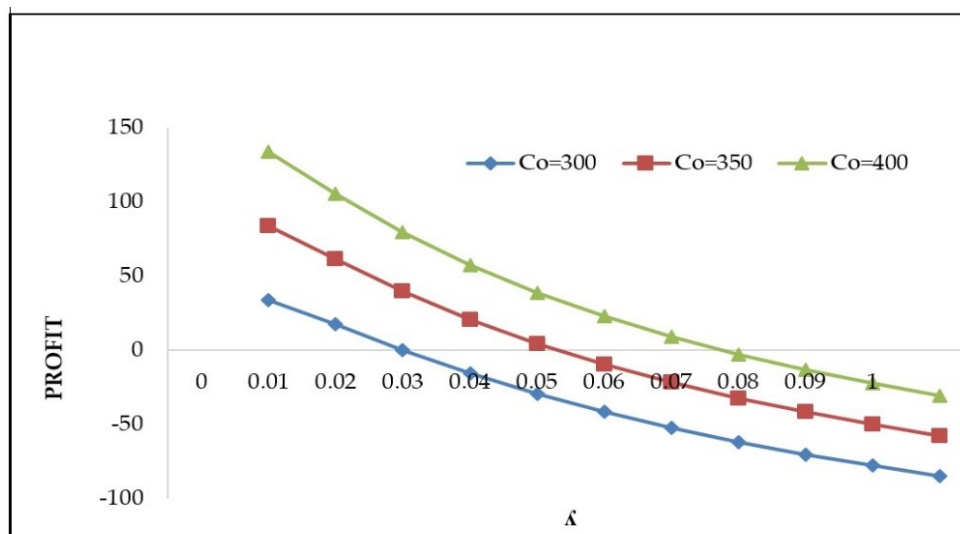


Figure 2: Profit for varied λ and C_0

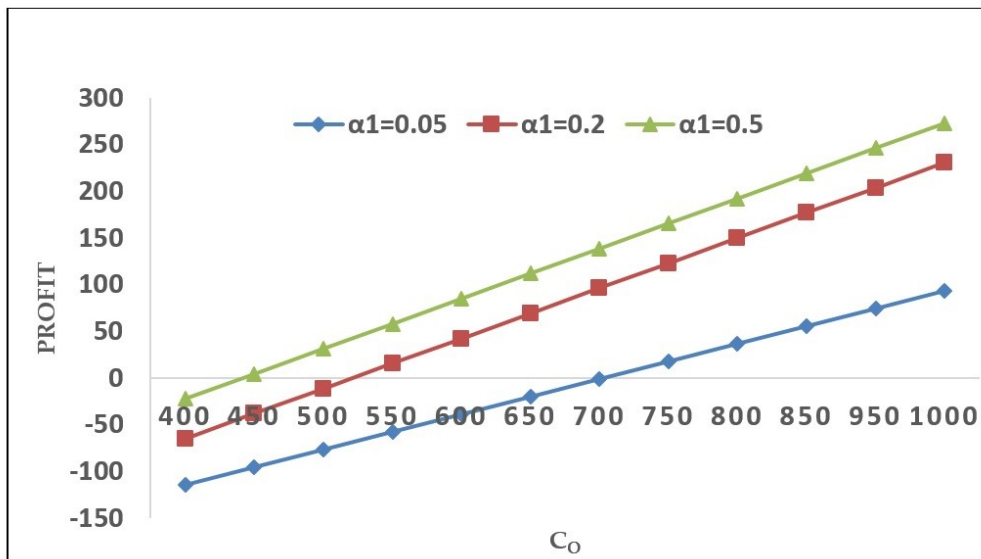


Figure 3: Profit for varied C_0 and α_1

It can be seen that, as the failure rate (λ) of operative unit increases, profit decreases and with increases in unit revenue (C_0), the profit increases. Similarly, in figure 3, the effect of repair rate (α_1) on profit function has revealed. As the value of repair rate (α_1) higher, the profit also become higher. Moreover, the change in profit due to activation rate (β) and cost per visit of repairman (C_3) is shown in Figure 4.

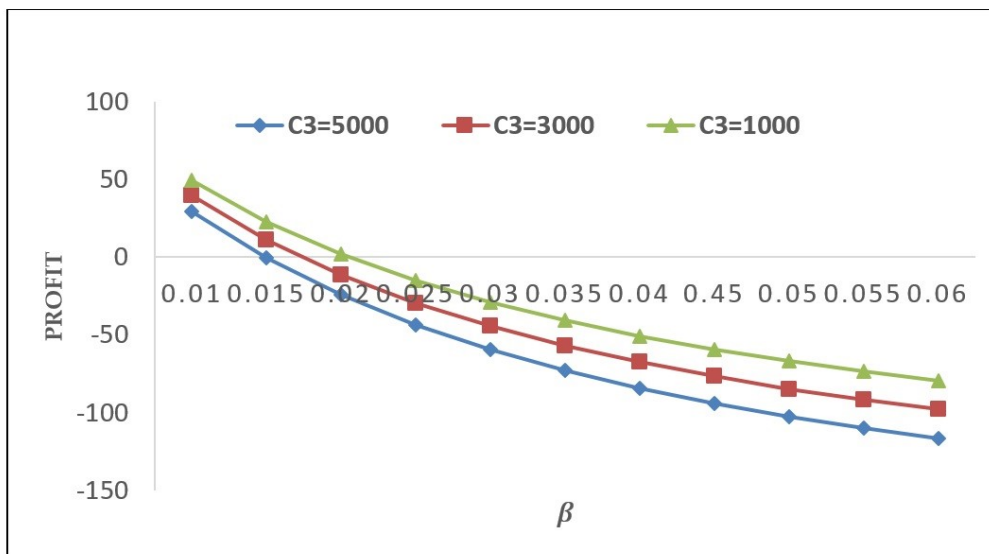


Figure 4: Profit for varied β and C_3

With increase in β and C_3 the profit decreases. Figure 5 reveals the change in profit w.r.t. C_2 and λ_1 respectively. With increase in C_2 and failure rate (λ_1) of warm standby unit, the profit decreases.

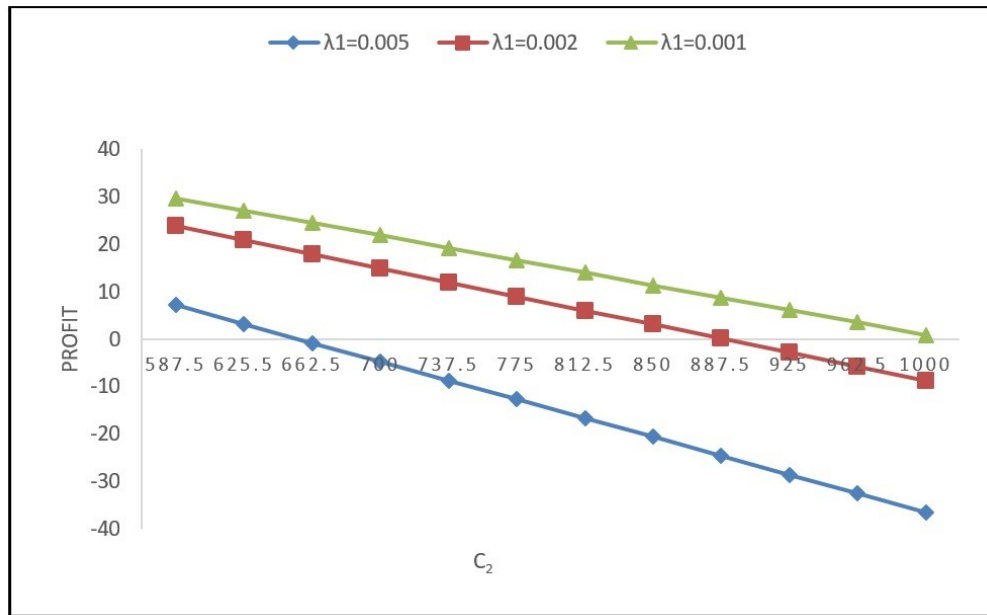


Figure 5: Profit for varied C₂ and λ₁

The bounds of λ, C_0, β and C_2 for profitable of system depicted in Table 1. It can be interpreted for $C_0 = 300$, failure rate should be less than equal to 0.003 for system to be profitable. Similarly, interpretation can be done for other value from previous figures 2, 3, 4 and located in Table 1.

Table 1: Profit Analysis of the model

Fig No.	Parameter value	Profit ≥ 0 if
2	$C_0 = 300$	$\lambda \leq 0.03$
	$C_0 = 350$	$\lambda \leq 0.05$
	$C_0 = 400$	$\lambda \leq 0.08$
3	$\alpha_1 = 0.05$	$C_0 \geq 450$
	$\alpha_1 = 0.02$	$C_0 \geq 550$
	$\alpha_1 = 0.5$	$C_0 \geq 700$
4	$C_3 = 5000$	$\beta \leq 0.015$
	$C_3 = 3000$	$\beta \leq 0.02$
	$C_3 = 1000$	$\beta \leq 0.025$
5	$\lambda = 0.005$	$C_2 \leq 662.5$
	$\lambda = 0.002$	$C_2 \leq 887.5$
	$\lambda = 0.001$	$C_2 \leq 1000$

9. CONCLUSION

A system reliability model that takes into account one operational, one cold standby, and one warm standby unit is examined. Various system measures have been drafted. The exponential case is studied numerically. Profit declines in accordance to the failure rate, activation rate, and cost per visit of the repairman. On the other hand, profit increases as revenue rises. Cut-off points for the system's cost and revenue per unit of time have also been determined in order to analyze the profitability element, which may aid them in making crucial judgments about the system's economics.

REFERENCES

- [1] Goel, L. R. and Gupta, P. (1983). Stochastic behaviour of a two-unit (dissimilar) hot standby system with three modes. *Microelectron. Reliab.*, 23: 10350–1040.
- [2] Tuteja, R. K., Arora, R. T. and Taneja, G. (1991). Stochastic behaviour of a two-unit system with two types of repairman and subject to random inspection. *Microelectron. Reliab.*, 31: 79–83.
- [3] Kumar, A., Kumar, S., Gupta, K. and Taneja, G. (1996). Probabilistic analysis of a two-unit cold standby system with instructions at need. *Microelectron. Reliab.*, 36: 273–277.
- [4] EL-Said, K. M. and EL-Said, M. S. (2005). Profit Analysis of a Two Unit Cold Standby System with Preventive Maintenance and Random Change in Units. *J. Math. Stat.*, 1: 71–77.
- [5] Parashar, B. and Taneja, G. (2007). Reliability and profit evaluation of a PLC hot standby system based on a master-slave concept and two types of repair facilities. *IEEE Trans. Reliab.*, 56: 534–539.
- [6] Mahmoud, M. A. W. and Moshref, M. E. (2009). On a two-unit cold standby system considering hardware, human error failures and preventive maintenance. *Math. Comput. Model.*, 51: 736–745.
- [7] Manocha, A. and Taneja, G. (2015). Stochastic analysis of a two-unit cold standby system with arbitrary distributions for life, repair and waiting times. *Int. J. Performability Eng.*, 11: 293–299.
- [8] Malhotra, and Taneja, G. (2015). Comparative study between a single unit system and a two-unit cold standby system with varying demand. *Springerplus*, 4: 1–17.
- [9] Manocha, A., Taneja, G. and Singh, S. (2017). Stochastic and cost-benefit analysis of two-unit hot standby database system. *Int. J. Performability Eng.*, 13: 63–72.
- [10] Adlakha, N., Taneja, G. and Batra, S. (2017). Reliability and cost-benefit analysis of a two-unit cold standby system used for communication through satellite with assembling and activation time. *Int. J. Appl. Eng. Res.*, 12: 9697–9702.
- [11] Batra, S. and Taneja, G. (2018). Optimization of number of hot standby units through reliability models for a system operative with one unit. *International Journal of Agricultural and Statistical Sciences*, 14: 0973–1903.
- [12] Batra, S. and Taneja, G. (2018). Reliability and optimum analysis for number of standby units in a system working with one operative unit. *International Journal of Applied Engineering Research*, 13: 2791–2797.
- [13] Batra, S. and Taneja, G. (2018). A reliability model for the optimum number of standby unit in a system working with two operative units. *Ciencia e Tecnica*, 33: 0254–0223.
- [14] Batra, S. and Taneja, G. (2019). Reliability modeling and optimization of the number of hot standby units in a system working with two operative units. *An international journal of advanced computer technology*, 10: 3059–3068.
- [15] Levitin, G., Finkelstein, M., and Xiang, Y. (2020). Optimal preventive replacement for cold standby systems with elements exposed to shocks during operation and task transfers. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 10: 2168–2216.
- [16] Malhotra, R., Dureja, T., and Goyal, A. (2021). Reliability analysis a two-unit cold redundant system working in a pharmaceutical agency with preventive maintenance. *In Journal of Physics: Conference Series*, 10: 1742–1750.
- [17] Jia, H., Peng, R., Yang, L., Wu, T., Liu, D., and Li, Y. (2022). Reliability evaluation of demand-based warm standby systems with capacity storage, *Reliability Engineering & System Safety*, 218: 108–132.

A TWO-STATE FEEDBACK RETRIAL QUEUEING SYSTEM HAVING TWO HETEROGENEOUS PARALLEL SERVERS AND IMPATIENT CUSTOMERS

Neelam Singla¹ and Harwinder Kaur^{2*}

¹ Department of Statistics, Punjabi University, Patiala-India, neelgagan2k3@pbi.ac.in

^{2*} University School of Business, Chandigarh University, Gharuan, Mohali-India,
harwinderkaurchahal@gmail.com

Abstract

Objective: In the present paper we consider a two-state retrial queueing system with feedback having two heterogeneous parallel servers and impatient customers. Transient state probabilities for exact number of arrivals and departures from the system will be obtained when both, one or none of the servers is busy. Numerical and graphical solutions will also be obtained. *Methods:* The difference-differential equations governing the system are solved recursively, Laplace transform is then used to obtain the transient state probabilities for exact number of arrivals and departures from the system. *Findings:* Time dependent probabilities are obtained when both, one and none of the servers is busy. Numerical and Graphical solutions are also obtained using MATLAB programming. *Novelty:* In past research, models considered arrivals and departures from the orbit whereas in present model arrivals and departures from the system are studied along with the concept of feedback. *Applications:* This type of model is implemented in computer systems.

1. INTRODUCTION

In addition to classical queueing systems, there exists a new class of queueing models, popularly known as retrial queueing models. Recently, a significant contribution has been provided in this direction. The retrial queueing systems are characterized by an arriving customer who is served instantly if it finds the server free else leaves the service area and joins the virtual queue (orbit) and repeats its demand for service after a random amount of time from the orbit. 'Basic problems of telephone traffic and the influence of repeated calls' published by [1] which is the initial work on retrial queues. Books [2], [3] are available as a great source for retrial queues.

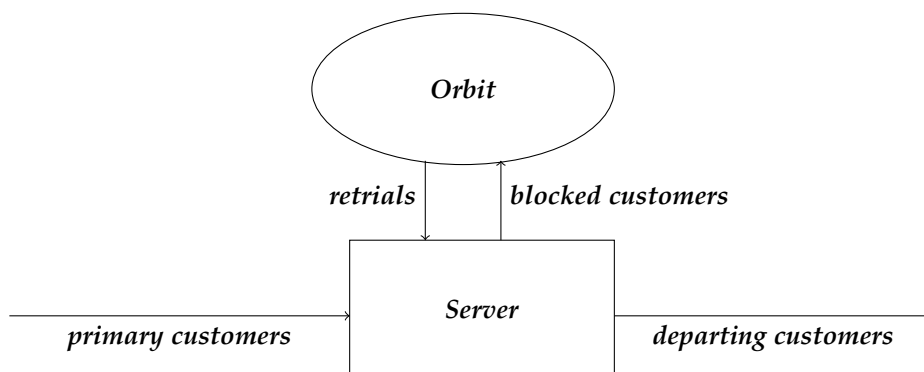


Figure 1: Basic Structure of a Retrial Queueing System

In [4] the author analyzed 'Some new results for the $M/M/1$ queue' in which a closed form solution with finite sums is obtained for the probability that exactly i number of arrivals, j services occur over a time interval t . In [5] authors worked on 'Performance analysis of a two-state queueing model with retrials' where they obtained the transient state probabilities for exact number of arrivals and departures from the system.

In some systems various servers possess different service rate depending on the requirements and other reasons, these servers are called heterogeneous or non-identical servers. Here the same type of job is rendered by different servers with different service rates. The expressions for the Laplace transforms of the waiting time as well as arbitrary moments are derived in [6]. 'Retrial queueing model with two heterogeneous server using matrix geometric method' done by [7] where the stationary analysis has been carried out using matrix geometric method.

In case of higher demand for any service, renege and balked i.e., impatient customers are usually observed. A wide study has been done in this direction for both standard and retrial queueing systems. In [8] blocking probability and the mean number of customers in the orbit was derived. 'A Single Server Retrial Queue with Impatient Customers' is analyzed in [9].

In some situations, customers seek service again as a result of dissatisfaction from the received service. This concept is known as feedback. For instance: when a message faces a failed transmission in multiple access telecommunication systems, it can be sent again. 'On multiserver feedback retrial queues with balking and control retrial rate' published by [10]. The author in [11] analyzed 'On an unreliable-server retrial queue with customer feedback and impatience'. The time dependent probability generating function was obtained using supplementary variable technique in [12].

The section wise description of the paper is as follows:

The model is described theoretically as well as mathematically in section 2. In section 3 the transient state probabilities are obtained. Verification of some results is given in section 4. The numerical and graphical results are obtained in section 5. Section 6 gives the busy period probabilities along with its graphical representation. Finally, the paper is concluded in section 7 and followed by references at the end.

2. MODEL DESCRIPTION

We considered a two-state retrial queueing model with feedback having two non-identical parallel servers and impatient customers. The primary or fresh customers arrive at system following Poisson process. On arrival if the entering customer finds any of the servers free it is served immediately else it may balk from the system or may join the orbit and retry for service as a repeated call or secondary customer. The secondary customers also follow Poisson process. The customers retrying from orbit in case getting busy servers may renege from the orbit. Also, if the customer after service feels unsatisfied may join the orbit in order to obtain satisfied service. Service times follow exponential distribution.

- Arrival Process: The primary calls arrive at the system following Poisson process with mean arrival rate λ .
- An arriving customer joins the first server with probability a_1 and second with probability a_2 .
- The Retrial Process: On arrival of a customer if any of the servers is free, it is served immediately. Otherwise, the customer joins the orbit and calls repeatedly until any of the servers is free. The retrial customers also follow Poisson process with parameter θ .
- Impatience: The fresh customers on encountering busy server may balk with probability $(1 - \beta)$, $\beta > 0$. Also, the customers retrying for service from orbit as secondary customers on encountering busy server may renege with probability $(1 - \alpha)$, $\alpha > 0$.
- Feedback Rule: After receiving service, the customer joins the orbit with probability γ (i.e., when unsatisfied) and departs from the system with probability $1 - \gamma$.

- The Service Process: Service times follow exponential distribution with parameters μ_1 for first server and μ_2 for second server.

The input flow of primary calls, intervals between repetitions, service times are statistically independent.

Laplace Transformation of $\bar{f}(s)$ of $f(t)$ is given by:

$$\bar{f}(s) = \int_0^{\infty} e^{-st} f(t) dt; \quad \text{Re}(s) > 0$$

The Laplace Inverse of

$$\frac{Q(p)}{P(p)} = \sum_{k=1}^n \sum_{l=1}^{m_k} \frac{t^{m_k-l} e^{a_k t}}{(m_k - l)!(l - 1)!} \times \left(\frac{d}{dp}\right)^{l-1} \frac{Q(p)}{P(p)} (p - a_k)^{m_k} \quad \forall p = a_k, a_i \neq a_k \text{ for } i \neq k$$

where

$$P(p) = (p - a_1)^{m_1} (p - a_2)^{m_2} \dots (p - a_n)^{m_n}$$

$Q(p)$ is a polynomial of degree $< m_1 + m_2 + m_3 + \dots + m_n - 1$

The Laplace Inverse of

$$\bar{N}_{n_1, n_2, n_3}^{a, b, c}(s) = \frac{1}{(s + a)^{n_1} + (s + b)^{n_2} + (s + c)^{n_3}} \quad is$$

$$N_{n_1, n_2, n_3}^{a, b, c}(t) = \sum_{l=1}^{n_3} \sum_{m=1}^l \frac{e^{-at} t^{n_3-l} (-1)^{m+1} \binom{l-1}{m-1} \left(\prod_{g_1=0}^{l-m-1} (n_1 + g_1)\right) \left(\prod_{g_2=0}^{m-2} (n_2 + g_2)\right)}{(n_3 - l)!(m - 1)!(b - a)^{n_2+m-1} (c - a)^{n_1+l-m}}$$

$$+ \sum_{l=1}^{n_2} \sum_{m=1}^l \frac{e^{-bt} t^{n_2-l} (-1)^{m+1} \binom{l-1}{m-1} \left(\prod_{g_1=0}^{l-m-1} (n_1 + g_1)\right) \left(\prod_{g_2=0}^{m-2} (n_3 + g_2)\right)}{(n_2 - l)!(m - 1)!(a - b)^{n_3+m-1} (c - b)^{n_1+l-m}}$$

$$+ \sum_{l=1}^{n_1} \sum_{m=1}^l \frac{e^{-ct} t^{n_1-l} (-1)^{m+1} \binom{l-1}{m-1} \left(\prod_{g_1=0}^{l-m-1} (n_2 + g_1)\right) \left(\prod_{g_2=0}^{m-2} (n_3 + g_2)\right)}{(n_1 - l)!(m - 1)!(a - c)^{n_3+m-1} (b - c)^{n_2+l-m}}$$

If $L^{-1}\{f(s)\} = F(t)$ and $L^{-1}\{g(s)\} = G(t)$, then

$$L^{-1}\{f(s)g(s)\} = \int_0^1 F(u)G(t - u)du = F * G,$$

$F * G$ is called the convolution of F and G .

2.1. The Two Dimensional State Model

2.1.1 Definitions

$P_{i,j,0}(t)$ = Probability that there are exactly i number of arrivals, j number of departures from the system by time t when both the servers are free.

$P_{i,j,1,k}(t)$ = Probability that there are exactly i number of arrivals, j number of departures from the system by time t and k^{th} ($k = 1$ or 2) server is busy .

$P_{i,j,2}(t)$ = Probability that there are exactly i number of arrivals, j number of departures from the system by time t and both the servers are busy.

$P_{i,j}(t)$ = Probability that there are exactly i number of arrivals, j number of departures from the system by time t .

$$P_{i,j}(t) = P_{i,j,0}(t) + P_{i,j,1,1}(t) + P_{i,j,1,2}(t) + P_{i,j,2}(t), \forall i, j; i \geq j$$

$$P_{i,j,1}(t) = P_{i,j,1,1}(t) + P_{i,j,1,2}(t)$$

$$P_{i,j,0}(t) = 0, i < j; P_{i,j,1,k}(t) = 0, (k = 1 \text{ or } 2) i < j; P_{i,j,2}(t) = 0, i < j$$

Initially

$$P_{0,0,0}(0) = 1;$$

$$P_{i,j,0}(0) = 0; P_{i,j,1,k}(0) = 0 (k = 1 \text{ or } 2); P_{i,j,2}(0) = 0; i \neq j$$

2.2. Difference-Differential Equations Governing the System

$$\frac{d}{dt}P_{i,j,0}(t) = -(\lambda + (i - j)\theta)P_{i,j,0}(t) + \mu_1(1 - \gamma)P_{i,j-1,1,1}(t) + \mu_1\gamma P_{i,j,1,1}(t) + \mu_2(1 - \gamma)P_{i,j-1,1,2}(t) + \mu_2\gamma P_{i,j,1,2}(t); \quad i \geq j \geq 0 \quad (1)$$

$$\frac{d}{dt}P_{1,0,1,1}(t) = -(\lambda + \mu_1)P_{1,0,1,1}(t) + \lambda a_1 P_{0,0,0}(t) + \theta a_1 P_{1,0,0}(t) \quad (2)$$

$$\frac{d}{dt}P_{1,0,1,2}(t) = -(\lambda + \mu_2)P_{1,0,1,2}(t) + \lambda a_2 P_{0,0,0}(t) + \theta a_2 P_{1,0,0}(t) \quad (3)$$

$$\frac{d}{dt}P_{i,j,1,1}(t) = -(\lambda + \mu_1 + (i - j - 1)\theta)P_{i,j,1,1}(t) + \lambda a_1 P_{i-1,j,0}(t) + (i - j)\theta a_1 P_{i,j,0}(t) + \mu_2(1 - \gamma)P_{i,j-1,2}(t) + \mu_2\gamma P_{i,j,2}(t); \quad i > 1, i > j \geq 0 \quad (4)$$

$$\frac{d}{dt}P_{i,j,1,2}(t) = -(\lambda + \mu_2 + (i - j - 1)\theta)P_{i,j,1,2}(t) + \lambda a_2 P_{i-1,j,0}(t) + (i - j)\theta a_2 P_{i,j,0}(t) + \mu_1(1 - \gamma)P_{i,j-1,2}(t) + \mu_1\gamma P_{i,j,2}(t); \quad i > 1, i > j \geq 0 \quad (5)$$

$$\frac{d}{dt}P_{i,j,2}(t) = -(\lambda\beta + \mu_1 + \mu_2 + (i - j - 2)\theta(1 - \alpha))P_{i,j,2}(t) + \lambda \{P_{i-1,j,1,1}(t) + P_{i-1,j,1,2}(t)\} + \lambda\beta(i - \delta_{i-2,j})P_{i-1,j,2}(t) + (i - j - 1)\theta \{P_{i,j,1,1}(t) + P_{i,j,1,2}(t)\} + (i - j - 1)\theta(1 - \alpha)P_{i,j-1,2}(t); \quad i \geq 2, i < j \geq 0 \quad (6)$$

Using Laplace Transform $\bar{f}(s)$ of $f(t)$ given by

$$\bar{f}(s) = \int_0^\infty e^{-st} f(t) dt; \quad Re(s) > 0$$

and using initial condition in equations (1) to (6), we have

$$(s + \lambda + (i - j)\theta)\bar{P}_{0,0,0}(s) = \mu_1(1 - \gamma)\bar{P}_{i,j-1,1,1}(s) + \mu_1\gamma\bar{P}_{i,j,1,1}(s) + \mu_2(1 - \gamma)\bar{P}_{i,j-1,1,2}(s) + \mu_2\gamma\bar{P}_{i,j,1,2}(s); \quad i \geq j \geq 0 \quad (7)$$

$$(s + \lambda + \mu_1)\bar{P}_{1,0,1,1}(s) = \lambda a_1 \bar{P}_{0,0,0}(s) + \theta a_1 \bar{P}_{1,0,0}(s) \quad (8)$$

$$(s + \lambda + \mu_2)\bar{P}_{1,0,1,2}(s) = \lambda a_2 \bar{P}_{0,0,0}(s) + \theta a_2 \bar{P}_{1,0,0}(s) \quad (9)$$

$$(s + \lambda + \mu_1 + (i - j - 1)\theta)\bar{P}_{i,j,1,1}(s) = \lambda a_1 \bar{P}_{i-1,j,0}(s) + (i - j)\theta a_1 \bar{P}_{i,j,0}(s) + \mu_2(1 - \gamma)\bar{P}_{i,j-1,2}(s) + \mu_2\gamma\bar{P}_{i,j,2}(s); \quad i > j \geq 0 \quad (10)$$

$$(s + \lambda + \mu_2 + (i - j - 1)\theta)\bar{P}_{i,j,1,2}(s) = \lambda a_2 \bar{P}_{i-1,j,0}(s) + (i - j)\theta a_2 \bar{P}_{i,j,0}(s) + \mu_1(1 - \gamma)\bar{P}_{i,j-1,2}(s) + \mu_1\gamma\bar{P}_{i,j,2}(s); \quad i > j \geq 0 \quad (11)$$

$$(s + \lambda\beta + \mu_1 + \mu_2 + (i - j - 2)\theta(1 - \alpha))\bar{P}_{i,j,2}(s) = \lambda \{ \bar{P}_{i-1,j,1,1}(s) + \bar{P}_{i-1,j,1,2}(s) \} + \lambda\beta(1 - \delta_{i-2,j})\bar{P}_{i-1,j,2}(s) + (i - j - 1)\theta \{ \bar{P}_{i,j,1,1}(s) + \bar{P}_{i,j,1,2}(s) \} + (i - j - 1)\theta(1 - \alpha)\bar{P}_{i,j-1,2}(s); \quad i > j \geq 0 \quad (12)$$

where

$$\delta_{i-2,j} = \begin{cases} 1; & i - 2 = j \\ 0; & \text{otherwise} \end{cases}$$

3. SOLUTION OF THE PROBLEM

Solving equations (7) -(12) recursively, we get:

$$\bar{P}_{0,0,0}(s) = \frac{1}{s + \lambda} \tag{13}$$

$$\bar{P}_{i,0,0}(s) = \frac{\mu_1 \gamma}{s + \lambda + i\theta} \bar{P}_{i,0,1,1}(s) + \frac{\mu_2 \gamma}{s + \lambda + i\theta} \bar{P}_{i,0,1,2}(s); \quad i \geq 1 \tag{14}$$

$$\bar{P}_{1,0,1,1}(s) = \frac{\lambda a_1}{s + \lambda + \mu_1} \bar{P}_{0,0,0}(s) + \frac{\theta a_1}{s + \lambda + \mu_1} \bar{P}_{1,0,0}(s) \tag{15}$$

$$\bar{P}_{1,0,1,2}(s) = \frac{\lambda a_2}{s + \lambda + \mu_2} \bar{P}_{0,0,0}(s) + \frac{\theta a_2}{s + \lambda + \mu_2} \bar{P}_{1,0,0}(s) \tag{16}$$

$$\begin{aligned} \bar{P}_{i,i,0}(s) = & \frac{\lambda}{s + \lambda} \left[\frac{\mu_1(1-\gamma)a_1}{s + \lambda + \mu_1} + \frac{\mu_2(1-\gamma)a_2}{s + \lambda + \mu_2} \right] \bar{P}_{i-1,i-1,0}(s) + \\ & \frac{\theta}{s + \lambda} \left[\frac{\mu_1(1-\gamma)a_1}{s + \lambda + \mu_1} + \frac{\mu_2(1-\gamma)a_2}{s + \lambda + \mu_2} \right] \bar{P}_{i,i-1,0}(s) + \\ & \frac{\mu_1(1-\gamma)\mu_2(1-\gamma)}{s + \lambda} \left[\frac{1}{s + \lambda + \mu_1} + \frac{1}{s + \lambda + \mu_2} \right] \bar{P}_{i,i-2,2}(s); \quad i \geq 1 \end{aligned} \tag{17}$$

$$\bar{P}_{i,i-1,1,1}(s) = \frac{\lambda a_1}{s + \lambda + \mu_1} \bar{P}_{i-1,i-1,0}(s) + \frac{\theta a_1}{s + \lambda + \mu_1} \bar{P}_{i,i-1,0}(s) + \frac{\mu_2(1-\gamma)}{s + \lambda + \mu_1} \bar{P}_{i,i-2,2}(s); \quad i \geq 2 \tag{18}$$

$$\bar{P}_{i,i-1,1,2}(s) = \frac{\lambda a_2}{s + \lambda + \mu_2} \bar{P}_{i-1,i-1,0}(s) + \frac{\theta a_2}{s + \lambda + \mu_2} \bar{P}_{i,i-1,0}(s) + \frac{\mu_1(1-\gamma)}{s + \lambda + \mu_2} \bar{P}_{i,i-2,2}(s); \quad i \geq 2 \tag{19}$$

$$\begin{aligned} \bar{P}_{i,1,1,1}(s) = & \frac{\lambda a_1}{s + \lambda + \mu_1 + (i-2)\theta} \bar{P}_{i-1,1,0}(s) + \frac{(i-1)\theta a_1}{s + \lambda + \mu_1 + (i-2)\theta} \bar{P}_{i,1,0}(s) + \\ & \frac{\mu_2(1-\gamma)}{s + \lambda + \mu_1 + (i-2)\theta} \bar{P}_{i,0,2}(s) + \frac{\mu_2 \gamma}{s + \lambda + \mu_1 + (i-2)\theta} \bar{P}_{i,1,2}(s); \quad i \geq 3 \end{aligned} \tag{20}$$

$$\begin{aligned} \bar{P}_{i,1,1,2}(s) = & \frac{\lambda a_2}{s + \lambda + \mu_2 + (i-2)\theta} \bar{P}_{i-1,1,0}(s) + \frac{(i-1)\theta a_2}{s + \lambda + \mu_2 + (i-2)\theta} \bar{P}_{i,1,0}(s) + \\ & \frac{\mu_1(1-\gamma)}{s + \lambda + \mu_2 + (i-2)\theta} \bar{P}_{i,0,2}(s) + \frac{\mu_1 \gamma}{s + \lambda + \mu_2 + (i-2)\theta} \bar{P}_{i,1,2}(s); \quad i \geq 3 \end{aligned} \tag{21}$$

$$\begin{aligned} \bar{P}_{i,0,2}(s) = & \frac{\lambda}{s + \lambda \beta + \mu_1 + \mu_2 + (i-2)\theta(1-\alpha)} \{ \bar{P}_{i-1,0,1,1}(s) + \bar{P}_{i-1,0,1,2}(s) \} + \\ & \frac{\lambda \beta}{s + \lambda \beta + \mu_1 + \mu_2 + (i-2)\theta(1-\alpha)} \bar{P}_{i-1,0,2}(s) + \\ & \frac{(i-1)\theta}{s + \lambda \beta + \mu_1 + \mu_2 + (i-2)\theta(1-\alpha)} \{ \bar{P}_{i,0,1,1}(s) + \bar{P}_{i,0,1,2}(s) \}; \quad i \geq 3 \end{aligned} \tag{22}$$

$$\bar{P}_{i,j,2}(s) = \sum_{k=1}^{i-j} \left\{ \prod_{p=k-1}^{i-j-2} \left(\frac{\lambda^{i-j-k} \beta^{i-j-(k+1)}}{s + \lambda \beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \eta'_k(s) \right\} \{ \bar{P}_{j+k,j,1,1}(s) + \bar{P}_{j+k,j,1,2}(s) \}$$

$$+ \sum_{k=1}^{i-j-1} \left\{ \prod_{p=k-1}^{i-j-2} \left(\frac{(\lambda\beta)^{i-j-k-1}k\theta(1-\alpha)}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \right\} \bar{P}_{j+k+1,j-1,2}(s);$$

$$i \geq j+2, j \geq 1 \quad (23)$$

where

$$\eta'_k = \begin{cases} 1 & k = 1 \\ 1 + \frac{(k-1)\theta\beta}{s + \lambda\beta + \mu_1 + \mu_2 + (k-2)\theta(1-\alpha)} & k = 2 \text{ to } i-j-1 \\ \frac{(k-1)\theta}{s + \lambda\beta + \mu_1 + \mu_2 + (k-2)\theta(1-\alpha)} & k = i-j \end{cases}$$

$$\begin{aligned} \bar{P}_{i,j,1,1}(s) &= \frac{\lambda a_1}{s + \lambda + \mu_1 + (i-j-1)\theta} \bar{P}_{i-1,j,0}(s) + \frac{(i-j)\theta a_1}{s + \lambda + \mu_1 + (i-j-1)\theta} \bar{P}_{i,j,0}(s) + \\ &\frac{\mu_2(1-\gamma)}{s + \lambda + \mu_1 + (i-j-1)\theta} \left[\sum_{k=0}^{i-j} \left\{ \prod_{p=k}^{i-j-1} \left(\frac{\lambda^{i-j-k}\beta^{i-j-(k+1)}}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \psi'_k(s) \right\} \right. \\ &\left. \left\{ \bar{P}_{j+k,j-1,1,1}(s) + \bar{P}_{j+k,j-1,1,2}(s) \right\} + \sum_{k=1}^{i-j} \left\{ \prod_{p=k-1}^{i-j-1} \left(\frac{(\lambda\beta)^{i-j-k}k\theta(1-\alpha)}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \right\} \bar{P}_{j+k,j-2,2}(s) \right] \\ &+ \frac{\mu_2\gamma}{s + \lambda + \mu_1 + (i-j-1)\theta} \left[\sum_{k=1}^{i-j} \left\{ \prod_{p=k-1}^{i-j-2} \left(\frac{\lambda^{i-j-k}\beta^{i-j-(k+1)}}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \eta'_k(s) \right\} \right. \\ &\left. \left\{ \bar{P}_{j+k,j,1,1}(s) + \bar{P}_{j+k,j,1,2}(s) \right\} + \sum_{k=1}^{i-j-1} \left\{ \prod_{p=k-1}^{i-j-2} \left(\frac{(\lambda\beta)^{i-j-k-1}k\theta(1-\alpha)}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \right\} \bar{P}_{j+k+1,j-1,2}(s) \right]; \end{aligned}$$

$$i \geq j+2, j > 1 \quad (24)$$

$$\begin{aligned} \bar{P}_{i,j,1,2}(s) &= \frac{\lambda a_2}{s + \lambda + \mu_2 + (i-j-1)\theta} \bar{P}_{i-1,j,0}(s) + \frac{(i-j)\theta a_2}{s + \lambda + \mu_2 + (i-j-1)\theta} \bar{P}_{i,j,0}(s) + \\ &\frac{\mu_1(1-\gamma)}{s + \lambda + \mu_1 + (i-j-1)\theta} \left[\sum_{k=0}^{i-j} \left\{ \prod_{p=k}^{i-j-1} \left(\frac{\lambda^{i-j-k}\beta^{i-j-(k+1)}}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \psi'_k(s) \right\} \right. \\ &\left. \left\{ \bar{P}_{j+k,j-1,1,1}(s) + \bar{P}_{j+k,j-1,1,2}(s) \right\} + \sum_{k=1}^{i-j} \left\{ \prod_{p=k-1}^{i-j-1} \left(\frac{(\lambda\beta)^{i-j-k}k\theta(1-\alpha)}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \right\} \bar{P}_{j+k,j-2,2}(s) \right] \\ &+ \frac{\mu_1\gamma}{s + \lambda + \mu_1 + (i-j-1)\theta} \left[\sum_{k=1}^{i-j} \left\{ \prod_{p=k-1}^{i-j-2} \left(\frac{\lambda^{i-j-k}\beta^{i-j-(k+1)}}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \eta'_k(s) \right\} \right. \\ &\left. \left\{ \bar{P}_{j+k,j,1,1}(s) + \bar{P}_{j+k,j,1,2}(s) \right\} + \sum_{k=1}^{i-j-1} \left\{ \prod_{p=k-1}^{i-j-2} \left(\frac{(\lambda\beta)^{i-j-k-1}k\theta(1-\alpha)}{s + \lambda\beta + \mu_1 + \mu_2 + p\theta(1-\alpha)} \right) \right\} \bar{P}_{j+k+1,j-1,2}(s) \right]; \end{aligned}$$

$$i \geq j+2, j > 1 \quad (25)$$

where

$$\psi'_k = \begin{cases} 1 & k = 0 \\ 1 + \frac{k\theta\beta}{s + \lambda\beta + \mu_1 + \mu_2 + (k-1)\theta(1-\alpha)} & k = 1 \text{ to } i-j-1 \\ \frac{k\theta}{s + \lambda\beta + \mu_1 + \mu_2 + (k-1)\theta(1-\alpha)} & k = i-j \end{cases}$$

$$\eta'_k = \begin{cases} 1 & k = 1 \\ 1 + \frac{(k-1)\theta\beta}{s + \lambda\beta + \mu_1 + \mu_2 + (k-2)\theta(1-\alpha)} & k = 2 \text{ to } i-j-1 \\ \frac{(k-1)\theta}{s + \lambda\beta + \mu_1 + \mu_2 + (k-2)\theta(1-\alpha)} & k = i-j \end{cases}$$

$$\begin{aligned} \bar{P}_{i,j,0}(s) &= \frac{\mu_1(1-\gamma)}{s + \lambda + (i-j)\theta} \bar{P}_{i,j-1,1,1}(s) + \frac{\mu_1\gamma}{s + \lambda + (i-j)\theta} \bar{P}_{i,j,1,1}(s) + \frac{\mu_2(1-\gamma)}{s + \lambda + (i-j)\theta} \bar{P}_{i,j-1,1,2}(s) \\ &+ \frac{\mu_2\gamma}{s + \lambda + (i-j)\theta} \bar{P}_{i,j,1,2}(s); \quad i > j > 1 \end{aligned} \quad (26)$$

Taking the Laplace inverse of (13)-(26), we get the transient state probabilities as:

$$P_{0,0,0}(t) = e^{-\lambda t} \quad (27)$$

$$P_{i,0,0}(t) = \mu_1\gamma e^{-(\lambda+i\theta)t} * P_{i,0,1,1}(t) + \mu_2\gamma e^{-(\lambda+i\theta)t} * P_{i,0,1,2}(t); \quad i \geq 1 \quad (28)$$

$$P_{1,0,1,1}(t) = \lambda a_1 e^{-(\lambda+\mu_1)t} * P_{0,0,0}(t) + \theta a_1 e^{-(\lambda+\mu_1)t} * P_{1,0,0}(t) \quad (29)$$

$$P_{1,0,1,2}(t) = \lambda a_2 e^{-(\lambda+\mu_2)t} * P_{0,0,0}(t) + \theta a_2 e^{-(\lambda+\mu_2)t} * P_{1,0,0}(t) \quad (30)$$

$$\begin{aligned} P_{i,i,0}(t) &= \left[\lambda a_1 \mu_1 (1-\gamma) e^{-\lambda t} \left\{ \frac{1}{\mu_1} - \frac{e^{-\mu_1 t}}{\mu_1} \right\} + \lambda a_2 \mu_2 (1-\gamma) e^{-\lambda t} \left\{ \frac{1}{\mu_2} - \frac{e^{-\mu_2 t}}{\mu_2} \right\} \right] * P_{i-1,i-1,0}(t) \\ &+ \left[\theta a_1 \mu_1 (1-\gamma) e^{-\lambda t} \left\{ \frac{1}{\mu_1} - \frac{e^{-\mu_1 t}}{\mu_1} \right\} + \theta a_2 \mu_2 (1-\gamma) e^{-\lambda t} \left\{ \frac{1}{\mu_2} - \frac{e^{-\mu_2 t}}{\mu_2} \right\} \right] * P_{i,i-1,0}(t) \\ \mu_1(1-\gamma)\mu_2(1-\gamma)e^{-\lambda t} &\left[\left\{ \frac{1}{\mu_1} - \frac{e^{-\mu_1 t}}{\mu_1} \right\} + \left\{ \frac{1}{\mu_2} - \frac{e^{-\mu_2 t}}{\mu_2} \right\} \right] * P_{i,i-2,2}(t); \quad i \geq 1 \end{aligned} \quad (31)$$

$$P_{i,i-1,1,1}(t) = \lambda a_1 e^{-(\lambda+\mu_1)t} P_{i-1,i-1,0}(t) + \theta a_1 e^{-(\lambda+\mu_1)t} P_{i,i-1,0}(t) + \mu_2(1-\gamma) e^{-(\lambda+\mu_1)t} P_{i,i-2,2}(t); \quad i \geq 2 \quad (32)$$

$$P_{i,i-1,1,2}(t) = \lambda a_2 e^{-(\lambda+\mu_2)t} P_{i-1,i-1,0}(t) + \theta a_2 e^{-(\lambda+\mu_2)t} P_{i,i-1,0}(t) + \mu_1(1-\gamma) e^{-(\lambda+\mu_2)t} P_{i,i-2,2}(t); \quad i \geq 2 \quad (33)$$

$$P_{i,1,1,1}(t) = \lambda a_1 e^{-(\lambda+\mu_1+(i-2)\theta)t} * P_{i-1,1,0}(t) + (i-1)\theta a_1 e^{-(\lambda+\mu_1+(i-2)\theta)t} * P_{i,1,0}(t) + \mu_2(1-\gamma) e^{-(\lambda+\mu_1+(i-2)\theta)t} * P_{i,0,2}(t) + \mu_2\gamma e^{-(\lambda+\mu_1+(i-2)\theta)t} * P_{i,1,2}(t); \quad i \geq 3 \quad (34)$$

$$P_{i,1,1,2}(t) = \lambda a_2 e^{-(\lambda+\mu_2+(i-2)\theta)t} * P_{i-1,1,0}(t) + (i-1)\theta a_2 e^{-(\lambda+\mu_2+(i-2)\theta)t} * P_{i,1,0}(t) + \mu_1(1-\gamma) e^{-(\lambda+\mu_2+(i-2)\theta)t} * P_{i,0,2}(t) + \mu_1\gamma e^{-(\lambda+\mu_2+(i-2)\theta)t} * P_{i,1,2}(t); \quad i \geq 3 \quad (35)$$

$$P_{i,0,2}(t) = \lambda e^{-(\lambda\beta+\mu_1+\mu_2+(i-2)\theta(1-\alpha))t} * \{P_{i-1,0,1,1}(t) + P_{i-1,0,1,2}(t)\} + \lambda\beta e^{-(\lambda\beta+\mu_1+\mu_2+(i-2)\theta(1-\alpha))t} * P_{i-1,0,2}(t) + (i-1)\theta\lambda e^{-(\lambda\beta+\mu_1+\mu_2+(i-2)\theta(1-\alpha))t} * \{P_{i,0,1,1}(t) + P_{i,0,1,2}(t)\}; \quad i \geq 3 \quad (36)$$

$$\begin{aligned} P_{i,j,2}(t) &= \lambda^{i-j-1} \beta^{i-j-2} \prod_{p=0}^{i-j-2} \left\{ e^{-(\lambda\beta+\mu_1+\mu_2+p\theta(1-\alpha))t} \frac{t^p}{p!} \right\} * \{P_{j+1,j,1,1}(t) + P_{j+1,j,1,2}(t)\} \\ &+ \sum_{k=2}^{i-j-1} \lambda^{i-j-k} \beta^{i-j-(k+1)} \prod_{p=k-1}^{i-j-2} \left\{ e^{-(\lambda\beta+\mu_1+\mu_2+p\theta(1-\alpha))t} \frac{t^{p-k+1}}{(p-k+1)!} \right\} * \{P_{j+k,j,1,1}(t) + P_{j+k,j,1,2}(t)\} \\ &+ \sum_{k=2}^{i-j-1} (\lambda\beta)^{i-j-k} (k-1)\theta \prod_{p=k-2}^{i-j-2} \left\{ e^{-(\lambda\beta+\mu_1+\mu_2+p\theta(1-\alpha))t} \frac{t^{p-k+2}}{(p-k+2)!} \right\} * \{P_{j+k,j,1,1}(t) + P_{j+k,j,1,2}(t)\} \\ &+ (i-j-1)\theta e^{-(\lambda\beta+\mu_1+\mu_2+(i-j-2)\theta(1-\alpha))t} * \{P_{i,j,1,1}(t) + P_{i,j,1,2}(t)\}; \quad i \geq j+2, j \geq 1 \end{aligned} \quad (37)$$

$$\begin{aligned}
 P_{i,j,1,1}(t) &= \lambda a_1 e^{-(\lambda+\mu_1+(i-j-1)\theta)t} * P_{i-1,j,0}(t) + (i-j)\theta a_1 e^{-(\lambda+\mu_1+(i-j-1)\theta)t} * P_{i,j,0}(t) + \\
 &\mu_2(1-\gamma)\lambda^{i-j-1}\beta^{i-j-2}e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \sum_{k=1}^{i-j-1} \lambda^{i-j-k}\beta^{i-j-(k+1)} \\
 &\prod_{p=k}^{i-j-1} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+1}} - \right. \\
 &\left. e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^{p-k} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+1}} \right\} * \{P_{j+k,j-1,1,1}(t) + P_{j+k,j-1,1,2}(t)\} + \\
 &\mu_2(1-\gamma)e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \sum_{k=1}^{i-j-1} (\lambda\beta)^{i-j-k} \prod_{p=k}^{i-j-1} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+1}} - \right. \\
 &\left. e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\} * \{P_{j+k,j-1,1,1}(t) + P_{j+k,j-1,1,2}(t)\} + \\
 &\mu_2(1-\gamma)e^{-(\lambda+\mu_1+(i-j-1)\theta)t} (i-j)\theta \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)} - \right. \\
 &\left. e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)t} \frac{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)} \right\} * \{P_{i,j-1,1,1}(t) + P_{i,j-1,1,2}(t)\} + \mu_2(1-\gamma)e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \\
 &\sum_{k=1}^{i-j} (\lambda\beta)^{i-j-k} k\theta(1-\alpha) \prod_{p=k-1}^{i-j-1} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+2}} - e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right. \\
 &\left. \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\} * P_{j+k,j-2,2}(t) + \mu_2\gamma e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \lambda^{i-j}\beta^{i-j-1} \\
 &\prod_{p=0}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p+1}} - e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^p \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-r+1}} \right\} \\
 &* \{P_{j+1,j,1,1}(t) + P_{j+1,j,1,2}(t)\} + \mu_2\gamma e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \sum_{k=2}^{i-j-1} \lambda^{i-j-k}\beta^{i-j-(k+1)} \\
 &\prod_{p=k-1}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+2}} - e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right. \\
 &\left. \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\}
 \end{aligned}$$

$$\begin{aligned}
 & * \left\{ P_{j+k,j,1,1}(t) + P_{j+k,j,1,2}(t) \right\} + \mu_2 \gamma e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \\
 & \sum_{k=2}^{i-j-1} (k-1)\theta(\lambda\beta)^{i-j-k} \prod_{p=k-2}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+3}} - e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right. \\
 & \left. \sum_{r=0}^{p-k+2} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+3}} \right\} * \left\{ P_{j+k,j,1,1}(t) + P_{j+k,j,1,2}(t) \right\} + \mu_2 \gamma e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \\
 & (i-j-1)\theta \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-2)\theta(1-\alpha)}{\beta}\right)} - \frac{e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-2)\theta(1-\alpha)}{\beta}\right)t}}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{(i-j-2)\theta(1-\alpha)}{\beta}\right)} \right\} \\
 & * \left\{ P_{i,j,1,1}(t) + P_{i,j,1,2}(t) \right\} + \mu_2 \gamma e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \sum_{k=2}^{i-j-1} k\theta(1-\alpha)(\lambda\beta)^{i-j-k-1} \\
 & \prod_{p=k-1}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+2}} - e^{-\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right. \\
 & \left. \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_2}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\} * P_{j+k+1,j-1,2}(t); \quad i \geq j+2, j > 1 \quad (38)
 \end{aligned}$$

$$\begin{aligned}
 P_{i,j,1,2}(t) &= \lambda a_2 e^{-(\lambda+\mu_2+(i-j-1)\theta)t} * P_{i-1,j,0}(t) + (i-j)\theta a_2 e^{-(\lambda+\mu_2+(i-j-1)\theta)t} * P_{i,j,0}(t) + \\
 & \mu_1(1-\gamma)\lambda^{i-j-1}\beta^{i-j-2}e^{-(\lambda+\mu_2+(i-j-1)\theta)t} \prod_{p=0}^{i-j-1} \left\{ \frac{1}{\left(\frac{\mu_1}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p+1}} - \right. \\
 & \left. e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^p \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-r+1}} \right\} * \left\{ P_{j,j-1,1,1}(t) + P_{j,j-1,1,2}(t) \right\} \\
 & + \mu_1(1-\gamma)e^{-(\lambda+\mu_2+(i-j-1)\theta)t} \sum_{k=1}^{i-j-1} \lambda^{i-j-k}\beta^{i-j-(k+1)} \prod_{p=k}^{i-j-1} \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+1}} - \right. \\
 & \left. e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^{p-k} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+1}} \right\} * \left\{ P_{j+k,j-1,1,1}(t) + P_{j+k,j-1,1,2}(t) \right\} \\
 & + \mu_1(1-\gamma)e^{-(\lambda+\mu_1+(i-j-1)\theta)t} \sum_{k=1}^{i-j-1} (\lambda\beta)^{i-j-k}
 \end{aligned}$$

$$\left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)^t} - \frac{e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)t}}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{(i-j-1)\theta(1-\alpha)}{\beta}\right)} \right\} * \{P_{i,j-1,1,1}(t) + P_{i,j-1,1,2}(t)\}$$

$$+ \mu_1(1-\gamma)e^{-(\lambda+\mu_2+(i-j-1)\theta)t} \sum_{k=1}^{i-j} (\lambda\beta)^{i-j-k} k\theta(1-\alpha)$$

$$\prod_{p=k-1}^{i-j-1} \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+2}} - e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\}$$

$$* P_{j+k,j-2,2}(t) + \mu_1\gamma e^{-(\lambda+\mu_2+(i-j-1)\theta)t} \lambda^{i-j} \beta^{i-j-1}$$

$$\prod_{p=0}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p+1}} - e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \sum_{r=0}^p \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-r+1}} \right\}$$

$$* \{P_{j+1,j,1,1}(t) + P_{j+1,j,1,2}(t)\} + \mu_1\gamma e^{-(\lambda+\mu_2+(i-j-1)\theta)t} \sum_{k=2}^{i-j-1} \lambda^{i-j-k} \beta^{i-j-(k+1)}$$

$$\prod_{p=k-1}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+2}} - e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right.$$

$$\left. \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\} * \{P_{j+k,j,1,1}(t) + P_{j+k,j,1,2}(t)\}$$

$$+ \mu_1\gamma e^{-(\lambda+\mu_2+(i-j-1)\theta)t} \sum_{k=2}^{i-j-1} (k-1)\theta(\lambda\beta)^{i-j-k}$$

$$\prod_{p=k-2}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+3}} - e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right.$$

$$\left. \sum_{r=0}^{p-k+2} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+3}} \right\} * \{P_{j+k,j,1,1}(t) + P_{j+k,j,1,2}(t)\} + \mu_1\gamma e^{-(\lambda+\mu_2+(i-j-1)\theta)t}$$

$$+ (i-j-1)\theta \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{(i-j-2)\theta(1-\alpha)}{\beta}\right)^t} - \frac{e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{(i-j-2)\theta(1-\alpha)}{\beta}\right)t}}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{(i-j-2)\theta(1-\alpha)}{\beta}\right)} \right\}$$

$$* \{P_{i,j,1,1}(t) + P_{i,j,1,2}(t)\}$$

$$\begin{aligned}
 & + \mu_1 \gamma e^{-(\lambda + \mu_2 + (i-j-1)\theta)t} \sum_{k=2}^{i-j-1} k\theta(1-\alpha)(\lambda\beta)^{i-j-k-1} \\
 & \prod_{p=k-1}^{i-j-2} \left\{ \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k+2}} - e^{-\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)t} \right. \\
 & \left. \sum_{r=0}^{p-k+1} \frac{t^r}{r!} \frac{1}{\left(\frac{\mu_2}{\beta} + \frac{\mu_1}{\beta} + \frac{p\theta(1-\alpha)}{\beta}\right)^{p-k-r+2}} \right\} * P_{j+k+1, j-1, 2}(t); \quad i \geq j+2, j > 1 \quad (39)
 \end{aligned}$$

$$\begin{aligned}
 P_{i,j,0}(t) = & \mu_1(1-\gamma)e^{-(\lambda+(i-j)\theta)t}P_{i,j-1,1,1}(t) + \mu_1\gamma e^{-(\lambda+(i-j)\theta)t}P_{i,j,1,1}(t) + \\
 & \mu_2(1-\gamma)e^{-(\lambda+(i-j)\theta)t}P_{i,j-1,1,2}(t) + \mu_2\gamma e^{-(\lambda+(i-j)\theta)t}P_{i,j,1,2}(t); \quad i > j > 1 \quad (40)
 \end{aligned}$$

4. VERIFICATION OF RESULTS

- Summing equations (13)-(26) over i and j we get:

$$\sum_{i=0}^{\infty} \sum_{j=0}^i \{ \bar{P}_{i,j,0}(s) + \bar{P}_{i,j,1,1}(s) + \bar{P}_{i,j,1,2}(s) + \bar{P}_{i,j,2}(s) \} = \frac{1}{s}$$

and hence

$$\sum_{i=0}^{\infty} \sum_{j=0}^i \{ P_{i,j,0}(t) + P_{i,j,1,1}(t) + P_{i,j,1,2}(t) + P_{i,j,2}(t) \} = 1$$

which is the verification of our results.

- Define $Q_{n,m}(t)$ = Probability that there are n customers in the orbit when m ($m = 0, 1, 2$) servers are busy at time t .

When server is free, it is represented by probability $Q_{n,0}(t)$

$$Q_{n,0}(t) = \sum_{j=0}^{\infty} P_{j+n, j, 0}(t)$$

The number of customers in the orbit in this case are calculated by using the following formula:

n = (number of arrivals - number of departures)

When one server is busy ($m = 1$), it is represented by the probability $Q_{n,m,k}(t)$

$$Q_{n,m,k}(t) = \sum_{j=0}^{\infty} P_{j+n+m, j, m, k}(t); \quad k = 1, 2$$

The number of customers in the orbit in this case are calculated by using the following formula:

n = (number of arrivals - number of departures - m)

When both servers are busy ($m = 2$), it is represented by the probability $Q_{n,m}(t)$

$$Q_{n,m}(t) = \sum_{j=0}^{\infty} P_{j+n+m, j, m}(t)$$

The number of customers in the orbit in this case are calculated by using the following formula:

$$n = (\text{number of arrivals} - \text{number of departures} - m)$$

Using above definitions in equations (1) to (6) and let $\mu_1=\mu_2=1$, $\gamma=0$, $\beta=1$ and $\alpha=1$ and using $Q_{n,1,1} + Q_{n,1,2} = Q_{n,1}$ and let $a_1 = a_2 = \frac{1}{2}$ and adding equations we get:

$$(\lambda + n\theta)Q_{n,0} = Q_{n,1}; \quad n \geq 0 \quad (41)$$

$$(\lambda + n\theta + 1)Q_{n,1} = \lambda Q_{n,0} + (n + 1)\theta Q_{n+1,0} + 2Q_{n,2}; \quad n \geq 0 \quad (42)$$

$$(\lambda + 2)Q_{n,2} = \lambda Q_{n,1} + (n + 1)\theta Q_{n+1,1} + \lambda(1 - \delta_{n,0})Q_{n-1,2}; \quad n \geq 0 \quad (43)$$

which coincides with the results (1)-(3) of [2].

5. NUMERICAL SOLUTION AND GRAPHICAL REPRESENTATION

Using MATLAB programming the following numerical results are generated for the case $\rho=0.8$, $\eta=0.6$, $\gamma=0.7$, $r_1=0.6$ ($r_2=1-r_1$), $a_1=0.5$ ($a_2=1-a_1$), $\beta=0.7$ and $\alpha=0.6$. Observing the tables below for various time instants t it could be concluded that the sum of probabilities approaches to 1.

Table 1: At $t=1$

$P_{0,0,0}$	$P_{1,0,0}$	$P_{1,1,0}$	$P_{1,0,1,1}$	$P_{2,0,1,1}$	$P_{1,0,1,2}$	$P_{2,0,1,2}$	$P_{2,0,2}$	$P_{3,0,2}$	Sum
0.4493	0.0444	0.0232	0.1393	0.0128	0.1525	0.0174	0.1016	0.0222	0.9627

Table 2: At $t=15$

$P_{8,5,0}$	$P_{8,6,0}$	$P_{8,7,0}$	$P_{8,8,0}$	$P_{8,4,1,1}$	$P_{8,5,1,1}$	$P_{8,6,1,1}$	$P_{8,7,1,1}$	$P_{7,4,1,2}$	$P_{7,5,1,2}$	$P_{8,3,1,2}$
0.0192	0.0409	0.0504	0.0272	0.0146	0.0349	0.0502	0.0307	0.0057	0.0057	0.0061

$P_{8,4,1,2}$	$P_{8,5,1,2}$	$P_{8,6,1,2}$	$P_{8,7,1,2}$	$P_{6,2,2}$	$P_{6,3,2}$	$P_{6,4,2}$	$P_{7,2,2}$	$P_{7,3,2}$	$P_{7,4,2}$	$P_{7,5,2}$
0.0221	0.0532	0.0772	0.048	0.0064	0.0111	0.0071	0.0052	0.0137	0.0179	0.0093

$P_{8,2,2}$	$P_{8,3,2}$	$P_{8,4,2}$	$P_{8,5,2}$	$P_{8,6,2}$	Sum
0.0066	0.0288	0.0778	0.1262	0.0963	0.8925

Table 3: $t=25$

$P_{0,0,0}$	$P_{1,0,0}$	$P_{8,5,0}$	$P_{8,6,0}$	$P_{8,7,0}$	$P_{8,8,0}$	$P_{8,6,1,1}$	$P_{8,7,1,1}$	$P_{8,5,1,2}$	$P_{8,6,1,2}$	$P_{8,7,1,2}$
0	0	0.0043	0.0377	0.1635	0.2894	0.0451	0.0966	0.0114	0.07	0.1549

$P_{8,4,2}$	$P_{8,5,2}$	$P_{8,6,2}$	Sum
0.0041	0.0254	0.0843	0.9867

Table 4: At $t=35$

$P_{0,0,0}$	$P_{4,3,0}$	$P_{6,4,0}$	$P_{8,5,0}$	$P_{8,7,0}$	$P_{8,8,0}$	$P_{8,6,1,1}$	$P_{8,7,1,1}$	$P_{8,7,1,2}$	$P_{8,5,2}$	Sum
0	0	0	0.0002	0.1247	0.6225	0.0111	0.0732	0.01188	0.0013	0.9518

The probabilities against time are graphically represented in following figures:

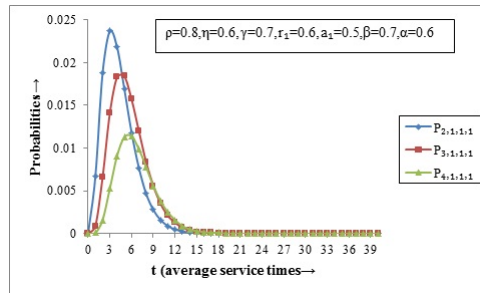


Figure 2

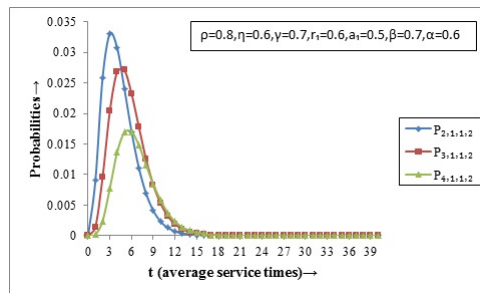


Figure 3

The probabilities $P_{2,1,1}$, $P_{3,1,1}$ and $P_{4,1,1}$ for both the servers 1 and 2 against time t are depicted in figures 2 and 3. From both the figures it is clearly interpreted that probabilities start increasing from 0 at $t=0$ in the beginning and then start decreasing. Also, the curve peaks are higher for lower number of arrivals. If we compare both the graphs, the probabilities are higher for second server than first because of the difference in r_1 and r_2 .

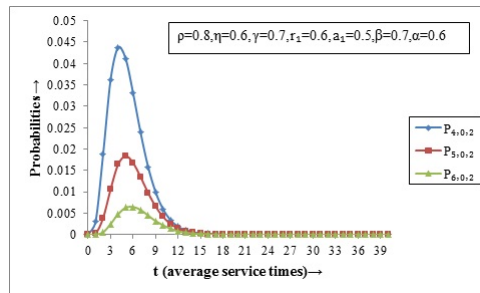


Figure 4

Figure 4 depicts the probabilities $P_{4,0,2}$, $P_{5,0,2}$ and $P_{6,0,2}$ against time t . Beginning with value 0 at $t=0$ the probabilities increases rapidly to their highest values and then decreases gradually. Also, the probabilities are higher for lower number of arrivals when both the servers are busy.

6. BUSY PERIOD PROBABILITIES

The probability that the server is busy is given as:

$$P(\text{Server is busy}) = \sum_{i>j \geq 0} (P_{i,j,1,1}(t) + P_{i,j,1,2}(t) + P_{i,j,2}(t)) \quad (44)$$

The probability that the system is busy is given as:

$$P(\text{System is busy}) = \sum_{i>j \geq 0} (P_{i,j,0}(t) + P_{i,j,1,1}(t) + P_{i,j,1,2}(t) + P_{i,j,2}(t)) \quad (45)$$

6.1. Numerical and Graphical Representation of Busy Period

Following [13] work and using MATLAB programming the numerical results are generated. The probability for system busy and server busy are studied by varying the value of ρ and keeping other parameters as constant ($\eta=0.6, \gamma=0.7, r_1=0.6 (r_2=1-r_1), a_1=0.5 (a_2=1-a_1), \beta=0.7$ and $\alpha=0.6$).

t	Probability(System Busy)			Probability(Server Busy)		
	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$
0	0	0	0	0	0	0
1	0.3122	0.4295	0.5269	0.2772	0.3855	0.4773
2	0.5073	0.6541	0.7571	0.4290	0.5674	0.6713
3	0.6358	0.7801	0.867	0.5286	0.6732	0.771
4	0.7235	0.8542	0.9229	0.5999	0.7413	0.8286
5	0.7848	0.8995	0.9527	0.6532	0.7874	0.864
6	0.8286	0.9281	0.9694	0.6940	0.8196	0.8866
7	0.8606	0.9467	0.979	0.7256	0.8425	0.9009

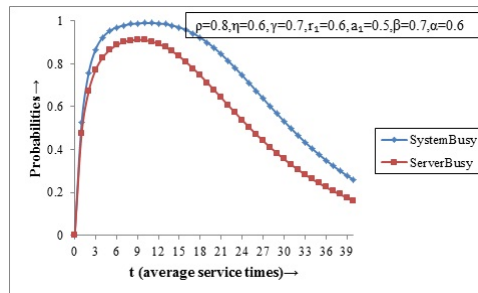


Figure 5

In figure 5 the probabilities for System busy and Server busy are plotted against time t for the case when $\rho=0.8, \eta=0.6, \gamma=0.7, r_1=0.6 (r_2=1-r_1), a_1=0.5 (a_2=1-a_1), \beta=0.7$ and $\alpha=0.6$. It is shown that both the probabilities increases rapidly and then decreases gradually Also, System busy remains higher than Server busy as required.

7. CONCLUSION

We considered a retrial queueing system with non-identical parallel servers having feedback and impatient customers. The transient state probabilities for exact number of arrivals and departures from the system are obtained when both, none or one servers are busy. Various results are also been verified. Numerical and graphical representations are provided in order to study the effect of change of various parameters.



8. ACKNOWLEDGMENT

The authors are highly thankful to the DST for developing computing facilities in the department of Statistics. The authors have made use of these facilities while doing this research work.

REFERENCES

- [1] JW Cohen. Basic problems of telephone traffic and the influence of repeated calls, philips telecom, 1957.
- [2] G I Falin and J G C Templeton. *Retrial queues*. CRC Press, 1997.
- [3] Jesús R. Artalejo and Antonio Gómez-Corral. *Retrial Queueing Systems: A Computational Approach*. Springer-Verlag Berlin Heidelberg, 2008.
- [4] Claude Dennis Pegden and Matthew Rosenshine. Some new results for the m/m/1 queue. *Management Science*, 28(7):821–828, 1982.
- [5] Neelam Singla and Sonia Kalra. Performance analysis of a two-state queueing model with retrials. *Journal of Rajasthan Academy of Physical Sciences*, 17(1):81–100, 2018.
- [6] Dmitry Efrosinin and Janos Sztrik. Performance analysis of a two-server heterogeneous retrial queue with threshold policy. *Quality Technology & Quantitative Management*, 8(3): 211–236, 2011.
- [7] M Seenivasan, G Rajarajan, and M Indumathi. Retrial queueing model with two heterogeneous server using matrix geometric method. In *IOP Conference Series: Materials Science and Engineering*, volume 1070, page 012028. IOP Publishing, 2021.
- [8] Jeong-Sim Kim. Retrial queueing system with collision and impatience. *Communications of the Korean Mathematical Society*, 25(4):647–653, 2010.
- [9] P C Garg and Sanjeev Kumar. A single server retrial queue with impatient customers. *Mathematical Journal of Interdisciplinary Sciences*, 1(1):67–82, 2012.
- [10] B Krishna Kumar and J Raja. On multiserver feedback retrial queues with balking and control retrial rate. *Annals of Operations Research*, 141(1):211–232, 2006.
- [11] Fu-Min Chang, Tzu-Hsin Liu, and Jau-Chuan Ke. On an unreliable-server retrial queue with customer feedback and impatience. *Applied Mathematical Modelling*, 55:171–182, 2018.
- [12] S Vanitha. M/g/1 feedback queue with two stage heterogeneous service and deterministic server vacations. *Int J Appl Eng Res*, 13:15899–15907, 2018.
- [13] Brian D Bunday. *Basic queueing theory*. Arnold Baltimore (Md), 1986.

ANALYSIS OF A NON-IDENTICAL COMPONENT-STRENGTHS SYSTEM BASED ON LOWER RECORD DATA

Amal S. Hassan , Doaa M. Ismail, and Heba F. Nagy 

Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, 12613, Egypt.

amal52_soliman@cu.edu.eg

doaa22ismail@gmail.com

heba_nagy_84@cu.edu.eg

Abstract

In engineering applications and reliability literature, stress-strength models play a crucial role. The goal of this study is to develop more accurate stress-strength models by addressing the reliability estimation in multi-component systems with non-identical component strengths and stress. In the context of lower record values, the system's reliability is assessed using both classical and Bayesian approaches. In classical estimation, the maximum likelihood estimator of the reliability function is constructed, and a simulation study based on measurements of precision is used to assess the behavior of various estimates. The Bayesian estimators of reliability under general entropy, logarithmic and precautionary loss functions are computed. The suggested Bayesian estimates are calculated using the Markov Chain Monte Carlo method through a simulation study because there is no one particular way to do it. We found through simulated research that the accuracy of measurements decreases as the number of records rises. The theoretical results are validated using an example from actual data sets.

Keywords: Exponentiated Pareto model; Lower record data; Bayesian inference; General entropy loss function.

MSC: 62N05; 62D99; 62F15; 62F40; 94A20

1. Introduction

Record values are crucial when collecting observations is challenging or when they are lost during experimental operations. Real-world problems needing destructive stress testing, industrial quality control trials, and statistics on the weather, the economy, and sports all depend critically on record values. Only observations that exceed or fall below the most recent extreme values are recorded in this case. The total number of observations is frequently much lower than the overall sample size and only successive severe items are measured. Suppose that $\{U_i, i=1,2,\dots\}$ is an unlimited sequence of independent and identically distributed (iid). An observation U_i is called a lower record value (LRV) if $U_i < U_j$ for each $i < j$.

The stress-strength (SS) model is a fundamental of reliability testing. When a stress Y surpasses strength X , the SS framework $\mathfrak{R} = P(X > Y)$ indicates the possibility of failure. In other words, the system keeps functioning as long as the stress does not outweigh its ability. Reference [1] was the author who initially presented this idea, and [2] later developed it. Many studies have

addressed inferences based on various methods and distributional presumptions; for some recent works, see [3-10].

To build a system with two or more components, the basic idea of $\mathfrak{R} = P(X > Y)$ can be changed. Reliability in a multi-component SS (MSS) model was first created in [11], who investigated the MSS model under the assumption that c out of t system components, where, $(1 \leq c \leq t)$ components survive a common random stress Y . The MSS is applicable to several fields, including communications, logistics, military, and manufacturing operations. For illustration, if only four of a car's eight cylinders are burning, it might be possible to drive the vehicle. It can therefore be expressed as 4 out of 8: G system.

Assume that if c ($1 \leq c \leq t$) or more of the components cooperate, a system with t similar components will function. In its operational setting, the system is subjected to a stress Y that is a random variable with cumulative distribution function (cdf) $G(y)$. The component strengths, or the minimal stresses necessary to manufacture failure, are iid random variables with cdf $F(x)$, then the reliability of the c-out-of-t system is represented by $\mathfrak{R}_{c,t}$ which is developed in [12], is given by:

$$\begin{aligned} \mathfrak{R}_{c,t} &= P[\text{at least } c \text{ of } (X_1, X_2, \dots, X_t) \text{ exceed } Y] \\ &= \sum_{i=c}^t \binom{t}{i} \int_0^\infty [1 - F(x)]^i F(x)^{t-i} dG(x). \end{aligned}$$

For various SS distributions and sampling procedures, many authors addressed the estimation of reliability in MSS models based on different sampling scheme, for example, see [13-19].

Due to the different topologies of system components, the assumption of comparable strength distributions may not be feasible in many actual circumstances. With systems that have backup components, this is frequently the situation. The strengths of different items, even those constructed of the same material, can vary. For instance, heat treating metals to acquire desired mechanical properties in the field of mechanics can lead to various types of breaking when the metal is quenched or cooled. As a result, the strengths of the components vary. Another example, if two different ropes are used to consolidate a rope, the tensile strengths of both ropes may not be evenly distributed. A model that at least incorporates non-identical random strengths for system components appears to be more realistic, see [20].

Assume a system has t components, of which t_1 are of kind 1, t_2 are of kind 2, ..., and the remainder $t_n = t - \sum_{i=1}^{n-1} t_i$ components are of kind n . Let $F_i(\cdot)$, $i = 1, 2, \dots, n$, be the cdf of the random strengths for components of the i^{th} kind. Assume that Y is a common stress with cdf $Q(\cdot)$ that all components are subjected to. The system will function as long as the c-out-of-the-t components can resist the stress. Reference [21] presented the system reliability $\mathfrak{R}_{c_1, \dots, c_n, t_1, \dots, t_n}$ with non-identical component strengths as follows:

$$\mathfrak{R}_{c_1, \dots, c_n, t_1, \dots, t_n} = \sum_{j_1=c_1}^{t_1} \dots \sum_{j_n=c_n}^{t_n} \left(\prod_{i=1}^n \binom{t_i}{j_i} \right) \int_0^\infty \prod_{i=1}^n (1 - F_i(x))^{j_i} (F_i(x))^{t_i - j_i} dQ(x), \quad (1)$$

where summation ranges over all possible combinations (j_1, j_2, \dots, j_n) with $0 \leq j_i \leq t_i$ for $i = 1, 2, \dots, n$ such $c \leq \sum_{i=1}^n j_i \leq t$. Each c_i indicates the minimal number of components of the i^{th} type that the system needs to function.

Considering the investigation of a system with two different sorts of components, the model (1) can be expressed as follows:

$$\mathfrak{R}_{c_1, c_2, t_1, t_2} = \sum_{j_1=c_1}^{t_1} \sum_{j_2=c_2}^{t_2} \binom{t_1}{j_1} \binom{t_2}{j_2} \int_0^{\infty} (1 - F_1(x))^{j_1} (F_1(x))^{t_1-j_1} (1 - F_2(x))^{j_2} (F_2(x))^{t_2-j_2} dQ(x). \quad (2)$$

In order to construct more realistic models, The Bayesian estimation of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ assuming the Weibull and exponential distributions on the strength and stress variates, respectively, was taken into consideration in [22] and [23]. The exponentiated Pareto distribution (EPD) was used to estimate $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ for non-identical MSS in [24]. Recently, [25] studied the case of non-identical component-strengths from the family of Kumaraswamy generalized distributions under upper record data. Reference [26] examined the estimation of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ when component strengths and stress follow inverse Lomax distribution based on complete sample. Reference [27] proposed the estimation of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ when component strengths and stress follow Weibull distributions under generalized progressive hybrid censoring scheme.

It is important to note that the majority of the work on the estimate of the SS reliability conducted to date requires to employ complete or censored samples and that record values are rarely used. Particularly in the estimation of MSS systems of non-identical component-strengths, we are interested in developing MSS models within the record scheme in the case of non-identical component-strengths, where component strengths and stress follow an EPD. A maximum likelihood estimator (MLE) of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ is derived under LRV and a simulation study is investigated. The general entropy loss function (GELF), the logarithmic loss function (LLF), and the precautionary loss function (PLF) are used to derive the Bayesian estimator of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$. Since these estimators are incapable of being reduced to simple closed forms, we use the Markov Chain Monte Carlo (MCMCO) approach for Bayesian estimates of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$. To show the relevance of our work, we also examined real data sets.

The following is how the rest of the essay is presented. The formulation of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$, and its MLE under LRV along with a numerical analysis is provided in Section 2. Section 3 discusses the Bayesian estimators of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$, through GELF, LLF, and PLF. The MCMCO technique is presented in Section 4. For the purposes of illustration, Section 5 includes real data sets. Final remarks are included in Section 6.

2. Model Description and Classical Estimation of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$

In this section, a model description of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ is provided. The MLE of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ is obtained in the presence of LRV. Numerical analysis is also carried out.

2.1. Expression of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$

Here, expression of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ is provided when the strength and stress random variables follow the EPD.

The EPD may be successfully used to assess numerous lifetime data sets, as argued in [28]. The EPD has a very flexible structure as a result of its decreasing or upside-down bathtub shape failure rates depending on shape parameters. This property provides advantages for modeling extreme events, particularly in hydrology. Furthermore, the EPD is a reasonable equivalent to the exponential distribution because of its heavier or lighter tail features. A variety of lifetime data might seem nice in the EPD. The EPD 's cdf is expressed by

$$F(x) = \left[1 - (1 + x)^{-\gamma} \right]^{\delta}, \quad x > 0, \gamma > 0, \delta > 0,$$

where δ and γ are the shape parameters. The associated probability density function (pdf) is given by:

$$f(x) = \delta\gamma \left[1 - (1+x)^{-\gamma} \right]^{\delta-1} (1+x)^{-(\gamma+1)}, \quad x > 0, \gamma > 0, \delta > 0.$$

Several scholars addressed the EP's research and applications, for instance see [29].

From the total of t system components in the model (2), let the first t_1 of first kind component strengths follow EPD (γ, δ_1) , while the remaining $t_2 = t - t_1$ of kind 2 component strengths follow EPD (γ, δ_2) . Also, suppose that Y follows EPD (γ, δ_3) independently. The respective distribution functions are as below:

$$F_i(x) = \left[1 - (1+x)^{-\gamma} \right]^{\delta_i}, \quad x, \gamma, \delta_i > 0, i = 1, 2, \tag{3}$$

$$Q(y) = \left[1 - (1+y)^{-\gamma} \right]^{\delta_3}, \quad y, \gamma, \delta_3 > 0. \tag{4}$$

By replacing F_1, F_2, Q given in (2) by (3) and (4), the formula of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ for such a system is as follows:

$$\begin{aligned} \mathfrak{R}_{c_1, c_2, t_1, t_2} &= \sum_{j_1=c_1}^{t_1} \sum_{j_2=c_2}^{t_2} \binom{t_1}{j_1} \binom{t_2}{j_2} \int_0^\infty (1 - (1 - (1+x)^{-\gamma})^{\delta_1})^{j_1} (1 - (1+x)^{-\gamma})^{\delta_1(t_1-j_1)} (1 - (1 - (1+x)^{-\gamma})^{\delta_2})^{j_2} \\ &\quad \times (1 - (1+x)^{-\gamma})^{\delta_2(t_2-j_2)} \delta_3 \gamma \left[1 - (1+x)^{-\gamma} \right]^{\delta_3-1} (1+x)^{-(\gamma+1)} dx. \end{aligned}$$

Let $z = 1 - (1+x)^{-\gamma}$, $dz = \gamma(1+x)^{-\gamma-1}$, then $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ is as follows

$$\mathfrak{R}_{c_1, c_2, t_1, t_2} = \sum_{j_1=c_1}^{t_1} \sum_{j_2=c_2}^{t_2} \binom{t_1}{j_1} \binom{t_2}{j_2} \delta_3 \int_0^1 (1-z^{\delta_1})^{j_1} (1-z^{\delta_2})^{j_2} z^{\delta_1(t_1-j_1) + \delta_2(t_2-j_2) + \delta_3-1} dz.$$

Using the binomial expansion for $(1-z^{\delta_1})^{j_1}$ and $(1-z^{\delta_2})^{j_2}$ leads to the following

$$\begin{aligned} \mathfrak{R}_{c_1, c_2, t_1, t_2} &= E_{j_1, j_2, m, n} \delta_3 \int_0^1 z^{\delta_1(m+t_1-j_1) + \delta_2(n+t_2-j_2) + \delta_3-1} dz \\ &= \frac{E_{j_1, j_2, m, n} \delta_3}{\delta_1(m+t_1-j_1) + \delta_2(n+t_2-j_2) + \delta_3}, \end{aligned} \tag{5}$$

where $E_{j_1, j_2, m, n} = \sum_{j_1=c_1}^{t_1} \sum_{j_2=c_2}^{t_2} \sum_{m=0}^{j_1} \sum_{n=0}^{j_2} \binom{t_1}{j_1} \binom{t_2}{j_2} \binom{j_1}{m} \binom{j_2}{n} (-1)^{m+n}$.

Note that expression (5) depends on the parameters δ_1, δ_2 and δ_3 .

2.2. The MLE of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ under LRV

In order to find the MLE of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ based on LRV, we first need to obtain the MLE of the parameters δ_1, δ_2 and δ_3 assuming γ is given known. So, let $\underline{r} = \{r_1, \dots, r_n\}$, $\underline{p} = \{p_1, \dots, p_m\}$ and $\underline{s} = \{s_1, \dots, s_w\}$ be three independent sets of LRV of sizes n, m , and w from EPD (γ, δ_1) , EPD (γ, δ_2) and EPD (γ, δ_3) respectively. The likelihood function of the observed records, according to [30], is defined by:

$$L(\underline{u} | \eta) = h(u_d) \prod_{i=1}^{d-1} \frac{h(u_i; \eta)}{H(u_i; \eta)}, \quad -\infty < u_d < \dots < u_1 < \infty. \tag{6}$$

Hence, the observed LRV data $\underline{r}, \underline{p}$ and \underline{s} , given η , based on (6), are given as below:

$$\begin{aligned}
 L(\underline{r}, \underline{p}, \underline{s} | \eta) &= L_1(\underline{r} | \gamma, \delta_1) L_2(\underline{p} | \gamma, \delta_2) L_3(\underline{s} | \gamma, \delta_3) \\
 &= \delta_1^n \delta_2^m \delta_3^w (\gamma)^{n+m+w} (1+r_n)^{-(\gamma+1)} (1+p_m)^{-(\gamma+1)} (1+s_w)^{-(\gamma+1)} [\tau_n]^{\delta_1-1} \\
 &\quad \times [\varphi_m]^{\delta_2-1} [\zeta_w]^{\delta_3-1} \prod_{i=1}^{n-1} (1+r_i)^{-(\gamma+1)} (\tau_i)^{-1} \prod_{j=1}^{m-1} (1+p_j)^{-(\gamma+1)} (\varphi_j)^{-1} \\
 &\quad \times \prod_{u=1}^{w-1} (1+s_u)^{-(\gamma+1)} (\zeta_u)^{-1},
 \end{aligned} \tag{7}$$

where $\tau_i = 1 - (1+r_i)^{-\gamma}$, $\varphi_j = 1 - (1+p_j)^{-\gamma}$, and $\zeta_u = 1 - (1+s_u)^{-\gamma}$, $i = 1, \dots, n$, $j = 1, \dots, m$, $u = 1, \dots, w$.

Consequently, the joint log-likelihood function, denoted by $\ln \ell$, is derived as:

$$\begin{aligned}
 \ln \ell &= n \ln(\delta_1) + m \ln(\delta_2) + w \ln(\delta_3) + (n+m+w) \ln(\gamma) - (\gamma+1) \ln(1+r_n) + \ln(1+p_m) + \ln(1+s_w) \\
 &\quad + (\delta_1 - 1) \ln[\tau_n] + (\delta_2 - 1) \ln[\varphi_m] + (\delta_3 - 1) \ln[\zeta_w] - \sum_{i=1}^{n-1} [(\gamma+1) \ln(1+r_i) + \ln(\tau_i)] \\
 &\quad - \sum_{j=1}^{m-1} [(\gamma+1) \ln(1+p_j) + \ln(\varphi_j)] - \sum_{u=1}^{w-1} [(\gamma+1) \ln(1+s_u) + \ln(\zeta_u)].
 \end{aligned}$$

Given that γ is a known, the following are the partial derivatives of $\ln \ell$ with respect to δ_1, δ_2 and δ_3 respectively

$$\frac{\partial \ln \ell}{\partial \delta_1} = \frac{n}{\delta_1} + \ln[\tau_n], \quad \frac{\partial \ln \ell}{\partial \delta_2} = \frac{m}{\delta_2} + \ln[\varphi_m], \quad \frac{\partial \ln \ell}{\partial \delta_3} = \frac{w}{\delta_3} + \ln[\zeta_w].$$

Then, the MLEs of δ_1, δ_2 and δ_3 denoted by $\hat{\delta}_1, \hat{\delta}_2$ and $\hat{\delta}_3$ are obtained by setting $\partial \ln \ell / \partial \delta_1, \partial \ln \ell / \partial \delta_2$ and $\partial \ln \ell / \partial \delta_3$ to be zero. Hence $\hat{\delta}_1, \hat{\delta}_2$ and $\hat{\delta}_3$ are obtained as

$$\hat{\delta}_1 = \frac{-n}{\ln[\tau_n]}, \quad \hat{\delta}_2 = \frac{-m}{\ln[\varphi_m]}, \quad \hat{\delta}_3 = \frac{-w}{\ln[\zeta_w]}. \tag{8}$$

Therefore, based on invariance property, we obtain the MLE of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ by inserting $\hat{\delta}_1, \hat{\delta}_2$ and $\hat{\delta}_3$ in (5) as follows

$$\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2} = \frac{E_{j_1, j_2, m, n} \hat{\delta}_3}{\hat{\delta}_1 (m + t_1 - j_1) + \hat{\delta}_2 (n + t_2 - j_2) + \hat{\delta}_3}. \tag{9}$$

2.3. Numerical Study

The MLE for the MSS variables is thoroughly numerically analysed in this subsection. In order to assess the accuracy of estimates for various parameter values and record numbers, two criteria are used: absolute biases (ABs) and mean squared errors (MSEs). The numerical research is performed in the following way:

- Create LRV samples based on the parameter values provided.
- The parameters values are selected as $(\delta_1, \delta_2, \delta_3) = (0.5, 0.5, 0.2), (1.1, 0.5, 0.2), (0.5, 0.5, 0.5)$ and $(1.1, 1, 0.2)$ for $\gamma = 1$ in all situations. The specified values for c-out-of-t systems are $(c_1, c_2, t_1, t_2) = (1, 1, 2, 2), (1, 2, 2, 2), (2, 1, 2, 2)$ and $(2, 2, 2, 2)$.
- The true values at $(c_1, c_2, t_1, t_2) = (1, 1, 2, 2)$ are 0.533, 0.871, 0.758 and 0.809, at $(c_1, c_2, t_1, t_2) = (1, 2, 2, 2)$ are 0.3, 0.746, 0.573 and 0.591, at $(c_1, c_2, t_1, t_2) = (2, 1, 2, 2)$ are 0.301, 0.760, 0.573 and 0.724, and at $(c_1, c_2, t_1, t_2) = (2, 2, 2, 2)$ are 0.2, 0.687, 0.477 and 0.562.
- The sample sizes of LRV samples (n, m, w) are selected to be $(2, 2, 2), (5, 5, 5), (7, 7, 7), (10, 10, 10), (2, 2, 3), (5, 5, 6), (7, 7, 8)$ and $(10, 10, 11)$.
- 5000 repetitions are used to evaluate the ABs and MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$.

- The simulated outcomes are shown in Table 1 and are illustrated in Figures 1–6.
- As the number of records increases, the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for all values of (c_1, c_2, t_1, t_2) decrease (Figure 1). For all true value of parameters, the MSE of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ decreases at $(c_1, c_2, t_1, t_2) = (2,2,2,2)$ when the number of records $n = m$ (Figure 2).

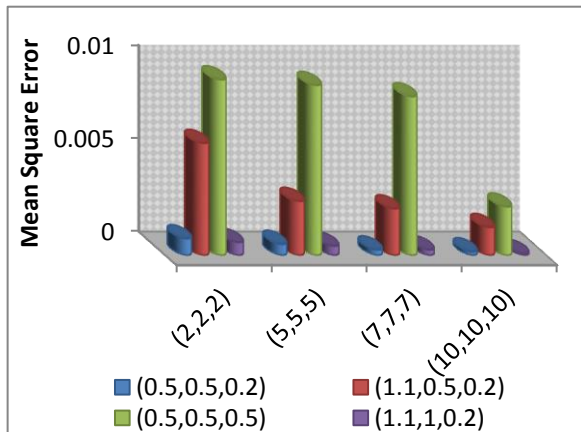


Figure 1: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for different $(\delta_1, \delta_2, \delta_3)$ values at $(c_1, c_2, t_1, t_2) = (1, 1, 2, 2)$ and $n = m = w$

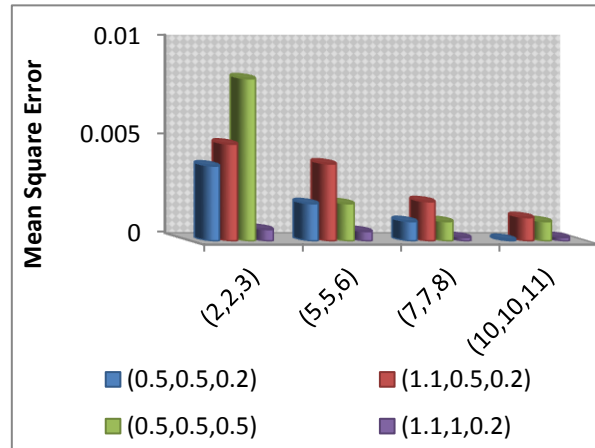


Figure 2: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for different $(\delta_1, \delta_2, \delta_3)$ values at $(c_1, c_2, t_1, t_2) = (2, 2, 2, 2)$ and $n = m$

- Figure 3 demonstrates that as the number of n, m and w increases, the ABs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for all actual values of $(\delta_1, \delta_2, \delta_3)$ are decreasing.
- Figure 4 illustrates that the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at $(c_1, c_2, t_1, t_2) = (2, 2, 2, 2)$ are larger than the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for others values of (c_1, c_2, t_1, t_2) .

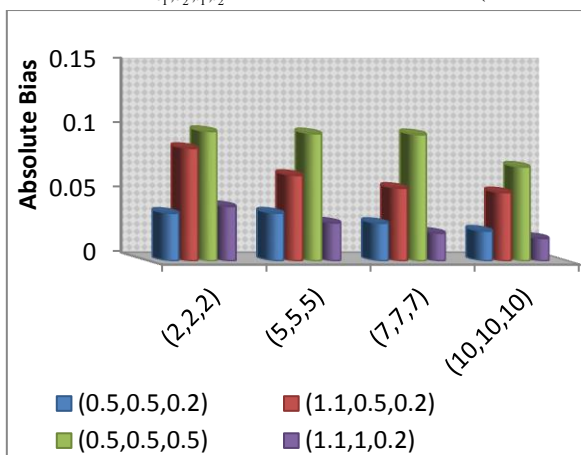


Figure 3: ABs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for different $(\delta_1, \delta_2, \delta_3)$ values at $(c_1, c_2, t_1, t_2) = (1, 2, 2, 2)$ and $n = m = w$

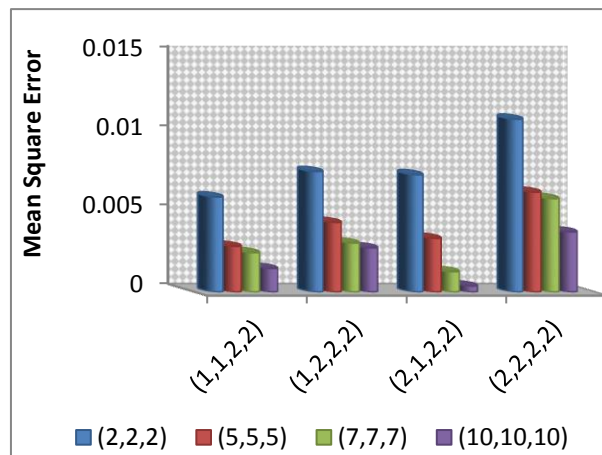


Figure 4: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for different (c_1, c_2, t_1, t_2) values at $n = m = w$ and $(\delta_1, \delta_2, \delta_3) = (1.1, 0.5, 0.2)$

- Figure 5 illustrates that the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at $(c_1, c_2, t_1, t_2) = (1, 1, 2, 2)$ are smaller than the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for others values of (c_1, c_2, t_1, t_2) .
- Figure 6 illustrates that the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ decrease when the true value of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ increases.

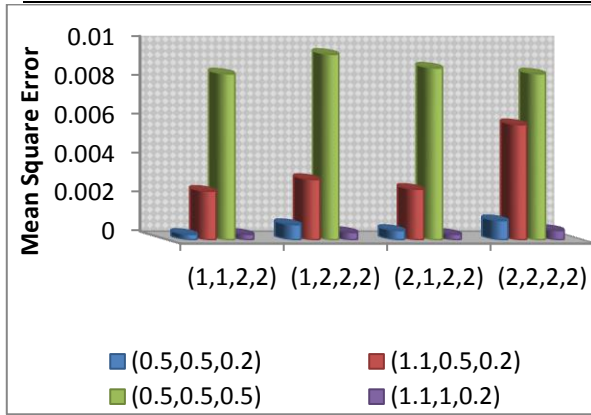


Figure 5: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for different $(\delta_1, \delta_2, \delta_3)$ and (c_1, c_2, t_1, t_2) values at $n = m = w = 7$

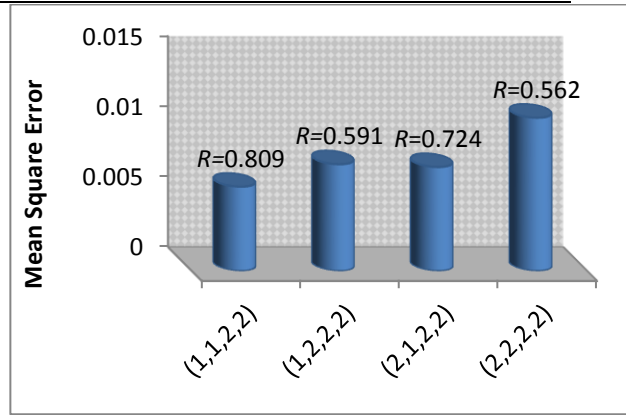


Figure 6: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for $(\delta_1, \delta_2, \delta_3) = (1.1, 0.5, 0.2)$ at $n = m = w = 2$

Table1: Numerical results of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for different values of $(\delta_1, \delta_2, \delta_3)$

$(\delta_1, \delta_2, \delta_3) = (0.5, 0.5, 0.2)$					$(\delta_1, \delta_2, \delta_3) = (1.1, 0.5, 0.2)$				
(c_1, c_2, t_1, t_2)	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE	(c_1, c_2, t_1, t_2)	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE
(1,1,2,2)	0.758	(2,2,2)	0.0303	0.0009	(1,1,2,2)	0.809	(2,2,2)	0.0772	0.0060
		(5,5,5)	0.0255	0.0006			(5,5,5)	0.0546	0.0029
		(7,7,7)	0.0197	0.0003			(7,7,7)	0.0506	0.0025
		(10,10,10)	0.0182	0.0002			(10,10,10)	0.0394	0.0015
		(2,2,3)	0.0244	0.0005			(2,2,3)	0.0284	0.0008
		(5,5,6)	0.0137	0.0002			(5,5,6)	0.0282	0.0007
		(7,7,8)	0.0075	0.0001			(7,7,8)	0.0205	0.0004
		(10,10,11)	0.0066	0.0001			(10,10,11)	0.0212	0.0004
(1,2,2,2)	0.573	(2,2,2)	0.0371	0.0013	(1,2,2,2)	0.591	(2,2,2)	0.0872	0.0076
		(5,5,5)	0.0372	0.0013			(5,5,5)	0.0664	0.0044
		(7,7,7)	0.0293	0.0008			(7,7,7)	0.0564	0.0031
		(10,10,10)	0.0233	0.0004			(10,10,10)	0.0530	0.0028
		(2,2,3)	0.0345	0.0011			(2,2,3)	0.0628	0.0039
		(5,5,6)	0.0307	0.0009			(5,5,6)	0.0484	0.0023
		(7,7,8)	0.0281	0.0007			(7,7,8)	0.0302	0.0009
		(10,10,11)	0.0118	0.0001			(10,10,11)	0.0305	0.0009
(2,1,2,2)	0.573	(2,2,2)	0.0306	0.0009	(2,1,2,2)	0.724	(2,2,2)	0.0863	0.0074
		(5,5,5)	0.0276	0.0007			(5,5,5)	0.0583	0.0034
		(7,7,7)	0.0238	0.0005			(7,7,7)	0.0364	0.0026
		(10,10,10)	0.0210	0.0003			(10,10,10)	0.0202	0.0004
		(2,2,3)	0.0233	0.0005			(2,2,3)	0.0303	0.0009
		(5,5,6)	0.0162	0.0003			(5,5,6)	0.0285	0.0008
		(7,7,8)	0.0128	0.0002			(7,7,8)	0.0272	0.0007
		(10,10,11)	0.0115	0.0001			(10,10,11)	0.0264	0.0007
(2,2,2,2)	0.477	(2,2,2)	0.0978	0.0095	(2,2,2,2)	0.562	(2,2,2)	0.1046	0.0109
		(5,5,5)	0.0530	0.0028			(5,5,5)	0.0794	0.0063
		(7,7,7)	0.0317	0.0010			(7,7,7)	0.0770	0.0059
		(10,10,10)	0.0305	0.0008			(10,10,10)	0.0619	0.0038
		(2,2,3)	0.0622	0.0038			(2,2,3)	0.0705	0.0049
		(5,5,6)	0.0437	0.0019			(5,5,6)	0.0629	0.0039
		(7,7,8)	0.0323	0.0010			(7,7,8)	0.0450	0.0020
		(10,10,11)	0.0211	0.0001			(10,10,11)	0.0346	0.0012

$(\delta_1, \delta_2, \delta_3) = (0.5, 0.5, 0.5)$					$(\delta_1, \delta_2, \delta_3) = (1.1, 1, 0.2)$				
(c_1, c_2, t_1, t_2)	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE	(c_1, c_2, t_1, t_2)	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE
(1,1,2,2)	0.533	(2,2,2)	0.0972	0.0094	(1,1,2,2)	0.871	(2,2,2)	0.0267	0.0007
		(5,5,5)	0.0958	0.0091			(5,5,5)	0.0222	0.0005
		(7,7,7)	0.0922	0.0085			(7,7,7)	0.0190	0.0003
		(10,10,10)	0.0515	0.0026			(10,10,10)	0.0183	0.0001
		(2,2,3)	0.0707	0.0050			(2,2,3)	0.0173	0.0003
		(5,5,6)	0.0628	0.0039			(5,5,6)	0.0130	0.0002
		(7,7,8)	0.0395	0.0015			(7,7,8)	0.0122	0.0001
		(10,10,11)	0.0251	0.0010			(10,10,11)	0.0113	0.0001
(1,2,2,2)	0.3	(2,2,2)	0.1002	0.0100	(1,2,2,2)	0.746	(2,2,2)	0.0423	0.0017
		(5,5,5)	0.0984	0.0096			(5,5,5)	0.0296	0.0008
		(7,7,7)	0.0977	0.0095			(7,7,7)	0.0217	0.0004
		(10,10,10)	0.0728	0.0053			(10,10,10)	0.0176	0.0002
		(2,2,3)	0.0717	0.0049			(2,2,3)	0.0228	0.0005
		(5,5,6)	0.0684	0.0046			(5,5,6)	0.0216	0.0004
		(7,7,8)	0.0558	0.0031			(7,7,8)	0.0128	0.0001
		(10,10,11)	0.0301	0.0018			(10,10,11)	0.0121	0.0001
(2,1,2,2)	0.301	(2,2,2)	0.0957	0.0091	(2,1,2,2)	0.760	(2,2,2)	0.0396	0.0015
		(5,5,5)	0.0954	0.0091			(5,5,5)	0.0288	0.0007
		(7,7,7)	0.0947	0.0088			(7,7,7)	0.0215	0.0003
		(10,10,10)	0.0701	0.0049			(10,10,10)	0.0170	0.0001
		(2,2,3)	0.0720	0.0051			(2,2,3)	0.0222	0.0004
		(5,5,6)	0.0636	0.0040			(5,5,6)	0.0211	0.0002
		(7,7,8)	0.0422	0.0017			(7,7,8)	0.0121	0.0001
		(10,10,11)	0.0288	0.0011			(10,10,11)	0.0118	0.0001
(2,2,2,2)	0.2	(2,2,2)	0.1028	0.0105	(2,2,2,2)	0.687	(2,2,2)	0.0430	0.0020
		(5,5,5)	0.0530	0.0098			(5,5,5)	0.0317	0.0009
		(7,7,7)	0.0317	0.0085			(7,7,7)	0.0220	0.0005
		(10,10,10)	0.0301	0.0050			(10,10,10)	0.0178	0.0003
		(2,2,3)	0.0906	0.0082			(2,2,3)	0.0230	0.0006
		(5,5,6)	0.0437	0.0019			(5,5,6)	0.0235	0.0005
		(7,7,8)	0.0323	0.0010			(7,7,8)	0.0130	0.0002
		(10,10,11)	0.0299	0.0010			(10,10,11)	0.0128	0.0002

3. Bayesian Estimation of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$

We will look in this section at the Bayesian estimator of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ under the assumption that δ_1, δ_2 and δ_3 are random variables.

Following [31], the prior distributions for δ_1, δ_2 and δ_3 are assumed to have the gamma distribution with the following pdfs

$$\pi_1(\delta_1) \propto \delta_1^{a_1-1} e^{-b_1 \delta_1}, \quad \pi_2(\delta_2) \propto \delta_2^{a_2-1} e^{-b_2 \delta_2}, \quad \text{and} \quad \pi_3(\delta_3) \propto \delta_3^{a_3-1} e^{-b_3 \delta_3},$$

where, the hyper-parameters; a_1, a_2, a_3, b_1, b_2 and b_3 are considered to be known. The joint prior distribution of $\eta = (\delta_1, \delta_2, \delta_3)$, assuming parameters independence is as follows:

$$\pi(\eta) = \delta_1^{a_1-1} \delta_2^{a_2-1} \delta_3^{a_3-1} e^{-(b_1 \delta_1 + b_2 \delta_2 + b_3 \delta_3)}.$$

Based on the observed samples, the joint density function of $\eta = (\delta_1, \delta_2, \delta_3)$ and the data are:

$$\begin{aligned} \pi(\eta | \underline{r}, \underline{p}, \underline{s}) &= \delta_1^{n+a_1-1} \delta_2^{m+a_2-1} \delta_3^{w+a_3-1} e^{-(b_1\delta_1+b_2\delta_2+b_3\delta_3)} \\ &\times (\gamma)^{n+m+w} (1+r_n)^{-(\gamma+1)} (1+p_m)^{-(\gamma+1)} (1+s_w)^{-(\gamma+1)} [\tau_n]^{\delta_1-1} \\ &\times [\varphi_m]^{\delta_2-1} [\zeta_w]^{\delta_3-1} \prod_{i=1}^{n-1} \frac{(1+r_i)^{-(\gamma+1)}}{(\tau_i)} \prod_{j=1}^{m-1} \frac{(1+p_j)^{-(\gamma+1)}}{(\varphi_j)} \prod_{u=1}^{w-1} \frac{(1+s_u)^{-(\gamma+1)}}{(\zeta_u)}. \end{aligned}$$

As a result, the posterior density function of $\eta = (\delta_1, \delta_2, \delta_3)$ can be expressed as

$$\pi^*(\eta | \underline{r}, \underline{p}, \underline{s}) = \frac{L(\underline{r}, \underline{p}, \underline{s} | \eta) \pi(\eta)}{\int_0^\infty \int_0^\infty \int_0^\infty L(\underline{r}, \underline{p}, \underline{s} | \eta) \pi(\eta) d\delta_1 d\delta_2 d\delta_3}.$$

The Bayesian estimator of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$, based on GELF, indicated by $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ is derived as follows:

$$\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2} = [E(\mathfrak{R}_{c_1, c_2, t_1, t_2})^{-\frac{1}{\xi}}]^{-\xi} = \left[\int_0^\infty \int_0^\infty \int_0^\infty (\mathfrak{R}_{c_1, c_2, t_1, t_2})^{-\frac{1}{\xi}} \pi^*(\eta | \underline{r}, \underline{p}, \underline{s}) d\delta_1 d\delta_2 d\delta_3 \right]^{-\xi}. \quad (10)$$

Additionally, the Bayesian estimator of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$, under LLF indicated by $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ is as follows:

$$\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2} = \exp(E(\log \mathfrak{R}_{c_1, c_2, t_1, t_2})) = \exp \left[\int_0^\infty \int_0^\infty \int_0^\infty \log \mathfrak{R}_{c_1, c_2, t_1, t_2} \pi^*(\eta | \underline{r}, \underline{p}, \underline{s}) d\delta_1 d\delta_2 d\delta_3 \right]. \quad (11)$$

The Bayesian estimator of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ for PLF indicated by $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ is as follows:

$$\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2} = \sqrt{E(\mathfrak{R}_{c_1, c_2, t_1, t_2}^2)} = \left[\int_0^\infty \int_0^\infty \int_0^\infty \mathfrak{R}_{c_1, c_2, t_1, t_2}^2 \pi^*(\eta | \underline{r}, \underline{p}, \underline{s}) d\delta_1 d\delta_2 d\delta_3 \right]^{0.5}. \quad (12)$$

It is difficult to find an explicit formula for (10)–(12) because the posterior density function $\pi^*(\eta | \underline{r}, \underline{p}, \underline{s})$ has a composite structure. In order to obtain Bayesian estimates, we calculate these integrations using the Metropolis-Hastings (M-H) technique using the MCMCO algorithm.

4. MCMCO Methodology

The MCMCO simulation is used to investigate the behavior $\mathfrak{R}_{c_1, c_2, t_1, t_2}$'s MSS. Bayes estimates (BE) under different loss functions are produced using gamma priors. The $\mathfrak{R}_{c_1, c_2, t_1, t_2}$'s BE accuracy is measured using the ABs, and MSEs. The various LRV options are $(n, m, w) = (2, 2, 2), (5, 5, 5), (7, 7, 7), (10, 10, 10), (2, 2, 3), (5, 5, 6),$ and $(7, 7, 8)$. The possible sets of hyperparameter values are considered to be: Prior I: $(2, 1.5, 3, 2, 1.5, 1.1)$ and Prior II: $(1, 1.4, 1, 2, 2.5, 3)$.

The outcomes are based on 5,000 replications. The M-H process is a popular subgroup of the MCMCO technique in the Bayesian literature for modeling departures from the posterior density and producing accurate anticipated results. The main difficulty with the MCMCO is getting the BEs of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ from GELF, LLF, and PLF using the M-H approach after simulating samples from the posterior density. It converges to the desired distribution using acceptance/rejection criteria. The M-H algorithm (see [32]) operates as follows:

- Set the starting parameter value of $\mathfrak{R}_{c_1, c_2, t_1, t_2}^0$ and the sample number N.
- For $i = 2$ to N, set $\mathfrak{R}_{c_1, c_2, t_1, t_2}^i = \mathfrak{R}_{c_1, c_2, t_1, t_2}^{i-1}$.
- Create u using the uniform (0,1).
- Choose a candidate parameter $\mathfrak{R}_{c_1, c_2, t_1, t_2}^*$ from the proposal density.

- e) If $u \leq \frac{\pi(\theta^*)g(\theta|\theta^*)}{\pi(\theta)g(\theta|\theta^*)}$, then set $\mathfrak{R}_{c_1, c_2, t_1, t_2}^i = \mathfrak{R}_{c_1, c_2, t_1, t_2}^*$; otherwise, set $\mathfrak{R}_{c_1, c_2, t_1, t_2}^i = \mathfrak{R}_{c_1, c_2, t_1, t_2}$.
- f) Return to step (b) and perform the aforementioned actions N times using $i = i + 1$.

Using the outputs of the study, which are shown in Tables 2, 3, and are illustrated by Figures 7–12, we come up with the following conclusions:

- The MSEs and ABs of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ estimates via the GELF, LLF and PLF decrease with increasing the record numbers n, m, w rises for all true values of (c_1, c_2, t_1, t_2) , (Tables 2, 3).
- The ABs of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ estimates via the GELF, LLF and PLF have the smallest values at $(c_1, c_2, t_1, t_2) = (1, 1, 2, 2)$, (Tables 2, 3).
- At true value $\mathfrak{R}_{c_1, c_2, t_1, t_2} = 0.748$, the MSE of $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ via PLF take the smallest values in case of prior I except at (7, 7, 8) (see Figure 7).
- At true value $\mathfrak{R}_{c_1, c_2, t_1, t_2} = 0.748$, the AB of $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at PLF gets the fewest values for a distinct number of records excepting at $(n, m, w) = (7, 7, 8)$ via prior I (see Figure 8).

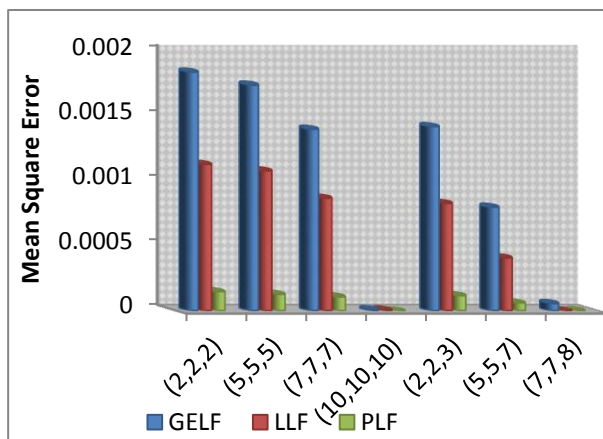


Figure 7: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at $(c_1, c_2, t_1, t_2) = (1, 2, 2, 2)$ for prior I

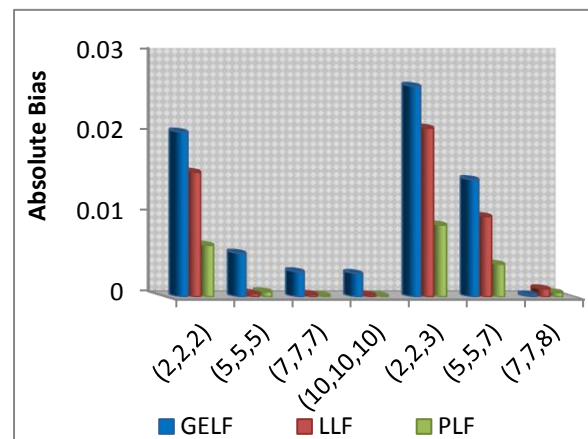


Figure 8: ABs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at $(c_1, c_2, t_1, t_2) = (1, 1, 2, 2)$ for prior I

- The MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ under the GELF, LLF and PLF, respectively, decrease as the number of records $n = m = w$ increases via prior II (see Figure 9).
- At true value $\mathfrak{R}_{c_1, c_2, t_1, t_2} = 0.760$, the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ under the GELF, LLF and PLF, respectively, get the least values for similar record values of (n, m) via prior II (see Figure 10).

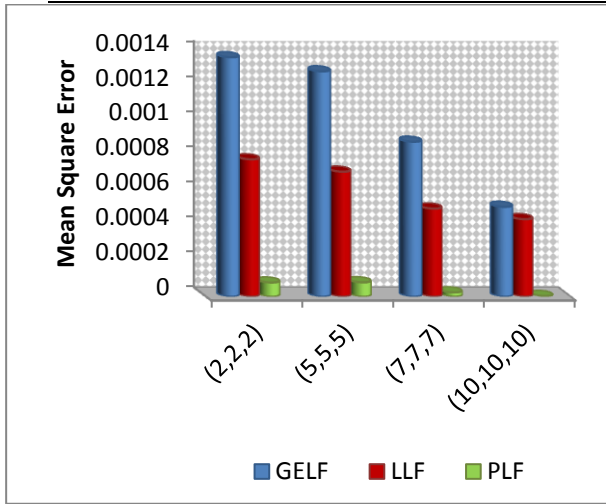


Figure 9: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, and $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at $(c_1, c_2, t_1, t_2) = (2, 1, 2, 2)$ for prior II

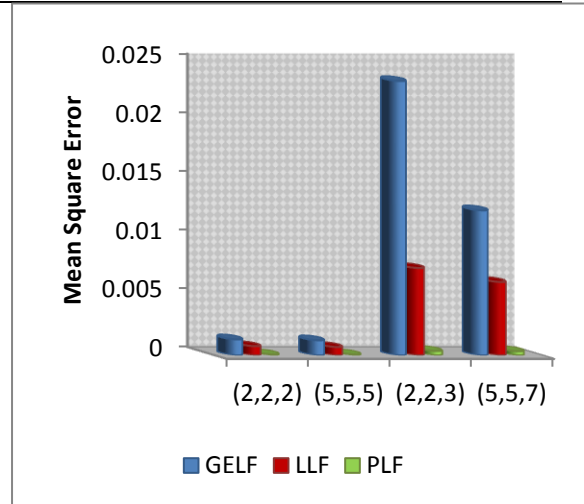


Figure 10: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, and $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at true value $\mathfrak{R}_{c_1, c_2, t_1, t_2} = 0.760$ for prior II

- At true value $\mathfrak{R}_{c_1, c_2, t_1, t_2} = 0.746$, the MSEs of $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ gets the smallest values compared to $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, and $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for similar record values except at $(n, m, w) = (10, 10, 10)$ via prior II (see Figure 11).
- Figure 12 illustrates that the ABs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ decrease as true value of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ increases for $(n, m, w) = (2, 2, 2)$.

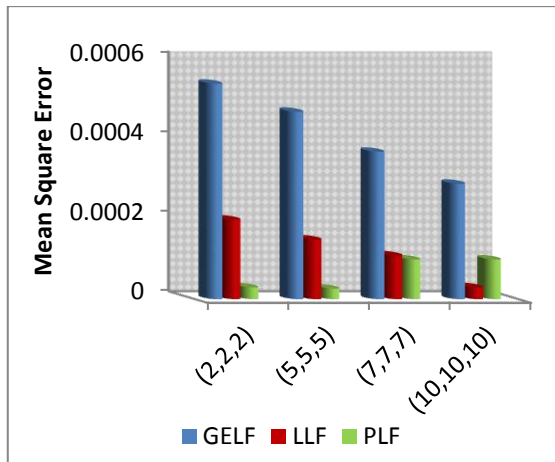


Figure 11: MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, and $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ at $(c_1, c_2, t_1, t_2) = (1, 2, 2, 2)$ for prior II

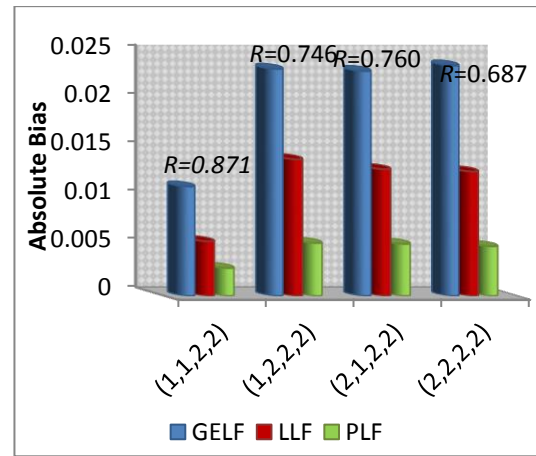


Figure 12: The ABs for all true values of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ at $n = m = w = 2$ for prior II

Table 2: Numerical results of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\check{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ for prior I

Loss function	$(c_1, c_2, t_1, t_2) = (1, 2, 2, 2)$				$(c_1, c_2, t_1, t_2) = (1, 2, 2, 2)$			
	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE
GELF	0.871	(2, 2, 2)	0.02036	0.00041	0.746	(2, 2, 2)	0.04300	0.00184
LLF			0.01541	0.00023			0.03373	0.00113
PLF			0.006413	0.00004			0.01229	0.00015
GELF	(5, 5, 5)	(5, 5, 5)	0.00542	2.9E-05	0.760	(5, 5, 5)	0.04175	0.00174
LLF			0.00048	2.3E-07			0.03295	0.00108
PLF			0.00067	4.5E-07			0.01169	0.00013
GELF	(7, 7, 7)	(7, 7, 7)	0.00314	9.9E-06	0.687	(7, 7, 7)	0.03750	0.00140

LLF			0.00036	1.3E-07			0.02955	0.00087
PLF			0.00023	5.4E-08			0.01078	0.00011
GELF			0.00293	8.6E-06			0.00323	1.0E-05
LLF	(10,10,10)		0.00030	1.3E-07	(10,10,10)		0.00289	8.3E-06
PLF			0.00020	4.2E-08			0.00005	3.2E-09
GELF			0.02601	0.00067			0.03778	0.00142
LLF	(2,2,3)		0.02085	0.00043	(2,2,3)		0.02893	0.00083
PLF			0.00891	0.00007			0.01107	0.00012
GELF			0.01449	0.00021			0.02832	0.00080
LLF	(5,5,7)		0.00995	0.00009	(5,5,7)		0.02031	0.00041
PLF			0.00407	1.6E-05			0.00797	6.3E-05
GELF			0.000286	8.1E-06			0.00768	5.9E-05
LLF	(7,7,8)		0.00103	1.0E-06	(7,7,8)		0.00161	2.6E-06
PLF			0.00058	3.3E-07			0.00177	3.1E-06
(c₁,c₂,t₁,t₂) = (2,1,2,2)					(c₁,c₂,t₁,t₂) = (1,2,2,2)			
Loss function	Real $\mathfrak{R}_{c_1,c_2,t_1,t_2}$	(n,m,w)	AB	MSE	Real $\mathfrak{R}_{c_1,c_2,t_1,t_2}$	(n,m,w)	AB	MSE
GELF	0.760		0.04292	0.00184	0.687		0.02976	0.00088
LLF		(2,2,2)	0.03398	0.00115		(2,2,2)	0.019006	0.00036
PLF			0.01359	0.00018			0.00830	6.8E-05
GELF			0.03351	0.00112			0.01627	0.00026
LLF		(5,5,5)	0.02530	0.00064		(5,5,5)	0.00513	2.6E-05
PLF			0.00979	9.5E-05			0.00013	1.8E-08
GELF			0.02497	0.00062			0.00552	3.0E-05
LLF		(7,7,7)	0.01697	0.00028		(7,7,7)	0.00244	5.9E-06
PLF			0.00664	4.4E-05			0.00016	2.6E-08
GELF			0.01358	0.00051			0.00491	6.11E-10
LLF		(10,10,10)	0.01511	0.00017		(10,10,10)	0.00235	2.04E-11
PLF			0.00544	3.2E-05			0.00015	2.03E-10
GELF			0.04676	0.00218			0.02680	0.01649
LLF		(2,2,3)	0.04676	0.00144		(2,2,3)	0.01649	0.00027
PLF			0.01530	0.00023			0.00663	4.41E-05
GELF			0.02963	0.00087			0.00534	2.8E-05
LLF		(5,5,7)	0.02181	0.00047		(5,5,7)	0.00193	3.7E-06
PLF			0.01002	0.00010			0.00187	3.5E-06
GELF			0.00357	1.2E-05			0.00017	3.11E-10
LLF		(7,7,8)	0.00270	7.3E-06		(7,7,8)	0.00105	6.55E-11
PLF			0.00031	9.9E-08			0.00135	3.44E-10

Table 3: Numerical results of $\hat{\mathfrak{R}}_{c_1,c_2,t_1,t_2}$, $\check{\mathfrak{R}}_{c_1,c_2,t_1,t_2}$, $\ddot{\mathfrak{R}}_{c_1,c_2,t_1,t_2}$ for prior II

(c₁,c₂,t₁,t₂) = (1,1,2,2)					(c₁,c₂,t₁,t₂) = (1,2,2,2)			
Loss function	Real $\mathfrak{R}_{c_1,c_2,t_1,t_2}$	(n,m,w)	AB	MSE	Real $\mathfrak{R}_{c_1,c_2,t_1,t_2}$	(n,m,w)	AB	MSE
GELF	0.871		0.01134	0.00012	0.746		0.02343	0.00054
LLF		(2,2,2)	0.00566	3.2E-05		(2,2,2)	0.01414	0.00020
PLF			0.002903	8.4E-06			0.005499	3.0E-05
GELF			0.00355	1.2E-05			0.02311	0.00047
LLF		(5,5,5)	0.00161	2.6E-06		(5,5,5)	0.01241	0.00015
PLF			0.00072	5.2E-07			0.00421	2.7E-05
GELF			0.00206	4.2E-06			0.02277	0.00037
LLF		(7,7,7)	0.00153	2.3E-06		(7,7,7)	0.01187	0.00011
PLF			0.00103	1.8E-07			0.00365	1.1E-04
GELF		(10,10,10)	0.00201	3.4E-06		(10,10,10)	0.02148	0.00029
LLF			0.00140	1.2E-07			0.01099	3.0E-05

PLF			0.00099	5.2E-08			0.00301	1.0E-04
GELF			0.02524	0.00164			0.03074	0.00426
LLF	(2,2,3)		0.02358	0.00121	(2,2,3)		0.03009	0.00077
PLF			0.00799	0.00080			0.00784	1.7E-05
GELF			0.02470	0.00135			0.02457	0.00333
LLF	(5,5,7)		0.02157	0.00117	(5,5,7)		0.02847	0.00051
PLF			0.00630	0.00060			0.00780	1.1E-05
GELF			0.02110	0.00124			0.01354	0.00251
LLF	(7,7,8)		0.01110	0.00101	(7,7,8)		0.02147	0.00039
PLF			0.00558	0.00038			0.00660	1.8E-06
$(c_1, c_2, t_1, t_2) = (2, 1, 2, 2)$				$(c_1, c_2, t_1, t_2) = (2, 2, 2, 2)$				
Loss function	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE	Real $\mathfrak{R}_{c_1, c_2, t_1, t_2}$	(n, m, w)	AB	MSE
GELF	0.760		0.02320	0.00136	0.687		0.02377	0.00056
LLF		(2,2,2)	0.01312	0.00078		(2,2,2)	0.01288	0.00016
PLF			0.00445	0.00008			0.00515	2.6E-05
GELF			0.03524	0.00128			0.02228	0.00035
LLF		(5,5,5)	0.02600	0.00071		(5,5,5)	0.01147	0.00012
PLF			0.00931	8.1E-05			0.00478	1.8E-06
GELF			0.03421	0.00088			0.02147	2.4E-05
LLF		(7,7,7)	0.02387	0.00050		(7,7,7)	0.01133	2.9E-06
PLF			0.00900	2.4E-05			0.00330	2.9E-08
GELF			0.03321	0.00051			0.02140	6.1E-08
LLF		(10,10,10)	0.02340	0.00044		(10,10,10)	0.01110	2.0E-07
PLF			0.00875	3.2E-06			0.00250	2.0E-08
GELF			0.03476	0.02336			0.02131	0.01356
LLF		(2,2,3)	0.03554	0.00744		(2,2,3)	0.01109	0.00124
PLF			0.02447	0.00037			0.00190	2.4E-05
GELF			0.03124	0.01235			0.02110	0.01254
LLF		(5,5,7)	0.02490	0.00625		(5,5,7)	0.01148	0.00120
PLF			0.02300	0.00035			0.00166	1.4E-05
GELF			0.03009	0.00147			0.01999	0.00124
LLF		(7,7,8)	0.02370	0.00420		(7,7,8)	0.01122	0.00110
PLF			0.02298	0.00021			0.00150	1.0E-06

Note that: E-0k stands for 10-k, k is integer

5. Actual Data Implementation

In this part, we illustrate our principles using three real datasets. We consider the real data sets reported in [33] where the data represent the time to break down (in minutes) of insulating fluids to electrodes at voltage levels 34 kV, 36 kV and 38 kV. The Kolmogorov-Smirnov (KS) test is used to separately fit each of the three datasets with the EPD along with the corresponding P-value (PV) (see Table 4). The empirical cdf and estimated pdf for these data are explained in Figure 13. At levels 34 kV, 36 kV and 38 kV, the times to break down are reported respectively as follows

Data Group I

0.96 4.15 0.19 0.78 8.01 31.75 7.35 6.5 8.27 33.91
 32.52 3.16 4.85 2.78 4.67 1.31 12.06 36.71 72.89

Data Group II

1.97 0.59 2.58 1.69 2.71 25.5 0.35 0.99 3.9 3.67
 2.07 0.96 5.35 2.9 13.77

Data Group III

0.47 0.73 1.4 0.74 0.39 1.13 0.09 2.38

Table 4: The K-S test and corresponding P-values for groups I, II and III

Data	K-S	PV
Group I	0.167	0.6013
Group II	0.185	0.6127
Group III	0.277	0.5013

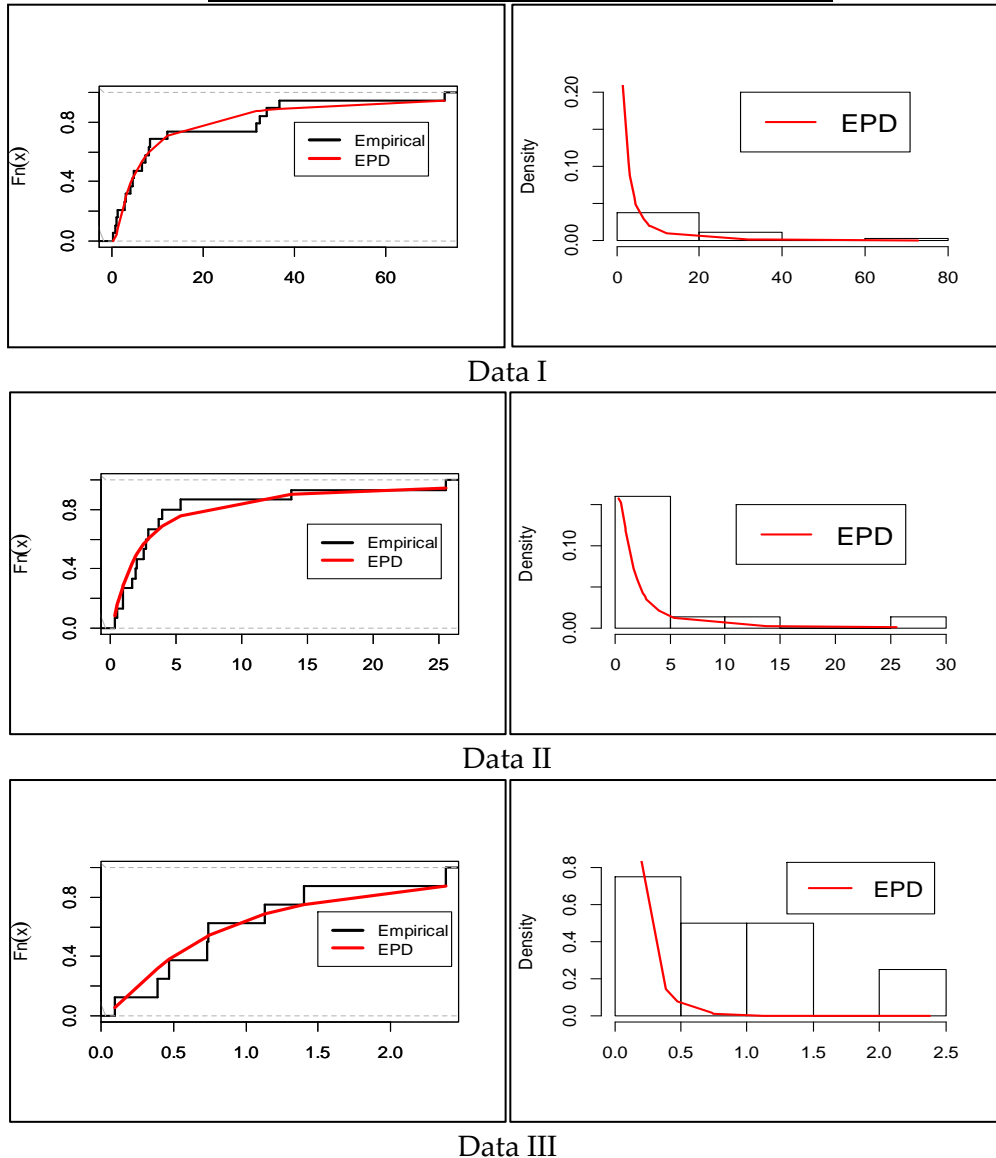


Figure 13: Characteristics and limitations of K-S test for the three data groups

We assume that electrical fluid of specimen considered being good if 1 out of 2 specimens are functioning properly at constant voltage. Form data group I, II and III, three sets of lower record values $\underline{r} = (0.96, 0.19)$, $\underline{p} = (1.97, 0.59, 0.35)$ and $\underline{s} = (0.47, 0.39, 0.09)$ are obtained, respectively. From \underline{r} , \underline{p} , and \underline{s} , we find that $n = 2$, $m = 3$, $w = 3$, then we calculate the estimates of $\mathfrak{R}_{c_1, c_2, f_1, f_2}$ using the ML and Bayesian approaches within GELF, LLF and PLF. Using the above LRVs, the MLE and BE of $\mathfrak{R}_{c_1, c_2, f_1, f_2}$, are calculated in Table 5.

Table 5: Bayes and ML estimates of $\mathfrak{R}_{c_1, c_2, f_1, f_2}$, for the real data

MLE of $\mathfrak{R}_{c_1, c_2, f_1, f_2}$	BE of $\mathfrak{R}_{c_1, c_2, f_1, f_2}$		
	GELF	LLF	PLF
0.5851	0.7673	0.7743	0.7956

6. Concluding Remarks

In the present work, we investigate the stress-strength reliability in a multi-component system with non-identical component strengths where both the stress and strength variables are the EPD. The ML and Bayesian procedures are used to analyse the reliability of MSS. Strength and stress distribution samples are used, and their measurements are presented in LRVs. We use MCMCO techniques in order to evaluate the accuracy of the various Bayesian estimates. The simulation study shows that for four choices of (c_1, c_2, t_1, t_2) , the MSEs and ABs decrease with the number of records, supporting the MLE's consistency characteristic of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$. Additionally, as the true value of $\mathfrak{R}_{c_1, c_2, t_1, t_2}$ increases, the MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ drop. Regarding the MCMCO approach, we deduce that the MSEs and ABs of $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ via PLF generally hold the lowest values in majority of cases. The ABs and MSEs of $\hat{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$, $\ddot{\mathfrak{R}}_{c_1, c_2, t_1, t_2}$ under different loss functions decrease as the number of records rises. The use of actual data demonstrates that our model's reliability estimates are very near to one, demonstrating its practical usefulness.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] Birnbaum, Z.W. (1956). *On a use of the Mann-Whitney statistic. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press Berkeley, CA, USA.
- [2] Birnbaum, Z.W. and McCarty, R.C. (1958), A distribution-free upper confidence bound for $\Pr\{Y < X\}$, based on independent samples of X and Y. *The Annals of Mathematical Statistics*: 558–562.
- [3] Hassan, A.S., Marwa, A.A., and Nagy, H.F. (2018), Estimation of P (Y<X) using record values from the generalized inverted exponential distribution. *Pakistan Journal of Statistics and Operation Research*, **14**(3): 645–660.
- [4] Safariyan, A., Arashi, M., and Arabi Belaghi, R. (2019), Improved estimators for stress-strength reliability using record ranked set sampling scheme. *Communications in Statistics-Simulation and Computation*, **48**(9): 2708–2726.
- [5] Sadeghpour, A., Salehi, M., and Nezakati, A. (2020), Estimation of the stress–strength reliability using lower record ranked set sampling scheme under the generalized exponential distribution. *Journal of Statistical Computation and Simulation*, **90**(1): 51–74.
- [6] Chaturvedi, A. and Malhotra, A. (2020), On estimation of stress-strength reliability using lower record values from proportional reversed hazard family. *American Journal of Mathematical and Management Sciences*, **39**(3): 234–251.
- [7] Al-Omari, A.I., Almanjahie, I.M., Hassan, A.S., and Nagy, H.F. (2020), Estimation of the stress-strength reliability for exponentiated Pareto distribution using median and ranked set sampling methods. *CMC-Computers, Materials & Continua*, **64**(2): 835–857.
- [8] Hassan, A.S., Al-Omari, A.I., and Nagy, H.F. (2021), Stress–strength reliability for the generalized inverted exponential distribution using MRSS. *Iranian Journal of Science and Technology, Transactions A: Science*, **45**(2): 641–659.
- [9] Akgul, F.G., Yu, K., and Senoglu, B. (2021), Estimation of the system reliability for generalized inverse Lindley distribution based on different sampling designs. *Communications in Statistics-Theory and Methods*, **50**(7): 1532–1546.
- [10] Hassan, A.S., Almanjahie, I.M., Al-Omari, A.I., Alzoubi, L., and Nagy, H.F. (2023), Stress–strength modeling using median-ranked set sampling: Estimation, simulation, and application. *Mathematics*, **11**(2): 318.
- [11] Bhattacharyya, G.K. and Johnson, R.A. (1974), Estimation of reliability in a multicomponent stress-strength model. *Journal of the American Statistical Association*, **69**(348): 966–970.
- [12] Bhattacharyya, G. and Johnson, R.A. (1974), Estimation of reliability in a multicomponent stress-strength model. *Journal of the American Statistical Association*, **69**(348): 966–970.

- [13] Ahmadi, K. and Ghafouri, S. (2019), Reliability estimation in a multicomponent stress–strength model under generalized half-normal distribution based on progressive type-II censoring. *Journal of Statistical Computation and Simulation*, **89**(13): 2505–2548.
- [14] Akgül, F.G. (2019), Reliability estimation in multicomponent stress–strength model for Topp-Leone distribution. *Journal of Statistical Computation and Simulation*, **89**(15): 2914–2929.
- [15] Pak, A., Khoolejani, N.B., and Rastogi, M.K. (2019), Bayesian inference on reliability in a multicomponent stress-strength bathtub-shaped model based on record values. *Pakistan Journal of Statistics and Operation Research*: 431–444.
- [16] Kohansal, A. (2019), On estimation of reliability in a multicomponent stress-strength model for a Kumaraswamy distribution based on progressively censored sample. *Statistical Papers*, **60**(6): 2185–2224.
- [17] Kayal, T., Tripathi, Y.M., Dey, S., and Wu, S.-J. (2020), On estimating the reliability in a multicomponent stress-strength model based on Chen distribution. *Communications in Statistics-Theory and Methods*, **49**(10): 2429–2447.
- [18] Hassan, A.S., Nagy, H.F., Muhammed, H.Z., and Saad, M.S. (2020), Estimation of multicomponent stress-strength reliability following Weibull distribution based on upper record values. *Journal of Taibah University for Science*, **14**(1): 244–253.
- [19] Hassan, A.S., Ismail, D.M., and Nagy, H.F. (2022), Reliability Bayesian analysis in multicomponent stress–strength for generalized inverted exponential using upper record data. *IAENG International Journal of Applied Mathematics*, **52**(3): 555–567.
- [20] Kotz, S. and Pensky, M., *The Stress-Strength Model and its Generalizations: Theory and Applications*. World Scientific. 2003.
- [21] Johnson, R.A. (1988), Stress-strength models for reliability. *Handbook of Statistics*, **7**: 27–54.
- [22] Pandey, M., Uddin, M.B., and Ferdous, J. (1992), Reliability estimation of an s-out-of-k system with non-identical component strengths: the Weibull case. *Reliability Engineering & System Safety*, **36**(2): 109–116.
- [23] Paul, R.K. and Uddin, M.B. (1997), Estimation of reliability of stress-strength model with non-identical component strengths. *Microelectronics Reliability*, **37**(6): 923–927.
- [24] Hassan, A.S. and Basheikh, H.M. (2012), Estimation of reliability in multi-component stress-strength model following exponentiated Pareto distribution. *The Egyptian Statistical Journal, Faculty of Graduate Studies for Statistical Research, Cairo University*, **56**(2): 82–95.
- [25] Rasethuntsa, T.R. and Nadar, M. (2018), Stress–strength reliability of a non-identical-component-strengths system based on upper record values from the family of Kumaraswamy generalized distributions. *Statistics*, **52**(3): 684–716.
- [26] Karam, N.S., Yousif, S.M., and Tawfeeq, B.J. (2020), Multicomponent inverse Lomax stress-strength reliability. *Iraqi Journal of Science*: 72–80.
- [27] Çetinkaya, Ç. (2021), Reliability estimation of a stress-strength model with non-identical component strengths under generalized progressive hybrid censoring scheme. *Statistics*, **55**(2): 250–275.
- [28] 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**(4): 887–904.
- [29] Shawky, A.I. and Abu-Zinadah, H.H. (2009), Exponentiated Pareto distribution: Different method of estimations. *International Journal of Contemporary Mathematical Sciences*, **4**(14): 677–693.
- [30] Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N., *Records*. Wiley, New York. 1998.
- [31] Chen, J. and Cheng, C. (2017), Reliability of stress–strength model for exponentiated Pareto distributions. *Journal of Statistical Computation and Simulation*, **87**(4): 791–805.
- [32] Lynch, S.M., *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Vol. 1. Springer. 2007.
- [33] Nelson, W.B., *Applied Life Data Analysis*. Vol. 521. John Wiley & Sons. 2003.

OPTIMAL SOFTWARE RELIABILITY PREDICTION USING CRITERIA WEIGHTS UNDER FUZZY DECISION- MAKING APPROACH

H. D. Arora, Anjali Naithani, Surbhi Gupta

•

Department of Mathematics, Amity Institute of Applied sciences,
Amity University Uttar Pradesh, Noida, U.P., India

hdarora@amity.edu, anathani@amity.edu, sgupta11@amity.edu

Abstract

Multi Criteria Decision Making techniques often face the challenge of determining criteria weights. The weights of criteria can significantly impact the outcomes of the decision-making process. Therefore, it is crucial to pay close consideration to the objectivity characteristics of criteria weights. Many weighting methods were discussed by various authors and utilized to solve various decision-making complications in Analytical Hierarchy Process (AHP), Entropy method, Weighing Score Method (WSM), Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Best worse method (MWM), VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR), Criteria Importance Through Intercriteria Correlation (CRITIC) method, ELECTRE, etc. This research article gives an overview of various weighting strategies that can be used in multi-criteria optimization and proposes a novel approach to determine criteria weights using Pythagorean fuzzy sets to handle uncertainties in the decision maker's preferences for allocating software reliability. The comparative analysis shows that the proposed weighting method has the advantage of being simple and straightforward in comparison to the existing weighting methods. The evaluation confirms that this novel approach is effective enough to determine objective weights.

Keywords: Criteria Weights, software reliability, MCDM, Entropy, AHP, Pythagorean fuzzy number

1. Introduction

Software systems have become an essential part of everyone's daily lives. The number of failures experienced by a specific user of software determines its reliability. The reliability goal is determined by the expectations of users and developers. Reliability refers to the degree to which the developments of software program enable it to perform its exact end-item use. software reliability, like functionality, usability, overall performance, serviceability, capability, maintainability, and documentation, is an important best trait. For software-primarily based systems utilized in protection-crucial packages which include computer relaying of energy transmission lines, software reliability assessment is critical. To give tools for estimating them, numerous software reliability models have been developed [1,2]. Early-stage software reliability prediction models are critical because they enable the early detection of cost overruns, issues with the software development process, and optimal development strategies [3]. Various components influence the overall operation of the software to varying degrees, and therefore information on component prominence can reveal critical information for software designers and test engineers. Additionally, software

dependability is estimated utilising antique data collected and an assumed distribution curve, making it inherently unstable. This model specifies how reliable software modules and programmes must be to maximise user utility while considering the system's financial and technical constraints. The reliability goal is determined by the expectations of users and developers. The degree to which the attributes of software enable it to perform its special end-item use is known as software reliability. Along with functionality, usability, overall performance, serviceability, capability, install ability, maintainability, and documentation, software reliability is a crucial quality trait.

Our research has concentrated on the problem of reliability allocation in a software system known as software hierarchy, which includes functions, programmes, and modules [4]. In this field, allocation of reliability is a relatively unexplored domain. To address the issue of reliability allocation, techniques of mathematical programming such as maximisation of reliability, minimization of cost, and multi-criterion decision-making models have been used. Numerous decision-making approaches have been discussed by various researchers for representing complex commercial or technical procedures.

Techniques for multiple-criteria decision-making (MCDM) have recently attained amazing popularity and wide use. Several researchers have employed AHP in testing consistency of software for reliability distribution or for issues with choosing suitable module. Sitorus et al. discussed the problem of choosing suitable method for mining and mineral processing in MCDM problems [5]. Methods such as AHP [6], weighted score method [7], VIKOR-TODIM [8], TOPSIS [9], DEMATEL [10], and other weighting methods have been proposed. Applications of decision making in public transport [11], location preferences of bridges [12], industrial symbiosis [13], and evaluating sustainable manufacturing strategies [14] were also discussed in the literature. In MCDM problem, selecting a suitable weighting method is a difficult task. Pamučar proposed full consistency method in MCDM which has cater smaller number of pairwise comparisons [15]. This method allows to produce realistic criteria weight coefficient values, which helps in reasonable judgement. These weighting methods are divided into several methods for directly and indirectly weighting criteria. A simulation study is provided to evaluate the effectiveness of approaches for transforming the ranks of multiple criteria into weights in multi-criteria decision-making [16,17]. Methods of weighting might be integrative, subjective, or both. Diakoulaki et al. suggested a method for calculating objective weights that is based on the quantification of two key MCDM concepts [18]. A sample of industrial businesses are subjected to the application of the suggested approach. The results demonstrated that the strategy guarantees a better compromise of the evaluated criteria when compared to those achieved by other sets of objective weights. Chatterjee et al. applied fuzzy AHP method for calculating the weights of functions, programmes, and modules to assess the dependability of the elements during the design and model phase of a software project [19]. In industrial health and risk evaluation, a new method has been presented [20]. The fuzzy VIKOR and Pythagorean fuzzy AHP are incorporated into the risk assessment process. To convert qualitative data into quantitative, Saaty proposed a numerical scale of one to nine where one denoting the 'equal importance' and nine 'great importance' [21]. FARE (Factor Relationship), a new technique for calculating the weights of the criteria based on the connections between all the criteria describing the phenomenon under consideration, is proposed [22]. An overview of several weighting approaches that can be used with multi-criteria optimization techniques is given in the work of Odu [23]. This approach concentrated on the utility of the various weighing methods for MCDM and suggested that subjective techniques are easy to compute as compared to objective one. Kasim presented subjective and objective weights methods for addressing MCDM problems applicable to real world scenario [24]. An MCDM-approach was proposed by Gupta et al. to rank a variety of SRGMs in software reliability [25]. To demonstrate the viability of the suggested strategy employing an entropy distance-based approach, an exploratory study is being conducted. In a hierarchy relating the expectations of clients, software technologists, and systems analyst for the software

system, Neha et al. described an allocation approach where preferences were assigned in terms of Pythagorean fuzzy numbers [26]. An example problem served as the basis for the proposed solution. For rating and assessing the services provided in hotels, Zoraghi et al. introduced a fuzzy MCDM prototype by incorporating both subjective and objective weights [27]. In the context of sustainable energy, Sahin proposed a comprehensive and comparative analysis of weighting MCDM methods [28].

This work proposes a novel technique for estimating the objective weights of criteria based on the elimination impact of criteria. This method employs a novel concept for weighing criteria. This method determines criterion weights based on the elimination impact of each criterion on alternative performance. Criteria with a greater impact on performance are given more weight. We provide some computer assessments to demonstrate effectiveness of our method after introducing it in a systematic manner. In practise, it is difficult for even a single decision-maker to provide numerical relative weights for various decision criteria.

This paper delivers an outline of various weighting methods that can be used with multi-criteria optimization techniques. The remainder of the paper is structured as follows: In section 2, we review the pioneered methods for determining criteria weights which will be helpful in the study for comparison purposes. Proposed methodology and algorithm are described in detail in section 3. Section 4 describes the proposed method's application using a real-life case study. Section 5 explains the comparison analysis of reliability allocation between the proposed method and the existing methods of determining criteria weights. Finally, section 6 concludes the paper by discussing prospects.

2. Methods

2.1. Determination of Criteria Weights

It might be challenging to select an appropriate weighting approach to resolve a multi-criteria decision problem. Some researchers have ignored the challenge of evaluating the criteria weights because they think that all decision-makers are aware of their importance. To avoid altering the MCDM models and provide accurate model outputs, it is necessary to take the validity of the acquired criteria weights into account. employing a variety of weighting methods. By enhancing the rationality, efficiency, and clarity of decision-making process, MCDM methodologies can aid in enhancing the quality of decisions. The authors pointed out that MCDM includes a variety of decision criteria and options, is recognised as a key component of recent operational research and decision science which comprises of multiple criteria and multiple decision alternatives. The methods used to establish the criteria weights that can be categorised as subjective and objective methods. While the objective approach chooses weights by arithmetic operations that disregard the decision makers' subjective judgement data, the subjective approach selects weights exclusively based on the considerations or judgements of decision makers. Because both subjective and objective approaches have advantages and disadvantages, an integrated or combined method appears to be preferable in determining weights of criteria [29]. To solve this problem, we investigated a technique for creating MCDM for ranking utilising fuzzy logic that relied on subjective and objective weights.

We put forward few essential objective and subjective methods of determining weights for solving MCDM problems in optimum allocation of software reliability.

2.1.1. Entropy Method

Because the decision matrix for a group of potential materials contains a certain amount of data, the entropy approach is used to determine the weight in a specific situation. Based on a predetermined decision matrix, entropy operates. Entropy is a measure of how much uncertainty a discrete

probability distribution can convey. Shannon developed an entropy approach that can be applied to determine the weights of the criteria in any MCDM problem [30]. The working algorithm based on Shannon's entropy [30] can be demonstrated in a series of steps:

Step 1: Create the decision matrix D_{ij} , which includes the evaluations of the options to be evaluated.

$$D_{ij} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ d_{31} & d_{32} & \dots & d_{3n} \\ \dots & \dots & \dots & \dots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{bmatrix}$$

Step 2: Normalize the decision matrix D_{ij} as

$$P_{ij} = \frac{D_{ij}}{\sum_{j=1}^n D_{ij}} \quad (1)$$

Step 3: Calculate the entropy for each selection criterion \mathfrak{E} as

$$\mathfrak{E}_i = \mathfrak{E}_0 \sum P_{ij} \times \ln(P_{ij}) \quad (2)$$

where \mathfrak{E}_0 is the entropy constant computed by $(\ln m)^{-1}$

Step 4: Compute diversity degree of the knowledge involved in the outcomes of the i^{th} criteria as

$$\check{D}_i = 1 - \mathfrak{E}_i \quad (3)$$

Step 5: Compute the normalised weights of the selected criteria as follows:

$$\theta_i = \frac{1 - \mathfrak{E}_i}{\sum_{i=1}^m (1 - \mathfrak{E}_i)} = \frac{\check{D}_i}{\sum_{i=1}^m \check{D}_i} \quad (4)$$

2.1.2. Analytical Hierarchy Process (AHP) Method

The AHP is a method that prioritises each choice by establishing the objectives or the hierarchy of relevance of attributes [31]. By condensing, dividing, and comparing numerous attributes, the AHP minimises cognitive errors. It can compare both qualitative and quantitative indices. As a result, it is frequently used in many different contexts, such as selection, evaluation, resource allocation, conflict resolution, priority and ranking, and optimization. The general procedure of finding weights using AHP is as follows:

Step 1: Creating a hierarchical structure with a goal at the top, attributes, or criteria at the second, and alternatives at the third.

Step 2: Determine the importance of various attributes or criteria in relation to the goal. A pairwise comparison matrix is created using a scale of relative importance (Table 1).

Step 3: Geometric mean (GM) suggested by Buckley [32] is employed to aggregate the pairwise comparison matrices as

$$\hat{r}_{ij} \sim = \sqrt[n]{(\hat{r}_{ij}^{1\sim} \otimes \hat{r}_{ij}^{2\sim} \dots \otimes \hat{r}_{ij}^{n\sim})} \quad (5)$$

where n stands for the DMs and $\hat{r}_{ij} \sim$ represents triangular values. Equation

$$\mathcal{P} \otimes \mathcal{Q} = (\mu_{\mathcal{P}}, \nu_{\mathcal{P}}, \gamma_{\mathcal{P}}) \otimes (\mu_{\mathcal{Q}}, \nu_{\mathcal{Q}}, \gamma_{\mathcal{Q}}) = (\mu_{\mathcal{P}} * \mu_{\mathcal{Q}}, \nu_{\mathcal{P}} * \nu_{\mathcal{Q}}, \gamma_{\mathcal{P}} * \gamma_{\mathcal{Q}}) \quad (6)$$

is used to multiply two fuzzy numbers.

Step 4: Fuzzy weights for all criteria are calculated using the formula

$$\varpi_i = \hat{r}_i \otimes (\hat{r}_1 \otimes \hat{r}_1 \otimes \dots \otimes \hat{r}_1)^{-1} \quad (7)$$

where \hat{r}_i is the vector summation of each $\hat{r}_{ij} \sim$.

Step 5: We need to de-fuzzified these fuzzy weights to get weights in crisp value for all criteria using centre of area formula as

$$\varpi_i = \frac{\mu_{\mathcal{P}} + \nu_{\mathcal{P}} + \gamma_{\mathcal{P}}}{3} \quad (8)$$

Step 6: Normalize these weights to get the weight total as 1. These weights can be further used for ranking of alternatives.

Table 1: Saaty's Scale Explanation

Linguistic Term	Importance	Explanation
Strongly low important (SLI)	0.142857	Values for inverse comparison
Very low important (VLI)	0.2	Values for inverse comparison
Low important (LI)	0.333333	Values for inverse comparison
Below average important (BAI)	0.5	Values for inverse comparison
Above average important (AAI)	2	Moderate advantage of the one element compared to the other
High important (HI)	3	High favouring of one element compared to the other
Very high important (VHI)	5	One element is given very high importance and has domination in practice compared to the other element
Strongly high important (SHI)	7	One element is favoured in comparison with the other based on strongly proved evidence and facts
Exactly equal (EE)	1	Both elements have equal contribution in the objective

2.1.3. Fuzzy AHP using Triangular Fuzzy Numbers (TFN)

Basic AHP has been enhanced by utilising fuzzy logic since it does not include vagueness for subjective judgments. Through the linguistic variables, which are represented as triangular numbers in F-AHP, pairwise comparisons of both criteria and alternatives were carried out [33]. Van Laarhoven and Pedrycz carried out one of the earliest implementations of fuzzy AHP [34]. For pairwise comparisons, they defined the triangle membership functions. Following that, Buckley contributed to the discussion by identifying the fuzziness of comparison ratios with triangle membership functions [32]. A novel technique for using triangular numbers in pair-wise comparisons was also presented by Chang [35].

A fuzzy number $\mathcal{F} = (l, m, u)$ is defined as TFN [36] if its membership function

$$\mu_A(z) = \begin{cases} \frac{z-l}{m-l}, & l \leq z \leq m \\ \frac{u-z}{u-m}, & m \leq z \leq u \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Graphically, TFN has been presented in Figure 1.

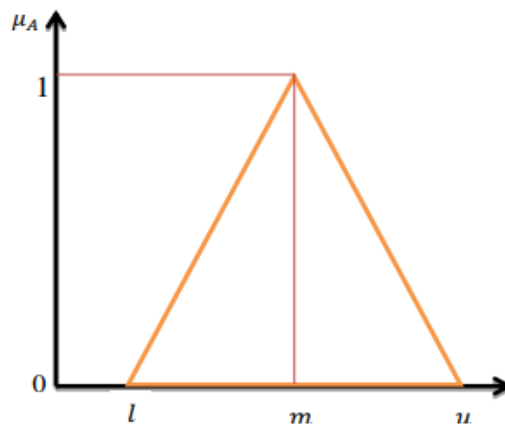


Figure 1: TFN

The laws of operation for addition, multiplication, and inverse are defined as follows:

$$\begin{aligned}(l_1, m_1, u_1) + (l_2, m_2, u_2) &= (l_1 + l_2, m_1 + m_2, u_1 + u_2) \\ (l_1, m_1, u_1) \times (l_2, m_2, u_2) &= (l_1 \times l_2, m_1 \times m_2, u_1 \times u_2) \\ (l_1, m_1, u_1)^{-1} &= \left(\frac{1}{u_1}, \frac{1}{m_1}, \frac{1}{l_1}\right)\end{aligned}$$

The membership function $\delta_{\mathcal{M}}$ of the fuzzy number \mathcal{M} can also be expressed [37] as

$$\mu_A(z) = \begin{cases} \mu_A^L(z), & a \leq z \leq b \\ 1, & b \leq z \leq c \\ \mu_A^U(z), & c \leq z \leq d \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where $\mu_A^L(z)$ and $\mu_A^U(z)$ are the lower and upper membership functions of fuzzy number A, respectively and $a \leq b \leq c \leq d$.

The procedure of fuzzy AHP suggested by Cheng's extent analysis is as follows:

- (1) Draw the hierarchal diagram.
- (2) Describe pair-wise comparisons in the form of fuzzy numbers.
- (3) Gather data as a fuzzy pairwise comparison matrix based on DMs judgement and expressed as

$$\begin{pmatrix} (1,1,1) & (l_{12}, m_{12}, u_{12}) & \dots & (l_{1n}, m_{1n}, u_{1n}) \\ (l_{21}, m_{21}, u_{21}) & (l_{22}, m_{22}, u_{22}) & \dots & (l_{2n}, m_{2n}, u_{2n}) \\ \dots & \dots & \dots & \dots \\ (l_{n1}, m_{n1}, u_{n1}) & (l_{n2}, m_{n2}, u_{n2}) & \dots & (1,1,1) \end{pmatrix}$$

- (4) Calculate \check{S}_i for every row of pair-wise comparison matrix using the expression

$$\check{S}_i = \sum_{j=1}^m \check{M}_{gi}^j \otimes \left[\sum_{i=1}^n \sum_{j=1}^m \check{M}_{gi}^j \right]^{-1} \quad (11)$$

where \check{M}_{gi}^j is TFN of pairwise comparison patterns. The values of $\sum_{j=1}^m \check{M}_{gi}^j$, $\sum_{i=1}^n \sum_{j=1}^m \check{M}_{gi}^j$ and $\left[\sum_{i=1}^n \sum_{j=1}^m \check{M}_{gi}^j \right]^{-1}$ are obtained by the expressions

$$\sum_{j=1}^m \check{M}_{gi}^j = \left(\sum_{j=1}^m l_j, \sum_{j=1}^m m_j, \sum_{j=1}^m u_j \right) \quad (12)$$

$$\sum_{i=1}^n \sum_{j=1}^m \check{M}_{gi}^j = \left(\sum_{i=1}^n l_i, \sum_{i=1}^n m_i, \sum_{i=1}^n u_i \right) \quad (13)$$

$$\left[\sum_{i=1}^n \sum_{j=1}^m \check{M}_{gi}^j \right]^{-1} = \left(\frac{1}{\sum_{i=1}^n u_i}, \frac{1}{\sum_{i=1}^n m_i}, \frac{1}{\sum_{i=1}^n l_i} \right) \quad (14)$$

- (5) Compute the magnitude of \check{S}_i w.r.t. each other. If $\mathbb{M}_1 = (l_1, m_1, u_1)$ and $\mathbb{M}_2 = (l_2, m_2, u_2)$ are two TFNs, then, the magnitude of \mathbb{M}_1 w.r.t. \mathbb{M}_2 is defined as follows

$$\mathcal{V}(\mathbb{M}_2 \geq \mathbb{M}_1) = \begin{cases} 1, & \text{if } m_2 \geq m_1 \\ 0, & \text{if } l_1 \geq u_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & \text{otherwise} \end{cases} \quad (15)$$

On the other hand, the magnitude of a TFN from k as another TFN is obtained by the following expression

$$\begin{aligned}\mathcal{V}(\check{M} \geq \check{M}_1, \check{M}_2, \dots, \check{M}_k) &= \mathcal{V}[(\check{M} \geq \check{M}_1) \text{ and } (\check{M} \geq \check{M}_2) \text{ and } \dots (\check{M} \geq \check{M}_k)] \\ &= \text{Min } \mathcal{V}(\check{M} \geq \check{M}_i), i = 1, 2, \dots, k\end{aligned} \quad (16)$$

- (6) Calculate the criteria and alternatives weight in pairwise comparison format using the expression

$$\check{d}^\dagger(A_i) = \text{Min } \mathcal{V}(\check{S}_i \geq \check{S}_k), k = 1, 2, \dots, n \text{ and } k \neq i \quad (17)$$

where A_i are n elements. The weight vector is now given by

$$\omega = [\check{d}^\dagger(A_1), \check{d}^\dagger(A_2), \dots, \check{d}^\dagger(A_n)]^T \quad (18)$$

- (7) Compute the final weight vectors after normalization as

$$\theta = [\check{d}(A_1), \check{d}(A_2), \dots, \check{d}(A_n)]^T \quad (19)$$

where θ is a non-fuzzy number.

2.1.4. Pythagorean Fuzzy Number (PFN) Method

We have evaluated the scale suggested by Gul [20] to determine the significance weight at every level of the hierarchical structure for the interval valued PFN. Let us assume that there are i alternatives.

Step 1: The compromised pairwise comparison matrix $A = (a_{ij})_{n \times n}$ is structured based on linguistic evaluations of experts using the scale proposed [38] in table 2.

Step 2: The difference matrices $\mathfrak{D} = [\mathfrak{d}_{ij}]_{n \times n}$ between the lower and upper values of the membership and non-membership functions are calculated using (20) and (21):

$$\mathfrak{d}_L = \mu_L^2 - \nu_U^2 \tag{20}$$

$$\mathfrak{d}_U = \mu_U^2 - \nu_L^2 \tag{21}$$

Table 2: Weighing scale for PFN

Linguistic Term	Lower value of membership degree (μ_L)	Upper value of membership degree (μ_U)	Lower value of non-membership degree (ν_L)	Upper value of non-membership degree (ν_U)
Strongly low important (SLI)	0	0	0.9	1
Very low important (VLI)	0.1	0.2	0.8	0.9
Low important (LI)	0.2	0.35	0.65	0.8
Below average important (BAI)	0.35	0.45	0.55	0.65
Above average important (AAI)	0.55	0.65	0.35	0.45
High important (HI)	0.65	0.8	0.2	0.35
Very high important (VHI)	0.8	0.9	0.1	0.2
Strongly high important (SHI)	0.9	1	0	0
Exactly equal (EE)	0.1965	0.1965	0.1965	0.1965

Step 3: Calculate relative multiplicative matrix $\mathfrak{M} = [m_{ij}]_{n \times n}$ with the help of (22) and (23):

$$m_L = \sqrt{1000^{\mathfrak{d}_L}} \tag{22}$$

$$m_U = \sqrt{1000^{\mathfrak{d}_U}} \tag{23}$$

Step 4: The determinacy value $\delta = [\delta_{ij}]_{n \times n}$ is calculated with the help of (24):

$$\delta_{ij} = 1 - (\mu_L^2 - \mu_U^2) - (\nu_U^2 - \nu_L^2) \tag{24}$$

Step 4: Compute matrix of weights $\omega = [\omega_{ij}]_{n \times n}$ is calculated by multiplying the relative multiplicative matrix with the determinacy value as

$$\omega_{ij} = \left(\frac{m_L + m_U}{2} \right) \delta_{ij} \tag{25}$$

Step 5: Normalize weights θ_i with the help of (26) as

$$\theta_i = \frac{\sum_{j=1}^n \omega_{ij}}{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij}} \tag{26}$$

These weights can be further used for ranking of alternatives.

3. Proposed Methodology and Algorithm

The assessment of the criteria weights is one of the most crucial phases of multicriteria evaluation. As criteria weights plays a significant role in MCDM, it is crucial to pay close attention to the factors associated with the criteria weights. Most of the current evaluation techniques for determining the weights of the criteria are based on the expert's subjective opinions. The selection of the criteria weights has a significant impact on the accuracy of the findings produced by multicriteria evaluation methods. In this segment, a technique based on the approach proposed by Keshavarz-Ghorabae et

al. is used to establish the weights of the criteria in a multi-criteria decision-making problem for optimal allocation of software reliability [39]. This technique is part of the objective weighting techniques used to determine criteria weights. The criteria weights in this method are determined by the exclusion effect of each criterion on the implementation of alternatives. In this analysis, the performances of the alternatives are calculated using a fundamental use of logarithmic measure with equal weights. We utilize the absolute deviation measure to define the effects of eliminating every condition. This metric indicates the variance between the performance of the alternative as a whole and its performance when a criterion is removed. The procedures for determining objective weights are as follows:

- The compromised pairwise comparison matrix $\mathcal{X} = (x_{ij})_{m \times n}$ is built using expert linguistic evaluations.
- Define fuzzy numbers to be used for pair-wise comparisons.
- Normalize the decision matrix as

$$\mathcal{N}_{ij} = \frac{x_{ij}}{\sum_{j=1}^n x_{ij}}; \tag{27}$$

$i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$

- Calculate the overall performance as follows:

$$\mathcal{P}_i = \ln \left[1 + \left\{ \frac{1}{n} \sum_{i=1}^n |\ln (\mathcal{N}_{ij})| \right\} \right] \tag{28}$$

- Using the formula, compute the performance of the alternatives by removing the impact of each criterion.

$$\mathcal{P}_{ij}^r = \ln \left[1 + \left\{ \frac{1}{n} \sum_{k, k \neq i}^n |\ln (\mathcal{N}_{ij})| \right\} \right] \tag{29}$$

- Evaluate the sum of absolute deviations between the overall performance and performance of the alternatives by eliminating impact of each criterion as

$$\omega_i = \sum_{j=1}^n |\mathcal{P}_{ij}^r - \mathcal{P}_i| \tag{30}$$

- Determine weights θ_i with the help of (31) as

$$\theta_i = \frac{\omega_i}{\sum_{i=1}^n \omega_i} \tag{31}$$

These weights can be further used for ranking of alternatives.

4. Allocation of System Reliability

The user is the ultimate arbiter of a system's performance and dependability. Before designing any strategy, it is essential to base it on users' opinions and perceptions of the dependability of distinct functions. The views of the user, the software manager and programmer may disagree. To accomplish the goal, it is necessary to incorporate all perspectives on how to assign reliability values to different software modules, programmes, and functions. First, we identify our problem and establish the system's reliability objective. This objective is based on what the user anticipates from the software. We consider the system's target reliability as 90% for the sake of our research. Based on this value, we allocate reliability to these modules, functions, and programmes of the system.

A hierarchy structure (figure 2) is formed based on the problem's objective. There are different tiers of modules, programmes, and functions in this structure. The user's decision-making judgement is converted into a fuzzy numerical value. Following user feedback on function, software technologists express their favourites for the programmes, and finally, computer programmer assign the inclinations to the autonomous modules. The total system reliability is shown at the top of this hierarchical structure. The user's perspective on the functions is the focus of the hierarchy's second level. Users give their opinions based on software's ability to produce the desired results. We have taken up four functions and denoted as $F_1, F_2, F_3,$ and F_4 . The viewpoint of a software engineer is shown at the third stage of the hierarchical structure. The programmes built for the user-preferred functions are shown at this level. Each function is allowed to have a different number of programmes. We assume programs at this level as $P_1, P_2,$ and P_3 . The perspective of the programmer

on the software system's modules is included in the hierarchy's bottom level. The system consists of 4 modules namely $M_1, M_2, M_3,$ and M_4 .

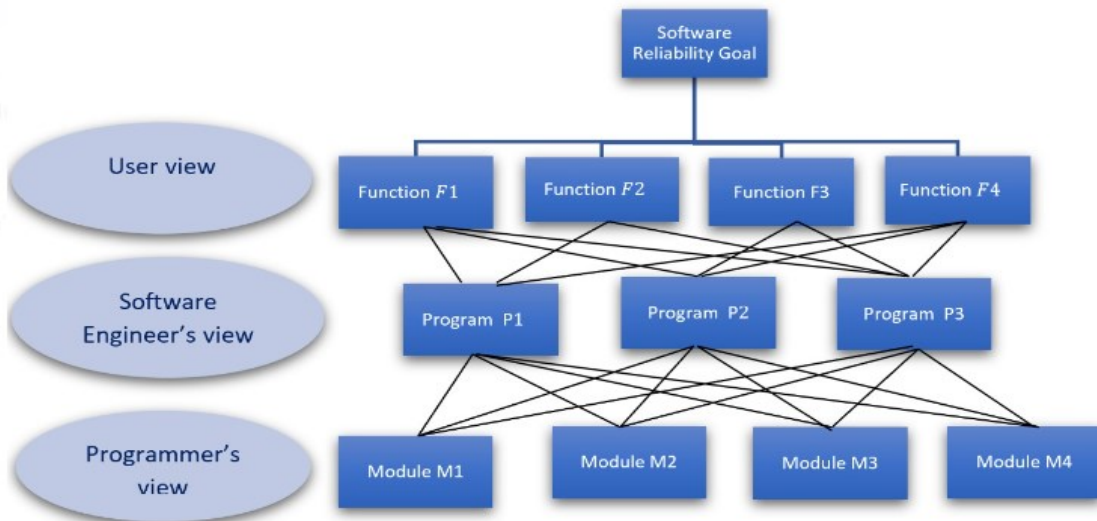


Figure 2: The software hierarchy of the proposed system for reliability allocation

We allocate the reliabilities after determining the weights for the alternatives based on the processes. We employ the following mathematical formula to determine the reliability to the priority weights:

$$\mathcal{R}T_i = \mathcal{R}T^{\theta_i}$$

where $\mathcal{R}T_i$: Reliability of the i^{th} element, $\mathcal{R}T$: Objective reliability of the system and θ_i : weights acquired.

Each element in the hierarchy is interconnected with multiple elements at the top levels. So, we choose the one that has the highest level of reliability. For example, if a program P_1 relates to four modules $M_1, M_2, M_3,$ and M_4 then reliability allocated will be

$$\mathcal{R}T_{P_1} = \text{Max} \{ \mathcal{R}T_{M_1}, \mathcal{R}T_{M_2}, \mathcal{R}T_{M_3}, \mathcal{R}T_{M_4} \}.$$

where $\mathcal{R}T_{P_1}$ represents reliability allocated to program P_1 and $\mathcal{R}T_{M_1}, \mathcal{R}T_{M_2}, \mathcal{R}T_{M_3},$ and $\mathcal{R}T_{M_4}$ are the reliabilities of modules $M_1, M_2, M_3,$ and M_4 associated with P_1 .

4.1. Application of the proposed framework

In this section, a simplified example has been used to illustrate how the suggested solution can be used to the reliable allocation problem. Figure 2 depicts the hierarchical structure of reliability of software.

It is assumed that the goal of our overall system reliability of the system to be 0.90. Assume a software with four functions has been created $F_1, F_2, F_3,$ and F_4 . The fuzzy values were assigned by DMs while performing pairwise comparing. As shown in table 1, linguistic statements were used for pairwise comparisons while collecting expert opinions using a scale of relative importance. The pairwise comparison matrix is attained as

Table 3: Comparison matrix of functions with linguistic statement using expert opinion

Software	F_1	F_2	F_3	F_4
F_1	EE	SLI	VLI	LI
F_2	SHI	EE	AAI	HI
F_3	VHI	BAI	EE	HI
F_4	HI	LI	LI	EE

The relative importance using Saaty scale is shown in table 4 as

Table 4: Comparison matrix with relative importance using expert opinion

Software	F ₁	F ₂	F ₃	F ₄
F ₁	1	0.1429	0.5	1
F ₂	7	1	0.333333	7
F ₃	2	3	1	2
F ₄	1	0.1429	0.5	1

The priority weights and allocation of reliabilities to these functions using proposed methodology is demonstrated in table 5 as

Table 5: Reliability allocation for functions

Functions	Weights (θ_i)	Reliability provision ($\mathcal{RT}_i = \mathcal{RT}^{\theta_i}$)
F ₁	0.017643	$(0.90)^{0.017643} = \mathbf{0.998143}$
F ₂	0.401141	$(0.90)^{0.401141} = 0.958616$
F ₃	0.315821	$(0.90)^{0.315821} = 0.967272$
F ₄	0.265395	$(0.90)^{0.265395} = 0.972425$

The weights and reliabilities determined for all offered functions in relation to the target reliability are shown in the table 5. It has been observed from the above table that the maximum reliability is assigned to the function F₁. i.e., 0.998143 followed by F₄, F₃, and F₂ in that order. Furthermore, we assign reliability to the programmes associated to each individual function using the target reliability of the functions as our benchmark. To meet the needs of each function, relative weights of the programmes are computed for reliability criteria. Programs P₁ and P₂ are required for function F₁, P₁ and P₃ are required for function F₂, P₂ and P₃ are required for function F₃, and P₁, P₂, P₃ are required for function F₄, as shown in table 6.

After converting linguistic statements to relative importance, we find weights and allocate reliability to these programs using our methodology as described in table 7.

The table 7 shows that programme P₁ has been given the highest reliability rating, which is 0.999606, followed by P₂ and P₃ in that order. When a single programme is linked to multiple functions, we choose the highest level of reliability possible. In a similar way, we assign reliability to four proposed modules. To determine the weights of these modules, evaluation patterns as studied are developed. Here, all these programs relate to these four modules such as M₁, M₂, M₃, and M₄ and assessment pattern of these modules w.r.t. each program in linguistic form is depicted in table 8.

Table 6: Comparison matrix of programs with linguistic statement using expert opinion

(a)			(b)		
F ₁	P ₁	P ₂	F ₂	P ₁	P ₃
P ₁	EE	SLI	P ₁	EE	BAI
P ₂	SHI	EE	P ₃	AAI	EE

(c)			(d)			
F ₃	P ₂	P ₃	F ₄	P ₁	P ₂	P ₃
P ₂	EE	BAI	P ₁	EE	SLI	BAI
P ₃	AAI	EE	P ₂	SHI	EE	LI
			P ₃	AAI	HI	EE

Table 7: Reliability allocation for programs

Programs	Weights (θ_i)	Reliability allocation ($\mathfrak{RT}_i = \mathfrak{RT}^{\theta_i}$)
P_1	0.00374, 0.107793, 0.072205	$Max \{(0.90)^{0.00374}, (0.90)^{0.99626}, (0.90)^{0.072205}\}$ = 0.999606
P_2	0.99626, 0.035899, 0.356939	$Max \{(0.90)^{0.99626}, (0.90)^{0.035899}, (0.90)^{0.356939}\}$ = 0.996225
P_3	0.892207, 0.964101, 0.570856	$Max \{(0.90)^{0.892207}, (0.90)^{0.964101}, (0.90)^{0.570856}\}$ = 0.941627

Table 8: Comparison matrix of modules with linguistic statement using expert opinion

(a)					(b)				
P_1	M_1	M_2	M_3	M_4	P_2	M_1	M_2	M_3	M_4
M_1	EE	BAI	SLI	VLI	M_1	EE	VLI	BAI	SLI
M_2	AAI	EE	BAI	SHI	M_2	VHI	EE	AAI	LI
M_3	SHI	AAI	EE	HI	M_3	AAI	BAI	EE	SLI
M_4	VHI	SLI	LI	EE	M_4	SHI	HI	SHI	EE

(c)				
P_3	M_1	M_2	M_3	M_4
M_1	EE	VLI	SLI	SHI
M_2	VHI	EE	AAI	BAI
M_3	SHI	BAI	EE	LI
M_4	SLI	AAI	HI	EE

After converting linguistic statements to relative importance, we find weights and allocate reliability to these modules using our methodology as described in table 9.

Table 9: Reliability allocation for modules

Programs	Weights (θ_i)	Reliability allocation ($\mathfrak{RT}_i = \mathfrak{RT}^{\theta_i}$)
M_1	0.048414, 0.034373, 0.220843	$Max \{(0.90)^{0.048414}, (0.90)^{0.034373}, (0.90)^{0.220843}\}$ = 0.996385
M_2	0.287318, 0.28954, 0.33712	$Max \{(0.90)^{0.287318}, (0.90)^{0.28954}, (0.90)^{0.33712}\}$ = 0.970182
M_3	0.374067, 0.134255, 0.260876	$Max \{(0.90)^{0.374067}, (0.90)^{0.134255}, (0.90)^{0.260876}\}$ = 0.985954
M_4	0.290201, 0.541832, 0.181161	$Max \{(0.90)^{0.290201}, (0.90)^{0.541832}, (0.90)^{0.181161}\}$ = 0.981094

From the above table, it has been observed that the module M_1 has been given the highest reliability rating, which is 0.996385, followed by M_3 , M_4 and M_2 in that order. Before developing the real system, all software components are given reliability targets using this allocation technique. These objectives must consider the normal user expectations as well as the specifications for the software's structure and reliability to be practical and significant. For software reliability allocation, this strategy enhances communication between users, software managers, and programmers.

5. Comparative Analysis

In this section, comparison has been carried out by finding criteria weights using existing methods as discussed in section 2. Weights calculation and reliability allocation for functions, programs and modules has been demonstrated in table 10-13 as shown below:

Table 10: Weight calculation and reliability allocation using entropy method

Functions	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
F_1	0.207905	0.978333
F_2	0.299162	0.968972
F_3	0.217526	0.977342
F_4	0.275407	0.97140
Programs	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
P_1	0.49999, 0.5000, 0.291107	0.969794
P_2	0.50000, 0.5000, 0.524813	0.948683
P_3	0.50000, 0.5000, 0.184080	0.980792
Modules	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
M_1	0.180953, 0.222048, 0.305404	0.981115
M_2	0.254397, 0.343202, 0.308463	0.973553
M_3	0.158464, 0.259382, 0.254434	0.983443
M_4	0.406186, 0.175368, 0.131699	0.98622

Table 11: Weight calculation and reliability allocation using AHP method

Functions	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
F_1	0.059249	0.993777
F_2	0.482803	0.950404
F_3	0.313851	0.967473
F_4	0.144097	0.984933
Programs	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
P_1	0.125016, 0.3333, 0.075064	0.992122
P_2	0.874984, 0.2500, 0.591723	0.974004
P_3	0.66660, 0.7500, 0.3332130	0.965502
Modules	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
M_1	0.066274, 0.059672, 0.16054	0.993733
M_2	0.31181, 0.233218, 0.358978	0.975727
M_3	0.488009, 0.106113, 0.249495	0.988882
M_4	0.133907, 0.60099, 0.230987	0.985991

Table 12: Weight calculation and reliability allocation using fuzzy AHP (TFN) method

Functions	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
F_1	0.056585	0.994056
F_2	0.376814	0.961077
F_3	0.348631	0.963934
F_4	0.21797	0.977296
Programs	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
P_1	NA, 0.50, 0.212266	0.977884
P_2	NA, 0.3163, 0.433994	0.9672
P_3	0.50, 0.6837, 0.35374	0.963416
Modules	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
M_1	0.130874, 0.072106, 0.233092	0.992432
M_2	0.317613, 0.265214, 0.277064	0.972444
M_3	0.343306, 0.156194, 0.285197	0.983678
M_4	0.208207, 0.506487, 0.204647	0.978669

Table 13: Weight calculation and reliability allocation using PFN method

Functions	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
F_1	0.020534	0.997839
F_2	0.542303	0.944464
F_3	0.31764	0.967087
F_4	0.119524	0.987486
Programs	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
P_1	0.03444, 0.316381, 0.040192	0.995774
P_2	0.965556, 0.14101, 0.709697	0.985253
P_3	0.683619, 0.85899, 0.250111	0.973992
Modules	Weights (θ_i)	Max. Reliability allocation ($\mathfrak{R}_i = \mathfrak{R}^{\theta_i}$)
M_1	0.020222, 0.020222, 0.333048	0.997872
M_2	0.366725, 0.202953, 0.206594	0.978844
M_3	0.437906, 0.047376, 0.34180	0.995021
M_4	0.175147, 0.729448, 0.118557	0.987586

Based on the comparison shown in table 10 to table 13 several finding can be made: first, function F_1 is assigned with maximum reliability by all the methods considered such as entropy method, AHP method, fuzzy AHP, PFN and our proposed technique. Secondly, maximum reliability is assigned to program P_1 by all except the entropy method and thirdly, for modules M_1 is chosen for the optimal allocation of software reliability. The above findings suggest that the suggested methodology provides software designers a productive and strategic method for producing highly robust software. The achievement of the best system reliability goal requires having a suitable reliability allocation method because development of a software is a significant cost factor of computer system. By incorporating users' opinions with those of software engineers and programmers, it also emphasises how crucial it is to comprehend users' roles in the software industry. As a result, this study improves communication between users, software engineers, and programmers. Also, employing this study DMs have more liberty and convenience in conveying their thoughts.

6. Conclusions and Future Scope

In MCDM, ascertaining criteria weights is a key issue. Weights are designed to convey the comparative importance of certain criteria. Real-world applications always involve varied degrees of criterion contribution to the outcomes being considered. Ignoring the issue will lead to wrong decisions. Pairwise comparisons are frequently employed as intermediate decision support when the DM finds it difficult to order the options as a whole and immediately regarding a criterion. Numerous techniques have been presented, and their efficacy has been compared, to estimate the preference values from the pairwise comparison matrix. It is essential to include users' perception while developing and designing software because they play a significant influence in the software market. This paper provides selective pioneered methods of computing priority weights and discusses the allocation of reliability of a software problem that arises during the designing and development phases of a programme. We compare the function based on the user's perception, compute the weights for each function, and assign reliability to them using a pairwise comparison matrix. We create a comparison matrix for modules considering the opinions of the programmers and a comparison matrix for programmes based on the decisions made by software engineers after assigning reliability to the functions. Then, based on the developed processes in the technique, reliability to programmes and modules is assigned using suggested approach. The study compares priority weights determined by entropy method, AHP, fuzzy AHP and Pythagorean fuzzy approach

with the weights computed from proposed methodology for validation. We may assert that incorporating fresh MCDM techniques based on innovative viewpoints could guarantee the reliability of outcomes. Decision-makers can get weights that are more trustworthy by incorporating weighting procedures. The success of the software system reliability objective is assured by the reliability allocation model proposed in this work. For reliability allocation, this strategy enhances communication between consumers, software engineers, and programmers. The study broadens developers' comprehension of the impact of users' opinions during the software development phase as well as engineers' and programmers' perspectives at various levels of the hierarchy. However, the study is only capable of considering at one decision maker's perspective at each level. Furthermore, because the study only considers a few methods for computing the weighing criteria for functions, programmes, and modules, future investigation will focus on different approaches of decision-making in the allocation of reliability to software and can incorporate more objective and subjective methods for determining priority weights. The suggested approach can enhance future study in a vague defined environment, such as fuzzy, Neutrosophic and Fermatean environments.

References

- [1] Brooks, W. D. & Motley, R. W. (1980). Analysis of discrete software reliability Models, *Technical Report RADC-TR-80-84*, Rome Air Development Center, New York.
- [2] Angus, J. E., Schafer, R. E., & Sukert, A. (1980). Software reliability model validation, *In Proc. Annu. Reliability and Maintainability Symp.*, San Francisco, CA, Jan., pp. 191-193.
- [3] Aggarwal, A. G., Dhaka, V., & Nijhawan, N. (2017). Reliability analysis for multi-release open-source software systems with change point and exponentiated Weibull fault reduction factor. *Life Cycle Reliability and Safety Engineering*, 6(1), 3–14. <https://doi.org/10.1007/s41872-017-0001-0>
- [4] Zahedi, F., & Ashrafi, N. (1991). Software reliability allocation based on structure, utility, price, and cost. *IEEE Transactions on Software Engineering*, 17(4), 345–356. <https://doi.org/10.1109/32.90434>
- [5] Sitorus, F., Cilliers, J. J., & Brito-Parada, P. R. (2019). Multi-criteria decision making for the choice problem in mining and mineral processing: Applications and trends. *Expert Systems with Applications*, 121, 393–417. <https://doi.org/10.1016/j.eswa.2018.12.001>
- [6] Sehra, S. K., Singh Brar, Y., & Kaur, N. (2012). Multi Criteria Decision Making Approach for Selecting Effort Estimation Model. *International Journal of Computer Applications*, 39(1), 10–17. <https://doi.org/10.5120/4783-6989>
- [7] Zha, S., Guo, Y., Huang, S., & Wang, S. (2020). A Hybrid MCDM Method Using Combination Weight for the Selection of Facility Layout in the Manufacturing System: A Case Study. *Mathematical Problems in Engineering*, 2020, 1–16. <https://doi.org/10.1155/2020/1320173>
- [8] Arya, V., & Kumar, S. (2020). A new picture fuzzy information measure based on shannon entropy with applications in opinion polls using extended VIKOR–TODIM approach. *Computational and Applied Mathematics*, 39(3). <https://doi.org/10.1007/s40314-020-01228-1>
- [9] Xu, Z., & Zhang, X. (2013). Hesitant fuzzy multi-attribute decision making based on TOPSIS with incomplete weight information. *Knowledge-Based Systems*, 52, 53–64. <https://doi.org/10.1016/j.knosys.2013.05.011>
- [10] Kobryń, A. (2017). DEMATEL as a weighting method in multi-criteria decision analysis. *Multiple Criteria Decision Making*, 12, 153–167. <https://doi.org/10.22367/mcdm.2017.12.11>
- [11] Ueasin, N. (2020). Decision-making on Public Transportation Services Based on the Socio-economic, Psychological, and Environmental Concern Factors. *The Open Transportation Journal*, 14(1), 22–31. <https://doi.org/10.2174/1874447802014010022>
- [12] Zagorskis, J., & Turskis, Z. (2020). Location Preferences of New Pedestrian Bridges Based on Multi-Criteria Decision-Making and GIS-Based Estimation. *The Baltic Journal of Road and Bridge*

Engineering, 15(2), 158–181. <https://doi.org/10.7250/bjrbe.2020-15.478>

[13] Alakaş, H. M., Gür, Ş., Özcan, E., & Eren, T. (2020). Ranking of sustainability criteria for industrial symbiosis applications based on ANP. *Journal of Environmental Engineering and Landscape Management*, 28(4), 192–201. <https://doi.org/10.3846/jeelm.2020.13689>

[14] Keshavarz-Ghorabae, M., Govindan, K., Amiri, M., Zavadskas, E. K., & Antuchevičienė, J. (2019). An integrated type-2 fuzzy decision model based on WASPAS and SECA for evaluation of sustainable manufacturing strategies. *Journal of Environmental Engineering and Landscape Management*, 27(4), 187–200. <https://doi.org/10.3846/jeelm.2019.11367>

[15] Pamučar, D., Stević, Ž., & Sremac, S. (2018). A New Model for Determining Weight Coefficients of Criteria in MCDM Models: Full Consistency Method (FUCOM). *Symmetry*, 10(9), 393. <https://doi.org/10.3390/sym10090393>

[16] Alfares, H. K., & Duffuaa, S. O. (2016). Simulation-Based Evaluation of Criteria Rank-Weighting Methods in Multi-Criteria Decision-Making. *International Journal of Information Technology & Decision Making*, 15(01), 43–61. <https://doi.org/10.1142/s0219622015500315>

[17] Németh, B., Molnár, A., Bozóki, S., Wijaya, K., Inotai, A., Campbell, J. D., & Kaló, Z. (2019). Comparison of weighting methods used in multicriteria decision analysis frameworks in healthcare with focus on low- and middle-income countries. *Journal of Comparative Effectiveness Research*, 8(4), 195–204. <https://doi.org/10.2217/cer-2018-0102>

[18] Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The critic method. *Computers & Operations Research*, 22(7), 763–770. [https://doi.org/10.1016/0305-0548\(94\)00059-H](https://doi.org/10.1016/0305-0548(94)00059-H)

[19] Chatterjee, S., Singh, J. B., & Roy, A. (2013). A structure-based software reliability allocation using fuzzy analytic hierarchy process. *International Journal of Systems Science*, 46(3), 513–525. <https://doi.org/10.1080/00207721.2013.791001>

[20] Gul, M. (2018). Application of Pythagorean fuzzy AHP and VIKOR methods in occupational health and safety risk assessment: the case of a gun and rifle barrel external surface oxidation and colouring unit. *International Journal of Occupational Safety and Ergonomics*, 26(4), 705–718. <https://doi.org/10.1080/10803548.2018.1492251>

[21] Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3), 234–281. [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5)

[22] Ginevičius, R. (2011). A new determining method for the criteria weights in multicriteria evaluation. *International Journal of Information Technology & Decision Making*, 10(06), 1067–1095. <https://doi.org/10.1142/s0219622011004713>

[23] Odu, G. O. (2019). Weighting methods for multi-criteria decision-making technique. *Journal of Applied Sciences and Environmental Management*, 23(8), 1449. <https://doi.org/10.4314/jasem.v23i8.7>

[24] Kasim, M. M. (2014). Determination of criteria weights in solving multi-criteria problems. *AIP Conference Proceedings*. <https://doi.org/10.1063/1.4903555>

[25] Gupta, A., Gupta, N., Garg, R., & Kumar, R. (2018). Evaluation, Selection and Ranking of Software Reliability Growth Models using Multi criteria Decision making Approach. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*. <https://doi.org/10.1109/ccaa.2018.8777644>

[26] Neha, N. A., Verma, V., Tandon, A., & Aggarwal, A. G. (2021). Software reliability allocation incorporating Pythagorean fuzzy theory and AHP. *International Journal of Industrial and Systems Engineering*, 38(1), 19. <https://doi.org/10.1504/ijise.2021.115128>

[27] Zoraghi, N., Amiri, M., Talebi, G., & Zowghi, M. (2013). A fuzzy MCDM model with objective and subjective weights for evaluating service quality in hotel industries. *Journal of Industrial Engineering International*, 9(1). <https://doi.org/10.1186/2251-712x-9-38>

[28] Şahin, M. (2020). A comprehensive analysis of weighting and multicriteria methods in the

context of sustainable energy. *International Journal of Environmental Science and Technology*, 18(6), 1591–1616. <https://doi.org/10.1007/s13762-020-02922-7>

[29] Liu H, & Kong F. (2005). A new MADM algorithm based on fuzzy subjective and objective integrated weights. *International Journal of Information and Systems Sciences*, 1: 420-427

[30] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

[31] Saaty, T. L., & Vargas, L. G. (1982). *The Logic of Priorities*. Springer Netherlands. <https://doi.org/10.1007/978-94-017-3383-0>

[32] Buckley, J. J. (1985). Fuzzy hierarchical analysis. *Fuzzy Sets and Systems*, 17(3), 233–247. [https://doi.org/10.1016/0165-0114\(85\)90090-9](https://doi.org/10.1016/0165-0114(85)90090-9)

[33] Kilincci, O., & Onal, S. A. (2011). Fuzzy AHP approach for supplier selection in a washing machine company. *Expert Systems with Applications*, 38(8), 9656–9664. <https://doi.org/10.1016/j.eswa.2011.01.159>

[34] van Laarhoven, P. J. M., & Pedrycz, W. (1983). A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems*, 11(1-3), 229–241. [https://doi.org/10.1016/s0165-0114\(83\)80082-7](https://doi.org/10.1016/s0165-0114(83)80082-7)

[35] Chang, D.-Y. (1996). Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, 95(3), 649–655. [https://doi.org/10.1016/0377-2217\(95\)00300-2](https://doi.org/10.1016/0377-2217(95)00300-2)

[36] Kaufmann, A., & Gupta, M. M. (1991). *Introduction to fuzzy arithmetic: theory and applications*. New York (N.Y.): Van Nostrand Reinhold.

[37] Chu, T.-C., & Lin, Y.-C. (2009). An interval arithmetic based fuzzy TOPSIS model. *Expert Systems with Applications*, 36(8), 10870–10876. <https://doi.org/10.1016/j.eswa.2009.01.083>

[38] Ilbahar, E., Karaşan, A., Cebi, S., & Kahraman, C. (2018). A novel approach to risk assessment for occupational health and safety using Pythagorean fuzzy AHP & fuzzy inference system. *Safety Science*, 103, 124–136. <https://doi.org/10.1016/j.ssci.2017.10.025>

[39] Keshavarz-Ghorabae, M., Amiri, M., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2021). Determination of Objective Weights Using a New Method Based on the Removal Effects of Criteria (MEREC). *Symmetry*, 13(4), 525. <https://doi.org/10.3390/sym13040525>

STRESS-STRENGTH RELIABILITY MODEL UNDER MULTIVARIATE NORMAL SETUP AND ITS APPLICATIONS

Babulal Seal¹ and Anirban Goswami^{*2}

¹Department of Mathematics and Statistics, Aliah University, Kolkata, West Bengal

²Regional Research Institute of Unani Medicine, Patna, Bihar

¹babulal_seal@yahoo.com, ²anirbanstat09@gmail.com

*Corresponding author

Abstract

We often see that in a system, the energy is supplied to the system by p_1 sources and its consumed through p_2 sources and the sources are linearly dependent with vector \mathbf{a}' and \mathbf{b}' . The overall representation of the two sets are related to vectors \mathbf{a} and \mathbf{b} , such that they are approximated by $\mathbf{a}'\mathbf{x}$ and $\mathbf{b}'\mathbf{y}$ as in principal component analysis. In this article, a stress strength reliability model $R = \Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$, when \mathbf{x} and \mathbf{y} are distributed dependently multivariate normal distribution is proposed, with \mathbf{a} and \mathbf{b} are two known vectors. MVUE and MLE of R are obtained. Through simulation studies, their performances are compared using different measures. The two-sided confidence intervals and lower bounds of R are obtained through exact and asymptotic distribution of maximum-likelihood estimators and using bootstrap procedure. Through simulation studies, the performances of these confidence intervals are empirically checked using their coverage and the accuracy. In this study, we proposed to choose the optimal sample size for an experiment assures an adequate power and level. Finally, we applied these interval estimators to a real data set.

Keywords: Stress-strength, Principal component, Maximum Likelihood Estimator (MLE), Minimum Variance Unbiased Estimator (MVUE), Confidence Intervals.

1. Introduction

The stress-strength model consists in estimating $R = \Pr(X > Y)$, the lifetime of a component which has a random strength X and it's subjected to random stress Y . In stress-strength model, the system fails if and only if, at any time, the stress is greater than its strength. Birnbaum was first introduced to this model [1] and was developed by Birnbaum and McCarty [2]. There has been a huge number of works in estimation of the reliability $R = P(X > Y)$ in the field of stress-strength models. It has several applications particularly in engineering ideas, like structures, deterioration of rocket motors, static fatigue of ceramic parts, fatigue failure of craft structures, and also in mechanical, civil engineering. The $R = \Pr(X > Y)$ has been formulated for the huge majority of the well-known statistical distributions when X and Y are independent random variables belonging to the univariate family and (X, Y) follows bivariate distribution with dependence between X and Y . The R has been established for the bulk of well-known statistical distributions, including Normal, uniform, exponential, gamma, beta, extreme value, Weibull, Laplace, etc [3-7].

This stress-strength reliability model may also be useful in clinical trial. Particularly when comparing two treatment effects, it may be more useful to draw conclusions regarding the unit's free measure, rather than comparing the means [8]. Simonoff, Hochberg and Reiser also used this function to find the effect of the treatment, if Y is the response for a control group, and X refers to a

treatment group [9]. A numerical procedure obtained by Birnbaum and McCarty based on the asymptotic distribution to find the sample size needed for setting up an upper confidence bound with the defined width and confidence coefficient [2]. Using this procedure Owen, Craswell and Hanson considered the same problem in case of bivariate normal distribution to obtain the sample size needed for specified confidence bound and the confidence coefficient [10]. Sen obtained the non parametric confidence bounds for $P(X<Y)$ based on independent samples [11]. Govindarazulu obtained two-sided confidence intervals for R when X and Y are independent and also dependent normal variates [12]. Church and Harris obtained confidence intervals for R in case of independent normal varieties [13]. Under the same assumptions, Downton derived the minimum variance unbiased estimator (MVUE) of R [14]. They are suggested that an alternative approximation to obtain the “best” estimate of R and its confidence intervals by Church and Harris. Woodward and Kelley obtained the uniformly minimum variance unbiased estimator (UMVUE) of R based on infinite series [15]. Mukherjee and Sharan obtained the UMVUE for R under the bivariate normal distribution [16] and also obtained their asymptotic variance when parameters other than the means are known and they proposed an estimator R based on maximum likelihood estimators when all the five parameters are unknown. Hor and Seal derived an alternative estimator viz. UMVUE of R under the same case of bivariate normal distribution [17].

All these above works were done under the univariate or bivariate setup. Gupta and Gupta first estimated the reliability under multivariate normal setup [18]. They considered the forms of R when $(\mathbf{x}_{p_1 \times 1}, \mathbf{y}_{p_2 \times 1})$ follows multivariate normal distribution with dependence vector between $\mathbf{x}_{p_1 \times 1}$ and $\mathbf{y}_{p_2 \times 1}$. Then, the reliability as $R = \Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$, where \mathbf{a}' and \mathbf{b}' are two vectors. This problem arises when a system in the energy is supplied to the system by p_1 sources and is consumed through p_2 sources and the sources of energy supplied and consumed are linearly dependent with known vector \mathbf{a}' and \mathbf{b}' . Under this set up, they obtained and compared the MVUE and MLE estimate of R with some interesting special cases. Enis and Geisser have demonstrated that, how to obtain the exact confidence bounds for R [19]. In this multivariate setup, Reiser and Farragi derived the lower confidence bounds for $R = P(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$ [20] and solved it iteratively and also derived an approximate lower confidence bounds for R. In a clinical trial require sample size calculations to determine the optimal number of participants (patients) to be included in the trial. Reiser and Guttman introduced the method to obtain the sample size for experiments concerned with inference on R, based on acceptance sampling theory in the univariate normal setup [21].

These two vectors \mathbf{a} and \mathbf{b} are to be chosen such that the multivariate behaviours are approximated by $\mathbf{a}'\mathbf{x}$ and $\mathbf{b}'\mathbf{y}$ as in principal component analysis. Thus, the Principal component analysis used to estimate the \mathbf{a}' and \mathbf{b}' where as Gupta and Gupta considered only spatial cases of \mathbf{a}' and \mathbf{b}' and compare the MVUE and MLE estimates of R using given mean vector and dispersion matrix [18]. The study is carried out on real data set. We do simulation studies to compare the performance of MVUE and MLE in terms of variance (VAR), mean square error (MSE) and mean absolute error (MAE). Then, it is shown that MVUE of R performs better than MLE.

We estimate $R = P(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$ under the multivariate normal setup, whereas Hor and Seal derived this under the bivariate normal distribution setup [17]. We choose some set of $\mu_1, \mu_2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}$ and to compute L_1 distance between the two distribution functions of MVUE and MLE. We may take different choices of parameters to obtain the L_1 distance, where the parameter is $\sqrt{n}\delta = \frac{-\sqrt{n}(\mathbf{b}'\mu_2 - \mathbf{a}'\mu_1)}{(\mathbf{a}'\Sigma_{11}\mathbf{a} - 2\mathbf{a}'\Sigma_{12}\mathbf{b} - \mathbf{b}'\Sigma_{22}\mathbf{b})^{\frac{1}{2}}}$. In this connection, the distributions function of the two estimators MVUE and MLE are derived in section 3. L_1 distance between two functions to compare two estimators is given in section 4. In section 6, we can deal with the problem to obtain two-sided confidence limits and lower bounds for R under the multivariate normal set up. Based on MVUE and MLE, we compare the performance between bootstrap and empirically interval estimator in terms of coverage and accuracy using simulation study. Finally, we applied these interval estimators to a real data set. Finally, in Section 7, we consider the problem of sample size determination.

2. Derivation of the Point Estimation of R

2.1. Maximum Likelihood Estimator of R

Let, $\mathbf{x}_{p_1 \times 1}$ and $\mathbf{y}_{p_2 \times 1}$ be two random vector such that the distribution of $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N_{p_1+p_2}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Where, $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}_{(p_1+p_2) \times 1}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}_{(p_1+p_2) \times (p_1+p_2)}$

Suppose we have known vectors \mathbf{a}' and \mathbf{b}' . Then, we want to find the reliability in terms of linear combination of $\mathbf{a}'\mathbf{x}$ and $\mathbf{b}'\mathbf{y}$ as $R = \Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y}) = \Pr(\mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y} > 0)$

Now, the distribution of $u = \mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y}$ follows $N(\mu_u, \sigma_u^2)$,

where, $\mu_u = E(\mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\mu}_1 - \mathbf{b}'\boldsymbol{\mu}_2$

and $\sigma_u^2 = Var(\mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a} - 2\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b} + \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b}$

Now, $R = \Pr(\mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y} > 0) = \Pr(u > 0)$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{1}{2}\left(\frac{u-\mu_u}{\sigma_u}\right)^2\right\} du = \int_{-\frac{\mu_u}{\sigma_u}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z^2\right\} dz = \Phi\left(\frac{\mu_u}{\sigma_u}\right)$$

The maximum likelihood estimator of $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}_{(p_1+p_2)}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}_{(p_1+p_2) \times (p_1+p_2)}$

define as $\begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{pmatrix}$ and $\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}$ respectively.

We have, $\widehat{\mu}_u = \mathbf{a}'\bar{\mathbf{x}} - \mathbf{b}'\bar{\mathbf{y}}$ and $\widehat{\sigma}_u^2 = \mathbf{a}'\mathbf{S}_{11}\mathbf{a} - 2\mathbf{a}'\mathbf{S}_{12}\mathbf{b} + \mathbf{b}'\mathbf{S}_{22}\mathbf{b}$

So, the maximum likelihood estimate of R is define as $R^* = \Phi\left(\frac{\widehat{\mu}_u}{\widehat{\sigma}_u}\right)$ (1)

2.2. Principal Component Estimation

Let us, compute the estimate of \mathbf{a}' and \mathbf{b}' by Principal component analysis. Principal component analysis explaining the variance-covariance structure $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ of a set of variables \mathbf{x} and \mathbf{y} through a linear combination (\mathbf{a}' and \mathbf{b}') of these variables, i.e, explain maximum variability. It is noted that, the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. We take the maximum likelihood estimate of $\boldsymbol{\Sigma}_{11}$ as \mathbf{S}_{11} and \mathbf{a}' as \mathbf{e}'_1 normalized eigenvectors of \mathbf{S}_{11} corresponding to λ_1 eigen value. Similarly, we have estimate of $\boldsymbol{\Sigma}_{22}$ as \mathbf{S}_{22} and \mathbf{b}' as \mathbf{l}'_1 normalized eigenvectors of \mathbf{S}_{22} corresponding to λ_1 eigen value. Then from (1) the estimate of R define as $R^* = \Phi\left(\frac{\widehat{\mu}_u}{\widehat{\sigma}_u}\right)$, where $\widehat{\mu}_u = \mathbf{e}'_1\bar{\mathbf{x}} - \mathbf{l}'_1\bar{\mathbf{y}}$ and $\widehat{\sigma}_u^2 = \mathbf{e}'_1\mathbf{S}_{11}\mathbf{e}_1 - 2\mathbf{e}'_1\mathbf{S}_{12}\mathbf{l}_1 + \mathbf{l}'_1\mathbf{S}_{22}\mathbf{l}_1$.

2.3. Minimum Variance Unbiased Estimator (MVUE) of R

Now, let us find out Minimum Variance Unbiased Estimator (MVUE) of $R = \Pr(\mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y} > 0)$. Here, it is assumed that the random sample $\begin{pmatrix} \mathbf{x}_\alpha \\ \mathbf{y}_\alpha \end{pmatrix}$, $\alpha = 1, 2, \dots, n$ are from multivariate normal distribution i.e. $\begin{pmatrix} \mathbf{x}_\alpha \\ \mathbf{y}_\alpha \end{pmatrix} \sim N_{p_1+p_2}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Then, $u_\alpha = (\mathbf{b}'\mathbf{y}_\alpha - \mathbf{a}'\mathbf{x}_\alpha) \sim N(\mu_u, \sigma_u^2)$, $\alpha = 1, 2, \dots, n$ be the random sample of size n. Now, (\bar{u}, S_u^2) is a complete sufficient statistic for (μ_u, σ_u^2) , where $\bar{u} = \frac{1}{n}\sum_{\alpha=1}^n u_\alpha$ and $S_u^2 = \frac{1}{n}\sum_{\alpha=1}^n (u_\alpha - \bar{u})^2$, the MVUE of R [22] is

$$\hat{R} = \int_c^1 \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-2}{2})} (1 - z^2)^{\frac{n-2}{2}-1} dz, \text{ where } c = \frac{\bar{u}}{(\sqrt{(n-1)s_u})}$$

Then the MVUE of $R = \Pr(\mathbf{a}'\mathbf{x} - \mathbf{b}'\mathbf{y} > 0)$ [18] is

$$\hat{R} = \begin{cases} 0 & \text{if } c > 1 \\ \frac{1}{2} \left(1 - B\left(c^2; \frac{1}{2}, \frac{n-2}{2}\right) \right) & \text{if } 0 < c \leq 1 \\ \frac{1}{2} \left(1 + B\left((-c)^2; \frac{1}{2}, \frac{n-2}{2}\right) \right) & \text{if } -1 < c \leq 0 \\ 1 & \text{if } c \leq -1 \end{cases} \quad (2)$$

where, $c = \frac{(e'_1 \bar{y} - l'_1 \bar{x})}{(\sqrt{(n-1)}(e'_1 s_{11} e_1 - 2e'_1 s_{12} l_1 + l'_1 s_{22} l_1))}$ and $B(k; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^k x^{\alpha-1} (1-x)^{\beta-1} dx$

2.4. Simulation Study

The simulation study we performed aim to compare the behaviors of two estimators of R, i.e. \hat{R} , the MVUE and $R^* = \Phi\left(\frac{\hat{\mu}_u}{\sigma_u}\right)$, the estimator based on maximum likelihood estimate of $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}_{(p_1, p_2)}$ and

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}_{(p_1+p_2) \times (p_1+p_2)}$$

For this purpose we compute the following measures:

- (i) Mean of \hat{R} and R^*
- (ii) Variance of \hat{R} and R^* : $E(\hat{R} - R)^2$ and $E(R^* - R)^2$
- (iii) Mean square error of \hat{R} and R^* : $\text{Var}(\hat{R}) + \text{Bias}(\hat{R}, R)^2$ and $\text{Var}(R^*) + \text{Bias}(R^*, R)^2$
- (iv) Mean absolute error of \hat{R} and R^* : $E(|\hat{R} - R|)$ and $E(|R^* - R|)$

It is difficult to obtain the analytical form of above expressions for different values of 'R'. So, we figure out these by using simulation study. Hence, we generate the random samples of size n from $\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N_{p_1+p_2}(\mu, \Sigma)$. For each of sample drown of size n, we compute the above measures by taking 500 replications each time.

For this purpose, here, R programming language is used.

Suppose, $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \sim N_5(\mu, \Sigma)$, where, $\mu' = (2, 4, 2, 1, 2)$; $\Sigma = \begin{pmatrix} 3.61 & 2.23 & -0.10 & 0.16 & 2.32 \\ 2.23 & 4.74 & 3.32 & -0.69 & 1.76 \\ -0.10 & 3.32 & 5.68 & -2.34 & -1.23 \\ 0.16 & -0.69 & -2.34 & 3.05 & 1.53 \\ 2.32 & 1.76 & -1.23 & 1.53 & 4.45 \end{pmatrix}$

Therefore, the estimated values of two known vectors based on first principal component are $\mathbf{a}' = (0.614, 0.789)$ and $\mathbf{b}' = (0.737, -0.491, -0.465)$. Now, we want to estimate $R = \Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y})$, then the exact value of stress strength reliability R is 0.886. We take the sample size of n up to 100 in order to achieved exact value for the reliability. Using (1) and (2), calculated mean, variance MSE and MAE of \hat{R} and R^* based on 500 repetitions are reported in Table 1. It can be observed that Variance of \hat{R} is lesser than the Variance of R^* in each sample size. Also, it is noted that the MSE's and MAE's of \hat{R} are less than MSE's and MAE's of R^* . However, the sample mean of \hat{R} is less than the R^* in each case. But, \hat{R} and R^* are under-estimates the true value of R, when sample size are small. It is also interesting to observe that, the Variance, MSE and MAE of \hat{R} and R^* are reduces as the sample size increases: when n=200 and almost achieved the true value of R.

Table 1: Sample Mean, Variance, MSE and MAE of \hat{R} and R^*

Sample Size	Sample Mean		Variance		MSE		MAE	
	\hat{R}	R^*	\hat{R}	R^*	\hat{R}	R^*	\hat{R}	R^*
10	0.672281	0.675052	0.105394	0.109709	0.150902	0.154031	0.247465	0.250765
20	0.730188	0.732219	0.094903	0.096861	0.119021	0.120346	0.190329	0.192359
30	0.755041	0.756586	0.082294	0.083551	0.099306	0.100158	0.152352	0.163332
40	0.793484	0.795067	0.059074	0.059788	0.067533	0.067956	0.11029	0.11728
50	0.804974	0.806287	0.054406	0.054921	0.060878	0.061181	0.097016	0.104556
60	0.812637	0.813826	0.048111	0.048495	0.053412	0.053621	0.909735	0.092666
70	0.838157	0.839244	0.032573	0.032796	0.034807	0.034926	0.066746	0.068149
80	0.844905	0.84585	0.030538	0.030711	0.032174	0.03227	0.062998	0.063118
90	0.84874	0.849622	0.026384	0.026519	0.027727	0.027797	0.055342	0.055777
100	0.855952	0.856766	0.021074	0.02117	0.021941	0.021988	0.04756	0.04864
200	0.881041	0.881482	0.003745	0.003751	0.003763	0.003765	0.018152	0.018332

3. Distribution Function of \hat{R} and R^*

In this section, we derive the distributions function of \hat{R} and R^* . We have

$$R = \begin{cases} 0 & \text{if } c > 1 \\ \frac{1}{2} \left(1 - B \left(c^2; \frac{1}{2}, \frac{n-2}{2} \right) \right) & \text{if } 0 < c \leq 1 \\ \frac{1}{2} \left(1 + B \left((-c)^2; \frac{1}{2}, \frac{n-2}{2} \right) \right) & \text{if } -1 < c \leq 0 \\ 1 & \text{if } c \leq -1 \end{cases}$$

where $B(x; \frac{1}{2}, \frac{n-2}{2})$ is c.d.f $B(\frac{1}{2}, \frac{n-2}{2})$ and it is clear that $0 \leq \hat{R} \leq 1$ for any real number of c.

Let the distribution function of R be $F_R(x)$, then

$$F_R(x) = 0 \text{ if } x < 0 \text{ and } F_R(x) = 1 \text{ if } x \geq 1$$

If $0 \leq x \leq \frac{1}{2}$, then the distribution function of \hat{R} is given by

$$\begin{aligned} F_{\hat{R}}(x) &= P(\hat{R} \leq x) = P[(\hat{R} \leq x) \cap \{(c > 0) \cup (c \leq 0)\}] = P[(\hat{R} \leq x) \cap (c > 0)] + P[(\hat{R} \leq x) \cap (c \leq 0)] \\ &= P[(\hat{R} \leq x) | (c > 0)]P[c > 0] + P[(\hat{R} \leq x) | (c \leq 0)]P[c \leq 0] \\ &= P\left[\frac{1}{2} - \frac{1}{2}B\left(c^2; \frac{1}{2}, \frac{n-2}{2}\right) \leq x\right]P\left[\frac{(b'\bar{y} - a'\bar{x})}{\sqrt{(n-1)(a'\Sigma_{11}a - 2a'\Sigma_{12}b + b'\Sigma_{22}b)}} > 0\right] \\ &= P\left[B\left(c^2; \frac{1}{2}, \frac{n-2}{2}\right) \geq 1 - 2x\right]P[(b'\bar{y} - a'\bar{x}) > 0] \end{aligned}$$

$$\text{, where } (b'\bar{y} - a'\bar{x}) \sim N_1((b'\mu_2 - a'\mu_1), \frac{1}{n}(a'\Sigma_{11}a - 2a'\Sigma_{12}b + b'\Sigma_{22}b))$$

$$\begin{aligned} &= P\left[c^2 \geq B_{\left(\frac{1}{2}, \frac{n-2}{2}\right)}^{-1}(1 - 2x)\right] \Phi\left(\frac{\sqrt{n}(b'\mu_2 - a'\mu_1)}{(a'\Sigma_{11}a - 2a'\Sigma_{12}b + b'\Sigma_{22}b)^{\frac{1}{2}}}\right) \\ &= \left\{P\left[c \geq \left(B_{\left(\frac{1}{2}, \frac{n-2}{2}\right)}^{-1}(1 - 2x)\right)^{\frac{1}{2}}\right] + P\left[c \leq -\left(B_{\left(\frac{1}{2}, \frac{n-2}{2}\right)}^{-1}(1 - 2x)\right)^{\frac{1}{2}}\right]\right\} \Phi(-\sqrt{n}\delta) \end{aligned}$$

$$\text{, where } \delta = \frac{-(b'\mu_2 - a'\mu_1)}{(a'\Sigma_{11}a - 2a'\Sigma_{12}b + b'\Sigma_{22}b)^{\frac{1}{2}}}$$

$$\begin{aligned}
 &= \{P[\frac{(b'\bar{y} - a'\bar{x})}{\sqrt{(n-1)(a'S_{11}a - 2a'S_{12}b + b'S_{22}b)}^{\frac{1}{2}}} \geq (B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}] + \\
 &P[\frac{(b'\bar{y} - a'\bar{x})}{\sqrt{(n-1)(a'S_{11}a - 2a'S_{12}b + b'S_{22}b)}^{\frac{1}{2}}} \leq -(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}]\} \Phi(-\sqrt{n}\delta) \\
 &= \{P[-\sqrt{n}\hat{\delta} \geq (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}] + P[-\sqrt{n}\hat{\delta} \leq -(n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}]\} \Phi(-\sqrt{n}\delta) \\
 &\quad , \text{where } \hat{\delta} = \frac{-\sqrt{(n-1)}(b'\bar{y} - a'\bar{x})}{\sqrt{n}(a'S_{11}a - 2a'S_{12}b + b'S_{22}b)^{\frac{1}{2}}} \\
 &= \{P[\sqrt{n}\hat{\delta} \leq -(n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}] + 1 - P[\sqrt{n}\hat{\delta} \leq (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}]\} \Phi(-\sqrt{n}\delta) \\
 &= \{F_{t'_{(n-1), \sqrt{n}\hat{\delta}}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}) + \\
 &[1 - F_{t'_{(n-1), \sqrt{n}\hat{\delta}}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}})]\} \Phi(-\sqrt{n}\delta) \tag{3}
 \end{aligned}$$

Using the standard distribution theory [23], if $x \sim N_p(\mu, \Sigma)$ then $a'x \sim N_1(a'\mu, a'\Sigma a)$. Let, \bar{x} and S be the unbiased estimator of μ and Σ respectively, then $a'\bar{x} \sim N_1(a'\mu, \frac{1}{n}(a'\Sigma a))$ and $S \sim W_p(n-1, \Sigma)$. Thus, we can write, $\frac{a'Sa}{a'\Sigma a} \sim \chi_{n-1}^2$, hence $\sqrt{n-1}\hat{\delta} \sim t'_{(n-1), \sqrt{n}\hat{\delta}}$ where $t'_{(n-1), \sqrt{n}\hat{\delta}}$ denotes the non-central t-distribution with $(n-1)$ d.f. We use the unbiased estimator of Σ instead of MLE, then $\sqrt{n}\hat{\delta} \sim t'_{(n-1), \sqrt{n}\hat{\delta}}$ with non-centrality parameter $\sqrt{n}\delta$ and $F_{t'_{(n-1), \sqrt{n}\hat{\delta}}}(\cdot)$ be the cdf of non-central t-distribution.

If $\frac{1}{2} < x < 1$, then the distribution function of \hat{R} is given by

$$\begin{aligned}
 F_{\hat{R}}(x) &= P(\hat{R} \leq x) = P[(\hat{R} \leq x) \cap \{(c > 0) \cup (c \leq 0)\}] = P[(\hat{R} \leq x) \cap (c > 0)] + P[(\hat{R} \leq x) \cap (c \leq 0)] \\
 &= P[(\hat{R} \leq x) | (c > 0)]P[c > 0] + P[(\hat{R} \leq x) | (c \leq 0)]P[c \leq 0] \\
 &= P[\frac{1}{2} - \frac{1}{2}B(c^2; \frac{1}{2}, \frac{n-2}{2}) \leq x]P[\frac{(b'\bar{y} - a'\bar{x})}{\sqrt{(n-1)(a'S_{11}a - 2a'S_{12}b + b'S_{22}b)}^{\frac{1}{2}}} > 0] + \\
 &P[\frac{1}{2} + \frac{1}{2}B((-c)^2; \frac{1}{2}, \frac{n-2}{2}) \leq x] \Phi(\sqrt{n}\delta) \\
 &= \Phi(-\sqrt{n}\delta) + P[B(c^2; \frac{1}{2}, \frac{n-2}{2}) \leq 2x - 1] \Phi(\sqrt{n}\delta) \\
 &= \Phi(-\sqrt{n}\delta) + P[c^2 \leq B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1)] \Phi(\sqrt{n}\delta) \\
 &= \Phi(-\sqrt{n}\delta) + P[-(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}} \leq c \leq (B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}}] \Phi(\sqrt{n}\delta) \\
 &= \Phi(-\sqrt{n}\delta) + P[-(n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}} \leq -\sqrt{n}\hat{\delta} \leq (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}}] \Phi(\sqrt{n}\delta) \\
 &= \Phi(-\sqrt{n}\delta) + P[(n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}} \geq \sqrt{n}\hat{\delta} \geq -(n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}}] \Phi(\sqrt{n}\delta) \\
 &= \Phi(-\sqrt{n}\delta) + [F_{t'_{(n-1), \sqrt{n}\hat{\delta}}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x - 1))^{\frac{1}{2}}) -
 \end{aligned}$$

$$F_{t'_{(n-1),\sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}})]\Phi(\sqrt{n}\delta)$$

Thus, the distribution function of \hat{R} is given by

$$F_{\hat{R}}(x) = 0 \text{ if } x < 0 \text{ and } F_{\hat{R}}(x) = 1 \text{ if } x \geq 1$$

If $0 \leq x \leq \frac{1}{2}$,

$$F_{\hat{R}}(x) = \{F_{t'_{(n-1),\sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}) + [1 - F_{t'_{(n-1),\sqrt{n}\delta}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}})]\}\Phi(-\sqrt{n}\delta)$$

If $\frac{1}{2} < x < 1$,

$$F_{\hat{R}}(x) = \Phi(-\sqrt{n}\delta) + [F_{t'_{(n-1),\sqrt{n}\delta}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}}) - F_{t'_{(n-1),\sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}})]\Phi(\sqrt{n}\delta) \quad (4)$$

The MLE estimate of R, $R^* = \Phi\left(\frac{-(\mathbf{b}'\bar{\mathbf{y}} - \mathbf{a}'\bar{\mathbf{x}})}{(\mathbf{a}'\mathbf{S}_{11}\mathbf{a} - 2\mathbf{a}'\mathbf{S}_{12}\mathbf{b} + \mathbf{b}'\mathbf{S}_{22}\mathbf{b})^{\frac{1}{2}}}\right)$

Let the distribution function of R^* be $F_{R^*}(x)$, then $F_{R^*}(x) = 0$ if $x < 0$ and $F_{R^*}(x) = 1$ if $x \geq 1$,

$$\begin{aligned} F_{R^*}(x) &= P(R^* \leq x) = P\left[\Phi\left(\frac{-(\mathbf{b}'\bar{\mathbf{y}} - \mathbf{a}'\bar{\mathbf{x}})}{(\mathbf{a}'\mathbf{S}_{11}\mathbf{a} - 2\mathbf{a}'\mathbf{S}_{12}\mathbf{b} + \mathbf{b}'\mathbf{S}_{22}\mathbf{b})^{\frac{1}{2}}}\right) \leq x\right] \\ &= P\left[\frac{-(\mathbf{b}'\bar{\mathbf{y}} - \mathbf{a}'\bar{\mathbf{x}})}{(\mathbf{a}'\mathbf{S}_{11}\mathbf{a} - 2\mathbf{a}'\mathbf{S}_{12}\mathbf{b} + \mathbf{b}'\mathbf{S}_{22}\mathbf{b})^{\frac{1}{2}}} \leq \Phi^{-1}(x)\right] = P[\sqrt{n}\hat{\delta} \leq \sqrt{(n-1)}\Phi^{-1}(x)] \\ &= F_{t'_{(n-1),\sqrt{n}\delta}}(\sqrt{(n-1)}\Phi^{-1}(x)) \end{aligned} \quad (5)$$

4. Distance Between $F_{\hat{R}}(\cdot)$ and $F_{R^*}(\cdot)$

Let us calculate the distance between two functions (L_1) $F_{\hat{R}}(x)$ and $F_{R^*}(x)$ [17],

$$U(n, \delta) = \int_0^1 |F_{R^*}(x) - F_{\hat{R}}(x)| dx, \text{ for different values of } n, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22}.$$

According to this method, this can be taken as a measure of deviation or equal deviation between \hat{R} and R^* .

Let, \bar{R} be any other estimator of R, then the maximum deviation between distribution of \bar{R} and \hat{R} as

$$M(n, \delta) = \sup_{\bar{R}} \int_0^1 |F_{\bar{R}}(x) - F_{\hat{R}}(x)| dx = \int_0^1 F_{\hat{R}}(x) dx, \text{ if } \int_0^1 F_{\bar{R}}(x) dx > \frac{1}{2}$$

$$= 1 - \int_0^1 F_R^\wedge(x) dx, \text{ if } \int_0^1 F_R^\wedge(x) dx \leq \frac{1}{2}$$

The ratio $R(n, \delta) = \frac{U(n, \delta)}{M(n, \delta)}$ has to be taken as a relative measure of deviation between \hat{R} and R^* , the maximum deviation between any other estimator of R, i.e. \bar{R} and \hat{R} . It's difficult to get the exact expression of this above measures. So, we compute these measure values numerically using R-programming. Here we take different choices of δ for $n = 20$. The results are reported in table 2. The results show that the overall output of MVUE and MLE of R are not too distant and the values for these differences are show in columns U(.), M(.) and R(.) of this table. From this table, it is seen that empirical values of the parameters and the performance of MVUE of R is better than MLE and also Figure 1 shows that, MVUE estimator of R is better than the other estimators, i.e. R^* . Also, L_1 distance and graphical impression show this.

5. Derivation of $Var(\hat{R})$ and $MSE(R^*)$

Now, $Var(\hat{R})$ and $MSE(R^*)$ are obtained by using equations (3), (4) and (5) as follows:

Since, $0 < \hat{R} < 1$, then $E(\hat{R}) = \int_0^1 \{1 - F_R^\wedge(x)\} dx$

If $0 \leq x \leq \frac{1}{2}$, then we have

$$E_1(\hat{R}) = \int_0^1 [1 - \{F_{t'_{(n-1), \sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}) + [1 - F_{t'_{(n-1), \sqrt{n}\delta}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}})]\} \Phi(-\sqrt{n}\delta)] dx$$

If $\frac{1}{2} < x < 1$, then we have

$$E_2(\hat{R}) = \int_0^1 [1 - \{\Phi(-\sqrt{n}\delta) + [F_{t'_{(n-1), \sqrt{n}\delta}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}}) - F_{t'_{(n-1), \sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}})]\} \Phi(\sqrt{n}\delta)] dx$$

So, $Var(\hat{R}) = \int_0^1 2x\{1 - F_R^\wedge(x)\} dx - \{E(\hat{R})\}^2 = \int_0^1 2x[1 - \{F_{t'_{(n-1), \sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}}) + [1 - F_{t'_{(n-1), \sqrt{n}\delta}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(1-2x))^{\frac{1}{2}})]\} \Phi(-\sqrt{n}\delta)] dx + \int_0^1 2x[1 - \{\Phi(-\sqrt{n}\delta) + [F_{t'_{(n-1), \sqrt{n}\delta}}((n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}}) - F_{t'_{(n-1), \sqrt{n}\delta}}(- (n-1)(B_{(\frac{1}{2}, \frac{n-2}{2})}^{-1}(2x-1))^{\frac{1}{2}})]\} \Phi(\sqrt{n}\delta)] dx - \{E_1(\hat{R}) + E_2(\hat{R})\}^2$

Similarly, we can determine $MSE(R^*) = \int_0^1 2x\{1 - F_{t'_{(n-1), \sqrt{n}\delta}}(\sqrt{(n-1)}\Phi^{-1}(x))\} dx -$

$$[\int_0^1 \{1 - F_{t'_{(n-1), \sqrt{n}\delta}}(\sqrt{(n-1)}\Phi^{-1}(x))\} dx]^2$$

From Figure 2 and 3 it is observe that the, values of $\hat{\text{Var}}(R)$ and $\text{MSE}(R^*)$ are almost close to zero of δ . Values of $\hat{\text{Var}}(R)$ are less as compared to other values of $\text{MSE}(R^*)$. Thus, the performance of MVUE of R is better than MLE.

Table 2: Performance of point estimators: δ and $100^* \{U(n, \delta), M(n, \delta), R(n, \delta)\}$

Non-negative values of δ				Negative values of δ			
δ	$U(n, \delta)$	$M(n, \delta)$	$R(n, \delta)$	δ	$U(n, \delta)$	$M(n, \delta)$	$R(n, \delta)$
3	0.087	99.865	0.087	-3	0.087	99.865	0.087
2.898	0.101	99.812	0.101	-2.898	0.101	99.812	0.101
2.797	0.115	99.742	0.115	-2.797	0.115	99.742	0.115
2.695	0.128	99.648	0.129	-2.695	0.128	99.648	0.129
2.593	0.141	99.525	0.141	-2.593	0.141	99.525	0.141
2.492	0.151	99.364	0.152	-2.492	0.151	99.364	0.152
2.39	0.157	99.157	0.158	-2.39	0.157	99.157	0.158
2.288	0.158	98.894	0.16	-2.288	0.158	98.894	0.16
2.186	0.153	98.561	0.155	-2.186	0.153	98.561	0.155
2.085	0.139	98.145	0.142	-2.085	0.139	98.145	0.142
1.983	0.116	97.632	0.119	-1.983	0.116	97.632	0.119
1.881	0.082	97.004	0.084	-1.881	0.082	97.004	0.084
1.78	0.036	96.243	0.038	-1.78	0.036	96.243	0.038
1.678	0.021	95.332	0.022	-1.678	0.021	95.332	0.022
1.576	0.089	94.252	0.094	-1.576	0.089	94.252	0.094
1.475	0.166	92.984	0.179	-1.475	0.166	92.984	0.179
1.373	0.25	91.511	0.273	-1.373	0.25	91.511	0.273
1.271	0.338	89.817	0.376	-1.271	0.338	89.817	0.376
1.169	0.424	87.89	0.482	-1.169	0.424	87.89	0.482
1.068	0.503	85.719	0.587	-1.068	0.503	85.719	0.587
0.966	0.571	83.3	0.685	-0.966	0.571	83.3	0.685
0.864	0.623	80.629	0.772	-0.864	0.623	80.629	0.772
0.763	0.663	77.702	0.853	-0.763	0.663	77.702	0.853
0.661	0.715	74.501	0.959	-0.661	0.715	74.501	0.959
0.559	0.843	70.977	1.188	-0.559	0.843	70.977	1.188
0.458	1.144	67.052	1.706	-0.458	1.144	67.052	1.706
0.356	1.635	62.739	2.606	-0.356	1.635	62.739	2.606
0.254	2.057	58.331	3.527	-0.254	2.057	58.331	3.527
0.153	1.877	54.406	3.45	-0.153	1.877	54.406	3.45
0.051	0.779	51.323	1.518	-0.051	0.779	51.323	1.518

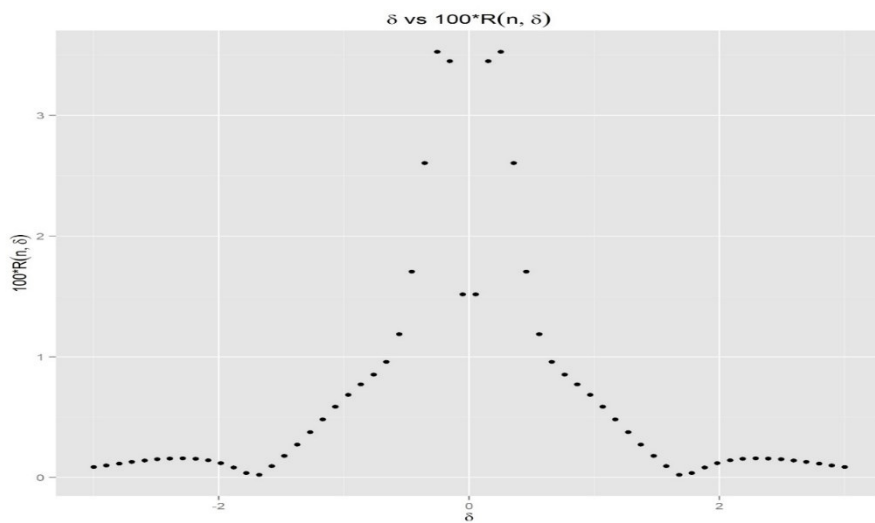


Figure 1: δ vs $100 * R(n, \delta)$

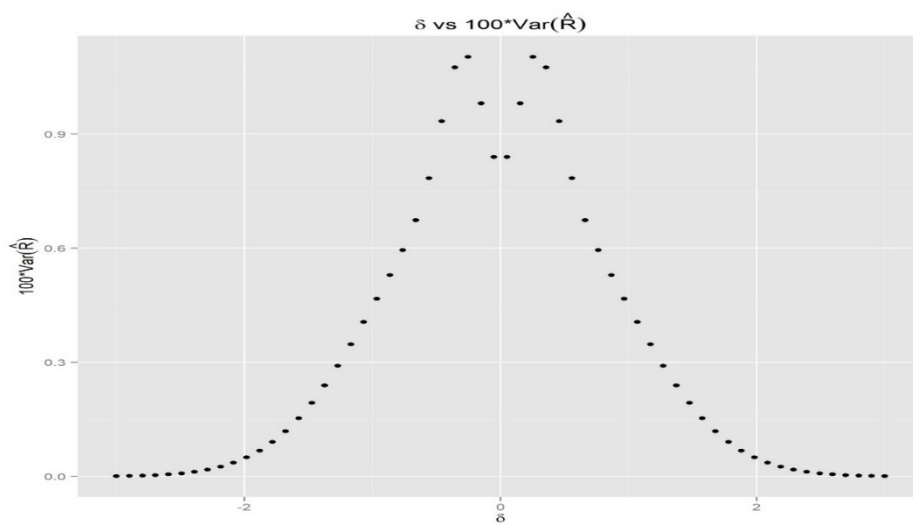


Figure 2: δ vs $100 * Var(\hat{R})$

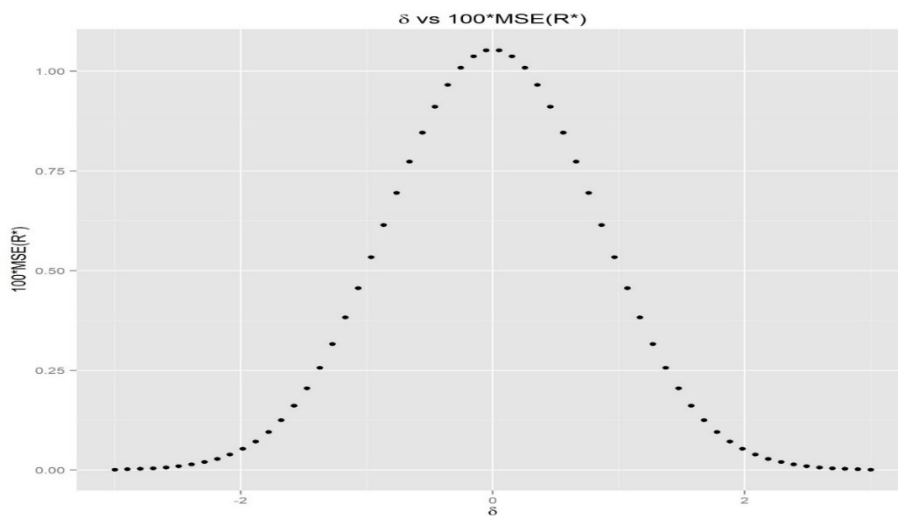


Figure 3: δ vs $100 * MSE(R^*)$

6. Confidence Intervals for R

6.1. Exact Two Sided Confidence Intervals for R

Here, we have $\sqrt{n}\hat{\delta} \sim t'_{(n-1),\sqrt{n}\hat{\delta}}$, then the tradition approach for finding the lower limit of R , we use the probability p_{δ_L} that $t'_{(n-1),\sqrt{n}\hat{\delta}}$ exceeds the value of $\sqrt{n}\hat{\delta}$ as

$$p_{\delta_L} = \Pr (t'_{(n-1),\sqrt{n}\hat{\delta}} > \sqrt{n}\hat{\delta}) = \alpha/2$$

or,

$$p_{\delta_L} = \Pr (t'_{(n-1),\sqrt{n}\hat{\delta}} < \sqrt{n}\hat{\delta}) = 1 - \alpha/2 \tag{6}$$

Similarly, we get the upper limit of δ as

$$p_{\delta_U} = \Pr (t'_{(n-1),\sqrt{n}\hat{\delta}} < \sqrt{n}\hat{\delta}) = \alpha/2 \tag{7}$$

Equation (6) and (7) can be solved numerically. Finally we get the $(1-\alpha)$ level confidence Intervals for δ as (δ_L, δ_U) .

Then, the $(1-\alpha)$ level confidence Intervals for R as $(\Phi(\delta_L), \Phi(\delta_U))$.

6.2. Exact Lower Confidence bound for R

In order to obtain the lower bound of the lower bound of R, we use the probability $p_{\delta_{LB}}$ that $t'_{(n-1),\sqrt{n}\hat{\delta}}$ exceeds the value of $\sqrt{n}\hat{\delta}$ as

$$p_{\delta_{LB}} = \Pr (t'_{(n-1),\sqrt{n}\hat{\delta}} > \sqrt{n}\hat{\delta}) = \alpha$$

or,

$$p_{\delta_{LB}} = \Pr (t'_{(n-1),\sqrt{n}\hat{\delta}} < \sqrt{n}\hat{\delta}) = 1 - \alpha \tag{8}$$

Thus, the $(1-\alpha)$ level confidence lower bound for δ can be obtained by solving equation (8). Then, the $(1-\alpha)$ level confidence lower bound for R is $(\Phi(\delta_{LB}))$.

6.3. Approximate Two Sided Confidence Intervals for R

From section 3, we have $\Pr(\mathbf{a}'\mathbf{x} > \mathbf{b}'\mathbf{y}) = \Phi \left[\frac{-(\mathbf{b}'\boldsymbol{\mu}_2 - \mathbf{a}'\boldsymbol{\mu}_1)}{(\mathbf{a}'\boldsymbol{\Sigma}_{11}\mathbf{a} - 2\mathbf{a}'\boldsymbol{\Sigma}_{12}\mathbf{b} - \mathbf{b}'\boldsymbol{\Sigma}_{22}\mathbf{b})^{\frac{1}{2}}} \right] = \Phi(\delta)$,

where $\sqrt{n}\hat{\delta} \sim t'_{(n-1),\sqrt{n}\hat{\delta}}$ with non-centrality parameter $\sqrt{n}\hat{\delta}$

In order to determine the two sided confidence Intervals, we use following well known approximation for large n [24] as

$$Z = \frac{[t'_{(n-1),\sqrt{n}\delta} - \sqrt{n}\delta]}{\left[1 + \frac{(t'_{(n-1),\sqrt{n}\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}} \sim N(0,1)$$

Using this,

$$Pr[-z_{\alpha/2} \leq \frac{[\sqrt{n}\hat{\delta} - \sqrt{n}\delta]}{\left[1 + \frac{(\sqrt{n}\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}} \leq z_{\alpha/2}] = 1-\alpha$$

or,

$$Pr[-z_{\alpha/2} \leq \frac{[\hat{\delta} - \delta]}{\left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}} \leq z_{\alpha/2}] = 1-\alpha$$

or,

$$Pr\left[\hat{\delta} - z_{\alpha/2} \left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}} \leq \delta \leq \hat{\delta} + z_{\alpha/2} \left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}\right] = 1-\alpha$$

Thus, an approximate (1- α) level confidence Intervals for δ is given by

$$(\delta_L, \delta_U) = \left\{ \hat{\delta} - z_{\alpha/2} \left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}, \hat{\delta} + \left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}} z_{\alpha/2} \right\}$$

Then, an approximate (1- α) level confidence Intervals for R is represented by

$$(\Phi(\delta_L), \Phi(\delta_U)) = \left\{ \Phi\left(\hat{\delta} - z_{\alpha/2} \left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}\right), \Phi\left(\hat{\delta} + z_{\alpha/2} \left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}\right) \right\}$$

Where, $z_{\alpha/2}$ upper critical value for the standard normal distribution

6.4. Approximate Lower Confidence bound for R

The lower bounds based on approximate results is given by

$$Pr(\delta_{LB} \leq \delta) = 1 - \alpha$$

or,

$$Pr\left(\frac{[\delta_{LB} - \hat{\delta}]}{\left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}} \leq \frac{[\delta - \hat{\delta}]}{\left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}}\right) = 1 - \alpha$$

or,

$$Pr\left(\frac{[\hat{\delta} - \delta]}{\left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}} \leq \frac{[\hat{\delta} - \delta_{LB}]}{\left[\frac{1}{n} + \frac{(\hat{\delta})^2}{2(n-1)}\right]^{\frac{1}{2}}}\right) = 1 - \alpha$$

$$\text{or, } Pr \left(z \leq \frac{\left[\hat{\delta} - \delta_{LB} \right]}{\left[\frac{1}{n} + \frac{\hat{(\delta)}^2}{2(n-1)} \right]^{\frac{1}{2}}} \right) = 1 - \alpha, \quad \text{or, } \frac{\left[\hat{\delta} - \delta_{LB} \right]}{\left[\frac{1}{n} + \frac{\hat{(\delta)}^2}{2(n-1)} \right]^{\frac{1}{2}}} = z_{1-\alpha}$$

$$\text{or, } \delta_{LB} = \hat{\delta} - z_{1-\alpha} \left[\frac{1}{n} + \frac{\hat{(\delta)}^2}{2(n-1)} \right]^{\frac{1}{2}}$$

So, an approximate (1-α) confidence lower bound for R as

$$\Phi(\delta_{LB}) = \Phi \left(\hat{\delta} - z_{1-\alpha} \left[\frac{1}{n} + \frac{\hat{(\delta)}^2}{2(n-1)} \right]^{\frac{1}{2}} \right)$$

6.5. Bootstrap confidence Intervals for R

In this subsection, we use the confidence intervals based on percentile bootstrap method. Efron suggests the procedure to find out the confidence intervals for a parameter [25] and corrects bias of percentile of bootstrap confidence intervals for R proposed by Efron [26]. It works as follows

- (1) Draw random sample $\begin{pmatrix} \mathbf{X}_\alpha \\ \mathbf{Y}_\alpha \end{pmatrix}$, $\alpha = 1, 2, \dots, n$ from multivariate normal distribution, where $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{p_1+p_2}(\mu, \Sigma)$
- (2) Generate bootstrap samples $\begin{pmatrix} \mathbf{X}_\alpha^* \\ \mathbf{Y}_\alpha^* \end{pmatrix}$, $\alpha = 1, 2, \dots, n$, by using random sample of $\begin{pmatrix} \mathbf{X}_\alpha \\ \mathbf{Y}_\alpha \end{pmatrix}$, $\alpha = 1, 2, \dots, n$.
- (3) Compute the bootstrap estimates of linear dependent vectors \mathbf{a}' and \mathbf{b}' using PC1, say \mathbf{e}^{*} and \mathbf{l}^{*} respectively. Also, Compute the bootstrap MLE estimates of $\mu_1, \mu_2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}$ by $\bar{\mathbf{x}}^*, \bar{\mathbf{y}}^*, \mathbf{S}_{11}^*, \mathbf{S}_{12}^*, \mathbf{S}_{22}^*$. Using these estimates compute the bootstrap estimate of R, say R_B^* .
- (4) Repeat steps 2 and 3, number of boot time B (B sufficiently large, i.e. 1000), thus we obtain the bootstrap distribution of $\{R_B^*\}$.
- (5) Estimate (1 - α) bootstrap percentile confidence intervals for R from $\{R_B^*\}$ by taking the $\left(\frac{\alpha}{2}\right)$ and $\left(1 - \frac{\alpha}{2}\right)$ quantiles as $\left(R_{B, \frac{\alpha}{2}}^*, R_{B, \left(1 - \frac{\alpha}{2}\right)}^*\right)$.
 or, (1 - α) bootstrap percentile lower bound for R as $R_{B, \alpha}^*$.

6.5. Simulation Study

In this section, we present simulation study to investigate the statistical properties of the interval estimators using the given matrix in section 2.4. The simulation study define as follows

- (1) Draw the random samples of size n from $\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{p_1+p_2}(\mu, \Sigma)$. For each of sample drawn of size n, considered different sample sizes (n=50, 100, 150,....etc). We compute the above measures by taking 500 replications each time.
- (2) Estimate MLE estimate of R using PC1 for different sample size and a Confidence Intervals (two-sided and lower bound) for R.
- (3) Compute exact, approximate and bootstrap confidence intervals using step 2, where number of boot time B=1000.

The results of the simulation study are recorded in Table 3-5. Figure 4-6, represent the exact, approximate and bootstrap confidence belt at 90%, 95% and 99% levels. It has been observed that for

a small sample size, the estimate of R is getting high and also confidence intervals. The results get better as the sample sizes increase and the reliability R gets closer to true value. The same phenomenon is observed for the exact, approximate and bootstrap confidence intervals. The overall band of exact and approximate confidence intervals is almost same, whereas bootstrap confidence intervals give the large confidence band for small sample size. But, exact, approximate confidence intervals and Bootstrap confidence intervals all are almost same for large sample size at 90%, 95% and 99% levels. All most the same variation found in confidence belt of exact, approximate CIs, but irregular variation in bootstrap CIs shows in Figure 4-6.

Table 3: Exact Confidence Intervals

Sample size	R*	90%			95%			99%		
		L	U	LB	L	U	LB	L	U	LB
50	0.8954	0.8738	0.9349	0.8820	0.8664	0.9394	0.8738	0.8511	0.9473	0.8574
100	0.8953	0.8782	0.9227	0.8838	0.8731	0.9262	0.8782	0.8629	0.9328	0.8671
150	0.8948	0.8785	0.9139	0.8828	0.8746	0.9168	0.8785	0.8669	0.9224	0.8701
200	0.8898	0.8751	0.9128	0.8797	0.8710	0.9159	0.8751	0.8627	0.9218	0.8661
250	0.8888	0.8743	0.9197	0.8800	0.8692	0.9233	0.8743	0.8588	0.9300	0.8631
300	0.8887	0.8732	0.9133	0.8782	0.8688	0.9166	0.8732	0.8599	0.9228	0.8636
350	0.8883	0.8703	0.9070	0.8747	0.8663	0.9101	0.8703	0.8583	0.9159	0.8616
400	0.8876	0.8653	0.9068	0.8704	0.8608	0.9103	0.8653	0.8517	0.9167	0.8554
450	0.8874	0.8677	0.9048	0.8722	0.8637	0.9079	0.8677	0.8556	0.9138	0.8589
500	0.8874	0.8676	0.9047	0.8721	0.8636	0.9079	0.8676	0.8556	0.9138	0.8589
550	0.8868	0.8688	0.9028	0.8729	0.8652	0.9057	0.8688	0.8579	0.9112	0.8609
600	0.8864	0.8640	0.9028	0.8729	0.8595	0.9057	0.8688	0.8503	0.9111	0.8609

Table 4: Approximate Confidence Intervals

Sample size	90%			95%			99%		
	L	U	LB	L	U	LB	L	U	LB
50	0.8741	0.9352	0.8823	0.8666	0.9395	0.8741	0.8512	0.9474	0.8576
100	0.8783	0.9228	0.8840	0.8732	0.9263	0.8783	0.8629	0.9328	0.8672
150	0.8786	0.9139	0.8829	0.8747	0.9169	0.8786	0.8669	0.9224	0.8701
200	0.8752	0.9129	0.8798	0.8711	0.9160	0.8752	0.8627	0.9218	0.8661
250	0.8744	0.9198	0.8802	0.8693	0.9234	0.8744	0.8589	0.9300	0.8632
300	0.8733	0.9134	0.8783	0.8689	0.9167	0.8733	0.8600	0.9228	0.8636
350	0.8704	0.9071	0.8748	0.8664	0.9101	0.8704	0.8584	0.9159	0.8616
400	0.8655	0.9069	0.8706	0.8609	0.9103	0.8655	0.8517	0.9167	0.8555
450	0.8678	0.9049	0.8723	0.8637	0.9080	0.8678	0.8556	0.9138	0.8590
500	0.8677	0.9048	0.8722	0.8637	0.9079	0.8677	0.8556	0.9138	0.8589
550	0.8689	0.9029	0.8730	0.8653	0.9057	0.8689	0.8580	0.9112	0.8609
600	0.8689	0.9028	0.8730	0.8652	0.9057	0.8689	0.8579	0.9112	0.8609

Table 5: Bootstrap Confidence Intervals

Sample size	90%			95%			99%		
	L	U	LB	L	U	LB	L	U	LB
50	0.8591	0.9592	0.8701	0.8495	0.9688	0.8591	0.8307	0.9876	0.8383
100	0.8804	0.9264	0.8855	0.8760	0.9308	0.8804	0.8674	0.9395	0.8709
150	0.8806	0.9149	0.8844	0.8773	0.9182	0.8806	0.8709	0.9246	0.8735
200	0.8518	0.9419	0.8617	0.8432	0.9505	0.8518	0.8263	0.9674	0.8331
250	0.8787	0.9188	0.8832	0.8749	0.9227	0.8787	0.8674	0.9302	0.8704
300	0.8757	0.9137	0.8799	0.8720	0.9174	0.8757	0.8649	0.9245	0.8678
350	0.8218	0.9629	0.8373	0.8082	0.9764	0.8218	0.7818	0.9896	0.7925
400	0.8676	0.9073	0.8720	0.8638	0.9111	0.8676	0.8564	0.9185	0.8594

STRESS-STRENGTH RELIABILITY MODEL UNDER MULTIVARIATE

450	0.8695	0.9059	0.8735	0.8660	0.9094	0.8695	0.8592	0.9162	0.8619
500	0.8692	0.9045	0.8731	0.8658	0.9079	0.8692	0.8592	0.9145	0.8619
550	0.8703	0.9038	0.8740	0.8671	0.9070	0.8703	0.8608	0.9133	0.8633
600	0.8705	0.9033	0.8741	0.8674	0.9065	0.8705	0.8612	0.9126	0.8637

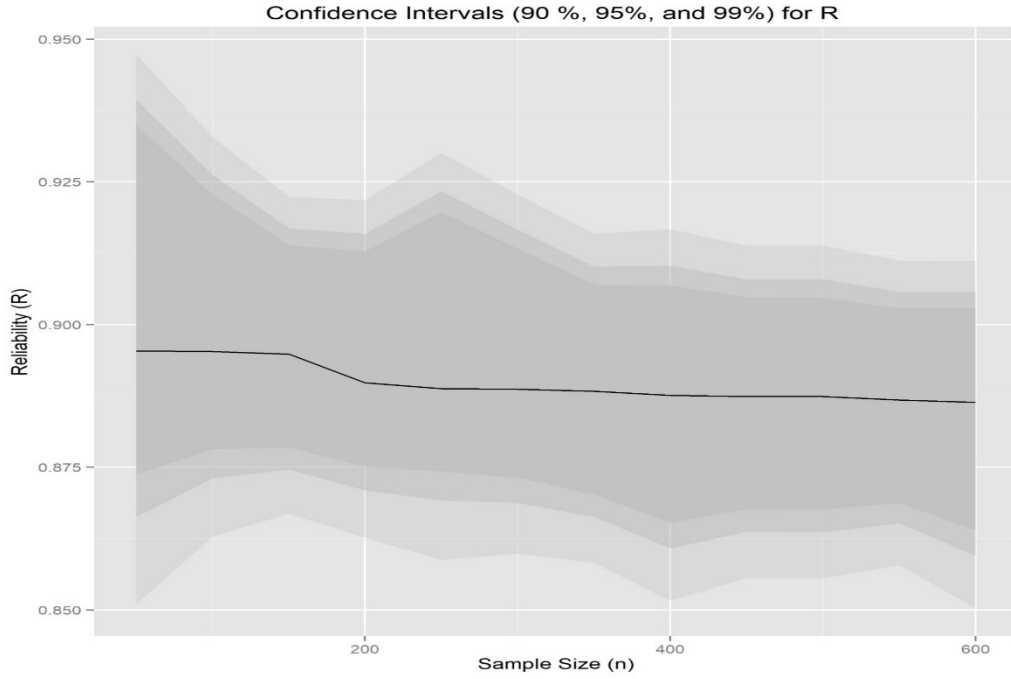


Figure 4: Exact Confidence Intervals

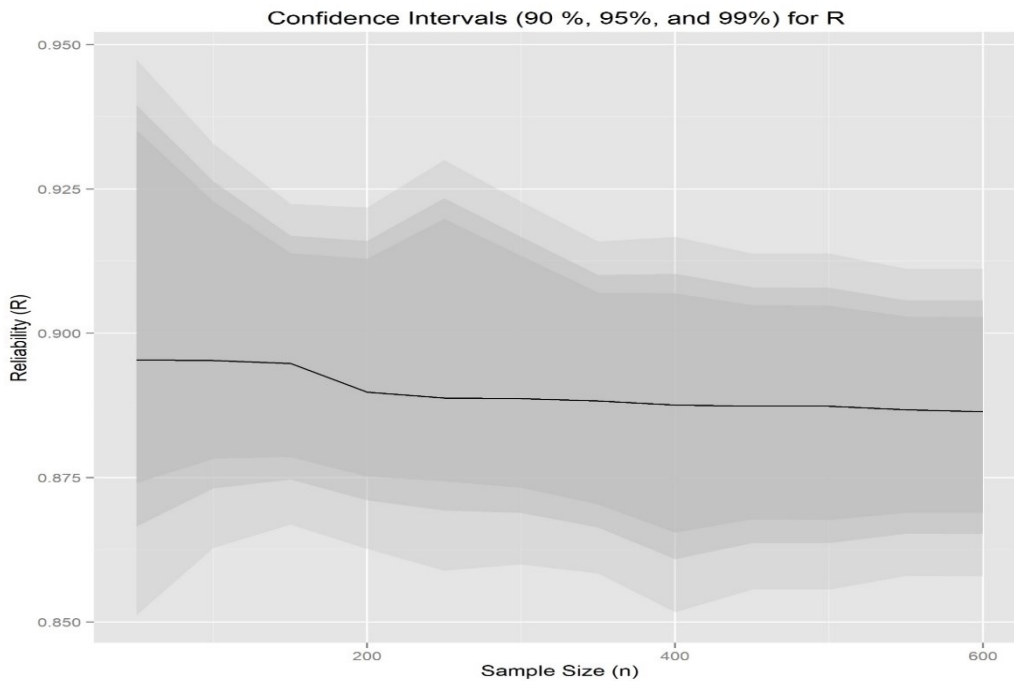


Figure 5: Approximate Confidence Intervals

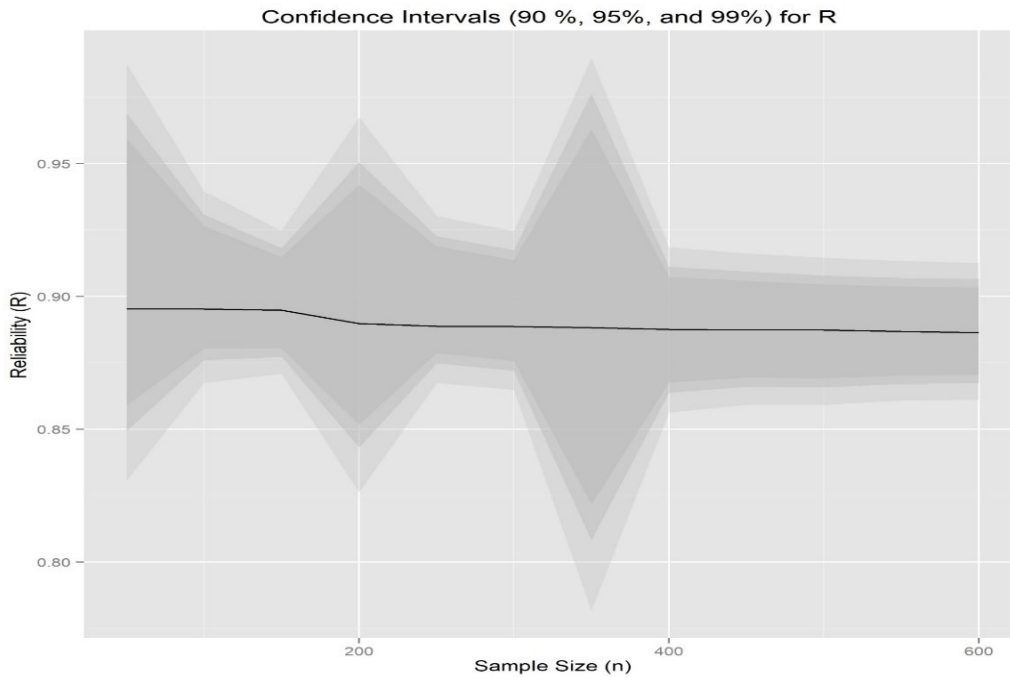


Figure 6: Bootstrap Confidence Intervals

6.6. Application on real data set

In this section, the above methods are applied to sample data set taken from Morrison [27]. This data set represent the level of three biochemical compounds found in the brain of twenty mice of the same strain in ten pairs. Both mice in each pair were in the same condition in terms of diet and care and one in each pair was randomly selected and received periodic administrations of the drug. The outcome of tests of the brains of the mice and consists of the amount of the compounds in micrograms per gram of the brain tissue. So, we want to determine the effect of the drug for changes in the level of three bio-chemical compounds found in the brain by estimating the probability.

Here, it is assumed that the $(x_1, x_2, x_3, y_1, y_2, y_3) \sim N_6(\mu, \Sigma)$. For the given data sets, proportion of variance of $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ for PC1: 0.9996 and 0.8662 respectively, then the MVUE of R is $\hat{R} = 0.6219$ and MLE of R is $R^* = 0.6325$. Also, we calculate that the above measure by principal component analysis as $\delta = 0.3386$, $U(n, \delta) = 3.266$, $M(n, \delta) = 60.846$, $R(n, \delta) = 5.367$, $\text{Var}(\hat{R}) = 0.0216$ and $\text{MSE}(R^*) = 0.0226$. The exact, approximate and bootstrap confidence intervals, using the sample data set reported in Table 6 and it is shows that that, the exact and approximate CIs are almost same band, but confidence band of bootstrap CIs is less than these.

Table 6: Confidence Intervals for the mice dataset

Confidence Intervals	90%			95%			99%		
	L	U	LB	L	U	LB	L	U	LB
Exact	0.4183	0.8067	0.4649	0.3788	0.8336	0.4183	0.3054	0.8790	0.3344
Approx.	0.4216	0.8092	0.4684	0.3819	0.8359	0.4216	0.3080	0.8807	0.3372
Bootstrap	0.3846	0.7381	0.4236	0.3507	0.7720	0.3846	0.2845	0.8382	0.3113

7. Sample Size Determination for Reliability

In a clinical study, the sample size calculation is to determine the number of subjects needed to have a desired power for detecting a clinically meaningful effect, i.e. the significant changes in clinical parameters. A study conducted with limited budget and/or some medical facilities, to choose the small number of subjects in respect of cost effectiveness and power. Suppose, we are interested in determining the minimum sample size before the study for effect of the drug of three bio-chemical compounds of mice [27].

In this above context, the hypotheses of interest are

$$H_0 : R = R_0 \text{ against } H_1 : R = R_1 (> R_0)$$

or,
$$H_0 : \delta = \Phi^{-1}(R_0) = \delta_0 \text{ against } H_1 : \delta = \Phi^{-1}(R_1) = \delta_1 (> \delta_0)$$

Then, the test statistic defined as $t = \sqrt{n}\hat{\delta}$ where, $t \sim t'_{(n-1), \sqrt{n}\delta}$, $\alpha =$ Type I error and $1-\beta =$ power of the test. Therefore, there exists an UMP invariant test [28], we reject H_0 when $t > c$, where 'c' is determined by

$$P_{H_0} (t > c) = \alpha$$

or,
$$P_{H_0} (t < c) = 1 - \alpha$$

or,
$$c = t'_{(1-\alpha), (n-1), \sqrt{n}\delta_0} \tag{9}$$

Then the power of test,

$$P_{H_1} (t > c) = 1 - \beta$$

or,
$$P_{H_1} \left(t'_{(n-1), \sqrt{n}\delta_1} < t'_{(1-\alpha), (n-1), \sqrt{n}\delta_0} \right) = \beta \tag{10}$$

We can get the sample size (n) by solving the equation (10) numerically for given value of α and β . The values of sample size n are reported in Table 7, in order to calculate sample size for two groups (i.e. treatment or control) are effect or not, we set the null hypothesis as $H_0 : R_0 = 0.5$ against the $R_0 > 0.5$.

Again, we consider the hypotheses,

$$H_0 : R = R_0 \text{ against } H_1 : R = R_1 (< R_0)$$

or,
$$H_0 : \delta = \delta_0 \text{ against } H_1 : \delta = \delta_1 (< \delta_0)$$

We get,

$$P_{H_0} (t < c) = \alpha$$

or,
$$c = t'_{\alpha, (n-1), \sqrt{n}\delta_0} \tag{11}$$

and,
$$P_{H_1} \left(t'_{(n-1), \sqrt{n}\delta_1} < t'_{\alpha, (n-1), \sqrt{n}\delta_0} \right) = 1 - \beta \tag{12}$$

In order to calculate the sample size, Reiser and Guttman used an approximation of a non-central t-distribution by a standard normal distribution [21], valid for large n as

$$n = \frac{z_{\beta}^2(1+\delta_c^2/2)}{(\delta_c - \delta_1)^2} = \frac{z_{(1-\alpha)}^2(1+\delta_c^2/2)}{(\delta_c - \delta_0)^2}, \text{ where } \delta_c = \frac{\delta_0 z_{\beta} + \delta_1 z_{\alpha}}{z_{\beta} + z_{\alpha}} \quad (13)$$

Example 1. Suppose the objective of the study is to compare a test drug with a control for changes in the level of three bio-chemical compounds found in the mice brain. Suppose the hypotheses of interest are

$$H_0 : R_0 = 0.5 \text{ against } H_1 : R_1 = 0.7 (> R_0)$$

or,

$$H_0 : \delta_0 = 0 \text{ against } H_1 : \delta_1 = 0.524 (> \delta_0)$$

There is no meaningful effect between test drug and control under H_0 and drug has an effect under H_1 . Then, by choosing $\alpha = 5\%$, and $\beta = 20\%$, we find $n \approx 24$ using (9) and (10). Thus, a total number of 24 subjects are required for achieving a 80% power for detection of a clinically meaningful effect at the 5% level of significance.

Example 2. Consider the example of Reiser and Guttman to determine sample size (n) [21] for the case

$$\text{Prodicer's risk} = \text{consumer's risk} = 0.05$$

To test,

$$H_0 : R_0 = 0.95 \text{ against } H_1 : R_1 = 0.90 (< R_0)$$

or,

$$H_0 : \delta_0 = 1.645 \text{ against } H_1 : \delta_1 = 1.282 (> \delta_0)$$

Here, Prodicer's risk and consumer's risk are equal, i.e. $z_{\alpha} = z_{\beta} = z_{0.05} = 1.645$, then we find $n \approx 170$ using (11) and (12). Similarly, we get the same result using (13).

Table 7: Sample Size Calculation table

$R_0:0.5$	Power = 1 - β								
	70%			80%			90%		
	Level of significant = α								
R_1	0.01	0.025	0.05	0.01	0.025	0.05	0.01	0.025	0.05
0.51	12934	9823	7489	15972	12491	9839	20715	16722	13628
0.55	517	393	299	638	499	393	827	667	544
0.6	129	98	75	159	124	98	206	166	135
0.65	57	44	33	70	55	43	90	73	59
0.7	32	24	19	39	31	24	50	40	33
0.75	21	16	12	25	19	15	31	25	20
0.8	14	11	8	17	13	10	21	17	14
0.85	10	8	6	12	9	7	15	12	10
0.9	8	6	5	9	7	5	11	9	7
0.95	6	4	3	7	5	4	8	6	5

7. Conclusions

Under the multivariate normal setup, MVUE of stress-strength model of reliability R is obtained, although the estimator based on MLE of $\mu_1, \mu_2, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}$. Simulation studies illustrate that, the Variance and MSE of two estimators reduces as the sample size increases and they almost achieved the true value of R. An application to the given real data set is described and shows that the same result as above. So, that the performance of MVUE of R is better than MLE in this case. In addition, the L_1 distance between distribution functions we see the improvement of such estimators. A difference in terms of MSE is much less as the values are given after multiplying by 100, though detailed calculations are required for other parametric values. Therefore, we may conclude that our

recommend estimator performs better.

The exact confidence intervals are preferable for marginally short band of confidence intervals than the approximate confidence intervals. The performance of bootstrap CIs is slightly worse than the exact and approximate CIs in terms of confidence band for small sample size. But the performance of bootstrap confidence intervals and other methods of CIs are almost same for large sample. Thus, the overall performance of the confidence interval is quite good for exact confidence intervals.

The sample size plays the important role using the stress strength reliability model in order to achieve minimum number of observation to evaluate the effectiveness of a new drug. The sample size should be massive enough to adequately answer the analysis question. The determination of the acceptable sample size involves applied statistical criteria additionally as in clinical studies. In order to calculate the sample size, it was necessary to choose the power, the significance level, to produce results that are clinically or experimentally meaningful. Under approximation, with δ_c we can calculate the sample size easily. But, it better to choose the exact method to get sample size.

References

- [1] Birnbaum, Z. W. (1956). On a use of the Mann-Whitney statistic. *Proc. of the third Berkeley Symposium on Mathematical Statistics and Probability*, 1:13-17.
- [2] Z. W. Birnbaum and R. C. McCarty. (1958). A distribution-free upper confidence bound for $P(Y < X)$ based on independent samples of X and Y . *Annals of Mathematical Statistics*, 29:558–562.
- [3] Nadarajah, S. (2002). Reliability for beta models. *Serdica Mathematical Journal*, 28: 267-282.
- [4] Nadarajah, S. (2003). Reliability for extreme value distributions. *Mathematical and Computer Modelling*, 37:915–922.
- [5] Nadarajah, S. (2003). Reliability for lifetime distributions. *Mathematical and Computer Modelling*, 37:683–688.
- [6] Nadarajah, S. (2004). Reliability for Laplace distributions. *Mathematical and Computer Modelling*, 2:169–183
- [7] Nadarajah, S. (2004). Reliability for Logistic distributions. *Engineering Simulation, to appear*, 26:81-98.
- [8] Weerahandi, S. and Johnson, R. A. (1992). Testing Reliability in a Stress-Strength Model When X and Y Are Normally Distributed. *Technometrics*, 34(1):83-91.
- [9] Simonoff, J. S., Hochberg, Y., and Reiser, B. (1986). Alternative estimation procedures for $\Pr(X < Y)$ in categorized data. *Biometrics*, 42:895–907.
- [10] Owen, D. B., Craswell, K.J. and Hanson, D.L. (1964). Nonparametric Upper Confidence Bounds for $\Pr\{Y < X\}$ and Confidence Limits for $\Pr\{Y < X\}$ When X and Y are Normal. *Journal of the American Statistical Association*, 59 (307):906-924.
- [11] Sen, P. K. (1967). A note on asymptotically distribution-free confidence bounds for $P(X < Y)$ based on two independent samples. *Sankhya*, 29:95-102.
- [12] Govindarazulu, Zakkula. (1967). Two-Sided Confidence Limits for $P(X < Y)$ Based on Normal Samples of X and Y . *Sankhyā*, 29(1/2):35-40.
- [13] Church, J.D., Harris, B. (1970). The estimation of reliability from stress strength relationship.
- [14] Downton, F. (1973). The estimation of $P(Y < X)$ in the normal case. *Technometrics*, 15(3): 551–558.
- [15] Woodward, W.A. and Kelley, G.D. (1977). Minimum Variance Unbiased Estimation of $P\{Y < X\}$ in the Normal Case. *Technometrics*, 19(1):95-98.
- [16] Mukherjee, S.P. and Sharan, L.K. (1985). Estimation of failure probability from a bivariate normal stress-strength distribution. *Microelectronics Reliability*, 25:699-702.
- [17] Hor, P. and Seal, B. (2015). Minimum variance unbiased estimation of stress–strength reliability under bivariate normal and its comparisons. *Communications in Statistics - Simulation and Computation*, 46(3):2447–2456.

- [18] Gupta, R. D. and Gupta, R. C. (1990). Estimation on $\Pr(a'x > b'y)$ in the Multivariate Normal Case. *Statistics: A Journal of Theoretical and Applied Statistics*, 21(1):91-97.
- [19] Enis, P. and Geisser, S. (1971). Estimation of the Probability that $Y < X$. *J. Amer. Statist. Assoc.*, 66: 162-168.
- [20] Reiser, B. and Farragi, D. (1994). Confidence bounds for $P(a'x > b'y)$. *Statistics*, 25:107–111.
- [21] Reiser, B. and Guttman, I. (1989). Sample size choice for reliability verification in strength-stress models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 17(3): 253-259.
- [22] Barton, D.E. (1961). Unbiased estimation of a set of probabilities. *Biometrika*, 48(1-2):227–229.
- [23] Mukhopadhyay, P. *Multivariate statistical analysis*, World Scientific, 2009.
- [24] Johnson, N.L. and Welch, B.L. (1939). Applications of the non-central t distribution. *Biometrika*, 31:362-389.
- [25] Efron, B. *The Jackknife, The Bootstrap, And Other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38 (Philadelphia, PA: SIAM), 1982.
- [26] Efron, B. (1987). Better bootstrap confidence intervals. *J. Am. Stat. Assoc.*, 82:171–185.
- [27] Morrison, D.F. *Multivariate statistical methods*, McGraw-Hill, New York, 2nd Edt., 1967.
- [28] Lehmann, E.L. and Romano, J.P. *Testing Statistical Hypotheses*, 3rd Edition. Springer, 2008.

GENERALIZATION OF LENGTH BIASED WEIGHTED GENERALIZED UNIFORM DISTRIBUTION AND ITS APPLICATIONS

Jismi Mathew

•

Department of Statistics, Vimala College (Autonomous), Thrissur, Kerala, India

jismij@gmail.com

Abstract

In this article, a generalization of length biased weighted generalized uniform distribution called Marshall Olkin length biased weighted generalized uniform distribution is introduced and studied. Some of the statistical properties of the new distribution such as hazard rate function, compounding, quantile function, moments, Renyi and Shannon entropies are discussed. The maximum likelihood estimation of the model parameters is done and a simulation study is conducted for confirming the validity of the estimates and also introduced a minification process with respect to the model and explored its sample path behaviour for different combinations of parameters. Further, the stress strength analysis is carried out and the estimate of the reliability is obtained based on a simulation study.

Keywords: Entropy, Length biased weighted generalized uniform distribution, Maximum likelihood method, Order statistics, Quantiles, Stress strength analysis.

1. Introduction

The theory of weighted distributions provides a collective access for the problems of model specification and data interpretation. Weighted distributions take into account the method of ascertainment, by adjusting the probabilities of the actual occurrence of events to arrive at a specification of the probabilities of those events as observed and recorded [14].

The uniform distribution is considered as the simplest probability model and is connected to all the distributions. Many characterizations and modifications of the generalized uniform distribution have been introduced and explored by various researchers [see, 19, 10, 8, 18]. Rather and Subramanian [16] introduced and studied the properties of length biased weighted generalized uniform distribution.

The probability density function and cumulative distribution function of length biased weighted generalized uniform distribution (LBWGU) are respectively, given by

$$g_{LBWGU}(x; \theta, \gamma) = \frac{(\theta+2)x^{\theta+1}}{\gamma^{\theta+2}}; \quad 0 < x < \gamma, \theta > -1 \quad (1)$$

$$G_{LBWGU}(x; \theta, \gamma) = \left(\frac{x}{\gamma}\right)^{\theta+2}; \quad 0 < x < \gamma, \theta > -1 \quad (2)$$

Marshall Olkin [11] introduced a new family of distributions by inserting a new shape parameter to the existing family of distributions. Let $G(x)$ be the cumulative distribution function (cdf) of a random variable X , then the cdf of the Marshall and Olkin family of distributions is

$$F(x) = \frac{G(x)}{1-(1-\beta)(1-G(x))} \quad (3)$$

The corresponding pdf of (3) is given by

$$f(x) = \frac{\beta g(x)}{[1-(1-\beta)(1-G(x))]^2} \tag{4}$$

where $\beta > 0$ is a shape parameter. Clearly, for $\beta = 1$, we obtain the baseline distribution, i.e., $F(x) = G(x)$.

Many authors have introduced various univariate distributions belonging to the Marshall-Olkin family of distributions such as Marshall-Olkin Weibull [5], Marshall-Olkin semi Burr and Marshall-Olkin Burr [7], Marshall-Olkin Frechet distribution [9], Marshall-Olkin generalized exponential distribution [17] and Marshall-Olkin extended generalized Lindley distribution [2]. Recently, introduced Marshall-Olkin form of additive Weibull distribution [1], reliability test plan for the Marshall-Olkin length biased Lomax distribution [12] and Marshall-Olkin length biased Maxwell distribution and its applications [13].

The rest of this paper is planned as follows. In section 2, the Marshall-Olkin length biased weighted generalized uniform (MOLBWGU) distribution is given, with plots of the pdf and cdf. The statistical properties of the new distribution are studied in section 3, including hazard rate function, moments, quantile function, compounding properties, order statistics and Renyi and Shannon entropies. Estimation of the model parameters are discussed in section 4. In section 5, the application of the distribution in time series analysis is discussed. In section 6, the stress strength analysis is carried out using a simulation study. Concluding remarks are presented in section 7.

2. Marshall-Olkin Length Biased Weighted Generalized Uniform Distribution

Let X follows length biased weighted generalized uniform distribution. A new distribution can be defined by inserting (2) in (3). The cdf obtained is

$$F_{MOLBWGU}(x; \theta, \gamma, \beta) = \frac{\left(\frac{x}{\gamma}\right)^{\theta+2}}{1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)}, \quad 0 < x < \gamma. \tag{5}$$

Based on (5), the survival function of the MOLBWGU distribution can be expressed as

$$S_{MOLBWGU}(x; \theta, \gamma, \beta) = \frac{\beta\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)}{1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)}, \quad 0 < x < \gamma. \tag{6}$$

where $\theta > -1$ and $\beta > 0$.

By putting (1) and (2) in (4), we obtain the pdf of the MOLBWGU distribution as

$$f_{MOLBWGU}(x; \theta, \gamma, \beta) = \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)\right)^2}, \quad 0 < x < \gamma. \tag{7}$$

We refer to this new distribution as the generalization of length biased weighted generalized uniform distribution with parameters θ, γ and β .

The shape of the pdf $f(x; \theta, \gamma, \beta)$ depends on parameter β . If $\beta \in (0,1)$ then the pdf is a bell shaped function on $(0, \gamma)$ with $f(0; \theta, \gamma, \beta) = 0$ and $f(\gamma, \theta, \gamma, \beta) = \frac{(\theta+2)\beta}{\gamma}$. In the case of $\beta > 1$ then the pdf is an increasing function on $(0, \gamma)$ with $f(0, \theta, \gamma, \beta) = 0$ and $f(\gamma, \theta, \gamma, \beta) = \frac{(\theta+2)\beta}{\gamma}$.

Remark 1. If $\beta = 1$, we obtain length biased weighted generalized uniform distribution introduced by Rather and Subramanian [16].

Remark 2. When $\theta = -1$ and $\beta = 1$, MOLBWGU distribution reduces to uniform distribution over $(0, \gamma)$.

Remark 3. When $\gamma = 1$ and $\beta = 1$, MOLBWGU distribution reduces to standard power function distribution.

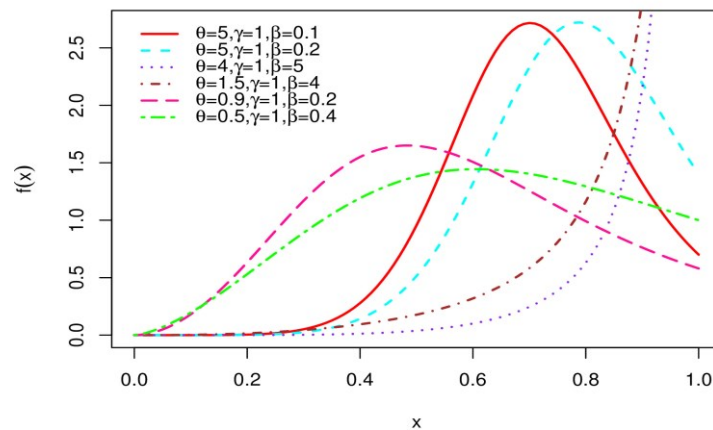


Figure 1 : Curves of the pdf of the MOLBWGU distribution for different values of the parameters.

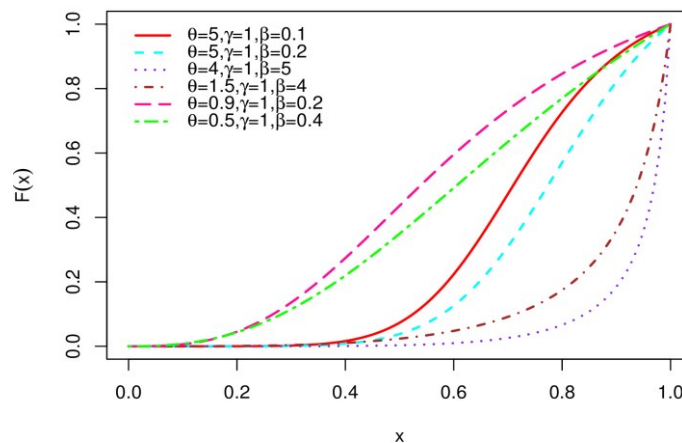


Figure 2 : Curves of the cdf of the MOLBWGU distribution for different values of the parameters.

3. Statistical Properties

This section is devoted to some statistical properties of the MOLBWGU distribution.

3.1. Hazard Rate Function

The hrf is given by,
$$h_{MOLBWGU}(x; \theta, \gamma, \beta) = \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)\right)^2 \left(1-\frac{\left(\frac{x}{\gamma}\right)^{\theta+2}}{1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)}\right)}$$

where, $0 < x < \gamma$.

For $\beta \in (0,1)$ and $\beta > 1$, the hrf is evidently increasing failure rate.

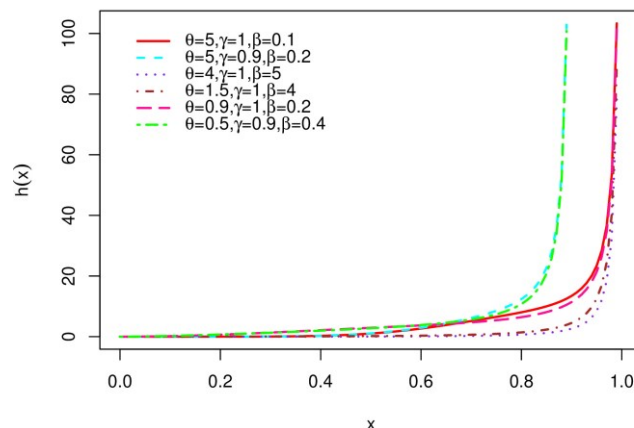


Figure 3: Curves of the hazard rate function of the MOLBWGU distribution for different values of the parameters.

The reverse hazard function of $MOLBWGU(\theta, \gamma, \beta)$ is given by,

$$r_{MOLBWGU}(x; \theta, \gamma, \beta) = \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)\right)^2 \left(\frac{\left(\frac{x}{\gamma}\right)^{\theta+2}}{1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)}\right)}$$

The reverse hazard rate function decreases with $r(0, \theta, \gamma, \beta) = 0$ and $r(\gamma, \theta, \gamma, \beta) = \frac{(\theta+2)\beta}{\gamma}$.

3.2. Compounding

The property that Marshall-Olkin family of distributions can be expressed as a compound distribution with exponential distribution as mixing density is useful in obtaining new parameter families of distribution in terms of existing ones, expressed Marshall-Olkin extended forms of Weibull, Lomax, linear exponential and exponential power family of distributions as a compound distribution [see 5, 4, 9].

Theorem 1. Let X be a continuous random variable with conditional survival function on $\Delta = \delta$ expressed as $\bar{F}(x|\delta) = \left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right) e^{-(1-\beta)\delta\left(\frac{x}{\gamma}\right)^{\theta+2}}$, $0 < x < \gamma$,

and let Δ follows a distribution function with probability density function

$$m(\delta) = \beta e^{-\beta\delta}, \delta > 0.$$

Then the random variable X has the $MOLBWGU(\theta, \gamma, \beta)$ distribution.

Proof: The unconditional survival function of the random variable X is given by,

$$\begin{aligned} \bar{F}(x) &= \int_{-\infty}^{\infty} \bar{F}(x|\delta)m(\delta)d\delta \\ &= \beta\left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right) \int_0^{\infty} e^{-[\beta+(1-\beta)\left(\frac{x}{\gamma}\right)^{\theta+2}]\delta} d\delta \\ &= \frac{\beta\left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)}{1-(1-\beta)\left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)}. \end{aligned}$$

which is the survival function of the $MOLBWGU(\theta, \gamma, \beta)$ distribution.

Theorem 2. Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. random variables with common survival function $G'(x)$. Let T be a geometric random variable independently distributed of $\{X_i, i \geq 1\}$ such that $P(T = n) = \beta(1 - \beta)^{n-1}, n = 1, 2, \dots, 0 < \beta < 1$. Let $Y_T = \min_{1 \leq i \leq T} X_i$. Then $\{Y_T\}$ is distributed as $MOLBWGU(\theta, \gamma, \beta)$ if and only if $\{X_i\}$ follows $LBWGU(\theta, \gamma)$.

Proof: The survival function of the random variable Y_T is

$$H'(x) = P(Y_T > x)$$

$$\begin{aligned}
 &= \sum_{n=1}^{\infty} P(Y_n > x)P(T = n) \\
 &= \sum_{n=1}^{\infty} [G'(x)]^n \beta(1 - \beta)^{n-1} \\
 &= \frac{\beta G'(x)}{1 - (1 - \beta)G'(x)} \\
 &= \frac{\beta(1 - (\frac{x}{\gamma})^{\theta+2})}{1 - (1 - \beta)(1 - (\frac{x}{\gamma})^{\theta+2})}.
 \end{aligned}$$

which is survival function of MOLBWGU(θ, γ, β) distribution.

Theorem 3. Let $\{X_i, i \geq 1\}$ be a sequence of i.i.d. random variables with common survival function $G'(x)$. Let T be a geometric random variable independently distributed of $\{X_i, i \geq 1\}$ such that $P(T = n) = \beta(1 - \beta)^{n-1}, n = 1, 2, \dots, 0 < \beta < 1$. Let $Z_T = \max_{1 \leq i \leq T} X_i$. Then $\{Z_T\}$ is distributed as MOLBWGU($\theta, \gamma, \frac{1}{\beta}$) if and only if $\{X_i\}$ follows LBWGU(θ, γ) distribution.

Proof: The distribution function of Z_T is

$$\begin{aligned}
 K(x) &= P(Z_T \leq x) \\
 &= \sum_{n=1}^{\infty} P(Z_n \leq x)P(T = n) \\
 &= \sum_{n=1}^{\infty} [G(x)]^n \beta(1 - \beta)^{n-1} \\
 &= \frac{\beta G(x)}{1 - (1 - \beta)G(x)} \\
 &= \frac{\beta(\frac{x}{\gamma})^{\theta+2}}{1 - (1 - \beta)(\frac{x}{\gamma})^{\theta+2}}.
 \end{aligned}$$

From this it follows that the survival function of the random variable Z_T is

$$K'(x) = \frac{\frac{1}{\beta}(1 - (\frac{x}{\gamma})^{\theta+2})}{1 - (1 - \frac{1}{\beta})(1 - (\frac{x}{\gamma})^{\theta+2})},$$

which implies that Z_T has MOLBWGU($\theta, \gamma, \frac{1}{\beta}$) distribution.

3.3. Order Statistics

Let (x_1, x_2, \dots, x_n) be a random sample of size n from MOLBWGU(θ, γ, β) distribution and let $x_{1:n}, x_{2:n}, \dots, x_{n:n}$ be the corresponding order statistics. Then the pdf of the j^{th} order statistic for the MOLBWGU distribution is given by

$$f_{j:n}(x) = \frac{n!}{(j-1)!(n-j)!} \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)\right)^2} \times \left[\frac{(\frac{x}{\gamma})^{\theta+2}}{1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)} \right]^{j-1} \left[\frac{\beta(1 - (\frac{x}{\gamma})^{\theta+2})}{1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)} \right]^{n-j}.$$

The MOLBWGU distribution has the following pdf for $x_{1:n}$

$$f_{1:n}(x) = \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)\right)^2} \left[\frac{\beta(1 - (\frac{x}{\gamma})^{\theta+2})}{1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)} \right]^{n-1}.$$

and the pdf for $x_{n:n}$ is given by

$$f_{n:n}(x) = n \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)\right)^2} \left[\frac{(\frac{x}{\gamma})^{\theta+2}}{1 - (1 - \beta)\left(1 - (\frac{x}{\gamma})^{\theta+2}\right)} \right]^{n-1}, \quad 0 < x < \gamma.$$

3.4. Quantile Function

The q^{th} quantile of MOLBWGU(θ, γ, β) distribution is given by

$$x_q = F^{-1}(q) = \gamma \left(\frac{q\beta}{1 - q(1 - \beta)} \right)^{\frac{1}{\theta+2}}, \quad 0 \leq q \leq 1.$$

where $F^{-1}(\cdot)$ is the inverse distribution function.

In particular the median of $MOLBWGU(\theta, \gamma, \beta)$ distribution is given by,

$$\text{median}(X) = \gamma \left(\frac{\beta}{1+\beta} \right)^{\frac{1}{\theta+2}}.$$

3.5. Moments

If X has the $MOLBWGU(\theta, \gamma, \beta)$ distribution, then the s^{th} order moment is obtained as

$$\begin{aligned} \mathbb{E}(X^s) &= \int_0^\gamma x^s \frac{(\theta+2)\beta\gamma^{-\theta-2}x^{\theta+1}}{\left(1-(1-\beta)\left(1-\left(\frac{x}{\gamma}\right)^{\theta+2}\right)\right)^2} dx \\ &= \frac{\beta(\theta+2)}{\gamma^{3(\theta+2)}} \int_0^\gamma \frac{x^{s+\theta+1}}{(\beta\gamma^{\theta+2}+(1-\beta)x^{\theta+2})^2} dx. \end{aligned}$$

where \mathbb{E} denotes the expectation.

Let $x^{\theta+2} = u$, above equation reduces to

$$\mathbb{E}(X^s) = \frac{\beta}{\gamma^{3(\theta+2)}} \int_0^{\gamma^{\theta+2}} \frac{u^{\frac{s}{\theta+2}}}{(\beta\gamma^{\theta+2}+(1-\beta)u)^2} du.$$

From Prudnikov [15],

$$\int_a^b \frac{(x-a)^{\alpha-1}}{(cx+d)^{\alpha+n+1}} dx = \frac{(b-a)^\alpha}{(ac+d)(bc+d)^\alpha} \sum_{k=0}^n nC_k \frac{B(\alpha+k, n-k+1)}{(bc+d)^k(ac+d)^{n-k}}.$$

where a, b, c and d are real numbers with $(ac+d)(bc+d) > 0$; Real part of $\alpha > 0$ and $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

Hence if $s/\theta + 2$ is a positive integer, we have,

$$\begin{aligned} \int_0^{\gamma^{\theta+2}} \frac{u^{\frac{s}{\theta+2}}}{(\beta\gamma^{\theta+2}+(1-\beta)u)^2} du &= \frac{1}{\gamma^{\theta+2+s}\beta^{1+s/(\theta+2)}} \sum_{k=0}^{s/\theta+2} s/\theta + 2 C_k \beta^k \\ &B\left(1+k+\frac{s}{\theta+2}, 1-k+\frac{s}{\theta+2}\right). \end{aligned}$$

Therefore

$$\mathbb{E}(X^s) = \frac{1}{\gamma^{4(\theta+2)+s\beta^{s/(\theta+2)}}} \sum_{k=0}^{s/\theta+2} s/\theta + 2 C_k \beta^k B\left(1+k+\frac{s}{\theta+2}, 1-k+\frac{s}{\theta+2}\right).$$

In particular,

$$\mathbb{E}(X) = \frac{1}{\gamma^{4\theta+9}\beta^{1/(\theta+2)}} \sum_{k=0}^{1/\theta+2} 1/\theta + 2 C_k \beta^k B\left(1+k+\frac{1}{\theta+2}, 1-k+\frac{1}{\theta+2}\right).$$

$$\mathbb{E}(X^2) = \frac{1}{\gamma^{4\theta+10}\beta^{2/(\theta+2)}} \sum_{k=0}^{2/\theta+2} 2/\theta + 2 C_k \beta^k B\left(1+k+\frac{2}{\theta+2}, 1-k+\frac{2}{\theta+2}\right).$$

3.6. Renyi and Shannon Entropies

The Renyi entropy is defined as

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \int_0^\infty f^\gamma(x) dx, \gamma > 0, \gamma \neq 1.$$

Then, $\int_0^\infty f^\gamma(x) dx = \int_0^\gamma \frac{\beta^\gamma(\theta+2)^\gamma}{\gamma^{3(\theta+2)\gamma}} \frac{x^{\gamma(\theta+1)}}{(\beta\gamma^{\theta+2}+(1-\beta)x^{\theta+2})^{2\gamma}} dx.$

Let $u = x^{\theta+2}$. Therefore

$$\int_0^\gamma \frac{x^{\gamma(\theta+1)}}{(\beta\gamma^{\theta+2}+(1-\beta)x^{\theta+2})^{2\gamma}} dx = \frac{1}{\theta+2} \int_0^{\gamma^{\theta+2}} \frac{u^{\frac{1}{\theta+2}(\gamma-1)(\theta+1)}}{(\beta\gamma^{\theta+2}+(1-\beta)u)^{2\gamma}} du.$$

Using the equation from Prudnikov [15] and if $\frac{(\gamma-1)(\theta+3)}{(\theta+2)}$ is a positive integer, the above integral becomes

$$\frac{1}{(\theta+2)\gamma^{(\theta+2)+(\gamma-1)(\theta+3)}\beta^{1+\frac{(\gamma-1)(\theta+3)}{(\theta+2)}}} \sum_{k=0}^{\frac{(\gamma-1)(\theta+3)}{(\theta+2)}} \frac{(\gamma-1)(\theta+3)}{(\theta+2)} C_k \beta^k B\left(1+k+\frac{(\gamma-1)(\theta+3)}{(\theta+2)}, 1-k+\frac{(\gamma-1)(\theta+3)}{(\theta+2)}\right).$$

Therefore the Renyi entropy is

$$I_R(\gamma) = \frac{1}{1-\gamma} \log \left[\frac{\beta^{\frac{(1-\gamma)}{(\theta+2)(\theta+2)(\gamma-1)}}}{\gamma^{(4\theta+9)\gamma-1}} \sum_{k=0}^{\frac{(\gamma-1)(\theta+3)}{(\theta+2)}} \frac{(\gamma-1)(\theta+3)}{(\theta+2)} C_k \beta^k \right. \\ \left. B\left(1+k+\frac{(\gamma-1)(\theta+3)}{(\theta+2)}, 1-k+\frac{(\gamma-1)(\theta+3)}{(\theta+2)}\right) \right].$$

Thus, the Shannon entropy is

$$E[-\log f(x)] = -\log[\beta(\theta+2)\gamma^{-3(\theta+2)}] - (\theta+1)E[\log(X)] + 2E[\log(\beta\gamma^{\theta+2} + (1-\beta)x^{\theta+2})].$$

4. Maximum Likelihood Estimation

The MLE method is used for the parameter estimation of MOLBWGU distribution. Let (x_1, x_2, \dots, x_n) be a random sample of size n from the MOLBWGU distribution. The likelihood function for the MOLBWGU distribution is,

$$L = \prod_{i=1}^n \frac{(\theta+2)\beta\gamma^{-\theta-2}x_i^{\theta+1}}{\left(1-(1-\beta)\left(1-\left(\frac{x_i}{\gamma}\right)^{\theta+2}\right)\right)^2}.$$

from which the log-likelihood function is obtained as

$$\log L = -2 \sum_{i=1}^n \log\left(1 - (1-\beta)\left(1 - \gamma^{-\theta-2}x_i^{\theta+2}\right)\right) + (\theta+1) \sum_{i=1}^n \log(x_i) \\ + n(-(\theta+2)\log(\gamma) + \log(\theta+2) + \log(\beta)).$$

The partial derivatives of this log-likelihood function is given by

$$\frac{\partial \log L}{\partial \theta} = -2 \sum_{i=1}^n \frac{(1-\beta)(\gamma^{-\theta-2}\log(\gamma)x_i^{\theta+2} - \gamma^{-\theta-2}x_i^{\theta+2}\log(x_i))}{1-(1-\beta)(1-\gamma^{-\theta-2}x_i^{\theta+2})} \\ + \sum_{i=1}^n \log(x_i) + n\left(\frac{1}{\theta+2} - \log(\gamma)\right). \\ \frac{\partial \log L}{\partial \gamma} = -2 \sum_{i=1}^n \frac{(-\theta-2)(1-\beta)\gamma^{-\theta-3}x_i^{\theta+2}}{1-(1-\beta)(1-\gamma^{-\theta-2}x_i^{\theta+2})} - \frac{(\theta+2)n}{\gamma}. \\ \frac{\partial \log L}{\partial \beta} = \frac{n}{\beta} - 2 \sum_{i=1}^n \frac{1-\gamma^{-\theta-2}x_i^{\theta+2}}{1-(1-\beta)(1-\gamma^{-\theta-2}x_i^{\theta+2})}.$$

The maximum likelihood estimator $(\hat{\theta}, \hat{\gamma}, \hat{\beta})$ of the parameters (θ, γ, β) can be obtained by solving the equations $\frac{\partial \log L}{\partial \theta} = 0$, $\frac{\partial \log L}{\partial \gamma} = 0$ and $\frac{\partial \log L}{\partial \beta} = 0$.

4.1. Simulation Study

In this section, some simulation results are provided to study the behaviour of the MLEs in terms of the sample size. For this purpose, a Monte Carlo simulation study is conducted for $MOLBWGU(\theta, \gamma, \beta)$ distribution. The results are obtained from 1000 Monte Carlo replications and the simulations are carried out using the statistical software R. In each replication, a random sample of size 25, 50, 100, 150, 200 is generated for different combinations of θ, γ and β . The initial values of parameters are $\theta = 1.2, \gamma = 0.3, \beta = 1.5$; $\theta = 2, \gamma = 1.5, \beta = 2.5$; $\theta = 0.5, \gamma = 0.5, \beta = 1.5$ and $\theta = 2, \gamma = 1, \beta = 2$. Then computed mean of the MLEs of the parameters, biases and mean square errors (MSEs) of the parameter estimates. Tables 1, 2, 3 and 4 gives the values of the estimates, biases and MSEs of the corresponding parameters. From the tables, it can be seen that, as sample size increases the bias and MSE of the estimates decreases.

Table 1: Estimates, Biases and MSEs for $\theta = 1.2$, $\gamma = 0.3$ and $\beta = 1.5$

Sample Size(n)	Parameters	Estimates	Biases	MSEs
25	θ	1.2155	0.0155	0.0431
	γ	0.3579	0.0579	0.1046
	β	1.5034	0.0034	0.0009
50	θ	1.2143	0.0143	0.0412
	γ	0.3565	0.0565	0.1051
	β	1.5022	0.0022	0.0003
100	θ	1.2135	0.0135	0.0159
	γ	0.3541	0.0541	0.1015
	β	1.5015	0.0015	0.0003
150	θ	1.2111	0.0111	0.0039
	γ	0.3509	0.0509	0.1012
	β	1.5009	0.0009	0.0002
200	θ	1.2097	0.0097	0.0027
	γ	0.3478	0.0478	0.1004
	β	1.5001	0.0001	0.0002

Table 2: Estimates, Biases and MSEs for $\theta = 2$, $\gamma = 1.5$ and $\beta = 2.5$

Sample Size(n)	Parameters	Estimates	Biases	MSEs
25	θ	2.6271	0.6271	0.4428
	γ	1.5096	0.0096	0.2635
	β	2.5242	0.0242	0.9715
50	θ	2.4927	0.4927	0.3527
	γ	1.5082	0.0082	0.0792
	β	2.5211	0.0211	0.6226
100	θ	2.3813	0.3813	0.3795
	γ	1.5065	0.0065	0.0489
	β	2.5175	0.0175	0.6535
150	θ	2.3145	0.3145	0.1529
	γ	1.5068	0.0068	0.0035
	β	2.5148	0.0148	0.3614
200	θ	2.2501	0.2501	0.1358
	γ	1.5045	0.0045	0.0012
	β	2.5122	0.0122	0.1428

Table 3: Estimates, Biases and MSEs for $\theta = 0.5$, $\gamma = 0.5$ and $\beta = 1.5$

Sample Size(n)	Parameters	Estimates	Biases	MSEs
25	θ	0.5312	0.0312	0.0803
	γ	0.5505	0.0505	1.5112
	β	1.5091	0.0091	0.1583
50	θ	0.5304	0.0304	0.0713
	γ	0.5447	0.0447	0.8218
	β	1.5082	0.0082	0.0685
100	θ	0.5275	0.0275	0.0752
	γ	0.5418	0.0418	0.7943
	β	1.5079	0.0079	0.0082
150	θ	0.5289	0.0289	0.0239
	γ	0.5345	0.0345	0.5728
	β	1.5074	0.0074	0.0047
200	θ	0.5217	0.0217	0.0249
	γ	0.5293	0.0293	0.5355
	β	1.5066	0.0066	0.0032

Table 4: Estimates, Biases and MSEs for $\theta = 2$, $\gamma = 1$ and $\beta = 2$

Sample Size(n)	Parameters	Estimates	Biases	MSEs
25	θ	2.1827	0.1827	0.3912
	γ	1.1578	0.1578	0.2014
	β	2.1666	0.1666	0.0009
50	θ	2.0915	0.0915	0.3373
	γ	1.1579	0.1579	0.2008
	β	2.1643	0.1643	0.0004
100	θ	2.0912	0.0912	0.1838
	γ	1.1458	0.1458	0.2007
	β	2.1712	0.1712	0.0003
150	θ	2.0774	0.0774	0.1425
	γ	1.1435	0.1435	0.2001
	β	2.1626	0.1626	0.0003
200	θ	2.0751	0.0751	0.0564
	γ	1.1315	0.1315	0.0197
	β	2.0125	0.0125	0.0002

5. Application in Autoregressive Time Series Modeling

In this section, some applications of MOLBWGU distribution in autoregressive time series modelling are provided. Now, we construct a first order autoregressive minification process with structure as follows,

$$X_n = \begin{cases} \varepsilon_n, & w.p. \beta \\ \min(x_{n-1}, \varepsilon_n), & w.p. 1 - \beta, \quad 0 \leq \beta \leq 1, n \geq 1, \end{cases} \tag{8}$$

where $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables following LBWGU(θ, γ) distribution independent of $\{x_{n-1}, x_{n-2}, \dots\}$. Then the process is stationary and is marginally distributed with MOLBWGU(θ, γ, β) distribution. This leads to the following theorem.

Theorem 4. In an AR(1) process with structure (8), $\{X_n, n \geq 0\}$ defines a stationary AR(1) minification process with MOLBWGU(θ, γ, β) marginal distribution iff $\{\varepsilon_n\}$ is a sequence of independently and identically distributed random variable with LBWGU(θ, γ) distributon.

Proof. Consider (8) in terms of survival function

$$\bar{F}_{X_n}(x) = \beta \bar{F}_{\varepsilon_n}(x) + (1 - \beta) \bar{F}_{X_{n-1}}(x) \bar{F}_{\varepsilon_n}(x).$$

Under stationary equilibrium it reduces to

$$\bar{F}_X(x) = \frac{\beta \bar{F}_{\varepsilon_n}(x)}{1 - (1 - \beta) \bar{F}_{\varepsilon_n}(x)}. \tag{9}$$

and hence

$$\bar{F}_{\varepsilon_n}(x) = \frac{\beta \bar{F}_X(x)}{1 - (1 - \beta) \bar{F}_X(x)}. \tag{10}$$

If ε_n follows LBWGU(θ, γ) from(9), we get

$$\bar{F}_X(x) = \frac{\beta \left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)}{1 - (1 - \beta) \left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)}$$

which is the survival function of MOLBWGU(θ, γ, β) .

Conversely, if we take

$$\bar{F}_{X_n}(x) = \frac{\beta \left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)}{1 - (1 - \beta) \left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)}$$

from (9) it can show that $\bar{F}_{\varepsilon_n}(x)$ is distributed LBWGU(θ, γ) with survival function $\left(1 - \left(\frac{x}{\gamma}\right)^{\theta+2}\right)$.

5.1. Sample Path

To study the behavior of the process we simulate the sample path for various values of β , the properties of sample path shows that the MOLBWGU AR(1) minification process can be used for modelling a rich variety of real data. Sample path of MOLBWGU AR(1) process for $\gamma = 0.5, \theta = 0.9$ and $\beta = 0.4, 0.5, 0.6$ and 0.8 is given in Figure 4.

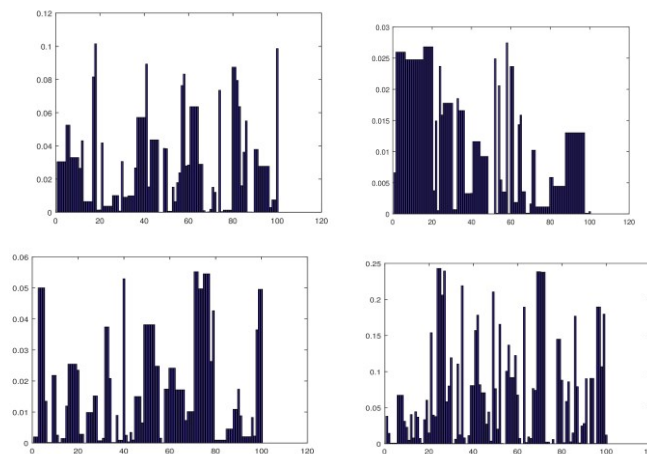


Figure 4: Sample path of the MOLBWGU AR(1) process for $\gamma = 0.5, \theta = 0.9$ and $\beta = 0.4, 0.5, 0.6$ and 0.8 .

6. Stress Strength Analysis

The stress strength reliability analysis can be regarded as an assessment of reliability of a system in terms of random variables X and Y , where X represents strength and Y represents the stress. If the stress exceeds strength the system would fail and the system will function if strength exceeds stress. The stress strength reliability can be defined as $R = P(X > Y)$. Gupta [6] obtained various results on the MO family in the context of reliability modelling and survival analysis. Then,

$$R = P(X > Y) = \int_{-\infty}^{+\infty} P(X > Y | Y = y) g_Y(y) dy$$

$$= \frac{\beta_1}{(\beta_2 - 1)^2} \left(-\ln \frac{\beta_1}{\beta_2} + \frac{\beta_1}{\beta_2} - 1 \right).$$

Let (x_1, x_2, \dots, x_m) and (y_1, y_2, \dots, y_n) be two independent random samples of sizes m and n from MOLBWGU distribution with tilt parameters β_1 and β_2 respectively, and common unknown parameters γ and θ .

The log likelihood function is given by

$$\log L = -2 \sum_{i=1}^m \log \left(1 - (1 - \beta_1)(1 - \gamma^{-\theta-2} x_i^{\theta+2}) \right) + (\theta + 1) \sum_{i=1}^m \log(x_i) + m(-(\theta + 2)\log(\gamma) + \log(\theta + 2) + \log(\beta_1))$$

$$- 2 \sum_{i=1}^n \log \left(1 - (1 - \beta_2)(1 - \gamma^{-\theta-2} y_i^{\theta+2}) \right) + (\theta + 1) \sum_{i=1}^n \log(y_i)$$

$$+ n(-(\theta + 2)\log(\gamma) + \log(\theta + 2) + \log(\beta_2))$$

Then MLE of β_1 and β_2 are the solutions of the nonlinear equations $\frac{\partial \log L}{\partial \beta_1} = 0$ and $\frac{\partial \log L}{\partial \beta_2} = 0$. The elements of Information matrix are given by,

$$I_{11} = -E \left(\frac{\partial^2 L}{\partial \beta_1^2} \right) = \frac{m}{3\beta_1^2}$$

$$I_{22} = -E \left(\frac{\partial^2 L}{\partial \beta_2^2} \right) = \frac{n}{3\beta_2^2}$$

$$I_{12} = I_{21} = -E \left(\frac{\partial^2 L}{\partial \beta_1 \partial \beta_2} \right) = 0.$$

By the property of MLE for $m \rightarrow \infty$ and $n \rightarrow \infty$,

$$(\sqrt{m}(\hat{\beta}_1 - \beta_1), \sqrt{n}(\hat{\beta}_2 - \beta_2)) \rightarrow N_2(0, \text{diag}\{a_{11}^{-1}, a_{22}^{-1}\}).$$

where $a_{11} = \lim_{m,n \rightarrow \infty} \frac{I_{11}}{m} = \frac{1}{3\beta_1^2}$ and $a_{22} = \lim_{m,n \rightarrow \infty} \frac{I_{22}}{n} = \frac{1}{3\beta_2^2}$

Now from [6] the 95% confidence interval for R is given by $\hat{R} \pm 1.96 \hat{\beta}_1 b_1(\hat{\beta}_1, \hat{\beta}_2) \sqrt{\frac{3}{m} + \frac{3}{n}}$,

where $b_1(\beta_1, \lambda_2) = \frac{\partial R}{\partial \beta_1} = \frac{\beta_2}{(\beta_1 - \beta_2)^3} \left[-2(\beta_1 - \beta_2) + (\beta_1 + \beta_2) \ln \frac{\beta_1}{\beta_2} \right]$ and $b_2(\lambda_1, \lambda_2) = \frac{\partial R}{\partial \beta_2} = \frac{\beta_1}{(\beta_1 - \beta_2)^3} \left[2(\beta_1 - \beta_2) - (\beta_1 + \beta_2) \ln \frac{\beta_1}{\beta_2} \right] = -\frac{\beta_1}{\beta_2} b_1(\beta_1, \beta_2)$.

6.1. Simulation Study

For the simulation study, generate $N=10,000$ sets of X -samples and Y -samples from the MOLBWGU distribution with parameters $(\beta_1, \gamma, \theta)$ and $(\beta_2, \gamma, \theta)$ respectively. The combinations of samples of sizes $m = 20, 25, 30$ and $n = 20, 25, 30$ are studied. The validity of the estimate of R is considered by using the following measures, namely average bias of the estimate (\bar{b}), average mean square error of the estimate (AMSE), average confidence interval of the estimate and coverage probability defined by,

1. Average bias (\bar{b}) of the estimates of R :

$$\frac{1}{N} \sum_{i=1}^N (\hat{R}_i - R).$$

2. Average mean square error of the estimates of R :

$$\frac{1}{N} \sum_{i=1}^N (\hat{R}_i - R)^2.$$

3. Average length of the asymptotic 95% confidence interval of R :

$$\frac{1}{N} \sum_{i=1}^N 2(1.96)\hat{\beta}_{1i} b_{1i}(\hat{\beta}_{1i}, \hat{\beta}_{2i}) \sqrt{\frac{3}{m} + \frac{3}{n}}$$

4. The coverage probability of the confidence intervals given by the proportion of such interval that include the parameter R .

The numerical values obtained for the measures are presented in Table 5 and 6. The average bias decreases as the sample size increases. The coverage probability is close to 0.95 as the sample size increases. This simulation results shows that the average bias, average MSE, average confidence interval and coverage probability do not show much variability for various parameter combinations.

Table 5: Average bias and average MSE of the simulated estimates of R for $\gamma = 1$ and $\theta = 0.9$

(β_1, β_2)								
(m, n)	Average Bias (\bar{b})				Average Mean Square Error (AMSE)			
	(0.2,0.4)	(0.4, 0.2)	(0.5, 0.1)	(0.4, 1.2)	(0.2, 0.4)	(0.4, 0.2)	(0.5, 0.1)	(0.4, 1.2)
(20, 20)	-0.0035	-0.0825	-0.0589	0.0286	0.0042	0.0115	0.0040	0.0052
(20, 25)	-0.0066	-0.0832	-0.0605	0.0289	0.0041	0.0117	0.0041	0.0052
(20, 30)	-0.0092	-0.0839	-0.0614	0.0279	0.0039	0.0117	0.0042	0.0050
(25, 20)	-0.0024	-0.0857	-0.0588	0.0233	0.0041	0.0115	0.0040	0.0041
(25, 25)	-0.0052	-0.0877	-0.0606	0.0241	0.0039	0.0118	0.0041	0.0043
(25, 30)	-0.0077	-0.0878	-0.0611	0.0234	0.0038	0.0118	0.0041	0.0042
(30, 20)	-0.0021	-0.0887	-0.0588	0.0199	0.0040	0.0116	0.0040	0.0036
(30, 25)	-0.0048	-0.0910	-0.0602	0.0204	0.0038	0.0118	0.0040	0.0036
(30, 30)	-0.0075	-0.0908	-0.0611	0.0205	0.0037	0.0118	0.0041	0.0037

Table 6: Average length of the confidence interval and coverage probability of the simulated 95% confidence intervals of R for $\gamma = 1$ and $\theta = 0.9$

(β_1, β_2)								
(m, n)	Average Confidence Length				Coverage Probability			
	(0.2,0.4)	(0.4, 0.2)	(0.5, 0.1)	(0.4, 1.2)	(0.2, 0.4)	(0.4, 0.2)	(0.5, 0.1)	(0.4, 1.2)
(20, 20)	0.3346	0.3505	0.3111	0.3239	0.9592	0.9839	0.9991	0.9805
(20, 25)	0.3167	0.3325	0.2960	0.3074	0.9585	0.9727	0.9898	0.9530
(20, 30)	0.3043	0.3201	0.2853	0.2956	0.9583	0.9611	0.9794	0.9299
(25, 20)	0.3179	0.3334	0.2951	0.3063	0.9582	0.9809	0.9799	0.9793
(25, 25)	0.2991	0.3145	0.2791	0.2889	0.9573	0.9613	0.9999	0.9533
(25, 30)	0.2860	0.3012	0.2675	0.2764	0.9568	0.9434	0.9696	0.9231
(30, 20)	0.3064	0.3215	0.2840	0.2939	0.9574	0.9720	0.9599	0.9783
(30, 25)	0.2867	0.3019	0.2671	0.2759	0.9557	0.9533	0.9599	0.9593
(30, 30)	0.2728	0.2879	0.2551	0.2631	0.9542	0.9264	0.9399	0.9333

7. Conclusion

In this paper, a generalization of LBWGU distribution namely MOLBWGU distribution is developed. Some of the statistical properties of the new distribution such as probability density function, hazard rate function, moments, quantile function, compounding, distribution of order statistics, Renyi and Shannon entropies are derived. We estimated the parameters of the distribution using maximum likelihood estimation method and a simulation study is conducted for proving the validity of the estimates.

Also we developed a minification process using the model and explored its sample path behavior for different combinations of parameters. To check the impact of stress on strength of devices and systems, the stress strength analysis is carried out and the estimate of the reliability is examined based on a simulation study.

References

- [1] Afify, A. Z., Cordeiro, G. M., Yousof, H. M., Saboor, A., and Ortega, E. M. (2018). The Marshall-Olkin additive Weibull distribution with variable shapes for the hazard rate. *Haceteppe Journal of Mathematics and Statistics*, 47 (2), 365-381.
- [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] Bhatt, M. B. (2014). Characterization of generalized uniform distribution through expectation. *Open Journal of Statistics*, 4 (8), 563-569.
- [4] Ghitany, M. E., Al-Awadhi, F. A., and Alkhalfan, L. (2007). Marshall Olkin extended Lomax distribution and its application to censored data. *Communications in Statistics - Theory and Methods*, 36 (10), 1855-1866.
- [5] Ghitany, M. E., Al-Hussaini, E. K., and Al-Jarallah, R. A. (2005). Marshall Olkin extended Weibull distribution and its application to censored data. *Journal of Applied Statistics*, 32 (10), 1025-1034.
- [6] Gupta, R. C., Ghitany, M. E., and Al-Mutairi, D. K. (2010). Estimation of reliability from Marshall Olkin extended Lomax distributions. *Journal of Statistical Computation and Simulation*, 80 (8), 937-947.
- [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 (3), 421-439.
- [8] Khan, M. I., and Khan, M. A. R. (2017). Characterization of generalized uniform distribution based on lower record values. *ProbStat Forum*, 10 (1), 23-26.
- [9] Krishna, E., Jose, K. K., Alice, T., and Ristic, M. M. (2013). The Marshall-Olkin Frechet distribution. *Communications in Statistics - Theory and Methods*, 42 (22), 4091-4107.
- [10] Lee, C. (2000). Estimations in a generalized uniform distribution. *Journal of the Korean Data and Information Science Society*, 11 (2), 319-325.
- [11] 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.
- [12] Mathew, J. (2020). Reliability test plan for the Marshall-Olkin length biased Lomax distribution. *Reliability: Theory and Applications*, 15 (2), 36-49.
- [13] Mathew, J., and Chesneau, C. (2020 a). Marshall Olkin length biased Maxwell distribution and its applications. *Mathematical and Computational Applications*, 25 (4), 65.
- [14] Patil, G. P. (2002). Weighted Distributions. *Encyclopedia of Environmetrics*, John Wiley and Sons, New York.
- [15] Prudnikov, A. P., Brychkov, Y. A., and Marichev, O. I. (1986). *Integrals and Series*. Gordon and Breach Sciences, Amsterdam, Netherland.

- [16] Rather, A. A., and Subramanian, C. (2018). Characterization and estimation of length biased weighted generalized uniform distribution. *International Journal of Scientific Research in Mathematical and Statistical Sciences*, 5 (5), 72-76.
- [17] Ristic, M. M., and Kundu, D. (2015). Marshall-Olkin generalized exponential distribution. *Metron*, 73 (3), 317-333.
- [18] Subramanian, C., and Rather, A. A. (2018). Transmuted generalized uniform distribution. *International Journal of Scientific Research in Mathematical and Statistical Sciences*, 5 (5), 25-32.
- [19] Tiwari, R. C., Yang, Y., and Zalkikar, J. N. (1996). Bayes estimation for the Pareto failure-model using Gibbs sampling. *IEEE Transactions on Reliability*, 45 (3), 471-476.

A NOVEL METHOD FOR SOFTWARE BUG REPORT ASSIGNMENT

LUKASZ CHMIELOWSKI^{1*,2}, PAVLO KONSTANTYNOV¹,
RYSZARD LUCZAK¹, MICHAL KUCHARZAK^{1,2}, ROBERT BURDUK²

•

^{1*}Nokia Solutions and Networks sp. z o.o., Poland

lukasz.chmielowski@nokia.com

pavlo.konstantynov.ext@nokia.com

ryszard.luczak@nokia.com

michal.kucharzak@nokia.com

² Wroclaw University of Science and Technology, Poland

robert.burduk@pwr.edu.pl

Abstract

During the development of software and electronic devices, it is inevitable to make mistakes. In large, developed companies, assigning a request to the right development team or even a department is not an easy task. Often, the creation of software bug reports and assignment to groups is also formalized by appropriate processes. The paper presents a novel method of software bug report assignment to a group of developers or analysts. A specific usage of organizational structure at the company is a key component of the proposed approach. There are presented results from real use application including both machine learning predictions and human decisions. Human predictions are not independent, the issues are raised as to why comparing the results of machine learning models with those of humans may be inappropriate and what factors influence human decisions. The work also covers conclusive research about potential benefits of the application of automated assignment of bug reports.

Keywords: Software bug assignment, Software bug triaging, Software bug report, Software bug, Text analysis

1. INTRODUCTION

Discussed problem concerns about assigning a software bug report to correct group automatically based on given data like description, system information in raw format or already processed by analyzing tools. Approaches similar to these presented in the paper may be applied to situations which work with other software development cases like related to feature requests, supporting questions or similar issues which should be handled during software development or maintenance. The approach might also be applied to any other different task related to machine learning tasks like classification or labeling in a similar context. It is expected that if a software bug occurs at unit testing level it should be handled by one of developers responsible for development of this unit. More challenging part is when a software bug occurs at the later stage of development or even in real customer use. For large and complex systems even pointing out department can be a complicated task [6]. Additionally, there is an assumption that the company is divided into at least two organization levels, like departments and divisions, as shown in Figure 1. Please note that the names "department" and "division" and relations between them are shown in Figure 1

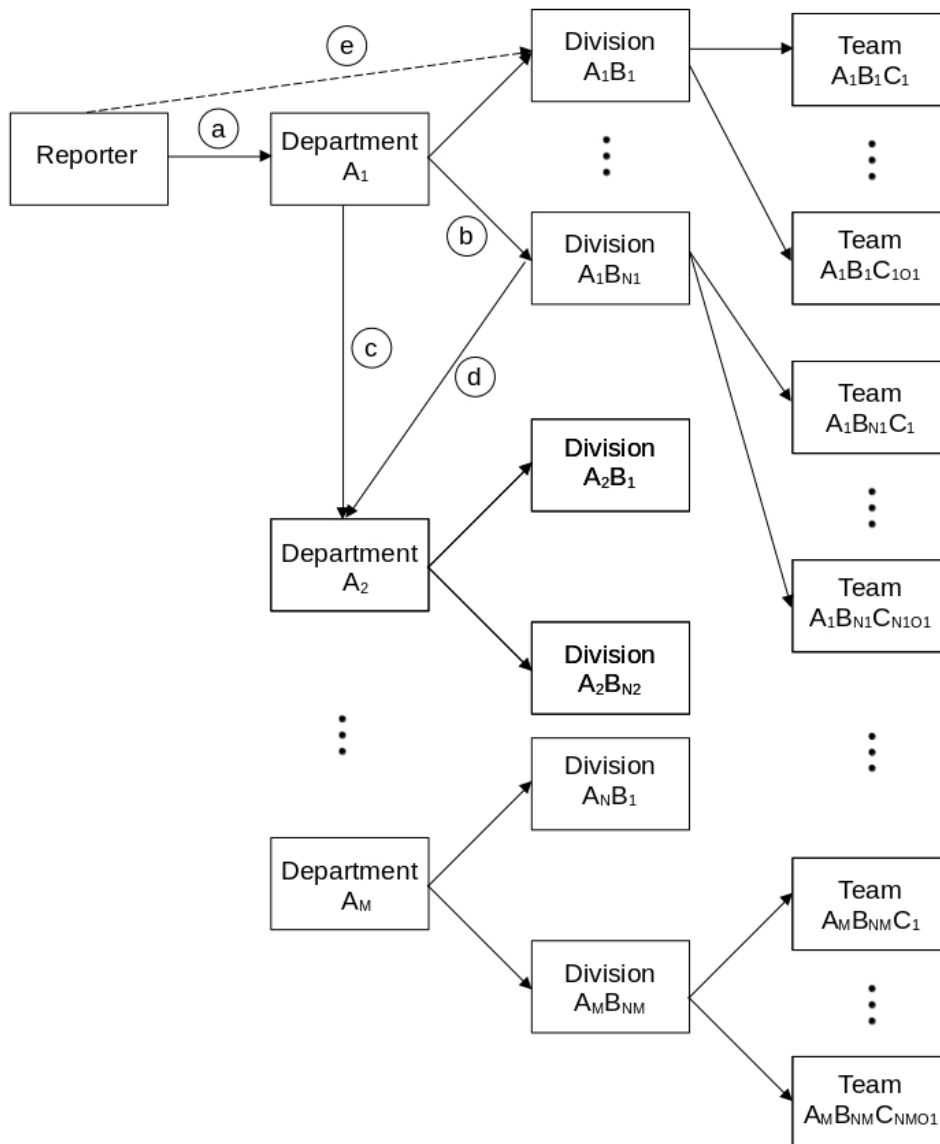


Figure 1: Flow chart of process of transferring reports of bugs inside company, 3 layers shown.

serve only to better illustrate the example. In real use cases testers or customer support engineers decide which department is most suitable for resolving issue (a), while reporting a software defect or anomaly. Next, a report is being assigned to one of the divisions inside the current department (b) or transferred to another one (c). The problem of assigning a software bug report to correct group in that context may be interpreted as:

- assigning to department (a) or (c),
- assigning to division in context of department (b),
- assigning to division directly (e).

2. RELATED WORKS

There is a plethora of ways to classify issues, i.e., classifying severity [7] or assigning the issue to a group which should handle particular case. As there may be numerous bug reports, not all of

them are handled simultaneously. Among others based on classification of severity decisions are made as to whether the bug will be fixed now, later or never. In [8] an approach to assign issue to specified components is presented. In above-mentioned work authors predict if reassignment of created bug report will occur. For that purpose they are using data which come from major projects Eclipse and Mozilla. In [2], bug report assignment is done directly to developers. In the scope was to build time oriented expert model which assigns more priority to developer who had worked on the similar bugs in the past. There are created activity profiles of people who deliver corrections with usage of factor for normalizing which uses the time of last usage of term by developer. [4] addresses bug report assignment to departments. It uses a specific time dependencies for creating train and test sets. Also one of the scope of that work was to investigate the impact of different way of preprocessing and vectorization on bug assignment accuracy. [9] considers bug report assignments to development teams. Moreover, the approach presented in [9], uses only selected cases to automate bug triaging. Selection is made based on confidence of the prediction. A threshold (cutoff) for the confidence is used as there exists a trade-off between accuracy and the number of predictions. [5] presents dual-output deep neural network which simultaneously predicts developer and team. Authors of that work also indicate the fact that this approach is robust against organizational changes as relations between teams and developers may change. In different kind of applications like disease detection and classification, a hierarchical concept of combination of machine learning models is used in [1]. There are two layers used. The purpose of the first of them is disease detection and the second one is its classification. The results also suggest that the hierarchical approach can outperform the flat one, especially in case of small amount of data. A general concept of hierarchical classification algorithms is described in [3]. It presents among others different type of structures for hierarchical problems.

Although in [9] was introduced the possibility of transferring bug reports to selected organization parts based on the thresholds, there is lack of publication which applies specific context and additional possibilities which can be gained due to known hierarchical structure, like for instance combination of models related to different levels especially with usage of tuning of thresholds.

3. RESEARCH QUESTIONS

- What are results of human predictions on department and inside department levels?
- What is the relation between results of machine learning predictions versus human predictions?
- What are the factors impacting human predictions?
- What are minimal requirements for solution applicability in software development company?
- What are the scenarios of deployment of application?
- What are main advantages and disadvantages of human versus computerized approaches?

4. PROPOSED SOLUTION

The novelty is that the cases incoming into department are being transferred into divisions (operation (b) in Figure 1) only in specific conditions. In that case a novel combination of machine learning models may be used, where one of the models predicts which division the issue should be addressed to, whereas the second one predicts whether the current department is the proper one. As a result, it is a transfer to proper division only in case if both of used models exceeded respective threshold. By the threshold we understand the cutoff confidence score based on the output of machine learning model. The novel features can be expressed as follows:

- assigning only selected cases of all cases incoming into department (or created inside) to specific divisions based on confidence level of prediction or state of issue,
- creating a model or other decision system based on at least two classification models, where at least one of them predicts division and at least one of them predicts department.

That approach is general and it is not limited to one way of creating a model. In a specific implementation in the company, predictions come from models prepared in a way similar to the approach presented in [4]. Fields used from bug report are title, description, product and release. The preprocessing phase uses methods for cleaning the text like removing chosen special characters, changing to lower letters, removing stopwords and part of content related to company template. The training set is built from data related to relevant cases from last 365 days up to date of creation of model. For each case the result of prediction was collected at the time when formally correct bug report appeared for the first time at A_1 . Different models are used for predictions at different levels of organization. Department is being predicted with the use of logistic regression classifier; division with use of support vector classifier with linear kernel. This production setup is being updated daily to get the newest available data for training as fast as possible.

5. RESULTS

The doctoral student conducted research on the possibility of automatic assignment of bug reports in selected cases. The results of the studies show what the effect on historical data would be if bug reports were sent to inside department from interface group A_1 . The predictions come from the cases which passed formal check of correctness of report filling at the time of use of machine learning model to make prediction and contain valid log content. Table 1 shows the number of cases above a certain threshold of a model which would be sent predicting department and respective precision. Similar research was conducted for model which predicts divisions inside the department. Table 2 indicates the number of cases which would be transferred from group A_1 to A_1B_1 and its precision. The results with combination of those two models are placed in Table 3. Based on above data, the decision about implementation pilot solution with thresholds 0.6 for department, 0.3 for division was made. Within that solution problem reports which meet the above-mentioned requirements were transferred. For those cases which did not meet these conditions were only placed information about suggested transfer possibilities. Selected results are presented in Tables 4 to 6 and Figures 2 to 4. Presented data do not show cases which were created at early stages of software development and discovered in later phases or even in customer use. The following notations are used in Table 6:

- Human only - percentage of cases where human prediction was correct, but ML model prediction was incorrect;
- ML only - percentage of cases where ML model prediction was correct, but human model prediction was incorrect;
- ML & Human - percentage of cases where both ML and human model predictions were correct;
- Both incorrect - percentage of cases where neither ML nor human model predictions were correct.

Additionally, we can see the benefits like that for cases in date ranges from November 2021 to January 2022 where the decision about transfers was made 79% of them were resolved¹ (or fix was not required²) inside department A_1 , but for cases where only decision about suggestion was made only 66% were resolved (or fix was not required) inside.

¹Resolved - resolved; not including internal department cases; ended inside department

²Fix not required - fix not required; not including internal department cases; ended inside department including inflow group A_1

Table 1: General flow of issues in organization.

Threshold set	Cases predicted as A_1	Precision of A_1 in the context of cases above threshold
0.2	266	58
0.3	244	61
0.4	183	64
0.5	130	68
0.6	71	67
0.7	48	71

Table 2: General flow of issues in organization.

Threshold set	Cases predicted as $A_1 B_1$	Percent of cases $A_1 B_1$ among accepted ended on A_1	Percent of cases $A_1 B_1$ among accepted
0.2	206	39	20
0.3	148	40	23
0.4	95	42	23
0.5	53	38	23
0.6	28	39	25

6. DISCUSSION ON REQUIREMENTS FOR APPLICATION OF SOLUTION INSIDE COMPANY

6.1. Minimal requirements

Although many people at glance think that such solutions have opportunities to be introduced only in case the predictions are better than human ones, this is not so simple as it is thought. In the case when machine learning predictions are better than it is rather obvious that is worth to make it application. Otherwise when it is worse, in some cases it may be also worth developing and applying such solutions. One of the reasons is that it may work as decision supporting system which does not make a binding decision, but only delivers suggestions which may be helpful for cases when a reporter has no idea how to address the problem, and sometimes may ignore suggestions when is sure where to address that or knows that the suggested target is wrong. Sometimes an issue with group overloading may occur, like for instance they currently handle too many bug fixes simultaneously, or have to deliver already committed new important features to product. Then, from the businesses perspective it may be reasonable to redirect cases to groups where it is less likely that the corrections will be delivered, but they may deliver detailed analysis or reject bug report as not valid. That effect may be gained due to tuning of mentioned in this work thresholds. Even if this change could lead to accuracy reduction, it may help to achieve business goals.

Table 3: General flow of issues in organization.

Threshold set for department A_1	Threshold set for A_1B_1	Cases predicted as A_1B_1	Precision of A_1B_1 in the context of cases above threshold and ended in A_1	Precision of A_1B_1 in the context of cases above threshold
0.3	0.3	88	36	25
0.3	0.4	60	38	26
0.3	0.5	40	38	25
0.3	0.6	26	41	27
0.4	0.3	65	34	25
0.4	0.4	44	34	25
0.4	0.5	31	34	26
0.4	0.6	21	33	24
0.5	0.3	49	40	29
0.5	0.4	34	40	29
0.5	0.5	21	43	33
0.5	0.6	13	44	30
0.6	0.3	29	57	41
0.6	0.4	20	64	45
0.6	0.5	11	75	55
0.6	0.6	7	80	57

Table 4: Chosen results transferred cases based on ML model decision.

Date range	type of resolution	Accuracy
November and December	Resolved	38 %
November and December	Fix not required	55 %
December	Resolved	50 %
December	Fix not required	80 %

6.2. Human factors

There are usually many validation aspects when comparing machine learning and human predictions in software bug report assignment process. Some of them are presented in the following paragraph. One of the most important ones is that the reporters may use an already introduced decision supporting system. The different aspect is that in cases when multiple groups delivered correction people can choose which one will be the final main one and may want to boost human or machine learning result if they wish so. What is more, reporters sometimes ask before creating reports where reports should be sent before creating formally one. At that step, many developers might be involved or even the solution might be known before the actual report is officially processed. Sometimes developer teams ask for verification of some functionality and create a report directly against them. In those last two cases the final group is known even before creating bug report. What is more, not always the best accuracy is the aim of introduction of such solutions. Last, but not least, recently detailed instructions on how to address the most common types of bug reports and responsibilities of divisions inside department were made to improve human decision making.

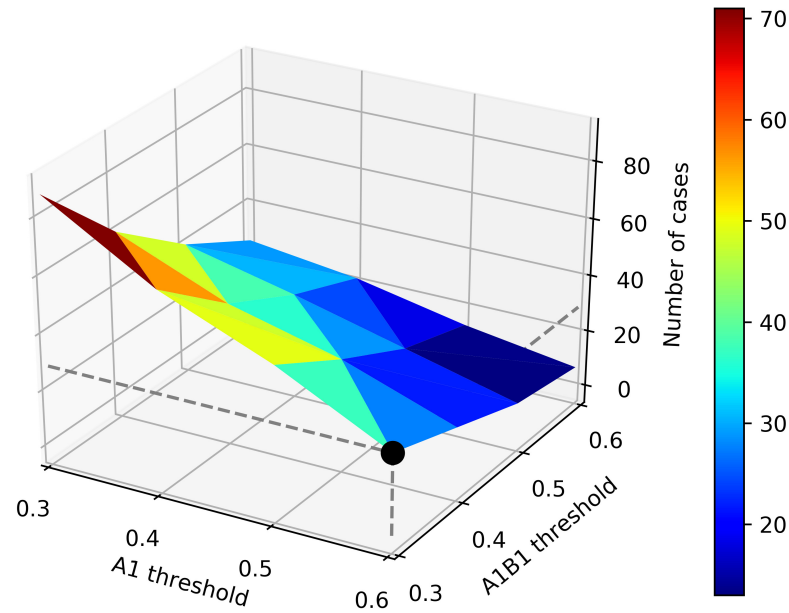


Figure 2: Number of bug reports meet threshold conditions in function of given thresholds.

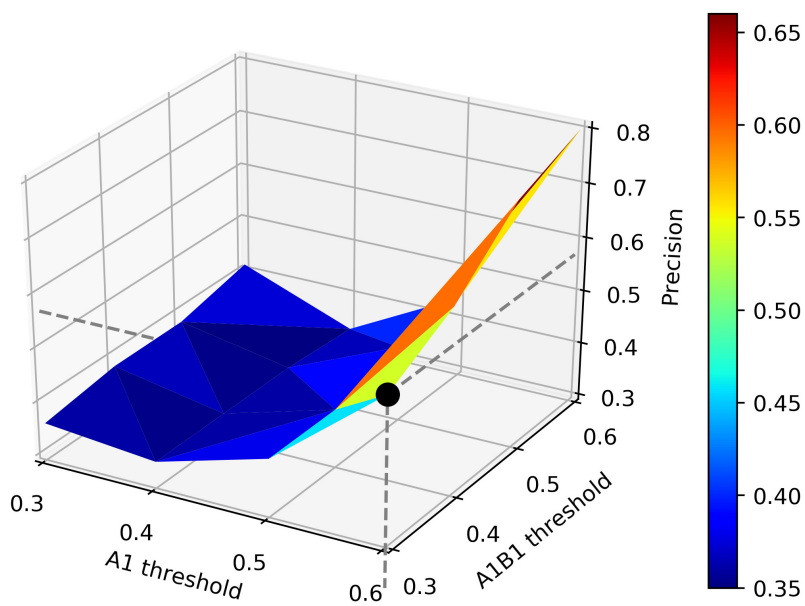


Figure 3: Precision of predictions of bug reports meet threshold conditions in function of given thresholds excluding cases which ended outside of department A_1 .

Table 5: Chosen results of suggested cases based.

Date range	type of resolution	Human accuracy	ML model accuracy
November and December	Resolved	54 %	38 %
November and December	Fix not required	65 %	35 %
December	Resolved	56 %	38 %
December	Fix not required	64 %	32 %

Table 6: Distribution of specific types of results of suggested cases.

Date range	type of resolution	Human only	ML only	ML & Human	Both incorrect
Nov and Dec	Resolved	31 %	15 %	23 %	32 %
Nov and Dec	Fix not required	40 %	10 %	25 %	25 %
Dec	Resolved	34 %	16 %	22 %	28 %
Dec	Fix not required	40 %	8 %	24 %	28 %

6.3. Advantages and disadvantages of such solutions

The main disadvantage of such solutions, which are often pointed out during automatic transfers, is the lack of analysis that would provide information on why such a decision was made. The second issue that will help minimize these defects is the implemented solution, which conveniently displays to developers' key information about the content of the base station configuration state logs at the time of collecting logs, provided of course the logs have collected in the correct way.

7. NEXT STEPS THAT HAVE BEEN TAKEN

Referring to the progress of implementation in industry, prepared an earlier pilot solution supporting the group A_1 , dealing with the handling of applications within the department, was gradually extended to handle more bug reports. The solution was to transfer selected bug reports from the group symbolically marked as A_1 to selected groups in the conditions specified by the machine learning model and if met the formal conditions for notification. This decision shall be taken automatically without the need for human verification. The model prepared was also used for transfers from department A_2 to the teams (A_1B_x) responsible for analysis in department A_1 as well as transfers to department A_2 . One of the models prepared is also used as one of the component models to solve suggesting to the reporter whether the report should be opened against department A_2 . In addition, it is also used as a component model for the automatic transfer solution applications between departments A_2 and A_3 . It was decided to remove a group responsible for initial investigation inside that department A_1 in June this year, thus fully abandoning one layer of analysis. In connection with these changes, the solution was adapted so that the submitted applications from department B were sent directly to groups A_1B_x . In addition, the system had to be adapted to indicate new groups after organizational changes, because on this layer the structure has also changed.

8. SUMMARY

The paper discusses problems related to methods of assignment of reports, feature requests, supporting questions or similar issues to group of employees, developers, organization unit, etc. The novelty introduced in this paper is related to the specific usage of the organizational structure in processes of handling (assigning) an issue. The paper shows possible scenarios of deployment of application supporting fault management with the use of solution based on

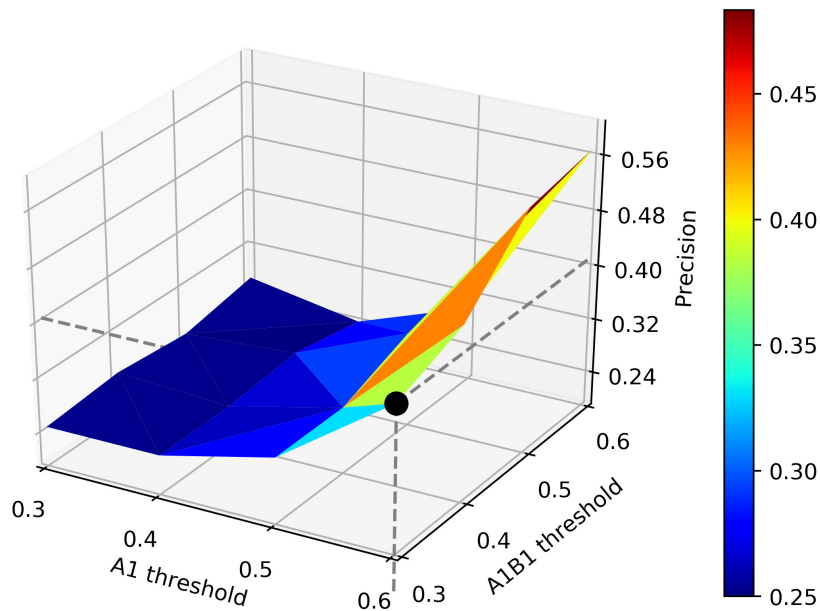


Figure 4: Precision of predictions of bug reports meet threshold conditions in function of given thresholds including cases which ended outside of department A_1 .

machine learning. The study demonstrates factors impacting human predictions, main advantages, and disadvantages of automated solution against human. Comparison of results between human and model predictions at both department and inside department levels are presented. Minimal requirements for the company in case of application of machine learning supporting system in the company are also defined.

ACKNOWLEDGEMENTS

This work has been carried out in cooperation between NOKIA and Wroclaw University of Science and Technology in context of a Ph.D. grant under the fourth edition of the "Implementation Doctorate Programme".

DECLARATION OF CONFLICTING INTERESTS

The Author(s) declare(s) that there is no conflict of interest.

REFERENCES

- [1] Guangzhou An, Masahiro Akiba, Kazuko Omodaka, Toru Nakazawa, and Hideo Yokota. "Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images". In: *Scientific Reports* 11.1 (Nov. 2021), p. 4250. ISSN: 2045-2322. DOI: 10.1038/s41598-021-83503-7. URL: <https://doi.org/10.1038/s41598-021-83503-7>.
- [2] Anjali, Devina Mohan, and Neetu Sardana. "Visheshagya: Time based expertise model for bug report assignment". In: *2016 Ninth International Conference on Contemporary Computing (IC3)*. 2016, pp. 1–6. DOI: 10.1109/IC3.2016.7880218.

- [3] Helyane Bronoski Borges, Carlos N. Silla, and Jlio Cesar Nievola. "An evaluation of global-model hierarchical classification algorithms for hierarchical classification problems with single path of labels". In: *Computers & Mathematics with Applications* 66.10 (2013). ICNC-FSKD 2012, pp. 1991–2002. ISSN: 0898-1221. DOI: <https://doi.org/10.1016/j.camwa.2013.06.027>. URL: <https://www.sciencedirect.com/science/article/pii/S089812211300432X>.
- [4] Lukasz Chmielowski and Michal Kucharzak. "Impact of Software Bug Report Preprocessing and Vectorization on Bug Assignment Accuracy". In: *Progress in Image Processing, Pattern Recognition and Communication Systems*. Ed. by Michal Choras, Ryszard S. Choras, Marek Kurzynski, Pawel Trajdos, Jerzy Pejas, and Tomasz Hyla. Cham: Springer International Publishing, 2022, pp. 153–162. ISBN: 978-3-030-81523-3. DOI: 10.1007/978-3-030-81523-3_15.
- [5] Christopher A. Choquette-Choo, David Sheldon, Jonny Proppe, John Alphonso-Gibbs, and Harsha Gupta. "A Multi-label, Dual-Output Deep Neural Network for Automated Bug Triaging". In: *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. 2019, pp. 937–944. DOI: 10.1109/ICMLA.2019.00161.
- [6] Mainak Dutta. *BEFORE AND AFTER OF DevOps: A PEEK INTO AGILE DevOps*. Nov. 2019. URL: <https://medium.com/%7B%5C%7Dmainakdutta76/before-and-after-of-devops-a-peek-into-agile-devops-3600c26129ac>.
- [7] S. Gujral, G. Sharma, S. Sharma, and Diksha. "Classifying bug severity using dictionary based approach". In: *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*. 2015, pp. 599–602. DOI: 10.1109/ABLAZE.2015.7154933.
- [8] A. Lamkanfi and S. Demeyer. "Predicting Reassignments of Bug Reports - An Exploratory Investigation". In: *2013 17th European Conference on Software Maintenance and Reengineering*. 2013, pp. 327–330. DOI: 10.1109/CSMR.2013.42.
- [9] Aindrila Sarkar, Peter C. Rigby, and Bela Bartalos. "Improving Bug Triaging with High Confidence Predictions at Ericsson". In: *2019 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. 2019, pp. 81–91. DOI: 10.1109/ICSME.2019.00018.

A NEW TWO PARAMETRIC GENERALIZED DIVERGENCE MEASURE AND ITS RESIDUAL

FAYAZ AHMED



Department of Statistics, University of Kashmir, Srinagar, India
fayazahmed4095@gmail.com

MIRZA ABDUL KHALIQUE BAIG



Department of Statistics, University of Kashmir, Srinagar, India
baigmak@gmail.com

Abstract

In this research article, we proposed a new two-parametric divergence measure and developed its weighted version. We also looked at its properties and specific cases with examples and also obtained some results and bounds for new two parametric weighted generalized divergence measure. With the aid of a numerical example that determine the distribution function and also studied some inequality for the new proposed divergence measure. The known divergence measure is the particular case of our proposed measure. The proposed measure uniquely characterized the distribution function using the proportional hazard rate model (PHRM). Its residual function is also being worked on.

Keywords: characterization result, Divergence measure, distribution function, proportional hazard rate model, Residual function.

1. INTRODUCTION

The idea of information measures plays an important role in the feild of information theory and other applied sciences. The conception of information measure(uncertainty) was first given by Shannon entropy [10]. He suggested a method for achieving the probability distribution inherent uncertainty and established that it is a main component of information theory, which today has numerous applications across many fields.

Suppose X is a hypothetical continous non-negative random variable, then the shannon's [10] entropy is defined as

$$H^S(X) = - \int_0^{\infty} f(x) \log f(x) dx \quad (1)$$

where f represent density function of X .

In addition, it can be written as

$$H^S(X) = E(-\log f(x))$$

The $H^S(X)$ is equal to the expected value of $(-\log f(x))$.

The concept of the distinction between the two density distributions is obtained from their function and is often used in Shannon entropy. The most common divergence applied in information theory is the Kullback-Leibler divergence, also known as relative entropy or Kullback-Leibler information divergence measure[4] (KL divergence). It is widely used in parameter estimation, contingency tables and ANOVA tables.

Suppose $f(x)$ and $g(x)$ are the two probability distributions for a continuous random variable X and Y , then the KL divergence[4] is given by.

$$D^{KL}(f||g) = \int_0^{\infty} f(x) \log \frac{f(x)}{g(x)} dx \quad (2)$$

Furthermore, it can be written as

$$D^{KL}(f||g) = E^{KL} \left[\log \frac{f(x)}{g(x)} \right]$$

Remarks

1. If $g(x)=1$, then it simply becomes Shannon's[10].
2. If $g(x)= f(x)$, then Kullback-Leibler[4] divergence is reduced to zero.

In this regard, Renyi[9] presented the generalization of Kullback-Leibler divergence[4] of order β , which is defined as

$$D^R(f||g) = \frac{1}{\beta-1} \log \int_0^{\infty} f(x)^\beta g(x)^{1-\beta} dx, \quad \beta \neq 1, \quad \beta > 0 \quad (3)$$

Remarks

1. If $\beta \rightarrow 1$, then it just becomes Kullback-Leibler[4] divergence. Since many researchers have developed distinct generalizations of Kullback-Leibler divergence[4] in distinct manners, Gupta and Nanda[8] introduced a new generalization of the KL divergence measure of order " β " defined as follows.

$$D^G(f||g) = \frac{1}{\beta-1} \log \int_0^{\infty} \left(\frac{f(x)}{g(x)} \right)^{\beta-1} f(x) dx, \quad \beta \neq 1, \quad \beta > 0 \quad (4)$$

Remarks

1. If $\beta \rightarrow 1$, then it simply becomes Kullback-Leibler divergence[4] Our goal is to develop a new Kullback-Leibler divergence[4] measure that seems to be two-parametric in nature.

Our proposed measure is defined as follows

$$D^{\alpha,\beta}(f||g) = \frac{1}{\beta-\alpha} \log \int_0^{\infty} \left(\frac{f(x)}{g(x)} \right)^{\beta-\alpha} f(x)^\alpha dx, \quad \beta \neq \alpha, \quad \alpha \geq 1, \quad \alpha, \beta > 0 \quad (5)$$

Additionally, it can be expressed as

$$D^{\alpha,\beta}(X||Y) = \frac{1}{\beta-\alpha} \log \int_0^{\infty} (f(x))^\beta (g(x))^{\alpha-\beta} dx \quad (6)$$

Remarks

1. When we take $f(x) = g(x)$ then divergence became zero
2. if $\alpha = 1$ (5) reduced to the Gupta and Nanda[8] of order β
3. if $\alpha = 1, \beta \rightarrow 1$ it converge to simply Kullback-Leibler divergence[4]
4. if $g(x) = 1$, (5) reduces to simply Shannon entropy[10]

The rest of the article consists of the following sections. In Section (2) we study the weighted generalized divergence measure and in Section (3) we study the weighted generalized residual divergence measure. In Section 4 we derive some characterization results using the proportional hazard rate model. Finally we studied some properties and bound the new proposed divergence measure.

2. WEIGHTED GENERALIZED DIVERGENCE MEASURE

We proposed the weighted generalised divergence measure in this section. Shannon’s measure and Kullback-Leibler divergence[4] both give the random variable identical weight in real-world scenarios, but this can be problematic. To solve that problem, weighted entropy first developed by Belis and Guiasu [1] . The weighted entropy is defined as

$$H_w^S(X) = - \int_0^{\infty} x f(x) \log f(x) dx \tag{7}$$

Remarks

1. If $x=1$, then, it becomes simply Shannon entropy[10].

A factor x that gives more weight to the random variable’s higher value provides as the expression of the weight function. Shift-dependent is the term for this measure. Suhov and Yasaei Sekeh Mirali et al.[11], Mirali et.al[6] and Rajesh et al.[7] and many researchers who have presented numerous weighted measures.

Based on the idea of weighted entropy, in recent years, Yasaei Sekeh et al.[12] gave weight to the Kullback-Leibler divergence[4] divergence, defined as

$$D_W^{KL}(f||g) = \int_0^{\infty} x f(x) \log \frac{f(x)}{g(x)} dx \tag{8}$$

Remarks

1. If $x=1$ then, it becomes simply Kullback-Leibler divergence[4].

Additionally, it can be written as

$$D_{KL}^w(f||g) = E_{KL} \left[X \log \frac{f(x)}{g(x)} \right]$$

Definition 2.1. Similar to (2) and based on (8) the weighted proposed measure is defined as

$$D_w^{\alpha,\beta}(X||Y) = \frac{1}{\beta - \alpha} \log \int_0^{\infty} x (f(x))^{\beta} (g(x))^{\alpha - \beta} dx \tag{9}$$

Remarks

1. If $x = 1$ then, it becomes reduced to (6).

The following example show how GDM and its weighted form differ from one another.

Example 2.1. Suppose X and Y are two non-negative continuous random variables with the density function as follows

1. $f_1(x) = 1, \quad 0 < x < 1$ and $g_1(x) = 2x \quad 0 < x < 1$

2. $f_2(x) = 1, \quad 0 < x < 1$ and $g_2(x) = 2(1 - x) \quad 0 < x < 1$

subsequently, the distribution function was characterized by the weighted generalized divergence measure.

Using (5) after simplification, we get

$$D^{1(\alpha,\beta)}(f||g) = \frac{1}{(\beta - \alpha)} \log 2^{\alpha-\beta} \left(\frac{1}{\alpha - \beta + 1} \right) = D^{2(\alpha,\beta)}(f||g) \tag{10}$$

Again using (9) after simplification, we get

$$D_w^{1(\alpha,\beta)}(f||g) = \frac{1}{(\beta - \alpha)} \log 2^{\alpha-\beta} \left(\frac{1}{\alpha - \beta + 2} \right) \tag{11}$$

and

$$D_w^{2(\alpha,\beta)}(f||g) = \frac{1}{(\beta - \alpha)} \log 2^{\alpha-\beta} \left(\frac{\Gamma(\alpha - \beta + 1)}{\Gamma(\alpha - \beta + 3)} \right) \tag{12}$$

Where, $t = n(\beta-\alpha)+1, \quad s = (\beta-\alpha)$, and $B(u,v) = \int_0^1 x^{u-1}(1-x)^{v-1}dx = \int_0^1 \frac{x^{u-1}}{(1+x)^{u+v}} = \frac{\Gamma u \Gamma v}{\Gamma(u+v)}$

which is known as complete beta function.

We can see from the above example that our proposed measure has the same value without weight but a different value when given weight,hence we draw the conclusion that the weighted measure uniquely determines the distribution.

3. WEIGHTED GENERALIZED RESIDUAL DIVERGENCE MEASURE

In this section, we discuss the generalized residual divergence measure and the weighted generalized residual measure. Shannon’s measure and the Kullback-Leibler divergence measure are inapplicable to a component that has survived for some units of time. To overcome this problem, [3] introduced a measure known as the residual measure of entropy. This measure of uncertainty for a random variable’s remaining lifetime $X_t = [X - t|X > t]$ is defined as

$$H^S(X;t) = - \int_t^\infty \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx \tag{13}$$

Remarks

1. If $t = 0$, then it simply becomes Shannon entropy[10].

The weighted form is represented by (13)[2] and defined as

$$H_w^S(X;t) = - \int_t^\infty x \frac{f(x)}{\bar{F}(t)} \log \frac{f(x)}{\bar{F}(t)} dx \tag{14}$$

Definition 3.1. The generalized residual divergence measure is defined as follows in accordance with (13) and based on (14)

$$D^{KL}(X||Y;t) = \int_t^{\infty} \frac{f(x)}{\bar{F}(t)} \log \left(\frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)} \right) dx \quad (15)$$

Remarks

1. If $t=0$ then it simply becomes Kullback-Leibler divergence[4].

The generalized residual divergence measure's weighted form is defined as

$$D_w^{KL}(X||Y;t) = \int_t^{\infty} x \frac{f(x)}{\bar{F}(t)} \log \left(\frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)} \right) dx \quad (16)$$

Remarks

1. If $x=1$ then it reduced to (15).

To proceed in this manner, the generalized residual divergence measure Gupta and Nanda[8], which is defined as

$$D^G(X||Y;t) = \frac{1}{\beta-1} \log \int_t^{\infty} \frac{f(x)}{\bar{F}(t)} \left(\frac{f(x)/\bar{F}(t)}{g(x)/\bar{G}(t)} \right)^{\beta-1} dx \quad (17)$$

It can be summarised as

$$D^G(X||Y;t) = \frac{1}{\beta-1} \log \int_t^{\infty} \left(\frac{f(x)}{\bar{F}(t)} \right)^{\beta} \left(\frac{g(x)}{\bar{G}(t)} \right)^{1-\beta} dx \quad (18)$$

We proposed the weighted form of (18), which is similar to (17) and is defined as

$$D_w^G(X||Y;t) = \frac{1}{\beta-1} \log \int_t^{\infty} x \left(\frac{f(x)}{\bar{F}(t)} \right)^{\beta} \left(\frac{g(x)}{\bar{G}(t)} \right)^{1-\beta} dx \quad (19)$$

Definition 3.2. Let X and Y be two non-negative random variables. The proposed weighted residual divergence measure is defined as

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta-\alpha} \log \int_t^{\infty} x \left(\frac{f(x)}{\bar{F}(t)} \right)^{\beta} \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (20)$$

Remarks 1. If $x=1$ and $\alpha=1$ then reduced to (18)

4. CHARACTERIZATION RESULT

Proportional hazard rate model (PHRM)

Definition 4.1. Assume that X and Y are two non-negative random variables with a survival function $\bar{F}(x)$ and $\bar{G}(x)$ respectively. Then the following relationship holds as

$$\bar{G}(x) = \bar{F}(x)^\mu \quad (\mu > 0) \quad \text{or} \quad \lambda_G(x) = \mu \lambda_F(x)$$

This mathematical relationship is crucial in the development of various statistical models. Many statistical disciplines, including medicine, reliability, economics, and survival analysis, employ this approach.

Theorem 1. For any $t > 0$, the following equality is true if random variables X and Y satisfy the proportional hazard rate model with proportionately constant $\mu > 0$.

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta - \alpha} \log \left(t \exp(\beta - \alpha) D_G(X||Y;t) + \int_t^\infty \left(\frac{\bar{F}(v)}{\bar{F}(t)} \right) \exp(\beta - \alpha) D_G^w(X||Y;t) dv \right) \quad (21)$$

Proof. Rewriting the (20) to

$$\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx = \int_t^\infty \left(\int_0^x v^0 dv \right) \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (22)$$

$$\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx = \int_t^\infty \left(\int_0^t v^0 dv + \int_t^x v^0 dv \right) \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (23)$$

$$\begin{aligned} \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx &= t \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \\ &+ \int_{v=t}^\infty \left(\int_{x=v}^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \right) dv \end{aligned} \quad (24)$$

$$\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx = \exp(\beta - \alpha) D^{\alpha,\beta}(X||Y;t) \quad (25)$$

Using the proportional hazard rate model in (24), we have

$$\int_t^\infty (f(x))^\beta (g(x))^{\alpha-\beta} = (\bar{F}(t))^\beta (\bar{G}(t))^{\alpha-\beta} \exp(\beta - \alpha) D^{\alpha,\beta}(X||Y;t) \quad (26)$$

$$\int_t^\infty (f(x))^\beta (g(x))^{\alpha-\beta} = (\bar{F}(t))^{\mu(\alpha-\beta)+\beta} \exp(\beta - \alpha) D^{\alpha,\beta}(X||Y;t) \quad (27)$$

using (23),(24) and (26) in (20) we obtained the desired result. ■

Theorem 2. If the proportional hazard rate model is satisfied by the two random variables X and Y and proportionally constant $\mu > 0$, and $D_w^{\alpha,\beta}(X||Y;t)$ is increasing for all $t > 0$, then $D_w^{\alpha,\beta}(X||Y;t)$ exclusively determines the $\bar{F}(t)$.

Proof. Rewriting (20) as

$$\exp(\beta - \alpha) D_G^w(X||Y;t) = \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (28)$$

Diff. (28) w.r.to t we have

$$\begin{aligned} \frac{\partial}{\partial t} \exp((\beta - \alpha) D_G^w(X||Y;t)) &= -t (\lambda_F(t))^\beta (\lambda_G(t))^{\alpha-\beta} \\ &+ \int_t^\infty x (f(x))^\beta (g(x))^{\alpha-\beta} \frac{\partial}{\partial t} (\bar{F}(t) \bar{G}(t)) \end{aligned} \quad (29)$$

$$\begin{aligned} \frac{\partial}{\partial t} \exp((\beta - \alpha)D_G^w(X|Y;t)) &= -t(\lambda_F(t))^\beta (\lambda_G(t))^{\alpha-\beta} \\ &+ [\beta\lambda_F(t) - (\alpha - \beta)\lambda_G(t)] \exp(\beta - \alpha)D_w^{\alpha,\beta}(X|Y;t) \end{aligned} \quad (30)$$

By using the (PHRM)

$$\begin{aligned} \frac{\partial}{\partial t} \exp((\beta - \alpha)D_G^w(X|Y;t)) &= -t(\lambda_F(t))^\beta (\mu\lambda_F(t))^{\alpha-\beta} \\ &+ [\beta\lambda_F(t) - (\alpha - \beta)\mu\lambda_F(t)] \exp(\beta - \alpha)D_w^{\alpha,\beta}(X|Y;t) \end{aligned} \quad (31)$$

If, we fixed $t > 0$ then $\lambda_F(t)$ is the solution of the equation $z(x_t) = 0$ where $z(0) = \exp(\beta - \alpha)D_w^{\alpha,\beta}(X|Y;t) \geq 0$ since we assumed that it is increasing in t if $x_t \rightarrow \infty$ the $z(\infty) = \infty$
 Now diff. w.r.to t we have

$$\frac{\partial}{\partial x_t} z(x_t) = \alpha t(\mu)^{\alpha-\beta}(x_t)^{(\alpha-1)} - [\mu(\alpha - \beta) + \beta] \exp(\beta - \alpha)D_w^{\alpha,\beta}(X|Y;t) \quad (32)$$

then $\frac{\partial}{\partial x_t} z(x_t) = 0$

$$x_t = \left[\frac{[\mu(\alpha - \beta) + \beta] \exp(\beta - \alpha)D_w^{\alpha,\beta}(X|Y;t)}{\alpha t(\mu)^{\alpha-\beta}} \right]^{\frac{1}{\alpha-1}} \quad (33)$$

then we can say that $x_t = x_0$ Here, x_t represents $\lambda_F(t)$ Hence $D_w^{\alpha,\beta}(X|Y;t)$ characterized the survival function uniquely. ■

5. PROPERTIES AND BOUNDS

Log-sum inequality

Definition 5.1. Let X be a bounded integer. If $f(x)$ and $g(x)$ are density functions and x is integrable, then

$$\int f(u) \log \frac{f(u)}{g(u)} du \geq \left[\int f(u) du \right] \log \left[\frac{\int f(u) du}{\int g(u) du} \right]$$

Definition 5.2. If $D_w^{\alpha,\beta}(X|Y;t)$ is increasing in t . Then, the survival function \bar{F} is said to have increasing(decreasing) $D_w^{\alpha,\beta}(X|Y;t)$ of order β and type α expressed as IWGRDM or DWGRDM. It indicate weather \bar{F} is increasing or decreasing if $D_w^{\alpha,\beta}(X|Y;t) \geq (\leq) 0$

Theorem 3. If the X and Y be the two non-negative random variables represent the life spans of two system components with density function $f(x)$ and $g(x)$ and survival functions respectively, $t > 0$, then $D_w^{\alpha,\beta}(X|Y;t)$ for $0 < \beta < \alpha$ where $\alpha = 1$ attains a lower bound as follows

$$D_w^{\alpha,\beta}(X|Y;t) \geq \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \log \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx + \frac{1}{\beta - \alpha} \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \log x dx \quad (34)$$

Proof. With the aid of log-sum inequality, we have

$$\int_t^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \log \frac{\left(\frac{f(x)}{\bar{F}(t)} \right)^\beta}{x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta}} dx \geq \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta dx \log \frac{\int_t^\infty \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta dx}{\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx} \quad (35)$$

$$\int_t^\infty \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \log \frac{\left(\frac{f(x)}{\bar{F}(t)}\right)^\beta}{x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta}} dx = -\log \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \quad (36)$$

$$\int_t^\infty \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \log \frac{\left(\frac{f(x)}{\bar{F}(t)}\right)^\beta}{x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta}} dx = (\beta - \alpha) D_w^{\alpha,\beta}(X||Y;t) \quad (37)$$

From L.H.S. we have

$$\begin{aligned} \beta \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \log \left(\frac{f(x)}{\bar{F}(t)}\right) - \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \log x dx - \beta \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \log \left(\frac{f(x)}{\bar{F}(t)}\right) \\ - (\alpha - \beta) \int_t^\infty \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \log \left(\frac{g(x)}{\bar{G}(t)}\right) dx \end{aligned} \quad (38)$$

Using (36) and (37), we derived the result. ■

Theorem 4. The upper bound $D_w^{\alpha,\beta}(X||Y;t)$ is valid for $0 < \beta < \alpha$ for the random variable X and Y with the support of $(0,k]$ and also with the pdf $f(x)$ and $g(x)$ and survival function respectively, $t > 0$ then

$$D_w^{\alpha,\beta}(X||Y;t) \leq \left[\frac{\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \log x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx}{\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx} + \log(k - t) \right] \quad (39)$$

Proof. We have a log-sum inequality for

$$\begin{aligned} \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \log x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \leq \int_t^k x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \\ \frac{\int_t^k x (f(x))^\beta (g(x))^{\alpha-\beta}}{\int_t^k (\bar{F}(t))^\beta (\bar{G}(t))^{\alpha-\beta} dx} \end{aligned} \quad (40)$$

$$\begin{aligned} \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \log x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx = \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \\ [(\beta - \alpha) D_w^{\alpha,\beta}(X||Y;t) - \log(k - t)] \end{aligned} \quad (41)$$

After simplifying, we got the desired result. ■

Theorem 5. Suppose X and Y be two random variables with weighted generalized residual divergence measure(WGRDM) for $0 < \beta < \alpha$ we get

$$D_w^{\alpha,\beta}(X||Y;t) \leq \frac{1}{\beta - \alpha} \left(\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx - 1 \right) \quad (42)$$

Proof. From the support of this inequality $\log \leq (x - 1)$ we have

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta - \alpha} \log \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)}\right)^\beta \left(\frac{g(x)}{\bar{G}(t)}\right)^{\alpha-\beta} dx \quad (43)$$

$$(\beta - \alpha)D_w^{\alpha,\beta}(X||Y;t) = \log \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (44)$$

$$(\beta - \alpha)D_w^{\alpha,\beta}(X||Y;t) = \left[\int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx - 1 \right] \quad (45)$$

After simplification, we obtained the result. ■

Theorem 6. Consequently, if the hazard rate $\lambda_F(t)$ or risk rate is decreasing in t then

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta - \alpha} \left(\log \int_t^\infty x \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx - \beta \log \lambda_F(t) \right) \quad (46)$$

Proof. The equation (20) is rewritten as

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta - \alpha} \log \int_t^\infty x \left(\frac{f(x)}{\bar{F}(t)} \right)^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (47)$$

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta - \alpha} \log \int_t^\infty x \left(\frac{\bar{F}(x)}{\bar{F}(t)} \right)^\beta (\lambda_F(x))^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (48)$$

Therefore, $\bar{F}(x) \leq \bar{F}(t)$ for $x \geq t$ which implies that $\lambda_F(x) \leq \lambda_F(t)$ then we have

$$D_w^{\alpha,\beta}(X||Y;t) = \frac{1}{\beta - \alpha} \log \int_t^\infty x \left(\frac{\bar{F}(t)}{\bar{F}(t)} \right)^\beta (\lambda_F(t))^\beta \left(\frac{g(x)}{\bar{G}(t)} \right)^{\alpha-\beta} dx \quad (49)$$

After simplifying, we got the desired result. ■

6. CONCLUSION

In this communication, we proposed a new two parametric weighted generalized divergence measure of order β and type α . The characterization result is justify by the numerical example that it uniquely determine the distribution and also we studied its residual function. Finally, we obtained the bound for the new proposed divergence measure and also studied its properties.

REFERENCES

- [1] Belis, M. and Guiasu, S. (1968). A quantitative-qualitative measure of information in cybernetic systems (Corresp.). *IEEE Transactions on Information Theory*, 14(4):593–594.
- [2] Di Crescenzo, A. and Longobardi, M. (2007). On weighted residual and past entropies. *arXiv preprint math/0703489*.
- [3] Ebrahimi, N. and Kirmani, S. N. A. (1996). A characterisation of the proportional hazards model through a measure of discrimination between two residual life distributions. *Biometrika*, 83(1):233–235.
- [4] Kullback, S. and Leibler, R. A. (1951). information and sufficiency. *Annals of mathematics statistics*, 22(1):79–86.

- [5] Kerridge, D. F. (1961). Inaccuracy and inference. *Journal of the Royal Statistical Society. Series B (Methodological)*,184–194.
- [6] Mirali, M.A.L.I.H.E H. and Fakoor, V. (2017). On weighted cumulative residual entropy. *Communications in Statistics-Theory and Methods*, 46(6):2857–2869.
- [7] Moharana, R. and Kayal, S. (2019). On Weighted Kullback–Leibler Divergence for Doubly Truncated Random Variables. *REVSTAT-Statistical Journal*, 17(3):297–320.
- [8] Gupta, R. D. and Nanda, A. (2002). α and β entropies and relative entropies of distribution. *Journal of statistics theory of application*, 1(30):177–190.
- [9] Renyi, A. (1961). On measures of entropy and information. Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. *In proceeding of the fourth Berkeley symposium on mathematical statistics and probability*, 1:547–561.
- [10] Shannon, C. E. (1948). A mathematical theory of communications. *Bell System Technical Journal*, 27:379–423.
- [11] Suhov, Y. and Salimeh, S. Y. (2015). Entropy-power inequality for weighted entropy. *arXiv preprint arXiv:1502.02188*.
- [12] Yasaei Sekeh, S. and Mohtashami Borzadaran, G. R. (2013). On Kullback-Leibler Dynamic Information. *Available at SSRN 2344078*.