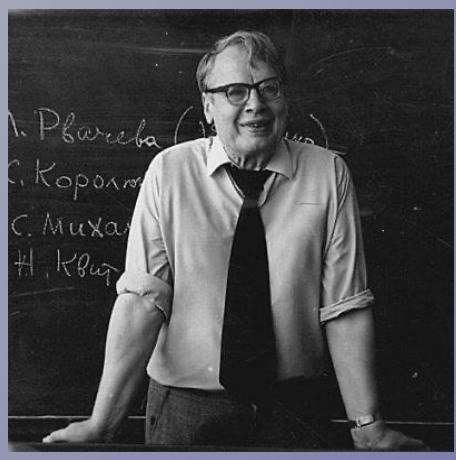
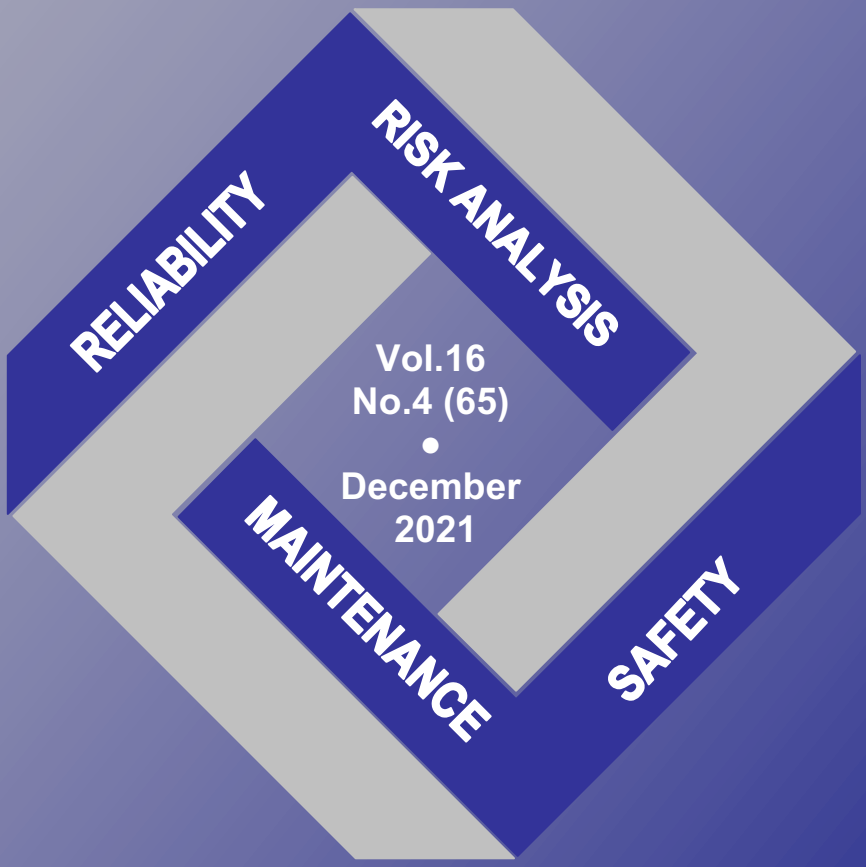
ELECTRONIC JOURNAL
OF INTERNATIONAL
GROUP ON RELIABILITY

Gnedenko Forum Publications



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

RELIABILITY:
THEORY&APPLICATIONS



ISSN 1932-2321

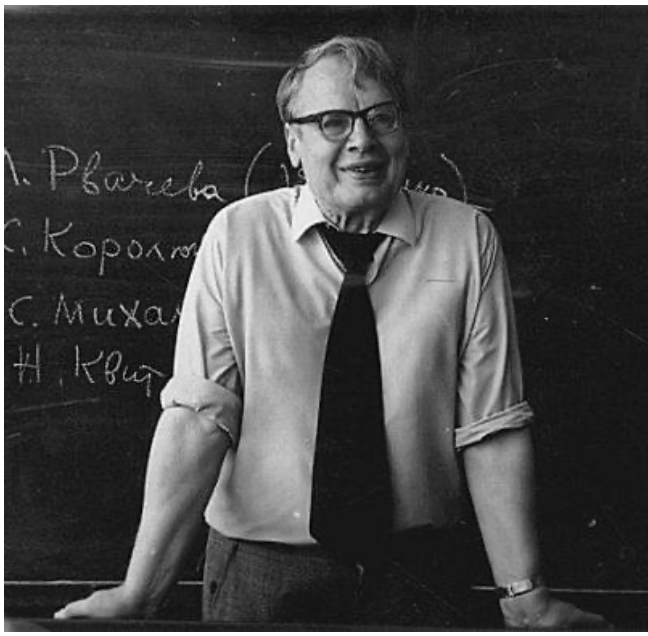
VOL.16 NO.4 (65)
DECEMBER, 2021

San Diego

ISSN 1932-2321

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





RELIABILITY:

THEORY & APPLICATIONS

Vol.16 No.4 (65),
December 2021

San Diego
2021

Editorial Board

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

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

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

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

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

The decision to accept or reject a manuscript is made by the Editor considering the referees' opinion and taking into account 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

In this paper, we introduce a new four parameters continuous probability distribution called transmuted sine-Dagum distribution obtained through the transmuted Sine-G family introduced by Sakthivel et al. [13]. We have obtained some distributional properties including moments, inverted moments, incomplete moments, central moments, and order statistics for proposed model. The reliability measures such as reliability function, hazard rate function, reversed hazard rate function, cumulative distribution function, mean waiting time and mean residual lifetime are studied in this paper. Further, we have discussed some income inequality measures including Lorenz curve, Bonferroni index and Zenga index. The maximum likelihood method is used to estimate the parameters of the proposed probability distribution. Finally, we demonstrated goodness of fit the proposed model with other suitable models in the literature using real life data sets.

This paper deals with an economic production quantity (EPQ) model in which production is random and having heterogeneous units of production. The production process is characterized by mixture of Weibull distribution. It is assumed that the demand is constant and the lifetime of the commodity is random and follows an exponential distribution. Assuming that the shortages are allowed and fully backlogged the instantaneous state of inventory in the production unit is derived. The minimizing the expected total production cost, the optimal production quantity, the production uptime and downtime are derived. Through sensitivity analysis it is observed that the random production with mixture distribution have a significant influence on the optimal production schedules and production quantity. It is also observed that the rate of deterioration can tremendously influence the optimal operating policies of the system. This model also includes some of the earlier models as particular cases. The model is extended to the case of without shortages. A comparison of the two models reveals that allowing shortages will reduce expected total cost of the model.

The regular increase in relative number of units of equipment, devices and installations (further - objects) electric power systems, which service life exceeds normative value and the consequences connected with this fact, including, including unacceptable ones, demand acceptance of drastic measures on increase of efficiency of their work. The main efforts today aimed at improving the methods of recognition and control of their technical condition. In other words, the problems of increasing the reliability of work and the safety of service brought to the fore quite justifiably. In the article, it is propos to carry out monitoring of the technical condition of overhead lines with a rated voltage of 110 kV and above monthly on the basis operational reliability parameters. New methods and algorithms for their estimation, comparison and ranking presented. As the operational reliability parameters are multidimensional, the existing methods for comparing and ranking one-dimensional statistical estimates for them are unacceptable, as the neglect preconditions of these methods conducts to essential growth of risk of the erroneous decision. The proposed new methods based on the fiducial approach, imitating modeling and the theory of statistical hypothesis testing. The cumbersomeness and laboriousness of manual calculation of operational reliability parameters, the science intensity of calculation methods is compensated by the transition to automated systems that provide information and methodological support with information about the technical condition of overhead lines. The recommended methods are included in the group of risk-focused approaches of increase the efficiency of the electric power systems.

Nowadays, large volumes of data generate by numerous business organizations due to digital communications, web applications, social media, internet of things, cloud and mobile computing. Such has turned the nature of classical data into big data. Loan risk analysis is one of the most importance financial tasks, where financial organizations predict loan risk through customer financial history and behavioral data. Financial institutions face loan risk related issues when they make a loan to a bad customer. As a result, financial institutions divide loan applications into loan risk and non-risk clusters before making a loan for avoiding the loan risk challenges. Clustering approach is a data mining technique that uses data behavior and nature to discover the unexpected loan without any external information. Clustering algorithms face efficiency and effectiveness challenges as a result of the primary characteristics of big data. Sampling is of the data reduction technique that reduces computation time and improves cluster quality, scalability and speed of clustering algorithm. This study suggests a Stratified Remainder linear Systematic Sampling Extension (SRSE) approach for loan risk analysis in big data clustering using a single machine execution. The SRSE sampling plan enhances the effectiveness and efficiency of the clustering algorithm by employing maximum variance stratum formulation, remainder linear systematic sampling and extending sampling results into final result through centroid distance metric. The performance of the SRSE-based clustering algorithm has been compared to existing K-means and K-means++ algorithms using Davies Bouldin score, Silhouette coefficient, SD Validity, Ray-Turi index and CPU time validation metric on risk datasets.

Wind energy is one of the abundant and renewable energy sources that can be harvested using wind turbines. Many factors affect the energy-yielding ability of wind turbines. The goal of this paper is to investigate the effects of uncertain weather conditions on the power generation ability of wind turbines. Uniquely, it presented the influence of the uncertain weather conditions and uncertain aerodynamic parameters of wind turbine on wind energy harvesting. The mathematical model of these factors and statistical analysis of their effects on the performance of wind turbines are presented using real-time data. It is found that the impact of uncertain weather conditions on annual average air density, and hence on the performance of wind turbines, is 1.33%. Whereas, the impact of variations of yearly average wind speed on the performance of wind turbines is found to be substantial. In particular, the annual uncertainty output power of wind turbines is found to be 32%. This investigation helps to find the mitigation mechanism and improve power generation efficiency from wind.

Hazard rate, and ageing intensity (AI) are measures or functions required for qualitative and quantitative analysis of ageing phenomena of a system with a well-defined statistical distribution respectively. In this paper, we reiterate upon the fact that in a few cases hazard rate and ageing intensity do not depict the same pattern as far as monotonicity is concerned. So, a question naturally arises which among hazard rate, and ageing intensity is a preferable measure for characterizing ageing phenomena of a system. As a consequence, an example involving two design systems are analyzed and is illustrated to answer the aforementioned question.

Dr. Surabhi Sengar, Dr. Mange Ram, Yigit Kasancoglu

This paper investigates the stochastic behavior of a redundant system having problem of waiting line in the maintenance section in terms of various aspects such as reliability, availability, sensitivity etc. The system under consideration has three parts X, Y and Z connected in series. Each part has two units. Out of which part X has one main unit and other applied redundant unit. Similarly part Y has one main unit and other cold redundant unit to support the system. When main units of both the part have failed, then redundant units start automatically. While part Z has two units connected in parallel configuration. Here a realistic situation is discussed that when main units and redundant units of part X and Y are failed and arrived for repair then due to unavailability of repair men a line is generated there and its affect on systems reliability. So the focus of the study is to investigate the nature of the system using supplementary variable technique with the application of copula methodology under the condition when all four units are in line for repair.

Certain modern developments in stochastic extreme value theory, on occasion of 110th birthday of Boris Vladimirovich Gnedenko

V. I. Piterbarg[1)], I. V. Rodionov[2)*]

•

[1)]Lomonosov Moscow state university, Moscow, Russia;
Scientific Research Institute of System Development of the Russian Academy of sciences;
International Laboratory of Stochastic Analysis and its Applications,
National Research University Higher School of Economics, Russia.
piter@mech.math.msu.su

[2)]Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia;
V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia
vecsell@gmail.com

Abstract

We present a short overview of developments of the last decade in asymptotic analysis of extrema of families of random variables. We focus on the methods of investigating the quality of approximations as given by Gnedenko's extreme value theorem, and its generalizations to the case of dependent random variables.

Keywords: Gnedenko, stochastic extreme value theory, modern developments, extreme value theorem.

I. Introduction

Let us consider a sequence $X_1, ..., X_n, ...$ of independent identically distributed random variables with the cumulative distribution function $F(x)$. Let us furthermore assume that there exists sequences of real numbers $a_n > 0$ and b_n such that the limit of the distribution functions of the sequence

$$\frac{\max(X_1, ..., X_n) - b_n}{a_n}$$

as $n \to \infty$ is non-degenerate, so that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \tag{1}$$

where the distribution function $G(x)$ takes more than two values. The maximum of random data is one of the key statistics in various applications, and so various possible forms of the function $G(x)$ were established early on in the 20s of the previous century, see [1]. But it was only in 1941 when Gnedenko, in a short note [2], published a rigorous mathematical statement describing all possible types of the distribution function $G(x)$, where the type of a distribution function $G(x)$ is understood to be a class of distributions obtained from $G(x)$ by shifting and scaling its argument. Let us state this result in modern notation.

*The work of I. V. Rodionov in section 3 was performed at the Steklov Mathematical Institute of Russian Academy of Sciences with the support of the Russian Science Foundation (grant 19-11-00290)

Theorem 1. *If* (1) *holds for some non-degenerate* G*, then there exist* $a > 0$ *and* b *such that* $G(ax+b) = G_\gamma(x)$*, where*

$$G_\gamma(x) = \exp\left(-(1+\gamma x)^{-1/\gamma}\right), \quad 1+\gamma x > 0, \tag{2}$$

the class of extreme value distributions with γ *real, and for* $\gamma = 0$ *the exponent on the right is interpreted as* $\exp(-e^{-x})$.

When $\gamma > 0$ this is the Frechet class, when $\gamma < 0$ this is the Weibull class, and for $\gamma = 0$ this is the Gumbel class, or, in this case, the standard Gumbel distribution.

The full proof of the theorem was published in 1943, [3] not in the Soviet Union for obvious reasons, but in Annals of Mathematics, in French. The English translation appeared in 1992 in a book Breakthroughs in Statistics published by Springer.

The importance of this work goes far beyond its use in the domain of applied probability theory and statistics. In our view, this is one of the cornerstones of the modern mathematical apparatus of the theory of probability. Every year since the result was discovered 80 years ago, a large number of papers that further develop mathematical methods in this area come out. It would not be an exaggeration to draw strong parallels with the central limit theorem which also came out of the needs of applications but since influenced the development of core mathematical methodologies of the whole of probability theory.

This short overview of the mathematical methods for asymptotic analysis of extrema of families of random variables is dedicated to the latest developments in this area, primarily covering the last decade, since the 100th birthday of Gnedenko that was widely celebrated by the mathematical community. We focus our attention on the areas that can be called a classical extension of the theory. Specifically we look into the quality of approximations given by Gnedenko's theorem, Theorem 1, and generalizations of the limit relation (2) to the case of dependent X_i that form a stochastic sequence or a random field on an integer lattice. There also exist various other generalizations of the original problem statement for limit distributions of maxima. This area of research mostly focuses on distributions of maxima of random processes and random fields in continuous time, extrema of vector sequences, and even functional limit theorems with follow-up analysis of the so-called max-stable stochastic processes. In short, here the focus is on limit distributions of maxima of random variables over various probabilistic structures. Among the latest on these topics the following are worth mentioning: [4] on the distribution of the maximum of a random number of random vectors, and [5], [6], [7] on max-stable processes and fields, as well as vector-valued random processes. All these papers have extensive literature reviews. From a somewhat different angle, considering triangular arrays, rather than sequences, of identically distributed random variables expands not only the class of possible limit distributions of normalized maxima, [8], [9], but also a class of distributions for which the limit distribution of the normalized maxima is non-degenerate, [11], [10]. There also exist results on the limit distribution of the maxima of stochastic sequences under non-linear normalization [12]. Lebedev in [13], [14] considers the problems of limit distributions of maxima of the particle scores in branching processes; the bibliography in these papers should also be perused. The author moves away from the classical conditions of the Gnedenko limit theorem, which is a substantial development of the theory of Lamperti-type maximal branching processes.

It is worth noting that we do not cover other types of convergence, focusing exclusively on convergence in distribution. We mention in passing one of the latest papers here, [15] and literature therein, on the iterated logarithm laws for almost sure convergence of sequences of maxima. Other works of I. Matsak on this topic are also of interest.

II. On the quality of convergence

The question of quantifying the quality of approximations in limit theorems of probability theory has many aspects. Broadly, the main topics of interest include convergence of moments; rates of convergence in limit theorems; rates of convergence for large and growing values of arguments (large deviations); convergence in probability and almost surely; asymptotic expansions and accompanying laws that improve the quality of approximations. An excellent review of relatively latest advances in the areas of moment convergence, rates of convergence, large deviations in the Gnedenko limit theorem, and sequences of normalized maxima can be found in Chapter 5 of a fairly current monograph [16]. Some of the more contemporary works covered in that review are also cited in our bibliography. The area of asymptotic expansions and accompanying measures (laws) is relatively mature, with only a few new developments appearing recently, mostly related to specific distributions important in certain applications, such as the Weibull distribution or the Normal distribution, see e.g. [17], [18].

It is important to point out that establishing asymptotic expansions and their accompanying laws is much easier for limit distributions of maxima of random variables than in the context of the central limit theorem [19], [20]. For maxima of *independent* random variables, deriving asymptotic expansions can basically just follow the approach developed by Gnedenko himself, or its somewhat more contemporary interpretations. A cumulative distribution function $F(x)$ from the maximum domain of attraction of the Gumbel distribution $MDA(\Lambda)$ can be characterized in terms of the von Mises function. As shown in [21], a distribution from $MDA(\Lambda)$ can be described via the von Mises representation. Specifically, under the assumption $F(x) < 1$ for all x, $F \in MDA(\Lambda)$ if and only if there exists $x_0 \geq 0$ such that $F(x)$ can be represented in the form

$$1 - F(x) = c(x) \exp\left\{ - \int_{x_0}^{x} \frac{g(t)}{f(t)} dt \right\}, \quad x \geq x_0, \tag{3}$$

where $f(x)$ is a positive absolutely continuous function on $[x_0, \infty)$, where $f'(x) \to 0$, $g(t) \to 1$ and $c(x) \to c > 0$ for $x \to \infty$. A similar statement can be made for a distribution bounded from the right. Normalizing sequences can be chosen as follows,

$$b_n = F^{\leftarrow}(1 - n^{-1}), \quad a_n = f(b_n).$$

It is obvious then that

$$\begin{aligned} F^n(a_n x + b_n) &= \left(1 - \exp\left(\log c(a_n x + b_n) + \int_{x_0}^{a_n x + b_n} \frac{g(t)}{f(t)} dt\right)\right)^n \\ &= \left(1 - \frac{1}{n} e^{-\gamma_n(x)}\right)^n, \end{aligned}$$

where

$$\gamma_n(x) := -\log n + \int_{b_n}^{a_n x + b_n} \frac{g(t)}{f(t)} dt - \log \frac{c(a_n x + b_n)}{c(b_n)}. \tag{4}$$

Let us denote

$$B_n(x) := e^{-e^{-\gamma_n(x)}} \mathbf{I}_{\{\gamma_n(x) \geq -\log\log n\}}, \tag{5}$$

where $\mathbf{I}$ is the indicator function. Paper [22] uses standard calculus techniques, under the assumption of $F(x) < 1$ for all x, to demonstrate that

$$P(M_n \leq a_n x + b_n) - B_n(x) = O\left(n^{-1} \log^2 n\right) \tag{6}$$

for $n \to \infty$, uniformly in $x \in \mathbb{R}$. This implies that the equality

$$P(M_n \leq a_n x + b_n) - \exp\left(-e^{-x}\right) = \exp\left(-e^{-x}\right) e^{-x} (\gamma_n(x) - x)(1 + o(1)) + O(n^{-1} \log^2 n) \tag{7}$$

holds uniformly on the set $\{x : \gamma_n(x) \geq -\log\log n\}$ as $n \to \infty$. Naturally, the idea of using the Taylor expansion applied to a power of the distribution function appears in various other works on distributions of maxima such as [23] and other references we cite.

Thus, the sequence $B_n(x)$ is the natural sequence of accompanying laws, i.e. signed measures, in Gnedenko's limit theorem. It gives an exponential-type rate of convergence to the distribution of the maximum. The same characterisation holds for the two other maximum domains of attraction, Frechét and Weibull. For them, an analogue to the representation (3) is obtained using Karamata representation for regularly varying functions, see for example [16], [24]. Expansions and accompanying laws can be derived along the same lines as our calculations above. We remind the reader that the Frechét maximum domain of attraction consists only of distributions with tails that are regularly varying at infinity, and the Weibull maximum domain of attraction with regularly varying tails at a finite right endpoint.

In [22] the authors consider the Gumbel maximum domain of attraction, where the double exponential gives a logarithmic rate of convergence only, which is often insufficient in applications. Another reason for considering this domain specifically is the fact that it is extremely broad, and various applications require splitting it into reasonable, in some sense, sub-domains. For example, this domain includes distributions whose tails are equivalent, for $x \to \infty$, to the tail of the Weibull distribution $\log(1 - F(x)) \sim -Cx^p$, $C, p > 0$, as well as log-Weibull tail, $\log(1 - F(x)) \sim -C(\log x)^p$, $C > 0, p > 1$. Moreover, the exponents in the asymptotics can be replaced by slowly varying at infinity functions. A wide variety of other distributions with heavier (slower decaying) or lighter (faster decaying) tails belongs to the same domain. The Weibull and log-Weibull classes of distributions are considered in detail in [22] as specific examples.

One of the principal recent approaches to the study of rates of convergence in the limit theorem for the maxima has been an introduction of additional conditions on the distribution tail behavior. Primarily this is the second-order condition suggested by de Haan [26]. Let us state this condition in terms of the function $\gamma_n(x)$.

The second-order condition for functions from $MDA(\Lambda)$ with an infinite right tail. *There exists a sequence $A(n)$ of constant sign, approaching zero as $n \to \infty$ and such that the limit*

$$\lim_{n\to\infty} \frac{e^{-\gamma_n(x)} - e^{-x}}{A(n)} = H(x) \tag{8}$$

exists and is not identically zero or infinite.

This formulation is based on Theorem 2.3.8, [16]. It follows from the second-order condition (see e.g. [16]) that $A(n)$ is a slowly varying at infinity function of non-positive index $\rho \le 0$. It is also known, see [27], that for the case of convergence to the Gumbel distribution we are considering here, the function H is equal to

$$H(x) = \frac{1}{\rho}\left(\frac{x^\rho - 1}{\rho} - \log x\right), \text{ if } \rho < 0, \tag{9}$$

and

$$H(x) = \frac{1}{2}\log^2 x \text{ if } \rho = 0.$$

Using the aforementioned Theorem 2.3.8, [16], and (4), one can obtain a somewhat different asymptotic expansion,

$$P(M_n \le a_n x + b_n) = \exp\left\{-e^{-x} - A(n)H(x)(1 + o(1)\right\}$$
$$\times \exp\left(-\frac{1}{n}\sum_{k=0}^{\infty} \frac{1}{(k+2)n^k}\left(\frac{1 - F(a_n x + b_n)}{1 - F(b_n)}\right)^{k+2}\right).$$

We note that if $\rho < -1$, the main contribution to the speed of convergence to the double exponential distribution is given by the second exponent. In the case $\rho = -1$, on the other hand, one needs to know the behavior of the function $A(n) = n^{-1}\ell(n)$ more precisely, i.e. how the slowly varying function $\ell(n)$ behaves. In the case $\rho > -1$, the second term in the first exponent is the main contributing factor to the rate of convergence.

Similar calculations can be carried out for the n-th order condition on the distribution tail introduced in [28]. Let us state a recent estimate by Drees and de Haan (see [29]) for the rate of convergence taking into account the accompanying law.

If the condition (8) is satisfied with $\rho < 0$, see (9), then for $b_n = F^{\leftarrow}\left(e^{-1/n}\right)$ and, correspondingly, $a_n = f(b_n)$, and for any $\varepsilon > 0$, the following holds,

$$\sup_x e^{(1-\varepsilon)x}\left|\frac{F^n(a_n x + b_n) - \exp(-e^{-x})}{A(n)} + \frac{1}{\rho}e^{-x+\rho x}e^{-e^{-x}}\right| \to 0$$

for $n \to \infty$.

Note that this can also be derived from the expansion (7).

It is also interesting to use the expansion (7) to study probabilities of large deviations in the Gnedenko limit theorem. For example, Corollary 2.1, [29] and Theorem 5.3.12, [16], under suitable restrictions, follow from the relation (7). [24] uses similar expansions for this purpose.

Scale in $MDA(\Lambda)$. As we already mentioned, the Gumbel maximum domain of attraction is extremely broad, and the idea of splitting it into parts and developing criteria for classifying distributions into these sub-domains is quite reasonable. [22] proposes one such classification of distributions with smooth tails, based on the von Mises representation. The first two "grades" in this scale are the generalized distributions of Weibull and log-Weibull type, defined by functions $f(t) = Ct^{1-p}$, $C, p > 0$, and $f(t) = Ct\log^{1-p} t$, $C > 0$, $p > 1$ in the representation (3), respectively. These distributions play an important role in financial and actuarial mathematics, in reliability theory, and other industrial applications. Yet the information on distribution tails obtained from the approximation provided by the Gnedenko theorem is far from complete. For example, insurance premiums directly depend on the specific type of the tail of the distribution that generates a given insurance event. Recently a number of studies appeared that aim to distinguish tails of Weibull and log-Weibull type distributions, see for example [30], [31], [32] and their bibliographies.

The continuation of the scale that begins with the two aforementioned classes of distributions can proceed as follows. Distribution tails with $f(t) = Ct(\log\log t)^{1-p}$, $p > 1$, are heavier than Weibull and log-Weibull type ones. (Here C denotes some constant that could be different in different contexts.) The number k of iterated logarithms in these expressions for $f(t)$ could be defined to be Gumbel's index for the distribution. Then tails of distributions of Weibull type have Gumbel's index $k = 0$, log-Weibull type distributions have index $k = 1$, and so on. More details can be found in [22].

Shubochkin in his thesis [33] determines convergence rates for approximations of distributions of normalized maxima and their accompanying laws, see (7).

The definition of the scale the we presented above is not the only reasonable option, and alternatives have been proposed. For example, Troshin in his thesis [34] considers an alternative definition of Gumbel's index, defined to be the smallest $k = 0, 1, ...$ such that the integral

$$\int_{x_0}^{\infty} \frac{a(t)dt}{t^2 \log t \log_{(2)} t ... \log_{(k)} t}.$$

converges. (Indices mean numbers of log ... log repeating.) The existence of such k for functions from $MDA(\Lambda)$ has been proved.

III. Models with dependence

One of the first follow-up questions that Gnedenko's limit theorem elicits is whether its results could be generalized to sequences of dependent and/or non-identically distributed random variables. First results of this type, mostly concerning distributions of maxima of stationary sequences, appeared back in the 60s and 70s in the works by Berman, Loynes, Cramér, and Leadbetter. The results obtained during this period are comprehensively covered in the monograph [35] by Leadbetter, Lindgren and Rootzén.

A situation when a maximum over some collection of random variables behaves like a maximum of independent random variables has a special name in the extreme value theory, and is called *extremal independence*. For example, if $\{X_i\}_{i\geq 1}$ is a sequence of random variables,

with F_i the distribution function of X_i, and $M_n = \max\{X_1, \dots, X_n\}$, then this sequence possesses extremal independence if

$$\sup_{x\in\mathbb{R}} \left| P(M_n \le x) - \prod_{i=1}^{n} F_i(x) \right| \to 0, \quad n \to \infty.$$

We note that if $\{X_i\}_{i\ge1}$ are identically distributed according to the (common) distribution function F, and this sequence is extremely independent, then the distribution of the normalized maximum converges to one of the three types of limit distributions from Gnedenko's limit theorem, as long as F satisfies the conditions of the theorem.

Stationary stochastic sequences provide an important example. Let us recall classical (in extreme value theory) sufficient conditions for extremal independence of a stationary sequence, namely the conditions D and D' from [35]. Let $\{X_i\}_{i\ge1}$ be a (strictly) stationary sequence with a marginal distribution function F. We say that it satisfies the condition $D(u_n)$ for a sequence u_n if for any integers $1 \le i_1 < \ldots < i_p$ and $j_1 < \ldots < j_q \le n$, for which $j_1 - i_p \ge l$, the following holds,

$$\left| P(\max(X_{i_1}, ..., X_{i_p}, X_{j_1}, ..., X_{j_q}) \le u_n) - P(\max_{k\in[p]} X_{i_k} \le u_n) P(\max_{k\in[q]} X_{j_k} \le u_n) \right| \le \alpha_{l,n}, \qquad (10)$$

where $\alpha_{l_n,n} \to 0$ as $n \to \infty$ for some index sequence $l_n = o(n)$. Furthermore, we say that the stationary sequence $\{X_i\}_{i\ge1}$ satisfies the condition $D'(u_n)$ for a sequence u_n if

$$\limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} P(X_1 > u_n, X_j > u_n) \to 0, \quad k \to \infty \qquad (11)$$

holds (here $[\cdot]$ is an integer part of a number). Then, if for independent copies of random variables$\{X_i\}_{i\ge1}$ the conditions of Gnedenko's limit theorem are satisfied for some sequences a_n and b_n, and the conditions $D(u_n(x))$ and $D'(u_n(x))$ are satisfied for the sequence $u_n = a_n x + b_n$ for any x, then the sequence $\{X_i\}_{i\ge1}$ is extremely independent.

The conditions D and D' play a foundational role in the extreme value theory for stochastic models with dependence. The conditions have been slightly modified in [36] and [37] to extend the result above to non-stationary sequences. Papers [38] and [39] extended it even further to stationary random fields on integer lattices, and [40] proved an equivalent result for non-stationary random fields in dimension 2. The main technique that was used in all these proofs was the so-called block method, where the domain of the stochastic process is split into non-overlapping intervals in such a way that maxima over the intervals are asymptotically independent. Applications of this method to random fields required very complicated versions of the conditions D and D' which hindered further progress along similar lines of attack. This issue was finally overcome in [41] (see full text in [42]). These papers derived the conditions for extremal independence of random variables that constitute so-called stochastic systems. A stochastic system here is defined as a sequence $(X_1(n), \dots, X_d(n)) \in \mathbb{R}^d$ of random vectors of varying dimensions, where $d = d(n)$ is some sequence of positive integers. We should also mention here [43] whose results can be used to derive asymptotics for the distribution of the maxima of a stochastic system under certain conditions. Stochastic systems generalize many models such as stochastic sequences, stochastic fields and triangular arrays. They are rich enough to even represent complicated objects such as random networks and graphs that are otherwise quite challenging to analyze.

Gaussian stochastic sequences, a special case of stochastic sequences, exhibit extremal independence under rather weak assumptions. Let $\{X_i\}_{i\ge1}$ be a Gaussian random sequence with mean 0 and covariance function $r(i)$, where $r(0) = 1$. Berman [44] found a simple condition for the convergence of the distribution of the maximum of the sequence $\{X_i\}_{i\ge1}$, with the same normalization as for the maximum of independent Gaussian variables in Gnedenko's limit theorem, to the Gumbel distribution. The Berman condition simply requires that

$$r(n) \ln n \to 0, \quad n \to \infty.$$

It has been established that the Berman's condition implies the conditions D and D' for stationary Gaussian sequences. Hüsler in [36] and [37] showed that under a certain generalization of the Berman's condition, a non-stationary Gaussian sequence is extremely independent. Pereira [45] obtained this result for Gaussian non-stationary random fields in $\mathbb{R}^2$, while Jakubowski and Soja-Kukieła in [46] extended it to Gaussian stationary fields of arbitrary dimension.

It is interesting that the Berman's condition is close to being necessary (as well as sufficient) for extremal independence of stationary Gaussian sequences. Specifically, Mittal and Ylvisaker in [47] showed that if $r(n)\ln n \to \gamma > 0$ as $n \to \infty$, then the limit distribution of the normalized maximum of a Gaussian stationary sequence is completely different, and is a convolution of the Gumbel distribution and the Gaussian one. In this case the sequence does not even possess a phantom distribution function (to be defined shortly).

The extremal independence property is far from being always satisfied, and processes that appear in applications often exhibit a high degree of dependence. It turns out that in many cases, the behavior of the maximum of a stationary sequence can be described in terms of the so-called extremal index. According to the definition from [35], a stationary sequence $\{X_i\}_{i\geq 1}$ has the extremal index $\theta \in [0,1]$, if for any $\tau > 0$ there exists a sequence $u_n(\tau)$ such that

$$n(1 - F(u_n(\tau))) \to \tau \quad \text{and } P(M_n \leq u_n(\tau)) \to e^{-\theta\tau}.$$

It follows, in particular, that

$$\left|P(M_n \leq u_n(\tau)) - F^{\theta n}(u_n(\tau))\right| \to 0, \quad n \to \infty,$$

so that the maximum of n terms of the stationary sequence behaves like the maximum of θn independent copies of X_1. It should now be obvious that the situation we considered just before corresponds to the case $\theta = 1$. The notion of the extremal index is of paramount importance in applications of extreme value theory, because one can often reduce a sequence of real-world observations to a stationary sequence, or simply consider it to be such.

Sufficient conditions for the existence of the extremal index were found by Chernick [48], and they look like this. Let us assume that for $\tau > 0$ a sequence $u_n(\tau)$ is defined such that $n(1 - F(u_n(\tau))) \to \tau$ as $n \to \infty$, and the condition $D(u_n(\tau))$ is satisfied for any such $\tau > 0$. Then, if for some τ the sequence $P(M_n \leq u_n(\tau))$ converges, then the extremal index exists for $\{X_i\}_{i\geq 1}$. It is interesting to note that the criterion for the existence of the extremal index has only been found relatively recently, [49], see also Proposition 11.4, [23]. Various other properties of the extremal index, as well as methods for its estimation, are covered in Chapter 10 of [50].

The extremal index provides a remarkably convenient mechanism for describing extremal dependence in stationary sequences. A single index, however, is not sufficient for describing extremal dependence of stationary random fields on integer lattices. Various attempts to extend the idea of the extremal index to random fields and use it to analyze extremal dependence have been undertaken in, for example, [51] and [52], with more complex models considered in [53]. However, [54] showed that the extremal index of a stationary random field on $\mathbb{Z}^d$ can materially depend on the direction of growth of a multi-index $\mathbf{n} = (n_1, \ldots, n_d)$.

A natural generalization of the notion of the extremal index is provided by the notion of a phantom distribution function, as discussed in [55]. We say that the distribution function G is the phantom distribution function for the stationary sequence$\{X_i\}_{i\geq 1}$ with the marginal distribution function F if

$$\sup_{x\in\mathbb{R}} |P(M_n \leq x) - G^n(x)| \to 0, \quad n \to \infty.$$

It is not hard to see that if G could be chosen to be of the form F^θ, then the extremal index of the sequence $\{X_i\}_{i\geq 1}$ is θ. The existence of a phantom distribution function for a stationary sequence is quite a common property. For example, it is shown in [49] that any α-mixing stationary sequence with a continuous marginal distribution function has a phantom distribution function. The same paper suggests a simple condition for the existence of the phantom distribution function: it exists if and only if for some sequence ν_n and $\gamma \in (0,1)$ the convergence$P(M_n \leq \nu_n) \to \gamma$ holds

as $n \to \infty$, and for all $T > 0$ the condition $B_T(\{\nu_n\})$:

$$\sup_{p,q\in\mathbb{N}:p+q\le T\cdot n} \left|P(M_{p+q} \le \nu_n) - P(M_p \le \nu_n)P(M_q \le \nu_n)\right| \to 0, \quad n \to \infty$$

is satisfied. Clearly the condition $B_T(\{\nu_n\})$ resembles Leadbetter's condition D. Theory of phantom distribution functions for models other than stationary sequences is still in its infancy. In this regard it is worth mentioning [54] where, for the first time ever, the question of existence of phantom distribution functions for stationary random fields on integer lattices is considered.

References

[1] Fisher, R. A. and Tippett L. H. C. (1928). Limiting forms of the frequence distributions of the largest or smallest member of a sample. *Proceedings of Cambridge Philosophical Society*, 24:180–190.

[2] Gnedenko, B.V. (1941). Limit theorems for maximum term in a random series. *Doklady of Russian Academy of Sciences*, 32(1):7–9. In Russian.

[3] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44:423–453.

[4] Hashorva, E., Padoan, S. A. and Rizzelli S. (2021). Multivariate extremes over a random number of observations. *Scandinavian Journal of Statistics*, 48(3):845–880.

[5] Debicki, K. and Hashorva, E. (2017). On extremal index of max-stable processes. *Probability and Mathematical Statistics*, 37(2):299–317.

[6] Hashorva, E. and Kume, A. (2021). Multivariate max-stable processes and homogeneous functionals. *Statistics and Probability Letters*, 173:109066.

[7] Hashorva, E. (2021). On Extremal Index of Max-Stable Random Fields (2021). *Lithuanian Mathematical Journal*, 61(2):217–238.

[8] Freitas, A. and Hüsler, J. (2003). Condition for the convergence of maxima of random triangular arrays. *Extremes* 6(4):381–394.

[9] Morozova, E. and Panov, V. (2021). Extreme Value Analysis for Mixture Models with Heavy-Tailed Impurity. *Mathematics*, 9(18):2208.

[10] Dkengne, P.S., Eckert, N. and Naveau, P. (2016). A limiting distribution for maxima of discrete stationary triangular arrays with an application to risk due to avalanches. *Extremes*, 19(1):25–40.

[11] Nadarajah, S. and Mitov, K. (2002). Asymptotics of Maxima of Discrete Random Variables. *Extremes*, 5:287–294.

[12] Mitov, K.V. and Nadarajah, S. (2021). Limit distributions for the maxima of discrete random variables under monotone normalization. *Lithuanian Mathematical Journal*, to appear.

[13] Lebedev, A.V. (2008). Maxima of random particles scores in Markov branching processes with continuous time. *Extremes*, 11(2):203–216.

[14] Lebedev, A.V. (2019). Multivariate Extremes of Random Scores of Particles in Branching Processes with Max-Linear Heredity. *Mathematical Notes*, 105:376–384.
With: Lebedev, A.V. (2020). Erratum to: Multivariate Extremes of Random Scores of Particles in Branching Processes with Max-Linear Heredity. *Mathematical Notes*, 107:1046.

[15] Matsak, I. (2019). Asymptotic Behavior of Maxima of Independent Random Variables. *Lithuanian Mathematical Journal*, 59:185–197.

[16] de Haan, L. and Ferreira, A. Extreme Value Theory. An Introduction. Springer, New York, 2006.

[17] Liu, C. and Liu, B. (2013). Convergence rate of extremes from Maxwell sample. *Journal of Inequalities and Applications*, 477.

[18] Peng, Z., Nadarajah, S. and Lin, F. (2010). Convergence Rate of Extremes for the General Error Distribution. *Journal of Applied Probability*, 47(3):668-679.

[19] Gnedenko, B. V. and Kolmogorov, A. N. Limit Theorems for Sums of Independent Random Variables. Addison-Wesley, Cambridge, Mass, 1954.

[20] Senatov, V. V. Cental Limit Theorem: Approximation accuracy and asymptotical expansions. "LIBROCOM", 2009 (in Russian)

[21] Balkema, A. A., and de Haan, L. (1972). On R. von Mises' condition for the domain of attraction of $\exp\{-e^{-x}\}$. *Annals of Mathematical Statistics*, 43:1352–1354.

[22] Piterbarg, V.I., Scherbakova, Yu. A. (2022). On accompanying measures and asymptotic expansions in limit theorems for maximum of random variables. *Theory of Probability and its Applications*, in press, see also https://arxiv.org/abs/2010.10972.

[23] Novak, S.Y. Extreme Value Methods with Applications to Finance, Chapman & Hall CRC, 2012.

[24] Resnick, S.I. Extreme values, regular variation, and point processes. Springer-Verlag, New York Berlin Heidelberg, 1987.

[25] Lin, F., Zhang, X., Peng, Z. and Jiang, Y. (2011). On the Rate of Convergence of STSD Extremes. *Communications in Statistics, Theory and Methods.* 40(10):1795–1806.

[26] de Haan, L. (1984). Slow variation and characterization of domains of attraction. In *Statistical Extremes and Applications* (Tiago de Oliveira, Ed.), D. Reidel, Dordrecht, 31–48.

[27] Resnick, S. and de Haan, L. (1996). Second-order regular variation and rates of convergence in extreme-value theory. *Annals of Probability*, 24(1):97–124.

[28] Wang, X.Q. and Cheng, S. H. (2005) General Regular Variation of the n-th Order and 2nd Order Edgeworth Expansions of the Extreme Value Distribution (I, II). *Acta Mathematica Sinica, English Series*, 2005, 21(5):1121–1130; 2006, 22(1):27–40.

[29] Drees, H., de Haan, L. and Li, D. (2003). On large deviations for extremes. *Statistics and Probability Letters*, 64:51–62.

[30] Goegebeur, Y. and Guillou, A. (2010). Goodness-of-fit testing for Weibull-type behavior. *Journal of Statistical Planning and Inference*, 140(6):1417–1436.

[31] Rodionov, I.V. (2018). On discrimination between classes of distribution tails. *Problems of Information Transmission*, 54(2):124–138.

[32] Kogut, N.S. and Rodionov, I.V. (2021). On tests for distinguishing distribution tails. *Theory of Probability and its Applications*, 66(3):348–363.

[33] Shubochkin, E. I. (2021). Asymptotical behavior in Gnedenko limit theorem for maximum of random variables from Gumbel maximum domain of attraction. Diploma thesis, Lomonosov Moscow State University (in Russian).

[34] Troshin, V. V. (2018). On rate of convergence in B. V. Gnedenko limit theorem. Diploma thesis, Lomonosov Moscow State University (in Russian).

[35] Leadbetter, M.R., Lindgren, G. and Rootzén, H. Extremes and Related Properties of Random Sequences and Processes. Springer, New York, 1983.

[36] Hüsler, J. (1983). Asymptotic approximation of crossing probabilities of random sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 63(2):257–270.

[37] Hüsler, J. (1986). Extreme Values of Non-Stationary Random Sequences. *Journal of Applied Probability*, 23(4):937–950.

[38] Leadbetter, M. and Rootzén, H. (1998). On Extreme Values in Stationary Random Fields. In: *Stochastic Processes and Related Topics.* Trends in Mathematics. Birkhäuser, Boston, 275–285.

[39] Ling, C. (2019). Extremes of stationary random fields on a lattice. *Extremes*, 22:391–411.

[40] Pereira, L. and Ferreira, H. (2006). Limiting crossing probabilities of random fields. *Journal of Applied Probability*, 3:884–891.

[41] Isaev, M., Rodionov, I.V., Zhang, R. and Zhukovskii, M.E. (2020). Extreme value theory for triangular arrays of dependent random variables. *Russian Mathematical Surveys*, 75(5):968–970.

[42] Isaev, M., Rodionov, I., Zhang, R. and Zhukovskii, M. Extremal independence in discrete random systems. arXiv:2105.04917.

[43] Arratia, R., Goldstein, L. and Gordon, L. (1989). Two Moments Suffice for Poisson Approximations: The Chen-Stein Method. *Annals of Probability*, 17(1):9–25.

[44] Berman, S.M. (1964). Limit theorems for the maximum term in stationary sequences. *Annals of Mathematical Statistics*, 35:502–516.

[45] Pereira, L. (2010). On the extremal behavior of a nonstationary normal random field. *Journal of Statistical Planning and Inference*, 140:3567–3576.

[46] Jakubowski, A. and Soja-Kukiela, N. (2019). Managing local dependencies in asymptotic theory for maxima of stationary random fields. *Extremes*, 22:293–315.

[47] Mittal, Y., and Ylvisaker, D. (1975). Limit distributions for the maxima of stationary Gaussian processes. *Stochastic Processes and their Applications*, 3:1–18.

[48] Chernick, M.R. (1981). A limit theorem for the maximum of autoregressive process with uniform marginal distributions. *Annals of Probability*, 9:145–149.

[49] Doukhan, P., Jakubowski, A. and Lang, G. (2015). Phantom distribution functions for some stationary sequences. *Extremes*, 18:697–725.

[50] Beirlant, T., Goegebeur, Y., Segers, J. and Teugels, J. Statistics of extremes. Theory and applications. Wiley, Wiley series in probability and statistics, London, 2004.

[51] Ferreira, H. and Pereira, L. (2008). How to compute the extremal index of stationary random fields. *Statistics and Probability Letters*, 78:1301–1304.

[52] Turkman, K.F. (2006). A note on the extremal index for space-time processes. *Journal of Applied Probability*, 43:114–126.

[53] Goldaeva, A.A. and Lebedev, A.V. (2018). On extremal indices greater than one for a scheme of series. *Lithuanian Mathematical Journal*, 58(4):384–398.

[54] Jakubowski, A., Rodionov, I. and Soja-Kukiela, N. (2021). Directional phantom distribution functions for stationary random fields. *Bernoulli*, 27(2):1028–1056.

[55] O'Brien, G. (1987). Extreme values for stationary and Markov sequences. *Annals of Probability*, 15:281–291.

RT&A, No 4 (65)
Volume 16, December 2021

MIKHAIL ANDREEVICH FEDOTKIN: A NONSTATISTICAL ANALYSIS OF THE FIRST 80 YEARS OF HIS LIFE

Andrei V. Zorine

•

Lobachevsky University
andrei.zorine@itmm.unn.ru

Abstract

Professor of the Lobachevsky University of Nizhni Novgorod celebrated his 80th anniversary in May, 2021. This paper touches some of his personal histories, and his scintific contributions.

Keywords: biography, M.A. Fedotkin, applied probabillity school in Nizhni Novgorod

This paper is dedicated to the life and scientific achievements of Mikhail Andreevich Fedotkin, on the occasion of his 80th anniversary. It is not an easy task to present the topic better than the protagonist did himself in his autobiographical book [1]. No one has a more complete knowledge of his life facts than he does. So, we may only revew several turning points in his biography, maybe in a half-joking tone sometimes, with all our love and respect to the man of the hour.

1. His life

May, 1th, 1941 was Thursday. Soviet Union was celebrating the May Day. Official governemental 'Pravda' newspaper reported about achievements of Soviet oil workers in socialist emulation. Workers of the V.V. Kuybyshev Locomotive Factory at Kolomna fulfilled the four-month norm at 109,6 percents. Vasily Smyslov recieved a USSR grandmaster rank from All-nion Committee for physical culture and sports. There are also summaries of the German commans on the sinking of convoys, reports by Reuters about evacuation of a part of Plymouth, on the actions of British aviation, reports on military operations in Africa and the Mediterranean Sea.

Figure 1: *Kulikovo Field surroundings, former Tuzhilki village and Kiselevka village marked with crosses in the bottom -right corner*

On the May Day, 1941 a child was born to Andrei Artemyevich and Ksenia Prokofievna Fedotkin. He was the seventh child in the family. He was named Mikhail.

His birthplace, Kiselevka village, lies in Central Russia, 400 kilometers to the south of Moscow, in the rich-soiled 'chernozem' fields. Not far from the place we find the famous historical Kulikovo Field (where the battle of Kulikovo took place on September, 8, 1380 between the armies of Golden Horde and joint Russian prinipalities under the command of Prince Dmitry Donskoy), Yasnaya Polyana (an estate where Leo Tolstoy lived and wrote his famous masterpieces

like 'War and Piece' and 'Anna Karenina'), and Konstantinovka village (where Sergei Yesenin was born, on of the most popular and well-known poets of the 20th century).

The place and epoch definitely influenced his life. He'd got interested in the game of chess and begged his elder brother to get him a set of pieces and a chessboard, he still keeps chessbooks on his bookshelves, and he solves chess compositions easily in his 80. The blazing war in Europe killed his father in 6 months after his birth, and the burden of his large family survival fell on his 36-years old mother. That issue of 'Pravda' oracular. And the region was unique: together with his school teacher and his classmates he went to old villages in the area to search for new historical knowledge about his birthplace. One of the villages was Lyapunovka. Later, using this data, he would reconstruct the genealogy of Lyapunov family who gave not only world-famous mathematician A.M. Lyapunov, but also several other renowned statesmen, scientists, doctors, and music-writers. Fedotkin even hipothesized a missing link in the family tree [1].

What else might have influenced his life trajectory? If one looks at Mikhail's natal horoscope just for fun, he will discover that almost all planets are in the constellation of Taurus together with the Sun, and only the Mars planet is in Aquarius. Astrology books claim that Mars in Aquarius signifies a searching man, eager for new approaches to even traditional problems. Aspiration to bring together like-minded people, his confederates, to give them an interesting task, challenging problems, to administrate them and equip with the necessary amenities. Should we trust this elder form of data-science? A quick check using a computer astronomy program shows that Mars was in Capricorn rather than in Aquarius that moment of time. But what is surprising, this characteristic of a 'Mars in Aquarius' fits Mikhail Fedotkin quite nicely.

He went regulary by feet to elementary school in Tuzhilki, a village at distance of 4 kilometers from Kiselevka. The school occupied the house which used to belong to Fedotkins family more than 10 years before that and now no more. After graduation from school in 1958 he finally enrolls to the Gorky State University. That year a novel educational program 'computational mathematics', focused on cybernetics and use of computers, was opened in the university's department of physics and mathematics.

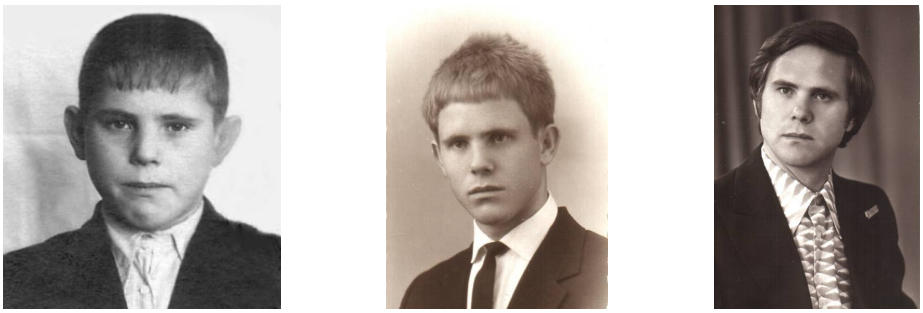


Figure 2: *At age of 7 y.r. (on the left), 17 y.o. (in the middle), 27 y.o. (on the right)*

In 1963 he was graduated from Gorky State University with diploma in mathematics and went to graduate school to specialize in theoretical cybernetics under the supervision of Yuri Isaakovich Neimark, one of the co-founders of the Research Institute for Applied Mathematics and Cybernetics (NII PMK in Russian) of the Gorky State University, and of the Department of Computational Mathematics and Cybernetics (1963). It was the first department of cybernetics in the Soviet Union. For example, the Moscow State University opened a similar department only 10 years later.

In his graduate research M.A. Fedotkin developed a mathematical theory of road-traffic control by means of traffic-lights. He defended his dissertation in 1968. His opponents at the defense were renowned scientists Igor Nikolaevich Kovalenko (1935 – 2019) and Alexander Dmitrievich Solovyev (1927 – 2001).

Beginning from 1968, on Mondays M.A.Fedotkin goes from Gorky to Moscow and back by train to be a listener at the Seminar in probability theory mathematical statistics, and stochastic processes, organized by academicians Andrei Nikolaevich Kolmogorov and Boris Vladimirovich

Gnedenko. He continues delevopment of his own mathematical methods and models of control for traffic flows. He gradually becomes one of top researchers in controlled queueing systems. All his research is done while in positions at NII PMK and at the Chair of Control Theory and Systems Dynamics of the Gorky State University. By year 1980 he writes his Doctoral dissertation. It was defended in 1984 in Moscow State University under the speciality 01.01.05 – Probability theory and mathematical statistics. The official opponents were academician V.S. Korolyuk, corresponding member of Academy S.V. Yablonsky and prof. G.P. Klimov.

As a newly-ranked Doctor of physical and mathematical sciences, he attains the opening of the Laboratory of methods of probability theory and mathematical statistics within NII PMK in 1985, and in 1986 he opens his own Chair of applied probability theory within the Department of Computational Mathematics and Cybernetics of the Gorky State University. Although Chairs of probability theory and mathematical statistics existed at several universities by that time, it was the first Chair in applied probability theory in the country. Its creation was voted for and supported by A. N. Kolmogorov, Yu. V. Prokhorov, B. V. Gnedenko, V. S. Korolyuk.

Figure 3: *M.A. Fedotkin read a lecture for MS students of Lobachevsky University*

For his creative growth, M.A. Fedotkin gives credit to the luck and happiness of meetings and conversations with many famous scientists, here are a few (in alphabetical order): T.A Azlarov, G.P. Basharin, Yu.K. Belyaev, L.N. Bolshev, A.A. Borovkov, N.P. Buslenko, B.V. Gnedenko, B.I. Grigelionis, V.M. Zolotarev, I.N. Kovalenko, A.N. Kolmogorov, V.S. Korolyuk, J.P. Kubilus, A.A. Lyapunov, N.N. Moiseev, N.N. Krasovsky, Yu.I. Neimark, B.N. Petrov, Yu.V. Prokhorov, Yu.A. Rozanov, T.A. Sarymsakov, B.A. Sevastyanov, A.G. Sigalov, S.H. Sirajdinov, A.V. Skorokhod, A.D. Solovyev, V.A. Statulevicius, I.A. Ushakov, A.N. Shiryaev, S.V. Yablonsky.

Apart of over 290 papers, he authored and co-authored text books and monographs on applied probability theory [2, 3, 4, 5, 6]. In the article [7] written by I.N. Kovalenko (2018), M.A. Fedotkin is called the leader of the Nizhni Novgorod mathematical school in the area of control of transportation flows.

2. His science

Scientific interests of M.A. Fedotkin lay in the following areas:

1) Dymanic systems for control of transportation flows.
2) Theory of controlled queueing systems with variable structure.
3) Adaptive stochastic control systems.

Figure 4: *Books written by M.A. Fedotkin*

4) Marked point processes and reliability theory.

5) Theory of functionals of sample paths of stochastic processes.

6) Cybernetic approach to construction, analysis, and optimization of probabilistic models of evolutionary experiments with control.

In his autobiographical book [1], M.A. gives extensive lists of references to his own works for each topic.

Below we'll review some of his contributions.

Application of Markov chains with incomes to transportation traffic control problems. In 1960 R.A. Howard published his book [8] (a Russian translation appeared soon in 1964) where a new branch of mathematics, dynamic programming, was applied to control problems for random processes. In his search for new approaches to transportation traffic control problems, M.A. Fedotkin proposed a mathematical model of a controlled intersection with a traffic-light in form of a multivariate stochastic process, and applied then Howard's policy iteration algorithm to generate optimal light-switching schemes. This model together with its numerical study was published in [9] and became a chapter in Candidate of sciences dissertation. Later, this research was continued by his Yu.I. Neimark and one Neimark's PhD students A.M. Preobrazhenskaya who also defended a dissertation on that topic in 1981.

Bartlett's traffic flow statistics. In the statistical theory of transportation flows M.A. Fedotkin also broke barriers and proposed a new viewpoint at the subject. He proposed a so-called 'non-local description' of a flow. In 1963 an English statistician M.S. Bartlett published a paper [10] with some sample data on inter-arrival times in vehicular flow near London and demonstrated that a classical Poisson model can't fit these data. Subsequent analysis of the data by D. Cox, P. Lewis, and others rejected several more models, e.g. a renewal process model. Fedotkin came to a conclusion that the reason for the failure was in desire to fit a traditional counting process model which observes each single arrival. He proposed to count arrivals only over large intervals of time. Since vehicles on a road move in groups with a slow vehicle in front, this idea seemed reasonable. Fedotkin splitted the Bartlett's data into 'groups of vehicles' and then he was able to fit a probability distribution for group size. In particular, the following probability distribution for the group size η worked:

$$\Pr(\{\eta = 1\}) = 1 - p, \qquad \Pr(\{\eta = k\}) = p(1-q)q^{k-2}, \qquad k = 2, 3, \ldots \quad (0 < p, q < 1). \tag{1}$$

It was called the Bartlett's distribution by M.A. Fedotkin and his disciples. Further development of this approach resulted in several heuristic flow partitioning algorithms, and in a queueing model for vehicular group formation that explained, after 20 years, why the Bartlett distribution (1) was likely [3, Ch. 10] to explain road traffic. This branch of study was continued by M.A. Fedotkin's co-workers E.V. Kuvykina, L.N. Anisimova, M.A. Rachinskaya, E.V. Kudryavtsev.

Non-local description of a flow. From a classical point of view, a flow is just a stochastic sequence $0 \leqslant \tau_1 \leqslant \tau_2 \leqslant \ldots$ of instants when new arrivals occur. Experience from having analysed

Bartlett's data led to the notion of a non-local description of a flow [11]. Basically, it can be defined as a marked point process

$$\{(\tau_i^{(\mathrm{obs})}, \eta_i^{(\mathrm{obs})}, \nu_i); i = 1, 2, \ldots\}$$

where $\tau_i^{(\mathrm{obs})}$ is an observation moment, $\eta_i^{(\mathrm{obs})}$ is the number of new arrivals during the time interval $(\tau_{i-1}^{(\mathrm{obs})}, \tau_i^{(\mathrm{obs})}]$, and $\nu_i \in M$ is a mark of all customers arriving during the interval, $i = 1, 2, \ldots$ (M denotes the set of possible marks). For example, to describe traffic flows we can include the traffic-light state into the marks. The flexibility of this new approach comes from the fact that, when the choice of the observational moments and marks is successful, we can build, analyze, and optimize quite sophisticated real-life systems not only in transportation traffic control, but also in mass production, information technologies, medicine etc.

Chung functionals. Reflecting of possible optimization problem statements in traffic control, M.A. Fedotkin came up with what he has called 'Chung functionals'. They were named after K.-L. Chung who extensively used taboo probabilities (transition probabilities with prohibited set) to study Markov chains. Let $\{X_n; n = 0, 1, \ldots\}$ be a denumerable Markov chain with the state-space S and $f(\cdot)\colon S \to \mathbb{R}$ a suitable function. Let S be partitioned into disjoint sets S_0 (admissible states), S_+ (target states), and S_- (forbidden stated). Define

$$\tau = \inf\{i\colon X_i \in S_+, S_j \notin S_-, j < i\}, \qquad \zeta = \sum_{i=0}^{\tau} f(X_i),$$
$$J(x; S_0, S_+, S_-) = \mathrm{E}(\zeta \mid \{X_0 = x, \tau < \infty\}).$$

Here the Chung functional $J(x; S_0, S_+, S_-)$ can be interpreted as the total income (or the total cost) of a Markovian random walk from the initial state $x \in S_0$ until exit from S_0 to S_+ without visits to the prohibited set S_- (think of the problem of unloading a crossroads without making even larger jams). In his paper [12] in the famous *Doklady AN SSSR*, recommended for publication by academician A.N. Kolmogorov, M.A. Fedotkin constructed an example where an infinite system of linear equations for the quantities $\{J(x; S_0, S_+, S_-)\colon x \in S_0\}$ might have several solutions, only one of them solved the original probabilistic model (before that, everybody believed such a system should have a unique solution).

This sort of optimization problems was applied to transportation traffic control by N.M. Golysheva and to priority queuing systems by A.V. Zorine.

Systems with varying structure and cybernetic approach Different particular models of qeueing situation with conflicting flows and algorithmic control, e.g. in road traffic control at intersections with complex crossing geometry, airtraffic control over takeoffs and landings of aircrafts, microwelding machines control for microchip production lines, led M.A. Fedotkin to invention of a unified framework for building adecvate queueing models, which are relatively feasible for analytical study and optimization. He called this framework a 'Theory of descrete systems with varying structure of service of quasi-regenerative flows' (this was the title of his doctoral dissertation). This framework assumed a non-local (integral) description of the source data, different operation modes of the server and the possibility of structural changes in at least one of the elements of a queueing system. Also, he advocated addition of new obligatory blocks to a typical queueing system, such as 'saturation flows', 'service algorithm' (explicitly spelled out as some mathematical entity, e.g. a graph), queuing discipline as a mathematical relation which specifies the actual amount of serviced customers as a function of the numbers in the queue, new arrivals, and saturation flows. This stage of development can be found in [13, 14].

Later he embedded this framework into even more general contept of an abstract control system — the term was introduces by pioneers of Soviet cybernetics Aleksei Andreevich Lyapunov and Sergey Vsevolodovich Yablonsky [15]. According to them, any abstract control system consists of only six functional elements: input and output poles, external and internal memories,

information processing units for each memory. The link between the two approaches became clear after conversations with A.A. Lyapunov during scientific events, and discussion of M.A.'s dissertation with S.V. Yablonsky, an official opponent. Fedotkin added a (possibly random) external environment and demonstrated that this approach can solve not only queueing problems, but also management problems in hispital administration [6, 16, 17]. Moreover, this approach allows to solve a hard problem of studying output flows from controlled queueing systems.

This area of research is the richest with respect to the number of produced models and disciples and followers. O.A.Vaganov, E.V. Kuvykina, N.V.Litvak, A.A. Vysotsky, A.N. Kudelin, A.V. Zorine, E.V. Proidakova, A.M. Fedotkin, M.A. Rachinskaya, E.V. Kudryavtsev.

References

[1] Fedotkin, M. A. Trajectories of Fate: a historical and biographical essay, Lobachevsky University Press, 2012 (in Russian: Traektorii sud'by)
[2] Fedotkin M.A. Basics of applied probability theory and statistics, Moscow, Vysshaya Schkola, 2006 (in Russian).
[3] Fedotkin M.A. Models in probability theory, Moscow, Fizmatlit, 2012 (in Russian)
[4] Zorine A.V., Fedotkin M.A. Monte-Carlo methods for parallel computing, Moscow, MSU Publishing, 2013 (in Russian)
[5] Fedotkin M.A. Lectures on analysis of random phenomena, Moscow, Fizmatlit, 2016 (in Russian)
[6] Fedotkin M.A. Non-traditional problems of mathematical modeling of experiments, Moscow, Fizmatlit, 2018 (in Russian)
[7] Kovalenko I. The FCFS-RQ system by Laslo Lacatos and its modifications. *Reliability Theory and Applications*, 49:51–56.
[8] Howard R.A. Dynamic Programming and Markov Processes. Wiley, 1960
[9] Neimark Yu.I., Fedotkin M.A., Preobrazhenskaya A.M. Operation of an automata with feedback controlling street traffic at an intersection. *Izvestiia AN SSSR. Ser. Tekhnicheskaya kibernetika*, 1968, N. 5. P 129–141.
[10] Bartlett M.S. The spectral analysis of point processes *Journal of Royal statistical society, Ser. B*. 1963. V. 25, No 2. P. 264–296.
[11] Fedotkin M.A. Incomplete description of flows of non-homogeneous calls In: *Queueing theory (Teoriia massovogo obsluzhivania)*, Moscow, MGU–VNIISI, 1981. P. 113–118. (in Russian)
[12] Fedotkin M.A. Algebraic properties of probability distributions for Chung functionals of homogeneous Markov chains with countable state-set *Doklady AN SSSR*, 1976, V. 7, P. 43–46. (in Russian)
[13] Fedotkin M.A. Optimal control of conflicting flows and marked point processes with selected discrete component. I. *Lietuvos matematikos rinknys*, T. 28, N. 4, 1988. P 784–794 (in Russian)
[14] Fedotkin M.A. Optimal control of conflicting flows and marked point processes with selected discrete component. II. *Lietuvos matematikos rinknys*, T. 29, N. 1, 1989. P 51–70 (in Russian)
[15] Lyapunov A.A., Yablonsky S.V. Theoretical problems of cybernetics In: *Problems of cybernetics: a collection of articles*, Moscow, Fizmatgiz, 1963, P. 5–22.
[16] Fedotkin M.A. Processes of servicing, and control systems. In: *Mathematical questiions in cybernetics (Matematicheskie voprosy kibernetiki)*, V. 6, 1996. P. 51–70. (in Russian)
[17] Fedotkin M.A. Non-local approach for definition of controlled random processes In: *Mathematical questiions in cybernetics (Matematicheskie voprosy kibernetiki)*, V. 7, 1998. P. 333-344. (in Russian)

MULTI-SERVER MARKOVIAN QUEUE WITH SUCCESSIVE OPTIONAL SERVICES

P. Vijaya Laxmi and E. Girija Bhavani

•

Department of Applied Mathematics, Andhra University, Visakhapatnam - 530 003, India
vijayalaxmiau@gmail.com
girijabhavaniedadasari@gmail.com

Abstract

In this study, we analyze a multi-server queueing model with two successive optional services. Each server provides FES as well as two optional services to each arriving customer, for a total of c servers. Every new customer requires the first essential service (FES). The customer may quit the system with probability $(1-r_0)$ or choose optional services supplied by the same server after finishing the FES. With probability r_0, customer chooses the first optional service (OS - 1). Following that, the customer has the option of joining the second optional service (OS - 2) with probability r_1 or leave the system with probability $(1-r_1)$. We obtain the steady-state probability distributions by applying matrix-geometric method. We also derive a number of performance measures of the queueing model. Sensitivity analysis is used to investigate the impact of various parameters on performance of the queueing model.

Keywords: queue, multi-server, first essential service, optional services, matrix-geometric method

I. Introduction

A common goal of service systems is reducing customer waiting times, which is usually achieved by using faster services or hiring more servers. Various fields like call centers, hospitals, supermarkets, and other situations that occur every day make use of multi-server queues. In classic works like Medhi [15] and Gross, Shortle, Thompson, and Harris [5], numerous results have been obtained in all aspects of the $M/M/c$ queue. The steady-state distribution of a truncated multi-channel queueing system with customers' impatience and general balk function has been considered by Abou-El-Ata and Hariri [1]. For more research topics regarding $M/M/c$ queues, refer to Kumar [9], Levy and Yechiali [13], Li and Stanford [11], Mora [16], Bouchentouf et al. [3] and the references therein.

Real-time service systems have instances where everyone needs the first essential service (FES) and only a few others need optional services provided by the same server. Madan [14] was the first to suggest an optional second service in an $M/G/1$ queueing system using the supplementary variable approach. Similarly, Ke [6] analyzed a queueing model using startup time, in which all arriving customers need FES, while some may require additional J optional services. Jain et al. [10] investigated multiple types of optional services and vacations for an unreliable server in an $M/G/1$ queue, in which the customer may prefer to select an optional service with probability r_1 or depart from the system with probability $(1-r_1)$. Further, the customer may also join for any one of i $(2 \leq i \leq l)$ optional services. In a study by Ke et al. [7], they examined an $M/M/c$ retrial queue with an additional optional service. In Yang et al. [19], they discussed an $M/M/R$ queueing model with a second optional channel and obtained steady-state probabilities and various system performance

measures by using a matrix-geometric method. Later, Ke et al. [8] extended this model to unlimited capacity. Research on a variety of queueing models dealing with optional services is available in Li and Wang [12], Yang and Chen [18], Anitha et al. [2], Chandrika and Kalaiselvi [4], etc.

There has been no research on a multi-server queueing model with two successive optional services despite the vast body of literature. The combination of multiple servers and successive optional services gives the queueing model more realism and versatility. In practice, there are several instances in which services are provided in stages, for example, once a customer enters multi-channel service facilities, they may proceed to the next stage in turn after finishing the previous stage. This applies to many different fields, including manufacturing systems, transportation systems, telecommunications, and many daily operations. The main objective of this study is to explore the steady-state behavior of an $M/M/c$ queue with one essential service and two successive optional services.

The remainder of this paper is organized as follows. Section 2 presents a model description. Section 3 contains the mathematical formulation of the proposed queueing model. In Section 4, we apply a matrix geometric approach to find the steady-state solution. The system characteristics are described in Section 5. Section 6 is devoted to present numerical illustrations of the queueing model through practical application. Finally, we wind up our study with conclusions in Section 7.

II. Description of the Model

Consider a multi-channel queueing model with infinite capacity, FES, and two successive optional services. The pictorial representation of the model is shown in Figure 1.

The following is a description of system's fundamental operation.

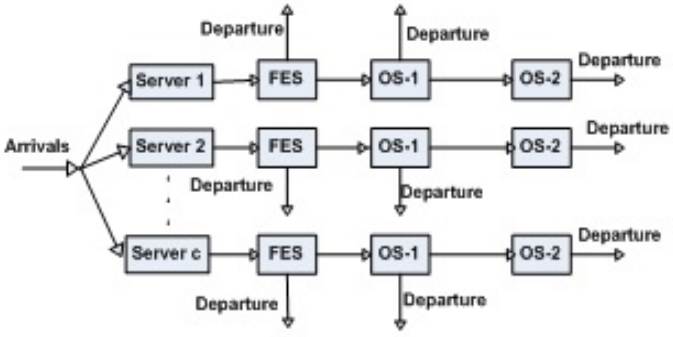


Figure 1: Model diagram

1. Arrival pattern follows Poisson process with parameter λ. There are c number of servers and each server provides FES as well as two optional services to each arriving customer.
2. After completing the FES, customer may leave the system with probability $(1 - r_0)$ or choose optional services provided by the same server. Customer opts for first optional service (OS - 1) with probability r_0 $(0 \leq r_0 \leq 1)$. After this, customer may join for second optional service (OS - 2) with probability r_1 $(0 \leq r_1 \leq 1)$ or may quit the system with probability $(1 - r_1)$. During FES, OS - 1 and OS - 2, the service times are exponentially distributed with rates of μ_0, μ_1, and μ_2, respectively.
3. The customer quits the system as soon as OS - 2 is completed, and the next consumer in the queue is allocated to FES. Each server can only serve one customer at a time and can only deliver one of three services (FES, OS - 1, OS - 2) at any given instant.
4. Upon arrival, the customer finds that all the servers are busy and must wait in the queue until one becomes available.

Practical Application

This model has real-time applications in motor vehicle service centers. The general services of vehicles include checking spark plugs, brake fluid, brake discs, checking for the normal functioning of lights, etc., which are mainly required by all vehicles. Engine oil replacement service is based on the distance traveled by the vehicle. Vehicles that have reached the certain miles enter the engine oil replacement service facility. After changing the oil, the mechanic performs the task of checking the oil filter. If it is misaligned or loose, it can be replaced. In this scenario, vehicles, mechanics, general service, engine oil replacement, and oil filter replacement, respectively, correspond to arrivals, servers, FES, OS - 1, OS - 2 in the queueing terminology.

III. Mathematical Formulation of the Model

Let $L(t)$ be the number of customers in the FES, $J_1(t)$ be the number of customers in OS - 1, and $J_2(t)$ be the number of customers in OS - 2 at time t. The process $\{(L(t), J_1(t), J_2(t)), t \geq 0\}$ defines a continuous-time Markov process with state space

$\chi = \{(i, j_1, j_2): i \geq 0, j_1 = 0,1,2,\ldots,c, j_2 = 0,1,2,\ldots,c.\}$.

It is noted that if $i + j_1 + j_2 \leq c$, the customer will receive the service immediately, if $i + j_1 + j_2 > c$, the newly arrived customer must wait in the queue.

We define the following steady-state probabilities for mathematical formulation.

P_{i,j_1,j_2} = Probability that i number of customers in the FES, j_1 number of customers in OS - 1, and j_2 number of customers in OS - 2, $i \geq 0,\ 0 \leq j_1,\ j_2 \leq c$.

Steady-State Equations:

Here, we develop the steady-state probability equations using the Markov process, which controls the dynamics of the queueing system as below.

Case I: When $j_1 = 0$ and $j_2 = 0$.

$$\lambda P_{0,0,0} = (1 - r_0)\mu_0 P_{1,0,0} + (1 - r_1)\mu_1 P_{0,1,0} + \mu_2 P_{0,0,1}, \quad (1)$$

$$(\lambda + i\mu_0)P_{i,0,0} = (i + 1)(1 - r_0)\mu_0 P_{i+1,0,0} + (1 - r_1)\mu_1 P_{i,1,0} + \mu_2 P_{i,0,1} + \lambda P_{i-1,0,0}, 1 \leq i \leq c - 1 \quad (2)$$

$$(\lambda + c\mu_0)P_{i,0,0} = c(1 - r_0)\mu_0 P_{i+1,0,0} + (1 - r_1)\mu_1 P_{i,1,0} + \mu_2 P_{i,0,1} + \lambda P_{i-1,0,0}, i \geq c. \quad (3)$$

Case II: When $1 \leq j_1 \leq c - 1$ and $j_2 = 0$.

$$(\lambda + j_1\mu_1)P_{0,j_1,0} = (1 - r_0)\mu_0 P_{1,j_1,0} + r_0\mu_0 P_{1,j_1-1,0} + (j_1 + 1)(1 - r_1)\mu_1 P_{0,j_1+1,0} + \mu_2 P_{0,j_1,1}, \quad (4)$$

$$(\lambda + i\mu_0 + j_1\mu_1)P_{i,j_1,0} = \lambda P_{i-1,j_1,0} + (i + 1)(1 - r_0)\mu_0 P_{i+1,j_1,0} + \mu_2 P_{i,j_1,1} + (i + 1)r_0\mu_0 P_{i+1,j_1-1,0} + (j_1 + 1)(1 - r_1)\mu_1 P_{i,j_1+1,0}, 1 \leq i \leq c - j_1 - 1, \quad (5)$$

$$(\lambda + (c - j_1)\mu_0 + j_1\mu_1)P_{i,j_1,0} = \lambda P_{i-1,j_1,0} + (c - j_1)(1 - r_0)\mu_0 P_{i+1,j_1,0} + (c - j_1 + 1)r_0\mu_0 P_{i+1,j_1-1,0} + (j_1 + 1)(1 - r_1)\mu_1 P_{i,j_1+1,0} + \mu_2 P_{i,j_1,1}, i \geq c - j_1. \quad (6)$$

Case III: When $j_1 = 0$ and $1 \leq j_2 \leq c - 1$.

$$(\lambda + j_2\mu_2)P_{0,0,j_2} = (1 - r_0)\mu_0 P_{1,0,j_2} + (1 - r_1)\mu_1 P_{0,1,j_2} + r_1\mu_1 P_{0,1,j_2-1} + (j_2 + 1)\mu_2 P_{0,0,j_2+1}, \quad (7)$$

$$(\lambda + i\mu_0 + j_2\mu_2)P_{i,0,j_2} = \lambda P_{i-1,0,j_2} + (i + 1)(1 - r_0)\mu_0 P_{i+1,0,j_2} + (1 - r_1)\mu_1 P_{i,1,j_2} + r_1\mu_1 P_{i,1,j_2-1}, 1 \leq i \leq c - j_2 - 1, \quad (8)$$

$$(\lambda + (c - j_2)\mu_0 + j_2\mu_2)P_{i,0,j_2} = \lambda P_{i-1,0,j_2} + (c - j_2)(1 - r_0)\mu_0 P_{i+1,0,j_2} + (1 - r_1)\mu_1 P_{i,1,j_2} + r_1\mu_1 P_{i,1,j_2-1} + (j_2 + 1)\mu_2 P_{i,0,j_2+1}, i \geq c - j_2. \quad (9)$$

Case IV: When $1 \leq j_1 \leq c - 1$, $1 \leq j_2 \leq c - 1$, and $j_1 + j_2 \leq c$.

$$(\lambda + j_1\mu_1 + j_2\mu_2)P_{0,j_1,j_2} = (1 - r_0)\mu_0 P_{1,j_1,j_2} + r_0\mu_0 P_{1,j_1-1,j_2} + (j_1 + 1)(1 - r_1)\mu_1 P_{0,j_1+1,j_2} + (j_1 + 1)r_1\mu_1 P_{0,j_1+1,j_2-1} + (j_2 + 1)\mu_2 P_{0,j_1,j_2+1},$$

$$j_1 + j_2 \le c - 1, \tag{10}$$

$$(\lambda + j_1\mu_1 + j_2\mu_2)P_{0,j_1,j_2} = r_0\mu_0 P_{1,j_1-1,j_2} + (j_1+1)r_1\mu_1 P_{0,j_1+1,j_2-1}, j_1 + j_2 = c. \tag{11}$$

$$\begin{aligned}(\lambda + i\mu_0 + j_1\mu_1 + j_2\mu_2)P_{i,j_1,j_2} &= \lambda P_{i-1,j_1,j_2} + (i+1)(1-r_0)\mu_0 P_{i+1,j_1,j_2} + \\ &(i+1)r_0\mu_0 P_{i+1,j_1-1,j_2} + (j_1+1)(1-r_1)\mu_1 P_{i,j_1+1,j_2} + (j_1+1)r_1\mu_1 P_{i,j_1+1,j_2-1} + \\ &(j_2+1)\mu_2 P_{i,j_1,j_2+1}, j_1 + j_2 \le c-1, 1 \le i \le c - j_1 - j_2 - 1,\end{aligned} \tag{12}$$

$$\begin{aligned}(\lambda + \beta\mu_0 + j_1\mu_1 + j_2\mu_2)P_{i,j_1,j_2} &= \lambda P_{i-1,j_1,j_2} + \beta(1-r_0)\mu_0 P_{i+1,j_1,j_2} + (\beta+1) \\ &r_0\mu_0 P_{i+1,j_1-1,j_2} + (j_1+1)\mu_1 P_{i,j_1+1,j_2} + (j_2+1)\mu_2 P_{i,j_1,j_2+1}, j_1 + j_2 \le c-1, \\ &i \ge c - j_1 - j_2.\end{aligned} \tag{13}$$

$$\begin{aligned}(\lambda + j_1\mu_1 + j_2\mu_2)P_{i,j_1,j_2} &= \lambda P_{i-1,j_1,j_2} + (\beta+1)r_0\mu_0 P_{i+1,j_1-1,j_2} + \\ &(j_1+1)r_1\mu_1 P_{i,j_1+1,j_2-1}, i \ge \beta + 1, \\ &j_1 + j_2 = c, \beta = c - j_1 - j_2.\end{aligned} \tag{14}$$

Case V: When $j_1 = c$ and $j_2 = 0$.

$$(\lambda + c\mu_1)P_{0,c,0} = r_0\mu_0 P_{1,c-1,0}, \tag{15}$$

$$(\lambda + c\mu_1)P_{i,c,0} = \lambda P_{i-1,c,0} + r_0\mu_0 P_{i+1,c-1,0}, i \ge 1. \tag{16}$$

Case VI: When $j_1 = 0$ and $j_2 = c$.

$$(\lambda + c\mu_2)P_{0,0,c} = r_1\mu_1 P_{0,1,c-1}, \tag{17}$$

$$(\lambda + c\mu_1)P_{i,0,c} = \lambda P_{i-1,0,c} + r_1\mu_1 P_{i,1,c-1}, i \ge 1. \tag{18}$$

IV. Matrix-Geometric Method

For the QBD process introduced in Section 3, obtaining a closed form solution is very difficult. We employ the matrix-geometric method to analyze the probabilities of the Markov chain in order to develop an effective and numerically stable solution. We can simply obtain the stationary probability vector using the matrix-geometric method because the transition rate matrix contains repeated block sub-matrices.

Applying the concept of Neuts [17], the infinitesimal generator **Q** for the process could be given as follows.

$$\mathbf{Q} = \begin{pmatrix} \mathbf{A}_0 & \mathbf{C} & & & & & & \\ \mathbf{B}_0 & \mathbf{A}_1 & \mathbf{C} & & & & & \\ & \mathbf{B}_1 & \mathbf{A}_1 & \mathbf{C} & & & & \\ & & \mathbf{B}_2 & \mathbf{A}_2 & \mathbf{C} & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & \mathbf{B}_c & \mathbf{A}_c & \mathbf{C} & \\ & & & & & \mathbf{B}_c & \mathbf{A}_c & \mathbf{C} \\ & & & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

We denote the transition probability from state (i, j_1, j_2) to the state $(\hat{\imath}, \hat{\jmath}_1, \hat{\jmath}_2)$ by $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)}$. The elements of the sub-matrix $\mathbf{A}_i$, $i \ge 0$ are given as:

- $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)} = -(\lambda + i\mu_0)$, for $0 \le i \le c-1, j_1 = j_2 = 0, \hat{\imath} = i, \hat{\jmath}_1 = j_1, \hat{\jmath}_2 = j_2$.
- $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)} = -(\lambda + c\mu_0)$, for $i \ge c, j_1 = j_2 = 0, \hat{\imath} = i, \hat{\jmath}_1 = j_1, \hat{\jmath}_2 = j_2$.
- $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)} = -(\lambda + i\mu_0 + j_2\mu_2)$, for $i \ge 0, 1 \le j_2 \le c, i + j_2 \le c, j_1 = 0, \hat{\imath} = i, \hat{\jmath}_1 = j_1, \hat{\jmath}_2 = j_2$.
- $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)} = -(\lambda + (c-j_2)\mu_0 + j_2\mu_2)$, for $i \ge 0, 1 \le j_2 \le c, i + j_2 > c, j_1 = 0, \hat{\jmath}_2 = j_2, \hat{\imath} = i, \hat{\jmath}_1 = j_1, \hat{\jmath}_2 = j_2$.
- $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)} = -(\lambda + i\mu_0 + j_1\mu_1)$, for $i \ge 0, 0 \le j_1 \le c\ i + j_1 \le c, j_2 = 0, \hat{\jmath}_1 = j_1, \hat{\imath} = i, \hat{\jmath}_1 = j_1, \hat{\jmath}_2 = j_2$.
- $P_{(i,j_1,j_2),(\hat{\imath},\hat{\jmath}_1,\hat{\jmath}_2)} = -(\lambda + (c-j_1)\mu_0 + j_1\mu_2)$, for $i \ge 0, i = \hat{\imath}, 0 \le j_1 \le c, i + j_1 > c, j_2 = 0, \hat{\jmath}_1 = j_1, \hat{\jmath}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = -(\lambda + (c - j_1 - j_2)\mu_0 + j_1\mu_1 + j_2\mu_2)$, for $i \geq 0, \hat{\imath} = i, j_1 + j_2 \leq c, 1 \leq j_1 \leq c-1, 1 \leq j_2 \leq c-1, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = j_1(1 - r_1)\mu_1$, for $i = 0, \hat{\imath} = i, 1 \leq j_1 \leq c, 0 \leq j_2 \leq c-1, j_1 + j_2 \leq c, \hat{j}_1 = j_1 - 1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = j_2\mu_2$, for $i = 0, \hat{\imath} = i, 1 \leq j_2 \leq c, 0 \leq j_1 \leq c-1, \hat{j}_2 = j_2 - 1$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = j_1 r_1 \mu_1$, for $i = 0, \hat{\imath} = i, 1 \leq j_1 \leq c, 0 \leq j_2 \leq c-1, \hat{j}_1 = j_1 - 1\ \hat{j}_2 = j_2 + 1$.

The elements of the sub-matrix $\mathbf{B}_i$, $i \geq 0$ are taken as:

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = i(1 - r_0)\mu_0$, for $1 \leq i \leq c-1, \hat{\imath} = i - 1, j_1 = j_2 = 0, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = c(1 - r_0)\mu_0$, for $i \geq c, \hat{\imath} = i - 1, j_1 = j_2 = 0, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = i(1 - r_0)\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_2 \leq c, i + j_2 \leq c, j_1 = 0, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = (c - j_2)(1 - r_0)\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_2 \leq c, i + j_2 > c, j_1 = 0, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = i(1 - r_0)\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_1 \leq c, i + j_1 \leq c, j_2 = 0, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = (c - j_1)(1 - r_0)\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_1 \leq c, i + j_1 > c, j_2 = 0, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = (c - j_1 - j_2)(1 - r_0)\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 1 \leq j_1 \leq c - 1, 1 \leq j_2 \leq c-1, \hat{j}_1 = j_1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = ir_0\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_2 \leq c, i + j_2 \leq c, j_1 = 0, \hat{j}_1 = j_1 + 1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = (c - j_2)r_0\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_2 \leq c, i + j_2 > c, j_1 = 0, \hat{j}_1 = j_1 + 1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = ir_0\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_1 \leq c, i + j_1 \leq c, j_2 = 0, \hat{j}_1 = j_1 + 1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = (c - j_1)r_0\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 0 \leq j_1 \leq c, i + j_1 > c, j_2 = 0, \hat{j}_1 = j_1 + 1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = (c - j_1 - j_2)r_0\mu_0$, for $i \geq 0, \hat{\imath} = i - 1, 1 \leq j_1 \leq c-1, 1 \leq j_2 \leq c-1, j_1 + j_2 \leq c, \hat{j}_1 = j_1 + 1, \hat{j}_2 = j_2$.

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = 0$, when $j_1 + j_2 = c$.

The elements of sub-matrix $\mathbf{C}$ are given as follows:

• $P_{(i,j_1,j_2),(\hat{\imath},\hat{j}_1,\hat{j}_2)} = \lambda$, for $i \geq 0, \hat{\imath} = i + 1, 0 \leq j_1 \leq c, 0 \leq j_2 \leq c, \hat{j}_1 = j_1$, and $\hat{j}_2 = j_2$.

Here, the sub-matrices $\mathbf{C}$, $\mathbf{A}_i$, $\mathbf{B}_i$, $i \geq 0$ are of order $1 + \sum_{n=1}^{c-1} n + 2c$.

4.1 Steady-state solution

Based on $\mathbf{Q}$ matrix structure, one can easily notice that the process $\{L(t), J_1(t), J_2(t), t \geq 0\}$ is a QBD process. As per the block structure of $\mathbf{Q}$, the stationary distribution of the process can be composed as segmented vectors, denoted as,

$$P_{i,j_1,j_2} = \lim_{t\to\infty} P\{L(t) = i, J_1(t) = j_1, J_2(t) = j_2\}, (i, j_1, j_2) \in \chi$$

According to Neuts (1981), the system is stable and the steady state probability vector exists if and only if $\tilde{\mathbf{Y}}\mathbf{C}\mathbf{e} < \tilde{\mathbf{Y}}\mathbf{B_c}\mathbf{e}$ where $\tilde{\mathbf{Y}}$ is an invariant probability of the matrix $\boldsymbol{\psi} = \mathbf{B_c} + \mathbf{A_c} + \mathbf{C}$. The equations $\tilde{\mathbf{Y}}\boldsymbol{\psi} = \mathbf{0}$ and $\tilde{\mathbf{Y}}\mathbf{e} = \mathbf{1}$ can be satisfied by $\tilde{\mathbf{Y}}$.

Under the stability condition, let $\mathbf{P}$ be the stationary probability vector of the generator $\mathbf{Q}$ satisfying the balance equation $\mathbf{PQ} = \mathbf{0}$ and $\mathbf{Pe}_i = 1$, where $\mathbf{0}$ is the row vector with all elements as zero and $\mathbf{e}_i$ is the column vector of appropriate dimension i with every element as one. The vector $\mathbf{P}$ partitioned as $\mathbf{P} = [\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \dots]$, where

$$\mathbf{P}_i = [P_{i,0,0}, P_{i,0,1}, P_{i,0,2}, \dots, P_{i,0,c}, P_{i,1,0}, P_{i,2,0}, \dots, P_{i,c,0}, P_{i,1,1}, P_{i,1,2}, \dots, P_{i,1,c-1},$$
$$P_{i,2,1}, P_{i,2,2}, \dots, P_{i,2,c-2}, \dots, P_{i,c-2,1}, P_{i,c-2,2}, P_{i,c-1,1}],\ i \geq 0.$$

When the stability criterion is met, the sub-vectors of $\mathbf{P}$ pertaining to various levels appear to satisfy

$$\mathbf{P}_i = \mathbf{P}_c\mathbf{R}^{i-c}, i \geq c, \tag{19}$$

where the matrix $\mathbf{R}$ is the minimal non-negative solution of the matrix quadratic equation

$$\mathbf{C} + \mathbf{R}\mathbf{A}_c + \mathbf{R}^2\mathbf{B}_c = \mathbf{0}. \tag{20}$$

The QBD process is positive recurrent if and only if the spectral radius $Sp(\mathbf{R}) < 1$. Further, it is rather complex to determine the explicit expression of the matrix $\mathbf{R}$ by solving equation (20). Neuts [17] has devised an iterative algorithm for calculating $\mathbf{R}$ numerically. We starting with initial iteration $\mathbf{R}_0 = 0$, and calculate the successive approximations using

$$\mathbf{R}_{i+1} = -(\mathbf{C} + \mathbf{R}_i^2\mathbf{B}_c)(\mathbf{A}_c)^{-1}, i \geq 0.$$

Now, $\mathbf{R}$ can be determined iteratively until it converges, i.e., $\lim_{i\to\infty}\mathbf{R}_i = \mathbf{R}$.

Using the equation $\mathbf{PQ} = \mathbf{0}$, the governing system of difference equations are expressed as follows

$$\mathbf{P}_0\mathbf{A}_0 + \mathbf{P}_1\mathbf{B}_1 = 0, \tag{21}$$
$$\mathbf{P}_{i-1}\mathbf{C} + \mathbf{P}_i\mathbf{A}_i + \mathbf{P}_{i+1}\mathbf{B}_{i+1} = 0, 1 \leq i \leq c-1, \tag{22}$$
$$\mathbf{P}_{i-1}\mathbf{C} + \mathbf{P}_i\mathbf{A}_c + \mathbf{P}_{i+1}\mathbf{B}_c = 0, i \geq c, \tag{23}$$

and the normalizing condition

$$\sum_{i=0}^{\infty} \mathbf{P}_i\mathbf{e}_i = 1. \tag{24}$$

After applying some mathematical manipulations to equations (21) to (23), we get

$$\mathbf{P}_{i-1} = \mathbf{P}_i\boldsymbol{\phi}_i, 1 \leq i \leq c, \tag{25}$$
$$\mathbf{P}_c[\boldsymbol{\phi}_c\mathbf{C} + \mathbf{A}_c + \mathbf{R}\mathbf{B}_c] = \mathbf{0}, \tag{26}$$

where

$$\boldsymbol{\phi}_1 = -\mathbf{B}_0(\mathbf{A}_0^{-1}), \boldsymbol{\phi}_i = -\mathbf{B}_i(\mathbf{A}_{i-1} + \boldsymbol{\phi}_{i-1}\mathbf{C})^{-1}, 1 \leq i \leq c.$$

Using equations (24) and (25), we obtain

$$\mathbf{P}_c[\sum_{j=1}^{c}\prod_{n=c}^{l}\boldsymbol{\phi}_n + (\mathbf{I} - \mathbf{R})^{-1}]\mathbf{e}_i = 1. \tag{27}$$

Solving equations (26) and (27) yields $\mathbf{P}_c$. We use equations (19) and (25) to get $\mathbf{P}_i$ for $i \geq 0$.

V. Performance Measures

An infinite capacity multi-server queueing system with two successive optional services has several system characteristics, such as the expected length of the system in FES, OS - 1, and OS - 2, the expected number of customers in the system, the expected number of idle servers, the expected number of busy servers, probability that the system is empty, can be obtained by using steady-state probabilities. The expressions of above are given as follows:

- Expected number of customers in FES

$$E[L_f] = \sum_{i=1}^{\infty} iP_{i,0,0} + \sum_{i=1}^{\infty} i\sum_{j_1=1}^{c-1} P_{i,j_1,0} + \sum_{i=1}^{\infty} i\sum_{j_2=1}^{c-1} P_{i,0,j_2} +$$

$$\sum_{i=1}^{\infty} i\sum_{j_1=1}^{c-1}\sum_{j_2=1}^{c-j_1} P_{i,j_1,j_2} + \sum_{i=1}^{\infty} iP_{i,c,0} + \sum_{i=1}^{\infty} iP_{i,0,c}.$$

- Expected number of customers in OS - 1

$$E[L_{s_1}] = \sum_{j_1=1}^{c-1} j_1\sum_{i=0}^{\infty} P_{i,j_1,0} + \sum_{j_1=1}^{c-1} j_1\sum_{i=0}^{\infty}\sum_{j_2=1}^{c-j_1} P_{i,j_1,j_2} + c\sum_{i=0}^{\infty} P_{i,c,0}.$$

- Expected number of customers in OS - 2

$$E[L_{s_2}] = \sum_{j_2=1}^{c-1} j_2 \sum_{i=0}^{\infty} P_{i,0,j_2} + \sum_{j_2=1}^{c-1} j_2 \sum_{i=0}^{\infty} \sum_{j_1=1}^{c-j_2} P_{i,j_1,j_2} + c \sum_{i=0}^{\infty} P_{i,0,c}.$$

- Expected number of customers in the system

$$E[L] = E[L_f] + E[L_{s_1}] + E[L_{s_2}].$$

- Expected number of idle servers

$$E[I] = \sum_{i+j_1+j_2=0}^{c-1} [c - max(i, j_1, j_2)] P_{i,j_1,j_2}.$$

- Expected number of busy servers

$$E[B] = c - E[I].$$

- Probability that the system is empty is $P_{0,0,0}$.

VI. Numerical Investigations

To understand the system long run behaviour change with the parameters, we have conducted some numerical studies on the system characteristics by changing the parameter values. Considering the practical application given in Section 2, we perform the sensitivity analysis using arbitrarily selected parameters $\lambda = 1.0$, $\mu_0 = 5.0$, $\mu_1 = 4.5$, $\mu_2 = 3.0$, $r_0 = 0.6$, $r_1 = 0.5$, $c = 4$, where

λ = The rate at which vehicles arrive at the service center,
μ_0 = Service rate for general services, including spark plugs check, brake fluid, brake discs, etc. (FES),
μ_1 = Service rate of engine oil replacement service (OS - 1),
μ_2 = Service rate of oil filters replacement service (OS - 2),
r_0 = Probability that vehicles taken for engine oil replacement,
r_1 = Probability that vehicles taken for oil filter replacement,
c = Amount of mechanics in the vehicle service center.

Table 1: Effect of r_0 on $E[L]$

	$E[L]$			
r_0	$\mu_2 = 3.0$	$\mu_2 = 3.2$	$\mu_2 = 3.4$	$\mu_2 = 3.6$
0.1	0.12143	0.11988	0.11848	0.11721
0.3	0.32241	0.31685	0.31179	0.30718
0.5	0.70647	0.69319	0.68101	0.66976
0.7	1.45122	1.4277	1.40562	1.38481
0.9	2.17964	2.16400	2.14999	2.13729

Table 2: Effect of r_1 on $E[L]$

	$E[L]$			
r_1	$\mu_2 = 3.0$	$\mu_2 = 3.2$	$\mu_2 = 3.4$	$\mu_2 = 3.6$
0.2	0.81844	0.81319	0.80852	0.80434
0.4	0.92491	0.91515	0.90642	0.89855
0.6	1.07338	1.06013	1.04821	1.03741
0.8	1.27478	1.25986	1.24638	1.23412
1.0	1.55389	1.54122	1.52988	1.51967

Tables 1 and 2 show the impact of the probabilities r_0 and r_1, on the expected length of the system $E[L]$ for different values of the service rates in OS - 1 (μ_1) and OS - 2 (μ_2). We observe that

- As r_0 (r_1) increases, the number of vehicles opting for engine oil (oil filter) replacement facility increases, which tends to increase the waiting time of vehicles at the service center. Hence $E[L]$ increases.
- Also, an increase in the service rate μ_1 (μ_2) reduces $E[L]$, which agrees with our intuition.

Table 3: Effect of λ on performance measures

	$r_0 = 0.0$ and $r_1 = 0.0$			$r_0 = 1.0$ and $r_1 = 1.0$		
λ	$E[L]$	$E[I]$	$E[B]$	$E[L]$	$E[I]$	$E[B]$
0.3	0.01522	3.98478	0.01522	2.19881	1.49894	2.50106
0.6	0.03085	3.96915	0.03085	2.33150	1.37036	2.62964
0.9	0.04685	3.95315	0.04685	2.47134	1.25348	2.74652
1.2	0.06318	3.93681	0.06319	2.61849	1.14669	2.85331
1.5	0.07982	3.92015	0.07985	2.77359	1.04824	2.95176
1.8	0.09671	3.90322	0.09677	2.93725	0.95641	3.04360

Table 3 shows the impact of arrival rate λ on expected number of vehicles at the service center $E[L]$, expected amount of idle mechanics $E[I]$, and expected amount of busy mechanics $E[B]$ in two situations as follows:

Case a: When no vehicle is opting for optional services ($r_0 = 0, r_1 = 0$)

Case b: When all arriving vehicles are opting both optional services ($r_0 = 1, r_1 = 1$)

It is observed that

- An increase in λ results in increase of $E[L]$, $E[B]$ and decrease of $E[I]$ for a fixed r_0 and r_1, as expected.
- Further, for a fixed λ, $E[L]$ and $E[B]$ are seen smaller when no vehicles adopting any optional service provided by a service center. On the other hand, $E[I]$ is smaller when all arriving vehicles choose both optional services, as anticipated.

Table 4: Effect of r_0 and r_1 on performance measures

	$r_1 = 0.3$		$r_1 = 0.5$		$r_1 = 0.7$	
r_0	$E[L]$	$E[I]$	$E[L]$	$E[I]$	$E[L]$	$E[I]$
0.3	0.23906	3.76450	0.31179	3.68605	0.40389	3.58498
0.5	0.54262	3.45912	0.68101	3.28845	0.85592	3.07158
0.7	1.38111	2.62083	1.40562	2.50599	1.51586	2.31483

The effect of the probability of opting OS - 1 and OS - 2 (r_0 and r_1) on $E[L]$ and $E[I]$ is shown in Table 4.

- For a fixed r_1, as r_0 increases, the number of vehicles opting for engine oil service grows, resulting in an increase in the number of vehicles waiting for service at the service center $E[L]$. Moreover, for a fixed r_0, the same trend is observed for $E[L]$ with increase in r_1.
- However, it can be seen that increase in r_0 (r_1) yields the lower $E[I]$. This is because an increase in these probabilities increases the vehicle service time.
- Also, considering the cases $r_0 < r_1 (= 0.7)$ and $r_0 > r_1 (= 0.3)$, we notice that $E[L]$ is higher when $r_0 < r_1$ for the chosen parameter values.

Table 5: Effect of λ on performance measures for different μ_0, μ_1, and μ_2

	$\mu_0 = 5.2, \mu_1 = 4.6, \mu_2 = 3.5$		$\mu_0 = 3.5, \mu_1 = 4.0, \mu_2 = 5.0$	
λ	$E[L]$	$E[I]$	$E[L]$	$E[I]$
0.4	0.45349	3.53054	0.47686	3.52990
0.8	0.77994	3.18595	0.82567	3.18415
1.2	1.03378	2.91530	1.09905	2.91338
1.6	1.24318	2.69213	1.32574	2.69058

The impact of λ on $E[L]$ and $E[I]$ for different μ_0, μ_1, and μ_2 is shown in Table 5. Here we depicted the comparison of cases $\mu_0 > \mu_1 > \mu_2$ and $\mu_0 < \mu_1 < \mu_2$. As shown in Table 3, increase of λ increases $E[L]$ and decreases $E[I]$. Evidently, from the table, expected size of vehicles at the service center can be reduced by taking $\mu_0 > \mu_1 > \mu_2$. This helps the service center managers to run the system effectively when the arrival rate of vehicles is high.

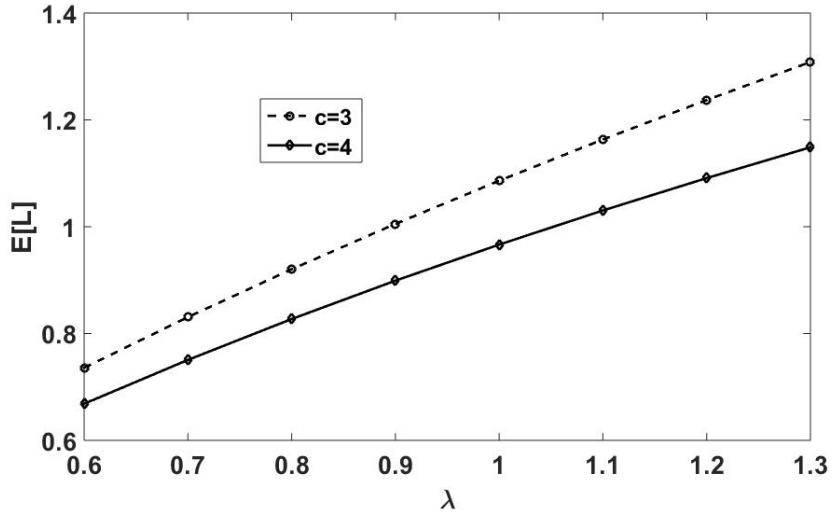


Figure 2: Effect of λ on $E[L]$ for different c

Figure 2 illustrates the impact of λ on $E[L]$ for various c values. As we seen in the tables, an increase in λ increases $E[L]$ for a fixed amount of mechanics c. Furthermore, an opposite effect is observed with the increase in c, this is due to the fact that increase of mechanics decreases vehicles waiting time. From this figure, we conclude that when the arrival rate of vehicles at service center is high, one can reduce the system size by increasing the number of mechanics, even though the service rates are kept constant.

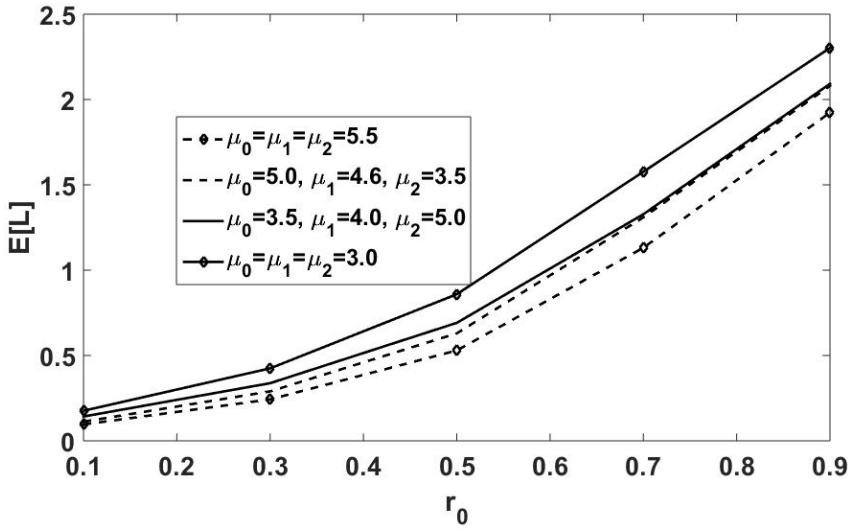


Figure 3: Effect of r_0 on $E[L]$

Figure 3 explores the impact of r_0 on $E[L]$ for different values of service rates. It is clear from the figure that as the number of vehicles opting for the first optional service facility increases, $E[L]$ increases. Subsequently, system performance can be ranked, with $\mu_0 = \mu_1 = \mu_2 = 5.5$ being best, followed by $\mu_0 = 5.0$, $\mu_1 = 4.6$, $\mu_2 = 3.5$ and then $\mu_0 = 3.5$, $\mu_1 = 4.0$, $\mu_2 = 5.0$, and lastly $\mu_0 = \mu_1 = \mu_2 = 3.0$.

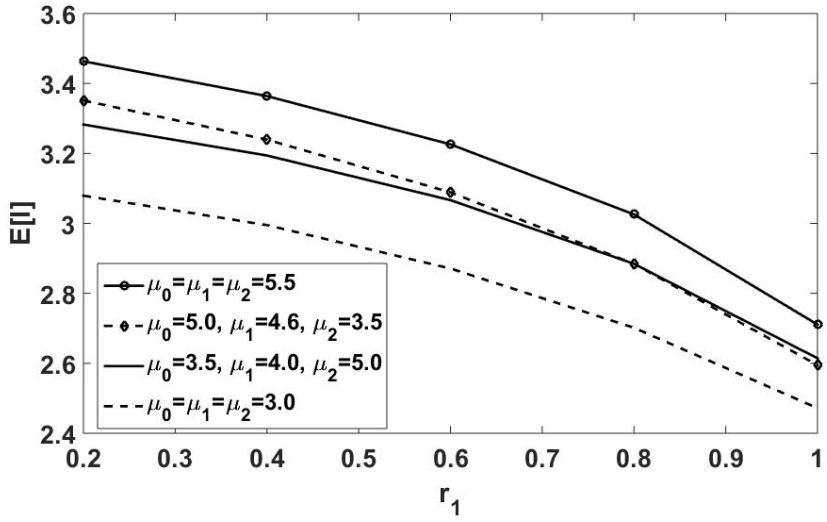


Figure 4: Effect of r_1 on $E[I]$

The impact of r_1 on expected number of idle servers $E[I]$ for different values of service rates in depicted in Figure 4. It is obvious that an increase in probability of vehicles choosing for oil filter service after getting engine oil service facility r_1 decreases the idle time of the mechanics at service center. Hence, $E[I]$ decreases. On the other hand, $E[I]$ is smaller when $\mu_0 = \mu_1 = \mu_2 = 3.0$ and higher when $\mu_0 = \mu_1 = \mu_2 = 5.5$. Also, it is quite interesting to note that while comparing the cases $\mu_0 = 5.0$, $\mu_1 = 4.6$, $\mu_2 = 3.5$ and $\mu_0 = 3.5$, $\mu_1 = 4.0$, $\mu_2 = 5.0$, $E[I]$ is observed higher for $\mu_0 = 5.0$, $\mu_1 = 4.6$, $\mu_2 = 3.5$ when $r_1 \leq 0.6$. At $r_1 = 0.7$ two curves are almost coincide and at $r_1 = 0.8$ two curves intersect each other. Further, for $r_1 > 0.8$, the trend is reversed and $E[I]$ is seen higher for $\mu_0 = 3.5$, $\mu_1 = 4.0$, $\mu_2 = 5.0$. This reveals the fact that as more and more vehicles are opting OS - 2 ($r_1 > 0.6$), by taking μ_2 as bigger than μ_0 and μ_1, $E[I]$ will be smaller (here r_0 value is chosen as 0.6).

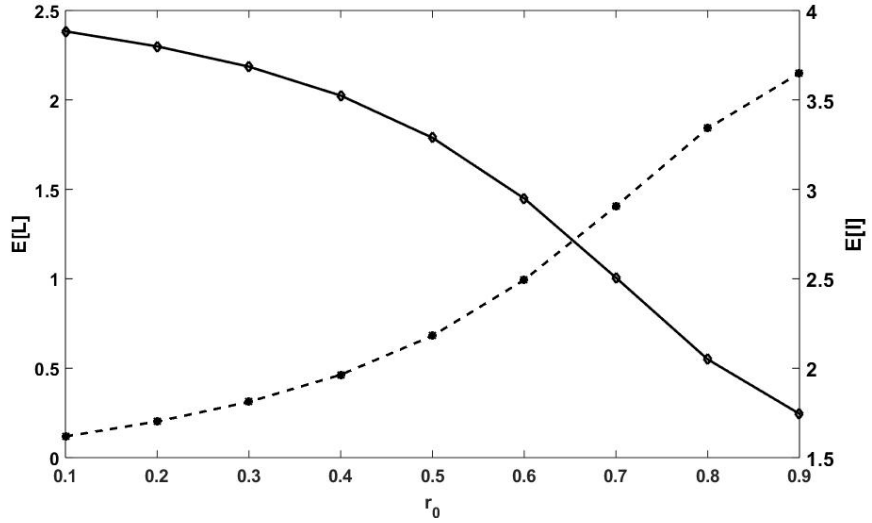


Figure 5: Effect of r_0 on $E[L]$ and $E[I]$

Figure 5 exhibits the effect of r_0 on $E[L]$ and $E[I]$. It demonstrates that $E[L]$ and $E[I]$ increases and decreases, respectively, with the increase in r_0. The point of intersection of two curves determines the value of r_0 at which $E[L]$ and $E[I]$ are the maximum and minimum, respectively. As a result, service center managers can optimize $E[L]$ or $E[I]$ by taking the appropriate measures using r_0 knowledge.

VII. Conclusion

In this study, we have carried out the analysis of $M/M/c$ queueing model with two successive optional services. Using QBD process and matrix geometric method, we have obtained the stationary probability distribution of the model. Further we have derived some performance measures of the model such as expected length of the system, expected number of idle servers and expected number of busy servers. Sensitivity analysis has been carried out by considering the practical application of the model. Through our numerical and graphical studies, it is observed that

- Expected number of vehicles at the service center increases with the increase of arrival rate and probability of opting optional services.
- System size can be reduced by increasing the amount of mechanics when the arrival rate is high.
- System size is smaller when $\mu_0 > \mu_1 > \mu_2$ for a constant arrival rate and optional service probabilities.
- When more number of vehicles opt for optional services, $E[L]$ can be decreased by taking equal higher service rates in FES, OS -1,OS -2.

This research work may extended further by incorporating the concepts of working vacations, customers' impatience, server breakdowns, etc.

References

[1] Abou-El-Ata, M. O. and Hariri, A. M. A. (1992). The $M/M/c/N$ queue with balking and reneging. *Computers & Operations Research*, 19(8):713–716.

[2] Anitha, K., Maragathasundari, S. and Bala, M. (2017). Queuing system of bulk arrival model with optional services in third stage and two different vacation policies. *International Journal of Mathematics and its Applications*, 5(4E):711–721.

[3] Bouchentouf, A. A., Cherfaoui, M. and Boualem, M. (2020). Analysis and performance evaluation of Markovian feedback multi-server queueing model with vacation and impatience. *American Journal of Mathematical and Management Sciences*, https://doi.org/:10.1080/01966324.2020.1842271

[4] Chandrika, K. U. and Kalaiselvi, C. (2013). Batch arrival feedback queue with additional multi optional service and multiple vacation. *International Journal of Scientific and Research Publications*,. 3(3):1–8.

[5] Gross, D., Shortle, J., Thompson, J. and Harris, C. (2013). *Fundamentals of Queueing Theory* ,John Wiley & Sons.

[6] Jain, M., Sharma, G. C. and Sharma, R. (2013). Unreliable server $M/G/1$ queue with multi-optional services and multi-optional vacations. *International Journal of Mathematics in Operational Research*, 5(2):145–169.

[7] Ke, J.C. (2008). An $M^{[X]}/G/1$ system with startup server and J additional options for service. *Applied Mathematical Modelling*, 32:443–458.

[8] Ke, J.C., Wu, C.H. and Pearn, W.L. (2011). Multi-server retrial queue with second optional service: algorithmic computation and optimisation. *International Journal of Systems Science*, 42(10): 1755–1769.

[9] Ke, J. C., Wu, C. H. and Pearn, W. L. (2013). Analysis of an infinite multi-server queue with an optional service. *Computers & Industrial Engineering*, 65:216–225.

[10] Kumar, R. (2013). Economic analysis of an $M/M/c/N$ queuing model with balking, reneging and retention of reneged customers. *OPSEARCH*, 50:383–403.

[11] Levy, Y. and Yechiali, U. (1976). An $M/M/s$ queue with servers vacations. *INFOR: Information Systems and Operations Research*, 14(2):153–163.

[12] Li, N. and Stanford, D. A. (2016). Multi-server accumulating priority queues with heterogeneous servers. *European Journal of Operational Research*, 252(3):866–878.

[13] Li, J. and Wang, J. (2006). An $M/G/1$ retrial queue with second multi-optional service, feedback and unreliable server. *Applied Mathematics-A Journal of Chinese Universities*, 21:252–262.

[14] Madan, K.C. (2000). An $M/G/1$ queue with second optional service. *Queueing Systems*, 3(4):37–46.

[15] Medhi, J. (2002) *Stochastic models in queueing theory*, Academic Press.

[16] Mora, J. N. (2019). Resource allocation and routing in parallel multi-server queues with abandonments for cloud profit maximization. *Computers & Operations Research*, 103:221–236.

[17] Neuts, M.F. (1981) Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach, John Hopkins University Press, Baltimore.

[18] Yang, D. Y. and Chen, Y. H. (2018). Computation and optimization of a working breakdown queue with second optional service. *Journal of Industrial and Production Engineering*, 35(3):181–188.

[19] Yang, D. Y., Wang, K. H. and Kuo Y. T. (2011). Economic application in a finite capacity multi-channel queue with second optional channel. *Applied Mathematics and Computation*, 217:7412–7419.

RT&A, No 4 (65)
Volume 16, December 2021

MARKOV RELIABILITY MODEL OF A WIND FARM

Victor Yu. Itkin

•

National University of Oil and Gas "Gubkin University"
itkin.v@gubkin.ru

Abstract

A Markov reliability model of a wind farm has been built using the example of Anholt wind farm, Denmark. Reliability indicators of wind turbine equipment are calculated as wind speed functions. Basing hourly measurements of the wind speed and the consumed electricity, two samples of duration time of the met and unmet demand of electricity were obtained. It has been found that these samples can be approximated with exponential mixture model of the probability distributions. The wind farm operation process has been approximated with a continues-time 5-states Markov process. As a result, stationary and non-stationary probabilities that the electricity demand will be met by wind power were estimated.

Keywords: wind farm, reliability, availability, met demand, unmet demand, exponential mixture model of distributions, continues-time finite-states Markov process.

1. Introduction

The share of renewable sources in the electricity market is constantly growing, however, at the current stage of development, they cannot guarantee the supply of electricity to consumers. The renewable sources depend on sufficiently significant random factors such as wind speed or insolation intensity. In this regard, it is necessary to use combined power supply systems that include both traditional and renewable energy sources.

A study of the reliability of such systems allows assessing how much power reserve of traditional energy sources is necessary to cover the deficit in the event of insufficient generation of renewable energy sources. It should take into account both instability of wind speed and usual equipment failures. In this paper, one of the most powerful wind farms, the Anholt wind farm in Denmark, is considered as an example. It was built in 2013 and consists of 111 wind turbines Siemens Gamesa Renewable Energy, SWT 3.6-120, the maximum capacity of each of them is 3.6 MW. The whole power plant can generate up to 400 MW, which is about 2.7 % of Denmark's electricity need.

There are numerous studies in literature on reliability models of wind farms and combined energy systems. In [1], a Markov model of combined power gas and thermal networks was built and the reliability of small business supply in Germany during a standard weekend day was investigated. In [2], the optimal parameters of a combine power plant consisting of gas and wind generators were evaluated. Wind energy was accounted for using a probability density, the estimate of which has not been included in this article. In [3, 4, 5], various models of the wind farm reliability were considered. These models take into account that the failure rates of wind turbine equipment are dependent on wind speed. Wind speed values were modelled with the Monte-Carlo method [3, 4] or with Markov chains [5]. In [6], reliability indicators of a wind farm were calculated with probability-generating functions. In [7], generated power of a wind farm was evaluated with a cubic model. The wind speed, as a random variable, was approximated by the Gnedenko-Weibull distribution. The distribution parameters were estimated according to statistical data, and all the measurements were considered independent, i.e. the correlation structure of measurements' series was not taken in account.

The correlation structure of a wind speed time series was studied in [8, 9] and other papers, where the short-terms forecasting problems of wind speed were considered. These problems were solved with neural network models, ARIMA models, etc. These models are not very accurate, so, in this paper, the generated power will be predicted based on real meteorological data. In [10] and [11], wind farm equipment reliability indicators were studied as functions of wind speed based on statistical data.

The models presented in some of the listed articles were utilized in our paper when processing a large amount of statistical data for a specific object – Anholt wind farm. As a result, the statistical patterns were identified, which made it possible to build a Markov model and estimate the reliability indicators of the wind farm.

The paper is organized as follows. In section 2, the reliability indicators of a turbine are estimated as a function of wind speed. In section 3, statistical data and mathematical models are investigated to evaluate the electricity demand and the wind farm capacity to produce it. Then a Markov model of the wind farm operation process has been built and its reliability indicators are estimated. In section 4, the results of the study are summarized.

2. Equipment Reliability of a Turbine

At first, let's estimate equipment reliability indicators of a wind farm. To do this, it is enough to consider any turbine. A stationary availability will be consider as a main reliability indicator:

$$K = \frac{T}{T+R},$$

where T is the mean time between failures (MTBF), R is the mean time to repair (MTTR).

In [10], it is found that the failure rates of wind farm equipment linearly depend on wind speed W in the speed range $W = 7 - 11$ m/s, but the result of extrapolating this dependence to a wider range of real wind speed in the region of Anholt Island is implausible: the failure rate forecast for small wind speeds is negative (Fig. 1). It is natural to assume the equipment does not fail when it is idle, i.e. if wind speed is zero, the failure rate must be zero too. This idea leads us to a quadratic model without a constant:

$$\lambda(W) = b_1 W + b_2 W^2, \tag{1}$$

where $\lambda(W)$ is the failure rate, 1/year, $b_1 = 0.353$ s/(year·m) and $b_2 = 0.0868$ s^2/(year·m^2) are the model coefficients fitted by least squares. Fig. 1 shows this model does not differ from the linear practically in speed range 7–11 m/s.

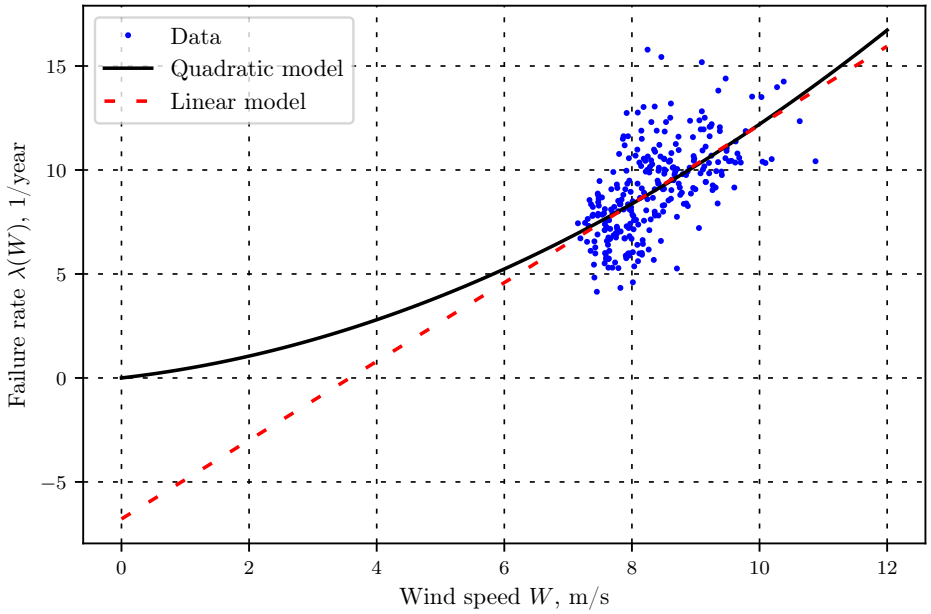


Figure 1: *Dependence of the Failure Rate on Wind Speed*

In [10], the estimations of failure rates (Table 1) and of mean times to repair (Table 2) for various equipment of a wind turbine are presented.

Table 1: *Failure rates $\bar{\lambda}_i$, 1/year*

Element \ Failure Type	Replacement	Major Repair	Minor Repair	No Cost Data
Blades	0.001	0.010	0.456	0.053
Contactor / Circuit Breaker / Relay	0.002	0.054	0.326	0.048
Controls	0.001	0.054	0.355	0.018
Electrical Components	0.002	0.016	0.358	0.059
Gearbox	0.154	0.038	0.395	0.046
Generator	0.095	0.321	0.485	0.098
Grease / Oil / Cooling Liq.	0	0.006	0.407	0.058
Heaters / Coolers	0	0.007	0.190	0.016
Hub	0.001	0.038	0.182	0.014
Other Components	0.001	0.042	0.812	0.150
Pitch / Hyd	0.001	0.179	0.824	0.072
Power Supply / Converter	0.005	0.081	0.076	0.018
Pumps/Motors	0	0.043	0.278	0.025
Safety	0	0.004	0.373	0.015
Sensors	0	0.070	0.247	0.029
Service Items	0	0.001	0.108	0.016
Tower / Foundation	0	0.089	0.092	0.004
Ransformer	0.001	0.003	0.052	0.009
Yaw System	0.001	0.006	0.162	0.020

Table 2: *Mean times to repair R_i, hours*

Element \ Failure Type	Replacement	Major Repair	Minor Repair	No Cost Data
Blades	288	21	9	28
Contactor / Circuit Breaker / Relay	150	19	4	5
Controls	12	14	8	17
Electrical Components	18	14	5	7
Gearbox	231	22	8	7
Generator	81	24	7	13
Grease / Oil / Cooling Liq.	–	18	4	3
Heaters / Coolers	–	14	5	5
Hub	298	40	10	8
Other Components	36	21	5	8
Pitch / Hyd	25	19	9	17
Power Supply / Converter	57	14	7	10
Pumps/Motors	–	10	4	7
Safety	–	2	2	2
Sensors	–	6	8	8
Service Items	–	2	7	9
Tower / Foundation	–	7	5	6
Ransformer	1	26	7	19
Yaw System	49	20	5	9

The failure rate of turbine subsystems should be as dependent on the wind speed as the failure rate of the whole turbine. Assuming that this dependence has the form (1), then it can be estimated using

$$\lambda_i(W) = \frac{\bar{\lambda}_i}{\sum\limits_k \bar{\lambda}_k}\lambda(W).$$

Let the equipment failures be independent. Then simultaneous failures of different elements are unlikely, they can be neglected. Therefore, the mean time to repair of the whole turbine can be calculated with the law of total probability:

$$R = \sum_k R_k p_k,$$

where p_k is the conditional probability of a failure of the k-th subsystem, if there is a failure of the whole turbine. These probabilities can be estimated as the multinomial distribution parameters:

$$p_k = \frac{\lambda_k(W)}{\sum\limits_i \lambda_i(W)} = \frac{\bar{\lambda}_k}{\sum\limits_i \bar{\lambda}_i}.$$

Remark. Assuming the failure flow is Poisson, i.e. the time between failures is exponentially distributed, then this result can be obtained more rigorously:

$$p_k = \mathrm{P}\{\xi_k < \xi_i, \forall i \neq k\} = \int\limits_0^\infty \prod_{i\neq k} \big(1 - F_i(t)\big)\, dF_k(t) = \int\limits_0^\infty \prod_{i\neq k} e^{-\lambda_i(W)t}\, \lambda_k(W) e^{-\lambda_k(W)t}\, dt = \frac{\bar{\lambda}_k}{\sum\limits_i \bar{\lambda}_i},$$

where ξ_j is the time between failures of the subsystem j, $F_j(t)$ is the cumulative distribution function of ξ_j.

In spite of the fact that the failure rates of the subsystems depend on wind speed, the probability p_k does not depend on it under assumption about equality of wind speed influence on all the equipment is true.

Thus, the mean time to repair of a turbine is

$$R = \frac{\sum\limits_k R_k \bar{\lambda}_k}{\sum\limits_i \bar{\lambda}_i} = 0.00153 \text{ years} = 13.5 \text{ hours}.$$

To estimate the mean time between failures, we will consider a wind turbine operation process, i.e. the number of work-repair cycles, as a renewal process. The mean number of failures to a point in time t is a renewal function $H(t)$, that, according elementary renewal theorem [12], is

$$H(t) \approx \frac{t}{T+R}.$$

If $t \to \infty$, then $\dfrac{H(t)}{t}$ is approximately equal to the failure rate $\lambda(W)$, so

$$T \approx \frac{1}{\lambda(W)} - R,$$

which implies

$$K(W) = 1 - \lambda(W)R = 1 - 0.000542\,W - 0.000133\,W^2. \tag{2}$$

3. Markov Modelling of Electricity Supply and Demand

We will use a continuous-time finite state Markov process to model a wind farm operation. Let consider a model with two states: in the state 0, the wind farm fully provides its consumers, and in state 1, there is not enough wind energy, it is necessary to use gas, coal, etc. Such model is correct, if the residence time of the process in each state is well approximated with the exponential law. To prof this assumption a large amount of statistical data was researched. We have used only open information sources, so it took some calculations to estimate the residence time distribution laws.

3.1. Estimation of Power Dependent on Wind Speed

Using experimental data from [13], a model of the dependence of the turbine power on the wind speed $P(W)$ in MW has been fitted with least squares given by (3) and represented at Fig. 2.

$$P(W) = 3.6\left(1 - e^{-\exp(-7.6+0.23W)W^{2.5}}\right). \tag{3}$$

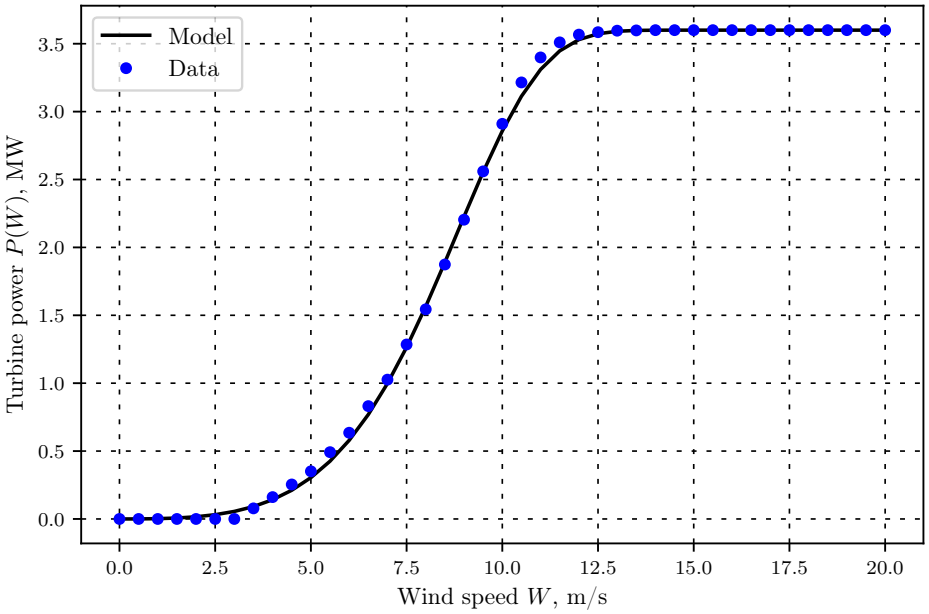


Figure 2: *Power Curve of the Turbine Siemens Gamesa Renewable Energy, SWT 3.6-120 [13]*

The wind speed is measured at the weather vane height, $h = 10$ m [14], and the power depends on the wind speed at the height of the turbine blades, i.e. at an altitude $z = 90$ m. To recalculate the wind speed, we will utilize the logarithmic model from [15]:

$$W = W_h \frac{\ln z - \ln z_0}{\ln h - \ln z_0}, \tag{4}$$

where $z_0 = 0.0002$ m is the roughness length [16]. Thus, knowing the wind speed at the height of the blades we can evaluate the power generated by a turbine.

The total generated power of the wind farm is equal to $N\,P(W)$, where N is a number of operable turbines, which may be less, than the nominal count $n = 111$ due equipment failures of some turbines. The random variable N has binomial distribution with a "success" probability $K(W)$, which may be calculated by the formula (2). To obtain a lower limit of the total generated power it is necessary to take a significance level $1 - \alpha$ and calculate a left quantile N_α. The number of the turbines is large enough, so the distribution of the variable N may be approximated by the normal law. Therefore, the low limit of the number of operable turbines is evaluated by

$$Z(W) = n\,K(W) + z_\alpha \sqrt{n\,K(W)\big(1 - K(W)\big)},$$

where z_α is a standard normal quantile. In this case, the real number of operable turbines will not be less $Z(W)$ with the probability $1 - \alpha$.

Thus, a lower limit of the power generated by the wind farm is

$$P_{Anholt}(W) = Z(W)P(W). \tag{5}$$

3.2. Met and Unmet Demand periods

The electricity generated by Anholt wind farm goes to the total network in Denmark, so it is impossible to point, what consumers receive this energy. That's why we have to evaluate the electricity demand from Anholt as a share of the national demand, assuming it corresponds to the share of the installed capacity of the Anholt wind farm ($Q_{Anholt} = 400$ MW) in the total installed capacity $Q_{Denmark}$ of all the power plants in Denmark. In open access, there is information about

the total installed capacity between 1990 and 2018 [17]. To predict the total installed capacity $Q_{Denmark}(y)$ for 2019 and 2020, a logarithmic model was used (Fig. 3):

$$Q_{Denmark}(y) = a_0 + a_1(y - 1989) + a_2 \ln(y - 1989),$$

where y is the time in years and the coefficients $a_0 = 8749.25$, $a_1 = 88.25$, $a_2 = 967.3$ are fitted by least squares.

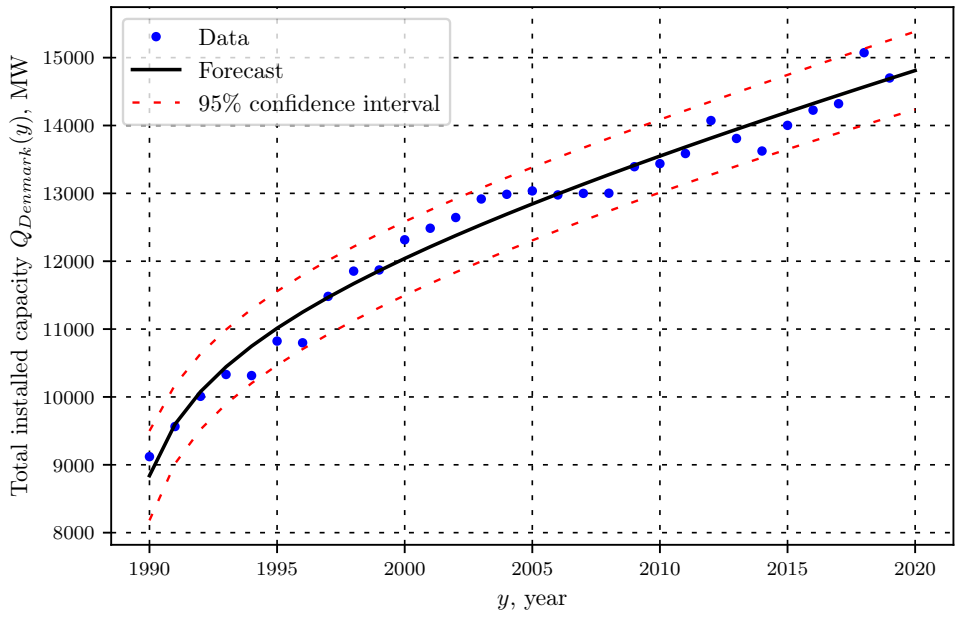


Figure 3: *Total Installed Electricity Capacity in Denmark*

According to this model, the total installed capacity was 14700 MW in 2019 and 14800 MW in 2020, so the Anholt share is evaluated as 2.72% and 2.70%, respectively.

Using hourly data on total electricity consumption in Denmark [18], we have obtained an evaluation of the electricity demand for the wind farm under study:

$$C_{Anholt}(t) = \frac{Q_{Anholt}}{Q_{Denmark}(y)} C_{Denmark}(t),$$

where t is the time in hours, $y = y(t)$ is the time in years, $C_{Denmark}(t)$ is the total electricity consumption in Denmark, MWh.

Using hourly data on wind speed in Anholt Island at the height $h = 10$ m we have obtained $W(t)$ – the wind speed at the height of the turbine blades for every hour t by the Eq. (4). Then by Eq. (5), we have calculated the power $P_{Anholt}(t) = P_{Anholt}(W(t))$ generated by all the turbines of the Anholt wind farm. A positive expression $P_{Anholt}(t) - C_{Anholt}(t)$ determines met demand periods, and a negative determines unmet ones. Fig. 4 shows a data fragment from March 28 to April 2, 2016.

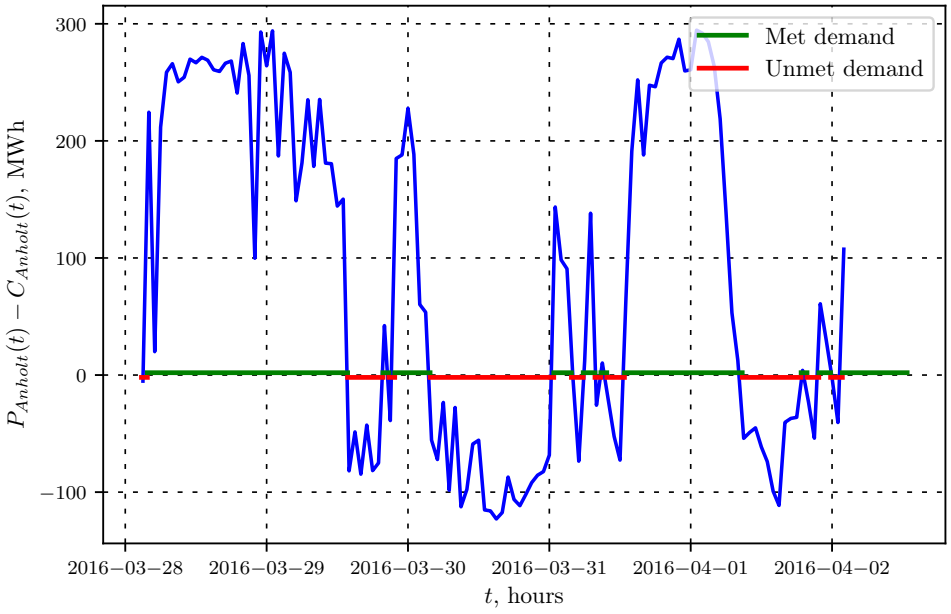


Figure 4: *Periods of Met and Unmet Demand*

The duration of the periods depends mainly on fluctuations in generated power, i.e. on wind speed. Although the demand fluctuations have 3 cycles (daily, weekly and seasonal), their amplitude is much less, than that of the power fluctuations. Fig. 5 shows a data fragment from January 9 to 21, 2016.

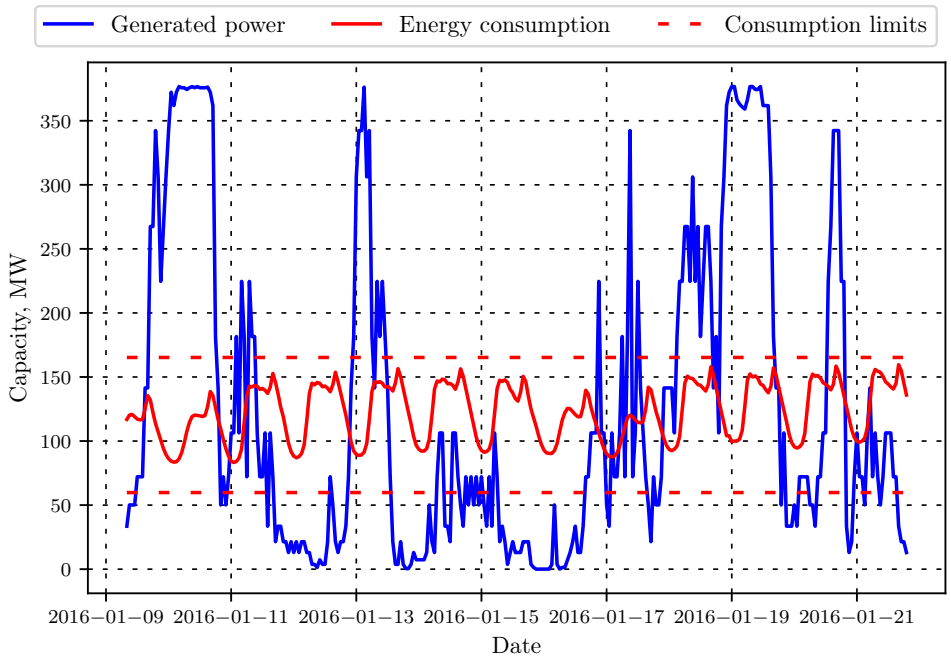


Figure 5: *Comparison of the Fluctuations' Amplitude in Supply and Demand*

After processing the data for 2016 – 2020, we have obtained two samples: the periods of met and unmet demand (the sample sizes are 2015 and 2016 observations, respectively). These measurements

are weakly correlated with each other (Fig. 6)), so we will consider them as independent observations

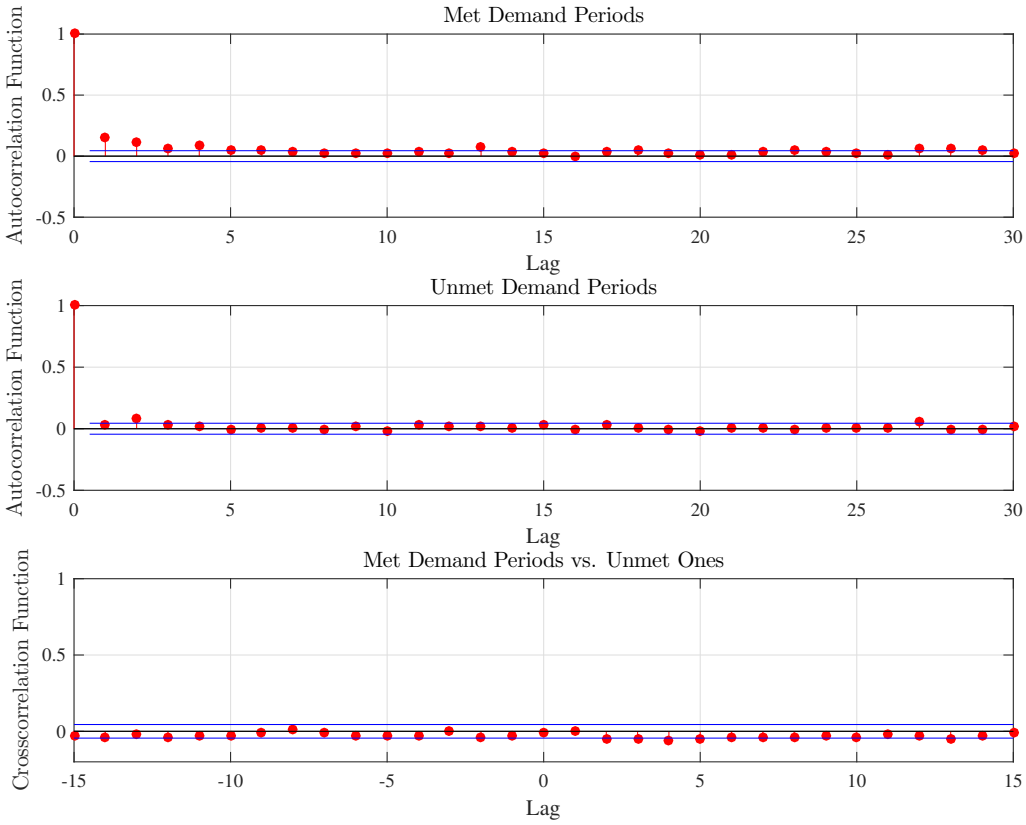


Figure 6: *Correlation Functions of the Periods*

The data may be described by the exponential mixture model, for which the probability density functions have the form

$$f(t) = \sum_{i=1}^{k} \frac{v_i}{\mu_i} e^{-\frac{t}{\mu_i}}.$$

The parameters μ_i are expectations of exponential distributions, from which the mixture model consists, and v_i are weight parameters of them. To estimate these parameters, the expectation-maximization algorithm (EM-algorithm) [19] was applied. It is a modification of the maximum likelihood method adapted for mixture distribution models. The model parameters for the met demand periods distribution are

$$\begin{aligned} \vec{v} &= [0.63,\ 0.19,\ 0.18]; \\ \vec{\mu} &= [2.2,\ 11.0,\ 43.6] \text{ hours}, \end{aligned}$$

and these ones for the unmet demand periods distribution are

$$\begin{aligned} \vec{v} &= [0.66,\ 0.34]; \\ \vec{\mu} &= [2.4,\ 24.0] \text{ hours}. \end{aligned}$$

Figures 7 and 8 show histograms and probability plots validating the models. The formal χ^2-test has not confirm goodness of fit of these models, because the samples are too large, and the test founds insignificant deviations of an empirical distribution from a hypothetical one. But if we reduce the sample size by 4 times (up to 500) by randomly discarding a part of the observations, the p-values will be 0.8 for met and 0.2 for unmet demand periods, that is, much greater than the significance level $\alpha = 0.05$.

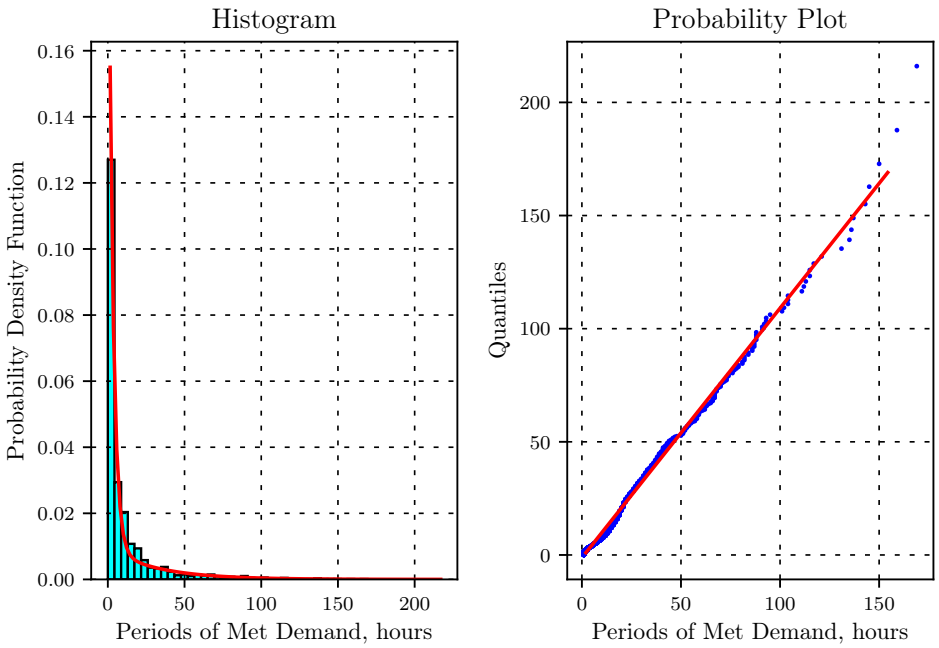


Figure 7: *Distribution Low of Met Demand Periods*

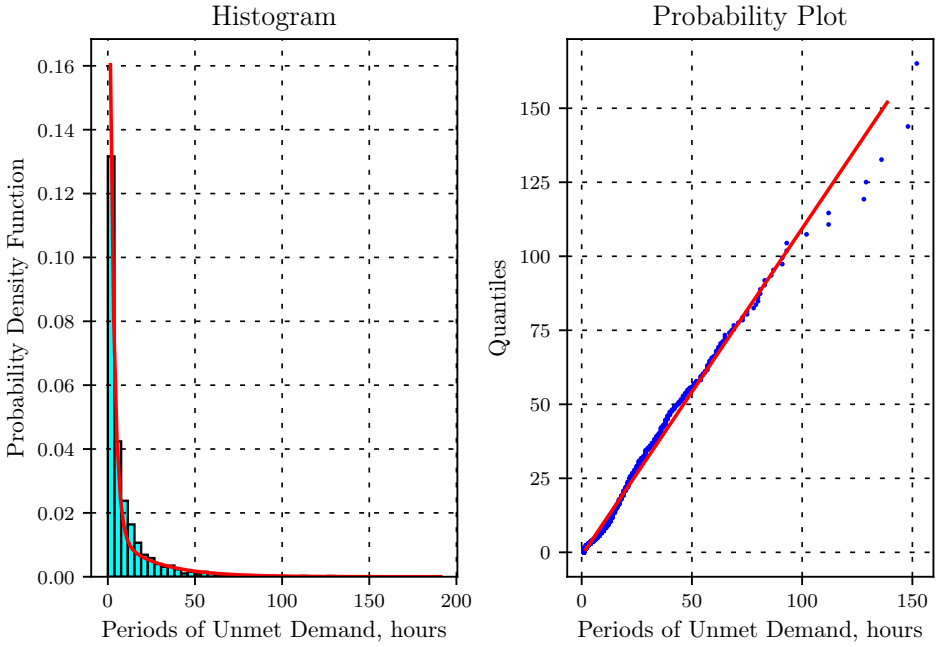


Figure 8: *Distribution Low of Unmet Demand Periods*

3.3. The Model

The exponential mixture model can be interpreted as follows: every real process state (0 – for met demand and 1 – for unmet one) consists from several fictive (3 or 2, respectively) states. The residence time in each of them has an exponential distribution. Let's number the fictive states: 0, 1 and 2 are the numbers of met demand states (shown by green vertices on the transition graph

in Fig. 9), 3 and 4 are the numbers of unmet demand states (shown by red vertices on the transition graph in Fig. 9).

The transition rates are inversely proportional to the mean residence times in the states from which the process leaves, and are directly proportional to the weight coefficients of the states into which the process comes:

$$\lambda_{ij} = \frac{v_j}{\mu_i}, \quad i, j = \overline{0, 4}.$$

Let's write the transition rates estimated by the data in the matrix form:

$$\Lambda = [\lambda_{ij}] = \begin{bmatrix} 0 & 0 & 0 & 0.30 & 0.15 \\ 0 & 0 & 0 & 0.06 & 0.03 \\ 0 & 0 & 0 & 0.02 & 0.01 \\ 0.26 & 0.08 & 0.07 & 0 & 0 \\ 0.03 & 0.01 & 0.01 & 0 & 0 \end{bmatrix}$$

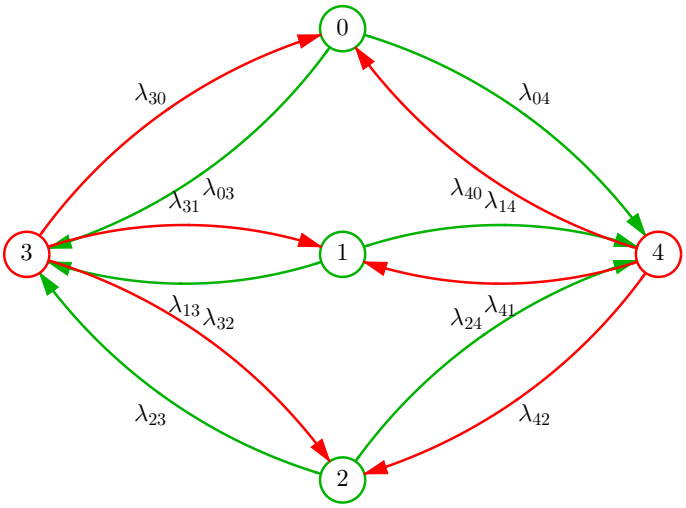


Figure 9: *Transition Graph of the Markov Model*

Because the residence time of every fictive states has approximately an exponential distribution, the process of transitions between them can be modelling as a Markov one. Therefore, the vector of the state probabilities $\vec{p}(t)$ is the solution of Kolmogorov system of equations [20]:

$$\frac{d\vec{p}}{dt} = \Lambda' \vec{p}, \quad \sum_{i=0}^{4} p_i(t) = 1. \tag{6}$$

A detailed derivation of this equation one can be find in numerous textbooks, for example in [21].

3.4. Model Investigation

To obtain an unique solution, it is necessary to set an initial condition by choosing one of the fictive states as the starting. With operational control at some time point t, we can attribute the last completed period to one of the fictitious states, because the EM-algorithm allows both estimating the parameters of a mixture of distributions and classifying observations.

The analytical solution of the system (6) is very cumbersome and is not presented here. It is recommended to apply numerical methods for solving the system (6), for example an explicit Runge-Kutta method.

The probabilities of the real states (0 – the met demand, 1 – unmet demand) are the sums of the fictive state probabilities:

$$\begin{aligned} p_0^*(t) &= \mathbf{P}\{P_{Anholt} \geq C_{Anholt}\} = p_0(t) + p_1(t) + p_2(t), \\ p_1^*(t) &= \mathbf{P}\{P_{Anholt} < C_{Anholt}\} = p_3(t) + p_4(t). \end{aligned}$$

Fig. 10 shows the plot of the met demand probability for various initial states. Green curves are for met demand initial states, red ones are for unmet demand initial states.

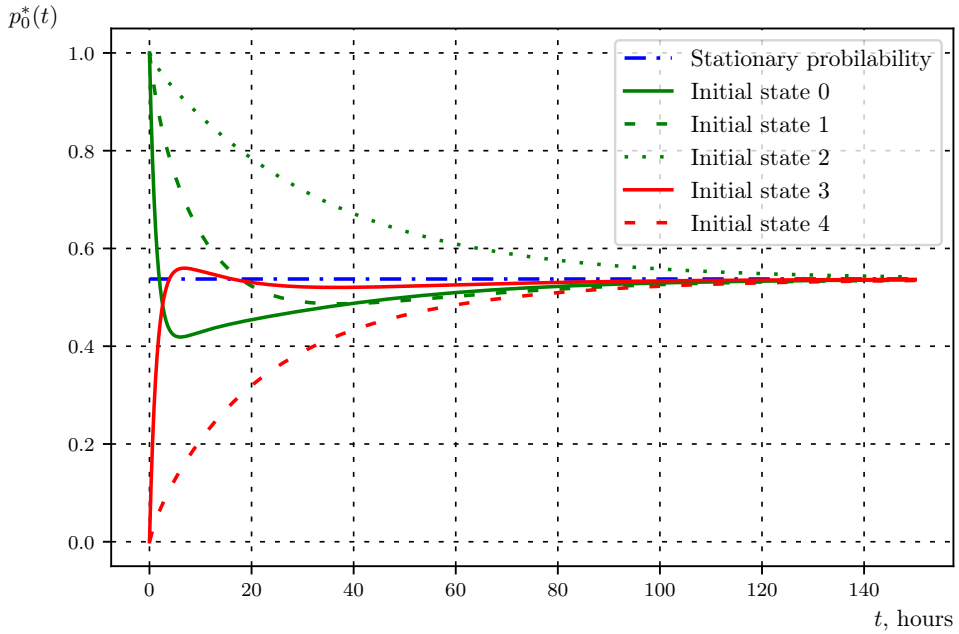


Figure 10: *Met Demand Probabilities for Various Initial States*

The stationary state probabilities $\vec{p} = \lim\limits_{t\to\infty} \vec{p}(t)$ can be obtain form the system (6) by setting

$$\frac{d\vec{p}}{dt} = \vec{0}, \text{ i.e. } \Lambda'\vec{p} = \vec{0}.$$

This is a system of linear algebraic equations. The rank of the system matrix is 4, which is 1 less than the number of states, but with the normalisation condition $\sum p_i = 1$, the system has the full rank and its unique solution is presented in Table 3.

Table 3: *Stationary Probabilities of the Process States*

i	0	1	2	3	4
p_i	0.066	0.099	0.373	0.075	0.387

The process converges to the stationary one, but does not it very quickly: only in 120 hours after a start time point, the state probabilities differ from its stationary values less than 0.01.

The stationary probability of the met demand does not depend on an initial state and is

$$p_0^* = p_0 + p_1 + p_2 \approx 0.537.$$

This value differs little from the total share of the met demand time for the studied period, which indicate the adequacy of the constructed model:

$$\frac{T_0^*}{T_0^* + T_1^*} \approx 0.544,$$

where T_0^* is the total time of the met demand, T_1^* is the total time of the unmet demand.

Thus, the power of the wind farm is insufficient to meet the electricity demand on average 54% of the time. At the rest of the time consumers have to use other energy sources additionally. At

the same time, the amount of energy generated during the periods of met demand is much greater than the required one. Although an industry technology for long-term storage of electricity, for example in the hydrogen form, has not yet been developed by now, investigations in this direction are being actively pursued. Conservation of energy during the period when it is generated in excess would make it possible to cover its deficit during periods of weak wind.

4. Conclusion

The electricity provision to consumers is determined both by the reliability of wind farm equipment and by weather conditions. This paper presents a mathematical model that makes it possible to assess the reliability indicators, such as failure rate, mean time to repair, and availability, based on the statistical data. The assessment took into account the dependence of the indicators on wind speed.

Based on statistical data and the already known mathematical models, the distributions of the met and unmet demand periods have been investigated. Using hourly data over 5-years period, we have found out that these distributions can be approximated by exponential mixture model, and, therefore, the process of the wind farm operation can modelled as a Markov process.

Thus, the model of the wind farm operating process taking into account both the random nature of wind speed as an energy source and usual failures of equipment is built. The main result of the study is the estimation of stationary probability of the met demand: it is approximately 0.537. It means that the power of the wind farm is insufficient to meet the electricity demand on average 54% of the time.

The amount of energy generated during the periods of met demand is much greater than the required one. Therefore, it is profitable to build an energy storage to save the energy excess. The author intends to study a reliability problem about a rational storage volume.

5. Acknowledgements

The author is grateful to Irina Vdovina, who found important statistical data used in this paper and performed some preliminary calculations as part of her master's thesis.

References

[1] Gaudenz, K. and Göran, A. (2006). The influence of combined power, gas, and thermal networks on the reliability of supply. The Sixth World Energy System Conference, Torino, Italy, July 10-12.

[2] Chen, Sh., Sun, G., Wei, Zh., Chen, Sh. and Li, Y. (2016). Multi-time Combined Gas and Electric System Optimal Power Flow Incorporating Wind Power. Energy Procedia. 100. 111-116. doi:10.1016/j.egypro.2016.10.147.

[3] Kim, H., Singh, C. and Sprintson, A. (2012). Simulation and Estimation of Reliability in a Wind Farm Considering the Wake Effect. IEEE Transactions on Sustainable Energy, 3, PP. 274–282.

[4] Nath, A., Paul, S., Rather, Z. and Mahapatra, S. (2019). Estimation of Offshore Wind Farm Reliability Considering Wake Effect and Wind Turbine Failure. 3866-3871. doi:10.1109/ISGT-Asia.2019.8880887.

[5] Nguyen, N. and Almasabi, S. and Mitra, J. (2019). Impact of Correlation Between Wind Speed and Turbine Availability on Wind Farm Reliability. IEEE Ransactions on Industry Applications. PP. 1-1. doi:10.1109/TIA.2019.2896152.

[6] Fazio, A.R., Russo, M. (2009). Wind farm modelling for reliability assessment. Renewable Power Generation, IET. 2. PP. 239 – 248. doi:10.1049/iet-rpg:20080005.

[7] Goldaev, S.V. and Radyuk, K.N. (2015). Raschet proizvoditel'nosti vetroenergeticheskoj ustanovki bol'shoj moshchnosti po usovershenstvovannoj metodike [Calculation of the ca-

pacity of a high power wind power plant using an improved procedure]. Izvestiya Tomskogo politekhnicheskogo universiteta. Inzhiniring georesursov. Vol. 326. № 8. (In Russian)

[8] Rakhmanov, N.R., Kurbatski, V.G., Guliev, G.B. and Tomin, N.V. (2015) Kratkosrochnoe prognozirovanie vyrabotki moshchnosti vetroenergeticheskih ustanovok dlya obespecheniya nadyozhnosti elektricheskih setej [Short-term forecasting of power generationwind power plants to ensure the reliability of electrical networks]. Metodicheskie voprosy issledovaniya nadezhnosti bol'shih sistem energetiki. Vypusk 66. Aktual'nye problemy nadezhnosti sistem energetiki [Methodological issues of reliability research large energy systems. Issue 66. Actual problems of reliability of energy systems]. Minsk, BITU, PP. 397-403 (In Russian).

[9] Kurbatsky, V.G. and Tomin, N.V. (2009). Use of the "ANAPRO" software to analyze and forecast operating parameters and technological characteristics on the basis of macro applications. 1 – 8. doi:10.1109/PTC.2009.5281905.

[10] Carroll J., McDonald A. and McMillian D. (2014). Reliability comparison of wind turbines with DFIG and PMG drive trains. IEEE Rans. Energy Convers., vol. PP, pp. 1–8, Dec.

[11] Wilson, G. and Mcmillan, D. (2014). Assessing Wind Farm Reliability Using Weather Dependent Failure Rates. Journal of Physics: Conference Series. 524. 012181. doi:10.1088/1742-6596/524/1/012181.

[12] Sevastýanov, B.A. (1975). Renewal theory. J. Math. Sci. 4, PP. 281—302. doi:10.1007/BF01097185

[13] The wind power. Wind Energy Market Intelligence. [online]. URL: https://www.thewindpower.net/turbine_en_79_siemens_swt-3.6-120.php. [Accessed 1 November 2021].

[14] National Centers for Environmental Information.[online] Data URL: https://www.ncei.noaa.gov/access/search/data-search/global-hourly. Description URL: https://www.ncei.noaa.gov/data/global-hourly/doc/isd-format-document.pdf. [Accessed 1 November 2021].

[15] Ryhlov, A.B. (2011) Ocenka parametrov zakonov izmeneniya srednej skorosti vetra s vysotoj v prizemnom sloe atmosfery na yugo-vostoke evropejskoj chasti Rossii dlya resheniya zadach vetroenergetiki [Estimation of the parameters of the laws of variation of the average wind speed with height in the surface layer of the atmosphere in the southeast of the European part of Russia for solving the problems of wind power]. Izv. Sarat. un-ta Nov. ser. Ser. Nauki o Zemle, №2.

[16] Mlyavaya, G.V. (2014). Vliyanie parametrov sherohovatosti podstilayushchej poverhnosti na skorost' vetra [Influence of the roughness length of an underlying surface on wind speed]. Ecologia i Geografia. №2 (323). PP. 181 -– 187.

[17] Un data. A world of information.[online]. URL: https://data.un.org. [Accessed 1 November 2021].

[18] Energi data service. Production and Consumption – Settlement. [online] URL: https://www.energidataservice.dk/tso-electricity/productionconsumptionsettlement. [Accessed 1 November 2021].

[19] Sundberg, R. (1976). An iterative method for solution of the likelihood equations for incomplete data from exponential families. Communications in Statistics – Simulation and Computation. 5 (1): 55–64. doi:10.1080/03610917608812007.

[20] Kolmogoroff, A. (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. 104, 415–458 . https://doi.org/10.1007/BF01457949 (In German).

[21] Rykov V.V., Itkin V.Yu. (2016) Nadyozhnost' tekhnicheskih sistem i tekhnogennyj risk (Reliability of technical systems and technogenic risk). Moscow: INFRA-M. 192 P. (In Russian).

Estimation procedures for a flexible extension of Maxwell distribution with data modeling

Abhimanyu Singh Yadav*[1], H. S. Bakouch[2], S. K. Singh[3] and Umesh Singh[4]

•

[1]Department of Statistics, Banaras Hindu University, Varanasi, India.
E-mail: [1]abhistats@bhu.ac.in, [3]singhsk64@gmail.com
[4]umeshsingh52@gmail.com
[2]Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt.
E-mail: [2]hassan.bakouch@science.tanta.edu.eg
*Corresponding Author

Abstract

In this paper, we introduce a flexible extension of the Maxwell distribution for modeling various practical data with non-monotone failure rate. Some main properties of this distribution are obtained, and then the estimation of the parameters for the proposed distribution has been addressed by maximum likelihood estimation method and Bayes estimation method. The Bayes estimators have been obtained under gamma prior using squared error loss function. Also, a simulation study is gained to assess the estimates performance. A real-life applications for the proposed distribution have been illustrated through different lifetime data.

Keywords: Family of Maxwell distributions, Entropy, Classical and Bayes estimation, Interval estimation, Asymptotic confidence length.

1. Introduction

The Maxwell distribution has broad application in statistical physics, physical chemistry, and their related areas. Besides Physics and Chemistry it has also a good number of applications in reliability theory. At first, the Maxwell distribution was used as lifetime distribution by [1]. The inferences based on generalized Maxwell distribution have been discussed by [2]. [3] considered the estimation of reliability characteristics for Maxwell distribution under Bayes paradigm. [4] discussed the prior selection procedure in case of Maxwell distribution. [5] studied the distributions of the product $|XY|$ and ratio $|X/Y|$, where X and Y are independent random variables having the Maxwell and Rayleigh distributions, respectively. [6] proposed the Bayesian estimation of the Maxwell parameters. [7] discussed the estimation procedure for the Maxwell parameters under progressive type-I hybrid censored data. Furthermore, several generalizations based on Maxwell distribution are advocated and statistically justified. Recently, two more extensions of Maxwell distribution has been introduced by [8], [9] and discussed the classical as well as Bayesian estimation of the parameter along with real-life applications.

A random variable Z follows the Maxwell distribution (MaD) with scale parameter α, denoted as $Z \sim MaD(\alpha)$, if its probability density function (PDF) and cumulative distribution function (CDF) are given by

$$f(z,\alpha) = \frac{4}{\sqrt{\pi}}\alpha^{\frac{3}{2}}z^2e^{-\alpha z^2} \quad z \geq 0, \alpha > 0 \tag{1}$$

and

$$F(z,\alpha) = \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2},\alpha z^2\right), \tag{2}$$

respectively, where $\Gamma(a,z) = \int_0^z p^{a-1}e^{-p}dp$ is the incomplete gamma function.

In this article, we propose a flexible extension of the Maxwell distribution. The objective of this article is to get some main properties of this distribution for showing its merit in modeling various practical data, and then estimate the unknown parameters using classical and Bayes estimation methods. Other motivations regarding the advantages of the distribution comes from its flexibility to model the data with non-monotone failure rates. The former aim is justified, where the proposed distribution provides better fit to the reliability/survival data comparing to the some known and recent versions of the Maxwell distribution. Further, the distribution is that having the nature of platykurtic, mesokurtic and leptokurtic, hence it can be used to model skewed and symmetric data as well. Also, the Bayes procedure under informative prior provides the more efficient estimates as compared to the maximum likelihood estimates (MLEs) concerning the estimation point of view. Another motivation for the confidence interval of the distribution parameters is that increasing the sample size decreases the width of confidence intervals, because it decreases the standard error, and this justified by simulation study and using sizes of four practical data sets.

The reminder of the considered work has been structured in the following manner. Section 2 provides some statistical properties related to the proposed model for purpose of data modeling. In Section 3, some types of entropy are investigated. The maximum likelihood (ML) and Bayes estimation procedures have been discussed in Section 4. Also, a simulation study is carried out to compare the performance of Bayes estimates with MLEs. In Section 5, we illustrate the application and usefulness of the proposed model by applying it to four practical data sets. Section 5 offers some concluding remarks.

2. The model and some of its properties

This section provides another generalization of the MaD using power transformation of Maxwell random variates for estimations issues of the distribution parameters and modeling practical data. For this purpose, consider the transformation $X = Z^{\frac{1}{\beta}}$, where $Z \sim MaD(\alpha)$, hence the resulting distribution of X is called as power Maxwell distribution (for short PMaD) and denoted by $X \sim PMaD(\alpha,\beta)$, where, α and β are the scale and shape parameters, respectively. The PDF and CDF of the PMaD are given by

$$f(x,\alpha,\beta) = \frac{4}{\sqrt{\pi}}\alpha^{\frac{3}{2}}\beta x^{3\beta-1}e^{-\alpha x^{2\beta}}, \quad x \geq 0, \alpha, \beta > 0, \tag{3}$$

$$F(x,\alpha,\beta) = \frac{2}{\sqrt{\pi}}\,\Gamma\left(\frac{3}{2},\alpha x^{2\beta}\right), \tag{4}$$

respectively. Plots of the PDF are given by Figure 1 for different choices of α and β. The plots show different kurtosis, positive skewness and symmetric shapes.

Some main mathematical and statistical properties of PMaD have been obtained in the following.

2.1. Behaviour with some reliability functions

This subsection, described the asymptotic nature of density and survival functions for the proposed distribution. To illustrate asymptotic behaviour, at first, we will show that $\lim_{x\to 0} f(x,\alpha,\beta) = 0$ and $\lim_{x\to\infty} f(x,\alpha,\beta) = 0$. Therefore, using (2.1)

$$\lim_{x\to 0} f(x,\alpha,\beta) = \frac{4}{\sqrt{\pi}}\alpha^{\frac{3}{2}}\beta \lim_{x\to 0} x^{3\beta-1}e^{-\alpha x^{2\beta}} = 0,$$

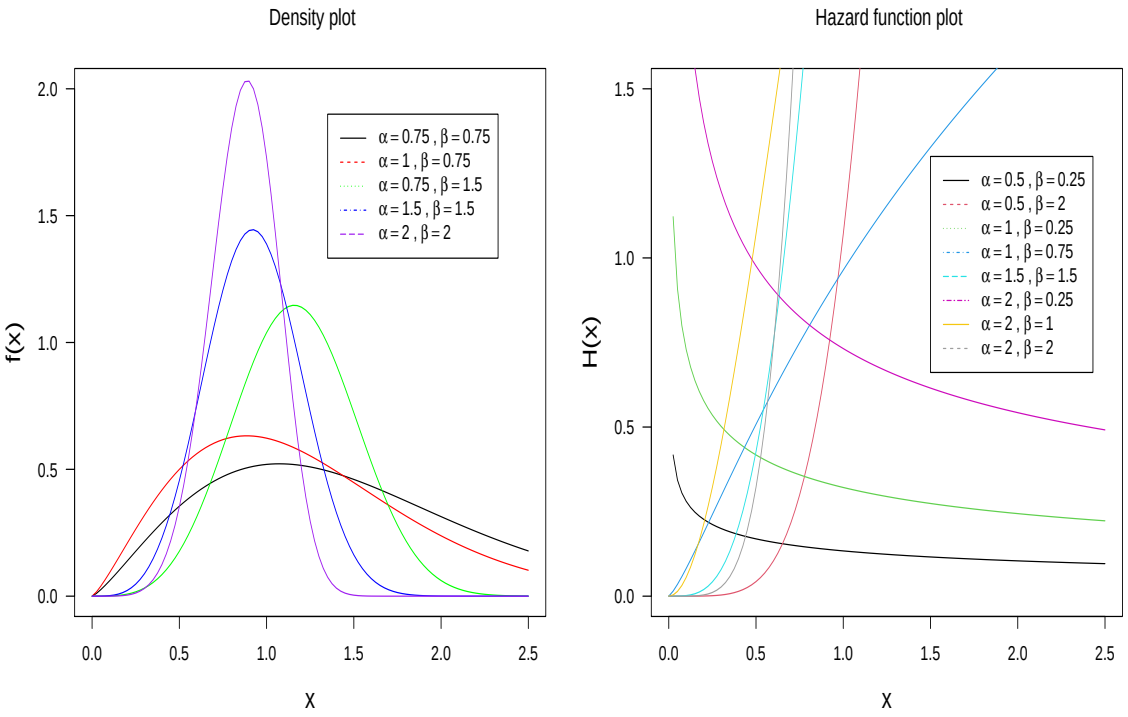


Figure 1: *Density function and hazard function plot for different choices of α and β.*

and

$$\lim_{x\to\infty} f(x,\alpha,\beta) = \frac{4}{\sqrt{\pi}}\alpha^{\frac{3}{2}}\beta \lim_{x\to\infty} x^{3\beta-1} \lim_{x\to\infty} e^{-\alpha x^{2\beta}} = 0$$

The characteristics based on reliability function and hazard function are very useful to study the pattern of any lifetime phenomenon. Let X be a random variable with PDF (2.1) and CDF (2.2), different reliability measures for the proposed distribution are obtained by following equations.

The reliability function $R(x)$ is given by

$$R(x) = P(X > x) = 1 - \frac{2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}, \alpha x^{2\beta}\right) \tag{5}$$

The mean time to system failure $M(x)$ is

$$M(x) = E(x) = \frac{2}{\sqrt{\pi}}\left(\frac{1}{\alpha}\right)^{\frac{1}{2\beta}}\Gamma\left(\frac{3\beta+1}{2\beta}\right) \tag{6}$$

The hazard function $H(x)$ is given as

$$H(x) = \frac{f(x,\alpha,\beta)}{1-F(x,\alpha,\beta)} = \frac{4\alpha^{\frac{3}{2}}\beta x^{3\beta-1}e^{-\alpha x^{2\beta}}}{\sqrt{\pi} - 2\Gamma\left(\frac{3}{2}, \alpha x^{2\beta}\right)} \tag{7}$$

The plots, in Figure 1, show that the proposed density is unimodel and positively skewed with monotone failure rate function for the different combination of the model parameters. The comparative behavior of the random variables can be measured by stochastic ordering concept

that is summarized in the next proposition.

Proposition: Let $X \sim PMaD(\alpha_1, \beta_1)$ and $Y \sim PMaD(\alpha_2, \beta_2)$, then the likelihood ratio is

$$\Phi = \frac{f_X(x)}{f_Y(x)} = \left(\frac{\alpha_1}{\alpha_2}\right)^{\frac{3}{2}} \left(\frac{\beta_1}{\beta_2}\right) x^{3(\beta_1-\beta_2)} e^{-(\alpha_1 x^{2\beta_1} + \alpha_2 x^{2\beta_2})}.$$

Therefore,

$$\Phi' = \log\left(\frac{f_X(x)}{f_Y(x)}\right) = \frac{1}{x}\left[3(\beta_1-\beta_2) - (\alpha_1 x^{2\beta_1} + \alpha_2 x^{2\beta_2})\right]$$

If $\beta_1 = \beta_2 = \beta$, then $\Phi' < 0$, which implies that the random variable X is a likelihood ratio order than Y, that is $X \leq_{lr} Y$. Also, if $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 < \beta_2$, then again $\Phi' < 0$, which shows that $X \leq_{lr} Y$. Other stochastic orderings behaviour follow using $X \leq_{lr} Y$, such as hazard rate order $(X \leq_{hr} Y)$, mean residual life order $(X \leq_{mrl} Y)$ and scholastically greater $(X \leq_{st} Y)$.

2.2. Moments and some conditional ones

Let $x_1, x_2, \cdots x_n$ be random observations from the PMaD(α, β). The r^{th} moment, μ_r', about origin is

$$\mu_r' = \int_{x=0}^{\infty} x^r f(x, \alpha, \beta) \quad dx = \frac{2}{\sqrt{\pi}} \left(\frac{1}{\alpha}\right)^{\frac{r}{2\beta}} \Gamma\left(\frac{3\beta + r}{2\beta}\right), \quad r \geqslant 1.$$

The coefficient of skewness and kurtosis measure the convexity of the curve and its shape. Using the moments above, the two earlier measures are obtained by moments based relations suggested by Pearson and given by

$$\beta_1 = \frac{\left[\mu_3' - 3\mu_2'\mu_1' + 2\left(\mu_1'\right)^3\right]^2}{\left[\mu_2' - \left(\mu_1'\right)^2\right]^3}$$

and

$$\beta_2 = \frac{\mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\left(\mu_1'\right)^2 - 3\left(\mu_1'\right)^4}{\left[\mu_2' - \left(\mu_1'\right)^2\right]^2}.$$

Numerical values of some measures above are calculated in Table 1 for different combination of the model parameters, and it is observed that the shape of the PMaD is right skewed and almost symmetrical for some choices of α, β. Also, it can has the nature of platykurtic, mesokurtic and leptokurtic, thus PMaD may be used to model skewed and symmetric data as well.

The mode (M_0) for PMaD (α, β) is obtained by solving the following expression $\frac{d}{dx} f(x, \alpha, \beta)|_{M_0} = 0$, which yields

$$M_0 = \left(\frac{3\beta - 1}{2\alpha\beta}\right)^{\frac{1}{2\beta}}.$$

Moreover, the median (M_d) of the proposed distribution can be calculated by using the empirical relation among the mean, median and mode. Thus, the median is,

$$M_d = \frac{1}{3}M_0 + \frac{2}{3}\mu_1' = \frac{1}{3}\left[\left(\frac{3\beta - 1}{2\alpha\beta}\right)^{\frac{1}{2\beta}} + \frac{4}{\sqrt{\pi}}\left(\frac{1}{\alpha}\right)^{\frac{1}{2\beta}} \Gamma\left(\frac{3\beta + 1}{2\beta}\right)\right].$$

The moment generating function (mgf) $M_X(t)$ for a PMaD random variable X is obtained as

$$M_X(t) = E(e^{tx}) = \frac{2}{\sqrt{\pi}} \sum_{i=0}^{\infty} \frac{1}{j!} \left(\frac{t}{\alpha^{2\beta}}\right)^r \Gamma\left(\frac{3\beta + r}{2\beta}\right).$$

Table 1: *Values of mean, variance, skewness, kurtosis, mode and coefficient of variation for different α, β*

α, β	$\mu_1^{'}$	μ_2	β_1	β_2	x_0	CV
when α fixed and β varying						
0.5, 0.5	3.0008	5.9992	2.6675	7.0010	1.0000	0.8162
0.5, 1.0	1.5962	0.4530	0.2384	3.1071	1.4142	0.4217
0.5, 1.5	1.3376	0.1499	0.0102	2.7882	1.3264	0.2894
0.5, 2.5	1.1780	0.0445	0.0481	2.7890	1.2106	0.1792
0.5, 3.5	1.1204	0.0211	0.1037	2.4351	1.1533	0.1298
when β fixed α varying						
0.5, 0.75	1.9392	1.1443	0.7425	3.8789	1.4057	0.5516
1.0, 0.75	1.2216	0.4541	0.7425	3.8789	0.8855	0.5516
1.5, 0.75	0.9323	0.2645	0.7425	3.8789	0.6758	0.5516
2.5, 0.75	0.6632	0.1338	0.7425	3.8789	0.4807	0.5516
3.5, 0.75	0.5299	0.0855	0.7425	3.8789	0.3841	0.5516
when both varying						
1, 1	1.1287	0.2265	0.2384	3.1071	1.0000	0.4217
2, 2	0.8723	0.0372	0.0102	2.7895	0.8891	0.2212
3, 3	0.8484	0.0163	0.0831	2.6907	0.8736	0.1506
4, 4	0.8509	0.0094	0.1069	1.9643	0.8750	0.1140
5, 5	0.8586	0.0062	0.0677	0.1072	0.8805	0.0915

For lifetime distributions, the conditional moments are of interest in prediction. Another application of conditional moments is the mean residual life (MRL). For this purpose, let X observed from PDF(2.1), the conditional moments, $E(X^r|X>k)$ and the conditional mgf $E(e^{tx}|X>k)$ are obtained as follows;

$$E(X^r|X>k)=\frac{\int_{x>k}x^r f(x,\alpha,\beta)dx}{\int_{x>k}f(x,\alpha,\beta)dx}=\frac{2\left(\frac{1}{\alpha}\right)^{\frac{r}{2\beta}}\Gamma\left(\frac{3\beta+r}{2\beta},\alpha k^{2\beta}\right)}{\sqrt{\pi}-2\Gamma\left(\frac{3}{2},\alpha k^{2\beta}\right)}$$

and

$$\begin{aligned}E(e^{tx}|X>k)&=\frac{\int_{x>k}e^{tx}f(x,\alpha,\beta)dx}{\int_{x>k}f(x,\alpha,\beta)dx}\\&=\frac{2\sum_{i=0}^{\infty}\frac{t^i}{i!}\left(\frac{1}{\alpha}\right)^{\frac{r}{2\beta}}\Gamma\left(\frac{3\beta+r}{2\beta},\alpha k^{2\beta}\right)}{\sqrt{\pi}-2\Gamma\left(\frac{3}{2},\alpha k^{2\beta}\right)},\end{aligned}$$

respectively. The MRL is the expected remaining life $X-x$, given that the equipment has survived to time k. The MRL function in terms of the first conditional moments is given as

$$m(x)=E[X-x|X>k]=\frac{2\left(\frac{1}{\alpha}\right)^{\frac{1}{2\beta}}\Gamma\left(\frac{3\beta+1}{2\beta},\alpha k^{2\beta}\right)}{\sqrt{\pi}-2\Gamma\left(\frac{3}{2},\alpha k^{2\beta}\right)}-x$$

3. Entropy measurements

In information theory, entropy measurement plays a vital role to study the uncertainty associated with the random variable. In this section, we discuss the different entropy measures for $PMaD$. For more detail about entropy measurement, see [10].

3.1. Renyi entropy

Renyi entropy of a r.v. X with PDF (2.1) is given as

$$R_E = \frac{1}{(1-\in)} \ln \left[\int_{x=0}^{\infty} \left\{ \frac{4}{\sqrt{\pi}} \alpha^{\frac{3}{2}} \beta x^{3\beta-1} e^{-\alpha x^{2\beta}} \right\}^{\in} dx \right]$$

Hence, after some algebra, we get

$$R_E = \frac{1}{(1-\in)} \left[\lambda \ln 4 - \frac{\lambda}{2} \ln \pi + \lambda \ln \beta - \frac{1-\lambda-2\beta}{2\beta} \ln \alpha - \frac{3\lambda\beta-\lambda+1}{2\beta} \ln \lambda + \ln \left(\frac{3\beta\lambda-\lambda+1}{2\beta} \right) \right].$$

3.2. Δ-entropy

The Δ entropy is also known as β entropy. The Δ entropy for a random variable X having PDF (2.1) is defined as

$$\Delta_E = \frac{1}{\Delta - 1} \left[1 - \int_{x=0}^{\infty} f^{\Delta}(x, \alpha, \beta) dx \right].$$

Using PDF (2.1) and after simplification, the expression for β-entropy is given by;

$$\Delta_E = \frac{1}{\Delta - 1} \left[1 - \left(\frac{4}{\sqrt{\pi}} \right)^{\Delta} \beta^{\Delta} \left(\frac{1}{\alpha} \right)^{\frac{1-\Delta-2\beta}{2\beta}} \left(\frac{\Gamma\left(\frac{3\Delta\beta-\Delta+1}{2\beta} \right)}{\Delta^{\frac{3\Delta\beta-\Delta+1}{2\beta}}} \right) \right]. \tag{8}$$

3.3. Generalized entropy

The generalized entropy is defined by

$$G_E = \frac{\nu_\lambda \mu^{-\lambda} - 1}{\lambda(\lambda - 1)} \quad ; \lambda \neq 0, 1,$$

where, $\nu_\lambda = \int_{x=0}^{\infty} x^\lambda f(x, \alpha, \theta) dx$ and $\mu = E(X)$. After some algebra, we get

$$G_E = \left(\frac{4}{\pi} \right)^{\frac{1-\lambda}{2}} \left[\frac{\Gamma\left(\frac{3\beta+\lambda}{2\beta} \right) \left\{ \Gamma\left(\frac{3\beta+1}{2\beta} \right) \right\}^{-\lambda}}{\lambda(\lambda-1)} \right], \quad \lambda \neq 0, 1. \tag{9}$$

4. Parameter estimation with a simulation study

Here, we describe the maximum likelihood estimation method and Bayes estimation method for estimating the unknown parameters α, β of the PMaD. The estimators obtained under these methods are not in nice closed form; thus, numerical approximation techniques are used to get the solution. Further, the performances of these estimators are studied through a Monte Carlo simulation.

4.1. Maximum likelihood estimation

The most popular and efficient method of classical estimation of the parameter(s) is maximum likelihood estimation. The estimators obtained by this method passes several desirable properties

such as consistency, efficiency etc. Let $X_1, X_2, \cdots, X_n$ be an iid random sample of size n taken from PMaD (α, β), then the likelihood function is

$$L(\alpha,\theta) = \prod_{i=1}^{n} \frac{4}{\sqrt{\pi}} \alpha^{\frac{3}{2}} \beta x_i^{3\beta-1} e^{-\alpha x_i^{2\beta}} = \frac{4^n}{\pi^{n/2}} \alpha^{\frac{3n}{2}} \beta^n e^{-\alpha \sum_{i=1}^{n} x_i^{2\beta}} \left(\prod_{i=1}^{n} x_i^{3\beta-1} \right),$$

hence the corresponding log-likelihood function is written as

$$\ln L(\alpha,\theta) = l = n \ln 4 - \frac{n}{2} \ln \pi + \frac{3n}{2} \ln \alpha + n \ln \beta - \alpha \sum_{i=1}^{n} x_i^{2\beta} + (3\beta - 1) \sum_{i=1}^{n} \ln x_i. \tag{10}$$

The MLEs of α and β are the solution of $\dfrac{\partial l}{\partial \alpha} = 0$ and $\dfrac{\partial l}{\partial \beta} = 0$, hence

$$\frac{3n}{2\alpha} - \sum_{i=1}^{n} x_i^{2\beta} = 0 \tag{11}$$

$$\frac{n}{\beta} - 2\alpha \sum_{i=1}^{n} x_i^{2\beta} \ln x_i + 3 \sum_{i=1}^{n} \ln x_i = 0. \tag{12}$$

The MLEs of the parameters are obtained by solving the two equations above simultaneously, and non-linear maximization techniques is used to get the solution.

4.1.1 Uniqueness of MLEs

The uniqueness of the MLEs discussed in the previous section can be checked by using following propositions.

Proposition 1: If β is fixed, then $\hat{\alpha}$ exists and is unique.

Proof: Let $L_\alpha = \dfrac{3n}{2\alpha} - \sum_{i=1}^{n} x_i^{2\beta}$, since L_α is continuous and it has been verified that $\lim\limits_{\alpha \to 0} L_\alpha = \infty$ and $\lim\limits_{\alpha \to \infty} L_\alpha = -\sum_{i=1}^{n} x_i^{2\beta} < 0$. This implies that L_α will have at least one root in interval $(0, \infty)$ and hence L_α is a decreasing function in α. Thus, $L_\alpha = 0$ has a unique solution in $(0, \infty)$.

Proposition 2: If α is fixed, then $\hat{\beta}$ exists and is unique.

Proof: Let $L_\beta = \dfrac{n}{\beta} - \alpha \sum_{i=1}^{n} x_i^{2\beta} \ln x_i + 3 \sum_{i=1}^{n} \ln x_i$, since L_β is continuous and it has been verified that $\lim\limits_{\beta \to 0} L_\beta = \infty$ and $\lim\limits_{\beta \to \infty} L_\beta = -2 \sum_{i=1}^{n} \ln x_i < 0$. This implies, as above, $\hat{\beta}$ exists and it is unique.

4.1.2 Fisher Information Matrix

Here, we derive the Fisher information matrix for constructing $100(1-\Psi)\%$ asymptotic confidence interval for the parameters using large sample theory. The Fisher information matrix can be obtained, by using equations (4.2) and (4.3), as

$$I(\hat{\alpha}, \hat{\beta}) = -E \begin{pmatrix} l_{\alpha\alpha} & l_{\alpha\beta} \\ l_{\beta\alpha} & l_{\beta\beta} \end{pmatrix}_{(\hat{\alpha}, \hat{\beta})} \tag{5.2.1}$$

where,

$$l_{\alpha\alpha} = -\frac{3n}{2\alpha^2}, \quad l_{\alpha\beta} = -2 \sum_{i=1}^{n} x_i^{2\beta} \ln x_i, \quad l_{\beta\beta} = -\frac{n}{\beta^2} - 4\alpha \sum_{i=1}^{n} x_i^{2\beta} (\ln x_i)^2.$$

The above matrix can be inverted and the diagonal elements of $I^{-1}(\hat{\alpha},\hat{\beta})$ provide the asymptotic variance of α and β, respectively. Now, two sided $100(1-\Psi)\%$ asymptotic confidence interval for α, β can be obtained as

$$\alpha \in [\hat{\alpha} - Z_{1-\frac{\Psi}{2}}\sqrt{var(\hat{\alpha})}, \hat{\alpha} + Z_{1-\frac{\Psi}{2}}\sqrt{var(\hat{\alpha})}],$$

$$\beta \in [\hat{\beta} - Z_{1-\frac{\Psi}{2}}\sqrt{var(\hat{\beta})}, \hat{\beta} + Z_{1-\frac{\Psi}{2}}\sqrt{var(\hat{\beta})}],$$

respectively.

4.2. Bayes estimation

In this subsection, the Bayes estimation procedure for the PMaD parameters has been developed. Here, we consider two independent gamma priors for both shape and scale parameter. The considered prior is very flexible due to its flexibility of assuming different shape. Thus, the joint prior $g(\alpha,\beta)$ is given by;

$$g(\alpha,\beta) \propto \alpha^{a-1}\beta^{c-1}e^{-b\alpha-d\beta} \quad ; \quad \alpha,\beta>0, \tag{13}$$

where a, b, c and d are the hyper-parameters of the considered priors. Using likelihood function of PMaD and equation above, the joint posterior density function $\pi(\alpha,\beta|x)$ is derived as

$$\begin{aligned}\pi(\alpha,\beta|x) &= \frac{L(x|\alpha,\beta)g(\alpha,\beta)}{\int_\alpha\int_\beta L(x|\alpha,\beta)g(\alpha,\beta)d\alpha\, d\beta} \\ &= \frac{\alpha^{\frac{3n}{2}+a-1}\beta^{n+c-1}e^{-\alpha\left(b+\sum_{i=1}^n x_i^{2\beta}\right)}e^{-d\beta}\left(\prod_{i=1}^n x_i^{3\beta-1}\right)}{\int_\alpha\int_\beta \alpha^{\frac{3n}{2}+a-1}\beta^{n+c-1}e^{-\alpha\left(b+\sum_{i=1}^n x_i^{2\beta}\right)}e^{-d\beta}\left(\prod_{i=1}^n x_i^{3\beta-1}\right)\quad d\alpha\quad d\beta}.\end{aligned} \tag{14}$$

In the Bayesian analysis, the specification of proper loss function plays an important role. We talk most frequently used the square error loss function (SELF) to obtain the estimators of the parameters, which defined as

$$L(\phi,\hat{\phi}) \propto \left(\phi-\hat{\phi}\right)^2, \tag{15}$$

where $\hat{\phi}$ is estimate of ϕ. Bayes estimators under SELF is the posterior mean and evaluated by

$$\hat{\phi}_{SELF} = \left[E(\phi|x)\right], \tag{16}$$

provided the expectation exist and finite. Thus, the Bayes estimators based on equation no. (4.5) under SELF are given by

$$\hat{\alpha}_{bs} = E_{\alpha,\beta|x}(\alpha|\beta,x) = \eta^{-1}\int_\alpha\int_\beta \alpha^{\frac{3n}{2}+a}\beta^{n+c-1}e^{-\alpha\left(b+\sum_{i=1}^n x_i^{2\beta}\right)}e^{-d\beta}\left(\prod_{i=1}^n x_i^{3\beta-1}\right)\, d\alpha\, d\beta, \tag{17}$$

and

$$\hat{\beta}_{bs} = E_{\alpha,\beta|x}(\beta|\alpha,x) = \eta^{-1}\int_\alpha\int_\beta \alpha^{\frac{3n}{2}+a-1}\beta^{n+c}e^{-\alpha\left(b+\sum_{i=1}^n x_i^{2\beta}\right)}e^{-d\beta}\left(\prod_{i=1}^n x_i^{3\beta-1}\right)\, d\alpha\, d\beta, \tag{18}$$

where $\eta^{-1} = \int_\alpha\int_\beta \alpha^{\frac{3n}{2}+a-1}\beta^{n+c-1}e^{-\alpha\left(b+\sum_{i=1}^n x_i^{2\beta}\right)}e^{-d\beta}\left(\prod_{i=1}^n x_i^{3\beta-1}\right)\, d\alpha\, d\beta$.

From equations (4.8) and (4.9), it is easy to observe that the posterior expectations are appearing in the form of the ratio of two integrals. Thus, the analytical solution of these expectations are not presumable. Therefore, any numerical approximation techniques may be implemented to

secure the solutions. Here, we used one of the most popular and quite effective approximation technique suggested by [11]. The detailed description is as follows.

$$(\hat{\alpha},\hat{\beta})_{Bayes} = \frac{\int_\alpha\int_\beta u(\alpha,\beta)e^{\rho(\alpha,\beta)+l}\quad d\alpha d\beta}{\int_\alpha\int_\beta e^{\rho(\alpha,\beta)+l}\quad d\alpha d\beta} \tag{19}$$

$$\begin{aligned} &= (\hat{\alpha},\hat{\beta})_{ml} + \frac{1}{2}[(u_{\alpha\alpha}+2u_\alpha\rho_\alpha)\tau_{\alpha\alpha} + (u_{\alpha\beta}+2u_\alpha\rho_\beta)\tau_{\alpha\beta} + (u_{\beta\alpha}+2u_\beta\rho_\alpha)\tau_{\beta\alpha} \\ &+ (u_{\beta\beta}+2u_\beta\rho_\beta)\tau_{\beta\beta}] + \frac{\alpha}{\beta}[(u_\alpha\tau_{\alpha\alpha}+u_\beta\tau_{\alpha\beta})(l_{111}\tau_{\alpha\alpha}+2l_{21}\tau_{\alpha\beta}+l_{12}\tau_{\beta\beta}) \\ &+ (u_\alpha\tau_{\beta\alpha}+u_\beta\tau_{\beta\beta})(l_{21}\tau_{\alpha\alpha}+2l_{12}\tau_{\beta\alpha}+l_{222}\tau_{\beta\beta})], \end{aligned} \tag{20}$$

where $u(\alpha,\beta)=(\alpha,\beta)$, $\rho(\alpha,\beta)=\ln g(\alpha,\beta)$ and $l=\ln L(\alpha,\beta|\underline{x})$,

$$l_{ab} = \frac{\partial^3 l}{\partial\alpha^a\partial\beta^b},\quad a,b=0,1,2,3\quad a+b=3,\quad \rho_\alpha=\frac{\partial\rho}{\partial\alpha},\quad \rho_\beta=\frac{\partial\rho}{\partial\beta}$$

$$u_\alpha = \frac{\partial u}{\partial\alpha},\quad u_\beta=\frac{\partial u}{\partial\beta},\quad u_{\alpha\alpha}=\frac{\partial^2 u}{\partial\alpha^2},\quad u_{\beta\beta}=\frac{\partial^2 u}{\partial\beta^2},\quad u_{\alpha\beta}=\frac{\partial^2 u}{\partial\alpha\partial\beta},$$

$$\tau_{\alpha\alpha}=\frac{1}{l_{20}},\ \tau_{\alpha\beta}=\frac{1}{l_{11}}=\tau_{\beta\alpha},\ \tau_{\beta\beta}=\frac{1}{l_{02}}.$$

Since $u(\alpha,\beta)$ is the function of α,β,

- If $u(\alpha,\beta)=\alpha$ in (4.11), then

$$u_\alpha=1,\quad u_\beta=0,\quad u_{\alpha\alpha}=u_{\beta\beta}=0,\quad u_{\alpha\beta}=u_{\beta\alpha}=0.$$

- If $u(\alpha,\beta)=\beta$ in (4.11), then

$$u_\beta=1,\quad u_\alpha=0,\quad u_{\alpha\alpha}=u_{\beta\beta}=0,\quad u_{\alpha\beta}=u_{\beta\alpha}=0,$$

and the rest derivatives based on likelihood function are obtained as

$$l_{30}=\frac{3n}{\alpha^3},\ \ l_{11}=-2\sum_{i=1}^n x_i^{2\beta}\ln x_i,\ \ l_{03}=\frac{2n}{\beta^3}-8\alpha\sum_{i=1}^n x_i^{2\beta}(\ln x_i)^3$$

$$l_{12}=-4\sum_{i=1}^n x_i^{2\beta}(\ln x_i)^2=l_{21}.$$

Using these derivatives the Bayes estimators of (α,β) are obtained by expressions

$$\begin{aligned} \hat{\alpha}_{bl} =&\hat{\alpha}_{ml} + \frac{1}{2}[(2u_\alpha\rho_\alpha)\tau_{\alpha\alpha} + (2u_\alpha\rho_\beta)\tau_{\alpha\beta}] + \frac{1}{2}[(u_\alpha\tau_{\alpha\alpha})(l_{30}\tau_{\alpha\alpha}+2l_{21}\tau_{\alpha\beta}+l_{12}\tau_{\beta\beta}) \\ &+ (u_\alpha\tau_{\beta\alpha})(l_{21}\tau_{\alpha\alpha}+2l_{12}\tau_{\beta\alpha}+l_{03}\tau_{\beta\beta})], \end{aligned} \tag{21}$$

$$\begin{aligned} \hat{\beta}_{bl} = \hat{\beta}_{ml} + \frac{1}{2}[(2u_\beta\rho_\alpha)\tau_{\beta\alpha} + (2u_\beta\rho_\beta)\tau_{\beta\beta}] + \frac{1}{2}[(u_\beta\tau_{\alpha\beta})(l_{30}\tau_{\alpha\alpha}+2l_{21}\tau_{\alpha\beta}+l_{12}\tau_{\beta\beta}) \\ + (u_\beta\tau_{\beta\beta})(l_{21}\tau_{\alpha\alpha}+2l_{12}\tau_{\beta\alpha}+l_{03}\tau_{\beta\beta})]. \end{aligned} \tag{22}$$

4.3. Simulation study

In this section, a Monte Carlo simulation study has been performed to assess the performance of the obtained estimators in terms of their mean square errors (MSEs). The MLEs of the parameters are evaluated by using $nlm()$ function, and also the MLEs of reliability characteristics are obtained by using invariance properties. The Bayes estimates of the parameters are evaluated by Lindley's

approximation technique. The hyper-parameters values are chosen in such a way that the prior mean is equal to the true value, and prior variance is taken as very small, say 0.5. All the computations are done by *R*3.4.1 software. At first, we generated 5000 random samples from the PMaD (α, β) using the Newton-Raphson algorithm for different variation of sample sizes as $n = 10$ (small), $n = 20, 30$ (moderate), $n = 50$ (large) for fixed $(\alpha = 0.75, \beta = 0.75)$ and secondly for different variation of (α, β) when sample size is fixed $(n = 20)$, respectively. Average estimates and mean square error (MSE) of the parameters are calculated for the above mentioned choices, and the corresponding results are reported in Table 2. The asymptotic confidence interval (ACI) and asymptotic confidence length (ACL) are also obtained and presented in Table 3. From this simulation study, it has been observed that the precision of MLEs and Bayes estimator are increasing when the sample size is increasing while average ACL is decreasing. The Bayes estimates under informative prior is more precise as compared to the MLEs especially for small sample sizes while for large sample the precision of the estimators is almost same for all the considered parametric choices.

Table 2: *Average estimates and mean square errors (in each second row) of the parameters and reliability characteristics based on simulated data.*

n	α, β	α_{ml}	β_{ml}	$M(t)_{ml}$	$R(t)_{ml}$	$H(t)_{ml}$	α_{bl}	β_{bl}
10	0.75,0.75	0.5070	1.1598	1.5119	0.9691	0.1663	0.5063	1.1028
		0.0631	0.2588	0.0164	0.0049	0.0947	0.0631	0.2027
20		0.6560	0.8848	1.4922	0.9343	0.2965	0.6521	0.8647
		0.0098	0.0326	0.0093	0.0014	0.0703	0.0105	0.0263
30		0.7096	0.8064	1.4883	0.9163	0.3504	0.7058	0.7951
		0.0022	0.0103	0.0071	0.0004	0.0010	0.0025	0.0087
50		0.7542	0.7453	1.4869	0.8988	0.3968	0.7514	0.7397
		0.0003	0.0031	0.0046	0.0001	0.0003	0.0003	0.0031
for fixed n and different α, β								
20	0.5,0.75	0.6603	0.6832	1.7380	0.9044	0.3400	0.6574	0.6716
		0.0261	0.0125	0.0585	0.0017	0.0099	0.0252	0.0117
	0.5, 1.5	0.7290	0.3033	4.6222	0.7871	0.3556	0.7258	0.3229
		0.0528	1.4330	11.9171	0.0402	0.1139	0.0513	1.3866
	1.5, 0.5	0.5090	2.9297	1.1531	0.9983	0.0207	0.5517	2.8634
		0.9907	6.6465	0.0242	0.1274	26.0695	0.9087	6.3006
	2.5,2.5	1.0448	0.5958	1.4084	0.7953	0.6393	1.2825	0.6727
		2.1402	3.6573	0.3860	0.0373	0.3553	1.5058	3.3715

Table 3: *Interval estimates and asymptotic confidence length (ACL) of the parameters.*

n	α, β	α_L	α_U	ACL_α	β_L	β_U	ACl_β
10	0.75,0.75	0.0874	0.9266	0.8393	0.5711	1.7485	1.1775
20	0.75,0.75	0.3209	0.9911	0.6703	0.5525	1.2171	0.6646
30	0.75,0.75	0.4263	0.9928	0.5665	0.5555	1.0574	0.5019
50	0.75,0.75	0.5290	0.9794	0.4505	0.5631	0.9275	0.3644
for fixed n and different α, β							
20	0.5, 0.75	0.3255	0.9951	0.6696	0.4142	0.9523	0.5381
	0.5, 1.5	0.3794	1.0785	0.6991	0.4819	1.7425	1.2429
	1.5, 0.5	0.4206	1.7812	0.76058	0.2260	1.8334	1.3807
	2.5, 2.5	0.5804	2.9509	0.9788	0.54133	2.7783	1.1365

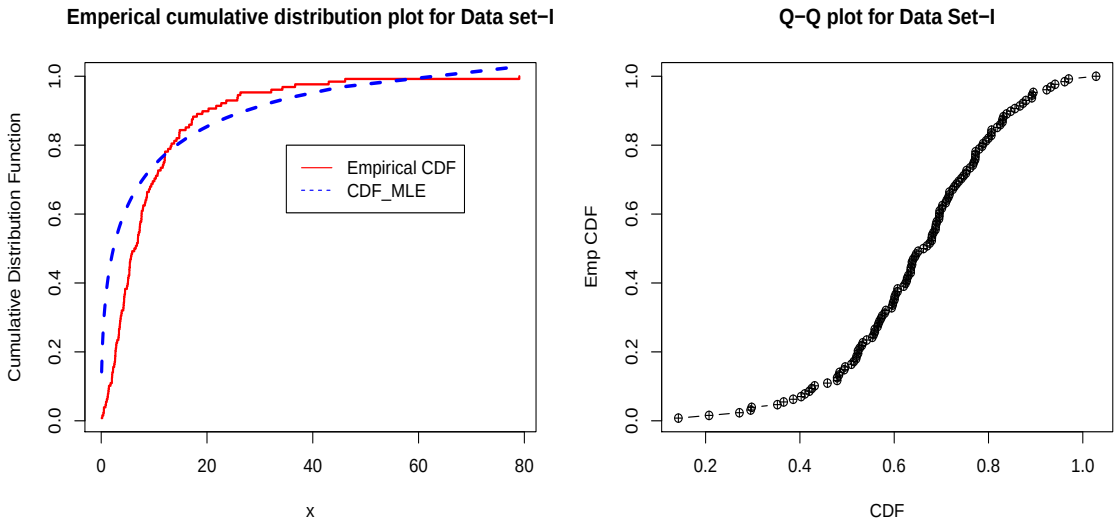


Figure 2: *Empirical cumulative distribution function and QQ plot for the data set-I.*

5. Practical data modeling

This section demonstrates the practical applicability of the proposed model in real-life scenario, especially for the survival/reliability data taken from different sources. The proposed distribution is compared with Maxwell distribution (MaD) and its different generalizations, such as, length biased Maxwell distribution (LBMaD), see [9], area biased Maxwell distribution (ABMaD), see [9], extended Maxwell distribution (EMaD), see [8] and generalized Maxwell distribution (GMaD), see [2]. For these models the estimates of the parameter(s) are obtained by method of maximum likelihood and the compatibility of PMaD has been discussed using model selection tools (which depend on the MLE) such as log-likelihood (-log L), Akaike information criterion (AIC), corrected Akaike information criterion (AICC), Bayesian information criterion (BIC) and Kolmogorov Smirnov (K-S) test. In general, the smaller values of these statistics indicate the better fit to the data.

The data sets description is as follows.

Data Set-I (Bladder cancer data): This data set represents the remission times (in months) of a 128 bladder cancer patients, and it was initially used by [12]. The same data set is used to show the superiority of extended Maxwell distribution by [8].

Data Set-II (Item failure data): This data set is taken from [13]. It shows 50 items put into use at initial time $t = 0$ and failure items recorded in weeks.

Data Set-III (Airborne communication transceiver): The data set was initially considered by [14]. It represent the 46 repair times (in hours) for an airborne communication transceiver.

Data Set-IV (Flood data). The data are the exceedances of flood peaks (in m^3/s) of the Wheaton River near Carcross in Yukon Territory, Canada. The data consist of 72 exceedances for the years 1958-1984, rounded to one decimal place. This data set was analyzed by [16].

Summary of the considered data sets is given in Table 5 and it can be seen that skewness is positive for all data sets which indicates that they have positive skewness which appropriately suited to the proposed model. This table also shows platykurtic, mesokurtic and leptokurtic nature of the data, which proves again the suitability of the proposed model to the data.

Table 4: *Goodness of fit values for different model.*

Bladder cancer data N=128							
Model	$\hat{\alpha}$	$\hat{\beta}$	-logL	AIC	AICC	BIC	K-S
PMaD	**0.7978**	**0.1637**	**366.3820**	**736.7639**	**732.8599**	**742.4680**	**0.3675**
MaD	0.0076	–	1014.4440	2030.8870	2028.9190	2033.7400	0.4144
LBMaD	98.6386	–	669.3668	1340.7340	1338.7650	1343.5860	0.4906
ABMaD	78.9109	–	767.8122	1537.6240	1535.6560	1540.4770	0.5608
ExMaD	0.8447	1.4431	412.1232	828.2464	824.3424	833.9504	0.8265
GMaD	0.7484	527.2314	426.6019	857.2037	853.2997	862.9078	0.7086
Item failure data N=50							
Model	$\hat{\alpha}$	$\hat{\beta}$	-logL	AIC	AICC	BIC	K-S
PMaD	**0.8339**	**0.1820**	**135.8204**	**275.6407**	**271.8961**	**279.4648**	**0.2625**
MaD	0.0104	–	367.8528	737.7056	735.7890	739.6177	0.4268
LBMaD	72.1146	–	315.1624	632.3248	630.4081	634.2368	0.5112
ABMaD	57.6917	–	374.1247	750.2494	748.3328	752.1615	0.5825
ExMaD	0.6186	1.0139	151.2998	306.5996	302.8550	310.4237	0.7327
GMaD	0.5400	534.1569	151.2643	306.5287	302.7840	310.3527	0.3920
Airborne communication transceiver N=46							
Model	$\hat{\alpha}$	$\hat{\beta}$	- logL	AIC	AICC	BIC	K-S
PMaD	**0.8735**	**0.2709**	**101.9125**	**207.8249**	**204.1040**	**211.4822**	**0.2136**
MaD	0.0406	–	245.1383	492.2766	490.3675	494.1052	0.5027
LBMaD	18.4603	–	237.4945	476.9890	475.0799	478.8176	0.5771
ABMaD	14.7683	–	284.7017	571.4034	569.4943	573.2320	0.6324
ExMaD	0.7290	0.8672	103.3052	210.6104	206.8895	214.2677	0.2989
GMaD	0.6015	122.7666	110.8521	225.7042	221.9833	229.3615	0.4392
River data N=72							
Model	$\hat{\alpha}$	$\hat{\beta}$	- logL	AIC	AICC	BIC	K-S
PMaD	0.805185	0.1504145	212.8942	429.7884	425.9623	434.3418	0.2760
MaD	0.005032	–	610.9235	1223.847	1221.904	1226.124	0.3821
LBMaD	149.0315	–	426.3076	854.6153	852.6724	856.8919	0.4113
ABMaD	119.2252	–	493.3271	988.6543	986.7114	990.9309	0.4529
ExMaD	0.697471	1.306933	251.9244	507.8487	504.0226	512.4021	0.7487
GMaD	0.648149	919.7356	251.2767	506.5534	502.7273	511.1068	0.4998

Table 5: *Summary of the data sets.*

Data	Min	Q1	Q2	Mean	Q3	Max	Kurtosis	Skewness
I	0.080	3.348	6.395	9.366	11.838	79.050	18.483	3.287
II	0.013	1.390	5.320	7.821	10.043	48.105	9.408	2.306
III	0.200	0.800	1.750	3.607	4.375	24.500	11.803	2.888
IV	0.100	2.125	9.500	12.204	20.125	64.000	5.890	1.473

Table 6: *ML and Bayes estimates of the four data sets.*

Data	α_{ml}	β_{ml}	α_{bl}	β_{bl}
I	0.7978	0.1637	0.7962	0.1639
II	0.8339	0.1820	0.8292	0.1821
III	0.8735	0.2709	0.8675	0.2703
IV	0.8052	0.1504	0.8023	0.1506

Table 7: *Interval estimates based on the four data sets.*

Data	α_L	α_U	ACL_α	β_L	β_U	ACL_β
I	0.6545	0.9411	0.2866	0.1373	0.1902	0.0529
II	0.5962	1.0717	0.4754	0.1376	0.2263	0.0888
III	0.6202	1.1269	0.5067	0.2081	0.3337	0.1256
IV	0.6126	0.9978	0.3852	0.1186	0.1822	0.0636

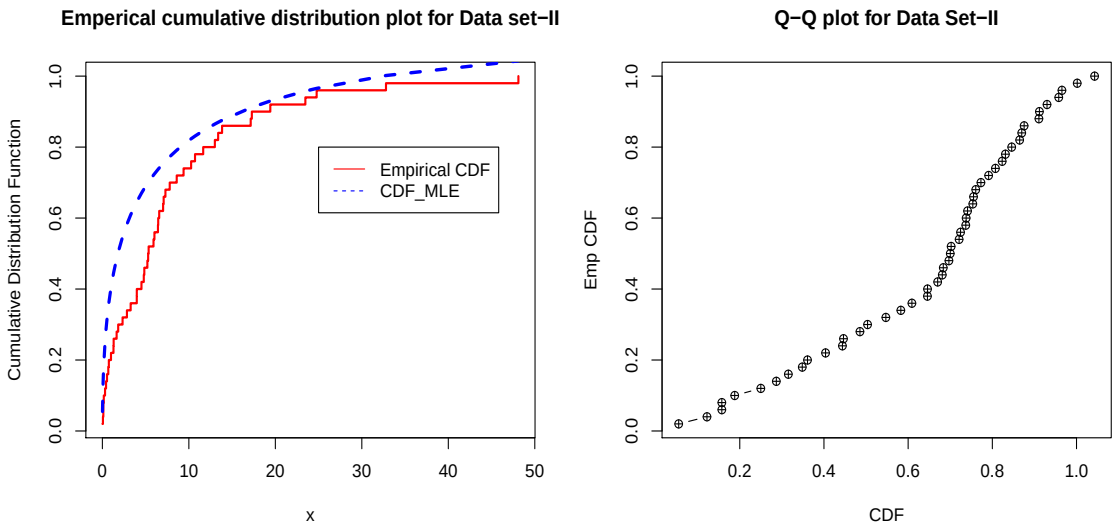


Figure 3: *Empirical cumulative distribution function and QQ plot for the data set-II.*

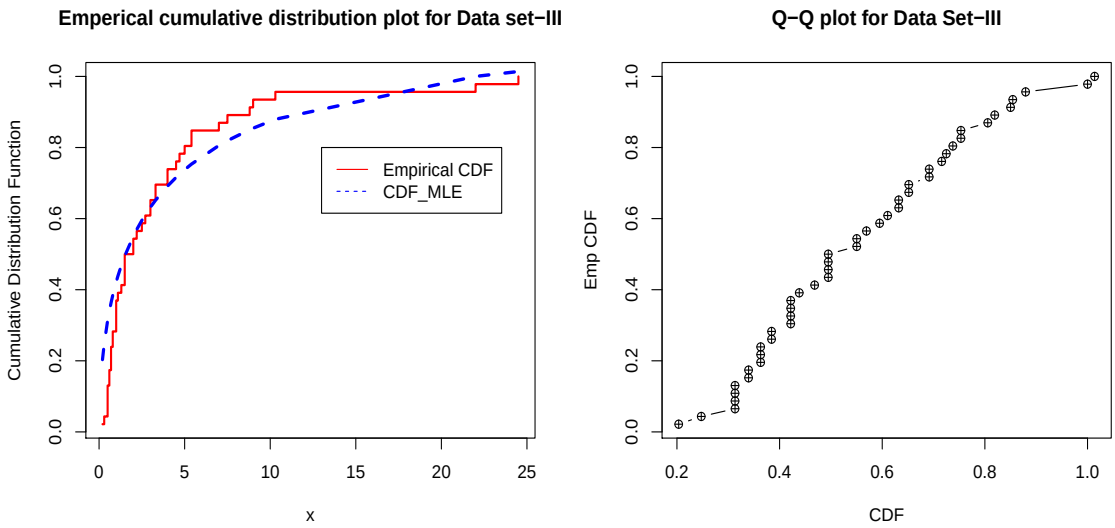


Figure 4: *Empirical cumulative distribution function and QQ plot for the data set-III.*

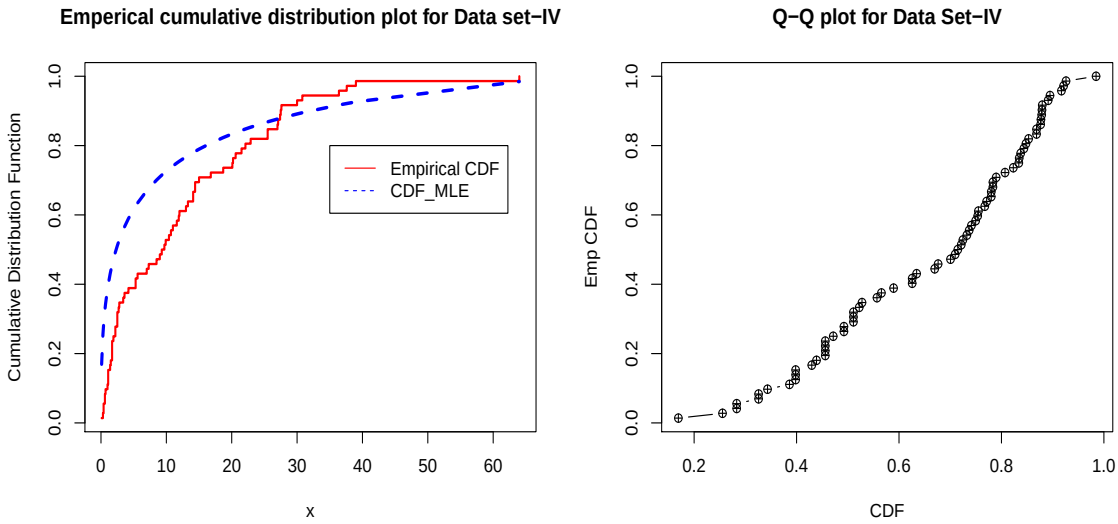


Figure 5: *Empirical cumulative distribution function and QQ plot for the data set-IV.*

From Table 4, it is clear that the proposed model (PMaD) has least value of the model selection tools, which reflects the merit of PMaD for modeling such four practical data sets than the the existing versions of the Maxwell distributions. The empirical cumulative distribution function (ECDF) plots and corresponding QQ plots for all the considered data set are plotted for *PMaD*, see Figures 2-5. From ECDF and QQ plots, it is clear that the considered data sets are adequately fitted to the proposed model. The point (ML and Bayes) estimates of the parameters for each data set are reported in Table 6. The Bayes estimates are calculated under non-informative prior, and it is observed that the obtained estimates (ML and Bayes) are almost same. The interval estimate of the parameter and corresponding asymptotic confidence length are also evaluated and presented in Table 7. This table shows that as the size of the data increases, the length of the interval is decreases, because it decreases the standard error, which support to our simulation part.

6. Conclusion

This article proposed the power Maxwell distribution (PMaD) as a flexible extension of the Maxwell distribution and studied some of its main properties for data modeling. We also study the skewness and kurtosis of the PMaD and found that it is capable of modeling the positively skewed as well as symmetric data. The unknown parameters of the PMaD are estimated by the maximum likelihood estimation (MLE) and Bayes estimation methods. The MLEs of the reliability function and hazard function are also obtained by using the invariance property. The 95% asymptotic confidence interval for the parameters are constructed using Fisher information matrix. The MLEs and Bayes estimators are compared through the Monte Carlo simulation and observed that Bayes estimators are more precise under informative prior. Finally, medical/reliability data have been used to show practical utility of the PMaD, and it is observed that it provides the better fit comparing to other versions of the Maxwell distributions. Thus, it can be recommended as an alternative model for the non-monotone failure rate models.

Acknowledgements

Authors are very grateful to the editor and reviewers for their recommendation to publish this articel in this reputed journal. The first author greatly acknowledges Banaras Hindu University, India for providing financial support in form of Seed grant under the Institute of Eminence

Scheme (scheme no. Dev. 6031).

References

[1] Tyagi, R. K., and Bhattacharya, S. K. (1989). A note on the MVU estimation of reliability for the Maxwell failure distribution. Estadistica 41:73-79.
[2] Chaturvedi, A., and U. Rani. (1998). Classical and Bayesian reliability estimation of the generalized Maxwell failure distribution. Journal of Statistical Research 32:113-20.
[3] Bekker, A., and Roux, J. J. (2005). Reliability characteristics of the Maxwell distribution: A Bayes estimation study. Communications in Statistics - Theory and Methods 34:2169-78.
[4] Radha R. K. and Vekatesan P. (2013). On the double prior selection for the parameter of Maxwell distribution, International Journal of Scientific & Engineering Research, Volume 4, Issue 5.
[5] Shakil M., Golam B. M. K. and Chang K. C. (2008). Distributions of the product and ratio of Maxwell and Rayleigh random variables, Statistical Papers, 49:729–747.
[6] Dey, S., and Maiti, S. S. (2010). Bayesian estimation of the parameter of Maxwell distribution under different loss functions. Journal of Statistical Theory and Practice 4:279-287.
[7] Tomer, S. K., and Panwar, M. S. (2015). Estimation procedures for Maxwell distribution under type I progressive hybrid censoring scheme. Journal of Statistical Computation and Simulation 85:339-356.
[8] Sharma, V. K., Bakouch, H. S. and Khushboo Suthar, K. (2017a). An extended Maxwell distribution: Properties and applications, Communications in Statistics - Simulation and Computation, 46:9, 6982-7007.
[9] Sharma, V. K., Dey, S., Singh, K. S. and Manzoor, U. (2017b). On Length and Area biased Maxwell distributions, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2017.1317804.
[10] Renyi A. (1961). On measures of entropy and information, in: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley.
[11] Lindley, D. V., (1980). Approximate Bayes method, Trabajos de estadistica, Vol. 31, 223-237, 1980.
[12] Lee, E. T. and Wang, J. W. (2003). Statistical Methods for Survival Data Analysis. Wiley, New York, DOI:10.1002/0471458546.
[13] Murthy D. N. P., Xie M., Jiang R. (2004). Weibull model. Wiley, New York.
[14] Chhikara, R. S. and Folks, J. L. (1977). The inverse gaussian distribution as a lifetime model, Technometrics 19: 461-468.
[15] Bonferroni C. E. (1930), Elementi di Statistica General, Seeber, Firenze.
[16] Choulakian, V., Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution, Technometrics 43(4), 478-484.

RT&A, No 4 (65)
Volume 16, December 2021

A Reliability-Inventory Problem Under N-policy of replenishment of component

Achyutha Krishnamoorthy

•

Centre for Research in Mathematics
C.M.S. College, Kottayam - 686001, India
achyuthacusat@gmail.com krishnamoorthy@cmscollege.ac.in

Abstract

In this paper a new process is introduced. To some extent it has resemblance with Queueing-Inventory (Inventory with positive service time) (see Sigman and Simchi-Levy [2] and Melikov and Molchanov [1]. We consider a k - out - of - n: G system of identical components, each of which has exponentially distributed life time with parameter λ, independent of the others. When the number of working components goes down to N ($k \leq N \leq n$) due to failures, an order for $n - k + 1$ items is placed. Replenishment time is exponentially distributed with parameter β. On replenishment, all failed components are instantaneously replaced by the new arrivals, subject to a maximum of $n - k + 1$. This process is investigated and its long run system state distribution derived explicitly. An associated optimization problem is discussed. Throughout this paper the k - out - of - n system is assumed to be COLD.

Keywords: COLD system, System Reliability, *k* - out - of - *n* System, Replenishment policy, Serial and Parallel Systems

1. Introduction

The purpose of this paper is to introduce a notion similar to Queueing - Inventory (QI), introduced in 1992 by two groups of researchers: Sigman and Simchi-Levy [2] and Melikov and Molchanov [1], independently of each other. Until then service time associated with providing an inventoried item was assumed to be negligible. In reality, that assumption is rarely valid. A brief description of QI is as follows. In classical queue, if the server is ready to serve and customers are waiting then the service starts. The notion of the requirement of some materials is totally missing in it. However, to provide service some item(s) is often required. It was Kazimirsky [7] who came up with the idea of an item required to provide service. In the absence of such an item(s) service cannot be given. In classical inventory, it is assumed that the service time is negligible. That is to say, if the item is of demand is available, the server provides it to the customer in a negligible amount of time and the customer leaves the system. In case the item is not available, customers may wait until the inventory gets replenished. Thus absence of inventory alone results in customers joining a queue of demands. The waiting space may be of finite or infinite capacity. On replenishment, a certain number of waiting customers equal to min{number waiting, number of inventoried items replenished}, leave the system with the inventory - it is assumed that each customer asks for exactly one unit of the item. The assumption of negligible service time is often unrealistic. This is the one that prompted Sigman and Simchi-Levy as well as Melikov and Molchanov to introduce positive service time. This results in the formation of queue even when inventory is available. The reader may refer to the recent survey paper by Krishnamoorthy et al [4] for further details on the work done up to 2018 in QI.

We consider a service providing system, namely a k-out-of-n system. Such a system has n identical components/units. The system continuously operates. When the number of operational component hits $k-1$, the system fails. We assume that the life times of these n units are independent and identically distributed random variables with exponential distribution having parameter λ. Up on the number of working components going down to $N(k \leq N \leq n)$ due to failures, an order for $n-k+1$ items are placed. Lead time is exponentially distributed with parameter β. The life time of components and lead time are independent random variables. On replenishment, all failed components are replaced by the new arrivals, subject to a maximum of $n-k+1$. This process is analysed to derive its long run system state distribution. In this paper the case of COLD system alone is analysed and an associated optimization problem discussed. The system is referred to as COLD if the components that were operational at the time of system failure, do not deteriorate further until the system is again put back to operation by replacement/repair of failed components. We can consider different types of replenishment policies and also systems of that are WARM or HOT. In a warm system, components that remain operational at the time of system failure continue to deteriorate, but at a slower rate than when the system is up. We restrict the discussion to COLD system because the very purpose of this work is to announce the above indicated new direction of thoughts. For this reason we also assumed that all distributions involved are exponential.

Next we present a brief discussion in the investigation done on the reliability of k - out - of - n : G system. This system is extensively investigated. Its particular cases, serial and parallel systems are of special interest. A detailed discussion on these can be found in Sivazlian and Stanfel [3]. Krishnamoorthy and Ushakumari [5] extended a repairable k - out - of - n : G system to the case of retrial of failed components for repair. Krishnamoorthy et al [6] introduced the N-policy of repair in k - out - of - n : G system and investigated the optimal number N of failed components that should accumulate in order to start the repair of failed components in a cycle to maximize the reliability of the system. Here a cycle is defined as the time interval that starts at the epoch all the n components are in working condition until the moment all components that fail during this time period are repaired and the system is back with all components in operational state.

Barlow and Heidtmann [9] present a linear-time algorithm and its short computer program in BASIC for the computation of reliability of a k - out - of - n : G system. We now turn to a few more recent investigations on k - out - of - n : G system. Zhang et al. [10] analyse a k - out - of - n : G system with repairman's single vacation and shut off rule. The working times and repair times of components follow exponential distributions, and the duration of the repairman's vacation is governed by a phase type distribution. Both transient and long run system availability are obtained. Time-dependent behavior of the system performance measures under different initial system states, are obtained. Monte Carlo simulation and special cases of the system are investigated to check the correctness of the results obtained. Ji-EunByun et al [11] investigate the reliability growth of k-out-of- N systems using matrix-based system reliability method. To increase the reliability of a specific system, using redundant components is a common method which is called redundancy allocation problem (RAP). Some of the RAP studies have focused on k-out-of-n systems. However, all of these studies assumed predetermined active or standby strategies for each subsystem. Mahsa Aghaei et al [12] propose a k - out - of - n series - parallel system when the redundancy strategy can be chosen for each subsystem. Because the optimization of RAP belongs to the NP-hard class of problems, a modified version of genetic algorithm (GA) is developed. The exact method and the proposed GA are implemented on a well-known test problem and the results obtained demonstrate the efficiency of the approach of the authors compared to the previous studies.

In this paper we introduce the concept of replacement of failed components through a purchase of new items that have the same life time distribution as the failed components. The

order for purchase is placed when the number of operational components in the system falls down to $N, k \leq N \leq n$. It takes an exponentially distributed amount of time, called the lead time, for the replenishment of items to take place. The order quantity is fixed at $n-k+1$. On physical realization of the order, failed components are replaced by the new arrivals. The time for replacement is assumed to be negligible. It may be noted that at most $n-k+1$ components need replacement at the time when replenishment takes place because operational components do not deteriorate when the system is down (COLD system). As a result none, one , ... up to a maximum of $N-k+1$ excess/spare components will be available as standby units. This means that the system is working now with all n components in operation and the remaining, if any, stay as spares. These are brought to operation, one at a time, as and when components fail. This process gets repeated.

The reader may wonder about the distinction from the classical queueing-inventory (QI) problem and may even ask the question: are they not the same if the number of customers in the QI is restricted to a finite number? *The answer is a firm NO. This is so because at a replenishment epoch the number of failed units of the k - out - of - n system can be smaller than $n-k+1$, the replenishment quantity. Thus there could be excess inventory to be stored, which are put into operation when failure of components of the system takes place. Those excess components alone have holding cost. However, in QI the inventory level may at most reach S at a replenishment epoch. Further all items held in the inventory have holding cost associated with them. Also notice that all components of the system that are in operation, deteriorate, though those on Şstandby (the excess remaining after failed components are replaced)Ť do not deteriorate (because the system is COLD).* Thus there are valid reasons for analysing the reliability-inventory (RI) problem presented in the previous paragraph.

The remaining part of this paper is arranged as follows. In section 2, the mathematical model of the problem is presented. The long run system state distribution is explicitly computed. In section 3, we compute a few distributions of interest, associated with the model. Section 4 provides a cost function involving the decision variable N. Its analysis is then presented. This cost function is shown to be convex. Thus there is a global optimum value for N. Finally a concluding section tells about future plans for extensions and generalizations.

Notations and abbreviations:
In the sequel the following notations and abbreviations are employed:
i.i.d - independent and identically distributed.
rv(s) - random variable(s).
CTMC - Continuous time Markov Chain.
IPV Ů initial probability vector.
$X(t)$ Ű Number of operational components in the system at time t.
$Y(t)$ Ű Number of spare/standby components available at time t.
Note that only when $X(t) = n$, the value of $Y(t)$ can be positive.

2. Mathematical Modeling and Analysis of the problem

The system under consideration is COLD: when the system fails in the absence of at least k operational components, the components that are still operational do not deteriorate until system again starts operation, with the failed components replaced by new ones. Though only one new component suffices to put the system back into operation, we follow the policy of replacing all failed components at the time when replenishment of the ordered items take place. The replenishment quantity is $n-k+1$. All of them may not be immediately required. Therefore the excess items are kept as spares/standby for future replacements as and when required. Life times of components are *i.i.d rv*s having exponential distribution with parameter λ. When number of operational components drops down to N, with $k \leq N \leq n$, an order is placed for $n-k+1$ new components. It takes an exponentially distributed time with parameter β for the materialization

of this order. This is referred to as lead time in inventory management. During this time, none, one, ... , up to a maximum of $N-k+1$ components may fail. Up on replenishment, all failed components are replaced by new ones and the system continues to operate. It may be noted as stated earlier, that all of these $n-k+1$ units may not be required to bring back the number of operational components in the system to n. Therefore only that much of these new components that are required.

With $X(t)$ defined as the number of operational components at time t and $Y(t)$, that of spares, we see that $\{(X(t),Y(t)),t\geq 0)\}$ is a two-dimensional $CTMC$ with state space $\{(i,0)|i=k-1,2,...,n\}\cup\{(n,j)|j=1,2,...,N-k+1\}$. This process is not skip - free to the right because, immediately after replenishment the number of operational components increases by at least $n-N$ (with none, one or more left as excess) and at most by $n-k+1$(without any unit left as standby). We employ the difference-differential equation technique to compute the long run system state distribution. The figure below provides the working of the system: 2 - out - of - 5 : G system.

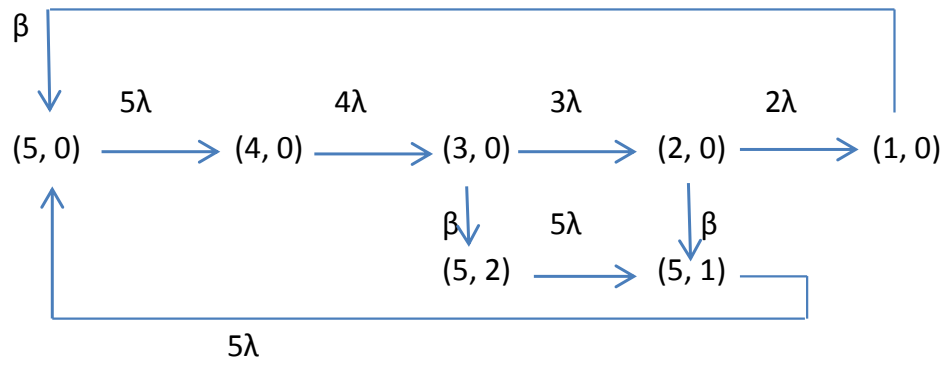


Figure 1: *Transition diagram of* 2 - *out* - *of* - 5 : G *system with* $N=3$ *when the failure rate is* λ.

The transition rate matrix of the 2 - out - of - 5 : G system is as given below:

$$\begin{array}{c} \\ (1,0)\\(2,0)\\(3,0)\\(4,0)\\(5,0)\\(5,1)\\(5,2)\end{array}\begin{array}{c}\begin{array}{ccccccc}(1,0)&(2,0)&(3,0)&(4,0)&(5,0)&(5,1)&(5,2)\end{array}\\ \begin{pmatrix}-\beta & & & & \beta & & \\ 2\lambda & -(\beta+2\lambda) & & & & \beta & \\ & 3\lambda & -(\beta+3\lambda) & & & & \beta \\ & & 4\lambda & -4\lambda & & & \\ & & & 5\lambda & -5\lambda & & \\ & & & & 5\lambda & -5\lambda & \\ & & & & & 5\lambda & -5\lambda\end{pmatrix}\end{array}$$

In this we have, $n=5$, $k=2$ and $N=3$. For that system we get the long run behavior of the system as

$$\begin{aligned} q_{4,0} &= \tfrac{5}{4}q_{5,0};\\ q_{3,0} &= \tfrac{4\lambda}{\beta+3\lambda}\tfrac{5}{4}q_{5,0};\\ q_{2,0} &= \tfrac{3\lambda}{\beta+2\lambda}\tfrac{4\lambda}{\beta+3\lambda}\tfrac{5}{4}q_{5,0};\\ q_{1,0} &= \tfrac{2\lambda}{\beta}\tfrac{3\lambda}{\beta+2\lambda}\tfrac{4\lambda}{\beta+3\lambda}\tfrac{5}{4}q_{5,0};\\ q_{5,2} &= \tfrac{\beta}{5\lambda}q_{3,0} = \tfrac{\beta}{5\lambda}\tfrac{4\lambda}{\beta+3\lambda}\tfrac{5}{4}q_{5,0}\\ q_{5,1} &= q_{5,0}-\tfrac{\beta}{5\lambda}q_{1,0} = q_{5,0}-\tfrac{\beta}{5\lambda}\tfrac{2\lambda}{\beta}\tfrac{3\lambda}{\beta+2\lambda}\tfrac{4\lambda}{\beta+3\lambda}\tfrac{5}{4}q_{5,0}. \end{aligned}$$

Now we add these limiting probabilities. Since their sum is 1, we immediately get $q_{5,0}$.

A different failure rate case also will be discussed in the numerical section; this one considers inverse variation of rate of failure with the number of operational components: when the number of components in operation is j, the failure rate is λ/j. This leads to more compact expressions for the system state probabilities. Therefore, we can expect a much nicer expression for the optimal N value as well. The figure below provides the working of the system: 2 - out - of - 5 : G system.

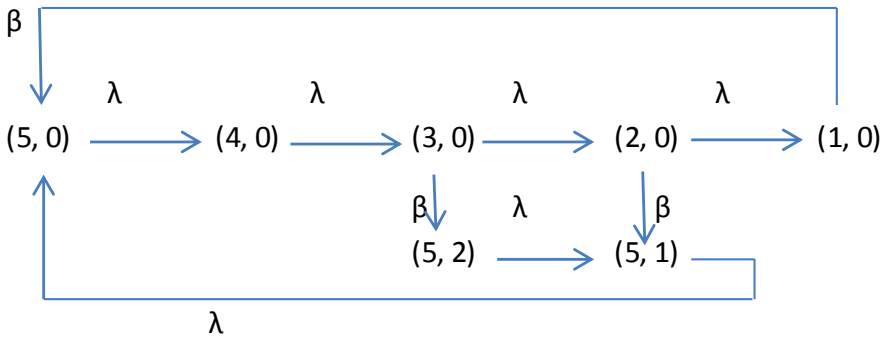


Figure 2: *Transition diagram of* 2 - *out* - *of* - 5 : G *system with* $N = 3$, *when the failure rate is* λ/j.

The transition rate matrix of the 2 - out - of - 5 : G system is as given below:

$$
\begin{array}{c}
 \\
(1,0) \\
(2,0) \\
(3,0) \\
(4,0) \\
(5,0) \\
(5,1) \\
(5,2)
\end{array}
\begin{array}{c}
\begin{array}{ccccccc}
(1,0) & (2,0) & (3,0) & (4,0) & (5,0) & (5,1) & (5,2)
\end{array} \\
\left(\begin{array}{ccccccc}
-\beta & & & & \beta & & \\
\lambda & -(\beta+\lambda) & & & & \beta & \\
 & \lambda & -(\beta+\lambda) & & & & \beta \\
 & & \lambda & -\lambda & & & \\
 & & & \lambda & -\lambda & & \\
 & & & & \lambda & -\lambda & \\
 & & & & & \lambda & -\lambda
\end{array}\right)
\end{array}
$$

Nevertheless, the long run system state probabilities are indicated below for the 2-out-of- 5 : G system, in the case where the failure rate is λ/j when j components are operating. We continue to use the same notation for the system state probability.

$$
\begin{aligned}
q_{5,1} &= (1+\tfrac{\lambda}{\lambda+\beta})q_{5,2}; \\
q_{5,0} &= (1+\tfrac{\lambda}{\lambda+\beta}+\tfrac{\lambda}{\beta}\tfrac{\lambda}{\lambda+\beta})q_{5,2}; \\
q_{4,0} &= (1+\tfrac{\lambda}{\beta})q_{5,2}; \\
q_{3,0} &= (\tfrac{\lambda}{\beta})q_{5,2}; \\
q_{2,0} &= (\tfrac{\lambda}{\beta})(\tfrac{\lambda}{\lambda+\beta})q_{5,2}; \\
q_{1,0} &= (\tfrac{\lambda}{\beta})^2(\tfrac{\lambda}{\lambda+\beta})q_{5,2}.
\end{aligned}
$$

These, together with the normalizing condition, gives $q_{5,2}$.

Assume that the process initially starts in state $(n,0)$. Up to the state $(N,0)$, the process is a pure death process, with linear death rates (depending on the number of operational components). An order for replenishment for $n-k+1$ units is placed on reaching $(N,0)$. The replenishment may precede next failure or may be after the next failure and so on, could be even after the system fails. Therefore there is a chance of system reliability getting affected. Since the system is COLD, no more working component fails until they are again put into operation which can happen

only after the replenishment. When replenishment takes place, all failed units are replaced instantaneously. Thus from $(N, 0)$ onwards the process is no more a pure death process nor it can be called a birth and death process because the replenishment is in bulk. Thus it is not skip-free to the right. Denote by $P_{ij}(t)$, the probability that the system is in state (i, j) at time t and $P'_{ij}(t)$ its derivative. Then the difference-differential equations satisfied by $P_{i,j}(t)$ are:

$$P'_{nj}(t) = -\lambda n P_{nj}(t) + \lambda n P_{nj+1}(t) + \beta P_{i0}(t); j = 0, 1, ..., i-k+1, i = k-1, ..., N;$$
$$P'_{i0}(t) = -(\lambda i + \beta) P_{i0}(t) + \lambda(i+1) P_{i+10}(t) \text{ for } i = k-1, ..., N;$$
$$P'_{i0}(t) = -\lambda i P_{i0}(t) + \lambda(i+1) P_{i+10}(t) \text{ for } i = N+1, ..., n-1.$$

These three systems of equations can be solved for computing the time dependent behavior of the system state probabilities (see Karlin and Taylor [13], Chapter 4). When transient effect fades, the system gets stabilized. Denote by q_{ij} the limit distribution, as $t \to \infty$, of $P_{ij}(t)$. The $CTMC$ under study is aperiodic and irreducible, though it may get absorbed into state $(k-1, 1)$, only to stay there for an exponentially distributed duration. Later on we will consider that state as an absorbing state for deriving the distribution of time during which the system provides failure free operation. Thus the above system of equations gives us:

$$n\lambda q_{n,j} = n\lambda q_{n,j+1} + \beta q_{i,0} \text{ for } j = 0, 1, ..., n-k+1-(n-i) : i = k-1, ..., N;$$
$$(\lambda i + \beta) q_{i,0} = \lambda(i+1) q_{i+1,0} \text{ for } i = k-1, ...N;$$
$$i\lambda q_{i,0} = (i+1)\lambda q_{i+1,0} \text{ for } i = N+1, ..., n-1.$$

These are recursively solved to arrive at the long run system state probability as given below.

Theorem 1. : With q_{ij} defined as the limit as $t \to \infty$ of $P_{ij}(t)$, we get:

$$q_{i,0} = \tfrac{i+1}{i} q_{i+1,0} \text{ for } i = N+1, ..., n-1;$$
$$q_{i,0} = \tfrac{\lambda(i+1)}{(\lambda i+\beta)} q_{i+1,0} \text{ for } i = k-1, ..., N \text{ and}$$
$$q_{n,j} = q_{nj+1} + \tfrac{\beta}{n\lambda} q_{i,0} \text{ for } j = 0, 1, ..., i-k+1 \text{ and } i = k-1, ..., N.$$

Proof. These show that we can express the system state probability in terms of $q_{n,0}$, for example. Then by total probability argument (the normalizing condition), we get $q_{n,0}$. Thus we have explicit analytical expressions for the system state probability. Next we use these to derive several system characteristics which, in turn, are used in analyzing a related optimization problem. ■

3. Performance Characteristics

- Mean number of operational components when the system is working (excluding spares, if any), $OCW = \sum_{j=1}^{j=N-k+1} n q_{n,j} + \sum_{i=k}^{n} i q_{i,0}$.
- Mean number of spare components, $SC = \sum_{j=1}^{N-k+1} j q_{n,j}$.
- Fraction of time system is down, $FTD = q_{k-1,0}$.
- Fraction of time the system is up, $FTU = 1 - q_{k-1,0}$.

FTU is the complement of FTD. Our objective is to make it as close to one as possible, subject to constraints of funds and at the same time the significance of the role of the machine. Thus N plays the most crucial role.

3.1. Related Distributions

In this section we derive a few distributions of interest that arise in the study of the system. We may assume, without loss of generality, that the system starts in state $(n, N-k)$. The distributions that are derived include the distribution of the time until first failure; distribution of the number of replenishments that take place before the system failure; distribution of the number of times the replenishment results in excess inventory and in particular, the distribution of the number of times the excess number of spares reached $N-k+1$ and those that resulted in no excess inventory.

Distribution of the time until first failure

We consider the Markov chain with state space $\{(i,0)|i=k,...,n\} \cup \{(n,j)|j=0,1,...,N-k+1\}$. Notice that we have dropped two states from the state space: The state $(k-1,0)$ is excluded because we want the distribution of the time during which the process remains continuously in the transient states of the Markov chain. Because of that, in consequence to a replenishment, the excess inventory/spare parts level cannot be zero. The initial probability vector γ of the Markov chain has entries 1 at the place corresponding to $(n,0)$ and 0 at the remaining positions. The reason for starting in state $(n,0)$ is that a new cycle starts after the machine failed. Thus the state $(k-1,0)$ is reached before replenishment of components. So after replacing all failed components by the new arrivals, the system is left with no spare unit. Our objective is to compute the distribution of the time T until the state $(k-1,0)$ is reached for the first time. This is given in the following theorem.

Theorem 2. Starting in state one of the states in the set, the distribution of the time T until absorption takes place is phase type with representation (γ, U) of order $n+N-2k+1$. U is that part in the infinitesimal generator of the Markov chain corresponding to the set of states $\{(i,0)|i=k,...,n\} \cup \{(n,j)|j=0,1,...,N-k+1\}$ and γ is the IPV vector with 1 at the position corresponding to the state $(n,0)$ and 0 at the remaining places.

NOTE We may relax the assumption that the initial state is $(n,0)$ by associating probabilities for starting in any state. In that case there will be corresponding changes in the IPV γ. However, for computing the distribution of the time till next failure (i.e., distribution of the time duration between two successive failures of the system), the state $(n,0)$ has to be the starting state. **Proof.** Write the difference - differential equations satisfied by the probabilities of the system occupying any state belonging to $\{(j,0)|j=k,...,n\} \cup \{(n,j)|j=0,1,...,N-k+1\}$. Now solve this matrix differential equation to get the tail distribution of T as $P(T>t) = \gamma e^{(Ut)}\mathbf{e}$, where $\mathbf{e}$ is a column vector of 1Šs having the same order as that of γ. Therefore $P(T<t) = 1-\gamma e^{(Ut)}\mathbf{e}$. The expected time to failure is given by - $\gamma U^{-1}\mathbf{e}$.(see Neuts [8]). ■

Distribution of the number of times the replenishment results in excess inventory before absorption to $(k-1,0)$

To compute this distribution we proceed as follows. We start at an epoch of replenishment that takes the state space to one of $(n,1),....,(n,N-k+1)$. These precisely correspond to those replenishments that take place while the system is in states $(k,0),...,(N,0)$, respectively. The IPV will be defined accordingly. Further we assume that the immediately preceding replenishment took place only after reaching the state $(k-1,0)$. The initial probability vector of the Markov chain associated with these states is $\Theta = (\theta_{k0},...,\theta_{n0},...,\theta_{nN-k})$ and at the remaining positions, including $(k-1,0)$ and $(n,0)$ the entries are all zeros. If we look at the time t (i.e., pre-event occurrence epoch), when the replenishment takes place during $[t,t+h)$ for h infinitesimally small, we notice that in the initial probability vector the only non-zero elements are $\theta_{k0},...,\theta_{N0}$. We introduce an additional component called *level*, as the first coordinate, into the state space of the process. We start at level 0 assuming that no replenishment order has so far materialized. It may happen that the process reaches $(0,k-1,0)$ before the materialization of the replenishment order

that was placed on reaching $(0, N, 0)$. In this case the required number turns out to be zero. We call this a *failure*. Suppose that replenishment against the order which was placed on reaching $(0, N, 0)$, materializes before dropping to $(0, k-1, 0)$. We label this as a *success*. Then the level goes up by 1 and the resulting state is an element of $\{(1, n, j)|j = 1, ..., N-k, N-k+1\}$. This is the first success. Thus the level, the first coordinate in the triplet, stands for the number of consecutive successes. The components start failing with passage of time and on reaching down to $(1, N, 0)$, the next replenishment order Is placed. The two possibilities thereafter are: i) replenishment only after the system breaks down (ie., state $(1, k-1, 0)$is reached) or ii) replenishment takes place before falling to state $(1, k-1, 0)$. In case the event mentioned as (ii) occurs, then we have the second success. The Śconsecutive success counting processŠ goes on like this. In this we notice that the time elapsed between consecutive replenishment epochs are *i.i.d.rv*s following the tail of the phase type distribution with representation $PH(\Theta, V)$ where V is the part of the infinitesimal generator corresponding to these transient states. It is important to note that, because $(k-1, 0)$ is absorbing state, we have not brought it into the above computational argument. For this reason the state $(., n, 0)$ also does not come into play.

Now back to the computation of the required probability distribution. Denote by Y, the random variable that represents the number of successes before the first failure where success and failure are in the context as described in the previous paragraph. Denote the tail of the $PH(\Theta, V)$ distribution described above by p and its complement by q. Then the distribution of Y is given by $P(Y = m) = p^m q$ for $m = 0, 1, ...$ which is the geometric distribution. We sum up these in the following theorem.

Theorem 3. The distribution of the number of times the replenishment results in excess inventory before absorption to $(k-1, 0)$ is given by the geometric distribution with parameter p where p is the tail of the $PH(\Theta, V)$ distribution which is the time until absorption into state $(k-1, 0)$ of the Markov chain describing the state space of the k - out - of - n : G system.

Corollary 1. From theorem 3.2, we conclude that the distribution of number of consecutive failures of the system between two successive failure free cycles is also geometrically distributed. Let Z denote this random variable. Then $P(Z = m) = q^m p$ for $m = 0, 1, 2, ...$

Corollary 2. From the state space description of the Markov chain of the k - out - of - n : G system, it is clear that the consecutive number of times the excess inventory is positive (i.e., it hits the set $\{1, 2, ..., N-k, N-k+1\}$ between two successive system failures, also has the geometric distribution: Denote this *rv* by D. Then $P(D = m) = p^m q$ for $m = 0, 1, 2, ...$

Remark 1. It can be easily proved that the distribution of the time between two successive visits to any state, say $(k-1, 0)$, is phase-type distributed with appropriate representation (see Theorem 3.1). A similar procedure can be adopted to compute the distribution of the time duration for successive visits to any state in the state space of the Markov chain.

4. An Optimization Problem

In this section we construct a cost function involving the decision variable N. The relevant costs are:
K - Fixed cost of placing an order for replenishment
C - Purchase cost/ unit item
h - Holding cost/excess units held/time
R - Penalty cost/time when system is down.
We consider the cost function: Average cost per unit time when the replenishment order for spare items is placed when number of operating components drops down to N,
$F(N) = [K + C(n-k+1)]/$(Expected time elapsed between two consecutive order placements)$+$
$h.\sum_{j=1}^{j=N-k+1} jq_{n,j} + Rq_{k-1,0}$
First we compute the expected length of a cycle. Here a cycle time is the time duration, starting

from an epoch at which state $(N,0)$ is reached to the next epoch at which that state is revisited. . Denote the length of this cycle by W. We have to compute $E(W)$. First we compute the distribution of W. Figure 1 gives an idea about W in the special case discussed therein. In the general case also the state space was described earlier. We incorporate a major modification in the order in which the state space appears and also an Şadditional elementŤ to it for the computation of the distribution of W : $\{(N,0),(N-1,0),...,,(k,0),(k-1,0),(n,N-k+1),...,(n,N-k),....,(n,1),(n,0),...,(n-1,0),...,(N+1,0),*\}$. In this $*$ is an absorbing state and the remaining states are transient. This $*$ is the same as the state $(N,0)$; however the intention of using a distinct notation is to indicate that the state $(N,0)$ is revisited. Thus we can compute the distribution of the distribution of the time duration elapsed, starting from $(N,0)$ back to $(N,0)$ for the first time after the next replenishment at the same level or a lower level followed by deterioration of components. The infinitesimal generator of the corresponding $CTMC$ is given below.

$$\mathcal{G} = \begin{bmatrix} Q & Q^* \\ \mathbf{0} & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} Z_{11} & Z_{12} & & \\ & Z_{22} & Z_{23} & \\ & & Z_{33} & \end{bmatrix}$$

and Q^* the column vector with entry $(N+1)\lambda$ in the last position.

$$Z_{11} = \begin{array}{c} \\ (N,0) \\ (N-1,0) \\ \vdots \\ (k,0) \end{array} \begin{array}{c} \begin{array}{cccccc} (N,0) & (N-1,0) & (N-2,0) & \dots & (k,0) \end{array} \\ \begin{pmatrix} -(\beta+N\lambda) & N\lambda & & & \\ & -(\beta+(N-1)\lambda) & (N-1)\lambda & & \\ & & \ddots & \ddots & \\ & & & & -(\beta+k\lambda) \end{pmatrix} \end{array}$$

$$Z_{12} = \begin{array}{c} \\ (N,0) \\ (N-1,0) \\ \vdots \\ (k,0) \end{array} \begin{array}{c} \begin{array}{ccccc} (k-1,0) & (n,N-k+1) & (n,N-k) & \dots & (n,1) \end{array} \\ \begin{pmatrix} & \beta & & & \\ & & \beta & & \\ & & & \ddots & \\ k\lambda & & & & \beta \end{pmatrix} \end{array}$$

$$Z_{22} = \begin{array}{c} \\ (k-1,0) \\ (n,N-k+1) \\ (n,N-k) \\ \vdots \\ (n,1) \end{array} \begin{array}{c} \begin{array}{ccccc} (k-1,0) & (n,N-k+1) & (n,N-k) & \dots & (n,1) \end{array} \\ \begin{pmatrix} -\beta & & & & \\ & -n\lambda & n\lambda & & \\ & & -n\lambda & & \\ & & & \ddots & \\ & & & & -n\lambda \end{pmatrix} \end{array}$$

$$Z_{23} = \begin{array}{c} \\ (k-1,0) \\ (n,N-k+1) \\ (n,N-k) \\ \vdots \\ (n,1) \end{array} \begin{array}{c} \begin{array}{cccc} (n,0) & \dots & (N+2,0) & (N+1,0) \end{array} \\ \begin{pmatrix} \beta & & & \\ 0 & & & \\ 0 & & & \\ \vdots & & & \\ n\lambda & & & \end{pmatrix} \end{array}$$

$$Z_{33} = \begin{array}{c} \\ (n,0) \\ (n-1,0) \\ \vdots \\ (N+2,0) \\ (N+1,0) \end{array} \begin{array}{c} \begin{array}{ccccc} (n,0) & (n-1,0) & \dots & (N+2,0) & (N+1,0) \end{array} \\ \begin{pmatrix} -n\lambda & n\lambda & & & \\ & -n\lambda & & & \\ & & \ddots & & \\ & & & -(N+2)\lambda & (N+2)\lambda \\ & & & & -(N+1)\lambda \end{pmatrix} \end{array}$$

It follows from it that the time until absorption to $*$ has the Coxian distribution with representation (δ, Q) where Q is that part of the infinitesimal generator sans the row and column corresponding to $*$. Its dimension is $n + N - 2(k-1)$ and δ is the initial probability vector with 1 at the first position and the remaining elements are 0sŠ. Its dimension is obvious from this description. Thus we have proved the following:

Theorem 4. The distribution of a cycle (starting from state (N,0), returning to it for the first time), has Coxian distribution with representation (δ, Q) of order $n + N - 2(k-1)$. Denoting by W the length of this cycle, we have $E(W) = \delta Q^{(-1)}\mathbf{e}$.

Now we go back to the cost function described above. We compute this for two cases:
(a) 2 - out - of - 5 : G system in which N can take values $2, 3, 4, 5$;
(b) 5 - out - of - 10 : G system in which N can take values $5, 6, 7, 8, 9$.
Fix the various costs as $K = \$10, C = \$1, h = \$3, R = \$20$.
We have computed the long run probability distribution of the system (a), as an illustration for the k - out - of - n : G system under N - policy for placing order for replenishment. First we take up that case. The expression for cost function is as follows: $F(N) = [10 + 1(5-2+1)]/$(Expected time elapsed between two consecutive order placements)$+3.\sum_{j=1}^{j=N-2+1} jq_{5,j} + 50q_{1,0}$. The results for various values of λ and β are summarized in the following table:

Table 1: *Effect of N on Cost Function for a* 2 *- out - of -* 5 : *G system.*

(λ, β)	$N=2$	$N=3$	$N=4$	$N=5$
$(1,1)$	28.0085	24.75	**24.444**	24.5
$(1,2)$	23.8617	**19.1453**	19.8726	21.0732
$(2,1)$	40.1261	**38.1405**	38.1920	38.3339

In the case when individual rate of failure is λ/j when the number of operating components is j, the system state probabilities are computed and given in section 2. (b) For this system the state space is$\{(i,0)|i = 4, ..., 10\} \cup \{(10,j)|j = 1, 2, ...N-4\}$.
The expression for cost function is as follows: $F(N) = [10 + 1(10-5+1)]/$(Expected time elapsed between two consecutive order placements)$+3.\sum_{j=1}^{j=N-2+1} jq_{10,j} + 50.q_{4,0}$. The results for various values of λ and β are summarized in the following table:

Table 2: *Effect of N on Cost Function for a* 5 *- out - of -* 10 : *G system.*

(λ, β)	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$
$(1,1)$	34.7209	34.0944	33.75	**33.6134**	33.64
$(1,2)$	28.9071	28.2580	**28.2174**	28.6125	29.3370
$(1.5,1)$	41.8150	41.6413	**41.6096**	41.6771	41.8181

A much more realistic but simple way of looking at the component deterioration would have been as follows: the rate of component deterioration is λ/j when j components in the system are

operational. This leads to the system deterioration rate as $j.\lambda/j = \lambda$. This is the case describing the load balance on the system as stronger when a larger number of components are operational which is more realistic. In this case the expression for the system state probability gets much more simplified and looks more elegant.

The long run system state probabilities in this case are:

$$q_{(k-1,0)} = (\frac{\lambda}{\beta})(\frac{\lambda}{\lambda+\beta})^{(N-k)}q_{(N,0)};$$
$$q_{(N-i,0)} = (\frac{\lambda}{\lambda+\beta})^{(N-i)}q_{(N,0)} \text{ for } i = 1,2,..,N-k;$$
$$q_{(N+1,0)} = q_{(N+2,0)} = ... = q_{(n-1,0)} = q_{(n,0)} = (\frac{\lambda+\beta}{\lambda})q_{(N,0)}$$

The case when failure rate is inversely proportional to the number of operating components, the system state probabilities can be deduced from the above or directly computed. These are as given below:

$$q_{(N+1,0)} = q_{(N+2,0)} = ... = q_{(n-1,0)} = q_{(n,0)} = (1+\frac{\beta}{\lambda})q_{(N,0)};$$
$$q_{(n,N-k+1)} = (\frac{\beta}{\lambda})q_{(N,0)};$$
$$q_{(n,N-k-j)} = (\frac{\beta}{\lambda})\{1+(\frac{\lambda}{\lambda+\beta})+...+(\frac{\lambda}{\lambda+\beta})^{(j+1)}\}q_{(N,0)} \text{ for } j = 0,1,2,..,N-k-1;$$
$$q_{(N-j,0)} = (\frac{\lambda}{\lambda+\beta})^{j}q_{(N,0)} \text{ for } j = 0,1,...,N-k$$
$$q_{(k-1,0)} = \frac{\lambda}{\beta}(\frac{\lambda}{\lambda+\beta})^{(N-k)}q_{(N,0)}.$$

Table 3: *Effect of N on Cost Function for a* 2 *- out - of -* 5 : *G system, when failure rate is* λ/j *when the number of operating components in the system is j.*

(λ,β)	$N=2$	$N=3$	$N=4$	$N=5$
$(1,1)$	11.4589	**7.4706**	7.7273	9.3385
$(1,2)$	8.9415	**5.7808**	7.2535	9.7473
$(1.5,1)$	16.67	11.5242	**10.8677**	11.5857

Table 1 shows that, for the $2-$out-of-5 system, the optimal values of N for the various combinations of (λ,β) given by $(1,1),(1,2)$ and $(2,1)$ are respectively, $4,3,3$ and the minimum costs are \$24.444, \$19.1453 and \$38.1405. In contrast to this, Table 3 shows pretty small values for the cost function. This shows the effect of reduced failure rate when the number of operating units is closer to the maximum value.

Table 4: *Effect of N on Cost Function for a* 5 *- out - of -* 10 : *G system, when failure rate is* λ/j *when the number of operating components in the system is j.*

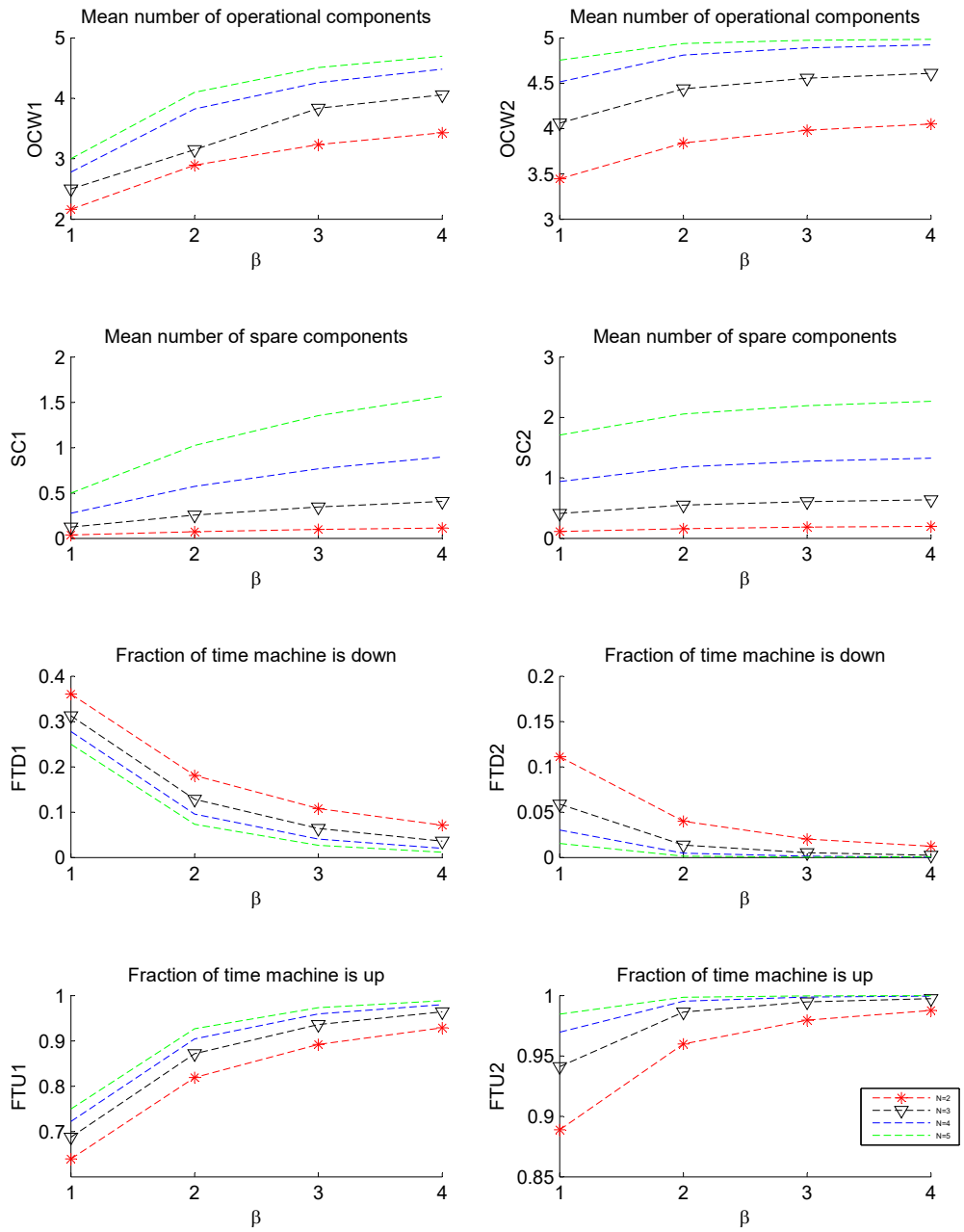
(λ,β)	$N=5$	$N=6$	$N=7$	$N=8$	$N=9$
$(1,1)$	6.5385	**5.4**	5.5306	6.5876	8.3679
$(1,2)$	4.2703	**4.2018**	5.1754	6.8345	9.0552
$(1.5,1)$	10.1739	8.4537	**7.9051**	8.3293	9.5734

Table 2 shows that, for the $5-$out-of-10 system, the optimal values of N for the various combinations of (λ,β) given by $(1,1),(1,2),1.5,1)$ are respectively, \$33.6134, \$28.2174 and \$41.6096. In contrast to this, when failure rate is inversely proportional to the number of operating units, the cost gets considerably reduced and a shift in the optimal N value is also observed.

In this case we can easily see that the holding cost of excess (spare) components increases with the increase in value of N because $N - k$ increases under this condition; the order for replenishment is placed when N is closer to n and so there is higher chance of replenishment taking place before the system reaches the state $(k-1, 0)$, thereby ensuring smooth functioning of the system thereby reducing the risk involved due to system failure. Conversely, if we move down N towards k, the reliability of the system can get seriously affected because the order materialization may get delayed. Consequently the number of operating components could get reduced to $k-1$, thus affecting system reliability. In other words the order for replenishment is placed when the number of operating components is closer to k. So the replenishment could get correspondingly delayed, endangering system reliability. Of course, one can argue that the replenishment time is exponentially distributed and so it lacks memory. In any case for the same parameter of the lead time exponential distribution, we will see the distinction through the examples. In the case of failure rate inversely proportional to number of operating components, we see that the cost function constructed is convex. In particular for parallel (1-out-of-n : G system) and serial (n-out-of-n : G system) systems we get the corresponding optimal N value from the general case.

The eight figures (titled as Figure 3) given below, provide a very clear picture of how the system performs. The first two among these indicate that, with faster replenishment rate the number of components in operation goes up in the two types of failure rates indicated. This trend is also seen to be true for the number of spares available (see the 3rd and 4th figures). Fraction of time the system is up, is considerably smaller when failure rate of the system is directly proportional to the number of operational components than when the system failure rate is inversely proportional to that number (the last two pair of figures). The third pair of figures tells us about the fraction of time the system is down in the two distinct scenarios.

Figure 3: *Effect of β and N on performance measures, when failure rate is λ and λ/j respectively.*



5. Concluding remarks

In this paper we considered a k - out - of - n : G system with N - policy for placing orders for replacement of failed components. The long run system state probability distribution is computed when failure rate is linear. The case of constant failure rate is shown to be a particular case of that. A number of distributions associated with the system are derived. In particular, the time duration between two successive failures of the system is shown to be of phase-type with appropriate representation. The distribution of consecutive number of failure free cycles (each replenishment taking place before the system drops to $(k-1,0)$, and thus system failure is averted) is shown to have geometric distribution. An optimization problem for determining the optimal value of the control variable N, is constructed and its optimal value is computed. Computational experience indicates that the function so constructed, is convex in N.

There are several extensions and generalizations of the problem investigated in this paper. For example, instead of exponential distribution any continuous distribution with non - negative part of real line as support which does not lack memory, could be introduced. However, this may result in the loss of $CTMC$ status for the system. The component life times also could be replaced by such distributions; however, this will lead to a very complex system. Yet another direction of investigation is the case of repair of failed components under N-policy. In this case, when the number of failed components reaches $n-N$, repair of failed units starts. Thus either a machinery/server for repair of failed components has to be hired. Questions such as immediate availability arises in this case just as the role played by the lead time in the model analysed. Also there arises the repair time. A comparison between the model analysed and the case of repair of failed components may lead to interesting results. There is a very important extension of the problem presented in this paper to what can be called Reliability - Queueing - Inventory problem. Another direction for future work is to have a permanent server for repair of failed components. He/she will also process items that can be used to replace failed components. The server does this while waiting for accumulation of $n-N$ failed components of the system. Work on these directions are in progress.

Acknowledgment: The author expresses his deep sense of gratitude to Ms. Anu Nuthan Joshua, Department of Mathematics, Union Christian College, Aluva, India, for carefully going through the manuscript and for correcting a few errors.

References

[1] Melikov, A.Z., Molchanov, A.A. (1992). Stock optimization in transportation/storage systems. *Cybernetics and Systems Analysis*, 28(3): 484 – 487.

[2] Sigman, K., Simchi-Levi, D. (1992) . Light traffic heuristic for an M/G/1 queue with limited inventory. *Annals of Operation Research*, 40: 371 – 380.

[3] Sivazlian B.D. Stanfel L.E. Analysis of Systems in Operations Research, Prentice-Hall international series in industrial and systems engineering, 1975 - 04 - 01.

[4] Krishnamoorthy, A., Dhanya Shajin, Viswanath C. Narayanan (2020). Inventory with Positive Service Time: a Survey, Queueing Theory 2, coordinated by Vladimir Anismov, Nikolaos Liminios. ISTE Editions.

[5] Krishnamoorthy, A.,ăUshakumari, P.V. (1999). Reliability of a k - out - of - n system with repair and retrial of failed units.ă*Topă*7, 293–304 doi:10.1007/BF02564728.

[6] Krishnamoorthy, A., ăUshakumari, P.V.,ăLakshmy, B. (2002). K - out - of - n - system with repair: The N - policy.*ăAsia -Pacific Journal of Operational Research*,ă19 (1): 47.

[7] Kazimirsky, A.V. (2006). Analysis of BMAP/G/1 queue with reservation of service. *Stochastic Analysis and Applications*, 24(4):703 – 718.

[8] Neuts, M.F.(1981). Matrix - Geometric Solutions in Stochastic Models: An Algorithmic Approach. The Johns Hopkins University Press, Baltimore.

[9] Barlow, R.E. and Heidtmann, K.D. (1984).Computing k - out - of - nSystem Reliability, *IEEE Transactions on Reliability*; 30(4):322 – 323.

[10] Zhang, Y., Wu, W., and Tang, Y.(2017).Analysis of an k -out-of-n : G system with repairmanŠs single vacation and shut off rule, *Operations Research Perspectives*,30(4): 29 – 38.

[11] Ji-EunByun, Hee-MinNoh, Junho Song (2017). Reliability growth analysis of k-out-of-N systems using matrix-based system reliability method, *Reliability Engineering and System Safety*, 165: 410 – 421.

[12] Mahsa Aghaei, Ali Zeinal Hamadani, Mostafa Abouei Ardakan (2017).Redundancy allocation problem for k-out-of-n systems with a choice of redundancy strategies,*Journal of Industrial Engineering International*, 13: 81 – 92

[13] Karlin S., Taylor H. E. (1975) *A first course in Stochastic Processes*, 2^nd ed., Academic Press, New York.

SAFETY vs. SECURITY – WHY ARCHITECTURE MAKES THE DIFFERENCE

Jens Braband
•
Siemens Mobility GmbH
Jens.braband@siemens.com

Hendrik Schäbe
•
TÜV Rheinland InterTraffic GmbH
Schaebe@de.tuv.com

Abstract

Cybersecurity plays an increasing role. This also holds true for safety systems. Hence, it is necessary to combine systems that fulfill security and safety requirements. These requirements are partially contradictory. Safety related software will not be changed in an ideal world, whereas security software needs almost permanent updates. This leads to problems that are hard to solve. Different approaches have been proposed by different authors. In this paper we will show, how a suitable architecture can be applied to satisfy the security as well as the safety requirements. We consider some examples of such architectures and show, how systems can be constructed that on the one hand side contain a "golden" code for safety that is not changed and on the other hand side security software that can easily be patched, not touching the "golden" code.

Keywords: Safety architecture, cybersecurity architecture, patching

I. Introduction

It is often claimed that safety and security shall be separated as much as possible but coordinated well. This is also the mantra of new TS 50701 [1]. Also, this is part of the requirements of the EN 50129 [2], see section 7.2 and table E.4 entry 1, which require the separation of safety and non-safety parts of the system. Since usually a security component is not a safety component in a first approximation, this is just what the standard requires.

But the problem is, how to achieve a good implementation of safety and security requirements. There are concepts, that that proper management is the optimal solution, others believe in coordinated lifecycles, and there are also other approaches [3,4,5,6].

However, in the view of the authors, architecture is the decisive factor. Safety and security architecture makes the difference. It is hard to generalize the good practices that are known, but this paper tries to give a few patterns.

II. The "Detect Single Faults" Pattern

Let us consider the first example, which is more or less a straightforward solution. We start with a well-known qualitative design pattern, which works for many safety-related systems. This is the often so-called "fail-safe" system, see EN 50129 [2 for the requirements. See also [7] for further research. This pattern has also its merits for safety vs. security.
Assume, one adds a single component K to a class 2 system S (or zone) according to EN 50159 [8], see figure 1.

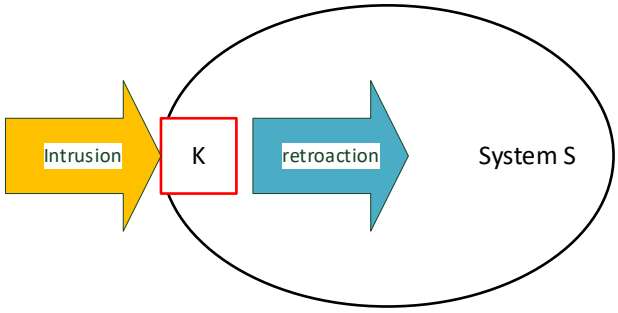


Figure 1: *Adding a security component K to a safety system S*

Even if one just adds K to S, some safety homework still needs to be done, as K may have impact on S even if K does not implement a function for S. For example K may increase latency, decrease reliability etc. The situation gets worse when K is connected to some outside network. Then, in addition a security risk assessment must be carried out, as intrusion may now be possible. Let us assume that this has been managed. However, the solution is not yet complete. One has still to make sure that the security functions implemented in K are functioning according to their specification and that they persist doing this.

Normally the safety standards require that the security mechanisms of K are monitored by S. But the detailed requirements depend on the function that K implements and its architecture.

Example 1: Single-Component-Architecture. Assume K is a filter or firewall just as a kind of gatekeeper that lets only permitted traffic pass (simple whitelisting). If there is a reasonable single failure mode that (partially) deactivates the function, then it is very likely that some monitoring has to be included and that results safely need to be checked – possibly this is not done by a technical function)

Example 2: Two-Component-Architecture: Assume K encrypts transparently all traffic from S to a neighboring zone S*, which has a counterpart K*. If now K fails to decrypt or encrypt any messages, then this will be immediately noticed at the other network when messages start missing. So given a sufficient traffic flow, it is highly unlikely that both components suffer from similar faults with the same effect within a few milliseconds. One can neglect this risk and there is no need to implement any additional monitoring on the safety level. However, it is necessary to ensure that there are no common causes in the two components or in supporting processes as e. g. maintenance that led to the same failure on both sides. Examples are the deactivation of encryption on both sides, use of default keys outdated algorithms etc.

III. The "Safety Channel" Pattern

In this section, we consider another possible architecture. Patching is a particular hot topic when considering systems that need to fulfil safety and security requirements as well. In safety

applications, everybody is reluctant to change the certified "golden code", while in security some applications shall be updated or patched every day. This seems to be a contradiction. It can be solved using an appropriate architecture.

Example 3: Assume one has a safety application, which needs to be protected by a virus checker (VC). This situation is equivalent to a situation, where 3rd party SW needs to be run on the same entity, be it a computer, a kernel or a virtual machine. Assume you need a majority of votes of the different entities for a safety critical decision, e.g., moving a switch.

The basic idea is to split the population of entities into two tribes: the entities labeled N are never changed - or only when the safety application needs to be updated, the entities labeled P can be patched as often as necessary. Of course, some integration tests need to be carried out before patching.. Additionally, the architecture contains voters V that check the outputs before execution. It goes without saying, that all components must be type checked before first operation.

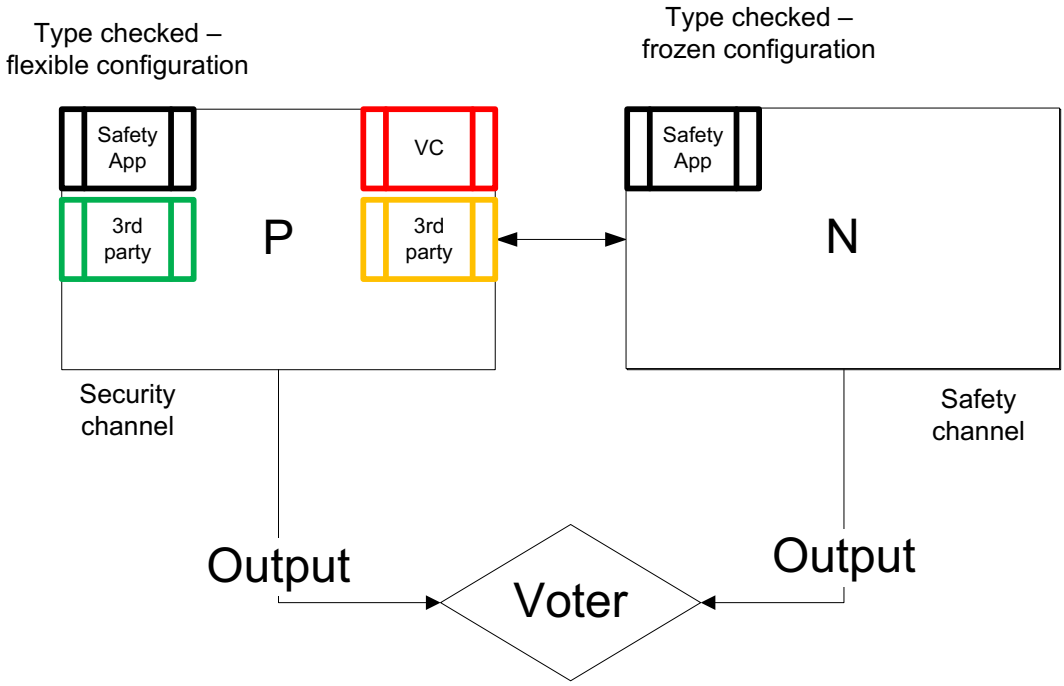


Figure 2: *Patch-friendly architecture*

Figure 2 shows this architecture for a 2002 configuration with one N and one P entity. Both have the same safety application, but on entity P has another software is implemented that needs regular updates. Entity N is sealed, physically and logically protected, e. g. tamper-proof. It is never touched (unless you want to change the safety app). In P the safety application is also never changed. On both channels this may be checked, e.g., by using a hash code on the application or other means as a command from the voter requiring a specific response from both channels, which must coincide. Functional differences in both channels can be detected by the voter command. Now whenever a safety decision needs to be taken, both N and P must agree, which is checked by the voter. So, a final decision of the system is only possible, if the unchanged safety application of N agrees with the decision of the application in P. And thus, it does not matter what other SW runs on P or if it is patched or not. If P was hacked or tampered with or the safety app influenced by the other apps, then P could not change the decision of N. But also, N can't take any decisions of its own, it always needs an agreement of P, the channel that is virus protected etc.

This may now be generalized e. g. to 2 N and 2 P channels demanding that always a majority agrees etc.

Note that some architecture elements have been left out for the clarity of the argument e. g. the inputs, power supply, communication as well as separation between the channels. So, there is still some homework to do. Nevertheless, this architecture is suitable to roll out patches with this architecture.

IV. The "Mixed Architecture" or "EN 50159" pattern

In some cases, it is not possible or not wanted do strictly divide the safety and the security components.

Assume that in the safety part some security functions are integrated, i.e. because on the application level some encryption or message authentication is running. This type of functions is implemented since measures described in EN 50159 [8] are implemented. One must note that EN 50159 is not a cybersecurity standard, it is for safety related communications. So, the measures, although partially the same as in cybersecurity, are dedicated against technical processes that might influence or degrade communication. The mechanisms are others than with hackers, see Braband and Schäbe [9]. Nevertheless, we arrive at the following architecture.

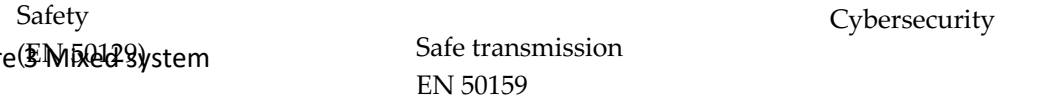


Figure 3 Mixed system

Figure 3: *Mixed Architecture*

Assume that such a system is connected to a similar one. This means, that the safe transmission part of the system as well as the safety parts should not be changed (golden code). All patches should be applied in the cybersecurity part, which then also must compensate for those measures for cybersecurity that cannot and will not be implemented in the safety block.

Example 4: Let us consider the following example. In the safe transmission block, there is an encryption algorithm to protect the data with regards to confidentiality and authenticity (like ETCS). Of course, this protection would not be complete since a hacker might attack the safety part of the system and get access to the data and the code. Therefore, the additional security module is necessary to ensure complete cybersecurity protection. In this example, the security part is split: a never change tribe consisting of the safety part – including the safe transmission part and a patch tribe consisting of the security part. This is a bit similar to example 2 described above. In order to cope with the increasing possibilities of hackers, the security parts is patched. This might even lead to a situation, where the encryption algorithms implemented in the safe transmission part will become superfluous, since in the security part an additional encryption method is installed. The system might thus evolve to the situation described in example 1. However, it is also possible, when carrying out an update of the safety part, also to update the safe transmission part by implementing a more efficient encryption algorithm. This makes it possible, to remove it then from the cybersecurity part.

From example 4, we see that it is only partially possible to design combined systems. Only methods that would not require permanent patching can be implemented in the safety part. Regarding the algorithms, they must work with a certain reserve, i.e., not only provide the simplest

and most basic solutions to the problems. In the safety part, encryption can be used, where a method needs to be chosen that cannot be broken within the next months. Here, one must also not only look for the time that the method can withstand brute force attacks, but backdoors and exploits need to be absent. But of course, these basis security measures in the safety part do not need to be perfect. The main security protection is implemented in the cybersecurity part. Other components as interface drivers, firewalls etc. should not be contained in the safety part since they are candidates for permanent patching.

The architecture discussed in [10] can also be seen as an example of this architecture. In [10] the authors proposed an architecture, where a complete separation of safety and security issues had been carried out. The security mechanisms generate conduits through which the safety functions could be implemented. This leads to the onion skin model shown on figure 4. The onion skin model can only be implemented in this rigorous manner, if a struct separation of safety functions and security functions is possible. In that case, it is the most efficient architectural solution.

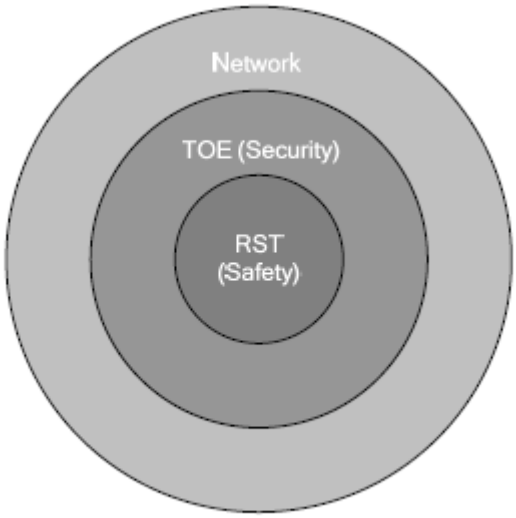


Figure 4: *The onion skin model, taken from [10]*

V. Conclusion

In this paper, we have presented examples system architectures that allow to fulfill the requirements arising from security as well as from safety. We have shown, that with the help of an appropriate architecture, the dilemma of conflicting requirements can be solved in an efficient manner. Surely, not each architecture is applicable for each situation. Therefore, we have presented several examples. We believe that by this approach the problem can be solved in an efficient manner.

References

[1] CENELEC: Railway applications – Cybersecurity, TS 50701, 2021.

[2] EN 50129 Railway applications - Communication, signalling and processing systems - Safety related electronic systems for signaling, 2018.

[3] Glas, B., Gebauer, C., Hänger, J., Heyl, A., Klarmann, J., Kriso, S., Vembar, P. and Wörz, P., (2015). Automotive safety and security integration challenges. In: Klenk, H., Keller, H. B., Plödereder, E. & Dencker, P. (Hrsg.), *Automotive - Safety & Security* 2014. Bonn: Gesellschaft für Informatik e.V.. (S. 13-28).

[4] Kharchenko V., Dotsenko, S.,Illiashenko O., Kamenskiy S., (2019), Integrated Cyber Safety & Security Management System: Industry 4.0 Issue, *10th Internaional Conference on Dependable Systems, Services and Technology (DESSERT)*, 5.-7. 6.2019, DOI: 10, 1109/DESSERT.2019.8770010

[5] Novak, T. and Treytl, A. (2008), Functional Safety and System Security in Automation Systems A Life Cycle Model, *Proceedings of ETFA - 13th IEEE Conference on Emerging Technologies & Factory Automation*, 311-318, DOI: 10.1109/ETFA.2008.4638412

[6] Novak, T., Treytl, A. and Palensky, P., (2007) Common Approach to Functional Safety and System Security in Building Automation and Control Systems, *Proceedings of ETFA. IEEE Conference on Emerging Technologies and Factory Automation 2007*, 1141-1148, (C) IEEE 2007, DOI 10.1109/EFTA.2007.4416910

[7] Gayen, J.T. and Schäbe, H. (2008). (Mis-) conceptions of safety principles, *ESREL Proceedings Safety, Reliability and Risk analysis*, 2: 1283-1291.

[8] CENELEC: Railway applications Communication, signalling and processing systems Safety-related communication in transmission systems, EN 50159, 2010.

[9] Braband, J. and Schäbe, H. (2016). Probability and security – pitfalls and chances, *Safety and reliability*, 36:3-12

[10] Bock, H.-H., Braband, J., Milius, B. and Schäbe, H. (2012). Towards an IT Security Protection Profile for Safety-Related Communication in Railway Automation, *Lecture Notes in Computer Science, Computer Safety, Reliability, and Security*, 7612: 137-148

EFFECTIVENESS RETENTION RATIO AND MULTISTATE SYSTEMS

Victor Netes
•
Moscow Technical University of Communications and Informatics
v.a.netes@mtuci.ru

Abstract

This paper analyzes approaches to dependability assessment of multistate systems in which partial failures can occur. It is shown that for many multistate systems it is advisable to use the effectiveness retention ratio as a dependability measure. The paper explains the meaning and advantages of this measure, and presents methods for its calculation for two classes of systems covering typical situations. They are additive systems in which the output effect is obtained by summing the output effects of the subsystems, and multimode systems that can perform some function or task in different modes depending on their state. Besides that, the presence and use of the effectiveness retention ratio in international and regional Euro-Asian standards are considered.

Keywords: dependability, multistate systems, partial failures, effectiveness retention ratio, international standards, additive and multimode systems.

I. Introduction

The traditional assumption in dependability theory is that there are two possible states of an item: up and down (definitions of these and other basic terms necessary for a proper understanding of this paper are given in Section II.). However, many complex systems can have intermediate states that they go to as a result of partial failures. These states are characterized by a loss of the ability to perform some, but not all, required functions, or by reduced performance. This led to the need to consider multistate systems. In recent years, a number of books have been published specifically dedicated to this topic: [1–4]. Besides that, a number of well-known handbooks [5–7] and monographs [8–10] have sections on multistate systems. They described different approaches and measures for such systems. Interestingly, in [6, 9], two approaches are described independently in different sections written by different authors. Thus, systematization is required.

This paper identifies and discusses two main approaches to assessing the dependability of multistate systems. The mathematical models used in each of the approaches are described, the corresponding dependability measures are given, the approaches are compared, i.e. their advantages and disadvantages are indicated. The first approach is based on the evaluation of system effectiveness. I. A. Ushakov, the founder and first editor-in-chief of the journal "Reliability: Theory & Applications", in the first paper of its first issue pointed out effectiveness ("performability") among "main directions of modern reliability theory" [11]. The principal

dependability measure that arises within this approach is the effectiveness retention ratio (ERR) [12, 13]. The paper gives its definition, explains the meaning and advantages. All these issues constitute the content of Section III.

Standardization plays an important role in any field of engineering. Dependability is no exception. Therefore, Section IV analyzes the presence and use of the ERR in standards. International and regional Euro-Asian standards are considered.

A number of methods can be used to calculate the ERR (see e.g. [5–7, 9, 10, 12, 13], some other works will be given below). Section V presents methods for its calculation for two classes of systems covering typical situations. They are additive systems in which the output effect is obtained by summing the output effects of the subsystems, and multimode systems that can perform some function or task in different modes depending on their state. In particular, for multimode systems, a technique is proposed that allows calculating the ERR in a fairly general situation and going beyond the two previously known special cases.

II. Some Terminology Remarks

Standardization plays an important role in any field of engineering. The leading international organization for the standardization of dependability is the International Electrotechnical Commission (IEC), or rather, its special technical committee No. 56 (TC56). In accordance with an agreement with the International Organization for Standardization (ISO), the TC56 develops dependability standards not only for the electrotechnical field, but address generic dependability issues across all disciplines, thus making it what is referred to as a horizontal committee [14].

An important area of standardization is terminology. It is necessary to ensure an unambiguous interpretation of terms and mutual understanding. This section is devoted to basic standardized dependability terms, and other aspects of standardization within the topic of this paper will be discussed later.

The standard [15] gives the general terminology used in the field of dependability. It is one of the parts (namely 192) of the International Electrotechnical Vocabulary (IEV), which is presented on the portal Electoropedia (also known as the “IEV Online”): https://electropedia.org/. The terms in [15] are generic and are applicable to all fields of dependability methodology. Unfortunately, some researchers and engineers do not know and do not use this standard. Some examples of incorrect and sometimes misleading use of terms related to dependability in modern information and communication technologies were given in [16]. Therefore, the definitions of the basic terms necessary for a proper understanding of this paper are given below.

Dependability of an item is its ability to perform as and when required. The same definition with the reference to [15] is repeated in the well-known standard [17]. There are two notes to this term in [15]. The first one states that dependability includes availability, reliability, recoverability, maintainability, and maintenance support performance (and, in some cases, other characteristics). The second note states that dependability is used as a collective term for the time-related quality characteristics of an item. In other words, dependability is an umbrella term for the above characteristics [14].

By the way, this definition was created as a result of long and active discussions, and some experts still disagree with it. The issue of developing such a definition was discussed in detail in [18].

The main states of an item are up state and down state. Up (or available) state is the state of being able to perform as required. Down (or unavailable) state is the state of being unable to perform as required, due to internal reason.

Reliability of an item is its ability to perform as required, without failure, for a given time interval, under given conditions. Availability of an item is its ability to be in a state to perform as

required. Reliability and availability may be quantified using appropriate measures. Some of them have the same words in their names:

- reliability is the probability of performing as required for the time interval (t_1, t_2), under given conditions;
- instantaneous (point) availability is the probability that an item is in a state to perform as required at a given instant;
- steady state (asymptotic) availability is the limit, if it exists, of the instantaneous availability when the time tends to infinity.

In other words, reliability is the probability of being in up state during the given time interval, and availability is the probability of being in up state at an instant of time (usually, very far from the original moment).

III. Two Approaches to Dependability Assessment of Multistate Systems

The traditional assumption in dependability theory is that there are two possible states of an item: up and down. Under this assumption, consider a system consisting of n elements. Then the states of the elements and the whole system can be expressed as binary variables. The following symbols are usually used for them. The indicator of the state of the ith element is denoted by x_i: $x_i = 1$, if the ith element is in up state, and $x_i = 0$, if the ith element is in down state. To describe the state of the system, the n-dimensional binary vector $\mathbf{x} = (x_1,...,x_n)$ is introduced. If we denote the two-element set {0, 1} by B, then the set of all states of the system S is B^n.

For the system, the structural function $\varphi : B^n \to B$ is defined [19]: $\varphi(\mathbf{x}) = 1$, if the state $\mathbf{x}$ is up state for the system, and $\varphi(\mathbf{x}) = 0$, if the state $\mathbf{x}$ is down state for the system. Usually, monotone systems (or coherent structures) are considered, which imposes certain restrictions on the function $\varphi(\mathbf{x})$ [19]. Namely, structural functions are 0-preserving, 1-preserving, and monotonic (in the terminology of Boolean functions).

The set of all states of the system S is divided into two disjoint subsets: the subset of up states $S_1 = \{\mathbf{x} \mid \varphi(\mathbf{x}) = 1\}$ and the subset of down states $S_0 = \{\mathbf{x} \mid \varphi(\mathbf{x}) = 0\}$. The main dependability measure in this case is the probability that the system is in up state, which is equal to the mathematical expectation of the structural function:

$$P = \mathbf{P}\{\varphi(\mathbf{x}) = 1\} = \mathbf{E}[\varphi(\mathbf{x})]. \quad (1)$$

However, many complex systems can have intermediate states that they go to as a result of partial failures. These states are characterized by a loss of the ability to perform some, but not all, required functions, or by reduced performance. This led to the need to consider multistate systems. There are two main approaches to assessing the dependability of such systems, which are discussed below.

The first of them originated in the late 1950s [20] and was developed in the 1960s. Its idea was well expressed in the classic monograph [21] (its original Russian edition was published in 1965). According to it, for complex systems "the reliability of the system should be understood to mean the stability of the efficiency with consideration of the reliability of the parts composing the system". However, this idea was not further developed in this book.

I. A. Ushakov made a significant contribution to the development and promotion of this approach. He considered system effectiveness, determined by taking into account the reliability of system's elements. This was the subject of his works [22, 23] and many subsequent ones, the corresponding sections were included in popular handbooks [5–7] and monographs [9, 10] (two handbooks by B. A. Kozlov and I. A. Ushakov have been translated into German, Bulgarian, and Czech; there was also a German version of [9]).

In practice, it is much more convenient to deal with dimensionless relative values. This leads to a dependability measure called the effectiveness retention ratio (ERR). It is defined as the ratio of the value of the effectiveness index of an item's intended use over a certain period of operation to the nominal value of this index, calculated on the assumption that the item did not affected by failures during the specified period. The ERR has a simple and clear meaning. For example, if $ERR = 0.95$, it means that due to failures, the effectiveness is reduced by an average of 5%.

The effectiveness index is usually defined as the expectation of the output effect of the system. Particular form of the output effect depends on the nature of the considered system. For example, it can be the quantity of released products for production systems, the amount of information transmitted, collected or processed for information and communication systems, and so on. For systems that perform individual tasks or jobs, the probability of successful completion of the task can be used as an index of effectiveness. Note that this index can also be represented as a mathematical expectation of the output effect. To do this, the output effect is assumed to be 1 if the task is completed and 0 otherwise. In this case, the ERR has a direct probabilistic meaning. It is equal to the probability that the task completion is not disrupted by failures [12].

Although I. A. Ushakov pointed out the expediency of relative effectiveness (for example, in [24]), the term ERR was not used in his works. The first book to discuss the ERR in detail was [12]. In English it was described in [13].

To construct a mathematical model when determining the ERR, the effectiveness function $\varphi(\mathbf{x})$ can be introduced. It generalizes the classical structural function, and can take not only the values 0 and 1, but also any value from the unit interval $I = [0, 1]$, so in this case $\varphi : B^n \to I$. The value $\varphi(\mathbf{x})$ is the relative output effect of the system in the state $\mathbf{x}$. Its maximum value, which is reached when all elements are in up state, is taken as one. Effectiveness functions, as well as structural functions, are 0-preserving, 1-preserving, and monotonic. Of course, the image of such a function is always a finite set, the number of its elements cannot be greater than 2^n. However, it is often unknown in advance and is determined during the dependability assessment.

This can be interpreted as the fuzzification of the failure criterion [25]. In other words, we can also consider the subsets of up and down states for the system, but they are complementary fuzzy ones with membership functions $\varphi(x)$ and $\overline{\varphi(\mathbf{x})} = 1 - \varphi(\mathbf{x})$ for the subsets of up and down states respectively.

The ERR can be expressed as the mathematical expectation of $\varphi(\mathbf{x})$, which is similar to the right member of the equality (1):

$$ERR = \mathbf{E}[\varphi(\mathbf{x})] = \sum_{\mathbf{x} \in S} \varphi(\mathbf{x}) p(\mathbf{x}), \tag{2}$$

where $p(\mathbf{x})$ is the probability that the system is in state $\mathbf{x}$.

The ERR can also be used for traditional two-state items, in which cases it is usually reduced to measures such as availability and reliability [12, 13]. This can make it easier to choose the right dependability measures.

Unfortunately, this approach and the ERR are little known outside of Russia, despite the above-mentioned publications in English and other languages and the fact that works on this topic by I. A. Ushakov, E. V. Dzirkal and V. A. Netes were mentioned in the survey [26] (these researchers based their works on extensive practical experience in assessing the dependability of complex information, control, and communication systems).

If necessary, besides the ERR, the probability $\mathbf{P}\{\varphi(\mathbf{x}) \geq u\}$ $(0 < u \leq 1)$ can be used as dependability measure. However, it is often difficult to reasonably choose the level u. Besides that, choosing a single value of u actually leads to the traditional scheme: then $\{\mathbf{x} \mid \varphi(\mathbf{x}) \geq u\}$ is the set of up states, and $\{\mathbf{x} \mid \varphi(\mathbf{x}) < u\}$ is the set of down states. This means that some partial failures are considered as complete ones, while others are not considered at all. In some situations, this may be

justified, but in most cases it can lead to a misconception about the system dependability. If we calculate such probabilities for several values $u_1, ..., u_k \in (0, 1]$, then the dependability assessment becomes more complicated, and its results are less clear and inconvenient for analysis.

The second approach emerged in the late 1970s (see [27–29], just name a few). It is used and described in a number of publications by authors from many countries (e.g. [6, 8, 9]). It is assumed that each element and the entire system can be in a finite set of states. Let $D_i = \{0, 1, ..., m_i\}$ be the set of states for the ith element and $D = \{0, 1, ..., m\}$ be the set of states for the system. In a frequently used special case, $m_i = m$, so $D_i = D \ \forall i$. The elements of the sets D_i and D are arranged in ascending order of the performance level. In this case, a generalized structural function can be introduced: $\varphi : D_1 \times ... \times D_n \to D$. Such functions are also monotonic, 0-preserving, and satisfy the condition $\varphi(m_1, ..., m_n) = m$.

The main dependability measures are the probabilities of keeping a given performance level; they are similar to the middle member of the equality (1):

$$P(\nu) = \mathbf{P}\{\varphi(\mathbf{x}) \geq \nu\} \quad (x_i \in D_i, \ \nu \in D). \tag{3}$$

The drawbacks of such measures for multistate systems were discussed above.

Comparing these two approaches to each other, the first immediately noticeable thing is the wider possibilities for describing the states of elements in the second approach. However, in the first approach, some parts composing the system can, if necessary, be considered as subsystems with more than two states. This can lead to the decomposition of systems, considered in particular in [6, 7, 9, 13]. However, this interesting issue is beyond the scope of this paper.

On the other hand, a less obvious but important circumstance is that the values of the cstructural function and the effectiveness function are expressed using different scales of measure (levels of measurement). These are the ordinal scale for the structural function and the absolute scale (the ratio scale having fixed natural 0 and 1) for the effectiveness function. The former is non-metric (qualitative), and the latter is metric (quantitative). Therefore, the values of the structural function can only be compared (equal, greater, or less), and arithmetic operations can be performed with the values of the effectiveness function. This explains and justifies the use of measures (2) and (3). Some works (e.g. [9]) have also introduced the mathematical expectation of the performance level similar to (2), but this does not make sense for ordinal variables.

III. Standardization

The standardization of ERR was specifically reviewed in [31]. However, it mainly discussed Russian and interstate standards. This section is mainly devoted to IEC standards. But first, a brief summary on the ERR in the interstate standards is given. These are regional standards adopted by the Euro-Asian Council for Standardization, Metrology and Certification (EASC) of the Commonwealth of Independent States. The terminology standard [32] gives the definition of the ERR. By the way, it first appeared in the Soviet terminology standard on dependability in 1983. The scope of application and recommendations for using the ERR are given in [33]. Besides that, this standard explains that the effectiveness of an item's intended use is understood as its property to create some useful result (output effect) during the operation under certain conditions.

The ERR is not included in the international terminology standard on dependability [15]. However, the previous version of such a standard [34] contained terms "effectiveness (performance)" and "capacity". They had the following definitions. Effectiveness (performance) is the ability of an item to meet a service demand of given quantitative characteristics. The note to this definition states that this ability depends on the combined aspects of the capability and the availability performance of the item. Capability is the ability of an item to meet a service demand

of given quantitative characteristics under given internal conditions. The note explains that internal conditions refer for example to any combination of faulty and non faulty sub-items. These concepts made it possible to go to the ERR, but this step was not taken, on the contrary, they were excluded during the development of [15].

As noted above, the ERR is primarily required for systems that might be affected by partial failures. This term is defined in [15] as a failure characterized by the loss of some, but not all, required functions. The note to this definition states that a partial failure may lead to a degraded state. The latter is defined as a state of reduced ability to perform as required, but with acceptable reduced performance.

Thus, the definition of partial failure in [15] is only suitable for multifunctional items. However, this concept also makes sense for single-function items, which may be in a degraded state with reduced performance. Therefore, it is advisable to expand this definition by stating it as follows: a failure characterized by the loss of the ability to perform some, but not all, required functions or by reduced performance (output effect).

There are two IEC standards in which the ERR is actually implicitly present. The first of them is [35]. Its subsection on availability contains paragraph 6.1.2.4, which deals with multistate systems with reference to [2]. It gives a simple example of such a system consisting of two elements. In fact, the measure introduced there is the ERR [31].

Another IEC standard that implies the ERR is [36]. It is devoted to communication network dependability. There are two network service scenarios of interest to network dependability. The first of them has the objective to determine the network dependability characteristics of end-to-end (E2E) network services from the perspective of network end-users. It is associated with the specific service paths selected for the E2E connections. The objective of the second scenario is to determine the network dependability characteristics of the entire network from the network operator or the network service provider perspective. Accordingly, two dependability measures are recommended: the E2E network availability and the full-end network availability.

The E2E network availability is the availability of an E2E network connection between a pair of nodes in question, including all available service paths. However, the full-end network availability is not really availability, that is, the probability that some item is in up state. It is the weighted sum of E2E availabilities for different pairs of nodes and actually turns out to be the ERR [37]. In this case, the output effect is defined as the number of connected pairs of users. In general, the feasibility of using the ERR for communication networks and some methods of its calculation were given in [37].

Notable that both of these examples in [35, 36] fit into the same fairly general scheme, which is discussed below.

IV. Calculation of the ERR

I. General Consideration

According to the above definition,

$$ERR = E/E_0,$$

where E is the index of effectiveness and E_0 is the nominal value of this index calculated under the condition that failures do not occur. However, this formula is usually not suitable for calculating in practice. On the contrary, if necessary, E can be calculated as the product of E_0 and the ERR. Formula (2) is suitable for calculations only with a small number of elements. Various

techniques that can be used to calculate the ERR are given in the above-mentioned publications [5–7, 9, 10, 12, 13, 22–25].

Here we present methods for calculating the ERR for two classes of systems, for which it is expressed in terms of dependability measures of subsystems. They cover typical situations and have not been published in English. In both cases, it is assumed that the system has a certain number of subsystems. Generally speaking, they can intersect, i.e. have common elements. Each subsystem is considered binary, i.e. it can be either in up state or in down state. Denote P_j the probability that the j-th subsystem is in up state. Depending on the situation, it can be its availability or reliability.

II. Additive Systems

The first class consists of so-called additive systems. For such systems, the output effect is obtained by summing the output effects of the subsystems. In particular, the systems mentioned above, which are considered in the standards [35, 36], belong to this class. It also includes multifunctional systems, in which it is possible to allocate subsystems responsible for performing functions, and the output effects for all functions are added up.

Let the system have k subsystems and each non-failed subsystem contributes to the overall output effect. Then

$$E = \sum_{j=1}^{k} E_j , \qquad E_0 = \sum_{j=1}^{k} E_{j0} ,$$

where E_j and E_{j0} are the effectiveness and the nominal effectiveness of the j-th subsystem. Also, for binary subsystems $E_j = P_j E_{j0}$.

From these equalities, the formula for calculating the ERR follows:

$$ERR = E \Big/ E_0 = \sum_{j=1}^{k} E_j \Big/ E_0 = \sum_{j=1}^{k} P_j E_{j0} \Big/ E_0 = \sum_{j=1}^{k} (E_{j0}/E_0) P_j = \sum_{j=1}^{k} w_j P_j,$$

where $w_j = E_{j0}/E_0$ is the "weight" of the j-th subsystem.

III. Multimodal Systems

The second class is multimodal systems. They have been known for a long time (see, for example, [38, 20, 23, 5–7]). In some publications (in particular, in [5–7]), they were called multifunctional, which does not fully correspond to the principle of their functioning. Indeed, such a system performs one function or task, but can do it in different modes, depending on its state. For each state, the mode that is possible for it, which gives the maximum output effect, is applied. Each mode corresponds to a specific subsystem that must be in up state in order for the system to operate in this mode.

An example is a communication network in which several paths with different performance parameters (bandwidth, delay, packet loss ratio, etc.) can be used to transmit information. The paths can be characterized by the probability of successful delivery (in time, without errors), and the best of the available paths is chosen.

Let there are m modes, the corresponding subsystems are $G_1, \dots, G_m$, and the corresponding values of the relative output effect are $v_1, \dots, v_m$. They are assumed to be numbered in decreasing order: $v_1 \geq v_1 \geq \cdots \geq v_m > 0$. If we denote by H_l the probability of performing the function in the lth mode, then

$$ERR = \sum_{l=1}^{m} v_l H_l.$$

Thus, the calculation of the ERR is reduced to the calculation of the probabilities H_l. It is clear that $H_1 = P_1$. For $l > 1$, H_l is the probability that the subsystem G_l is in up state, and all subsystems with smaller numbers have failed.

Formulas for calculating the probabilities H_l for $l > 1$ were known for two special cases [5–7, 23].

1. Each element can be included in only one subsystem, i.e. the subsystems are pairwise disjoint. Then

$$H_l = P_l \prod_{j=1}^{l-1} (1 - P_j).$$

2. Each subsequent subsystem is contained in the previous one, that is $G_1 \supset G_2 \supset \cdots \supset G_m$. This means that the 1st (the best) mode require all elements, the 2nd mode require fewer elements, and so on. Then

$$H_l = P_l - P_{l-1}.$$

Let all subsystems are series ones. Then

$$P_l = \prod_{i \in G_l} p_i,$$

where $p_i = \mathbf{P}\{x_i = 1\}$ is the probability that the ith element is in up state. In this case, a general technique can be proposed for calculating the probabilities H_l. It is based on their representation in the following form:

$$H_l = \mathbf{P}\left[\left(1 - \prod_{i \in G_1} x_i\right) \ldots \left(1 - \prod_{i \in G_{l-1}} x_i\right) \prod_{i \in G_l} x_i = 1\right] = \mathbf{E}\left[\left(1 - \prod_{i \in G_1} x_i\right) \ldots \left(1 - \prod_{i \in G_{l-1}} x_i\right) \prod_{i \in G_l} x_i\right].$$

The expression in the brackets in the right member of this formula should be transformed so that there are no repeated variables x_i in it, i.e. that they are all different. This can be done by using the following equalities:

$$(1 - xy)\cdot x = (1 - y)\cdot x, \quad (1 - xy)\cdot(1 - x) = (1 - x), \quad (1 - xy)\cdot(1 - xz) = 1 - x\cdot(y + z - yz).$$

They are valid for any variables $x, y, z \in B$ since they are idempotent. After that, the final result is obtained by substituting p_i instead of x_i in the resulting expression. This follows from the properties of the mathematical expectation and the equality $p_i = \mathbf{E}[x_i]$.

This technique is similar to the one used in [39] to calculate the interval reliability of communication networks.

V. Conclusion

This paper analyzes approaches to dependability assessment of multistate systems in which partial failures can occur. Along the way, the incompleteness of the definition of partial failure in the basic

terminology standard on dependability IEC 60050-192:2015 is revealed and its extended formulation is proposed.

It is shown that for many multistate systems it is advisable to use the effectiveness retention ratio as a dependability measure. It is defined as the ratio of the value of the effectiveness index of an item's intended use over a certain period of operation to the nominal value of this index, calculated on the assumption that the item did not affected by failures during the specified period. The effectiveness index is usually defined as the expectation of the output effect of the system. This approach can be interpreted as the fuzzification of the failure criterion. Unfortunately, this measure is not very well known (especially outside of Russia), and it is used less often than it deserves. Its usage is recommended in reginal interstate (Euro-Asian) standards adopted by EASC (GOST 27.002–2015 and GOST 27.003–2016). It is also implicitly present in two international standards IEC 61703:2016 and IEC 62673:2013, which do not quite correctly attribute it to availability measures.

The paper explains the meaning and advantages of the effectiveness retention ratio, and presents methods for its calculation for two classes of systems covering typical situations. They are additive systems in which the output effect is obtained by summing the output effects of the subsystems, and multimode systems that can perform some function or task in different modes depending on their state. Additive systems include, in particular, the systems considered in the above-mentioned IEC standards. For multimode systems, a technique is proposed that allows calculating the ERR in a fairly general situation and going beyond the two previously known special cases.

The author hopes that this paper will contribute to the dissemination of information about the effectiveness retention ratio and its wider application.

Acknowledgements

I would like to honor the memory of my teachers I. A. Ushakov and E. V. Dzirkal, who did a lot to develop and promote the approach described here.

References

[1] Lisnianski, A., Levitin, G. Multi-state System Reliability: Assessment, Optimization and Applications, World Scientific, Singapore, 2003.

[2] Lisnianski, A., Frenkel, I., Ding, Y. Multi-state System Reliability Analysis and Optimization for Engineers and Industrial Managers, Springer-Verlag, London, 2010.

[3] Natvig, B. Multistate Systems Reliability Theory with Applications, John Wiley & Sons, Chichester, 2011.

[4] Recent Advances in Multi-state Systems Reliability, eds. Lisnianski, A., Frenkel, I., Karagrigoriou, A., Springer, 2018.

[5] Kozlov, B., Ushakov, I. Reliability Handbook, Holt, Rinehart and Winston, New York, 1970.

[6] Reliability of Technical Systems: Handbook, ed. Ushakov, I. A. Radio i Sviaz, Moscow, 1985 (in Russian).

[7] Handbook of Reliability Engineering, ed. Ushakov, I. A. John Wiley & Sons, New York, 1994.

[8] Beichelt, F., Franken. P. Zuverlässigkeit und Instandhaltung. Mathematische Methoden, VEB Verlag Technik, Berlin, 1983.

[9] Reinschke, K., Ushakov, I. A. Application of Graph Theory in Reliability Analysis, Radio i Svyaz, Moscow, 1988 (in Russian).

[10] Gnedenko, B. V., Ushakov, I. A. Probabilistic Reliability Engineering, John Wiley & Sons,

New York, 1995.

[11] Ushakov, I. (2006). Reliability: past, present, future. *Reliability: Theory & Applications*, 1: 10–16.

[12] Dzirkal, E. V. Setting and Testing Requirements to the Reliability of Complex Items, Radio i Svyaz, Moscow, 1981 (in Russian).

[13] Netes, V. A. (2012). Effectiveness retention ratio: a dependability measure for complex systems. *Dependability*, 4:24–33.

[14] Van Hardeveld, T. International Perspectives on Reliability, https://reliabilityweb.com/articles/entry/international_perspectives_on_reliability.

[15] IEC 60050-192:2015. International Electrotechnical Vocabulary – Part 192: Dependability.

[16] Netes, V. (2020). Modern network technologies and dependability. In: Proceedings of the 3rd International Science and Technology Conference "Modern Network Technologies" (MoNeTec-2020).

[17] ISO 9000:2015. Quality Management Systems. Fundamentals and Vocabulary.

[18] Netes, V. A., Tarasyev, Yu. I., Shper, V. L. (2014). How we should define what "dependability" is, *Dependability*, 4:15-26.

[19] Barlow, R. E., Proschan, F. Mathematical Theory of Reliability, John Wiley & Sons, New York, 1965.

[20] Fridell, H. G., Jack, H. G. (1959). System operational effectiveness (reliability, performance, maintainability). In: Proceedings of the 5th National Symposium on Reliability and Quality Control in Electronics.

[21] Gnedenko, B. V., Belyayev, Yu. K., Solovyev, A. D. Mathematical Methods in Reliability Theory, Academic Press, New York and London, 1969.

[22] Ushakov, I. A. (1960). An estimate of effectiveness of complex systems. In: Reliability of Radioelectronic Equipment, Sovietskoye Radio, Moscow (in Russian).

[23] Ushakov, I. A. (1966). Performance effectiveness of complex systems. In: On Reliability of Complex Technical Systems, Sovietskoye Radio, Moscow (in Russian).

[24] Ushakov, I. A. Theory of System Reliability, Drofa, Moscow, 2009 (in Russian).

[25] Netes, V. A. (1976). Method of estimating complex systems dependability and its application to tree-like information networks. *Transactions of ZNIIS*, 2:17–23 (in Russian).

[26] Rukhin, A. L., Hsieh, H. K. (1987). Survey of Soviet works in reliability. *Statistical Science*, 2(4):484–503.

[27] Barlow, R. E., Wu, A. S. (1978). Coherent systems with multi-state components, *Mathematics of Operations Research*, 3:275–281.

[28] El-Neweihi, E., Proschan, F., Sethuraman, J. (1978). Multistate coherent systems, *Journal of Applied Probability*, 15:675–688.

[29] Ross, S. M. (1979). Multivalued state component systems, *Annals of Probability*, 7:379–383.

[30] Reinschke, K. Zuverlässigkeit von Systemen. Bd. I. Systeme mit Endlich Vielen Zuständen, VEB Verlag Technik, Berlin, 1979.

[31] Netes, V. A. (2021). Effectiveness retention ratio and its standardization, *Dependability*, 2:3–8.

[32] GOST 27.002–2015. Dependability in Technics. Terms and Definitions (in Russian).

[33] GOST 27.003–2016. Industrial Product Dependability. Contents and General Rules for Specifying Dependability Requirements (in Russian).

[34] IEC 50(191):1990. International Electrotechnical Vocabulary – Chapter 191: Dependability and Quality of Service.

[35] IEC 61703:2016. Mathematical Expressions for Reliability, Availability, Maintainability and Maintenance Support Terms.

[36] IEC 62673:2013. Methodology for Communication Network Dependability Assessment

and Assurance.

[37] Netes, V. (2020). Dependability measures for access networks and their evaluation. In: Proceedings of the 26th Conference of Open Innovation Association FRUCT.

[38] Zagor, H. I., Curtin, K., Greenberg H. (1958). Reliability of multi-moded systems. *Electronic Industries,* 4:101–104,164–165.

[39] Netes, V. (2021). The interval reliability, its usage and calculation for information and communication systems and networks. In: Proceedings of the 29th Conference of Open Innovation Association FRUCT.

Kernel Sampling Based Parameter Estimation in Detected Community in Weighted Graph in Big Data

Ram Milan, Diwakar Shukla

•

Department of Computer Science and Applications, Dr. Harisingh Gour Vishwavidyalaya (A Central University), Sagar M.P[1, 2.]
rammilan.in@gmail.com[1], diwakarshukla@rediffmail.com[2]

Abstract

The social media platforms are such examples of big-data where the volume, velocity, and variety are visualized over time domain. Registered users of such platforms bear frequent communication with others and that could be identified as a community. Many methods (algorithms) exist in literature to detect such likely groups of frequent communication. This paper presents contribution to estimate parameters of detected communities using sampling procedure. A Kernel sampling procedure is suggested in the setup of detected community environment. A method is suggested whose efficiency has been estimated using calculations of confidence interval. Simulation procedure is used to obtain the lower and upper limits of confidence intervals with the help of multiple samples.

Keywords: Community Detection, Weighted Graph, Big Data, Internet Technology, 4G, Sampling, Simulations Confidence interval.

I. Introduction

With the expansion of social media platforms and technologies, large numbers of users are interacting with each other by forming groups, based on commonness of characters. Some most popular social networking sites are Face-book, Twitter, Instagram and Whatsapp etc. Where users register them self and communicate with the likeminded peoples. This motivates to think over for the identification of phenomena of community formation and community detection. The formation is usually on commonness but detection needs scientific methodologies.

One can assume that each registered user, on social networking platforms, is a vertex of a graph and his social communication with other people represents an edge of a graph. The quantum of connectivity with each other varies exponentially over time which generates voluminous data in a small span of time. The communication defers in modes like text, voice, image, videos, and many other similar which reveal variety in data. Moreover, in a fraction of time, growth of data on social networking platforms is immensely high which reveal velocity characteristics.

The community size and type detection is one such aspect which generates information in terms of popularity and security. Dongsheng Duan Li et al.[1] suggested algorithms for community mining assuming each user a vertex and density of connecting edges a community. An approach to community

discovery based on evaluation of partition matrix has also been considered along with detection of change points. Pizzuti et al. [2] used Genetic algorithms approach for detecting communities in social media platform with mathematical approach using concept of graph theory. The Nan Du, et al. [3] detected community development in large scale social networks. An efficient approach based on faster algorithm for obtaining close community structure was suggested due to Newman et al.[4]. A community may be subdivided into small sub communities whose formation and analysis performed by Ferrara E. [5] .The graph theoretical application for community designing and analysis was attempted by Fortunato S. [6].

Communication at the social networking platform when become highly frequent, close and intense then it reaches up to sentimental level. Deitrick et al. [7] suggested sentiment analysis approach on data obtained through social media platform. Leskovec et al.[8] considered several algorithms for network community detection. A methodological survey based contributions over community detection procedures are due to Plantie et al. [9] and Uthayasankar et al. [10]. This paper focuses on developing parameter estimation approach as a posterior application to the detected community.

II. Graph Based Rules

The methodology of community detection targets to the detection of groups of vertices within which connections are dense. Consider a graph G which is set of vertices V (G), and set of Edges E (G). One can construct rules for cliques and kernel formation based on collection of vertices and corresponding edges as under.

III. Community detection in weighted graph

The clique is referring to a kind of cohesive sub structure whose maxima provide a tool for community detection. The overlapping maximal clique is kernel. In view to N.Du, et al.[3] some of rule are as under:

Rule 1. $S \subseteq V(G)$, $\forall u,v \varepsilon S, u \neq v$, such that $(u,v) \varepsilon E$, then S is a clique in G. if any other S′ is a clique and $S' \supseteq S$ iff S′= S, S is a maximal clique of G.

Rule 2. For a given vertex v, N(v)= {u|(v,u) ε E (G) }, we call N (v) is the set of all neighbors of v. Given set $S \subseteq V(G)$, $N|s = \cup N(v_i) - S$, $V_i \varepsilon S$, N|s is the set of all neighbors of S.

Rule 3. Let Com (G) be the set of all components in G. the giant component is denoted by C_g and M (C_g) is the set of all the maximal cliques of C_g. We use $V_m \subseteq V(G)$ to represent the set of all vertices covered by M (C_g).

Rule 4. Let $P_0, P_1, --, -- P_{n-1}$ be the sub graph of G such that $\forall P_i, P_j$, $V(p_i) \cap V(P_j) = \phi$, and $V(P_0) \cup, ---, V(P_{n-1}) = V(G)$. For any pair of P_i and P_j , if $|E(P_i)| > |(N|P_i \cap P_j)|$, P_i is defined as a community of G.

Rule 5. Given vertex $v_i \varepsilon V_m$, define $C_i = \{S|S \varepsilon M(C_g), V_i \varepsilon S\}$ to be the set of all maximal cliques containing V_i and C the set of all $C_{i'}$'s. $\forall C_i, C_j \varepsilon C$,if $\frac{|Ci \cap cj|}{cj} >= f$ which is a threshold to describe the extent to which C_i overlaps with C_j, we call C_j is contained in C_i , denoted by $C_j < C_i$ .If c_i is not contained by any other element in C, C_i is called the kernel of G and Vi is the center of C_i .

Rule 6. Let K be the set of all kernels in G. $V_k = \{ V_i \mid V_i \varepsilon K, K_j \varepsilon K \}$ is the set of all vertices covered by K. and $I_k = \cup (K_i \cap k_j)$, $k_i, k_j \varepsilon K$, $i \neq j$ is the union of all the vertices that any pair of element in K has in common.

IV. Problem undertaken

Assume, using any of existing algorithms several communities have been detected. One may be interested to estimate unknown parameter of characteristics associated with edge between any pair of vertices, within the community formed in graphical population structure of a social media platform in the setup of big data. For example, large numbers of registered users are on social networking platform then the average time consumed between any pair of users within a community is a problem to work out. Being a large data setup, growing fast over time and space, the estimation of such is time and cost consuming. This paper considers a solution approach for a problem described herein using sampling procedure.

V.A Graphical Structure:

Assume a fig. 1 where enumeration of cliques is taken into consideration. Among constituted cliques, there exist maximal clique which is a complete sub graph which can represent closed relationship for single entity in a given network.

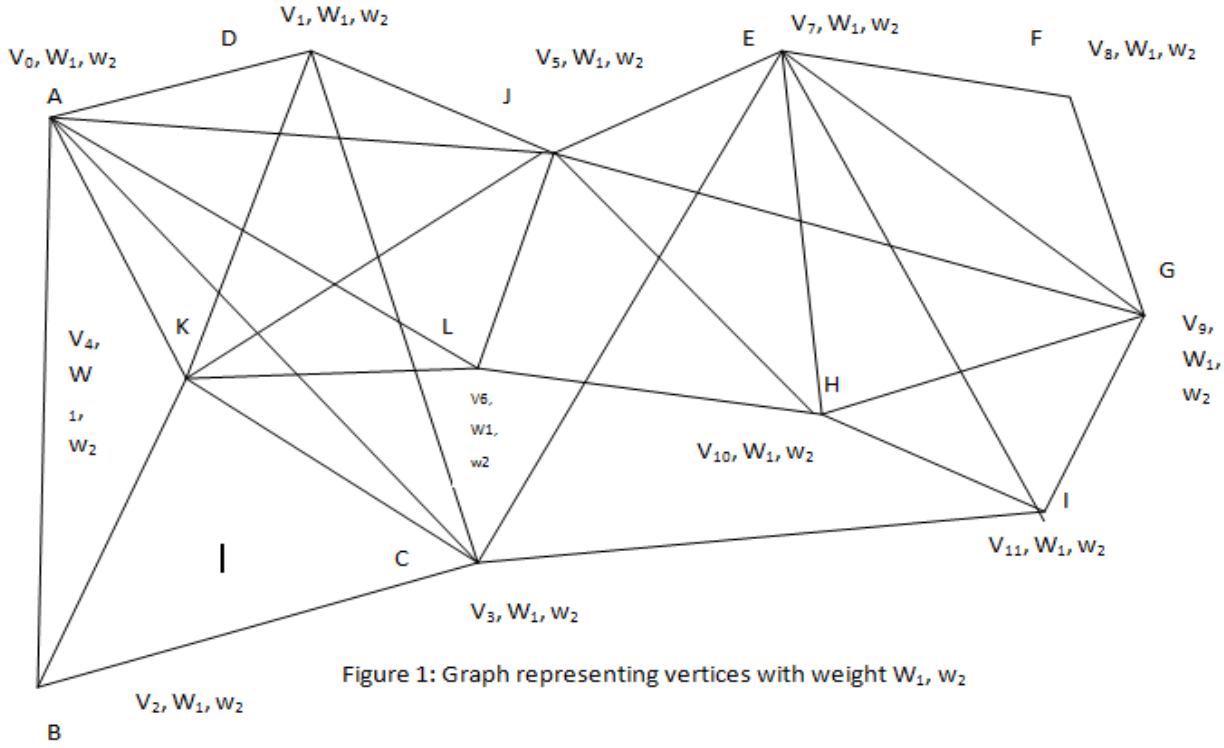


Figure 1: Graph representing vertices with weight W_1, w_2

For enumerate the cliques of a graph using rules 1-5: one can get:

$\mathbf{C_0}$ = {(V_0, W_1,w_2), (V_1,W_1,w_2), (V_4,W_1,w_2), (V_5,W_1,w_2)}, {(V_0, W_1,w_2), (V_1,W_1,w_2), (V_3,W_1,w_2), (V_4,W_1,w_2)}, {(V_0, W_1,w_2), (V_2,W_1,w_2), (V_3,W_1,w_2), (V_4,W_1,w_2)}, {(V_0, W_1,w_2), (V_6,W_1,w_2), (V_4,W_1,w_2), (V_5,W_1,w_2)} V_0 being as the center.

$\mathbf{C_1}$={(V_0, W_1,w_2), (V_1,W_1,w_2), (V_4,W_1,w_2), (V_5,W_1,w_2)}, {(V_0, W_1,w_2), (V_1,W_1,w_2), (V_3, W_1, w_2),(V_4,W_1,w_2)}

$\mathbf{C_2}$ = {(V_0, W_1, w_2), (V_2, W_1, w_2), (V_4, W_1, w_2), (V_3, W_1, w_2)}

$\mathbf{C_3}$ = {(V_0, W_1, w_2), (V_2, W_1, w_2), (V_4, W_1, w_2), (V_3, W_1, w_2)},

$\mathbf{C_4}$={(V_0, W_1,w_2), (V_1,W_1,w_2), (V_4,W_1,w_2), (V_6,W_1,w_2)}, {(V_0, W_1,w_2), (V_2,W_1,w_2), (V_3,W_1,w_2), (V_4, W_1, w_2)}, {(V_1, W_1, w_2), (V_4, W_1, w_2), (V_5, W_1, w_2), (V_6, W_1, w_2)}

$\mathbf{C_5}$= {(V_5, W_1, w_2), (V_1, W_1, w_2), (V_4, W_1, w_2), (V_6, W_1, w_2)}

$\mathbf{C_6}$= {(V_5, W_1, w_2), (V_1, W_1, w_2), (V_4, W_1, w_2), (V_6, W_1, w_2)}

$\mathbf{C_7}$={(V_7, W_1,w_2), (V_8,W_1,w_2), (V_9,W_1,w_2), (V_{10},W_1,w_2)}, {(V_7, W_1,w_2), (V_9,W_1,w_2), (V_{11},W_1,w_2), (V_{10}, W_1, w_2)},

$\mathbf{C_8}$ = {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

$\mathbf{C_9}$= {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

$\mathbf{C_{10}}$= {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

$\mathbf{C_{11}}$= {(V_7, W_1, w_2), (V_8, W_1, w_2), (V_9, W_1, w_2), (V_{10}, W_1, w_2)}

C_8, C_9 C_{10}, C_{11} are contained by C_7.

Therefore C_0 and C_7 are two different kernels respectively with weight associated with vertices.

VI. Parameter estimation

Consider the following graph in figure 3 where first weight the age of the users registered in the social networking sites and the other weight is the number of hours of the social networking sites used. In figure 2, social media communities detected through algorithms and unknown parameters existence are given from which one can extract sample based implementation.

Community 1	**Community2 ……**	**Community p**
Unknown parameter θ_1, θ_2, θ_3, θ_4,	Unknown parameter θ_1, θ_2, θ_3, θ_4,	Unknown parameter θ_1, θ_2, θ_3, θ_4,
Sampl	Sampl	Sampl

Figure 2: Social media communities & unknown parameters

Consider the graph as population having kernel based k groups classification likes below:-

Table 1: Kernel based groups

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	…..K^{th}
K_{e1}	K_{e2}	K_{e3}	K_{e4}	K_{e5}	K_{e6}	K_{e7}	K_{e8}	K_{e9}	K_{e10}	K_{e11}	K_{e12}	….. K_{en}

VII. Kernel Sampling:

One can consider the graphical population of vertices (node) and edges G= (V, E) divided into k Kernel based groups, derived from given a graphical population (see table 1). This constitutes setup of Kernel Sampling. Assume the strata sizes are N_1, N_2, N_3……N_k such that $\sum_{i=1}^{k} N_i = N$

Let the total size of population is N from which a sample of size of population n (n< N) is drawn which is divided into Kernel based group wise as n_1, n_2, n_3,….n_k. Such that $\sum_{i=1}^{k} n_i$=n. Let the sample means are $\bar{m}_1$,$\bar{m}_2$, $\bar{m}_3$,…,…. $\bar{m}_k$ of the k strata respectively P.V. Sukhatme[11] and Cochran [12].

Consider vertices of graph G= (V, E) having two variables W_2: number of hours the user is consuming social media website is used in a month (auxiliary variable) and W_1: the age of user (in completer years) as main variable. The unknown parameter is average number of hours consumed by a user W_2. It may assume that mean age of users $\bar{W}_1$ in population is known (due to registration data while creating account on social networking sites). The i^{th} Kernel based group has size N_i and pair of values (W_{1ij}, W_{2ij}) where W_{1ij}, W_{2ij} are j^{th} value i^{th} Kernel based group relating to number of hours consumed by users and ages fo users.

$$\bar{W}_1=\frac{1}{N}=\sum_{i=1}^{K}\sum_{l=1}^{NI} W_{1ij} \quad \text{(Known parameter)} \tag{4.1}$$

$$\bar{W}_2=\frac{1}{N}=\sum_{i=1}^{K}\sum_{j=1}^{NI} W_{2ij} \quad \text{(Unknown parameter and to be estimated)} \tag{4.2}$$

Moreover some other symbols are as under:

$\bar{W}_{1i}$: Population mean of i^{th} strata of variable W_1

$\bar{W}_{2i}$: Population mean of i^{th} strata of variable W_2

Estimation method under Kernel Sampling:

To estimate unknown$\bar{W}_2$, the random samples of sizes n_i are drawn from i^{th} group N_i paired values (w_{1ij} , w_{2ij}) such that

$$\bar{w}_{1i} = \frac{1}{n_i}\sum_{j=1}^{ni} w_{1ij} \tag{4.3}$$

$$\bar{w}_{2i} = \frac{1}{n_i}\sum_{j=1}^{ni} w_{2ij} \tag{4.4}$$

and (w_{1ij} , w_{2ij}) are pair of sample observations from i^{th} group

Method to use for estimation of $\bar{W}_2$is

M = $\sum_{i=1}^{k} \phi(z_i, \acute{z}_i)\bar{W}_{2i}$, where ϕ(z_i ,z'_i)= (z_i .z'_i) and $z_i = \bar{w}_{1i}$, $\acute{z}_i = \frac{1}{\bar{W}_{2i}}$ and $\bar{W}_{2i}$ assumed known. (4.5)

The Mean Square Error of method M is

MSE (M)= $\sum_{i=1}^{k} Z_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right)(S_i^*)^2$ (4.6)

$R=\bar{W}_1/\bar{W}_2$; $Z_i=\frac{N_i}{N}$; (4.7)

Where $(S_i^*)^2 = [S_{iW1}^2+R^2S_{iW2}^2-2RS_{iw1w2}]$

$S_{iW1}^2=\frac{1}{N_i-1}\sum_{i=1}^{K}(W_{1ij}-\bar{W}_{1i})^2$, $S_{iW2}^2=\frac{1}{N_i-1}\sum_{i=1}^{K}(W_{2ij}-\bar{W}_{i2})^2$; (4.8)

$\bar{W}_{1i}=\frac{1}{N_1}\sum_{j=1}^{N1}W_{1ij}$; $\bar{W}_{2i}=\frac{1}{N_2}\sum_{j=1}^{N2}W_{2ij}$; $S_{iw1w2}=\frac{1}{N_{i-1}}\sum_{i=1}^{K}(W_{1ij}-\bar{W}_{1i}).(W_{2ij}-\bar{W}_{i2})$ (4.9)

The estimate of $(S_i^*)^2$is est(MSE) = $\sum_{i=1}^{k}Z_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)(S_i^*)^2$

Where $(S_i^*)^2$ =$[S_{iW1}^2+R2S_{iW2}^2-2RSiw1w2]$, the S_{iW1}^2, S_{iW2}^2, Siw1w2 are estimated from sample and r=$\frac{\bar{w}_1}{\bar{w}_2}$ exist in sample.

The 95% Confidence interval for estimating $\bar{W}_1$ is:
P [M- 1.96 $\sqrt{MSE(M)}$< M+1.96 $\sqrt{MSE(M)}$]=0.95 (4.10)

VIII. Simulation procedure for confidence interval

Step I: Draw a random sample of size n
Step II: Compute the lower limit and upper limit of confidence interval
Step III: Repeat step I and II for k times (K=200)
Step IV: Compute the less than type and more than type cumulative frequency over all k samples for lover limit and upper limit of confidence interval.
Step V: Plot data of step IV on graph. The perpendicular from point of intersection on the x-axis is the simulated value of lover limit and upper limit of confidence interval for parameter to be estimated.

IX. Numerical illustration:

Consider figure 2 having 11 vertices and consisting of data in the tuple (Vi, W_{1i}, W_{2i}). The relationship of vertices is in the form of edges which is used to constitute form clique and kernel.

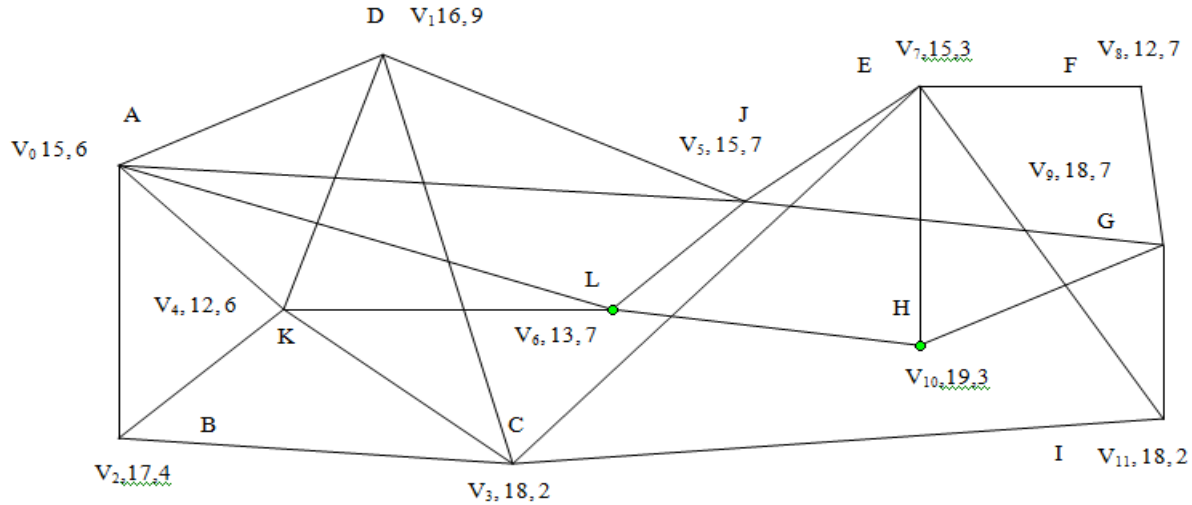


Figure 3: Graph with weight representing Age and hours of use.

The figure has 2 kernels C_0 and C_7.The Kernel constituted based group structure of graphical population is as under. From figure 4 we are extracting samples from group 1(C_0) and group 2(C_7).

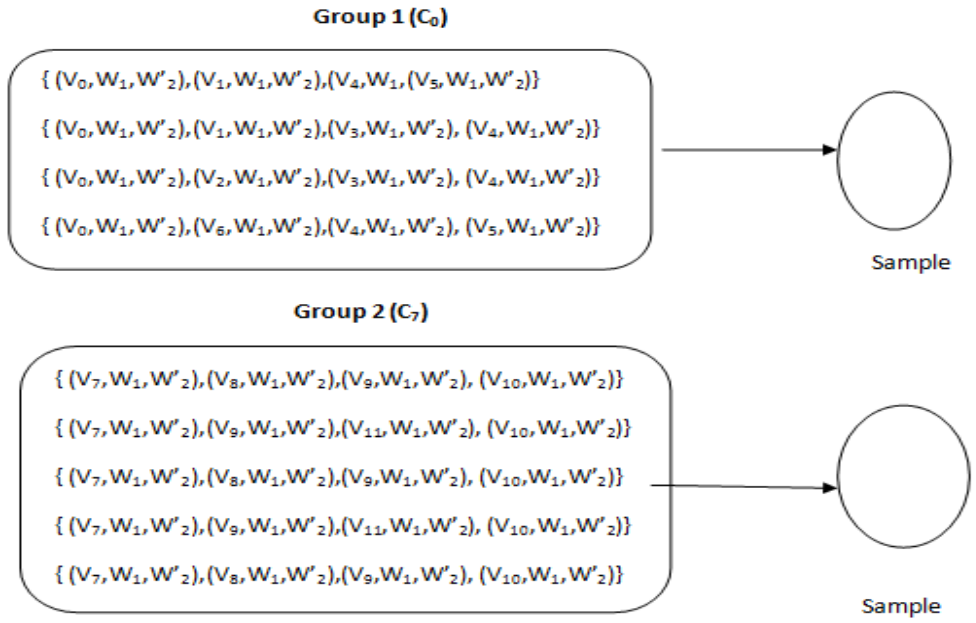


Figure 4: Representing Kernels and Samples

As per figure 3, the representation of the vertices with weight (W1: ages of users) and (W2: time consumed by users) are given below in terms of V_i=(W_{1i} ,W_{2i})

V_0=(15,6) , V_1=(16,9), V_2=(17,4) V_3=(18,2) V_4=(12,6) V_5=(15,7)

V_6=(13,7), V_7=(15,3), V_8=(12,7), V_9=(18,7) V_{10}=(19,3), V_{11}=(18,2)

The group 1 Kernels contains 16 tuple (N_1=16) and group 2 contains 20 tuple (N_2=20).

A random sample of size n_1=6 is drawn from N_1=16. Similarly, random sample of size n_2 is drawn from N_2=20 ($n_1 < N_1$, $n_2 < N_2$). Using these sample values, the objective is to estimate unknown population mean$\overline{W}_1$.

Table 2: Description of population parameters

	Group size N_i	Group mean W_1	Group mean W_2
Group I	N_1=16 $Z_1=N_1/N$ =0.44	$\bar{W}_{1G1}$= 14.75 S^2_{W1G1}=4.2	$\bar{W}_{2G1}$=5.93 S^2_{W2G1}=3.02
Group II	N_2=20 $Z_2=N_2/N$ =0.55	$\bar{W}_{1G2}$=16.6 S^2_{W1G2}=6.25	$\bar{W}_{2G2}$=4.5 S^2_{W2G2}=4.47
	$N=N_1+N_2$=36	R= $\frac{\bar{W}_1}{\bar{W}_2}$=14.75/16.6=0.88	

Table 3: Sample based computation (First Sample)

	Sample size	Sample values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_0,15,6),(V_1,16,9),(V_4,12,6),(V_5,15,7),(V_3,18,2),(V_6,13,7)	$\bar{w}_{1G1}$=14.83 $S^2_{W1G1}=4.51$	$\bar{w}_{2G1}$=6.16 S^2_{W2G1}=5.47	r_1=2.4	$(S_1^*)^2$=12.92	[8.70-19.67]
Group II	n_2=4	(V_7,15,3),(V_8,12,7) (V_9,18,7),(V_{10},19,3)	$\bar{w}_{1G2}$=16.0 $S^2_{W1G2}=10$	$\bar{w}_{2G2}$=5.0 $S^2_{W2G2}=5.33$	r_2=3.2	$(S_2^*)^2$=131.2	
M=14.19			Est.(MSE)=7.89				

Table 4: Sample based computation (Second sample)

	Sample size	Sample values(V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est. .$(S_i^*)^2$	
Group I	n_1=6	(V_0,15,6),(V_4,12,6), (V_5,15,7) (V_3,18,2) ,(V_6,13,7), (V_1,16,9)	$\bar{w}_{1G1}$=14.83 S^2_{W1G1}=4.51	$\bar{w}_{2G1}$=6.16 S^2_{W2G1}=5.47	r_1=2.40	$(S_1^*)^2$=12.82	[10.98-17.98]
Group II	n_2=4	(V_7,15,3) ,(V_{11},18,2) (V_8,12,7), (V_9,18,7)	$\bar{w}_{1G2}$=15.75 S^2_{W1G2}=8.25	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.31	$(S_2^*)^2$=52.31	
M=14.48			Est.(MSE)=3.37				

Table 5: Sample based computation (Third sample)

	Sample size	Sample Values(V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_3,18,2) (V_4,12,6) (V_5,15,7) (V_6,13,7) (V_1,16,9) (V_0,15,6)	$\bar{w}_{1G1}$=14.83 $S^2_{W1G1}=4.51$	$\bar{w}_{2G1}$=6.16 S^2_{W2G1}=5.47	r_1=2.40	$(S_1^*)^2$=12.77	[13.19-22.47]
Group II	n_2=4	(V_7,15,3) (V_{11},18,2) (V_9,18,7) (V_{10},19,3)	$\bar{w}_{1G2}$=17.5 S^2_{W1G2}=3.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.66	$(S_1^*)^2$=52.57	
M=17.831			Est.(MSE)=5.56				

Table 6: Sample based computation (Fourth Sample)

	Sample size	Sample Values(V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_5,15,7) (V_4,12,6) (V_6,13,7) (V_0,15,6) (V_1,16,9) (V_2,17,4)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1} = 3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=22.75	[8.74-18.84]
Group II	n_2=4	(V_{10},19,3) (V_9,18,7) (V_8,12,7) (V_7,15,3)	$\bar{w}_{1G2}$=16 S^2_{W1G2} = 10	$\bar{w}_{2G2}$=5 S^2_{W2G2}5.33	r_2=3.2	$(S_1^*)^2$=103.9	
M=13.79			Est.(MSE)=6.66				

Table 7: Sample based computation (Fifth Sample)

			Mean				95% C.I.
	Sample Size	Sample Values(V_0,w_1,w_2)	w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6), (V_3,18,2), (V_6,13,7) (V_2,17,4) (V_1,16,9) (V_0,15,6)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}5.86	r_1=2.67	$(S_1^*)^2$=69.31	[11.48-19.9]
Group II	n_2=4	(V_8,12,7) ,(V_9,18,7) (V_{10},19,3) ,(V_{11},18,2)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$=55.57	
M=15.69			Est.(MSE)=4.64				

Table 8: Sample based computation (Sixth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_0,15,6) (V_2,17,4) (V_5,15,7) (V_1,16,9) (V_3,18,2)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$=49.05	[7.91-23.81]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$=260.11	
M=15.86			Est.(MSE)= 16.52				

Table 9: Sample based computation (Seventh Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_0,15,6) (V_1,16,9), (V_4,12,6) (V_5,15,7) (V_2,17,4) (V_3,18,2)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$=49.43	[9.48-21.46]
Group II	n_2=4	(V_7,15,3) (V_8,12,7) (V_9,18,7) (V_{10},19,3)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=5.41	r_2=3.36	$(S_1^*)^2$=140.67	
M=15.47			Est.(MSE)= 9.37				

Table 10: Sample based computation (Eighth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_3,18,2) (V_4,12,6), (V_5,15,7) (V_6,13,7) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=14.83 S^2_{W1G1}=5.44	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.54	$(S_1^*)^2$=55.55	[10.1,19.54]
Group II	n_2=4	V_7,15,3) (V_8,12,7) (V_9,18,7) (V_{11},18,2)	$\bar{w}_{1G2}$=15.75 S^2_{W1G2}=8.25	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.31	$(S_1^*)^2$=79.59	
M=14.82			Est.(MSE)=5.82				

Table 11: Sample based computation (Ninth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_3,18,2), (V_6,13,7) (V_0,15,6) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.67	$(S_1^*)^2$=59.16	[8.93-21.59]
Group II	n_2=4	(V_8,12,7) (V_{10},19,3), (V_9,18,7) (V_{11},18,2)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=5.00 S^2_{W2G2}=8	r_1=3.35	$(S_1^*)^2$=155.78	
M=15.26			Est(MSE)=10.45				

Table 12: Sample based computation (Tenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_6,13,7) (V_5,15,7), (V_4,12,6) (V_0,15,6) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=21.55	[8.86-23.98]
Group II	n_2=4	(V_7,15,3) (V_{10},19,3) (V_{11},18,2) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.26	$(S_1^*)^2$=242.52	
M=16.42			Est.(MSE)=14.95				

Table 13: Sample based computation (Eleventh Sample)

		Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_6,13,7) (V_5,15,7), (V_4,12,6) (V_3,18,2) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.60	$(S_1^*)^2$ =56.98	[8.87-22.11]
Group II	n_2=4	(V_8,12,7) (V_{10},19,3) (V_{11},18,2) (V_9,18,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$ =172.80	
M=15.49			Est.(MSE)=11.43				

Table 14: Sample based computation (Twelfth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2		est .$(S_i^*)^2$	
Group I	n_1=6	(V_2,17,4) (V_1,16,9), (V_0,15,6) (V_4,12,6) (V_5,15,7) (V_6,13,7)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=22.76	[8.19-19.39]
Group II	n_2=4	(V_7,15,3) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=5.00 S^2_{W2G2}=5.33	r_2=3.2	$(S_1^*)^2$=129.42	
M=13.79			Est.(MSE)=8.19				

Table 15: Sample based computation (Thirteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_2,17,4) (V_1,16,9), (V_3,18,2) (V_5,15,7) (V_6,13,7) (V_0,15,6)	$\bar{w}_{1G1}$=15.66 S^2_{W1G1}=3.06	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.68	$(S_1^*)^2$= 39.85	[9.23-22.19]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$= 172.80	
M=15.71			Est.(MSE=11.10				

Table 16: Sample based computation (Fourteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_5,15,7) (V_4,12,6), (V_6,13,7) (V_3,18,2) (V_2,17,4) (V_1,16,9)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.83 S^2_{W2G1}=6.15	r_1=2.60	$(S_1^*)^2$=56.98	[9.82-24.82]
Group II	n_2=4	(V_{10},19,3) (V_{11},18,2) (V_7,15,3) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.26	$(S_1^*)^2$=229.74	
M=17.32			Est.(MSE)=14.85				

Table 17: Sample based computation (Fifteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_1,16,9) (V_0,15,6), (V_4,12,6) (V_5,15,7) (V_2,17,4) (V_3,18,2)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$ =49.43	[9.27-22.43]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$ =172.8	
M=15.85			Est.(MSE)=11.29				

Table 18: Sample based computation (Sixteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_6,13,7) (V_5,15,7), (V_4,12,6) (V_0,15,6) (V_2,17,4) (V_3,18,2)	$\bar{w}_{1G1}$=15.0 S^2_{W1G1}=5.2	$\bar{w}_{2G1}$=5.33 S^2_{W2G1}=3.85	r_1=2.81	$(S_1^*)^2$=66.45	[14.31-23.43]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_7,15,3)	$\bar{w}_{1G2}$=17.50 S^2_{W1G2}=3.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.66	$(S_1^*)^2$=70.39	
M=18.87			Est.(MSE)=5.46				

Table 19: Sample based computation (Seventeenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2		r_i	
Group I	n_1=6	(V_5,15,6) (V_4,12,6), (V_6,13,7) (V_2,17,4) (V_1,16,9) (V_0,15,6)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=26.08	[8.84-24]
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_7,15,3) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.91	r_2=4.26	$(S_1^*)^2$=242.52	
M=16.42			Est.(MSE)= 15.04				

Table 20: Sample based computation (Eighteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_3,18,2), (V_2,17,4) (V_1,16,9) (V_0,15,6) (V_5,15,7)	$\bar{w}_{1G1}$=15.5 S^2_{W1G1}=4.3	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.73	$(S_1^*)^2$=-0.476	[9.35-22.35
Group II	n_2=4	(V_{11},18,2) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.75 S^2_{W1G2}=10.24	$\bar{w}_{2G2}$=4.75 S^2_{W2G2}=6.91	r_2=3.52	$(S_1^*)^2$=172.80	
M=15.85			Est.(MSE)=10.37				

Table 21: Sample based computation (Nineteenth Sample)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_4,12,6) (V_3,18,2), (V_2,17,4) (V_1,16,9) (V_0,15,6) (V_6,13,7)	$\bar{w}_{1G1}$=15.16 S^2_{W1G1}=5.36	$\bar{w}_{2G1}$=5.66 S^2_{W2G1}=5.86	r_1=2.67	$(S_1^*)^2$=60.92	[9.05-20.73]
Group II	n_2=4	(V_7,15,3) (V_{10},19,3) (V_9,18,7) (V_8,12,7)	$\bar{w}_{1G2}$=16.00 S^2_{W1G2}=10.0	$\bar{w}_{2G2}$=5.00 S^2_{W2G2}=5.33	r_2=3.2	$(S_1^*)^2$=129.42	
M=14.89			Est.(MSE)=8.90				

Table 22: Sample based computation (Twenty Samples)

	Sample size	Sample Values (V_0,w_1,w_2)	Mean				95% C.I.
			w_1	w_2	r_i	est .$(S_i^*)^2$	
Group I	n_1=6	(V_2,17,4) (V_1,16,9), (V_4,12,6) (V_5,15,7) (V_6,13,7) (V_0,15,6)	$\bar{w}_{1G1}$=14.66 S^2_{W1G1}=3.46	$\bar{w}_{2G1}$=6.5 S^2_{W2G1}=2.7	r_1=2.25	$(S_1^*)^2$=22.96	[13.4-21.42]
Group II	n_2=4	(V_9,18,7) (V_7,15,3) (V_{10},19,3) (V_{11},18,2)	$\bar{w}_{1G2}$=17.50 S^2_{W1G2}=3.0	$\bar{w}_{2G2}$=3.75 S^2_{W2G2}=4.9	r_2=4.66	$(S_1^*)^2$=63.47	
M=17.41			Est.(MSE)=4.23				

Table 23: For Confidence interval calculations

For lower limit of Confidence Interval				For upper limit of Confidence Interval			
Class Interval	Probability over 200 samples	LTT	MTT	Class Interval	Probability over 200 samples	LTT	MTT
Below 8.0	0.08	0.08	1.00	Below 17.0	0.05	0.05	1.00
8.0-9.0	0.38	0.46	0.92	17.0-18.0	0.13	0.18	0.95
9.0-10.0	0.29	0.75	0.54	18.0-19.0	0.16	0.34	0.82
10.0-11.0	0.14	0.89	0.25	19.0-20.0	0.18	0.52	0.66
11.0-12.0	0.04	0.93	0.11	20.0-21.0	0.23	0.75	0.48
12.0-13.0	0.05	0.98	0.07	21.0-22.0	0.14	0.89	0.25
13.0-14.0	0.01	0.99	0.02	22.0-23.0	0.09	0.98	0.11
Above 14.0	0.01	1	0.01	23.0-24.0	0.01	0.99	0.02
LTT: Less Than Type; MTT: More Than Type				Above 24.0	0.01	1.00	0.01

Probability = $\left(f_i \middle/ \sum f_i\right)$; f_i : frequency of i^{th} class interval

$\sum f_i$: total frequency; P[A]= probability of event A.

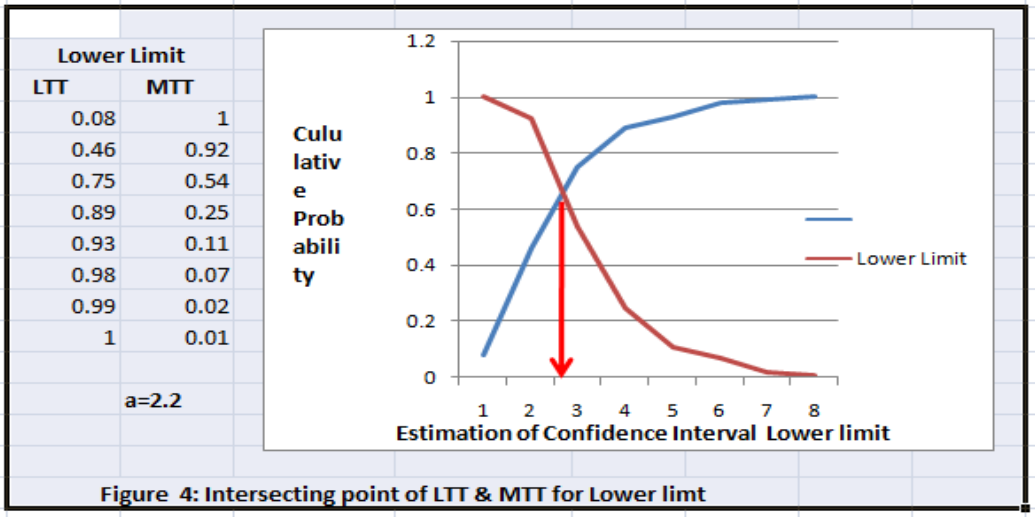


Figure 4: Intersecting point of LTT & MTT for Lower limt

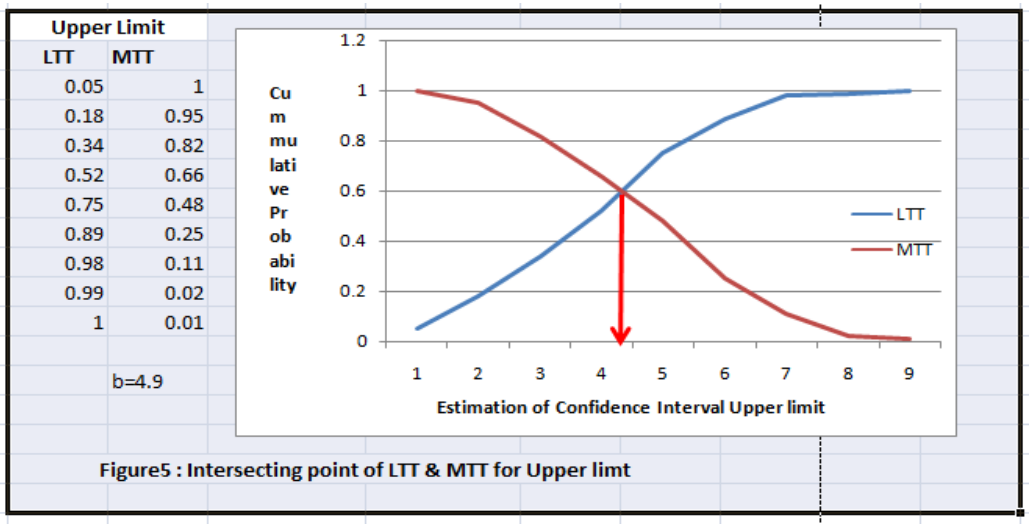


Figure5 : Intersecting point of LTT & MTT for Upper limt

a=2.2, b=4.9

Confidence Interval = P[a<2.3<b]=0.95 ; where P[A] is probability of event A.

Other Computations: - $(S_1^*)^2$ =17.64, $(S_2^*)^2$= 39.08

MSE (M) = $Z_1^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right)(S_1^*)^2 + Z_2^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right)(S_2^*)^2$ = 4.70

X. Conclusion

In this paper, a graphical structure of population has been considered and using the Kernel creation procedure rules and closed communities have been detected. The closeness is based on criteria of click formation. In order to estimate the unknown population parameter (average hours used) a scheme named after as Kernel Sampling estimation method is used. The 95% confidence intervals have been computed. It has been found that 95% confidence intervals are catching the true values. The simulation procedure suggested herein provides the well predicted estimated interval. This contribution opens up avenues and opportunities to think for mixing of community detection and parameter estimation.

References:

[1]. Dongsheng Duan YUhua Li et al. "Community Mining on Dynamic Weighted Directed Graphs", CNIKM 09 November 6, 2009, Hong Kong, China.

[2]. C. Pizzuti, "Community detection in social networks with genetic algorithms" Annual conference on Genetic and Evolutionary Computation, pages1137-1138, 2008.

[3]. Nan. Du, B.Wu, Xin Pei et al. "Community detection in large scale social networks" , SNA-KDD, pages 16-25, August 12, 2007, California, USA.

[4]. M.E.J. Newman. "Fast algorithm for detecting community structure in networks" Phys Rev E Stat Nonilin Soft Matter Phys, 2004.

[5]. Ferrara E. "A large-scale community structure analysis in Face-book. EPJ Data Sci ;1(9),2012

[6]. Fortunato S. Community detection in graphs. Phys Rep 2010; 4863(3-5):75-174.

[7]. Deitrick W, Valyou B, et al. "Enhancing sentiment analysis on twitter using community detection". Communications and Network 2013;5(3):192-7.

[8]. Leskovec J, Lang KJ, Mahoney MW. Empirical comparison of algorithms for network community detection". In: International conference on World Wide Web (WWW); 2010.p. 631-40.

[9]. Plantie M, Crampes M. "Survey on social community detection", In: Social media retrieval, computer Communications and Networks. Springer: 2013. P 65-85.

[10]. Uthayasankar Sivarajah et al. "Critical analysis of Big data challenges and analytical methods", Journal of Business research, (2017), P 263-286.

[11]. P.V. Sukhatme, B.V.Sukhatme et al. "Sampling Theory of Surveys with Applications", IOWA State University Press and Indian Society of Agricultural Statistics (New Delhi), 1984.

[12]. Cochran W.G (2005) , Sampling Techniques, John Willey and Sons, New York.

RELIABILITY OF A BIG CITY SEWER NETWORK

Baranov L. A.
•
Russian University of Transport (MIIT)
baranov.miit@gmail.com

Ermolin Y.A.
•
Russian University of Transport (MIIT)
ermolin.y@yandex.ru

Shubinsky I. B.
•
NIIAS
igor-shubinsky@yandex.ru

Abstract

The ramified sewerage system for receiving and transferring household and industrial sewage typical for a large city is considered. Consideration is restricted to the sub-system of sewage conveyance (sewer network). A sewer network is defined as a combination of underground pipes (sewers) passing sewage through the force of gravity. A review of the literature reveals that there is currently no universally acceptable definition or measure for the reliability of urban sewer network. The aim of this article is to propose the physically obvious reliability index, and to develop an engineering methodology for its calculating. The relative raw sewage volume that could be potentially discharged to the environment as a result of component failures in the sewer network is proposed as a measure of overall system reliability. A simple method for quick and proper calculation of this volume is presented. The basis for this method is a representation of the sewer network by a combination of Y-like fragments. Each such fragment is formally substituted by a fictitious equivalent sewer that has a failure rate leading to the same output for the same input. A sequential application of this approach reduces the problem of estimating the discharged sewage volume to an elementary sub-problem with a simple solution is. The proposed approach is based on the reliability theory. The notions "failure flow" and "repair flow" are used. These flows are taken stationary with known parameters. Numerical examples are used to demonstrate the proposed approach.

Keywords: Sewer network; Reliability; Sewage discharge; Y-like network fragment; Decomposition-equivalence method.

I. Introduction

The proper functioning of the urban sewage disposal system is a primary determinant of the city's ecological and sanitary-hygienic conditions. Confronting problems associated with the sewer network maintenance as a subsystem of an entire sewage disposal system, is a necessary step for improving operation efficiency in an urban waste water disposal system as a whole. In recent years, in response to increasing congestion in urban sewer networks and the adverse environmental impact of such congestion, substantial attention has been focused on working out the proposals to improve waste water disposal processes. A critical issue in the evaluation and effective implementation of

these proposals is the development of the best, in some specified sense, sewage disposal strategies. In practice sometimes, very significant improvements in management efficiency could be accomplished simply by better maintenance of the waste water disposal system.

There is a great deal of research dedicated to the reliability problems of water supply systems reported in the literature [1-6]. From the latest publications we emphasize the work [2], which provides an in-depth review of the relevant research literature in the context of the mathematical methods for measuring water distribution system reliability. However, as note in other works, for example, [4]: "A review of the literature reveals that there is currently no universally acceptable definition or measure of the reliability of water distribution systems … For a large system … it is extremely difficult to analytically compute the mathematical reliability".

By contrast, the reliability problems of the sewage disposal systems are still uninvestigated [7-13]. Therefore, any effort to comprehend, set up and refine the issue of sewer network reliability takes on great significance. The final objective of these investigations is to develop sewer network design, reconstruction and maintenance methods with due regard for reliability.

II. Short description of the object and problem statement

An urban wastewater disposal system is a network of structurally and technologically interconnected structures intended for sewage collection and its conveyance to the purification facilities.

Usually the city sewage disposal system is designed and constructed according to the head-and-gravity concept. This means that the sewage passes through underground sewers having a specified fall by gravity, and pumping stations lift sewage in areas where gravity flow is impossible. (As a rule, the sewage pumping station is designed as a system providing a redundancy of the pumping equipment. Because of this, in the following, we assume that the pumping stations are absolutely reliable). By this means the sewer network, by nature, is a peculiar water distribution system. The reliability of such systems is often defined by heuristic guidelines, like having all pipe diameters greater than a minimum prescribed value. By using such guidelines it is implicitly assumed that reliability will be assured, but the level of reliability provided is not quantified or measured. Thus, the question: "Is the system reliable?" is usually well understood and easy to answer, while the question "What is its reliability level?" is not straightforward. As a result, only limited confidence can be placed on such rules, as reliability is not considered explicitly.

The underground pipes, as sewer network components, are subject to so many influences that it is difficult, if not impossible, to predict their combined effect in advance. These influences include the corrosive action of the soil and sewage, ground movements, the weather, etc. Most of these factors are random, and are characterized by significant variability. These circumstances adversely affect sewer network reliability. Currently, traditional wastewater disposal system design and maintenance methods usually fail to account for this situation.

A determination of the timeline and the sequence of a sewerage modernization plan is an important problem of the applied reliability theory. The strategy development for the object reconstruction falls into two stages. At the first stage, the object technical condition is established, and a need for renovation is determined. The second stage is job scheduling for the specific network elements requiring repair or replacements.

Depending on the purpose of the study and the specifics of an object, its technical state may be estimated, from the viewpoint of a reliability, using different quantitative measures: for example, by the average time between failures or by the probability of trouble-free functioning over a given period of time. We note that these traditional measures accepted in theory, as applied to sewer networks, provide not enough information because it is very hard to interpret them physically.

Here the specific reliability index is proposed. This index is intended for functioning

efficiency estimation of tree-like hierarchical structures; an urban sewer network is a typical representative of such structure. The damage due to raw sewage discharged to the environment resulting from sewer network failures is considered as a quantitative reliability measure [7].

At the moment there are no universally accepted procedures for assessment of the economic and ecological damage due to raw sewage discharge resulting from sewer network failures. Within any particular region, sewer basin or city district this damage and the methods of assessing it may differ significantly and may change with the time. What is considered acceptable for one area or time period may not be appropriate for another area or time. In any case however, it is evident that this damage is dependent on the volume of raw sewage discharged to the environment (in actual practice this sewage is usually pumped over into a suitable nearly manhole by a mobile emergency pumping plant).

For this reason, the volume of raw sewage potentially discharged from the sewer network to the environment over some time period (for example, one year) may be taken as a measure of the damage caused by the network unreliability, and, therefore, as an indirect measure of the sewer network reliability.

Thus, the problem reduces to finding of the raw sewage volume potentially discharged from the sewer network.

III. Reliability analysis of the sewer network fragment

Systems like a urban sewer network, are often described in terms of a graph, with links representing the pipes (sewers), and nodes representing connections between the pipes. The behavior of a sewer network is governed by the physical laws that describe the flow relationships in the pipes (laws of conservation), and the network layout.

Two features of a sewer network should be pointed out: 1) a sewage gravitates through each sewer in one direction only, and 2) the hydraulic elements used to link different sewer basins are lacking. This means particularly that the sewage entering into any network inlet, may be piped to a certain its outlet by a strictly specified sequence of sewers, i.e. along the only path. Thus, mathematically, the graph of an urban sewer network is a simply connected, oriented, and acyclic graph; in theory such graphs are also known as tree-like graphs.

We consider the three-component Y-like sewer network fragment shown in Fig. 1.

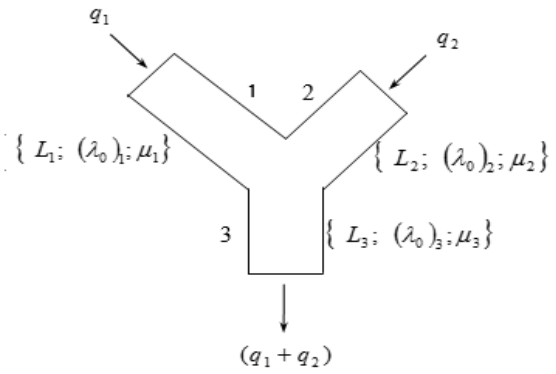


Figure. 1: *Y-like sewer network fragment.*

Each enumerated sewer of this fragment is characterized by its length $(L_1; L_2; L_3)$; in addition, we suppose that the unidirectional sewage flow rates in the inlets of sewer 1and 2 (q_1 and q_2, respectively) are known, constant and equal to the mean values calculated by averaging the

historical data obtained over a long period of time. We assume that, from time to time, each sewer fails, is repaired and, thereafter, put back in service again. Thus, each sewer can be either up (operable) or down (failed). In terms of the reliability theory, this means that so-called failure and repair flows are both acting on each sewer.

For a mathematical description of these flows, what is meant by term "failure" must be ascertained.

The exact definition of failure is somewhat fluid and depends on the level of detail of the required analysis, and has a variety of meanings to different individuals. In actual practice, a disturbance of the normal operation of the sewer can be manifested as a reduction of its capacity caused by cracks in the pipe, sewer breaks under extreme mechanical load, increasing rates of infiltration, repeated overflows, etc. Here, we shall define "failure" as an event implying a need for immediate overhaul or replacement of the pipe. In other words, the failure of a sewer is defined as an event when the sewer capacity becomes equal to zero, and consequently, all sewage entering into the sewer discharges to the environment.

The repair is taken here to mean that a renewal process reaches completion and the sewer is returned to service.

Usually, such events are documented with accompanying parameters. This information is systematically renewed, statistically processed and stored in relevant data bases. In the following, we assume that these data (in particular, the mean time to failure and mean time to repair) are known and available for analysis.

Both of these flows are characterized by their rates. Physically, the failure rate is the mean number of failures in a unit of time. The repair rate is defined similarly. In line with a much used assumption, we suppose that the failure flow as well as the repair flow are exponentially distributed flows [4]. From this it follows that the specific failure rate (the failure rate per unit sewer length) for each sewer $((\lambda_0)_1,(\lambda_0)_2,(\lambda_0)_3$, respectively) is constant. Analogously, the repair rates for sewers 1, 2 and 3 (μ_1,μ_2,μ_3) are constant as well. We assume that all these values are given.

The problem is stated as follows: given the values of all quantities listed above, it is necessary to estimate the volume of raw sewage discharged from the sewer network to the environment over some time period (one year in this study).

In order to solve this problem, we must first bring out the possible states of the system taken as a whole. These states are enumerated and listed below; what is meant by each state is explained in parentheses, and, next, the associated probability p_i of the system residing in state i is introduced:

0: (sewers 1, 2 and 3 up) - p_0;

1: (sewer 1 down, sewers 2 and 3 up) - p_1;

2: (sewer 2 down, sewers 1 and 3 up) - p_2;

3: (sewer 3 down, sewers 1 and 2 up) - p_3;

4: (sewers 1 and 2 down, sewer 3 up) - p_4;

5: (sewers 1 and 3 down, sewer 2 up) - p_5;

6: (sewers 2 and 3 down, sewer 1 up) - p_6;

7: (sewers 1, 2 and 3 down) - p_7.

With time, under the influence of failure and repair flows, the system goes from one state to another accidentally. This process is conveniently described by the use of the state space graph [14] (see Fig. 2), in which the possible system states are represented by circles with their number inside. The arrows indicate the transitions between states. The associated failure or repair rate is placed by

an arrow; in this case $\lambda_1 = (\lambda_0)_1 L_1, \lambda_2 = (\lambda_0)_2 L_2$, and $\lambda_3 = (\lambda_0)_3 L_3$.

Such a graph gives a descriptive idea of the changing system states. As an example, we consider state 4 (sewers 1 and 2 down, sewer 3 up). From this state the system departs to state 1 if the renewal process of sewer 2 reaches completion (the repair rate is μ_2), to state 2 when sewer 1 is returned to service (μ_1), and to state 7 when sewer 3 breaks down as well (λ_3). We note incidentally that the transition of the system, for example, from state 4 to state 0 is impossible due to the features of the exponentially distributed flow.

With the state space graph in hand it becomes possible to find all state probabilities $p_i(t)$ as functions of time. For this purpose so-called Kolmogorov's equations are formed [14].

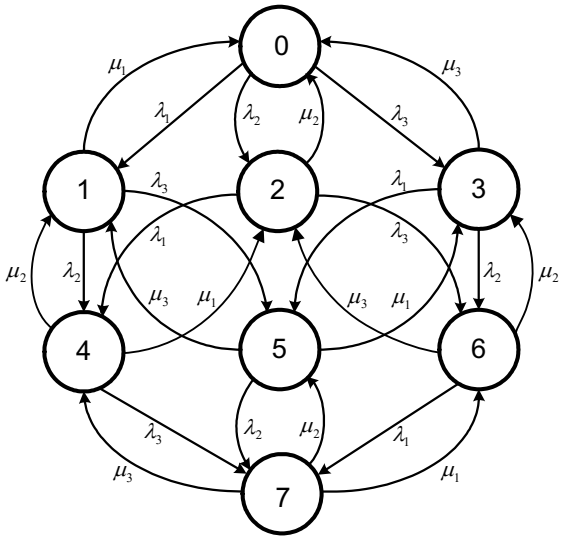


Figure. 2: *State space graph for the Y-like network fragment.*

For the graph shown in Fig. 2 these equations take the form:

$$\begin{cases} \frac{dp_0(t)}{dt} = \mu_1 p_1(t) + \mu_2 p_2(t) + \mu_3 p_3(t) - (\lambda_1 + \lambda_2 + \lambda_3) p_0(t), \\ \frac{dp_1(t)}{dt} = \lambda_1 p_0(t) + \mu_2 p_4(t) + \mu_3 p_5(t) - (\lambda_2 + \lambda_3 + \mu_1) p_1(t), \\ \frac{dp_2(t0}{dt} = \lambda_2 p_0(t) + \mu_1 p_4(t) + \mu_3 p_6(t) - (\lambda_1 + \lambda_3 + \mu_2) p_2(t), \\ \frac{dp_3(t)}{dt} = \lambda_3 p_0(t) + \mu_1 p_5(t) + \mu_2 p_6(t) - (\lambda_1 + \lambda_2 + \mu_3) p_3(t), \\ \frac{dp_4(t)}{dt} = \lambda_2 p_1(t) + \lambda_1 p_2(t) + \mu_3 p_7(t) - (\lambda_3 + \mu_1 + \mu_2) p_4(t), \\ \frac{dp_5(t)}{dt} = \lambda_3 p_1(t) + \lambda_1 p_3(t) + \mu_2 p_7(t) - (\lambda_2 + \mu_1 + \mu_3) p_5(t), \\ \frac{dp_6(t)}{dt} = \lambda_3 p_2(t) + \lambda_2 p_3(t) + \mu_1 p_7(t) - (\lambda_1 + \mu_2 + \mu_3) p_6(t), \\ \frac{dp_7(t)}{dt} = \lambda_3 p_4(t) + \lambda_2 p_5(t) + \lambda_1 p_6(t) - (\mu_1 + \mu_2 + \mu_3) p_7(t). \end{cases} \quad (1)$$

When a system state is changed, transition processes on the probabilities $p_i(t)$ take place. But, as shown this in [15] for real values λ_i and μ_i, these processes are very rapid. Usually, engineering practice uses so-called stationary probabilities. The probability of the system residing in state i assumes that stochastic transition process is stationary. By this is meant that all probabilities are independent of time (otherwise, they are also known as the stationary or limiting probabilities [14]. They may be obtained from Eqs. (1) taking all derivatives with respect to time equal to zero. In line with a common procedure [14], we form a set of linear algebraic equations for stationary probabilities p_i:

$$\begin{cases} p_0(\lambda_1+\lambda_2+\lambda_3)=\mu_1 p_1+\mu_2 p_2+\mu_3 p_3; \\ p_1(\lambda_2+\lambda_3+\mu_1)=\lambda_1 p_0+\mu_2 p_4+\mu_3 p_5; \\ p_2(\lambda_1+\lambda_3+\mu_2)=\lambda_2 p_0+\mu_1 p_4+\mu_3 p_6; \\ p_3(\lambda_1+\lambda_2+\mu_3)=\lambda_3 p_0+\mu_1 p_5+\mu_2 p_6; \\ p_4(\lambda_3+\mu_1+\mu_2)=\lambda_2 p_1+\lambda_1 p_2+\mu_3 p_7; \\ p_5(\lambda_2+\mu_1+\mu_3)=\lambda_3 p_1+\lambda_1 p_3+\mu_2 p_7; \\ p_6(\lambda_1+\mu_2+\mu_3)=\lambda_3 p_2+\lambda_2 p_3+\mu_1 p_7; \\ p_7(\mu_1+\mu_2+\mu_3)=\lambda_3 p_4+\lambda_2 p_5+\lambda_1 p_6. \end{cases} \quad (2)$$

Due to the fact that the set of Eqs. (2) fails to involve constant terms, there are infinitely many different solutions satisfying Eqs. (2). In order to be able to choose the unique solution in terms of p_i, it is necessary to substitute any one of the equations in (2) by the normalizing condition:

$$\sum_{i=0}^{7} p_i = 1 \quad 3)$$

which reflects the fact that the considered system is in any one state at all time.

We can notice the following rules for forming each individual equation by inspecting the set (2) and the associated graph (Fig. 2). The left-hand side of the equation contains the product of the probability of residing in state i and of the summarized rate of all flows departing the system from the i th state. The right-side of the equation is the sum of products of the probability of the state from which it is possible to arrive to state i, and of the corresponding failure or repair flow rate. Thus, given the state space graph, forming a set of equations allows us to calculate the stationary probabilities.

Solving the set of Eqs. (2), we have the values of all stationary probabilities. Physically, the value of $p_i(0 \le p_i \le 1)$ obtained is the relative mean time of the system residing in state i. We point out that this method is known as the state-enumeration method [14].

We calculate the stationary probabilities for eight possible states for the graph shown in Fig. 2. We assume: $L_1 = 1$ km, $L_2 = 1.5$ km, $L_3 = 2$ km. Let also $(\lambda_0)_1 = 0.42$ 1/(yr·km), $(\lambda_0)_2 = 0.37$ 1/(yr·km), $(\lambda_0)_3 = 0.3$ 1/(yr·km). Then $\lambda_1 = (\lambda_0)_1 L_1 = 0.42$ 1/yr, $\lambda_2 = (\lambda_0)_2 L_2 = 0.56$ 1/yr and $\lambda_3 = (\lambda_0)_3 L_3 = 0.6$ 1/yr.1/yr. For the sake of calculation simplicity it is assumed as well that $\mu_1 = \mu_2 = \mu_3 = \mu$. We take $\mu = 0.02$ 1/h = 175.2 1/yr. By substituting these values in Eqs.(2) (replacing one of them by (3)) and solving the set (2) for p_i, we obtain: $p_0 = 0.9$; $p_1 = 2.376 \cdot 10$

$^{-3}$; $p_2 = 3.168 \cdot 10^{-3}$; $p_3 = 3.394 \cdot 10^{-3}$; $p_4 = 7.594 \cdot 10^{-6}$; $p_5 = 8.136 \cdot 10^{-6}$; $p_6 = 1.085 \cdot 10^{-5}$; $p_7 = 2.596 \cdot 10^{-8}$.

With every system state one can associate a certain volume ΔQ_i of raw sewage discharged to the environment that can be represent by the following correspondence relations:

$$\begin{aligned} &0 \to \Delta Q_0 = 0; && 4 \to \Delta Q_4 = (q_1 + q_2)T; \\ &1 \to \Delta Q_1 = q_1 T; && 5 \to \Delta Q_5 = (q_1 + q_2)T; \\ &2 \to \Delta Q_2 = q_2 T; && 6 \to \Delta Q_6 = (q_1 + q_2)T; \\ &3 \to \Delta Q_3 = (q_1 + q_2)T; && 7 \to \Delta Q_7 = (q_1 + q_2)T; \end{aligned} \tag{4}$$

where T is the interval of time for which the discharged sewage volume Q_d is to be estimated.

For the sake of concreteness, we assume that $q_1 = 0.4$ m^3/s, $q_2 = 0.6$ m^3/s and $T = 1$ yr = $31.536 \cdot 10^6$ s. Then $\Delta Q_0 = 0$, $\Delta Q_1 = 126.144 \cdot 10^5$ m^3, $\Delta Q_2 = 189.216 \cdot 10^5$ m^3, $\Delta Q_3 = \Delta Q_4 = \Delta Q_5 = \Delta Q_6 = \Delta Q_7 = 315.360 \cdot 10^5$ m^3. The raw sewage volume is calculated as expectation of the random variable

$$Q_d = \sum_{i=0}^{7} \Delta Q_i p_i = 1.978 \cdot 10^5 \text{ m}^3 \tag{5}$$

that is 0.63 % of the total volume of sewage $(q_1 + q_2)T = 315.360 \cdot 10^5$ m^3 that entered the inlets of the considered network during the year.

Thus, the problem formulated for the sewer network, shown in Fig. 1, is solved.

A more realistic and much used approach proceeds from the fact that the probabilities of simultaneous failure of two or more sewers are extremely low. This fact is easy to verify by analyzing the results of numerical calculations cited above. Taking this into account and assuming that these probabilities are equal to zero, it can be seen that for the network fragment shown in Fig. 1 only four possible states are available, namely:

0: (sewers 1, 2 and 3 up) - p_0;

1: (sewer 1 down, sewers 2 and 3 up) - p_1;

2: (sewer 2 down, sewers 1 and 3 up) - p_2;

3: (sewer 3 down, sewers 1 and 2 up) - p_3.

The corresponding state space graph is shown in Fig. 3,a.

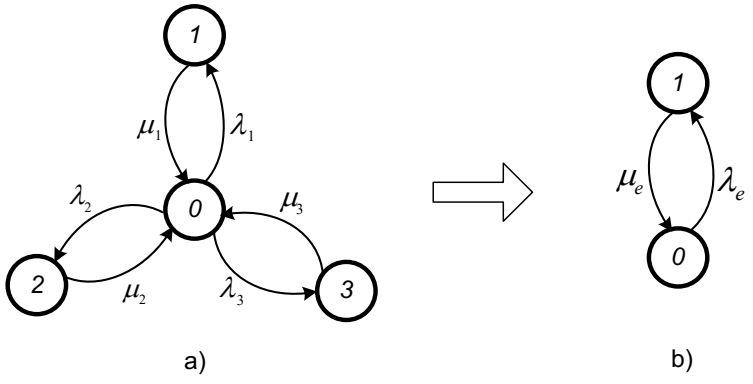


Figure. 3: *Simplified state space graph for the Y-like fragment a) and its transformation b).*

The set of equations written with respect to the stationary probabilities takes the form:

$$\begin{cases} p_0(\lambda_1+\lambda_2+\lambda_3)=\mu_1 p_1+\mu_2 p_2+\mu_3 p_3; \\ p_1\mu_1=\lambda_1 p_0; \\ p_2\mu_2=\lambda_2 p_0; \\ p_3\mu_3=\lambda_3 p_0. \end{cases} \tag{6}$$

It is possible to solve this set of equations analytically. Granting that , we get:

$$\begin{aligned} p_0&=\frac{1}{1+\gamma_1+\gamma_2+\gamma_3}; \\ p_1&=\frac{\gamma_1}{1+\gamma_1+\gamma_2+\gamma_3}; \\ p_2&=\frac{\gamma_2}{1+\gamma_1+\gamma_2+\gamma_3}; \\ p_3&=\frac{\gamma_3}{1+\gamma_1+\gamma_2+\gamma_3}, \end{aligned} \tag{7}$$

where dimensionless parameters characterizing the rate of the "failure-repair" process for each sewer of Y-like network fragment are introduced.

By analogy with (4) we can write for volumes :

$$\begin{aligned} &0\to\Delta Q_0=0; \quad 2\to\Delta Q_2=q_2T; \\ &1\to\Delta Q_1=q_1T;\ 3\to\Delta Q_3=(q_1+q_2)T, \end{aligned} \tag{8}$$

and to make an estimate of the raw sewage discharge Q_d as:

$$Q_d=\sum_{i=0}^{3}\Delta Q_i p_i=\frac{(\gamma_1+\gamma_3)q_1+(\gamma_2+\gamma_3)q_2}{1+\gamma_1+\gamma_2+\gamma_3}T. \tag{9}$$

For data used in this numerical example, the calculation by (9) yields: $Q_d=1.969\cdot10^5\,\mathrm{m}^3$, that is coincident practically with the result (5) obtained above.

IV. Equivalenting of the network fragment

Difficulties emerge when we estimate the raw sewage discharge resulting from sewer network failures for a sufficiently branched, multicomponent sewer network. The problem is that the number of the possible states rapidly increases with number n of network elements (sewers), and equals 2^n. For example, for $n=15$ we have 32768 possible states. The high order of the problem presents difficulties in solving an associated set of equations, equals to the number of possible states, in actual practice. Below is proposed an approach that provides a way of simplifying the procedure of estimating the discharged sewage volume for a sufficiently branched sewer network.

First of all, we recall that the mean relative time of the system residing in the inoperable state having only two possible states (up and down), is numerically equal [16]:

$$p=\frac{\lambda}{\lambda+\mu}=\frac{\gamma}{1+\gamma} \tag{10}$$

where $\gamma=\lambda/\mu$.

Return again to Fig. 1 and imagine a fictitious sewer 123 with unknown, for now, failure (λ_e) and repair (μ_e) rates, at the inlet of which the sewage flow rate (q_1+q_2) is the case, that substitute, in some sense, the Y-like network fragment shown in Fig. 1.

Schematically, such substitution is represented in Fig. 4.

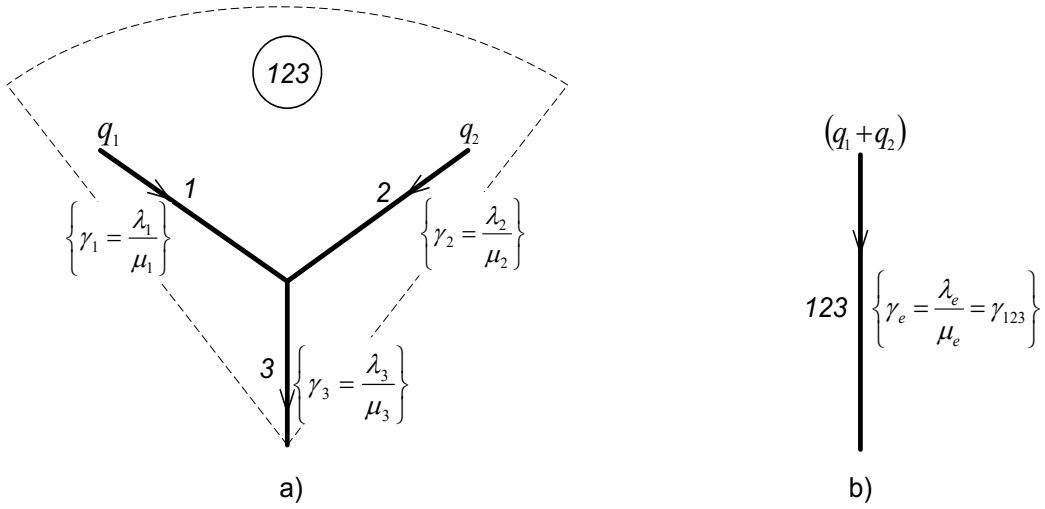


Figure. 4: *Three-component network a) and its equivalent b).*

The state space graph corresponding to Fig. 4b) is shown in Fig. 3b).

It is easy to verify that the volume of raw sewage $(Q_d)_{123}$ discharged from this sewer for time T, is:

$$(Q_d)_{123} = \frac{\gamma_e}{1+\gamma_e}(q_1+q_2)T \tag{11}$$

We call attention to the fact that, at given flow rates q_1 and q_2 at the inlets of Y-like network fragment, the volume of discharged sewage for time T is dependent on the dimensionless parameter γ_e of fictitious sewer only. In this case, under equivalenting of Y-like network fragment, is no need to find λ_e and μ_e separately, but their ratio only.

We find γ_e leading to the same output for the same input. To this end we equate (11) to (9) and solve the equation obtained for γ_e This leads to

$$\gamma_e = \frac{(\gamma_1+\gamma_3)q_1 + (\gamma_2+\gamma_3)q_2}{(1+\gamma_2)q_1 + (1+\gamma_1)q_2} \tag{12}$$

Usually, in actual practice the mean time to failure is far in excess of mean time to repair, that $\gamma << 1$, and, then, Eq. (12) can be written as:

$$\gamma_e \approx \frac{(\gamma_1+\gamma_3)q_1 + (\gamma_2+\gamma_3)q_2}{q_1+q_2} \tag{13}$$

Thus, the Y-like sewer system shown in Fig. 4a) is superseded formally with an equivalent fictitious sewer 123, having the dimensionless parameter $\gamma_e = \gamma_{123}$ and sewage flow rate at the inlet (q_1+q_2) (see Fig. 4b)).

Sometimes, the cases occur when at one point of network more than two (generally, k) sewers are connected (Fig. 5a)).

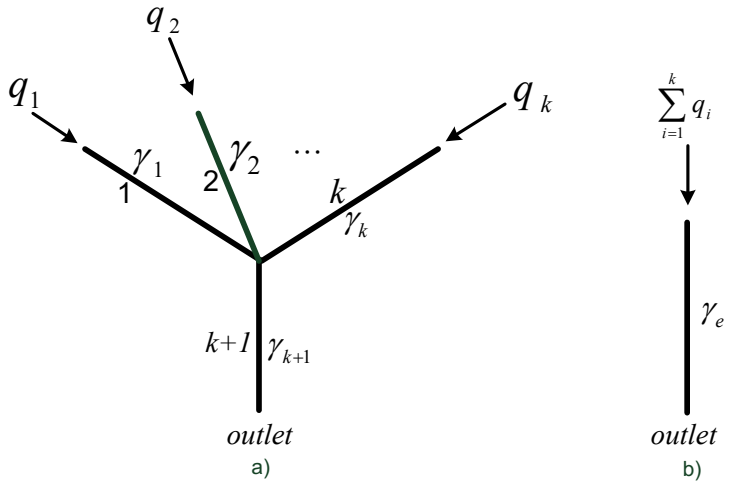


Figure. 5:. *Extension of an equivalenting procedure*

In this case γ_e must be calculated by formula [16]:

$$\gamma_e = \gamma_{k+1} + \frac{\sum_{i=1}^{k} \gamma_i q_i}{\sum_{i=1}^{k} q_i} \tag{14}$$

and, then, the system depicted in Fig. 5a) can be superseded by one sewer as shown in Fig. 5b).

Now we find out a physical meaning of the dimensionless parameter γ_e.

Because $\gamma_e << 1$, the Eq. (11) may be rewritten as $(Q_d)_{123} \approx \gamma_e (q_1 + q_2)T$. It is evident that cofactor $(q_1 + q_2)T = Q$ in the right-hand side of this expression is a total volume of raw sewage that entered the inlets of the considered network at time T . Then, γ_e is a part of Q that is not conveyed to the network outlet, i. e. is discharged to the environment. When multiplied by 100, physically shows the raw sewage discharge resulting from sewer network failures, expressed as a percentage of total sewage volume entered to its inlets. By virtue of the fact that γ_e is varied from 0 (absolutely reliable network) to 1 (theoretically, completely inoperable network), the parameter γ_e, in our opinion, may be used as an objective, single-valued measure of the sewer network reliability.

We emphasize that the sewer network fragment of Fig. 4a) (or Fig.5a)) is a structure-forming component in the sense that any arbitrary complicated dendritic sewer network may be thought of as a composition of such components that substantially reduces and simplifies a body of calculations in estimating raw sewage discharged from the network. Below we give a technique of how to apply this approach.

V. Decomposition-equivalence technique

We shall call this procedure as the "decomposition-equivalence technique". It is more convenient to demonstrate this technique by the following example.

Consider the network in Fig. 6a) consisting of seven sewers, each determined by the values

λ_i and μ_i, and, hence, by the value $\gamma_i = \lambda_i / \mu_i \ (i = 1 \div 7)$. In addition, the sewage flow rate at the network inlets (q_1, q_2, q_3, q_4) will be considered to have constant values.

It is necessary to estimate the raw sewage volume discharged from the network throughout the year as a consequence of possible failures.

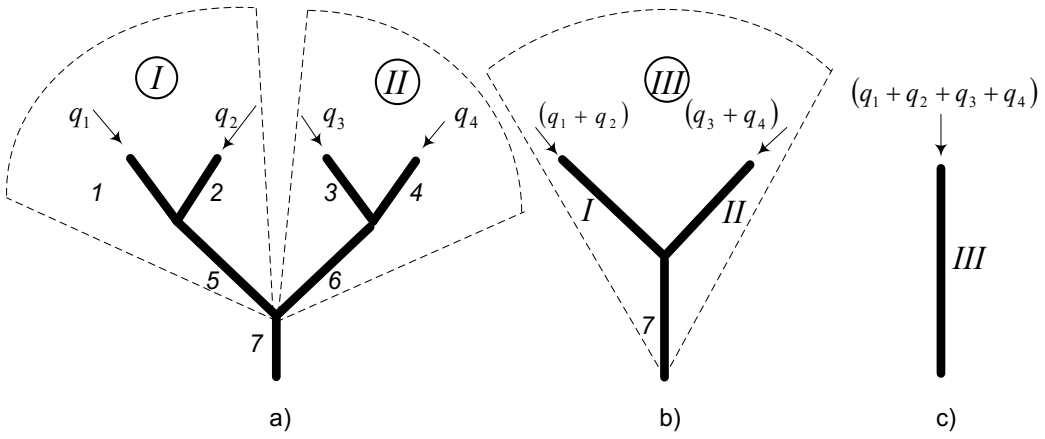


Figure. 6:.*Decomposition-equivalence technique.*

First we consider the contours I and II in Fig. 6a). Either contour includes the Y-like system, and, consequently, can be substituted by one equivalent sewer with its associated value of parameter γ calculated according to the method proposed above. Using Eq. (13), we have γ_I for contour I. Similarly, with assigned notations, for contour II we have γ_{II}.

The results obtained enable one to present the initial network in the form shown in Fig. 6b). But this is an Y-like system (contour III) again. Using Eq. (13), we have finally the parameter $\gamma_{III} = \gamma_e$ of one equivalent sewer substituting the initial network (see Fig. 6c)). Thus, the problem is solved.

As may be seen from this example, unlike the state-enumeration method here, there is no need to solve an unwieldy set of equations. The problem reduces to a sequence of simple computations using, at every stage, the results of a preceding step.

Although this methodology has been applied to a comparatively simple case, it can be extended easily to multicomponent networks.

VI. Applications

The method developed in this paper may be used to solve many practical problems. Some of these, in the form of numerical examples, are considered below in a deliberately simplified but well realistic statement.

6.1. Problem 1.The sewer network shown in Fig. 7a) is given.

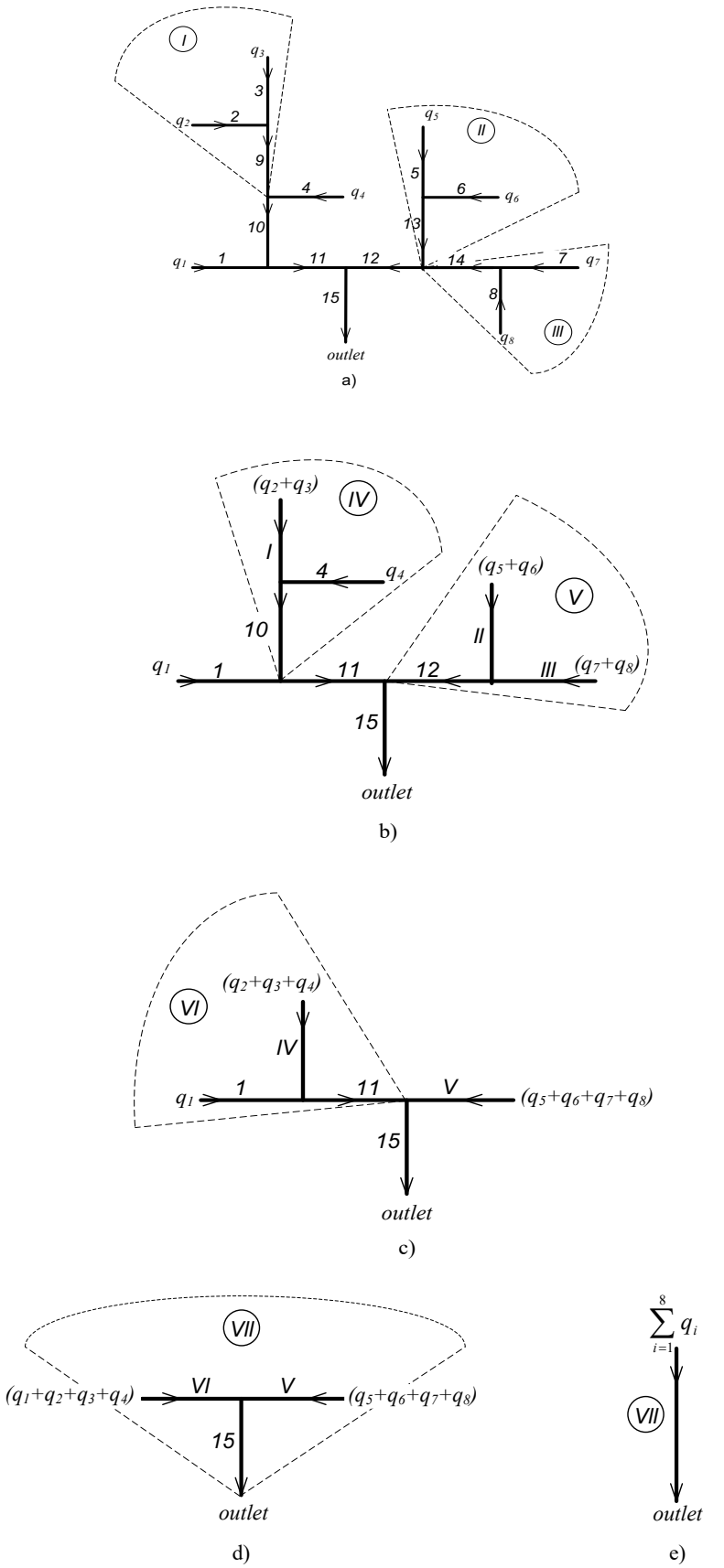


Figure 7: *Initial sewer network a) and its sequential transformations b), c), d), e).*

The network consists of 15 enumerated sewer sections; the number of inlets is equal to 8. The direction of the sewage flow through an each sewer is shown by the arrow. There is a need to estimate a reliability level of this network (in the sense of the proposed criterion).

To carry out the calculations we need some data. Such data are represented in Table 1.

Table 1: *Input data for calculations*

Section, i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Failure rate λ_i, (1/yr)	0.52	0.68	0.79	0.91	1.34	0.83	0.75	0.03	0.85	0.62
Repair rate μ_i, (1/yr)	220	220	220	220	220	220	220	220	200	150
Parameter $\gamma_i = \lambda_i / \mu_i\ (\times 10^3)$	2.36	3.09	3.59	4.14	6.09	3.77	3.41	0.14	4.25	4.13

11	**12**	**13**	**14**	**15**
0.84	1.10	0.03	0.50	0.05
120	120	200	200	90
7.00	9.17	0.15	2.50	0.56

Besides, the inlets sewage flow rates in Table 2 are shown.

Table 1: *Network inlets sewage flow rate.*

Inlet, i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Sewage flow rate $q_i(\times 10^2)$, (m^3/s)	3	9	6	4	5	1	5	7

In addition, without loss of generality, we assume that the length of each sewer section is equal to 1 km. We note also that all values are hypothetical, convenient for calculations only.

First we consider the contours I, II and III (Fig. 7a)) at the network periphery. Either contour includes the Y-like system, and, consequently, can be substituted by one equivalent sewer with its associated value of parameter γ calculated according to the method proposed above. Using Eq. (13) where now, taking account of the new notations, and the data from Table 1 and Table 2 we have for contour I:

$$\gamma_I = \frac{(\gamma_2 + \gamma_9)q_2 + (\gamma_3 + \gamma_9)q_3}{q_2 + q_3} = 7.54 \cdot 10^{-3}$$

Similarly, for contour II:

$$\gamma_{II} = \frac{(\gamma_5 + \gamma_{13})q_5 + (\gamma_6 + \gamma_{13})q_6}{q_5 + q_6} = 5.85 \cdot 10^{-3}$$

and for contour III:

$$\gamma_{III} = \frac{(\gamma_7 + \gamma_{14})q_7 + (\gamma_8 + \gamma_{14})q_8}{q_7 + q_8} = 4.00 \cdot 10^{-3}$$

The results obtained enable one to present the initial network in the form shown in Fig. 7b). But here are the Y-like systems (contours IV and V) again. Using Eq. (13) we have the parameter γ_{IV} for contour IV:

$$\gamma_{IV} = \frac{(\gamma_1 + \gamma_{10})(q_2 + q_3) + (\gamma_4 + \gamma_{10})q_4}{q_2 + q_3 + q_4} = 10.95 \cdot 10^{-3}$$

and for contour V:

$$\gamma_V = \frac{(\gamma_{II} + \gamma_{12})(q_5 + q_6) + (\gamma_{III} + \gamma_{12})(q_7 + q_8)}{q_5 + q_6 + q_7 + q_8} = 13.79 \cdot 10^{-3}$$

As a result, the structure shown in Fig. 7b) substitutes by the structure depicted in Fig. 7c) where the Y-like sub-system (contour VI) may be selected. Equivalenting this contour again by one sewer section with the parameter

$$\gamma_{VI} = \frac{(\gamma_1 + \gamma_{11})q_1 + (\gamma_{IV} + \gamma_{11})(q_2 + q_3 + q_4)}{q_1 + q_2 + q_3 + q_4} = 16.78 \cdot 10^{-3}$$

we are going to the Fig. 7d).

But the structure shown in Fig. 7d) is the Y-like fragment (contour VII) in itself that may be substituted by one sewer (see Fig. 7e)). Thus, finally we have the parameter γ_{VII} of one equivalent sewer substituting the initial network depicted in Fig. 7a):

$$\gamma_{VII} = \frac{(\gamma_{VI} + \gamma_{15})(q_1 + q_2 + q_3 + q_4) + (\gamma_V + \gamma_{15})(q_5 + q_6 + q_7 + q_8)}{q_1 + q_2 + q_3 + q_4 + q_5 + q_6 + q_7 + q_8} = 16.00 \cdot 10^{-3}$$

The sequence of "decomposition-equivalence" operations is completed. Hence, in this case, $\gamma_e = \gamma_{VII} = 0.016$. This means that, on the average, 1.6 % of the total sewage volume that entered the network inlets during time T , discharges from the sewer network to the environment arising from the network component failures. The accuracy of this measure increases as T increase, that is characteristic for probabilistic problems at all.

6.2. Problem 2.Let us assume that specialists analyzing the results obtained in preceding Problem 1 come to the conclusion that the raw sewage discharge from the sewer network (Fig. 7a)) is much too large, and, consequently, the network reliability needs to be increased. The question concerning replacement of some components by a new sewer pipe is discussed, but it is possible to replace only one sewer because available funds are limited. On the present evidence, it may be argued that the failure rate for a new sewer (manufacturer's data) is $\lambda_n = 0.02$ 1/yr; the repair rate μ_n is taken to be equal to 200. It is desired to identify the preferential alternative.

First of all, we compute $\lambda_n = \lambda_n / \mu_n = 0.1 \cdot 10^{-3}$. As before, we will take the discharged sewage volume as an efficiency index of the alternative to be accepted. Calculate this quantity assuming that the replacement of sewer section 1 in the initial network (Fig. 7a)) has just been made. For this purpose, we substitute the input data (associated with the sewer 1) of the Problem 1, for one another (corresponding to the new sewer), namely $\gamma_n = 0.1 \cdot 10^{-3}$. Carrying out the relevant calculations, we obtain the discharged sewage volume expressed as a percentage of the total sewage entered the network: 1.582 %. By repeating the similar calculations with respect to each network section we come to the results represented in Table 3.

Table 3: *Example table Result of calculations.*

Section to be replaced	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Relative sewage volume discharge from network,%	1.582	1.532	1.547	1.559	1.524	1.590	1.558	1.559
	9	**10**	**11**	**12**	**13**	**14**	**15**	
	1.444	1.408	1.220	1.191	1.599	1.527	1.523	

Referring to Table 3, it is seen that the smallest volume of sewage to be discharged from the network occurs when the network's section 12 is replaced (in Table 3 this is highlighted in bold print). It is obvious that, under otherwise equal conditions, this alternative is preferable from the viewpoint of the reliability index accepted in this work.

The problems considered are simple as well; for this reason, the results seem to be trivial. Note, however, that the simplicity of the examples makes it possible to see the potential of the proposed method for practical use.

VII. Conclusion

Although sewer reliability depicts a fairly complete reliability measure of the sewer network, it is convenient to use a single index to represent the composite effect of the component reliabilities. We propose to assess sewer network reliability as a whole by a volume of raw sewage discharged from the system because of failures of its components for an appreciable length of time. The traditional method for solving such problems is the so-called state-enumeration method. But, for the multicomponent networks, this generates a need to solve a set of equations having very high order, which renders the method unsuitable for many practical applications. The approach proposed in this work makes it possible to circumvent these difficulties by using the concept of equivalent sewer. As a result, the problem reduces to a sequential consideration of elementary sub-problems the solution of which is easily accomplished.

As the methodology is applicable for sewer networks, each component of which can be either up (operable) or down(failed), additional research is need for extending the method for more complex cases.

In our view, similar problems exist also in the course of maintenance of oil, gas and other pipeline systems. Such a setting and solving of problems may also be of interest for specialists working with general reliability issues.

References

[1] Bao, Y., Mays, L. W., Model for water distribution system reliability, Journal of Hydraulic Engineering, 116 (1990), 1119-1137.

[2] Engelhardt, M. O., Skipworth, P. J., Savic, D. A., Saul, A. J., Walters, O. A., Rehabilitation strategies for water distribution networks: a literature review with a UK perspective, Urban Water, 2 (2000), 153-170.

[3] Fujiwara, O., Ganesharajah, T., Reliability assessment of water supply systems with storage and distribution networks, Water Resources Research, 29 (1993), 2917-2924.

[4] Mays, L. W. (Ed.), Reliability analysis of water distribution systems, Congress Cataloging-in-Publication Data, American Society of Civil Engineering, 1989.

[5] Su, Y. C., Mays, L. W., Duan, N., Lansey, K. E., Reliability-based optimization model for water distribution systems, Journal of Hydraulic Engineering, 114 (1987), 1539-1556.

[6] Wagner, J. M., Shamir, U., Marks, D. H., Water distribution reliability. Analytical methods, Journal of Water Resources Planning and Management Division, 114 (1988), 253-275.

[7] Ermolin, Y. A., Estimation of raw sewage discharge resulting from sewer network failures, Urban Water, 4 (2001), 271-276.

[8] Kwietniewski, M., Lesniewski, M., The reliability assessment of a typical structure fragment of a stormwater collection network including uncertainty, Progress in Biomedical Optics and Imaging, 2006, T. 619 II.

[9] Ermolin Y. A., Reliability Estimation of Urban Wastewater Disposal Networks, in: Gregory I. Hayworth (Ed.), Reliability Engineering Advances, Nova Science Publishers, Inc., New York, 2009, pp. 379-397.

[10] Jin, Y., Mukherjee, A., Modeling blockage failures in sewer systems to support maintenance decision making, Journal of Performance of Constructed Facilities, 6 (2010), 622-633.

[11] Yermolin, Yu. A., Alexeyev, M. I., Reliable operation of wastewater collection systems and ways of its improvement, Water Supply and Sanitary Technique, 1 (2012), 13-16. (in Russian).

[12] Ermolin, Y. A.. Alexeev, M. I. Reliability measure of a sewer network. Water and Ecology,

2(2018), 51-58.

[13] Alexeev, M. I., Baranov, L. A., Ermolin, Y. A. Risk-based approach to evaluate the reliability of a city sewer network. Water and Ecology, 3(2020), 3-7. (in Russian).

[14] Ventzel, E. S., Operations Analysis: Problems, Principles and Methodology, "VysshajaShkola", Moscow, 2001. (in Russian).

[15] Alekseev, M. I., Ermolin, Yu. A., Stationarization of the "failure-restoration" process at the analyzing of complex system' reliability, "Bulletin of Civil Engineers", 1 (2014), 84-87. (in Russian).

[16] Baranov, L. A., Ermolin, Y. A., Shubinsky, I. B. On a reliability of tree-like transportation networks. Reliability: Theory & Applications, 2(2021), 115-123. DOI: https://doi.org/10.24412/1932-2321.1-2021-262-115-123.

RT&A, No 4 (65)
Volume 16, December 2021

Reliability analysis for a class of exponential distribution based on progressive first-failure censoring

Kambiz Ahmadi

Department of Computer Sciences, Faculty of Mathematical Sciences, Shahrekord University, Iran
K.Ahmadi@sku.ac.ir

Abstract

Based on progressively first-failure censored data, the problem of estimating parameters as well as reliability and hazard rate functions for a class of an exponential distribution is considered. The classic and Bayes approaches are used to estimate the parameters. The maximum likelihood estimates and exact confidence interval as well as exact confidence region for parameters are developed based on this censoring scheme. Also, when the parameters have discrete and continuous priors, several Bayes estimators with respect to different symmetric and asymmetric loss functions such as squared error, linear-exponential (LINEX) and general entropy are derived. Finally two numerical examples are presented to illustrate the methods of inference developed in this paper.

Keywords: Bayes estimator, Confidence region, Exponential distribution, Maximum likelihood estimator, Loss function, Progressive first-failure censoring scheme

1. Introduction

In many life test studies, it is common that the lifetimes of the test units may not be able to record exactly. Censoring is very common in reliability data analysis, in the past several decades. It usually applies when the exact lifetimes are known for only a portion of the products and the remainder of the lifetimes has only partial information. There are several types of censoring schemes in survival analysis and the type-II censoring scheme is one of the most common for consideration. In type-II censoring, the test terminates after a predetermined number of failure occurs in order to save time or cost, but the conventional type-II censoring scheme does not has the flexibility of allowing removal of units at points other than the terminal point of the experiment. For this reason, a more general censoring scheme called progressive type-II right censoring is proposed. Although, progressive censoring scheme was introduced long ago in the statistical literature, in recent years the progressive censoring scheme has received considerable attention in the statistical literature, see for instance [1], [2] and [3]. For an exhaustive list of references and further details on progressive censoring, readers are referred to [4]. In some cases, the lifetime of products is quite long and so the experimental time of the progressive type-II censoring scheme can still be too long. In order to give an efficient experiment, the other test methods are proposed by statisticians where one of them is the progressive first-failure censoring scheme. It can be described as follows.

Suppose that n independent groups with k items within each group are put on a life test and experimenter decides beforehand the quantity m, the number of units to be failed. At the time of the first failure, $X^{\mathbf{r}}_{1;m,n,k}$, r_1 groups and the group in which the first failure is observed are randomly removed. r_2 groups and the group with observed failure are randomly removed as soon as the second failure, $X^{\mathbf{r}}_{2;m,n,k}$, has occurred. The procedure is continued until all r_m groups and the group with observed failure are removed at the time of the m-th failure, $X^{\mathbf{r}}_{m;m:k:n}$. Then $X^{\mathbf{r}}_{1;m,n,k} < X^{\mathbf{r}}_{2;m,n,k} < \ldots < X^{\mathbf{r}}_{m;m,n,k}$ are called progressively first-failure censored order

statistics with the censoring scheme $\mathbf{r} = (r_1, r_2, \ldots, r_m)$. We notice that if $k = 1$, progressively first-failure censored reduces to the progressive type-II censoring. Also, if $k = 1$ and $r_1 = r_2 = \cdots = r_{m-1} = 0, r_m = n - m$, it reduce to the type-II censoring. Wu et al. [5] and Wu and Yu [6] obtained the maximum likelihood estimates (MLEs), exact confidence intervals and exact confidence regions for the parameters of the Gompertz and Burr type XII distributions based on first-failure censored sampling, respectively. Wu and Kuş [7] studied the Weibull distribution under progressive first-failure censoring to make some classical inference on the parameters of a Weibull distribution and they proved that the progressive first-failure censoring scheme had shorter expected test times than the progressive type-II censoring scheme. Abou-Elheggag [8] studied the Rayleigh distribution under progressive first-failure censoring. He derived the maximum likelihood estimates and Bayes estimates of scale parameter, survival and hazard rate functions. Also, one can refer to [9], [10], [11],[12] and [13].

To simplify the notation, we will use X_i in place of $X^{\mathbf{r}}_{i;m,n,k}$. Let $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ be a progressive first-failure censored sample from a continuous population with the cumulative distribution function (CDF), $F(x)$, the probability density function (PDF), $f(x)$, and $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ is an observed value of $\mathbf{X}$. The joint pdf of $\mathbf{X}$ is given by [7] as follows

$$f_{1,2,\ldots,m}(\mathbf{x}) = Ak^m \prod_{i=1}^{m} f(x_i)[1 - F(x_i)]^{k(r_i+1)-1}, \quad 0 < x_1 < x_2 < \ldots < x_m < \infty, \tag{1}$$

where $A = n(n - r_1 - 1)(n - r_1 - r_2 - 2) \cdots (n - r_1 - r_2 - \cdots - r_{m-1} - m + 1)$.

In this paper, our main object is to study the classical and Bayes estimation procedures for the parameter(s) of a general class of exponential- type distribution based on a progressively first-failure censored sample.

The rest of this paper is organized as follows. In Section 2, the model is described. Some classical estimation, such as maximum likelihood estimation and interval estimation are presented in Section 3. Section 4 develops the Bayes estimators for different loss functions such as squared error, LINEX and general entropy. One illustrative example and a simulation study via a Monte Carlo method are conducted in Section 5. Finally, we conclude the paper in Section 6.

2. Model description

Suppose the lifetime random variable T has a continuous distribution with two parameters as α and λ, and with the PDF and CDF as

$$f(t; \alpha, \lambda) = \alpha \psi(t; \lambda) \exp\{-\alpha \Psi(t; \lambda)\}, \qquad 0 < t < \infty, \tag{2}$$
$$F(t; \alpha, \lambda) = 1 - \exp\{-\alpha \Psi(t; \lambda)\}, \tag{3}$$

where $\psi(t; \lambda) = \frac{\partial \Psi(t;\lambda)}{\partial t}$, $\Psi(t; \lambda)$ is increasing in t with $\Psi(0; \lambda) = 0$ and $\Psi(\infty; \lambda) = \infty$. The corresponding reliability and hazard rate functions becomes:

$$R(t) = \exp\{-\alpha \Psi(t; \lambda)\}, \tag{4}$$
$$h(t) = \alpha \psi(t; \lambda), \tag{5}$$

respectively. This general form for lifetime model including some well-known and useful models such as Burr XII distribution with $\Psi(t; \lambda) = \ln(1 + t^\lambda)$, Gompertz distribution with $\Psi(t; \lambda) = \frac{e^{\lambda t} - 1}{\lambda}$, Weibull distribution with $\Psi(t; \lambda) = t^\lambda$, two parameters bathtub-shaped lifetime distribution (see [14]) with $\Psi(t; \lambda) = e^{t^\lambda} - 1$ and so on. For more details, we refer the reader to [15].

3. Classical estimation

In this section, we consider the maximum likelihood estimation and interval estimation for the unknown parameters when the data are progressively first-failure censored.

3.1. Point estimation

Let $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ be a progressive first-failure censored sample from (2), with censoring scheme $(r_1, r_2, \ldots, r_m)$. From (1) the likelihood function is given by

$$L(\alpha, \lambda; \mathbf{x}) = Ak^m \alpha^m \exp\{-\alpha k \sum_{i=1}^{m}(r_i+1)\Psi(x_i;\lambda)\} \prod_{i=1}^{m} \psi(x_i;\lambda). \tag{6}$$

By setting the derivatives of the log-likelihood function with respective to α or λ to zero, the MLE of λ, say $\hat{\lambda}$, is the solution to the following likelihood equation

$$\sum_{i=1}^{m} \frac{(\partial/\partial\lambda)\psi(x_i;\lambda)}{\psi(x_i;\lambda)} = \frac{m\sum_{i=1}^{m}(r_i+1)(\partial/\partial\lambda)\Psi(x_i;\lambda)}{\sum_{i=1}^{m}(r_i+1)\Psi(x_i;\lambda)}, \tag{7}$$

and the MLE of α, say $\hat{\alpha}$, can be obtained as

$$\hat{\alpha} = \frac{m}{k\sum_{i=1}^{m}(r_i+1)\Psi(x_i;\hat{\lambda})}. \tag{8}$$

It is not easy to solve the equation(7) analytically in order to achieve the MLE of λ. Some numerical methods can be employed such as the Newton-Raphson method. Finally, using the invariance property, the MLEs of $R(t)$ and $h(t)$ are obtained as

$$\hat{R}(t) = \exp\{-\hat{\alpha}\Psi(t;\hat{\lambda})\}, \quad t > 0,$$

and

$$\hat{h}(t) = \hat{\alpha}\psi(t;\hat{\lambda}), \quad t > 0,$$

respectively.

3.2. Interval estimation

Let $Y_i = k\alpha\Psi(X_i;\lambda)$ for $i = 1, 2, \ldots, m$. It can be seen that $Y_1 < Y_2 < \ldots < Y_m$, are the progressive first-failure censored order statistics from an exponential distribution with mean 1. Consider the following transformation:

$$\begin{aligned} Z_1 &= nY_1, \\ Z_i &= (n - r_1 - r_2 - \cdots - r_{i-1} - i + 1)(Y_i - Y_{i-1}), \quad i = 2, 3, \ldots, m. \end{aligned}$$

The generalized spacings $Z_1, Z_2, \ldots, Z_m$ are independent and identically distributed as an exponential distribution with mean 1 (see [1]). Hence, for $j = 1, 2, \ldots, m-1$,

$$\tau_j = 2\sum_{i=1}^{j} Z_i = 2k\alpha[\sum_{i=1}^{j}(r_i+1)\Psi(X_i;\lambda) + \sum_{i=j+1}^{m}(r_i+1)\Psi(X_j;\lambda)] \tag{9}$$

and

$$\gamma_j = 2\sum_{i=j+1}^{m} Z_i = 2k\alpha \sum_{i=j+1}^{m}(r_i+1)[\Psi(X_i;\lambda) - \Psi(X_j;\lambda)] \tag{10}$$

are independently Chi-squared distributed with $2j$ and $2(m-j)$ degrees of freedom, respectively. We consider the following pivotal quantities:

$$\eta_j = \frac{j}{m-j}.\frac{\sum_{i=j+1}^{m}(r_i+1)(\Psi(X_i;\lambda) - \Psi(X_j;\lambda))}{\sum_{i=1}^{j}(r_i+1)\Psi(X_i;\lambda) + \sum_{i=j+1}^{m}(r_i+1)\Psi(X_j;\lambda)}, \quad j = 1, 2, \ldots, m-1, \tag{11}$$

$$\xi = 2k\alpha\sum_{i=1}^{m}(r_i+1)\Psi(X_i;\lambda). \tag{12}$$

It is clearly that η_j has a F distribution with $2(m-j)$ and $2j$ degrees of freedom and ξ has a Chi-squared distribution with $2m$ degree of freedom. Meanwhile, η_j and ξ are independent. To construct an exact confidence interval for λ and the joint confidence region for the parameters α and λ, we need the following lemma.

Lemma 1. Suppose that for $x_1 < x_2 < \cdots < x_m$,

$$w_j(\lambda) = \frac{\sum_{i=j+1}^{m}(r_i+1)(\Psi(x_i;\lambda)-\Psi(x_j;\lambda))}{\sum_{i=1}^{j}(r_i+1)\Psi(x_i;\lambda)+\sum_{i=j+1}^{m}(r_i+1)\Psi(x_j;\lambda)}, \quad j=1,2,\ldots,m-1. \tag{13}$$

Then $w_j(\lambda)$ is strictly increasing in λ, if function $\frac{\Psi'(t;\lambda)}{\Psi(t;\lambda)}$ is strictly increasing in t, where $\Psi'(t;\lambda)$ is $(\partial/\partial\lambda)\Psi(t;\lambda)$.

Proof. Let $w_j(\lambda) = w_{1j}(\lambda)/w_{2j}(\lambda)$, where

$$w_{1j}(\lambda) = \sum_{i=j+1}^{m}(r_i+1)\frac{\Psi(x_i;\lambda)}{\Psi(x_j;\lambda)} - \sum_{i=j+1}^{m}(r_i+1), \tag{14}$$

$$w_{2j}(\lambda) = \sum_{i=1}^{j}(r_i+1)\frac{\Psi(x_i;\lambda)}{\Psi(x_j;\lambda)} + \sum_{i=j+1}^{m}(r_i+1). \tag{15}$$

Since $w_{1j}(\lambda)$ and $w_{2j}(\lambda)$ are positive, the proof is obtained if we can show that $w_{1j}(\lambda)$ and $w_{2j}(\lambda)$ are strictly increasing and decreasing functions in λ, respectively. It is observed that

$$w'_{1j}(\lambda) = \frac{1}{\Psi^2(x_j;\lambda)}\sum_{i=j+1}^{m}(r_i+1)[\Psi'(x_i,\lambda)\Psi(x_j;\lambda)) - \Psi(x_i;\lambda)\Psi'(x_j;\lambda)] > 0, \tag{16}$$

$$w'_{2j}(\lambda) = \frac{1}{\Psi^2(x_j;\lambda)}\sum_{i=1}^{j}(r_i+1)[\Psi'(x_i;\lambda)\Psi(x_j;\lambda)) - \Psi(x_i;\lambda)\Psi'(x_j;\lambda)] \le 0, \tag{17}$$

Since, when $\frac{\Psi'(t;\lambda)}{\Psi(t;\lambda)}$ is strictly increasing in t, then $\frac{\Psi'(x_i;\lambda)}{\Psi(x_i;\lambda)} < \frac{\Psi'(x_j;\lambda)}{\Psi(x_j;\lambda)}$ for $i=1,2,\ldots,j-1$, and $\frac{\Psi'(x_i;\lambda)}{\Psi(x_i;\lambda)} > \frac{\Psi'(x_j;\lambda)}{\Psi(x_j;\lambda)}$ for $i=j+1,j+2,\ldots,m$. ■

Remark 1. For all of well-known lifetime distributions mentioned in Section 2, it can be shown that $\frac{\Psi'(t;\lambda)}{\Psi(t;\lambda)}$ is strictly increasing in t. For instance, when $\Psi(t;\lambda) = \ln(1+t^\lambda)$, it turns out to be Burr XII distribution and see [5].

Let $F_{v_1,v_2}(p)$ is the percentile of F distribution with v_1 and v_2 degrees of freedom with the right-tail probability p. An exact confidence interval for the parameter λ, and the joint confidence region for the parameters α and λ are given in the following theorems, respectively.

Theorem 1. Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ be a progressive first-failure censored sample from (2), with censoring scheme $(r_1, r_2, \ldots, r_m)$, $\frac{\Psi'(t;\lambda)}{\Psi(t;\lambda)}$ is strictly increasing in t, and

$$W_j(\lambda) = \frac{j}{m-j}w_j(\lambda), \quad j=1,2,\ldots,m-1, \tag{18}$$

where $w_j(\lambda)$ is defined in (13). Then, for any $0<\nu<1$ and $j=1,2,\ldots,m-1$, when $F_{2(m-j),2j}(\frac{\nu}{2})$ and $F_{2(m-j),2j}(1-\frac{\nu}{2})$ are in the range of the function $W_j(\lambda)$

$$\varphi_j(\mathbf{X}, F_{2(m-j),2j}(1-\frac{\nu}{2})) < \lambda < \varphi_j(\mathbf{X}, F_{2(m-j),2j}(\frac{\nu}{2})) \tag{19}$$

is a $100(1-\nu)\%$ confidence interval for λ, where $\varphi_j(\mathbf{X}, u)$ is the solution of λ for equation $W_j(\lambda) = u$.

Proof. By Lemma 1, $W_j(\lambda)$ is a strictly increasing function in λ and since $F_{2(m-j),2j}(1-\frac{\nu}{2})$ and $F_{2(m-j),2j}(\frac{\nu}{2})$ are in the range of function $W_j(\lambda)$, then equations $W_j(\lambda) = F_{2(m-j),2j}(\frac{\nu}{2})$ and $W_j(\lambda) = F_{2(m-j),2j}(1-\frac{\nu}{2})$ have unique solutions with respect to λ. Also we know that η_j has an F distribution with $2(m-j)$ and $2j$ degrees of freedom. Thus for $0 < \nu < 1$,

$$P\big(F_{2(m-j),2j}(1-\frac{\nu}{2}) < \eta_j < F_{2(m-j),2j}(\frac{\nu}{2})\big) = 1-\nu$$

is equivalent to

$$P\Big(\varphi_j\big(\mathbf{X}, F_{2(m-j),2j}(1-\frac{\nu}{2})\big) < \lambda < \varphi_j\big(\mathbf{X}, F_{2(m-j),2j}(\frac{\nu}{2})\big)\Big) = 1-\nu.$$

■

Theorem 2. Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ be a progressive first-failure censored sample from (2), with censoring scheme $(r_1, r_2, \ldots, r_m)$, $\frac{\Psi'(t;\lambda)}{\Psi(t;\lambda)}$ is strictly increasing in t. Then, for any $0 < \nu < 1$ and $j = 1, 2, \ldots, m-1$, when $F_{2(m-j),2j}(\frac{1+\sqrt{1-\nu}}{2})$ and $F_{2(m-j),2j}(\frac{1-\sqrt{1-\nu}}{2})$are in the range of function $W_j(\lambda)$, a $100(1-\nu)\%$ confidence region for (α, λ) is given by

$$\begin{cases} \varphi_j\big(\mathbf{X}, F_{2(m-j),2j}(\frac{1+\sqrt{1-\nu}}{2})\big) < \lambda < \varphi_j\big(\mathbf{X}, F_{2(m-j),2j}(\frac{1-\sqrt{1-\nu}}{2})\big) \\ \frac{\chi^2_{2m}(\frac{1+\sqrt{1-\nu}}{2})}{2k\sum_{i=1}^m (r_i+1)\Psi(x_i;\lambda)} < \alpha < \frac{\chi^2_{2m}(\frac{1-\sqrt{1-\nu}}{2})}{2k\sum_{i=1}^m (r_i+1)\Psi(x_i;\lambda)} \end{cases}, \tag{20}$$

where $\chi^2_{v_1}(p)$ is the percentile of Chi-squared distribution with v_1 degree of freedom with the right-tail probability p and $\varphi_j(\mathbf{X}, u)$ is defined in Theorem 1 .

Proof. For $0 < \nu < 1$,

$$\begin{aligned} &P\Big(\varphi_j\big(\mathbf{X}, F_{2(m-j),2j}(\frac{1+\sqrt{1-\nu}}{2})\big) < \lambda < \varphi_j\big(\mathbf{X}, F_{2(m-j),2j}(\frac{1-\sqrt{1-\nu}}{2})\big), \\ &\qquad \frac{\chi^2_{2m}(\frac{1+\sqrt{1-\nu}}{2})}{2k\sum_{i=1}^m (r_i+1)\Psi(x_i;\lambda)} < \alpha < \frac{\chi^2_{2m}(\frac{1-\sqrt{1-\nu}}{2})}{2k\sum_{i=1}^m (r_i+1)\Psi(x_i;\lambda)}\Big) = \\ &P\Big(F_{2(m-j),2j}(\frac{1+\sqrt{1-\nu}}{2}) < \eta_j < F_{2(m-j),2j}(\frac{1-\sqrt{1-\nu}}{2})\Big) \\ &\qquad P\Big(\chi^2_{2m}(\frac{1+\sqrt{1-\nu}}{2}) < \xi < \chi^2_{2m}(\frac{1-\sqrt{1-\nu}}{2})\Big) = \sqrt{1-\nu}\sqrt{1-\nu} = 1-\nu, \end{aligned}$$

and the first equality follows from the fact that η_j and ξ are independent. ■

It is observed that Theorems 1 and 2 provides the different confidence intervals and confidence regions, respectively for various j. We can derive optimal confidence interval and region based on different criteria such as shortest interval length and minimum region area.

4. Bayes Estimation

The Bayesian approach in statistical inference provides an alternative choice for parameters estimation. We consider the Bayesian estimates of the unknown parameters α and λ as well as reliability function $R(t)$ and hazard rate function $h(t)$ under symmetric and asymmetric loss functions.

The loss function plays a critical role in Bayes perspective. In many practical situations, usually symmetric loss function such as squared error loss function is taken into consideration to produce Bayes estimates. In most cases, it is done for convenience but may not be appropriate in many real life situations. Since under this loss function overestimation and underestimation are equally penalized which is not a good criteria from practical point of view. As an example, in

reliability estimation overestimation is considered to be more serious than the underestimation. Thus, it is important to consider Bayes estimates under asymmetric loss function. The squared error loss function is defined as

$$L_1\left(f(\mu),\hat{f}(\mu)\right)=\left(\hat{f}(\mu)-f(\mu)\right)^2,$$

with $\hat{f}(\mu)$ begins an estimator of $f(\mu)$. Here $f(\mu)$ denotes some parametric function of μ. Bayes estimator, say $\hat{f}_{SB}(\mu)$ is evaluated by the posterior mean of $f(\mu)$.

One of the most commonly used asymmetric loss function is LINEX loss function which introduced first by [16] and further properties of this loss function have been investigated by [17] . It is defined as follows:

$$L_2\left(f(\mu),\hat{f}(\mu)\right)=e^{c\Delta}-c\Delta-1,\qquad c\neq 0,$$

where $\Delta=\hat{f}(\mu)-f(\mu)$. When c is negative, underestimation is more serious than overestimation and it is opposit for positive c. The Bayes estimator of $f(\mu)$ for the loss function L_2 can be obtained as $\hat{f}_{LB}(\mu)=-\frac{1}{c}\ln\left\{E_\mu\left(e^{-cf(\mu)}|data\right)\right\}$, provided that $E_\mu(.)$ exists and is finite. Another useful asymmetric loss function is the general entropy loss which is a generalization of the entropy loss and is given as

$$L_3\left(f(\mu),\hat{f}(\mu)\right)\propto\left(\frac{\hat{f}(\mu)}{f(\mu)}\right)^{-q}-q\ln\left(\frac{\hat{f}(\mu)}{f(\mu)}\right)-1,\quad q\neq 0.$$

For this loss function, overestimation is heavily penalized when q is positive, and vice versa. The Bayes estimator of $f(\mu)$ under general entropy loss function is obtained as

$$\hat{f}_{EB}(\mu)=\left\{E\left[(f(\mu))^{-q}|data\right]\right\}^{-\frac{1}{q}},$$

provided that $E_\mu(.)$ exists and is finite.

4.1. Prior distribution and posterior analysis

In this subsection, we need to assume some prior distributions for the unknown parameters. Under the assumption that two parameters α and λ are unknown, specifying a general conjugate joint prior forα and λ is not easy task. In this case, we develop the Bayesian set-up by considering the idea of [18] regarding the choice of prior distributions. We assume that for $j=1,2,\ldots,M$, λ has a discrete prior say,

$$P(\lambda=\lambda_j)=\theta_j,\ \sum_{j=1}^{M}\theta_j=1, \tag{21}$$

while the conditional distribution of α given λ_j has a conjugate prior distribution, with density

$$g(\alpha|\lambda_j)=\beta_j\exp\{-\alpha\beta_j\},\quad \alpha,\beta_j>0, \tag{22}$$

where $\beta_j,\ \ j=1,2,\ldots,M$, are hyper-parameters. Combining (6) and (22), the conditional posterior of the parameter α,

$$\pi(\alpha|\mathbf{x},\lambda_j)=\frac{g(\alpha|\lambda_j)L(\alpha,\lambda_j;\mathbf{x})}{\int_\alpha g(\alpha|\lambda_j)L(\alpha,\lambda_j;\mathbf{x})\,d\alpha}, \tag{23}$$

takes the form

$$\pi(\alpha|\mathbf{x},\lambda_j)=\frac{1}{\Gamma(m+1)}c_j^{m+1}\alpha^m\exp\{-\alpha c_j\},\ \ j=1,2,\ldots,M, \tag{24}$$

where $c_j = k\sum_{i=1}^{m}(r_i+1)\Psi(x_i;\lambda_j)+\beta_j$. Also by applying (6), (21), (22) and the discrete version of Bayes theorem, the marginal posterior distribution of λ can be expressed as

$$p_j = P(\lambda=\lambda_j|\mathbf{x}) = \frac{\int_\alpha P(\lambda=\lambda_j)g(\alpha|\lambda_j)L(\alpha,\lambda_j;\mathbf{x})\,d\alpha}{\sum_{j=1}^{M}\int_\alpha P(\lambda=\lambda_j)g(\alpha|\lambda_j)L(\alpha,\lambda_j;\mathbf{x})\,d\alpha}$$
$$= \frac{\beta_j\theta_j c_j^{-(m+1)}\prod_{i=1}^{m}\psi(x_i;\lambda_j)}{\sum_{j=1}^{M}\beta_j\theta_j c_j^{-(m+1)}\prod_{i=1}^{m}\psi(x_i;\lambda_j)},\quad j=1,2,\ldots,M. \tag{25}$$

Therefore, the Bayes estimators of α and λ under the squared error loss function L_1 are

$$\hat{\alpha}_{SB} = (m+1)\sum_{j=1}^{M}\frac{p_j}{c_j}, \tag{26}$$

$$\hat{\lambda}_{SB} = \sum_{j=1}^{M}p_j\lambda_j, \tag{27}$$

respectively. Also, the Bayes estimators of $R(t)$ and $h(t)$ against the loss function L_1 are given respectively, by

$$\hat{R}_{SB}(t) = \sum_{j=1}^{M}p_j\big[1+\frac{\Psi(t;\lambda_j)}{c_j}\big]^{-(m+1)}, \tag{28}$$

$$\hat{h}_{SB}(t) = (m+1)\sum_{j=1}^{M}\frac{p_j\psi(t;\lambda_j)}{c_j}. \tag{29}$$

For the loss function L_2, the Bayes estimators of α, λ, $R(t)$ and $h(t)$ are respectively obtained as

$$\hat{\alpha}_{LB} = -\frac{1}{c}\ln\left[\sum_{j=1}^{M}p_j(1+\frac{c}{c_j})^{-(m+1)}\right], \tag{30}$$

$$\hat{\lambda}_{LB} = -\frac{1}{c}\ln\left[\sum_{j=1}^{M}p_j e^{-c\lambda_j}\right], \tag{31}$$

$$\hat{R}_{LB}(t) = -\frac{1}{c}\ln\left[\sum_{j=1}^{M}\sum_{l=0}^{\infty}\frac{(-1)^l}{\Gamma(l+1)}p_j c^l(1+\frac{l\Psi(t;\lambda_j)}{c_j})^{-(m+1)}\right], \tag{32}$$

$$\hat{h}_{LB}(t) = -\frac{1}{c}\ln\left[\sum_{j=1}^{M}p_j(1+\frac{c\psi(t;\lambda_j)}{c_j})^{-(m+1)}\right]. \tag{33}$$

Finally, against the loss function L_3, the Bayes estimators of α, λ, $R(t)$ and $h(t)$ can be expressed as

$$\hat{\alpha}_{EB} = \left[\frac{\Gamma(m+1-q)}{\Gamma(m+1)}\sum_{j=1}^{M}p_j c_j^q\right]^{-\frac{1}{q}},\quad m+1>q, \tag{34}$$

$$\hat{\lambda}_{EB} = \left[\sum_{j=1}^{M}p_j\lambda_j^{-q}\right]^{-\frac{1}{q}}, \tag{35}$$

$$\hat{R}_{EB}(t) = \left[\sum_{j=1}^{M}p_j(1-\frac{q\Psi(t;\lambda_j)}{c_j})^{-(m+1)}\right]^{-\frac{1}{q}}, \tag{36}$$

$$\hat{h}_{EB}(t) = \left[\frac{\Gamma(m+1-q)}{\Gamma(m+1)}\sum_{j=1}^{M}p_j(\frac{c_j}{\psi(t;\lambda_j)})^q\right]^{-\frac{1}{q}},\quad m+1>q, \tag{37}$$

respectively.

4.2. The choice of hyper-parameters

The priors specification are completed by specifying λ_j, θ_j and hyper-parameters β_j, for $j = 1, 2, \ldots, M$, in practice. The values of λ_j and θ_j are fairly straightforward to specify, but sometimes it is not always possible to know the values of the hyper-parameters β_j, in prior (22). In practice, the values of β_j are difficulty to know, since it is necessary to condition prior beliefs about α on each λ_j, $j = 1, 2, \ldots, M$. Thus, the estimation problem for hyper-parameters β_j, $j = 1, 2, \ldots, M$, is considered in this subsection.

There are different methods to estimate the hyper-parameters $\beta_j, j = 1, 2, \ldots, M$. First, we consider the maximum likelihood type-II method (see [19, pp. 99]).

Let $U_i = \Psi(X_i; \lambda), i = 1, 2, \ldots, m$. It can be shown that $U_1 < U_2 < \cdots < U_m$, are the progressive first-failure censored order statistics with censoring scheme $(r_1, r_2, \ldots, r_m)$, from conditional density

$$f_U(u; \alpha) = \alpha e^{-\alpha u}, \quad u > 0. \tag{38}$$

For given λ_j, the marginal PDF and CDF of U are given by

$$f_U(u) = \int_\alpha g(\alpha|\lambda_j) f_U(u; \alpha)\, d\alpha = \frac{\beta_j}{(\beta_j + u)^2}, \quad u > 0, \tag{39}$$

$$F_U(u) = 1 - \frac{\beta_j}{\beta_j + u}, \quad u > 0, \tag{40}$$

respectively. From (1), the log-likelihood function of $\mathbf{U} = (U_1, U_2, \ldots, U_m)$, can be written as

$$logL(\beta_j; \mathbf{u}) = \ln(Ak^m) + nk \ln(\beta_j) - \sum_{i=1}^{m} (k(r_i + 1) + 1) \ln(\beta_j + u_i). \tag{41}$$

By setting the derivative of the log-likelihood function with respective to β_j to zero, the MLE of β_j, is the solution to the likelihood equation $\frac{1}{\beta_j} = H(\beta_j)$, where

$$H(\beta_j) = \frac{1}{nk} \sum_{i=1}^{m} \frac{k(r_i + 1) + 1}{\beta_j + u_i}, \tag{42}$$

and it is unique (see Appendix). Most of the standard iterative process can be used for finding the MLE. We propose a simple iterative scheme to finding the MLE of β_j. Start with an initial guess of β_j, say $\beta_j^{(0)}$, then obtain $\beta_j^{(1)} = 1/H(\beta_j^{(0)})$, and proceeding in this way iteratively to obtain $\beta_j^{(N)} = 1/H(\beta_j^{(N-1)})$. Stop the iterative procedure, when $|\beta_j^{(N)} - \beta_j^{(N-1)}| < \varepsilon$, some pre-assigned tolerance limit.

Another useful alternative method to estimate the hyper-parameters $\beta_j, j = 1, 2, \ldots, M$, is based on the idea of [20]. By applying (22), the expected value of the reliability function $R(t)$ conditional on $\lambda = \lambda_j$, can be written as

$$E(R(t)) = \int_\alpha R(t) g(\alpha|\lambda_j)\, d\alpha = \frac{\beta_j}{\beta_j + \Psi(t; \lambda_j)}, \quad j = 1, 2, \ldots, M. \tag{43}$$

For given time t, by considering $E(R(t)) = \hat{R}(t)$, the estimate of β_j is

$$\hat{\beta_j} = \frac{\hat{R}(t)\Psi(t; \lambda_j)}{1 - \hat{R}(t)}, \quad j = 1, 2, \ldots, M. \tag{44}$$

Similarly, we can use the expected value of the hazard rate function $h(t)$ conditional on $\lambda = \lambda_j, j = 1, 2, \ldots, M$. It can be shown that

$$E(h(t)) = \int_\alpha h(t) g(\alpha|\lambda_j)\, d\alpha = \frac{\psi(t; \lambda_j)}{\beta_j}, \tag{45}$$

and

$$\hat{\beta}_j = \frac{\psi(t;\lambda_j)}{\hat{h}(t)}, \quad j = 1,2,\ldots,M. \tag{46}$$

It is obviously the second method to estimate the hyper-parameters $\beta_j, j = 1,2,\ldots,M$, depend on the value of MLEs $\hat{\alpha}$ and $\hat{\lambda}$. Therefore, the author recommends the first method.

5. Data analysis

To illustrate the above procedures, we present the analysis of one real data set. Also, we report some numerical experiments performed to evaluate behavior of the different estimators.

Example 1.(Real Data) In this example, we analyze a data set from [21], which represents the number of 1000s of cycles to failure for electrical appliances in a life test. The complete data have been used earlier by [22]. They showed that the bathtub-shaped distribution is suitable to fitting the data. The CDF of the bathtub-shaped distribution is form(3), where $\Psi(t;\lambda) = e^{t^\lambda} - 1, t > 0$. It can be shown that $\frac{\Psi'(t;\lambda)}{\Psi(t;\lambda)}$, is strictly increasing in t (see [14]).

Table 1: *progressively first-failure censored sample of size 8 out of 20 groups.*

i	1	2	3	4	5	6	7	8
x_i	0.014	0.034	0.059	0.061	0.069	0.142	0.165	1.270
r_i	4	0	3	0	0	2	3	0

The data are randomly grouped into 20 groups with $k = 3$ items within each group. The progressively first-failure censored sample is given in Table 1. For this example, 12 groups of failure times are censored, and 8 first-failures are observed. By applying, (19) and (20), the 95% exact confidence intervals (CI) for λ, confidence regions (CR) for (α,λ), are obtained and the length of confidence intervals (LCI) and area for confidence regions (ACR) are presented in Table 2, where $A(\lambda) = \sum_{i=1}^{8}(r_i+1)(e^{x_i^\lambda} - 1)$.

Table 2: *The 95% confidence intervals and regions and their some properties for λ and (α,λ).*

j	CI	CR	LCI	ACR
1	$0.3933 < \lambda < 1.7034$	$0.3397 < \lambda < 1.8545$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	1.3101	1.3904
2	$0.3694 < \lambda < 1.4175$	$0.3192 < \lambda < 1.5198$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	1.0481	1.0334
3	$0.3538 < \lambda < 1.3167$	$0.3044 < \lambda < 1.4039$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	0.9629	0.9081
4	$0.2317 < \lambda < 1.0946$	$0.1920 < \lambda < 1.1696$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	0.8629	0.6805
5	$0.1391 < \lambda < 0.9320$	$0.1092 < \lambda < 1.0014$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	0.7929	0.5232
6	$0.1302 < \lambda < 0.9750$	$0.0963 < \lambda < 1.0462$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	0.8448	0.5708
7	$0.0212 < \lambda < 0.7932$	$0.0109 < \lambda < 0.8646$, $\frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)}$	0.7720	0.4094

From Table 2, It is observed that, the 95% optimal confidence interval for λ is $(0.0212, 0.7932)$, and the optimal confidence region for (α,λ) is given by

$$0.0109 < \lambda < 0.8646 \ , \ \frac{1.0114}{A(\lambda)} < \alpha < \frac{5.2012}{A(\lambda)},$$

and $ACR = \int_{0.0109}^{0.8646} \frac{4.1898}{A(\lambda)}\, d\lambda = 0.4094$.

Since there is no prior information about α, to compute the Bayes estimates, we estimate the hyper-parameters $\beta_j, j = 1,2,\ldots,8$, using the maximum likelihood type-II method. The values of β_j and p_j, for each given λ_j and θ_j, $j = 1,2,\ldots,8$, are summarized in Table 3. The MLEs as well as Bayes estimates of α, λ, reliability function $R(t)$, and hazard rate function $h(t)$, for $t = 0.5$, are presented in Table 4.

Table 3: *Prior information, hyper-parameter values and the posterior probabilities.*

j	1	2	3	4	5	6	7	8
λ_j	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75
θ_j	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
β_j	3.5605	3.1398	2.7814	2.4735	2.2073	1.9756	1.7727	1.5942
p_j	0.0308	0.0549	0.0859	0.1206	0.1532	0.1778	0.1897	0.1871

Table 4: *The ML and the Bayes estimates of $\alpha, \lambda, R(t)$ and $h(t)$, with $t = 0.5, c = 1$ and $q = 1$.*

$\hat{\alpha}$	$\hat{\alpha}_{SB}$	$\hat{\alpha}_{LB}$	$\hat{\alpha}_{EB}$	$\hat{\lambda}$	$\hat{\lambda}_{SB}$	$\hat{\lambda}_{LB}$	$\hat{\lambda}_{EB}$
0.4800	0.4252	0.4132	0.3674	0.7200	0.6268	0.6220	0.6099
$\hat{R}(t)$	$\hat{R}_{SB}(t)$	$\hat{R}_{LB}(t)$	$\hat{R}_{EB}(t)$	$\hat{h}(t)$	$\hat{h}_{SB}(t)$	$\hat{h}_{LB}(t)$	$\hat{h}_{EB}(t)$
0.6697	0.6871	0.6833	0.6753	0.7700	0.6584	0.6267	0.5570

Table 5: *The estimated MSE values of the estimators of α and λ.*

n	m	k	C.S	$\hat{\alpha}$	$\hat{\alpha}_{SB}$	$\hat{\alpha}_{LB}$	$\hat{\alpha}_{EB}$	$\hat{\lambda}$	$\hat{\lambda}_{SB}$	$\hat{\lambda}_{LB}$	$\hat{\lambda}_{EB}$
20	10	1	I	0.0026	0.0018	0.0017	0.0018	0.0300	0.0088	0.0082	0.0069
			II	0.0034	0.0025	0.0024	0.0023	0.0089	0.0054	0.0052	0.0047
			III	0.0027	0.0021	0.0021	0.0019	0.0111	0.0059	0.0056	0.0051
		5	I	0.0216	0.0031	0.0029	0.0020	0.0919	0.0088	0.0081	0.0063
			II	0.0025	0.0020	0.0020	0.0016	0.0262	0.0080	0.0075	0.0064
			III	0.0040	0.0025	0.0024	0.0018	0.0342	0.0082	0.0077	0.0065
30	10	1	I	0.0023	0.0018	0.0018	0.0016	0.0463	0.0089	0.0082	0.0067
			II	0.0034	0.0025	0.0024	0.0023	0.0079	0.0049	0.0048	0.0043
			III	0.0024	0.0021	0.0020	0.0018	0.0106	0.0058	0.0056	0.0051
		5	I	0.0658	0.0041	0.0038	0.0024	0.0955	0.0086	0.0079	0.0061
			II	0.0025	0.0021	0.0020	0.0016	0.0216	0.0077	0.0072	0.0062
			III	0.0066	0.0029	0.0028	0.0020	0.0291	0.0079	0.0074	0.0063
30	15	1	I	0.0016	0.0011	0.0011	0.0012	0.0156	0.0073	0.0069	0.0060
			II	0.0021	0.0017	0.0017	0.0016	0.0051	0.0037	0.0036	0.0034
			III	0.0016	0.0013	0.0013	0.0013	0.0059	0.0042	0.0040	0.0038
		5	I	0.0037	0.0016	0.0016	0.0012	0.0390	0.0080	0.0074	0.0061
			II	0.0011	0.0010	0.0010	0.0009	0.0139	0.0068	0.0065	0.0057
			III	0.0018	0.0013	0.0013	0.0011	0.0167	0.0070	0.0066	0.0058

Example 2.(Simulation study) To evaluate the performance of the MLEs and Bayes estimators, a simulation study using Monte Carlo method is performed. In this example, we exclusively focus on the bathtub-shaped distribution. For comparison purpose different n, m, k, and censoring schemes(C.S) are considered. We present the results for $\alpha = 0.1$ and $\lambda = 0.5$. For generating progressively first-failure censored samples, we use the algorithm suggested in [23]. We take into consideration that the progressive first-failure censored order statistics $X^{\mathbf{r}}_{1;m,n,k}, X^{\mathbf{r}}_{2;m,n,k}, ..., X^{\mathbf{r}}_{m;m,n,k}$ is a progressively type-II censored sample from a population with distribution function $1-(1-F(x))^k$. We considered the following censoring schemes:

- Scheme I: $r_m = n - m, r_i = 0$, for $i \neq m$.
- Scheme II: $r_1 = n - m, r_i = 0$, for $i \neq 1$.
- Scheme III: $r_{\frac{m}{2}} = n - m, r_i = 0$, for $i \neq \frac{m}{2}$ if m is even, and $r_{\frac{m+1}{2}} = n - m, r_i = 0$, for $i \neq \frac{m+1}{2}$ if m is odd.

The Bayes estimates are obtained for $c = 1, q = 1$, and λ_j and θ_j were given in previous example. The performance of all estimators has been compared numerically in terms of their mean squared errors (MSEs). In each case, for a particular censoring scheme the estimated MSEs are computed over 10,000 simulations. The simulation study was conducted in R software (R x64 4.0.3) and the R code can be obtained on request from the author. Based on tabulated the estimated MSEs, following conclusions can be drawn from Tables 5 and 6.

1. For all censoring schemes, it can be observed that the Bayes estimators are superior to MLE for the parameters α, λ. We also observe that Bayes estimators of $h(t)$ perform better than MLEs of $h(t)$.
2. It is clearly observed that the performance of all estimators of $R(t)$ are very fine in respect to MSE in all situations.
3. In the case of λ, when n and m are fixed, the censoring scheme $(n - m, 0, \ldots, 0)$ posses the smallest estimated MSE values.

4. For fixed n and k, it is observed that as m increases the performance of all estimators improve in terms of the estimated MSE values.

Table 6: *The estimated MSE values of the estimators of $R(t)$ and $h(t)$, fot $t = 0.75$.*

n	m	k	$C.S$	$\hat{R}(t)$	$\hat{R}_{SB}(t)$	$\hat{R}_{LB}(t)$	$\hat{R}_{EB}(t)$	$\hat{h}(t)$	$\hat{h}_{SB}(t)$	$\hat{h}_{LB}(t)$	$\hat{h}_{EB}(t)$
20	10	1	I	0.0040	0.0038	0.0040	0.0043	0.0039	0.0039	0.0037	0.0033
			II	0.0048	0.0053	0.0056	0.0062	0.0046	0.0036	0.0034	0.0034
			III	0.0038	0.0049	0.0051	0.0055	0.0040	0.0036	0.0034	0.0031
		5	I	0.0060	0.0072	0.0075	0.0082	0.1122	0.0114	0.0100	0.0054
			II	0.0028	0.0048	0.0050	0.0054	0.0084	0.0060	0.0056	0.0040
			III	0.0033	0.0058	.0060	0.0065	0.0176	0.0084	0.0076	0.0048
30	10	1	I	0.0032	0.0042	0.0043	0.0046	0.0058	0.0050	0.0047	0.0036
			II	0.0047	0.0055	0.0058	0.0064	0.0047	0.0038	0.0036	0.0035
			III	0.0034	0.0048	0.0050	0.0054	0.0040	0.0037	0.0036	0.0031
		5	I	0.0118	0.0091	0.0095	0.0107	0.4210	0.0161	0.0135	0.0064
			II	0.0028	0.0050	0.0052	0.0055	0.0078	0.0062	0.0058	0.0040
			III	0.0043	0.0066	0.0069	0.0075	0.0320	0.0103	0.0091	0.0056
30	15	1	I	0.0026	0.0027	0.0027	0.0029	0.0022	0.0020	0.0020	0.0020
			II	0.0032	0.0043	0.0044	0.0048	0.0029	0.0022	0.0022	0.0022
			III	0.0024	0.0035	0.0037	0.0039	0.0023	0.0019	0.0019	0.0018
		5	I	0.0025	0.0047	0.0048	0.0051	0.0210	0.0065	0.0059	0.0036
			II	0.0015	0.0031	0.0031	0.0033	0.0028	0.0029	0.0028	0.0022
			III	0.0018	0.0039	0.0040	0.0042	0.0071	0.0046	0.0043	0.0030

6. Conclusions

Lifetime studies are very important to assess the reliability of products. This article investigates the problem of reliability analysis for a class of an exponential distribution based on progressive first failure censoring. It is note that many well-known and useful lifetime distributions which have wide application in reliability theory and failure time modeling as well as other related fields, are included in this class of exponential distribution. Both classical and Bayesian point estimations have been developed. Additionally, the exact confidence interval and region respectively for α and (α, λ) have been conducted. In the future, we can study the problem of predicting times to failure of units censored in multiple stages in progressive first failure censored sample based on model (2).

References

[1] Balakrishnan, N. and Aggarwala, R. *Progressive Censoring: Theory, Methods, and Applications*. Birkhäuser, Boston, 2000.

[2] Balakrishnan, N. (2007). Progressive censoring methodology: an appraisal (with discussions). *Test* **16**, 211 – 259.

[3] Ahmadi, K., Rezaei, M. and Yousefzadeh, F. (2015). Estimation for the generalized half-normal distribution based on progressive type-II censoring. *J. Stat. Comput. Simul.* **85**, 1128–1150.

[4] Balakrishnan, N. and Cramer, E. *The art of progressive censoring: applications to reliability and quality*. Birkhäuser, New York, 2014.

[5] Wu, J. W., Hung, W. L. and Tsai, C. H. (2003). Estimation of the parameters of the Gompertz distribution under the first-failure censored sampling plan. *Statistics* **37**, 517-525.

[6] Wu, J. W. and Yu, H. Y. (2005). Statistical inference about the shape parameter of the Burr type XII distribution under the failure censored sampling plan. *Appl. Math. Comput.* **163**, 443-482.

[7] Wu, S. J., Kuş, C. (2009). On estimation based on progressive first-failure censored sampling. *Comput. Statist. Data Anal.* **53**, 3659-3670.

[8] Abou-Elheggag, N. A. (2013). Estimation for Rayleigh distribution using progressive first-failure censored data. *J. Statist. Appl. Prob.* **2**, 171–182.

[9] Soliman, A. A., Abd Ellah, A. H., Abo-Elheggag, N. A. and Abd-Elmougod, G. A. (2012). Estimation of the parameters of life for Gompertz distribution using progressive first-failure censored data. *Comput. Statist. Data Anal.* **51**, 2065–2077.

[10] Mahmoud, M. A. W., Soliman, A. A. E., Abd Ellah, A. H. and El-Sagheer, R. M. (2013). Bayesian estimation using MCMC approach based on progressive first-failure censoring from generalized Pareto distribution. *American Journal of Theoretical and Applied Statistics* **2**, 128–141

[11] Dube, M., Krishna, H. and Garg, R. (2016). Generalized inverted exponential distribution under progressive first-failure censoring. *J. Stat. Comput. Simul.* **86**, 1095–1114.

[12] Mohammed, H. S., Ateya, S. F. and AL-Hussaini(2017). Estimation based on progressive first-failure censoring from exponentiated exponential distribution. *J. Appl. Stat.* **44**, 1479-1494.

[13] Bi, Q., Ma, Y. and Gui, W. (2020). Reliability estimation for the bathtub-shaped distribution based on progressively first-failure censoring sampling. *Comm. Statist. Simulation Comput*, In press, doi:10.1080/03610918.2020.1746338.

[14] Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Stat. Prob. Lett.* **49**, 155–161.

[15] Wang, L. and Shi, Y. (2013). Reliability analysis of a class of exponential distribution under record values *J. Comput. Appl. Math* **239**, 367–379.

[16] Varian, H. R. (1975). A Bayesian Approach to Real Estate Assessment, In Studies in Bayesian Econometrics and Statistics in Honor of L. J. Savage (Eds S.E. Feinberge and A. Zellner), North Holland, Amsterdam, 195–208.

[17] Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss function. *J. Amer. Statist. Assoc.* **81**, 446–451.

[18] Soland, R. M. (1969). Bayesian Analysis of the Weibull process with unknown scale and shape parameters. *IEEE Trans. Reliab.* **18**, 181–184.

[19] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second ed., Springer, New Yourk.

[20] Soliman, A. A., Abd Ellah, A. H., and Sultan, K. S. (2006). Comparison of estimates using record statistics from Weibull model: Bayesian and non-Bayesian approaches. *Comput. Statist. Data Anal.* **51**, 2065–2077.

[21] Lawless, J. F. *Statistical Models and Methods for Lifetime Data* Wiley, New York, 2003.

[22] Sarhan, A. M., Hamilton, D. C. and Smith, B. (2012). Parameter estimation for a two-parameter bathtub-shaped lifetime distribution. *Appl. Math. Model.* **36**, 5380–5392.

[23] Balakrishnan, N. and Sandhu, R. A. (1995). A simple algorithm for generating progressively type-II generated samples. *American Statistician* **49**, 229–230.

Appendix

We show that the equation

$$\frac{1}{\beta_j} - H(\beta_j) = 0, \tag{47}$$

where $H(\beta_j)$ is defined in (42), has only one root. By considering

$$Q_1(\beta_j) = \sum_{i=1}^{m} k(r_i+1)\frac{u_i}{\beta_j(\beta_j+u_i)},$$
$$Q_2(\beta_j) = \sum_{i=1}^{m} \frac{1}{\beta_j+u_i},$$

the equation (47) is equivalent to $Q_1(\beta_j) - Q_2(\beta_j) = 0$. The functions $Q_1(\beta_j)$ and $Q_2(\beta_j)$ are strictly decreasing and convex, since

$$\frac{\partial Q_1(\beta_j)}{\partial \beta_j} = -\sum_{i=1}^{m} k(r_i+1)\frac{u_i}{\beta_j^2(\beta_j+u_i)^2} < 0,$$
$$\frac{\partial^2 Q_1(\beta_j)}{\partial \beta_j^2} = 2\sum_{i=1}^{m} k(r_i+1)\frac{u_i^2}{\beta_j^3(\beta_j+u_i)^3} > 0,$$
$$\frac{\partial Q_2(\beta_j)}{\partial \beta_j} = -\sum_{i=1}^{m} \frac{1}{(\beta_j+u_i)^2} < 0,$$
$$\frac{\partial^2 Q_2(\beta_j)}{\partial \beta_j^2} = 2\sum_{i=1}^{m} \frac{1}{(\beta_j+u_i)^3} > 0.$$

Also,

$$\lim_{\beta_j\to 0} Q_1(\beta_j) = +\infty, \quad \lim_{\beta_j\to +\infty} Q_1(\beta_j) = 0,$$
$$\lim_{\beta_j\to 0} Q_2(\beta_j) = \sum_{i=1}^{m} \frac{1}{u_i}, \quad \lim_{\beta_j\to +\infty} Q_2(\beta_j) = 0,$$
$$\lim_{\beta_j\to \infty} \frac{Q_2(\beta_j)}{Q_1(\beta_j)} = +\infty,$$

thus the equation (47), has only one root.

RT&A, No 4 (65)
Volume 16, December 2021

Transmuted Sine-Dagum Distribution and its Properties

K.M. Sakthivel and K. Dhivakar

•

Department of Statistics, Bharathiar University, Coimbatore - 641 046, Tamil Nadu, India.
sakthithebest@buc.edu.in

Abstract

In this paper, we introduce a new four parameters continuous probability distribution called transmuted sine-Dagum distribution obtained through the transmuted Sine-G family introduced by Sakthivel et al. [13]. We have obtained some distributional properties including moments, inverted moments, incomplete moments, central moments and order statistics for proposed model. The reliability measures such as reliability function, hazard rate function, reversed hazard rate function, cumulative distribution function, mean waiting time and mean residual life time are studied in this paper. Further, we have discussed some income inequality measures including Lorenz curve, Bonferroni index and Zenga index. The maximum likelihood method is used to estimate the parameters of the proposed probability distribution. Finally, we demonstrated goodness of fit the proposed model with other suitable models in the literature using real life data sets.

Keywords: Dagum distribution, Sine-G family, Reliability function, Order statistics, Lorenz curve, Maximum likelihood method.

.

I. Introduction

The lifetime model is playing a vital role in different fields such as life sciences, biological sciences, environmental sciences, medicine, finance and actuarial science. The last three decades, the development and applications of new probability distributions for lifetime data are remarkable in the literature. In this information era, the data generated from different fields are voluminous and dynamic in nature. Therefore, the need for generating new family of probability distributions is inevitable. As a result, the generating new family of probability distribution has attracted many researchers. The main advantage of generating new family of probability distributions is provide better flexibility and better fit at the cost of one or more additional parameters. The following are a list of few well-known generating new family of probability distributions: exponential family is introduced by Gupta et al. [6], Marshall-Olkin family is introduced by Marshal and Olkin [9], transmuted family is introduced by Shaw and Bucklay [14], Kumaraswamy-G family is introduced by Cordeiro and Castro [3], Topp-Leone family is introduced by Al-Shomrani et al. [1], Power Lindley-G family is introduced by Hassan and Nassr [7] and gamma-G family is introduced by Cordeiro et al. [4], to mention a few.

Dagum distribution is introduced by Camilo Dagum [5] in the year 1977 for modeling income data. It has been extensively used in various fields including wealth data, reliability analysis, survival analysis and meteorological data. The Dagum distribution is an alternative to log-normal, Pareto and generalized beta distributions. This distribution is also called Burr-XII distribution, particularly in the actuarial literature.

A continuous random variable X is said to have Dagum distribution with three parameters σ, θ and β if its probability density function and cumulative distribution function are given respectively as

$$f(x;\sigma,\theta,\beta) = \sigma\theta\beta\, x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1};\ x>0,\sigma>0,\theta>0 \text{ and } \beta>0. \tag{1}$$

and

$$F(x;\sigma,\theta,\beta) = \left(1+\sigma x^{-\theta}\right)^{-\beta};\ x>0,\sigma>0,\theta>0 \text{ and } \beta>0. \tag{2}$$

where σ is scale parameter, while θ and β are shape parameters. It is to be noted that if σ=1 the Dagum distribution becomes Burr III distribution and if θ=1, the Dagum distribution becomes log-logistic or Fisk distribution.

In this paper, we introduce a new four parameter continious probability distribution namely transmuted Sine-Dagum distribution. This generating probability distribution provides better fit and flexibility for real life problem.

This paper is organized as follows: In Section 1, a brief introduction and need for the generating family of distributions is given. In Section 2, we present the transmuted Sine-G family and the proposed probability distribution namely transmuted Sine-Dagum distribution. In Section 3, we discuss some reliability measures like reliability function, hazard rate function, reversed hazard rate function, cumulative hazard function, mean waiting time, mean residual life function and mean past life time. In Section 4, we present some distributional properties including moments, inverted moments, incomplete moments, central moments and order statistics. The income inequality measures are discussed in Section 5. The method of maximum likelihood estimation is presented in Section 6. In Section 7, the real time application is presented. Finally, the concluding remarks are presented in Section 8.

II. Transmuted Sine-G Family

Transmuted Sine-G family is introduced by Sakthivel and Rajkumar [13]. This transmuted Sine-G family is the mixture of Sine function and quadratic rank transmuted map. The probability density function of transmuted Sine-G family of distributions is given by

$$f(x;\lambda) = \frac{\pi}{2}h(x)cos\left(\frac{\pi}{2}H(x)\right)\left[(1+\lambda)-2\lambda sin\left(\frac{\pi}{2}H(x)\right)\right];x>0,\lambda>0. \tag{3}$$

and the corresponding cumulative distribution function is given by

$$F(x;\lambda) = (1+\lambda)sin\left(\frac{\pi}{2}H(x)\right)-\lambda\left[sin\left(\frac{\pi}{2}H(x)\right)\right]^2;x>0,\lambda>0. \tag{4}$$

where, λ is the parameter of transmuted Sine-G family of distributions. If $\lambda=0$, the transmuted Sine-G family is becomes Sine-G family.

I. Transmuted Sine-Dagum Distribution

A continuous random variable X is said to be follow the transmuted Sine-Dagum distribution with parameters σ, θ, β and λ, (i.e.,) $X \sim TSD(X;\sigma,\theta,\beta,\lambda)$, then the probability density function of X is of the form

$$\begin{aligned} f(x;\sigma,\theta,\beta,\lambda) =& \frac{\pi}{2}\sigma\theta\beta\, x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) \\ & \times\left[(1+\lambda)-2\lambda\, sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right]; \\ & x>0,\sigma>0,\theta>0,\beta>0 \text{ and } -1\le\lambda\le 1 \end{aligned} \tag{5}$$

The above equation can be rewritten as

$$f(x;\sigma,\theta,\beta,\lambda) = (1+\lambda)\frac{\pi}{2}\sigma\theta\beta\, x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\pi\sigma\theta\beta\, x^{-\theta-1} \times \left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) \quad (6)$$

The cumulative distribution function is given by

$$F(x;\sigma,\theta,\beta,\lambda) = \left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]; \quad x>0,\sigma>0,\theta>0,\beta>0 \text{ and } -1\le\lambda\le 1. \quad (7)$$

where θ is scale parameter; σ and β are shape parameters and λ is the parameter of quadratic rank transmutation map.

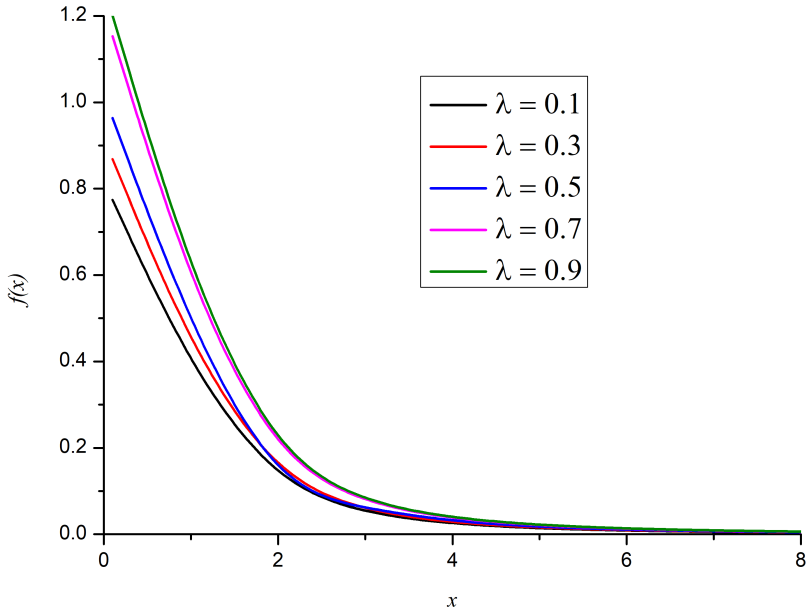


Figure 1: *Pdfs of transmuted Sine-Dagum distribution for fixed values of* $\sigma = 0.5, \theta = 1, \beta = 2$ *and different values of* λ.

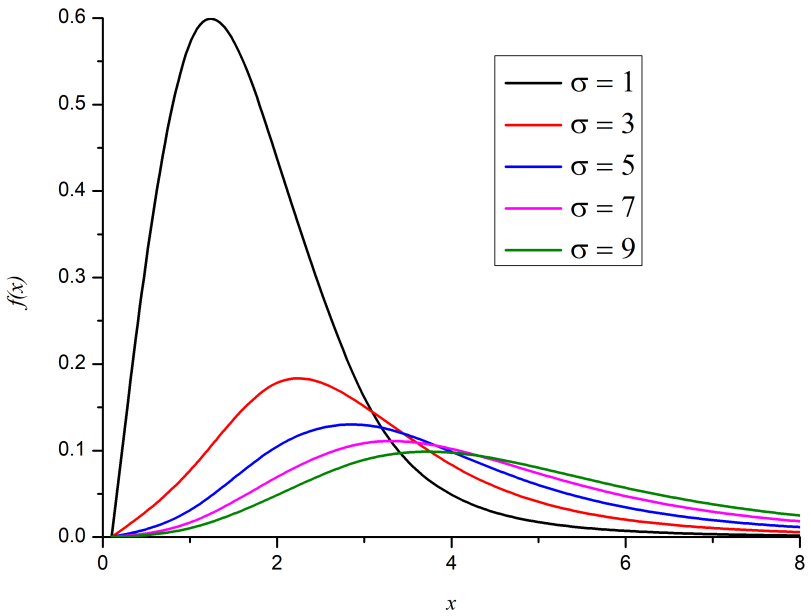


Figure 2: *Pdfs of transmuted Sine-Dagum distribution for fixed values of* $\theta = 2, \beta = 3, \lambda = 0.5$ *and different values of* σ.

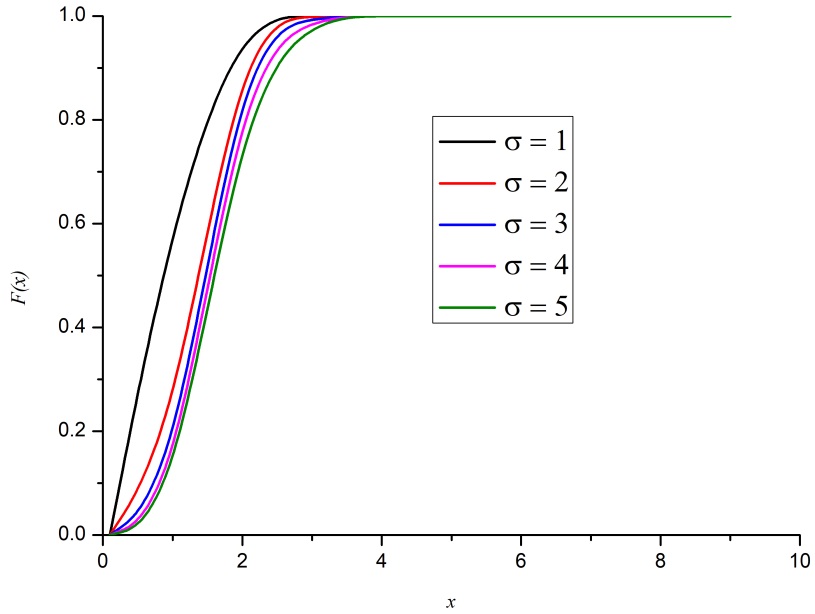


Figure 3: *Cdfs of transmuted Sine-Dagum distribution for fixed values of* $\theta = 4, \beta = 2, \lambda = 1$ *and different values of* σ.

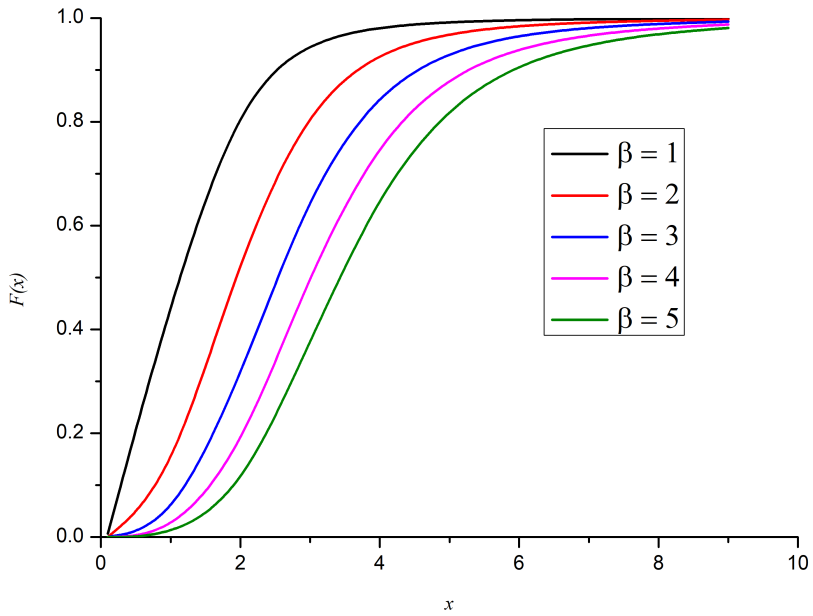


Figure 4: *Cdfs of transmuted Sine-Dagum distribution for fixed values of* $\sigma = 4, \theta = 2, \lambda = 0.7$ *and different values of* β.

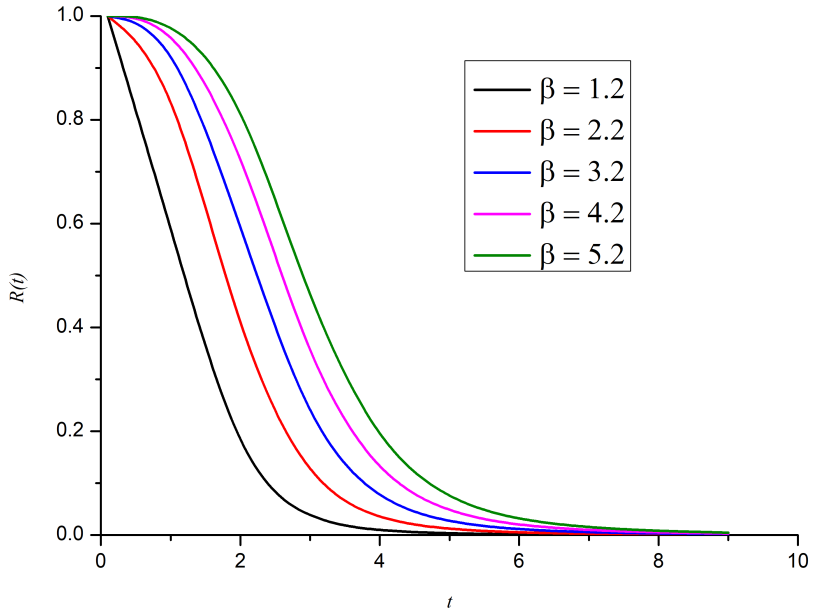


Figure 5: *Reliability function of transmuted Sine-Dagum distribution for fixed values of* $\sigma = 3.5, \theta = 2.2, \lambda = 0.8$ *and different values of* β.

III. Reliability Measures

In this Section, we deal with some reliability measures for transmuted Sine-Dagum distribution. If X∼TSD (X;σ,θ,β,λ), then the reliability measures of random variable X are given by;

I. Reliability function

The reliability function is defined by

$$R(x;\sigma,\theta,\beta,\lambda) = 1 - F(x;\sigma,\theta,\beta,\lambda)$$

$$= 1 - \left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]. \tag{8}$$

II. Hazard rate function

The hazard rate function is defined by

$$h(x,\sigma,\theta,\beta,\lambda) = \frac{f(x,\sigma,\theta,\beta,\lambda)}{1 - F(x,\sigma,\theta,\beta,\lambda)}$$

$$= \frac{\frac{\pi}{2}\sigma\theta\beta\ x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\left[(1+\lambda) - 2\lambda\ sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right]}{1 - \left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]}. \tag{9}$$

III. Reversed hazard rate function

The reversed hazard rate function is given by

$$r(x,\sigma,\theta,\beta,\lambda) = \frac{f(x,\sigma,\theta,\beta,\lambda)}{F(x,\sigma,\theta,\beta,\lambda)}$$

$$= \frac{\frac{\pi}{2}\sigma\theta\beta\ x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\left[(1+\lambda) - 2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right]}{\left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]}. \tag{10}$$

IV. Cumulative hazard function

The cumulative hazard function is defined by

$$H(x,\sigma,\theta,\beta,\lambda) = -log\ R(x,\sigma,\theta,\beta,\lambda)$$

$$= -log\left[1 - \left((1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right)\right]. \tag{11}$$

V. Mean waiting time

The mean waiting time is defined by

$$\varphi(x) = x - \left[\frac{1}{F(x)}\int_0^x sf(s)ds\right]$$

$$\begin{aligned}\varphi(x) =& x - \left[\frac{1}{F(x)}\int_0^x \frac{\pi}{2}\sigma\theta\beta\, s^{-\theta}\left(1+\sigma s^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma s^{-\theta}\right)^{-\beta}\right)\right.\\ &\left.\times\left((1+\lambda) - 2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma s^{-\theta}\right)^{-\beta}\right)\right)\right]ds\end{aligned}$$

Therefore, the mean waiting time of transmuted Sine-Dagum distribution is given by

$$= x - \frac{\begin{array}{c}\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{\pi}{2})^{2n}}{(2n)!}\left[B\left(1-\frac{1}{\theta},2\beta n+\beta+\frac{1}{\theta};y\right)\right.\\ \left.-\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{\pi}{2})^{2n+1}}{(2n+1)!}2\lambda\beta\sigma^{\frac{1}{\theta}}B\left(1-\frac{1}{\theta},4\beta n+2\beta+\frac{1}{\theta};y\right)\right]\end{array}}{\left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]}. \tag{12}$$

VI. Mean residual life function

The mean residual life function is defined by

$$\phi(x) = \frac{1}{S(x)}\int_x^{\infty} xf(x)dx - x$$

$$= \frac{\begin{array}{c}\int_0^{\infty} x\left[\frac{\pi}{2}\sigma\theta\beta\, x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right.\\ \left.\times\left((1+\lambda)-2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)\right]dx - x\end{array}}{1-\left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]}$$

Therefore, the mean residual life function of transmuted Sine-Dagum distribution is given by

$$= \frac{\begin{array}{c}\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{\pi}{2})^{2n}}{(2n)!}\left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{1}{\theta}}B\left(1-\frac{1}{\theta},2\beta n+\beta+\frac{1}{\theta}\right)\right.\\ \left.-\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{\pi}{2})^{2n+1}}{(2n+1)!}\lambda\beta\sigma^{\frac{1}{\theta}}B\left(1-\frac{1}{\theta},4\beta n+2\beta+\frac{1}{\theta}\right)\right]\end{array}}{1-\left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right]} - x. \tag{13}$$

VII. Mean past lifetime

The mean past lifetime of the component can be defined as

$$K(x) = E\left[x - X|X \le x\right] = \frac{\int_0^x F(t)dt}{F(x)} = x - \frac{\int_0^x tf(t)dt}{F(x)}$$

$$K(x) = x - \frac{\begin{array}{c}\int_0^x t\left[\frac{\pi}{2}\sigma\theta\beta\, t^{-\theta-1}\left(1+\sigma t^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma t^{-\theta}\right)^{-\beta}\right)\right.\\ \left.\times\left((1+\lambda)-2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma t^{-\theta}\right)^{-\beta}\right)\right)\right]dt\end{array}}{\left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)^2\right]}$$

Therefore, the mean past life time is given by

$$= x - \frac{\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{1}{\theta}} B\left(1 - \frac{1}{\theta}, 2\beta n + \beta + \frac{1}{\theta}; y\right) - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda \beta \sigma^{\frac{1}{\theta}} B\left(1 - \frac{1}{\theta}, 4\beta n + 2\beta + \frac{1}{\theta}; y\right) \right]}{\left[(1+\lambda) sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda \left(sin \frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2 \right]}. \tag{14}$$

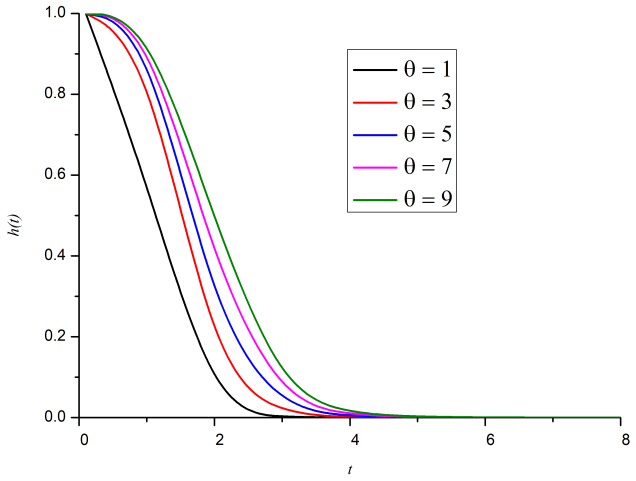


Figure 6: *Reliability function of transmuted Sine-Dagum distribution for fixed values of* $\sigma = 1.4, \beta = 2.4, \lambda = 0.7$ *and different values of* θ.

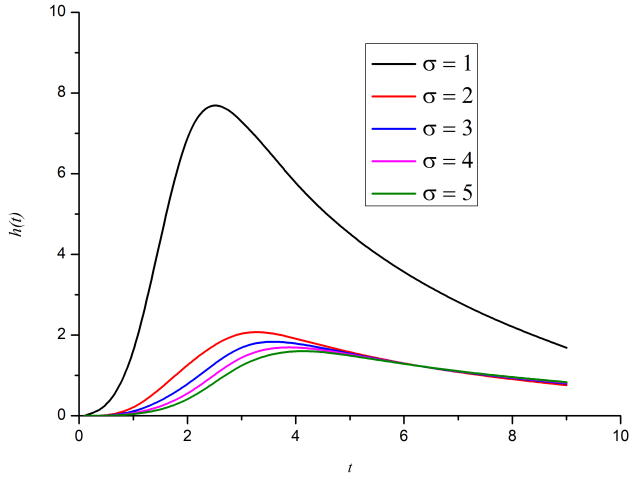


Figure 7: *Hazard rate function of transmuted Sine-Dagum distribution for fixed values of* $\theta = 4, \beta = 6, \lambda = 0.5$ *and different values of* σ.

IV. Distributional Properties

I. Moments

The r^{th} moment of transmuted Sine-Dagum distribution of the random variable X is obtained as

$$\mu_r^{'} = \int_0^{\infty} x^r \left[(1+\lambda)\frac{\pi}{2}\sigma\theta\beta\ x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) \right.$$
$$\left. -\lambda\pi\sigma\theta\beta\ x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right] dx$$

Using the Taylor series of the sine and cosine functions for moments, we have

$$sin\ x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}, \qquad cos\ x = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

Therefore, we have

$$cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}\left(1+\sigma x^{-\theta}\right)^{-2\beta n}}{(2n)!}$$

$$sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}\left(1+\sigma x^{-\theta}\right)^{-2n\beta-\beta}}{(2n+1)!}$$

Hence, the r^{th} moment of transmuted Sine-Dagum distribution is given by

$$\mu_r^{'} = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!}\left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 2\beta n+\beta+\frac{r}{\theta}\right)\right.$$
$$\left. -\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!}\lambda\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 4\beta n+2\beta+\frac{r}{\theta}\right)\right] \tag{15}$$

We have obtained the mean and variance of this distribution as

$$\mu_1^{'} = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!}\left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 2\beta n+\beta+\frac{1}{\theta}\right)\right.$$
$$\left. -\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!}\lambda\pi\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 4\beta n+2\beta+\frac{1}{\theta}\right)\right]$$

and

$$v(x) = \left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!}\left((1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{2}{\theta}} B\left(1-\frac{2}{\theta}, 2\beta n+\beta+\frac{2}{\theta}\right)\right.\right.$$
$$\left.\left. -\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} 2\lambda\pi\beta\sigma^{\frac{2}{\theta}} B\left(1-\frac{2}{\theta}, 4\beta n+2\beta+\frac{2}{\theta}\right)\right)\right] -$$
$$\left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!}\left((1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 2\beta n+\beta+\frac{1}{\theta}\right)\right.\right.$$
$$\left.\left. -\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!}\lambda\pi\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 4\beta n+2\beta+\frac{1}{\theta}\right)\right)\right]^2 \tag{16}$$

The moment generating function of transmuted Sine-Dagum distribution is given by

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left((1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 2\beta n + \beta + \frac{r}{\theta}\right) \right.\right.$$
$$\left.\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} 2\lambda\pi\beta\sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}\right)\right)\right] \tag{17}$$

The characteristic function is given by

$$\Phi_X(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left((1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 2\beta n + \beta + \frac{r}{\theta}\right) \right.\right.$$
$$\left.\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} 2\lambda\beta\sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}\right)\right)\right] \tag{18}$$

and the cumulant generating function is given by

$$K_X(t) = log\left[\sum_{r=0}^{\infty} \frac{t^r}{r!} \left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 2\beta n + \beta + \frac{r}{\theta}\right) \right.\right.\right.$$
$$\left.\left.\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda\pi\beta\sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}\right) \right]\right]\right]. \tag{19}$$

II. Inverted moments

The inverted moment is defined by

$$\mu_r^* = \int_{-\infty}^{\infty} x^{-r} f(x) dx$$

Thus, the inverted moment for this distribution is given by

$$\mu_r^* = \int_0^{\infty} x^{-r} \left[(1+\lambda) \frac{\pi}{2} \sigma\theta\beta \; x^{-\theta-1} \left(1 + \sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\pi\sigma\theta\beta \; x^{-\theta-1} \right.$$
$$\left. \times \left(1 + \sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right)\right] dx$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{r}{\theta}} B\left(1 + \frac{r}{\theta}, 2\beta n + \beta - \frac{r}{\theta}\right) \right.$$
$$\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda\pi\beta\sigma^{-\frac{r}{\theta}} B\left(1 + \frac{r}{\theta}, 4\beta n + 2\beta - \frac{r}{\theta}\right)\right]. \tag{20}$$

III. Incomplete moments

The r^{th} incomplete moment is defined by

$$m_r(x) = \int_0^x s^r f(s) ds$$

$$m_r(x) = \int_0^x s^r \left[\frac{\pi}{2} \sigma\theta\beta \; s^{-\theta-1} \left(1 + \sigma s^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1 + \sigma s^{-\theta}\right)^{-\beta}\right) - \lambda\pi\sigma\theta\beta \; s^{-\theta-1} \right.$$
$$\left. \times \left(1 + \sigma s^{-\beta-1}\right) cos\left(\frac{\pi}{2}\left(1 + \sigma s^{-\theta}\right)^{-\beta}\right) sin\left(\frac{\pi}{2}\left(1 + \sigma s^{-\theta}\right)^{-\beta}\right)\right] ds$$

Hence, the r^{th} incomplete moments of transmuted Sine-Dagum distribution is given by

$$= \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 2\beta n + \beta + \frac{r}{\theta}; y\right) \right.$$
$$\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\pi/2)^{2n+1}}{(2n+1)!} 2\lambda \beta \sigma^{\frac{r}{\theta}} B\left(1 - \frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}; y\right) \right]. \quad (21)$$

IV. Central moments

The central moment is defined by

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu_1^{'})^r f(x) dx = \sum_{m=0}^{r} \binom{r}{m} (-1)^m (\mu_1^{'})^m \mu_{r-m}^{'}$$

Therefore, the central moments of transmuted Sine-Dagum distribution is given by

$$\mu_r = \sum_{m=0}^{r} \binom{r}{m} (-1)^m \times \left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{1}{\theta}} B\left(1 - \frac{1}{\theta}, 2\beta n + \beta + \frac{1}{\theta}\right) \right.\right.$$
$$\left.\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda \pi \beta \sigma^{\frac{1}{\theta}} B\left(1 - \frac{1}{\theta}, 4\beta n + 2\beta + \frac{1}{\theta}\right) \right] \right]$$
$$\times \left[\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda) \frac{\pi}{2} \beta \sigma^{\frac{r-m}{\theta}} B\left(1 - \frac{r-m}{\theta}, 2\beta n + \beta + \frac{r-m}{\theta}\right) \right.\right.$$
$$\left.\left. - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda \pi \beta \sigma^{\frac{r-m}{\theta}} B\left(1 - \frac{r-m}{\theta}, 4\beta n + 2\beta + \frac{r-m}{\theta}\right) \right] \right]. \quad (22)$$

V. Order statistics

The pdf of the j^{th} order statistics for transmuted Sine-Dagum distribution is given by

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)(n-j)!} \left[\frac{\pi}{2} \sigma\theta\beta \; x^{-\theta-1} \left(1 + \sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) \right.$$
$$\left. \times \left((1+\lambda) - 2\lambda sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right)\right) \right]$$
$$\times \left[(1+\lambda) sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) - \lambda \left(sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right)\right)^2 \right]^{j-1}$$
$$\times \left[1 - \left((1+\lambda) sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) - \lambda \left(sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right)\right)^2 \right) \right]^{n-1} \quad (23)$$

The pdf of the smallest order statistics $X_{(1)}$ is given by

$$f_{X_{(1)}}(x) = n \left[1 - \left((1+\lambda) sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) - \lambda \left(sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right)\right)^2 \right) \right]^{n-1}$$
$$\times \left[\frac{\pi}{2} \sigma\theta\beta \; x^{-\theta-1} \left(1 + \sigma x^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right) \right.$$
$$\left. \times \left((1+\lambda) - 2\lambda sin\left(\frac{\pi}{2}\left(1 + \sigma x^{-\theta}\right)^{-\beta}\right)\right) \right] \quad (24)$$

The pdf of the largest order statistics $X_{(n)}$ is given by

$$\begin{aligned} f_{X_{(n)}}(x) =& n\left[\left((1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right)\right]^{n-1} \\ &\times\left[\frac{\pi}{2}\sigma\theta\beta\ x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1}cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right. \\ &\left.\times\left((1+\lambda)-2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)\right] \end{aligned} \tag{25}$$

The pdf of the median order statistics is given by

$$\begin{aligned} f_{m+1:n}(x) =& \frac{(2m+1)}{m!m!}\left[\left((1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right)\right]^m \\ &\times\left[1-\left((1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)-\lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2\right)\right]^m \\ &\times\left[\frac{\pi}{2}\sigma\theta\beta\ x^{-\theta-1}\left(1+\sigma x^{-\theta}\right)^{-\beta-1}cos\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right. \\ &\left.\times\left((1+\lambda)-2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)\right]. \end{aligned} \tag{26}$$

V. Income inequality measures

In this section, we deal with different basic income inequality measures including Lorenz curve, Bonferroni index and Zenga index. The following measures are given below.

I. Lorenz curve

The lorenz curve was introduced by Lorenz [8] in the year 1905. it is widely used in economic and many other fields and its defined by

$$L(x) = \frac{1}{\mu}\int_0^x sf(s)ds$$

Hence, the Lorenz curve of transmuted Sine-Dagum distribution is given by

$$\begin{aligned} L(x) =& \frac{1}{\mu}\int_0^x s\left[\frac{\pi}{2}\sigma\theta\beta\ s^{-\theta-1}\left(1+\sigma s^{-\theta}\right)^{-\beta-1}cos\left(\frac{\pi}{2}\left(1+\sigma s^{-\theta}\right)^{-\beta}\right)\right. \\ &\left.\times\left((1+\lambda)-2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma s^{-\theta}\right)^{-\beta}\right)\right)\right]dx \end{aligned}$$

$$\begin{aligned} =& \frac{\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{\pi}{2})^{2n}}{(2n)!}\left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{r}{\theta}}B\left(1-\frac{r}{\theta},2\beta n+\beta+\frac{r}{\theta};y\right)\right.}{\left.-\sum_{n=0}^{\infty}\frac{(-1)^n(\pi/2)^{2n+1}}{(2n+1)!}2\lambda\beta\sigma^{\frac{r}{\theta}}B\left(1-\frac{r}{\theta},4\beta n+2\beta+\frac{r}{\theta};y\right)\right]} \\ &\frac{\sum_{n=0}^{\infty}\frac{(-1)^n(\frac{\pi}{2})^{2n}}{(2n)!}\left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{1}{\theta}}B\left(1-\frac{1}{\theta},2\beta n+\beta+\frac{1}{\theta}\right)\right.}{\left.-\sum_{n=0}^{\infty}\frac{(-1)^n(\pi/2)^{2n+1}}{(2n+1)!}2\lambda\beta\sigma^{\frac{1}{\theta}}B\left(1-\frac{1}{\theta},4\beta n+2\beta+\frac{1}{\theta}\right)\right]}. \end{aligned} \tag{27}$$

II. Bonferroni index

The Bonferroni index was introduced by Bonferroni [2] in the year 1930. it is ratio of Lorenz curve and cumulative distribution function of the distribution. The Bonferroni index is defined as

$$B(x) = \frac{L(x)}{F(x)}$$

Hence, the Bonferroni index of transmuted Sine-Dagum distribution is given by

$$B(x) = \frac{\delta}{\eta}$$

where

$$\delta = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}; y\right) - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda\pi\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}; y\right) \right]$$

$$\eta = \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 2\beta n + \beta + \frac{1}{\theta}\right) - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} 2\lambda\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 4\beta n + 2\beta + \frac{1}{\theta}\right) \right] \times \left[(1+\lambda) sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)^2 \right].$$

III. Zenga index

Zenga index is introduced by Zenga [15] in the year 1980. The Zenga index is defined by

$$Z = 1 - \frac{\bar{\mu}_{(x)}}{\mu^{+}_{(x)}}$$

where

$$\bar{\mu}_{(x)} = \frac{1}{F(x)} \int_0^x s f(s) ds$$
$$\mu^{+}_{(x)} = \frac{1}{1-F(x)} \int_0^{\infty} x f(x) dx$$

Therefore, we get

$$\bar{\mu}_{(x)} = \frac{\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 2\beta n + \beta + \frac{r}{\theta}; y\right) - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda\pi\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 4\beta n + 2\beta + \frac{r}{\theta}; y\right) \right]}{\left[(1+\lambda) sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)^2 \right]}$$

$$\mu^{+}_{(x)} = \frac{\sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 2\beta n + \beta + \frac{1}{\theta}\right) - \sum_{n=0}^{\infty} \frac{(-1)^n (\frac{\pi}{2})^{2n+1}}{(2n+1)!} \lambda\pi\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 4\beta n + 2\beta + \frac{1}{\theta}\right) \right]}{1 - \left[(1+\lambda) sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)^2 \right]}$$

Hence, the Zenga index of transmuted Sine-Dagum distribution is given by

$$Z = 1 - \frac{\gamma}{\delta}$$

where

$$\gamma = \sum_{n=0}^{\infty} \frac{(-1)^n(\frac{\pi}{2})^{2n}}{(2n)!} \left[(1+\lambda)\frac{\pi}{2}\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 2\beta n+\beta+\frac{r}{\theta}; y\right)\right.$$
$$\left. - \sum_{n=0}^{\infty} \frac{(-1)^n(\frac{\pi}{2})^{2n+1}}{(2n+1)!}\lambda\pi\beta\sigma^{\frac{r}{\theta}} B\left(1-\frac{r}{\theta}, 4\beta n+2\beta+\frac{r}{\theta}; y\right)\right]$$
$$\times \left[1-\left((1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)^2\right)\right]$$

$$\delta = \sum_{n=0}^{\infty} \frac{(-1)^n(\frac{\pi}{2})^{2n}}{(2n)!} \left[\frac{(1+\lambda)\pi}{2}\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 2\beta n+\beta+\frac{1}{\theta}\right)\right.$$
$$\left. - \sum_{n=0}^{\infty} \frac{(-1)^n(\frac{\pi}{2})^{2n+1}}{(2n+1)!}\lambda\pi\beta\sigma^{\frac{1}{\theta}} B\left(1-\frac{1}{\theta}, 4\beta n+2\beta+\frac{1}{\theta}\right)\right]$$
$$\times \left[(1+\lambda)sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right) - \lambda\left(sin\left(\frac{\pi}{2}\left(1+\sigma x^{-\theta}\right)^{-\beta}\right)\right)^2\right].$$

VI. Parameter Estimation

Let $x_1, x_2, \ldots, x_n$ be a random sample from the transmuted Sine-Dagum distribution then the likelihood function is given by

$$L(\sigma,\theta,\beta,\lambda;\underline{x}) = \prod_{i=1}^{n}\left[\frac{\pi}{2}\sigma\theta\beta\, x_{(i)}^{-\theta-1}\left(1+\sigma_{(i)}^{-\theta}\right)^{-\beta-1} cos\left(\frac{\pi}{2}\left(1+\sigma_{(i)}^{-\theta}\right)^{-\beta}\right)\right.$$
$$\left. \times \left((1+\lambda) - 2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma_{(i)}^{-\theta}\right)^{-\beta}\right)\right)\right] \quad (28)$$

Hence, the log likelihood function is given by

$$L(\sigma,\theta,\beta,\lambda;\underline{x}) = n\, log\frac{\pi}{2} + n\, log\sigma + n\, log\theta + n\, log\beta + (-\theta-1)\sum_{i=1}^{n} logx_{(i)}$$
$$+ (-\beta-1)\sum_{i=1}^{n} log\left(1+\sigma x_{(i)}^{-\theta}\right) + \sum_{i=1}^{n} log\, cos\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)$$
$$+ \sum_{i=1}^{n} log\, cos\left((1+\lambda) - 2\lambda\, sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)\right)$$

The MLE of parameters σ, θ, β and λ are obtained from the following equations

$$\frac{\partial logL}{\partial\sigma} = 0, \frac{\partial logL}{\partial\theta} = 0, \frac{\partial logL}{\partial\beta} = 0 \text{ and } \frac{\partial logL}{\partial\lambda} = 0$$

That is,

$$\frac{\partial logL}{\partial\sigma} = \frac{n}{\sigma} + \sum_{i=1}^{n}\frac{(-\beta-1)x_{(i)}^{-\theta}}{(1+\sigma x_i^{-\theta})} + \sum_{i=1}^{n}\frac{sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)\frac{\pi}{2}\beta\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta-1}x_{(i)}^{-\theta}}{cos\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)} = 0 \quad (29)$$

$$\frac{\partial logL}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} logx_{(i)} + \sum_{i=1}^{n} \frac{(-\beta-1)x_{(i)}^{-\theta} log\ x_{(i)}}{\left(1+\sigma x_{(i)}^{-\theta}\right)} + \sum_{i=1}^{n} \frac{sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right) \frac{\pi}{2}\beta\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta-1} \sigma x_{(i)}^{-\theta} log\ x_{(i)}}{cos\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)} - \sum_{i=1}^{n} \frac{2\lambda cos\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right) \beta\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta-1} \sigma x_{(i)}^{-\theta} logx_{(i)}}{cos\left((1+\lambda) - 2\lambda\ sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)\right)} = 0 \tag{30}$$

$$\frac{\partial logL}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^{n} log\left(1+\sigma x_i^{-\theta}\right)^{-\beta} - \sum_{i=1}^{n} \frac{sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right) \frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta} log\left(1+\sigma x_{(i)}^{-\theta}\right)}{cos\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)} = 0 \tag{31}$$

and

$$\frac{\partial logL}{\partial \lambda} = \frac{1 - 2sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)}{(1+\lambda) - 2\lambda sin\left(\frac{\pi}{2}\left(1+\sigma x_{(i)}^{-\theta}\right)^{-\beta}\right)} = 0. \tag{32}$$

The above mentioned four nonlinear equations are difficult to solve analytically. Therefore, these equations can be solved through iteration methods like Newton-Raphshon method etc., However, we estimate the parameters using *R* software.

VII. Applications

The data set is about the time-to-failure of a 100 cm polyster/viscose in a textile experiment to evaluate the tensile fatigue characteristics of the yarn when its strain level is 2.3 percentage. This data set is used early by Quesenberry and Kent [12], Pal and Tiensuwan [11] and Nasiru et al. [10].

We have fitted the model based on the minimum value of different goodness of fit measures values represented by -2log likelihood, corrected Akaike Information Criterion (CAIC), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC). We have compared the Topp-Leone generated Dagum distribution with other competitive statistical models like Exponentiated generalized exponential Dagum distribution (EGEDD), Exponentiated generalized Dagum distribution (EGDD), Dagum distribution (DD), Exponentiated generalized exponential Burr distribution (EGEBD), Exponentiated generalized Burr distribution (EGBD), Exponentiated generalized exponential Frechet distribution (EGEFD), Exponentiated generalized Frechet distribution (EGFD), Mc-Dagum distribution (McD), exponentiated Kumaraswamy Dagum distribution (EKDD) for this yarn data. The transmuted Sine-Dagum distribution provides better fit for tensile fatigue characteristics of the yarn data compared to other above mentioned competitive statistical models. The details are given in the following tables.

Table 1: *Summary of statistics for tensile fatigue characteristic of the yarn data*

n	Mean	Median	Minimum	Maximum	Q_1	Q_3
100	222.0	195.5	15.0	829.0	129.2	283.0

Table 2: *The values of estimated parameters for tensile fatigue characteristic of the yarn data*

Model	Estimated value of the parameters
TSDD	$\hat{\sigma}$=10868.26, $\hat{\theta}$=1.732, $\hat{\beta}$=1.0574, λ=-0.473
EGEDD	$\hat{\sigma}$=0.026, $\hat{\sigma}$=75 .310, $\hat{\beta}$=0.017, $\hat{\theta}$=3.513, $\hat{c}$=45.692, $\hat{d}$=0.090
EGDD	$\hat{\alpha}$=1.992, $\hat{\beta}$=10.480, $\hat{\theta}$=4.733, $\hat{c}$=75.487, $\hat{d}$=0.223
DD	$\hat{\alpha}$=19.749, $\hat{\beta}$=11.599, $\hat{\theta}$=1.126
EGEBD	$\hat{\lambda}$=35.463, $\hat{\beta}$=35.965, $\hat{\theta}$=4.859, $\hat{c}$=15.667, $\hat{d}$=0.070
EGBD	$\hat{\beta}$=24.801, $\hat{\theta}$=4.196, $\hat{c}$=73.9120, $\hat{d}$=0.258
EGEFD	$\hat{\alpha}$=20.662, $\hat{\lambda}$=34.477, $\hat{\theta}$=5.217, $\hat{c}$=16.438, $\hat{d}$=0.65
EGFD	$\hat{\alpha}$=10.537, $\hat{\theta}$=5.239, $\hat{c}$=21.341, $\hat{d}$=0.140
McD	$\hat{\lambda}$=0.027, $\hat{\delta}$=0.600, $\hat{\beta}$=98.780, $\hat{a}$=0.333, $\hat{b}$=25.042, $\hat{c}$=46.276
EKD	$\hat{\alpha}$=546.109, $\hat{\lambda}$=39.413, $\hat{\delta}$=5.188, $\hat{\phi}$=0.203, $\hat{\theta}$=31.169

Table 3: *Statistical model selection for tensile fatigue characteristic of the yarn data*

Model	-2LL	AIC	AICC	BIC
TSDD	1252.689	1260.689	1261.11	1271.11
EGEDD	1256.34	1268.336	1269.553	1283.967
EGDD	1306.14	1316.137	1317.040	1329.163
DD	1298.52	1304.517	1304.938	1312.333
EGEBD	1261.74	1271.745	1272.648	1284.771
EGBD	1306.06	1314.056	1314.694	1324.447
EGEFD	1261.52	1271.523	1272.426	1284.549
EGFD	1333.76	1341.757	1342.395	1352.177
McD	1256.4	1268.399	1269.616	1284.030
EKD	1307.92	1317.913	1318.816	1330.938

VIII. Conclusion

In this paper, we have presented a new transmuted Sine-Dagum distribution using transmuted Sine-G family of distributions. We have studied some reliability measures like reliability function, hazard rate function, reverse hazard rate function, cumulative hazard rate function, second failure rate function, mean waiting time, mean past lifetime and mean residual life. We have obtained some distributional properties like moments, moment generating function, characteristic function, cumulant generating function, incomplete moments, central moments and order statistics. We have also investigated some income inequality measures including Lorenz curve, Bonferroni index and Zenga index for proposed new probability distribution. The maximum likelihood method is used to estimate the parameters of proposed new probability distribution. Finally, we have analysed a real lifetime data set for proposed probability distribution. The proposed distribution fits well for this data compared to other competitive models.

References

[1] Al Shomrani, A., Arif, O., Shawky, A., Hanif, S., and Shahbaz, M.Q., (2016). Topp -Leone family of distributions: some properties and applications, Pakistan Journal of Statistics and Operation Research, Vol. 12(3), pp. 443-451.

[2] Bonferroni, C., (1930). Elmenti di statistica generale, Libreria Seber, Firenze.

[3] Cordeiro G.M., and De Castro, M., (2013). A new family of generalized distributions, Journal of Statistical Computation and Simulation, Vol. 81, pp. 883-893.

[4] Cordeiro, G.M., Silva, R.B., and Nascimento A.D.C., (2020). The gamma-G family of distributions, Recent Advances in Lifetime and Reliability Models, Vol. 29, pp. 149- 177.

[5] Dagum, C., (1977). A new model of personal income distribution: specification and estimation, Economie Appliquée, Vol. 30, pp. 413-437.

[6] Gupta, R. C., Gupta P.L., and Gupta, R.D., (1998). Modeling failure time data by Lehmann alternatives, Communication in Statistics Theory and Methods, Vol. 27, pp. 887-904.

[7] Hassan, A.S., and Nassr, S.G., (2019). Power Lindley-G family of distribution, Annals of Data Science, Vol. 16(2), pp. 189-210.

[8] Lorenz, M.O., (1905). Methods of measuring the concentration of wealth, American Statistical Association, Vol. 9, pp. 209-219.

[9] Marshall, A.W., and Olkin, I., (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, Biometrika, Vol. 84(3), pp. 641-652.

[10] Nasiru, S., Mwita, P.N., and Ngesa, O., (2019). Exponentiated generalized exponential Dagum distribution, Journal of King Saud University-Science, Vol. 31(3), pp. 362-371.

[11] Pal, M., and Tiensuwan, M., (2014). The beta transmuted exponentiated Weibull geometric distribution, Austrian Journal of Statistics, Vol. 43 (2), pp. 133–149.

[12] Quesenberry, C.P., and Kent, J., (1982). Selecting among probability distributions used in reliability, Technometrics, Vol. 24 (1), pp. 59–65.

[13] Sakthivel, K.M., and Rajkumar, J., (2021). Transmuted sine-G family of distributions: theory and applications, Statistics and Applications, (Accepted: 10 August 2021).

[14] 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, Research Report.

[15] Zenga, La curtosi (Kurtosis), (1996). Statistica, Vol. 56, pp. 87-101.

* * *

EPQ MODELS WITH MIXTURE OF WEIBULL PRODUCTION EXPONENTIAL DECAY AND CONSTANT DEMAND

V. Sai Jyothsna Devi[1], K. Srinivasa Rao[2]

•

Department of Statistics[1, 2], Andhra University, Visakhapatnam, India
jyothsna.vudatha@gmail.com[1], ksraoau@yahoo.co.in[2]

Abstract

This paper deals with an economic production quantity (EPQ) model in which production is random and having heterogeneous units of production. The production process is characterized by mixture of Weibull distribution. It is assumed that the demand is constant and the lifetime of the commodity is random and follows an exponential distribution. Assuming that the shortages are allowed and fully backlogged the instantaneous state of inventory in the production unit is derived. The minimizing the expected total production cost, the optimal production quantity, the production uptime and downtime are derived. Through sensitivity analysis it is observed that the random production with mixture distribution have a significant influence on the optimal production schedules and production quantity. It is also observed that the rate of deterioration can tremendously influence the optimal operating policies of the system. This model also includes some of the earlier models as particular cases. The model is extended to the case of without shortages. A comparison of the two models reveals that allowing shortages will reduce expected total cost of the model.

Keywords: Stochastic production, Mixture of Weibull Distribution, Exponential decay, Production Schedules, Sensitivity analysis.

I. Introduction

In production quantity models much emphasis is given for the lifetime of the commodity. In many production processes the lifetime of the commodity is random and can be characterized by a probability distribution. The literature on inventory models for deteriorating items are reviewed by Pentico and Drake (2011), Ruxian Lie et al (2010), Goyal and Giri (2001), Raafat (1991) and Nahmias (1982). The exponential decay models of inventory are studied by Ghare and Schrader (1963), Shah and Jaiswal (1977), Cohen (1977), Aggarwal (1978), Dave and Shah (1982), Pal (1990), Kalpakam and Sapna (1996), Giri and Chaudhari (1999). The exponential decay is used when the rate of deterioration is constant which coincide with the deterioration of several perishable items such as medicine, sea foods, vegetable oils, cement and paints. Hence it is reasonable to assume exponential decay of the product.

Another important consideration in EPQ models is the rate of production and it is studied by several authors Perumal and Arivarignan (2002), Pal and Mandal (1997), Sen and Chakrabarthy (2007), Lin and Gong(2006), Maity et al(2007), Hu and Liu(2010), Uma Maheswararao et al (2010), Venkata Subbaiah et al (2011), Essey and Srinivasa Rao (2012), Ardak and Borade (2017), Anindya Mandal, Brojeswar Pal and Kripasindhu Chaudhuri (2020), Sunit Kumar, Sushil Kumar and

Rachna Kumari (2021). In all these papers they assumed that the production is deterministic and having finite rate. However, in many production processes the production is not deterministic and random.

Stochastic production is a reality in the modern technological industrial developments. One of the major consideration for scheduling the production and determining the optimal production quantity lies on several factors such as availability of raw material, power supply, man power skill level, machine tool wear which are governed by laws of chance and become stochastic. Because of the stochastic factors the production process in many industries is random and can be characterized by a probability distribution.

Recently Sridevi et al. (2010), Srinivasa Rao et al. (2010), Laxmana Rao et al. (2015), Srinivasa Rao et al. (2017), Madhulatha et al. (2017), Punyavathi et al. (2020) have developed and analyzed production quantity models with random production. In all these papers they assumed that the production is homogeneous even though governed by stochastic nature i.e., all the production is done in one unit or in a single machine. But in practice several of the products are produced by different machines or in different units which are operated under different conditions. Hence these heterogeneous production processes can be characterized by mixes of probability distributions. It is also observed that in each unit the production rate may be increasing/decreasing/remains constant. This type of variable rate of production can be represented by Weibull probability distribution. Hence in this paper we develop and analyze stochastic production quantity models assuming that the production is random and follows a two component Weibull mixture distribution. It is also further assumed that the demand is constant and in the production backorders are allowed and fully backlogged.

Using the differential equations the production quantity at a given time is derived. With suitable costs the total expected production cost is derived. By minimizing the total expected production cost the optimal production schedules, the production quantities are derived. Through sensitivity analysis the effect of the change in parameters and cost on optimal production schedules and production quantity is discussed. This model is extended to the case of without shortages.

II. Assumptions

For developing the model the following assumptions are made:

- The demand rate is constant say k. (1)
- The production is random and follows a mixture of two-parameter Weibull distribution. The instantaneous rate of production is:

$$R(t) = \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}; \alpha_1, \alpha_2 > 0, \beta_1, \beta_2 > 0, 0 \le p \le 1 \qquad (2)$$

- Lead time is zero.
- Cycle length is T. It is known and fixed.
- Shortages are allowed and fully backlogged.
- A deteriorated unit is lost.
- The lifetime of the item is random and follows a exponential distribution with probability density function:

$$f(t) = \theta e^{-\theta t};\ \theta > 0, t > 0$$

Therefore the instantaneous rate of deterioration is

$$h(t) = \theta;\ \theta > 0 \qquad (3)$$

The following notations are used for developing the model.
Q is the production quantity
A is setup cost
C is cost per unit
h Inventory holding cost per unit per unit time
π Shortages cost per unit per unit time

III. EPQ Model with Shortages

Consider a production system in which the stock level is zero at time $t = 0$. The stock level increases during the period $(0, t_1)$, due to production after fulfilling the demand and deterioration. The production stops at time t_1 when stock level reaches S. The inventory decreases gradually due to demand and deterioration in the interval (t_1, t_2). At time t_2 the inventory reaches zero and backorders accumulate during the period (t_2, t_3). At time t_3 the production again starts and fulfills the backlog after satisfying the demand. During (t_3, T) the backorders are fulfilled and inventory level reaches zero at the end of cycle T. The Schematic diagram representing the inventory level is given in Figure 1.

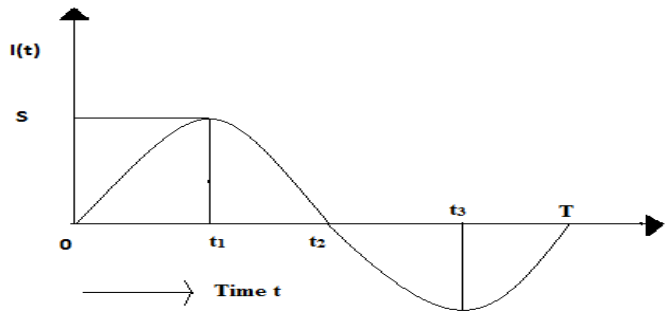


Figure 1: *Schematic Diagram representing the inventory level*

Let *I(t)* be the inventory level of the system at time '*t*' ($0 \le t \le T$). The differential equations governing the instantaneous state of *I(t)* over the cycle of length are:

$$\frac{d}{dt}I(t) + h(t)I(t) = \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}} - k; 0 \le t \le t_1 \quad (4)$$

$$\frac{d}{dt}I(t) + h(t)I(t) = -k;\ t_1 \le t \le t_2 \quad (5)$$

$$\frac{d}{dt}I(t) = -k;\ t_2 \le t \le t_3 \quad (6)$$

$$\frac{d}{dt}I(t) = \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}} - k;\ t_3 \le t \le T \quad (7)$$

Where, *h(t)* is as given in equation (3), with the initial conditions $I(0) = 0$, $I(t_1) = S$, $I(t_2) = 0$ and $I(T) = 0$
Solving the differential equations, the on hand inventory at time '*t*' is obtained as:

$$I(t) = Se^{\theta(t_1-t)} - e^{-t\theta}\int_t^{t_1}\left[\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}} - k\right]e^{u\theta}du\,; 0 \le t \le t_1 \quad (8)$$

$$I(t) = Se^{\theta(t_1-t)} - ke^{-t\theta}\int_{t_1}^{t} e^{u\theta}\,du;\ t_1 \le t \le t_2 \quad (9)$$

$$I(t) = k(t_2 - t); t_2 \le t \le t_3 \quad (10)$$

$$I(t) = \int_t^T \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}dt + k(T-t); t_3 \le t \le T \quad (11)$$

Production quantity Q in the cycle of length T is:

$$Q = \int_0^{t_1} R(t)dt + \int_{t_3}^{T} R(t)dt$$

$$= \int_0^{t_1} \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}dt + \int_{t_3}^{T} \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}dt \quad (12)$$

From equation (8) and using the initial condition *I(0) = 0*, we obtain the value of 'S' as:

$$S = e^{-\theta t_1}\int_0^{t_1}\left(\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}\right)e^{u\theta}du - \frac{k}{\theta}\left(1 - e^{-\theta t_1}\right) \quad (13)$$

When $t = t_3$, then equations (10) and (11) become:

$$I(t_3) = k(t_2 - t_3) \quad (14)$$

and

$$I(t_3) = k(T - t_3) - \int_{t_3}^{T} \frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}du \quad (15)$$

Equating the equations (14) and (15) and on simplification one can get:

$$t_2 = \frac{1}{k}\int_{t_3}^{T}\left[\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}\right]du + T = x(t_3)\ say \quad (16)$$

Let $K(t_1, t_2, t_3)$ be the total production cost per unit time. Since the total production cost is the sum of the set up cost, cost of the units, the inventory holding cost. Hence the total production cost per unit time become:

$$K(t_1, t_2, t_3) = \frac{A}{T} + \frac{CQ}{T} + \frac{h}{T}\left[\int_0^{t_1} I(t)dt + \int_{t_1}^{t_2} I(t)dt\right] + \frac{\pi}{T}\left[\int_{t_2}^{t_3} -I(t)dt + \int_{t_3}^{T} -I(t)dt\right] \quad (17)$$

Substituting the values of *I(t)* given in equations (8), (9), (10) and (11) and Q given in equation (12) in equation (17) one can obtain $K(t_1, t_2, t_3)$ as:

$$K(t_1,t_2,t_3)=\frac{A}{T}+\frac{C}{T}\left[\int_0^{t_1}\frac{p\alpha_1\beta_1t^{\beta_1-1}e^{-\alpha_1t^{\beta_1}}+(1-p)\alpha_2\beta_2t^{\beta_2-1}e^{-\alpha_2t^{\beta_2}}}{pe^{-\alpha_1t^{\beta_1}}+(1-p)e^{-\alpha_2t^{\beta_2}}}dt+\int_{t_3}^{T}\frac{p\alpha_1\beta_1t^{\beta_1-1}e^{-\alpha_1t^{\beta_1}}+(1-p)\alpha_2\beta_2t^{\beta_2-1}e^{-\alpha_2t^{\beta_2}}}{pe^{-\alpha_1t^{\beta_1}}+(1-p)e^{-\alpha_2t^{\beta_2}}}dt\right]$$
$$+\frac{h}{T}\left[\int_0^{t_1}\left[Se^{\theta(t_1-t)}-e^{-t\theta}\int_t^{t_1}\left(\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}-k\right)e^{u\theta}du\right]dt+\int_{t_1}^{t_2}\left[Se^{\theta(t_1-t)}-ke^{-t\theta}\int_{t_1}^{t}e^{u\theta}\,du\right]dt\right]$$
$$-\frac{\pi}{T}\left[k\int_{t_2}^{t_3}(t_2-t)dt+\int_{t_3}^{T}\left[\left(\int_t^T\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}du\right)dt\right]+k\int_{t_3}^{T}(T-t)\right] \quad (18)$$

Substituting the value of S given in equation (13) in the total production cost equation (18), we obtain:

$$K(t_1,t_2,t_3)=\frac{A}{T}+\frac{C}{T}\left[\int_0^{t_1}\frac{p\alpha_1\beta_1t^{\beta_1-1}e^{-\alpha_1t^{\beta_1}}+(1-p)\alpha_2\beta_2t^{\beta_2-1}e^{-\alpha_2t^{\beta_2}}}{pe^{-\alpha_1t^{\beta_1}}+(1-p)e^{-\alpha_2t^{\beta_2}}}dt+\int_{t_3}^{T}\frac{p\alpha_1\beta_1t^{\beta_1-1}e^{-\alpha_1t^{\beta_1}}+(1-p)\alpha_2\beta_2t^{\beta_2-1}e^{-\alpha_2t^{\beta_2}}}{pe^{-\alpha_1t^{\beta_1}}+(1-p)e^{-\alpha_2t^{\beta_2}}}dt\right]$$
$$+\frac{h}{T}\left[\frac{1-e^{-\theta t_2}}{\theta}\int_0^{t_1}\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}e^{u\theta}du+\frac{k}{\theta^2}\left(1-e^{-t_2\theta}\right)-\frac{k}{\theta}t_2-\int_0^{t_1}e^{-\theta t}\left(\int_t^{t_1}\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}e^{u\theta}du\right)dt\right]$$
$$-\frac{\pi}{T}\left[\frac{k}{2}[(T-t_3)^2-(t_2-t_3)^2]+\int_{t_3}^{T}\left(\int_t^T\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}du\right)dt\right] \quad (19)$$

Substituting the value of $'t_2'$ given in equation (16) in the total production cost equation (19), we obtain:

$$K(t_1,t_2,t_3)=\frac{A}{T}+\frac{C}{T}\left[\int_0^{t_1}\frac{p\alpha_1\beta_1t^{\beta_1-1}e^{-\alpha_1t^{\beta_1}}+(1-p)\alpha_2\beta_2t^{\beta_2-1}e^{-\alpha_2t^{\beta_2}}}{pe^{-\alpha_1t^{\beta_1}}+(1-p)e^{-\alpha_2t^{\beta_2}}}dt+\int_{t_3}^{T}\frac{p\alpha_1\beta_1t^{\beta_1-1}e^{-\alpha_1t^{\beta_1}}+(1-p)\alpha_2\beta_2t^{\beta_2-1}e^{-\alpha_2t^{\beta_2}}}{pe^{-\alpha_1t^{\beta_1}}+(1-p)e^{-\alpha_2t^{\beta_2}}}dt\right]$$
$$+\frac{h}{T}\left[\frac{1-e^{-\theta x(t_3)}}{\theta}\int_0^{t_1}\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}e^{u\theta}\,du+\frac{k}{\theta^2}\left(1-e^{-\theta x(t_3)}\right)\right.$$

$$-\frac{k}{\theta}x(t_3)-\int_0^{t_1}e^{-\theta t}\left(\int_t^{t_1}\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}e^{u\theta}du\right)dt\Bigg]$$

$$-\frac{\pi}{T}\Bigg[\frac{k}{2}[(T-t_3)^2-(x(t_3)-t_3)^2]$$

$$+\int_{t_3}^{T}\left(\int_t^{T}\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}du\right)dt\Bigg] \quad (20)$$

IV. Optimal Production Schedules of the Model

In this section we obtain the optimal policies of the system under study. To find the optimal values of t_1 and t_3, we obtain the first order partial derivatives of $K(t_1,t_3)$ given in equation with respect to t_1 and t_3 and equate them to zero. The condition for minimization of $K(t_1,t_3)$ is

Where D is the Hessian matrix

$$D=\begin{vmatrix}\frac{\partial^2K(t_1,t_3)}{\partial {t_1}^2} & \frac{\partial^2K(t_1,t_3)}{\partial t_1\partial t_3}\\ \frac{\partial^2K(t_1,t_3)}{\partial t_1\partial t_3} & \frac{\partial^2K(t_1,t_3)}{\partial {t_3}^2}\end{vmatrix}>0$$

Differentiating $K(t_1,t_3)$ given in equation (20) with respect to t_1 and equating to zero, we get

$$\left\{\frac{C}{T}\left[\frac{p\alpha_1\beta_1{t_1}^{\beta_1-1}e^{-\alpha_1{t_1}^{\beta_1}}+(1-p)\alpha_2\beta_2{t_1}^{\beta_2-1}e^{-\alpha_2{t_1}^{\beta_2}}}{pe^{-\alpha_1{t_1}^{\beta_1}}+(1-p)e^{-\alpha_2{t_1}^{\beta_2}}}\right]\right.$$

$$\left.+\frac{h}{T}\left[\frac{1-e^{-x(t_3)\theta}}{\theta}\left[\frac{p\alpha_1\beta_1{t_1}^{\beta_1-1}e^{-\alpha_1{t_1}^{\beta_1}}+(1-p)\alpha_2\beta_2{t_1}^{\beta_2-1}e^{-\alpha_2{t_1}^{\beta_2}}}{pe^{-\alpha_1{t_1}^{\beta_1}}+(1-p)e^{-\alpha_2{t_1}^{\beta_2}}}\right]e^{t_1\theta}\right]\right\}=0 \quad (21)$$

Differentiating $K(t_1,t_3)$ given in equation (20) with respect to t_3 and equating to zero, we get

$$\left\{-\frac{C}{T}\left[\frac{p\alpha_1\beta_1{t_3}^{\beta_1-1}e^{-\alpha_1{t_3}^{\beta_1}}+(1-p)\alpha_2\beta_2{t_3}^{\beta_2-1}e^{-\alpha_2{t_3}^{\beta_2}}}{pe^{-\alpha_1{t_3}^{\beta_1}}+(1-p)e^{-\alpha_2{t_3}^{\beta_2}}}\right]\right.$$

$$+\frac{h}{T}\left[\frac{1}{\theta}\left[\frac{p\alpha_1\beta_1{t_3}^{\beta_1-1}e^{-\alpha_1{t_3}^{\beta_1}}+(1-p)\alpha_2\beta_2{t_3}^{\beta_2-1}e^{-\alpha_2{t_3}^{\beta_2}}}{pe^{-\alpha_1{t_3}^{\beta_1}}+(1-p)e^{-\alpha_2{t_3}^{\beta_2}}}\right]\left(1-e^{-\theta x(t_3)}\right)\right.$$

$$-\frac{e^{-\theta x(t_3)}}{k}\left[\frac{p\alpha_1\beta_1{t_3}^{\beta_1-1}e^{-\alpha_1{t_3}^{\beta_1}}+(1-p)\alpha_2\beta_2{t_3}^{\beta_2-1}e^{-\alpha_2{t_3}^{\beta_2}}}{pe^{-\alpha_1{t_3}^{\beta_1}}+(1-p)e^{-\alpha_2{t_3}^{\beta_2}}}\right]\int_0^{t_1}\left[\frac{p\alpha_1\beta_1u^{\beta_1-1}e^{-\alpha_1u^{\beta_1}}+(1-p)\alpha_2\beta_2u^{\beta_2-1}e^{-\alpha_2u^{\beta_2}}}{pe^{-\alpha_1u^{\beta_1}}+(1-p)e^{-\alpha_2u^{\beta_2}}}\right]e^{u\theta}\,du\Bigg]$$

$$-\frac{\pi}{T}\left[k(t_3-T)+(x(t_3)-t_3)\left[\left[\frac{p\alpha_1\beta_1{t_3}^{\beta_1-1}e^{-\alpha_1{t_3}^{\beta_1}}+(1-p)\alpha_2\beta_2{t_3}^{\beta_2-1}e^{-\alpha_2{t_3}^{\beta_2}}}{pe^{-\alpha_1{t_3}^{\beta_1}}+(1-p)e^{-\alpha_2{t_3}^{\beta_2}}}\right]+k\right]\right.$$

$$-\int_{t_3}^{T}\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}}+(1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}}+(1-p)e^{-\alpha_2 u^{\beta_2}}}du\Bigg]\Bigg\}=0 \qquad (22)$$

Solving the equations (21) and (22) simultaneously, we obtain the optimal time at which production is stopped t_1^* of t_1 and the optimal time t_3^* of t_3 at which the production is restarted after accumulation of backorders.

The optimum production quantity Q^* of Q in the cycle of length T is obtained by substituting the optimal values of t_1^*, t_3^* in equation (12).

V. Numerical Illustration

In this section we discuss the solution procedure of the model through a numerical illustration by obtaining the production uptime, production downtime, optimum production quantity and the total production cost of an inventory system. Here, it is assumed that the production is of deteriorating nature and shortages are allowed and fully backlogged. For demonstrating the solution procedure of the model the parameters are considered as A = Rs.300/-, C = Rs.10\-, h = Re.0.2\-, π = Rs.3.3\-, T = 12 months. For the assigned values of production parameters (α_1, α_2, β_1, β_2, p) = (11, 15, 0.55, 2, 0.5), deterioration parameter θ = 3, demand rate k = 3.3. The values of parameters above are varied further to observe the trend in optimal policies and the results are obtained are shown in Table 1. Substituting these values the optimal production quantity Q^*, the production uptime, production downtime and total production cost are computed and presented in Table 1.

From Table 1 it is observed that the deterioration parameter and production parameters have a tremendous influence on the optimal values of production times, production quantity and total production cost.

When the ordering cost 'A' increases from 300 to 345, the optimal production quantity Q^* decreases from 33.867 to 33.863, the optimal production down time t_1^* remains constant, the optimum production uptime t_3^* increases from 3.685 to 3,686, the total production cost per unit time K^* increases from 80.793 to 84.529. As the cost parameter 'C' increases from 10 to 11.5, the optimal production quantity Q^* increases from 33.867 to 33.872, the optimal production down time t_1^* and optimal production uptime t_3^* remains constant, the total production cost per unit time K^* increases from 80.793 to 82.451. As the holding cost 'h' increases from 0.2 to 0.23, the optimal production quantity Q^*, the optimal production down time t_1^*, the optimal production uptime t_3^* remains constant, the total production cost per unit time K^* decreases from 80.793 to 80.755. As the shortage cost 'π' increases from 3.3 to 3.795, the optimal production quantity Q^* increases from 33.867 to 33.966, the optimal production down time t_1^* remains constant, the optimal production uptime t_3^* decreases from 3.685 to 3.655, the total production cost per unit time K^* increases from 80.793 to 87.753.

As the production parameter 'α_1' varies from 11 to 12.65, the optimal production quantity Q^* increases from 33.867 to 39.086, the optimal production down time t_1^* increases from 1.274 to 1.277, the optimal production uptime t_3^* decreases from 3.685 to 3.628, the total production cost per unit time K^* increases from 80.793 to 93.146.As the production parameter 'α_2' varies from 15 to 17.25, the optimal production quantity Q^*, the optimal production down time t_1^*, the optimal production uptime t_3^*, the total production cost per unit time K^* remains constant.

Table 1: *Numerical Illustration*

A	C	h	π	T	α_1	α_2	β_1	β_2	θ	k	p	t_1^*	t_3^*	Q^*	K^*
300	10	0.2	3.3	12	11	15	0.55	2	3	3.3	0.5	1.274	3.685	33.867	80.793
315												1.274	3.685	33.866	82.039
330												1.274	3.686	33.865	83.284
345												1.274	3.686	33.863	84.529
	10.5											1.274	3.685	33.869	81.346
	11											1.274	3.685	33.87	81.898
	11.5											1.274	3.685	33.872	82.451
		0.21										1.274	3.685	33.867	80.781
		0.22										1.274	3.685	33.867	80.768
		0.23										1.274	3.685	33.867	80.755
			3.465									1.274	3.675	33.9	83.1
			3.63									1.274	3.665	33.933	85.42
			3.795									1.274	3.655	33.966	87.753
					11.55							1.275	3.666	35.599	84.778
					12.1							1.276	3.647	37.338	88.895
					12.65							1.277	3.628	39.086	93.146
						15.75						1.274	3.685	33.867	80.793
						16.5						1.274	3.685	33.867	80.793
						17.25						1.274	3.685	33.867	80.793
							0.578					1.275	3.648	36.366	88.423
							0.605					1.276	3.609	38.996	97.002
							0.633					1.277	3.565	41.973	107.394
								2.1				1.274	3.685	33.867	80.793
								2.2				1.274	3.685	33.867	80.793
								2.3				1.274	3.685	33.867	80.793
									3.15			1.274	3.685	33.867	80.805
									3.3			1.274	3.685	33.867	80.816
									3.45			1.274	3.685	33.867	80.826
										3.465		1.274	3.689	33.853	79.886
										3.63		1.274	3.693	33.841	79.062
										3.795		1.274	3.696	33.83	78.31
											0.525	1.274	3.685	33.818	80.753
											0.55	1.274	3.685	33.772	80.714
											0.575	1.274	3.685	33.728	80.677

As the production parameter 'β_1' varies from 0.55 to 0.633, the optimal production quantity Q^* increases from 33.867 to 41.973, the optimal production down time $t_1{}^*$ increases from 1.274 to 1.277, the optimal production uptime $t_3{}^*$ decreases from 3.685 to 3.565, the total production cost per unit time K^* increases from 80.793 to 107.394. As the production parameter 'β_2' varies from 2 to 2.3, the optimal production quantity Q*, the optimal production down time $t_1{}^*$, the optimal production uptime $t_3{}^*$ and the total production cost per unit time K^* remains constant. As the production parameter 'p' varies from 0.5 to 0.575, the optimal production quantity Q^* decreases from 33.867 to 33.728, the optimal production down time $t_1{}^*$ and the optimal production uptime $t_3{}^*$ remains constant, the total production cost per unit time K^* decreases from 80.793 to 80.677.

As the deterioration parameter 'θ' varies from 3 to 3.45, the optimal production quantity Q^*, the optimal production down time $t_1{}^*$ and the optimal production uptime $t_3{}^*$ remains constant, the total production cost per unit time K^* increases from 80.793 to 80.826.

As the demand rate parameter 'k' increases from 3.3 to 3.795 the optimal production quantity Q^* decreases from 33.867 to 33.83, the optimal production down time $t_1{}^*$ remains constant, the optimal production uptime $t_3{}^*$ increases from 3.685 to 3.696, the total production cost per unit time K^* decreases from 80.793 to 78.31.

VI. Sensitivity Analysis of the Model

Sensitivity analysis is carried to explore the effect of changes in model parameters and costs on the optimal policies, by varying each parameter (-15%, -10%, -5%, 0%, 5%, 10%, 15%) at a time for the model under study. The results are presented in Table 2. The relationships between the parameters and the optimal values of the production schedule are shown in Figure 2.

Table 2: *Sensitivity Analysis of the Model - With Shortages*

Variation Parameters	Optimal Policies	-15%	-10%	-5%	0%	5%	10%	15%
A	$t_1{}^*$	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	$t_3{}^*$	3.684	3.684	3.685	3.685	3.685	3.686	3.686
	Q^*	33.871	33.869	33.868	33.867	33.866	33.865	33.863
	K^*	77.058	78.303	79.548	80.793	82.039	83.284	84.529
C	$t_1{}^*$	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	$t_3{}^*$	3.684	3.685	3.685	3.685	3.685	3.685	3.685
	Q^*	33.862	33.864	33.866	33.867	33.869	33.87	33.872
	K^*	79.138	79.689	80.241	80.793	81.346	81.898	82.451
h	$t_1{}^*$	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	$t_3{}^*$	3.685	3.685	3.685	3.685	3.685	3.685	3.685
	Q^*	33.867	33.867	33.867	33.867	33.867	33.867	33.867
	K^*	80.832	80.819	80.806	80.793	80.781	80.768	80.755
π	$t_1{}^*$	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	$t_3{}^*$	3.715	3.705	3.695	3.685	3.675	3.665	3.655
	Q^*	33.766	33.8	33.834	33.867	33.9	33.933	33.966
	K^*	73.95	76.218	78.499	80.793	83.1	85.42	87.753
α_1	$t_1{}^*$	1.271	1.272	1.273	1.274	1.275	1.276	1.277
	$t_3{}^*$	3.74	3.722	3.704	3.685	3.666	3.647	3.628
	Q^*	28.718	30.427	32.143	33.867	35.599	37.338	39.086
	K^*	69.61	73.212	76.939	80.793	84.778	88.895	93.146

Variation Parameters	Optimal Policies	-15%	-10%	-5%	0%	5%	10%	15%
α_2	t_1^*	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	t_3^*	3.685	3.685	3.685	3.685	3.685	3.685	3.685
	Q^*	33.867	33.867	33.867	33.867	33.867	33.867	33.867
	K^*	80.793	80.793	80.793	80.793	80.793	80.793	80.793
β_1	t_1^*	1.27	1.272	1.273	1.274	1.275	1.276	1.277
	t_3^*	3.776	3.748	3.719	3.685	3.648	3.609	3.565
	Q^*	27.639	29.56	31.58	33.867	36.366	38.996	41.973
	K^*	63.978	68.832	74.257	80.793	88.423	97.002	107.394
β_2	t_1^*	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	t_3^*	3.685	3.685	3.685	3.685	3.685	3.685	3.685
	Q^*	33.867	33.867	33.867	33.867	33.867	33.867	33.867
	K^*	80.793	80.793	80.793	80.793	80.793	80.793	80.793
θ	t_1^*	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	t_3^*	3.685	3.685	3.685	3.685	3.685	3.685	3.685
	Q^*	33.867	33.867	33.867	33.867	33.867	33.867	33.867
	K^*	80.749	80.766	80.78	80.793	80.805	80.816	80.826
k	t_1^*	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	t_3^*	3.67	3.676	3.68	3.685	3.689	3.693	3.696
	Q^*	33.917	33.899	33.882	33.867	33.853	33.841	33.83
	K^*	84.168	82.916	81.798	80.793	79.886	79.062	78.31
p	t_1^*	1.274	1.274	1.274	1.274	1.274	1.274	1.274
	t_3^*	3.685	3.685	3.685	3.685	3.685	3.685	3.685
	Q^*	34.029	33.972	33.918	33.867	33.818	33.772	33.728
	K^*	80.929	80.881	80.836	80.793	80.753	80.714	80.677

VII. Observations

The major observations drawn from the numerical study are:

- t_1^* and t_3^* are less sensitive, Q^* is slightly sensitive and K^* is moderately sensitive to changes of ordering cost 'A'.
- t_1^* and t_3^* are less sensitive, Q^* is slightly sensitive and K^* is moderately sensitive to changes of cost per unit 'C'.
- t_1^*, t_3^* and Q^* are less sensitive, K^* is slightly sensitive to changes of holding cost 'h'.
- t_1^* is less sensitive, t_3^* and Q^* are slightly sensitive and K^* is highly sensitive to change in parameter 'π'.
- t_1^* and t_3^* are slightly sensitive, Q^* and K^* are highly sensitive to change in the production parameter 'α_1'.
- t_1^*, t_3^*, Q^* and K^* are less sensitive to change in the production parameter 'α_2'.
- t_1^* and t_3^* are slightly sensitive, Q^* and K^* are highly sensitive to change in the production parameter 'β_1'.
- t_1^*, t_3^*, Q^* and K^* are less sensitive to change in the production parameter 'β_2'.
- t_1^* and t_3^* are less sensitive, Q^* and K^* are slightly sensitive to change in the production parameter 'p'.
- t_1^*, t_3^* and Q^* are less sensitive, K^* is slightly sensitive to change in the deterioration parameter 'θ'.
- t_1^* is less sensitive, t_3^* and Q^* are slightly sensitive and K^* is highly sensitive to change in the demand parameter 'k'.

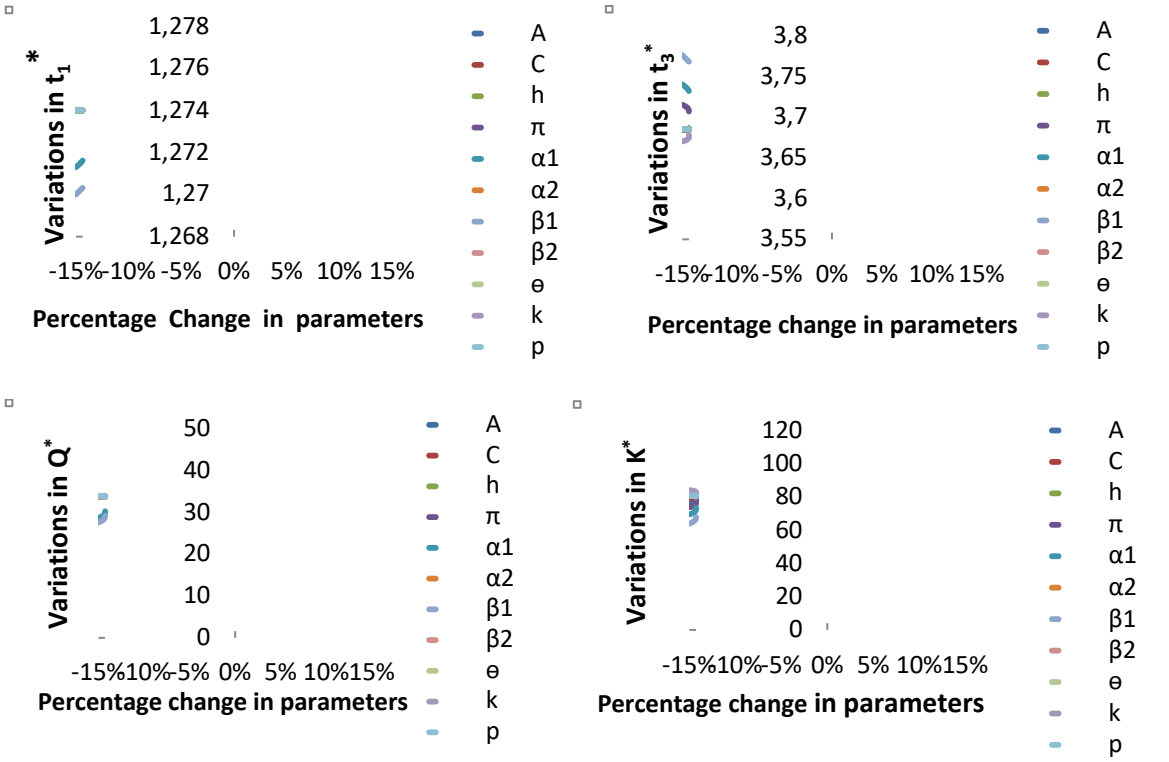


Figure 2: *Relationship between parameters and optimal values with shortages*

VIII. EPQ Model without Shortages

In this section the inventory model for deteriorating items without shortages is developed and analyzed. Here, it is assumed that shortages are not allowed and the stock level is zero at time $t=0$. The stock level increases during the period $(0, t_1)$ due to excess production after fulfilling the demand and deterioration. The production stops at time t_1 when the stock level reaches S. The inventory decreases gradually due to demand and deterioration in the interval (t_1, T) . At time T the inventory reaches zero. The schematic diagram representing the instantaneous state of inventory is given in Figure 3.

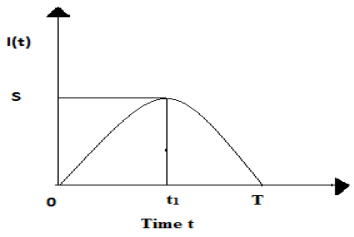


Figure 3: *Schematic diagram representing the inventory level*

Let $I(t)$ be the inventory level of the system at time 't' $(0 \le t \le T)$. Then the differential equations governing the instantaneous state of $I(t)$ over the cycle of length T are:

$$\frac{d}{dt}I(t) + h(t)I(t) = \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}} - k;\ 0 \le t \le t_1 \qquad (23)$$

$$\frac{d}{dt}I(t) + h(t)I(t) = -k;\ t_1 \le t \le T \tag{24}$$

Where, *h(t)* is as given in equation (3), with the initial conditions *I(0) = 0*, $I(t_1) = S$ and *I(T) =0*. Substituting *h(t)* given in equation (3) in equations (23) and (24) and solving the differential equations, the on hand inventory at time 't' is obtained as:

$$I(t) = Se^{\theta(t_1-t)} - e^{-t\theta}\int_t^{t_1}\left[\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}} - k\right]e^{u\theta}du\,;0 \le t \le t_1 \tag{25}$$

$$I(t) = Se^{\theta(t_1-t)} - ke^{-t\theta}\int_{t_1}^{t} e^{u\theta}\,du; t_1 \le t \le T \tag{26}$$

Production quantity Q in the cycle of length T is

$$Q = \int_0^{t_1} R(t)dt$$

$$= \int_0^{t_1} \frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}dt \tag{27}$$

From equation (25) and using the initial conditions *I(0) = 0*, we obtain the value of 'S' as

$$S = e^{-\theta t_1}\int_0^{t_1}\left(\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}\right)e^{u\theta}du - \frac{k}{\theta}\left(1 - e^{-\theta t_1}\right) \tag{28}$$

Let $K(t_1)$ be the total production cost per unit time. Since the total production cost is the sum of the set up cost, cost of the units, the inventory holding cost. Therefore the total production cost per unit time becomes

$$K(t_1) = \frac{A}{T} + \frac{CQ}{T} + \frac{h}{T}\left[\int_0^{t_1} I(t)dt + \int_{t_1}^{T} I(t)dt\right] \tag{29}$$

Substituting the values of *I(t)* and Q from equations ,(25), (26) and (27) in equation (29), we obtain $K(t_1)$ as

$$K(t_1) = \frac{A}{T} + \frac{C}{T}\int_0^{t_1}\left[\frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}\right]dt$$

$$+\frac{h}{T}\left[\int_0^{t_1}\left[Se^{\theta(t_1-t)} - e^{-t\theta}\int_t^{t_1}\left[\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}du - k\right]e^{u\theta}du\right]dt\right.$$

$$\left.+\int_{t_1}^{T}\left[Se^{\theta(t_1-t)} - ke^{-t\theta}\int_{t_1}^{t} e^{u\theta}\,du\right]dt\right] \tag{30}$$

Substituting the value of S given in equation (28) in the total cost equation (30), we obtain

$$K(t_1) = \frac{A}{T} + \frac{C}{T}\int_0^{t_1}\left[\frac{p\alpha_1\beta_1 t^{\beta_1-1}e^{-\alpha_1 t^{\beta_1}} + (1-p)\alpha_2\beta_2 t^{\beta_2-1}e^{-\alpha_2 t^{\beta_2}}}{pe^{-\alpha_1 t^{\beta_1}} + (1-p)e^{-\alpha_2 t^{\beta_2}}}\right]dt$$

$$+\frac{h}{T}\left[\left(1-e^{-\theta T}\right)\left[\frac{1}{\theta}\int_0^{t_1}\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}e^{u\theta}du + \frac{k}{\theta^2}\right]\right.$$

$$\left.-\frac{k}{\theta}T - \int_0^{t_1} e^{-t\theta}\left(\int_t^{t_1}\frac{p\alpha_1\beta_1 u^{\beta_1-1}e^{-\alpha_1 u^{\beta_1}} + (1-p)\alpha_2\beta_2 u^{\beta_2-1}e^{-\alpha_2 u^{\beta_2}}}{pe^{-\alpha_1 u^{\beta_1}} + (1-p)e^{-\alpha_2 u^{\beta_2}}}e^{u\theta}du\right)dt\right] \quad (31)$$

IX. Optimal Production Schedules of the Model

In this section we obtain the optimal policies of the inventory system under study. To find the optimal values of t_1, we equate the first order partial derivatives of $K(t_1)$ with respect to t_1 equate them to zero. The condition for minimum of $K(t_1)$ is

$$\frac{\partial^2 K(t_1)}{\partial {t_1}^2} > 0$$

Differentiating $K(t_1)$ with respect to t_1 and equating to zero, we get

$$\left\{\frac{C}{T}\left[\frac{p\alpha_1\beta_1 {t_1}^{\beta_1-1}e^{-\alpha_1 {t_1}^{\beta_1}} + (1-p)\alpha_2\beta_2 {t_1}^{\beta_2-1}e^{-\alpha_2 {t_1}^{\beta_2}}}{pe^{-\alpha_1 {t_1}^{\beta_1}} + (1-p)e^{-\alpha_2 {t_1}^{\beta_2}}}\right]\right.$$

$$\left.+\frac{h}{T}\left[\frac{\left(1-e^{-\theta T}\right)e^{t_1\theta}}{\theta}\left[\frac{p\alpha_1\beta_1 {t_1}^{\beta_1-1}e^{-\alpha_1 {t_1}^{\beta_1}} + (1-p)\alpha_2\beta_2 {t_1}^{\beta_2-1}e^{-\alpha_2 {t_1}^{\beta_2}}}{pe^{-\alpha_1 {t_1}^{\beta_1}} + (1-p)e^{-\alpha_2 {t_1}^{\beta_2}}}\right]\right]\right\} = 0 \quad (32)$$

Solving the equation (32), we obtain the optimal time t_1^* of $t_1$at which the production is to be stopped.

The optimal production quantity Q^* of Q in the cycle of length T is obtained by substituting the optimal values of t_1 in equation (27).

X. Numerical Illustration

In this section we discuss the solution procedure of the model through a numerical illustration by obtaining the production time, optimum production quantity and the total production cost of an inventory system. For demonstrating the solution procedure of the model the parameters are considered as A = Rs.310\-, C = Rs.15\-, h = Re.0.2\-, (α_1, α_2, β_1, β_2, p) = (11, 14, 0.55, 3, 0.5), θ= 3, k=3.3 and T=12 months. The values of parameters above are varied further to observe the trend in optimal policies and the results are obtained are shown in Table 3. Substituting these values the optimal production quantity Q^*, the production time and total production cost are computed and presented in Table 3

From Table 3 it is observed that the deterioration parameters and production parameters have a tremendous influence on the optimal values of the model.

Table 3: *Numerical Illustration*

A	C	h	T	α_1	α_2	β_1	β_2	θ	k	p	t_1^*	Q^*	K^*
310	15	0.2	12	11	14	0.55	3	3	3.3	0.5	5.495	28.771	61.871
325.5											5.495	28.771	63.163
341											5.495	28.771	64.454
356.5											5.495	28.771	65.746
	15.75										5.496	28.775	63.698
	16.5										5.497	28.778	65.477
	17.25										5.499	28.782	67.281
		0.21									5.495	28.772	61.875
		0.22									5.495	28.772	61.879
		0.23									5.495	28.772	61.883
				11.55							5.5	30.217	63.686
				12.1							5.501	31.598	65.402
				12.65							5.503	33.01	67.143
					14.7						5.493	28.767	61.859
					15.4						5.492	28.763	61.825
					16.1						5.491	28.761	61.788
						0.578					5.499	30.157	63.607
						0.605					5.504	31.56	65.366
						0.633					5.509	33.088	67.281
							3.15				5.496	28.773	61.89
							3.3				5.496	28.775	61.907
							3.45				5.497	28.777	61.923
								3.15			5.495	28.772	61.886
								3.3			5.495	28.773	61.901
								3.45			5.496	28.773	61.92
									3.465		5.495	28.771	61.86
									3.63		5.495	28.771	61.85
									3.795		5.495	28.771	61.839
										0.525	5.495	28.722	61.806
										0.55	5.494	28.675	61.743
										0.575	5.494	28.629	61.681

When the ordering cost 'A' increases from 310 to 356.5, the optimal production quantity Q^* and the optimal production down time t_1^* remains constant, the total production cost per unit time K^* increases from 61.871 to 65.746. As the cost parameter 'C' increases from 15 to 17.25, the optimal production quantity Q^* increases from 28.771 to 28.782, the optimal production down time t_1 increases from 5.495 to 5.499, the total production cost per unit time K^* increases from 61.871 to 67.281. As the inventory holding cost 'h' increases from 0.2 to 0.23, the optimal production quantity Q^* increases from 28.771 to 28.772, the optimal production down time t_1^* remains constant, the total production cost per unit time K^* increases from 61.871 to 61.883.

As the production parameter 'α_1' varies from 11 to 12.65, the optimal production quantity Q^* increases from 28.771 to 33.01, the optimal production down time t_1^* increases from 5.495 to 5.503, the total production cost per unit time K^* increases from 61.871 to 67.143. As the production

parameter 'α_2' varies from 14 to 16.1, the optimal production quantity Q^* decreases from 28.771 to 28.761, the optimal production down time t_1^* decreases from 5.495 to 5.491, the total production cost per unit time K^* decreases from 61.871 to 61.788. As the production parameter 'β_1' varies from 0.55 to 0.633, the optimal production quantity Q^* increases from 28.771 to 33.088, the optimal production down time t_1^* increases from 5.495 to 5.509, the total production cost per unit time K^* increases from 61.871 to 67.281. As the production parameter 'β_2' varies from 3 to 3.45, the optimal production quantity Q^* increases from 28.771 to 28.773, the optimal production down time t_1^* increases from 5.495 to 5.497, the total production cost per unit time K^* increases from 61.871 to 61.923. As the production parameter 'p' varies from 0.5 to 0.575 the total production quantity Q^* decreases from 28.771 to 28.629, the optimal production down time t_1^* decreases from 5.495 to 5.494, the total production cost per unit time K^* decreases from 61.871 to 61.681.

As the deterioration parameter 'θ' varies from 3 to 3.45, the optimal production quantity Q^* increases from 28.771 to 28.773, the optimal production down time t_1^* increases from 5.495 to 5.496, the total production cost per unit time K^* increases from 61.871 to 61.92.

As the demand parameter 'k' varies from 3.3 to 3.795, the total production quantity Q^* remains constant, the optimal production down time t_1^* remains constant, the total production cost per unit time K^* decreases from 61.871 to 61.839.

XI. Sensitivity Analysis of the Model

The sensitivity analysis is carried to explore the effect of changes in model parameters and costs on the optimal policies, by varying each parameter (-15%, -10%, -5%, 0%, 5%, 10%, 15%) at a time for the model under study. The results are presented in Table 4. The relationship between the parameters and the optimal values of the production schedule is shown in Figure 4.

Table 4: *Sensitivity analysis of the model - Without Shortages*

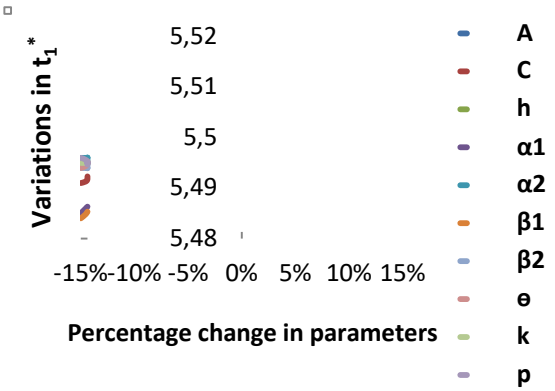
Variation Parameters	Optimal Policies	-15%	-10%	-5%	0%	5%	10%	15%
A	t_1^*	5.495	5.495	5.495	5.495	5.495	5.495	5.495
	Q^*	28.772	28.772	28.771	28.771	28.771	28.771	28.771
	K^*	57.996	59.288	60.579	61.871	63.163	64.454	65.746
C	t_1^*	5.491	5.492	5.494	5.495	5.496	5.497	5.499
	Q^*	28.761	28.765	28.768	28.771	28.775	28.778	28.782
	K^*	56.465	58.266	60.068	61.871	63.698	65.477	67.281
h	t_1^*	5.495	5.495	5.495	5.495	5.495	5.495	5.495
	Q^*	28.771	28.771	28.771	28.771	28.772	28.772	28.772
	K^*	61.859	61.863	61.867	61.871	61.875	61.879	61.883
α_1	t_1^*	5.485	5.489	5.491	5.495	5.5	5.501	5.503
	Q^*	24.537	26.203	27.358	28.771	30.217	31.598	33.01
	K^*	56.463	58.599	60.076	61.871	63.686	65.402	67.143
α_2	t_1^*	5.496	5.496	5.496	5.495	5.493	5.492	5.491
	Q^*	28.775	28.775	28.774	28.771	28.767	28.763	28.761
	K^*	61.901	61.885	61.880	61.871	61.859	61.825	61.788

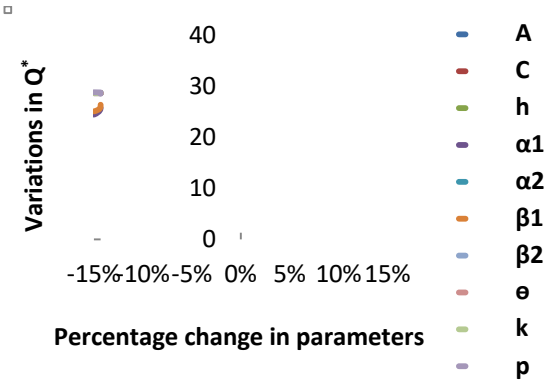
β_1	t_1^*	5.484	5.487	5.491	5.495	5.499	5.504	5.509
	Q^*	25.171	26.286	27.453	28.771	30.157	31.56	33.088
	K^*	57.359	58.756	60.219	61.871	63.607	65.366	67.281
β_2	t_1^*	5.494	5.494	5.494	5.495	5.496	5.496	5.497
	Q^*	28.768	28.769	28.77	28.771	28.773	28.775	28.777
	K^*	61.811	61.831	61.851	61.871	61.89	61.907	61.923
θ	t_1^*	5.494	5.495	5.495	5.495	5.495	5.495	5.496
	Q^*	28.77	28.771	28.771	28.771	28.772	28.773	28.773
	K^*	61.832	61.845	61.857	61.871	61.886	61.901	61.92
k	t_1^*	5.495	5.495	5.495	5.495	5.495	5.495	5.495
	Q^*	28.771	28.771	28.771	28.771	28.771	28.771	28.771
	K^*	61.903	61.892	61.882	61.871	61.86	61.85	61.839
p	t_1^*	5.496	5.495	5.495	5.495	5.495	5.494	5.494
	Q^*	28.936	28.878	28.823	28.771	28.722	28.675	28.629
	K^*	62.082	62.009	61.939	61.871	61.806	61.743	61.681

XII. Observations

The major observations drawn from the numerical study are:

- t_1^* is less sensitive, Q^* is slightly sensitive and K^* is moderately sensitive to the changes in ordering cost 'A'.
- t_1^* and Q^* are slightly sensitive and K^* is moderately sensitive to the changes in cost per unit 'C'.
- t_1^* is less sensitive, Q^* and K^* are slightly sensitive to the changes in holding cost 'h'.
- t_1^* is slightly sensitive, Q^* and K^* are highly sensitive to the change in the production parameter 'α_1'.
- t_1^*, Q^* and K^* are slightly sensitive to the change in the production parameter 'α_2'.
- t_1^* is slightly sensitive, Q^* and K^* are moderately sensitive to the change in the production parameter 'β_1'.
- t_1^*, Q^* and K^* are slightly sensitive to the change in the production parameter 'β_2'.
- t_1^*, Q^* and K^* are slightly sensitive to the change in the production parameter 'p'.
- t_1^*, Q^* and K^* are slightly sensitive to the change in the deterioration parameter 'θ'.
- t_1^* and Q^* are less sensitive, K^* is slightly sensitive to change the demand parameter 'k'.





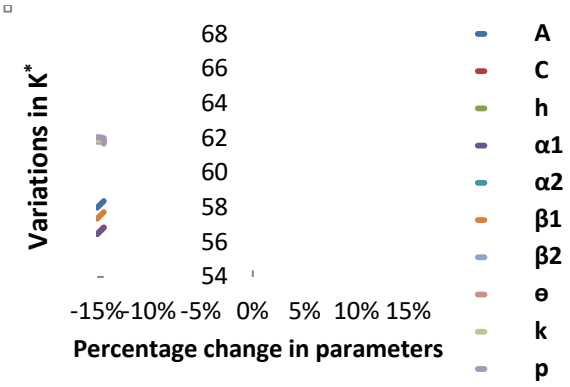


Figure 4: *Relationship between parameters and optimal values without shortages*

XIII. Conclusions

This paper introduces a new EPQ model with random production having mixture of two component Weibull production rate and exponential decay having constant demand. The mixture of two parameter Weibull distribution characterises the heterogeneous process more close to reality. By using the historical data we can estimate the replenishment and deterioration distribution parameters. The production manager can estimate the optimal production downtime and uptime with the distributional data of production and deterioration parameters. The Weibull rate of production can include increase/ decrease/constant rates for different values of parameters. Sensitivity analysis is used to understand the change in the parameters of Weibull rates of production and exponential deterioration. It is observed that random production and deterioration have significant influence on optimal values of the production schedule and production quantity. This model also includes some of the earlier models as particular cases. This model can be used to analyse production processes where the production is done in two different units/ machines and rate of deterioration is constant. It is possible to extend this model with other types of demand functions such as stock dependent demand, time and selling price dependent demand which will be taken up elsewhere. This paper is useful for analyzing optimal production schedules for the industries dealing with deteriorating items such as sea foods and edible oil. This model also includes some of the earlier EPQ models as particular cases for specific values of the parameters.

Funding

No funding was provided for the research.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

References

[1] Pentico, D. W. and Drake, M. J, "A survey of deterministic models for the EOQ and EPQ with partial backordering", European Journal of Operational Research, Vol. 214, Issue. 2, pp. 179-198, 2011.

[2] Ruxian, LI., Lan, H. and Mawhinney, R. J, "A review on deteriorating inventory study", Journal of Service Science Management, Vol. 3, No. 1, pp. 117-129, 2010.

[3] Goyal, S. K and Giri, B. C, "Recent trends in modeling of deteriorating inventory", European Journal of operational Research, Vol. 134, No.1, pp. 1-16, 2001.

[4] Raafat, F. "Survey of literature on continuously deteriorating inventory models", Journal of the Operational Research Society, Vol. 42, No. 1, pp. 27-37, 1991.

[5] Nahmias, S, "Perishable inventory theory: A review", OPSEARCH, Vol. 30, No. 4, pp. 680-708, 1982.

[6] Ghare, P. M and Schrader, G. F, "A model for exponentially decaying inventories", Journal of Industrial engineering, Vol. 14, pp. 238-2430, 1963.

[7] Shah, Y. and Jaiswal, M. C, "An order level inventory model for a system with a constant rate of deterioration", OPSEARCH, Vol. 14, pp. 174-184, 1977.

[8] Cohen, M. A, "Joint pricing and ordering for policy exponentially decaying inventories with known demand", Naval Research Logistics. Q, Vol. 24, pp. 257-268, 1977.

[9] Aggarwal, S. P, "A note on an order level inventory model for system with constant rate of deterioration", OPSEARCH, Vol. 15, No. 4, pp.184–187, 1978.

[10] Dave, U and Shah, Y.K, "A probabilistic inventory model for deteriorating items with time proportional demand", Journal of Operational Research Society, Vol. 32, pp. 137-142, 1982.

[11] Pal, M, 'An inventory model for deteriorating items when demand is random", Calcutta Statistical Association Bulletin, Vol. 39, pp. 201-207, 1990.

[12] Kalpakam, S and Sapna, K. P, "A lost sales (S-1, S) perishable inventory system with renewal demand", Naval Research Logistics, Vol. 43, pp. 129-142, 1996.

[13] Giri, B. C and Chaudhuri, K. S, "An economic production lot-size model with shortages and time dependent demand", IMA Journal of Management Mathematics, Vol. 10, No.3, pp. 203-211, 1999.

[14] Perumal, V. and Arivarignan, G, "A production model with two rates of productions and back orders", International Journal of Management system, Vol. 18, pp. 109-119, 2002.

[15] Pal, M. and Mandal, B, "An EOQ model for deteriorating inventory with alternating demand rates", Journal Of Applied Mathematics and Computing, Vol. 4, No.2, pp. 392-397, 1997.

[16] Sen,S. and Chakrabarthy, T, "An order- level inventory model with variable rate of deterioration and alternating replenishing rates considering shortages", OPSEARCH, Vol. 44(1), pp. 17-26, 2007.

[17] Lin, G. C and Gong, D. C, "On a production-inventory system of deteriorating items subject to random machine breakdowns with a fixed repair time", Mathematical and Computer Modeling, Vol. 43, Issue. 7-8, pp. 920-932, 2006.

[18] Maity, A. K., Maity., K., Mondal, S and Maiti, M, "A Chebyshev approximation for solving the optimal production inventory problem of deteriorating multi-item", Mathematical and Computer Modelling, Vol. 45, No. 1, pp. 149-161, 2007.

[19] Hu, F and Liu, D, "Optimal replenishment policy for the EPQ model with permissible delay in payments and allowable shortages", Applied Mathematical Modelling, Vol. 34 (10), pp. 3108-3117, 2010.

[20] Uma Maheswara Rao, S. V., Venkata Subbaiah, K. and Srinivasa Rao. K, "Production inventory models for deteriorating items with stock dependent demand and Weibull decay", IST Transaction of Mechanical Systems-Theory and Applications, Vol. 1, No. 1(2), pp. 13-23, 2010.

[21] Venkata Subbaiah, K., Uma Maheswara Rao, S.V. and Srinivasa Rao, K, "An inventory model for perishable items with alternating rate of production", International Journal of Advanced Operations Management, Vol. 3, No. 1, pp. 66-87, 2011.

[22] Essey, K. M and Srinivasa Rao, K, "EPQ models for deteriorating items with stock dependent demand having three parameter Weibull decay", International Journal of Operations Research, Vol.14, No.3, pp. 271-300, 2012.

[23] Ardak, P.S. and Borade, A.B, "An economic production quantity model with inventory dependent demand and deterioration", International journal of engineering and technology, Vol.9, No.2, pp. 955-962, 2017.

[24] Anindya Mandal, Brojeswar Pal and Kripasindhu Chaudhuri, "Unreliable EPQ model with variable demand under two-tier credit financing", Journal of Industrial and Production Engineering, Vol.37, No. 7, pp. 370–386, 2020.

[25] Sunit Kumar, Sushil Kumar and Rachna Kumari, "An EPQ model with two-level trade credit and multivariate demand incorporating the effect of system improvement and preservation technology", Malaya Journal of Matematik, Vol. 9, No. 1, pp. 438-448, 2021.

[26] Sridevi, "Inventory model for deteriorating items with Weibull rate of replenishment and selling price dependent demand", International Journal of Operational Research, Vol. 9(3), pp. 329-349, 2010.

[27] Srinivasa Rao, K., Nirupama Devi, K. and Sridevi, G, "Inventory model for deteriorating items with Weibull rate of production and demand as function of both selling price and time", Assam Statistical Review, Vol. 24, No.1, pp.57-78, 2010.

[28] Lakshmana Rao, A. and Srinivasa Rao, K, "Studies on inventory model for deteriorating items with Weibull replenishment and generalized Pareto decay having demand as function of on hand inventory", International Journal of Supply and Operations Management, Vol. 1, Issue. 4, pp. 407-426, 2015.

[29] Srinivasa Rao et al, "Inventory model for deteriorating items with Weibull rate of replenishment and selling price dependent demand", International Journal of Operational Research, Vol. 9(3), pp. 329-349, 2017.

[30] Madhulatha, D. "Economic production quantity model with generalized Pareto rate of production and Weibull decay having selling price dependent demand", Journal of Ultra scientist of physical sciences,Vol.29, No.11, pp. 485-500,2017.

[31] Punyavathi, B. "On an EPQ model with generalized pareto rate of replenishment and deterioration with constant demand", International Journal of Scientific and Engineering Research", Vol.11, Issue 1, pp. 1191-1208, 2020.

ESTIMATION, COMPARISON AND RANKING OF OPERATIONAL RELIABILITY INDICATORS OF OVERHEAD TRANSMISSION LINES OF ELECTRIC POWER SYSTEMS

Farhadzadeh E.M., Muradaliyev A.Z., Abdullayeva S.A.

•

Azerbaijan Scientific-Research and Design-Prospecting Institute of Energetic
e-mail: elmeht@rambler.ru

Abstract

The regular increase in relative number of units of equipment, devices and installations (further - objects) electric power systems, which service life exceeds normative value and the consequences connected with this fact, including, including unacceptable ones, demand acceptance of drastic measures on increase of efficiency of their work. The main efforts today aimed at improving the methods of recognition and control of their technical condition. In other words, the problems of increasing the reliability of work and the safety of service brought to the fore quite justifiably. In the article, it is propos to carry out monitoring of the technical condition of overhead lines with a rated voltage of 110 kV and above monthly on the basis operational reliability parameters. New methods and algorithms for their estimation, comparison and ranking presented. As the operational reliability parameters are multidimensional, the existing methods for comparing and ranking one-dimensional statistical estimates for them are unacceptable, as the neglect preconditions of these methods conducts to essential growth of risk of the erroneous decision. The proposed new methods based on the fiducial approach, imitating modeling and the theory of statistical hypothesis testing. The cumbersomeness and laboriousness of manual calculation of operational reliability parameters, the science intensity of calculation methods is compensated by the transition to automated systems that provide information and methodological support with information about the technical condition of overhead lines. The recommended methods are included in the group of risk-focused approaches of increase the efficiency of the electric power systems.

Keywords. Operational reliability, overhead lines, estimation, comparison, ranking, classification, fiducial approach, risk of the erroneous decision.

I. Introduction

Increase of the efficiency of the electric network enterprises (further – ENE) is one of the most important problems of electric power systems. Formed and intensively developing in the energy economy a new scientific direction "Development of the asset management system for electric network companies." In our opinion, this direction is the most fully reflected in the concept [1], where it is noted: “the organization of the activities of the electric network companies in a competitive environment brings to the fore the economic criteria with *unconditional assurance of reliability”*.

In [2], at the same time, it is rightly noted: "the activity on the transmission of the electric power has monopoly character. Because of this, there is practically no motivation to increase the efficiency of work.".

And, at last, it is well known, that today the service life of more than 50% of ENE objects exceeds the normative value and there is no hope of decrease this value, or at least not exceeding it [3]. The relevance of the problem of reliability of objects, the service life of which exceeds the normative value is clearly confirmed by the materials of international conferences [4, 5]. It is necessary to consider:

- in opinion of leading experts, the efficiency of work today is determined not only by efficiency, but also reliability and safety [6];
- the consequences of failures in EPS are increasingly unacceptable and violate energy security;
- the social importance of increase in electricity tariffs determines the systematic support by the state of the electric power companies, with an increase in their operating costs;
- there is no opportunity, and necessity of mass replacement and modernization of objects, which service life, exceeds normative value.

And if all this is taken into account, today the major problem EPS is *increase of reliability and safety of objects, which service life exceeds normative value.* The methodical approach of authors to the decision of this problem based on some assumptions, allowing using the approaches accepted in others, completely, different systems. First of all, it is offered to agree with a known postulate according to which «*the person creates objects similar to himself*». This opinion allows using the approaches used for increase of vital functions of the specialist of a pension age. The main thing here is the increase in intensity of the control of a state of health. For electric power objects this sounds as « increase in intensity of the control of a technical condition» (further - the TC). And this is well known to us on the recommendation of transition from scheduled precautionary repair to repair on the TC [7], detailed explanations of expediency of the account of the TC [8, 9, 10] and increases of intensity of the control. An analogue of the operational control of the TC is the monthly preparation of the form of 3-tech (energy) [11], which characterizes the TC of power units of thermal power stations.

However, the opportunity does not always exist. An illustrative example of such an object is overhead transmission lines (further - OHL) with a rated voltage of 110 kV and above. By development of methods and algorithms of an estimation of parameters of operative reliability and safety in the illustrative purposes these will be used OHL.

II. Methodical Features Of An Estimation Of Parameters Of Operative Reliability OHL

Formulas of an estimation of parameters of reliability OHL well known and are used for the analysis of the reasons of occurrence of emergency switching-off. Increase of a faultlessness of recommendations is reached by using of statistical data for a number of years of supervision. Reference books and technical literature provide estimates of specific damageability, average downtime in emergency and planned (capital) repairs. Estimations are average, as a rule, on a class of a voltage, occasionally - for the reasons of refusal or a material of support. Specific damageability is led OHL in the extent of 100 km.

In other words, classification of statistical data is spent traditionally on one, a maximum three attributes. The algorithm of use of these estimations, on the recommendation of M.N. Rozanov (1984) consists in the following: «it is necessary to calculate failure of refusals for of some years of operation, to construct a graph and extrapolate frequency of refusals approximately for five years forward. The resulting failure refusals should be used in estimation the reliability of newly constructed objects". In table 1 according to Guk Y.V. (1974) intervals of change of estimations of specific number of damages OHL.

Table 1. Intervals of change of estimations of specific number of damage OHL.

Voltage, kV	Specific number of damages, on 100 km/years	
	Stable	Unstable
100; 154	0,5-1,7	5-7
220	0,25-1,5	1-2
330	0,15-1,6	0,5-1,5
500	0,2-1,1	0,15-2,5

These data are interesting in that the intervals of change in the specific number of damages crossed. This testifies to the inexpediency of their classification on the basis of "voltage class". The above-stated allows to conclude, that existing methods of an estimation of parameters of reliability OHL are unacceptable for an estimation of parameters of operative reliability.

In table 2 recommended main parameters of the operative reliability OHL and formulas for their calculation.

Table 2. Recommended parameters of operative reliability OHL.

Name of a parameter	Symbol	Unit of measure	Formula of an estimation
Parameter of a stream of refusals	λ_i^*	open/km.month	$\left[\sum_{i=1}^{N_i} n_{\Sigma,i,j}\right] \Big/ b_L \sum_{j=1}^{N_i} L_j$
Parameter of a stream of stable refusals	ω_i^*	open/km.month	$\left[\sum_{i=1}^{N_i} n_{y,i,j}\right] \Big/ b_L \sum_{j=1}^{N_i} L_j$
Average monthly probability of stable refusal	$R_{y,i}^*$	relative unit	$\frac{\omega_i^*}{\lambda_i^*}$
Average monthly duration downtime in emergency repair	$M_i^*(\tau_a)$	hour/month	$\left[\sum_{j=1}^{n_{y,i}} \tau_{e,i,j}\right] \Big/ n_{st,i}$
Average monthly coefficient downtime in emergency repair	$K_{\Pi,i}^*$	relative unit	$K_{P,i}^* = \frac{M_i^*(\tau_a)\cdot\omega_i^*}{T_i}$

The note: N_i ---number of working OHL in i-th month; $n_{\Sigma,i,j}$ - number of refusals of j-th OHL in i-th month; L_j - length of j-th OHL; $n_{st,i,j}$ - number of stable refusals of j-th OHL in i-th month; $\tau_{e,i,j}$ - duration of downtime in emergency repair of j-th OHL in i-th month; b_L - coefficient characterizing the length of a conditional line (L_y). It is calculated as $b_L=1/L_y$.

III. Methodical features of comparison of estimations of parameters of operative reliability OHL.

As noted above, “mechanical” comparison of estimates of parameters of operative reliability is associated with a high risk of erroneous decisions and caused by casual character of compared estimations. The relevance of the following operational problem is doubtless: *estimation of a degree of influence on reliability OHL EPS of performance of preliminary made decision concerning reorganization of operation, change of system of maintenance service and repair (further - STMR). This* estimation is spent by comparing the parameters of operative reliability before and after these changes.

Is no less actual the task: *the control of a infallibility of observed* regularity *of change of parameters of operative reliability over time with the purpose of use of this law for forecasting reliability*. The accounting of casual character of statistical indicators can be spent on the basis of the verification theory of statistical hypotheses. But, since the main task of the risk-focused approaches is decrease the risk of an erroneous decision, it is unacceptable to neglect the requirements, in accordance with which:

– the law of distribution of random variables must be known;
– application of criteria for testing statistical hypotheses calculated for one-dimensional random variables to multidimensional, ones sharply increases the risk of erroneous decisions.

Considered at the analysis of reliability OHL random variables far do not always correspond to the normal law, and estimations of parameters of operative reliability are multidimensional.

Overcoming of these difficulties is reached by application a fiducial approach, imitating modeling and the principles of the theory of testing statistical hypotheses [12]. With a high degree of accuracy, modeling algorithms for of some parameters of operative reliability can be replaced developed express method based on the approximation of fiducial distributions of some nonlinear function [13].

Unlike formulas for an estimation of parameters of operative reliability OHL, algorithms of comparison of the average monthly estimations of parameters of operative reliability are identical. Let's designate an estimation (*) the generalized parameter (P) operative reliability as P*. The algorithm of comparison of the average monthly estimation P* in i-th month P_i^* with the average monthly estimation in preceded (i-1) month P_{i-1}^* as follows:

If $P_{i-1}^* \le P_i^*$, otherwise

and $\left|L_{st,(i-1)}\right| \le \left|L_{st,i}\right|$ otherwise

and $P_i^* > \overline{P_{(i-1)}^{**}}$, that $H \Rightarrow H_2$

otherwise $H \Rightarrow H_0$

and $P_{i-1}^* < \underline{P_i^{**}}$, that $H \Rightarrow H_1$

otherwise $H \Rightarrow H_0$

and $\left|L_{st,(i-1)}\right| \le \left|L_{st,i}\right|$ otherwise

and $P_i^* \ge \underline{P_{(i-1)}^{**}}$, that $H \Rightarrow H_2$ (1)

otherwise $H \Rightarrow H_0$

and $P_{i-1}^* > \overline{P_i^{**}}$, that $H \Rightarrow H_2$

otherwise $H \Rightarrow H_0$

For illustrative purposes, formulas for calculating the boundary values of the fiducial interval of estimations $K_{P,(i-1)}^*$ and $K_{P,i}^*$ given in table 3.

Table 3. Formulas for estimating the boundary values of the fiducial interval: conditional downtime coefficients in emergency repair of OHL of EPS for (i-1) and i- th months of the year

Indicator	Evaluation formulas	Note
$\underline{K_{P,(i-1)}^{**}}$	$K_{P,(i-1)}^* \cdot \left(1 - A_\alpha / \sqrt{J_{(i-1)}}\right)$	$J_{(i-1)} = b_L \sum_{j=1}^{N_{(i-1)}} L_j$
$\overline{K_{P,(i-1)}^{**}}$	$K_{P,(i-1)}^* \cdot \left(1 + A_\beta / \sqrt{J_{(i-1)}}\right)$	
$\underline{K_{P,i}^{**}}$	$K_{P,i}^* \cdot \left(1 - A_\alpha / \sqrt{J_i}\right)$	$J_i = b_L \sum_{j=1}^{N_i} L_j$
$\overline{K_{P,i}^{**}}$	$K_{P,i}^* \cdot \left(1 + A_{(1-\beta)} / \sqrt{J_i}\right)$	

IV. Methodical Features Of Ranging Of Estimations Of Parameters Of Operative Reliability OHL ENE ESP.

The analysis of change of operative reliability OHL ESP is of course important, at least from the point of view of an estimation of influence on reliability of work carried out to improve the operation, STMR OHL. Clearly, that all these actions are spent in certain ENE and on certain OHL. Recognition of this OHL ENE carried out based on operating experience, intuitively, and often subjectively. This approach is not accidental. Simply there is no "help", allowing to reveal these «weak parts» and to lower risk of the erroneous decision. Below is a method and algorithm for solving this problem.

In the section discusses the varieties of the attribute by which it is expedient to spend ranging for revealing ENE, TC OHL, which demands intervention. To range them, i.e. to place in order of decreasing operational reliability, also does not represent special work. Difficulty consists that all these estimations of parameters of operative reliability OHL ENE have random character. This take place because of random character of their refusals. In other words, the observed difference of estimations, as well as their difference from parameters of operative reliability OHL EPS, can be casual, and classification itself is useless. We have met this fact at the analysis of specific number OHL of various voltage classes (see table 1).

But before to consider an opportunity of overcoming noted above difficulty, it is necessary to have in view of, that:

– the recommended method and algorithm do not depend on type of a parameter of operative reliability. Therefore we shall keep sense of estimations $P^{*}_{\Sigma,j}$, $\underline{P^{**}_{\Sigma,j}}$, $\overline{P^{**}_{\Sigma,j}}$, $P^{*}_{\nu,i,j}$, $\underline{P^{**}_{i,j}}$, $\overline{P^{**}_{i,j}}$. We take into account, that j - a ordinal number of month, j=1,12, i - a ordinal number of the varieties ν-th of the attribute, i=1,m_ν; ν - a serial number of an attribute, ν=1,m_a; m_a - number of consideration attributes;

– boundary values of the fiducial interval randomly differ from each other with a significance level of 2α. In other words, random character of a divergence of parameters of operative reliability OHL EPS and of some ENE yet does not mean random character of a divergence of all parameters of operative reliability OHL of these ENE.

The recommended method and algorithm of ranging of estimations of parameters of operative reliability OHL ENE EPS reduced to following sequence of calculations.

– the estimations $P^{*}_{\nu,i,j}$ located in ascending order are compared with $\underline{P^{**}_{\Sigma,j}}$. Are allocated in the first group of an estimation $P^{*}_{\nu,i,j}$ not exceeding $\underline{P^{**}_{\Sigma,j}}$, as estimations not casually differing from the estimation $P^{*}_{\Sigma,j}$;

– the remained estimations $P^{*}_{\nu,i,j}$ are compared with $\overline{P^{**}_{\Sigma,j}}$. The estimations $P^{*}_{\nu,i,j}$, not exceeding $\overline{P^{**}_{\Sigma,j}}$ are allocated the second group and characterized as an estimation casually differing from the estimation $P^{*}_{\Sigma,j}$;

– the part of estimations $P^{*}_{\nu,i,j}$ which exceeds $\overline{P^{**}_{\Sigma,j}}$ belongs to the third group. A rating of the first group (r1) ENE is agreed to estimate as "good", the second groups (r2) ENE - as "satisfactory", and the third group (r3) ENE - as "unsatisfactory".

Objective ranging of operative reliability OHL ENE involves overcoming another difficulty - presence of many parameters. This difficulty overcomes by transition to an integrated parameter. The recommended methodology for calculating the integral parameter was uses at the analysis of operative reliability of power units of thermal power stations [14].

The essence of recommendations is reduced:

– to the transition from the estimation of a parameter of operative reliability to the estimation of its probability on statistical function of distribution $F^*(P^*_{v,i,j})$. The value $F^*(P^*_{v,i,j})$ characterizes size of "wear" and is defined under the formula $Iz^*(P^*_{v,i,j}) = i_\nu / m^2_{in}$, where i - a ordinal number of ν -th parameter ENE in ranging data series for j-th month; ν=1,m_{in}, m_{in} - number of indicators of operative reliability;

– to calculation of an integrated parameter of operative reliability under the formula

$$In^*(Iz) = M^*\left[Iz^*(P_{i,j})\right] = \sum_{\nu=1}^{m_\nu} i_\nu \Big/ m^2_{in} \quad (2)$$

Not less significant it is necessary to consider a problem of an estimation of a degree of increase of operative reliability as a result of prospective change in STMR OHL ENE.

Let's consider methodology of the decision of this problem using the example of average duration of downtime in emergency repair $M^*_\Sigma(\tau_e)$. Suppose, that according on statistical data for j-th month of work it established, that:

– a monthly average estimation $M^*_{\Sigma,j}(\tau_{e,j}) = \sum_{i=1}^{n_\Sigma} \tau_{e,i,j} \Big/ n_\Sigma$, where n_Σ - frequency of downtime in emergency repair;

– as a result of classification of statistical data OHL on ENE, their rangings in order of decreasing in operative reliability and an estimation of character of a divergence with $M^*_{\Sigma,j}(\tau_{e,j})$, it is established, that monthly average estimations $M^*_{\nu,j}(\tau_{e,i})$ of each of three groups ESP are accordingly equal:

$$M^*_{г1,j}(\tau_{e,j}) = \sum_{i=1}^{n_{г1}} \tau_{e,i,j} \Big/ n_{г1};$$

$$M^*_{г2,j}(\tau_{e,j}) = M^*_{\Sigma,j}(\tau_{e,j}) = \sum_{i=1}^{n_{г2}} \tau_{e,i,j} \Big/ n_{г2};$$

$$M^*_{г3,j}(\tau_{e,j}) = \sum_{i=1}^{n_{г3}} \tau_{e,i,j} \Big/ n_{г3}$$

where $n_{г1}$, $n_{г2}$, $n_{г3}$ - accordingly number of realizations of duration of emergency downtime in first, second and third groups ESP

Since in the third group are placed ENE, monthly average duration of downtime time in emergency repair in which nonrandom exceeds $M^*_{\Sigma,j}(\tau_{e,j})$, that, naturally, it is necessary to provide, first of all, restoration of wear OHL ENE of the third group. Obviously, as a result of restoration of wear, monthly average value $M^*_{г3,j}(\tau_{e,j})$ will randomly differ from the OHL ENE of the first group. At the same time, thus decrease in duration of downtime in emergency repair ENE of the third group will be equal:

$$\Delta\sum_{i=1}^{n_{г3}} \tau_{e,i,j} = \sum_{i=1}^{n_{г3}} \tau_{e,i,j} - M^*_{г1,j}(\tau_{e,j}) n_{г3,j} \quad (3)$$

And the relative size of this decrease is equal:

$$\delta M^*_{\Sigma,j}(\tau_{e,j}) = \frac{n_{г3,j}\left[M^*_{г3,j}(\tau_{e,j}) - M^*_{г1,j}(\tau_{e,j})\right]}{n_{\Sigma,j} \cdot M^*_{\Sigma,j}(\tau_{e,j})} \quad (4)$$

By analogy, formulas can be received of an estimation of relative value of change of operative reliability OHL EPS and for other parameters. Results of these transformations are shown in table 4.

Table 4. Formulas of an estimation of possible change of operative reliability OHL EPS as a result of restoration of wear

Parameter	Formulas for calculating a possible increase in operational reliability
$\delta\lambda^*_{\Sigma,j}$	$\left[b_L(\lambda^*_{r3,j}-\lambda^*_{r1,j})\sum_{i=1}^{N_{r3,j}}L_i\right]\Big/ b_L\lambda^*_{\Sigma,j}\sum_{i=1}^{N_\Sigma}L_i$
$\delta\omega^*_{\Sigma,j}$	$\left[b_L(\omega^*_{r3,j}-\omega^*_{r1,j})\sum_{i=1}^{N_{r3,j}}L_i\right]\Big/ b_L\omega^*_{\Sigma,j}\sum_{i=1}^{N_\Sigma}L_i$
$\delta M^*_{\Sigma,j}(\tau_{a,j})$	$\dfrac{n_{r3,j}\left[M^*_{r3,j}(\tau_{a,j})-M^*_{r1,j}(\tau_{a,j})\right]}{n_{\Sigma,j}M^*_{\Sigma,j}(\tau_{a,j})}$
$\delta K^*_{\Pi,\Sigma,j}$	$\dfrac{\omega_{r3,j}\left[M^*_{r3,j}(\tau_{a,j})-M^*_{r1,j}(\tau_{a,j})\right]}{\omega_{\Sigma,j}M^*_{\Sigma,j}(\tau_{a,j})}$
$\delta R^*_{\Sigma,j}$	$\dfrac{n_{r3,j}\left[R^*_{r3,j}-R^*_{r1,j}\right]}{n_{\Sigma,j}R^*_{\Sigma,j}}$

V. Methodical Features Of Benchmarking Of Operative Reliability ENE EPS.

Benchmarking of ENE EPS belongs to the category internal ones. Remind, that internal benchmarking [15]:

- it is a kind of the comparative analysis;
- its essence consists in revealing most and the least effective same type objects;
- it is least costly a kind of research;
- its main task - to reveal objects, which increase in the efficiency of which to the greatest extent increases the efficiency of the system as a whole;
- owing to the simplicity, is the best way of decrease in risk of the erroneous decision;
- the greatest effect takes place only at the regular comparative analysis. A single use only leads to temporary success.

The comparative analysis of operative reliability of set OHL ENE EPS allows to correctly solve many operational problems at level EPS. But it is absolutely insufficient for ENE themselves. And indeed. Certainly, it is important to management ENE to know, how they govern ENE differs from others ENE EPS; as reliability of work OHL in billing month has changed; how effective were the new approaches to the recognition of hazardous defects. But, first of all, ENE leader must know exactly, where to direct efforts for increase of rating ENE.

As noted above, "mechanical" classification OHL ENE is connected with high risk of the erroneous decision. For revealing of "weak parts», that reduce the ENE rating, it is necessary:

- For each ENE, whose TC rating is assessed as unsatisfactory:
  - to classify OHL and statistical data about their idle time in emergency repair on each of n_p-1 the signs attributes (except for an attribute name ENE) and to its varieties. Preliminary with the continuous character of change of varieties of attributes, they are transformed in discrete;
  - for each varieties of an attribute the estimation of a parameter of operative reliability $P^*_{\nu,i,j}$ is calculated;

* the maximal value of estimations $P^*_{\nu,i,j}$ is defined, with ν=1,m_i, where m_i - number of varieties i-th an attribute $P^*_{max,i,j} = \max\left\{P^*_{1,i,j},.....,P^*_{m_i,i,j}\right\}$. Obvious, that is the most significant varieties of the i-th attribute;
* the most significant attribute is defined;

$$P^*_{max,j} = \max\left\{P^*_{max,1,i,j};P^*_{max,2,i,j};.....;P^*_{m_i,n_{II},j}\right\}$$

– Attributes are defined, estimations of parameters operative reliability of which exceed the estimation $P^*_{r3,j}$;

– On these attributes classification OHL and statistical data about their downtime in emergency repair is carried out, according to which the estimation $P^*_{max,j}$ is spent calculated;

– Further calculations are carried out similarly to the above. Calculations completed by consideration of all possible and expedient classifications and all ENE with a unsatisfactory rating.

As a result of the calculations, for each ENE with a unsatisfactory rating are determined set of OHL, that determine this rating. This list recommended for restoration TC OHL.

Vi. Formation Of Information And Methodical Support For Management ENE And EPS.

By increase of reliability the risk-focused of the approach naturally is increase science intensity, cumbersomeness and laboriousness of manual calculation. Check of reliability is carried out by a method of the decision of "a return problem» when recommendations are trivial. Efforts are required only for preparation in the tabulated in form of monthly data on automatic emergency switching-off OHL. In day of input of the information Chief engineers EPS and ENE receive the specialized forms containing information on operative reliability, accordingly, OHL EPS and ENE and the recommendation on increase of an work efficiency [16].

For illustrative purposes, fig. 1 shows the form of information about the monthly average operative reliability OHL EPS. It is necessary to note, that the content of this form depends as on time of use of opportunities of the automated monitoring system of operative reliability OHL, and on the "interest" of the ENE and EPS managers.

Date of the analysis *Chief engineer EPS*
05 may 2021 год

Data
about operative reliability OHL in April 2021

1. Monthly average value of a parameter of operative reliability is estimated as ***satisfactory****;*
2. In comparison with ***March*** *operative reliability has decreased on* ***15,3%;***
3. In the table And are resulted located by way of decreasing of operative reliability ENE EPS.
Table A. ENE EPS operational reliability details

Name ENE	*Integrated parameter of wear, r.u.*	*Relative change over time, %*	*Rating ENE*
ENE 5	*0,053*	*5,7*	*good*
ENE 7	*0,12*	*10,2*	*satisfactory*
ENE 2	*0,23*	*25,6*	*satisfactory*

4. Recommendations on increase of operative reliability.
4.1. In ENE, the rating of operative reliability OHL which is unsatisfactory, to lead selective survey OHL (see Table B)
Table B. List of OHLs subject to examination

Name	
ENE	Overhead transmission lines
ENE 2	*B 26*
	B 8
ENE 9	*B 36*
	B 49

4.2. On specified in table B OHL to lead diagnostics of a technical condition using modern methods of the analysis.
4.3. To organize for personnel ENE with unsatisfactory technical condition OHL short-term courses of improvement of qualification.

Fig. 1. The form of data on operative reliability OHL EPS.

Conclusion

1. Set of parameters of operative reliability and the formula of their estimation is proposed;
2. The new algorithm of an estimation of expediency of classification of multivariate data about refusals and duration of downtime OHL is developed. At each stage of classification total OHL is represented by three groups. The first group includes OHL, an estimation of a indicator of which operative reliability is not casual less estimations of a similar indicator for total OHL. The second group includes OHL, the estimation of a indicator of which operative reliability casually differs from a similar indicator for total OHL. The third group includes OHL, an estimation of a indicator of which operative reliability is not casual more estimations of a similar indicator for total OHL;
3. Methods and algorithms for comparing of two multivariate estimations of parameters of operative reliability OHL are developed, allowing to estimate character of change of reliability of these OHL in time;
4. Methods and algorithms of ranging of parameters of operative reliability OHL the electric network enterprises are developed, allowing to identify OHL, that demanding operational inspection;
5. Methods and algorithms of calculation of an estimation of an integrated parameter of operative reliability are developed;
6. Monthly information and methodical support of a technical management of the electric network enterprises and an electric power system is provided. Methodical support includes recommendations on increase of reliability OHL.

References

[1] Volkova I.O. Concept of construction of a control system by actives of the electric network companies in Russia on the basis on benchmarking. Economy and management of a power complex. Scientific and technical statements. SPbSPU №2, 2008, p.101-105;
[2] Karpov N.V. Strategy for managing the production of industrial actives of the electric network company. The bulletin of Omsk University Series "Economics", 2019, v.17, №2. ***DOI 10.25513/1812-3988.2019 s112-123****;*
[3] Grabchak E.P. Estimation of a technical condition of the power equipment in conditions of digital economy // Reliability and safety in power. 2017, v10, №4, p.268-274;
[4] V Scientifically Practical Conference «Control of a technical condition of the equipment of objects of electric power industry» / Electricity. Transmission and distribution - 2018;
[5] VI Scientifically Practical Conference «Control of a technical condition of the equipment of objects of electric power industry // Electricity. Transmission and distribution -2019;
[6] Danilov G.A. Improving the quality of functioning of overhead transmission lines. Monograph. Danilov G.A., Denchik Yu.N., Ivanov N.N., Sitnikov G.V. 3rd ed. M.- Berlin: Direct-Media, 2019, 558 p.;
[7] STO-34.01-24-003-2017. UNEG Production Assets Management System, 2017. Federal Network Company;
[8] STO-34.01-23.1-001-2017. Volume and rate of test of an electric equipment;
[9] STO-34-01-24-002-2018. The organization of maintenance service and repair of objects of electric power industry of the UNEG, 2018, FNC;
[10] STO-34.01-35-001-2020 Methodical instructions on carrying out of technical survey of the equipment of substations, overhead transmission lines. "Rosseti", 2020, 20 p.;
[11] RD 34.09.453 Typical algorithm of calculation the technical and economic indicators of powerful heating thermal power stations SPO "Soyuztekhenergo": Updated 01.01.2021;

[12] Farhadzadeh E.M., Muradalyiev A.Z., Abdullayeva S.A. Fidutsial the approach at comparison of the same objects. Kiev. Electronic modelling №1, 2019, p.55-66, ***DOI: 10.15407/emodel.42.01.013****;*

[13] Farhadzadeh E.M., Muradaliyev A.Z., Dzhalagova E.I., Abdullayeva S.A. Method and algorithm of comparing the efficiency of gas piston power plants of electric power systems M: Izv. RAS "Energetika" No 2, 2019, p.106-17. **DOI: 10.1134/S0002331019020067;**

[14] Farhadzadeh E.M., Muradaliyev A.Z., Farzaliyev Yu.Z., Abdullayeva S.A. Comparison and ranging steam turbine installations of power units TES by operating efficiency. Heat power engineering, 2018, No. 10, p. 41-49. **DOI: 10.1134/S0040363618100028**;

[15] Cane M.M., Ivanov B.V., etc. Systems, methods and their tools of a quality management. 2nd ed. Id. "Peter", 2017, p. 576.

[16] Farhadzadeh E.M., Muradaliyev A.Z., Rafiyeva T.K., Rustamova A.A. Ensuring of reliability of methodical support of objects of electric power systems. M:, Electricity, No.2, 2020, p. 4-9. **DOI: 10.24160/0013-5380-2020-2-4-9**

RT&A, No 4 (65)
Volume 16, December 2021

On Transmuted Exponential-Topp Leon Distribution with Monotonic and Non-Monotonic Hazard Rates and its Applications

Aminu Suleiman Mohammed[1,2*] , Fidelis Ifeanyi Ugwuowo[2]

•

[1]Department of Statistics, Ahmadu Bello University, Zaria
[2]Department of Statistics, University of Nigeria, Nsukka
Email: fidelis.ugwuowo@unn.edu.ng
*Corresponding author: Email: mohammedas@abu.edu.ng

Abstract

For the last decade, inspired by the increasing demand for probability distributions in numerous fields, many generalized distributions have been studied. Most of these distributions are developed by adding one or more parameter(s) to the standard probability distributions to make them flexible in capturing the sensitive parts of a dataset. The Topp-Leone distribution (TL) is one of the continuous probability distributions used in modelling lifetime datasets and sometimes is called J-shaped distribution. In this paper, we proposed a new lifetime distribution named transmuted Exponential- Topp Leon distribution in short (TE-TLD) which possessed different density shapes. Some properties of the distribution were presented in an explicit form and the parameters of the distribution are estimated by the method of maximum likelihood. The hazard function of the TE-TLD can be monotonic or non-monotonic failure rate which makes it more robust in terms of studying failure rates. The TE-TLD outperformed other distributions with the same underlying baseline distribution when applied to real datasets in the study. Furthermore, the likelihood ratio test (LRT) shows that the additional parameter(s) are significant which further proves the robustness of the TE-TLD over the nested distributions in the study.

Keywords: Topp Leon distribution, failure rate, Maximum Likelihood, generalized distributions

1. Introduction

In reliability and survival analysis, lifetime distributions such as Exponential, Weibull among others, play an important role in modelling lifetime data. Most of these distributions have infinite support in theory, as the lifetime of a system or item can be infinite. On the other hand, distributions with finite support will be appropriate in modelling data sets that are generated as a result of limited power supply, the design life of the system, among others [1]. For the past few years, inspired by the increasing demand for probability distributions in numerous fields, many generalized distributions have been studied. Most of these distributions are developed by adding one or more parameter(s) to the standard probability distributions to make them robust in capturing the sensitive parts of a dataset. For example, [2] proposed Beta Exponential distribution which has three parameters and was found to be more flexible than the classical Exponential distribution.

The Topp-Leone distribution (TL) is one of the continuous probability distributions used in modelling lifetime datasets and sometimes is called J-shaped distribution. This distribution has a closed-form and was proposed by [3]. However, the J-shaped distribution had not received much attention due to some of its complexity until [4] studied some properties of the distribution which include moments, central moments, and characteristic function. This work led to increasing interest in studying TL distribution. For instance, [1] studied and explored some reliability

measures and their stochastic orderings, a comprehensive study on flat-toppedness of the TL distribution was studied by [5], study on record values by [6], the moments of the order statistics of the TL distribution was studied by [7], the goodness-of-fit tests for the TL distribution are evaluated by [8] and [9] proposed Topp-Leone-Exponential distribution which has two parameters and is skewed to the right.

Current kinds of literature pay more attention to propose more flexible distributions but give less concern to the hazard function of the distributions. In reliability analysis, hazard rate plays an important role to characterize life phenomena and as well guides in model selection [10]. Furthermore, many systems exhibit failure rates that are non-monotonic. For instance, the failure rate pattern of numerous electronic components comprises of three phases: initial phase (or burn-in) where failure is high at the start of the product life cycle due to design and manufacturing problems and decreases to a constant level, the middle phase (flat region) with an approximately constant hazard rate, and the final phase (or wear-out stage), from where the hazard rate starts to increase: This failure (hazard) rates are "U" or bathtub shaped. The Exponential, Weibull, Gamma among other distributions, and some of their extensions allow only monotone failure rates and are unable to produce bathtub shape and thus cannot appropriately describe the datasets with this feature. These have opened room for more research that can account for monotone and non-monotone hazard rate function.

In this research, we developed an extension of Topp Leon distribution named transmuted Exponential-Topp Leon distribution (TE-TLD) which possessed both monotonic and non-monotonic failure rate shapes and its density function can be left-skewed, right-skewed, Bathtub, or J-shape. The cumulative distribution and probability density function of Topp Leon distribution are respectively given as;

$$G(x,\alpha) = x^{\alpha}(2-x)^{\alpha} \tag{1}$$

and

$$g(x,\alpha) = 2\alpha x^{\alpha-1}(1-x)(2-x)^{\alpha-1} \tag{2}$$

where, $0 < x < 1 \, and \, \alpha > 0$.

Based on the work of [11], the cdf and pdf of Transmuted Exponential-G family of distributions are respectively given by;

$$F(x;\lambda,\theta,\xi) = \left(1-(1-G(x,\xi))^{\lambda}\right)\left(1+\theta\,(1-G(x,\xi))^{\lambda}\right) \tag{3}$$

and

$$f(x;\lambda,\theta,\xi) = \frac{g(x,\xi)}{1-G(x,\xi)}\lambda\,(1-G(x,\xi))^{\lambda}\left(1-\theta+2\theta\,(1-G(x,\xi))^{\lambda}\right) \tag{4}$$

Where, $G(x,\xi)$ and $g(x,\xi)$ are the baseline cdf and pdf respectively depending on a vector parameter ξ whereas, $\lambda > 0$, $-1 \leq \theta \leq 1$ are two additional parameters i.e scale and transmuted (shape) parameter respectively.

2. Transmuted Exponential-Topp Leon Distribution

Substituting equations (1) and (2) into (3) yields cumulative distribution function (cdf) of the transmuted Exponential-Topp Leon distribution (TE-TLD).

$$F(x) = \left[1-(1-x^{\alpha}(2-x)^{\alpha})^{\lambda}\right]\left[1+\theta\,(1-x^{\alpha}(2-x)^{\alpha})^{\lambda}\right] \tag{5}$$

and the associated probability density function (pdf) is given by;

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

where, $0 < x < 1,\ \alpha\,\lambda > 0\, and\ -1 \leq \theta \leq 1.$

A useful linear representation for the pdf of TE-TLD is given as;

$$f(x) = 2\sum_{j=0}^{n} A_j x^{\alpha(1+j)-1}(1-x)(2-x)^{\alpha(1+j)-1} \tag{7}$$

$$where\ A_j = (-1)^j \alpha\lambda \left\{ (1-\theta) \begin{pmatrix} \lambda - 1 \\ j \end{pmatrix} + 2\theta \begin{pmatrix} 2\lambda - 1 \\ j \end{pmatrix} \right\}$$

2.1. Distribution validity check

Fact 1: The TE-TLD is a valid density function.

$\int_0^1 f(x;\alpha\,\lambda\,\theta)dx = 1$

Proof:

$\int_0^1 \left(2\alpha\lambda x^{\alpha-1}(1-x)(2-x)^{\alpha-1}\left(1-x^{\alpha}(2-x)^{\alpha}\right)^{\lambda-1}\left[1-\theta+2\theta\left(1-x^{\alpha}(2-x)^{\alpha}\right)^{\lambda}\right]\right)dx$

$let\ u = 1 - x^{\alpha}(2-x)^{\alpha},\ \ as\ x \to 0,\ u \to 1\ and\ x \to 1,\ u \to 0$

$\frac{du}{dx} = -\left(x^{\alpha}\alpha(2-x)^{\alpha-1}(-1) + (2-x)^{\alpha}\alpha x^{\alpha-1}\right) = -\left(x^{\alpha}\alpha(2-x)^{\alpha}\left(\frac{-1}{(2-x)} + \frac{1}{x}\right)\right)$

$\frac{du}{dx} = -x^{\alpha}\alpha(2-x)^{\alpha}\left(\frac{-x+2-x}{(2-x)x}\right) = \frac{-2\alpha x^{\alpha}(2-x)^{\alpha}(1-x)}{(2-x)x}$

$dx = \frac{-du}{2\alpha x^{\alpha-1}(2-x)^{\alpha-1}(1-x)}$

$\int_0^1 \left(2\alpha\lambda x^{\alpha-1}(1-x)(2-x)^{\alpha-1}u^{\lambda-1}\left[1-\theta+2\theta u^{\lambda}\right]\right)\frac{du}{2\alpha x^{\alpha-1}(2-x)^{\alpha-1}(1-x)}$

$\int_0^1 \lambda u^{\lambda-1}(1-\theta+2\theta u^{\lambda})du$

$let\ v = u^{\lambda},\ \frac{dv}{du} = \lambda u^{\lambda-1},\ du = \frac{dv}{\lambda u^{\lambda-1}}$

$\int_0^1 \lambda u^{\lambda-1}(1-\theta+2\theta v)\frac{dv}{\lambda u^{\lambda-1}}$

Finally,

$\int_0^1 (1-\theta+2\theta v)dv = \left(v - \theta v + \theta v^2\right)_0^1 = 1$

2.2. Graphical illustration of the pdf and cdf of TE-TLD

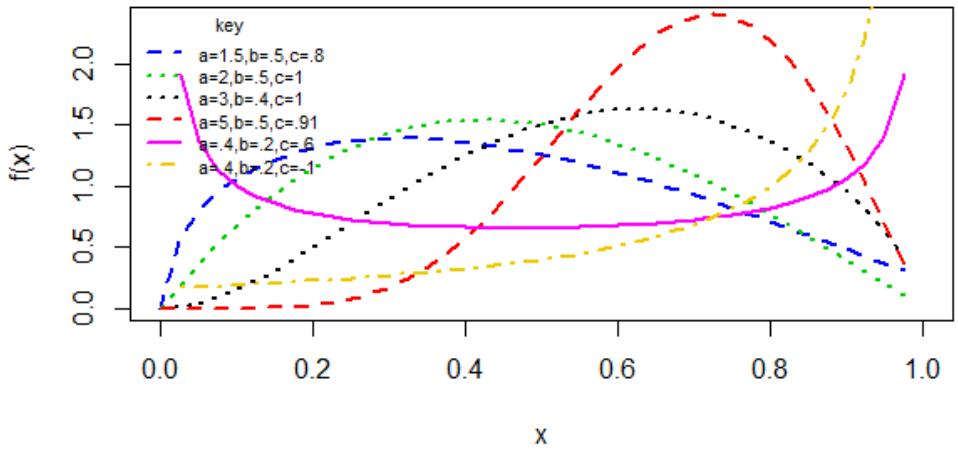


Figure 1: *The pdf plot of TE-TLD*

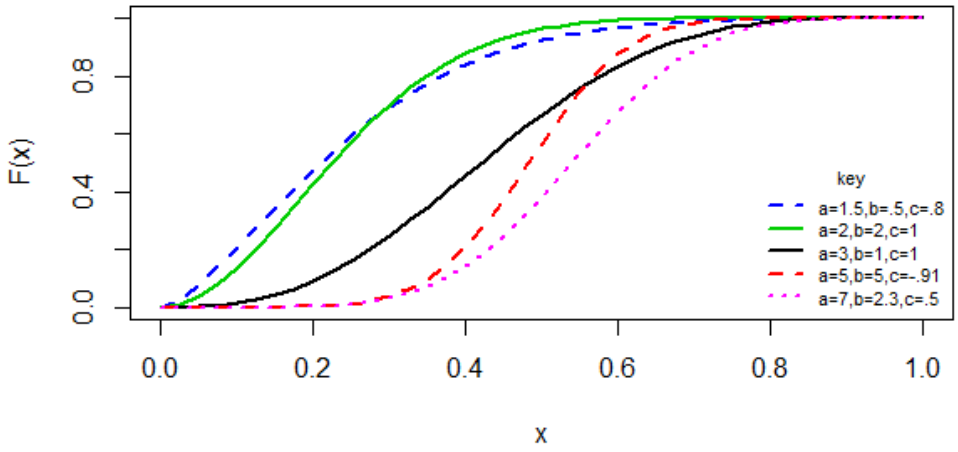


Figure 2: *The cdf plot of TE-TLD*

The shape of the density function corresponding to the TE-TLD may be characterized as follows;

- 1. For $\alpha = 1.5,\ \lambda = .5,\ \theta = 0.8\, and\, \alpha = 2,\ \lambda = .5,\ \theta = 1$, f(x) has a positive skewed shape.
- For $\alpha = 3,\ \lambda = .4,\ \theta = 1\, and\, \alpha = 5,\ \lambda = .5,\ \theta = .91$, f(x) has a negative skewed shape.
- For $\alpha = .4,\ \lambda = .2,\ \theta = .6$, f(x) has a bathtub shape.
- For $\alpha = .4,\ \lambda = .2,\ \theta = -1$, f(x) has a J-shaped.

3. Statistical properties of TE-TLD

In this section, some basic properties of TE-TLD are provided in an explicit form.

3.1. Moments and Moment generating function

The r^{th} Moments of TE-TLD is given by;

$$\mu_r^{'} = 2^r \sum_{j=0}^{\infty} A_j 4^{\alpha(1+j)} Be\left(r+\alpha(1+j), \alpha(1+j); \frac{1}{2}\right) - 2^{r+1} \sum_{j=0}^{\infty} A_j 4^{\alpha(1+j)} Be\left(r+\alpha(1+j)+1, \alpha(1+j); \frac{1}{2}\right) \tag{8}$$

$where\ A_j = (-1)^j \alpha\lambda \left\{(1-\theta)\begin{pmatrix} \lambda-1 \\ j \end{pmatrix} + 2\theta \begin{pmatrix} 2\lambda-1 \\ j \end{pmatrix}\right\}$

The r^{th} Moment about the Mean is given by;

$$\begin{aligned} E(x-\mu)^r = \quad & \alpha\lambda \sum_{j=0}^{\infty} \sum_{k=0}^{r} \psi_{j,k} 2^{r+1+2\alpha(1+j)-k-1} Be\left(r+\alpha(1+j)-k, \alpha(1+j); \tfrac{1}{2}\right) \\ & - \alpha\lambda \sum_{j=0}^{\infty} \sum_{k=0}^{r} \psi_{j,k} 2^{r+1+2\alpha(1+j)-k} Be\left(r+\alpha(1+j)-k+1, \alpha(1+j); \frac{1}{2}\right) \end{aligned} \tag{9}$$

where,

$$\psi_{j,k} = (-1)^{j+k} \begin{pmatrix} r \\ k \end{pmatrix} \mu^k \left\{(1-\theta)\begin{pmatrix} \lambda-1 \\ j \end{pmatrix} + 2\theta \begin{pmatrix} 2\lambda-1 \\ j \end{pmatrix}\right\}$$

and, The moment generating function of TE-TLD is given by;

$$M_x(t) = \sum_{r,j=0}^{\infty} \frac{2^r t^r}{r!} A_j 4^{\alpha(1+j)} Be\left(r+\alpha(1+j), \alpha(1+j); \frac{1}{2}\right) - \sum_{r,j=0}^{\infty} \frac{2^r t^r}{r!} A_j 2^{2\alpha(1+j)+1} Be\left(r+\alpha(1+j)+1, \alpha(1+j); \frac{1}{2}\right) \tag{10}$$

$where\ A_j = (-1)^j \alpha\lambda \left\{(1-\theta)\begin{pmatrix} \lambda-1 \\ j \end{pmatrix} + 2\theta \begin{pmatrix} 2\lambda-1 \\ j \end{pmatrix}\right\} and Be(.,.,u)$ is an incomplete Beta function .

3.2. Survival and Hazard function of TE-TLD

The survival and hazard function are respectively given as;

$$S(x) = 1 - \left[1 - (1 - x^{\alpha}(2-x)^{\alpha})^{\lambda}\right]\left[1 + \theta\,(1 - x^{\alpha}(2-x)^{\alpha})^{\lambda}\right] \tag{11}$$

and,

$$h(x) = \frac{2\alpha\lambda x^{\alpha-1}(1-x)(2-x)^{\alpha-1}\left(1-x^{\alpha}(2-x)^{\alpha}\right)^{\lambda-1}\left[1-\theta+2\theta\left(1-x^{\alpha}(2-x)^{\alpha}\right)^{\lambda}\right]}{1-\left[1-\left(1-x^{\alpha}(2-x)^{\alpha}\right)^{\lambda}\right]\left[1+\theta\left(1-x^{\alpha}(2-x)^{\alpha}\right)^{\lambda}\right]} \tag{12}$$

The plots of the survival and hazard function for some selected values of parameters are respectively displayed in Figures 3 and 4 as shown below;

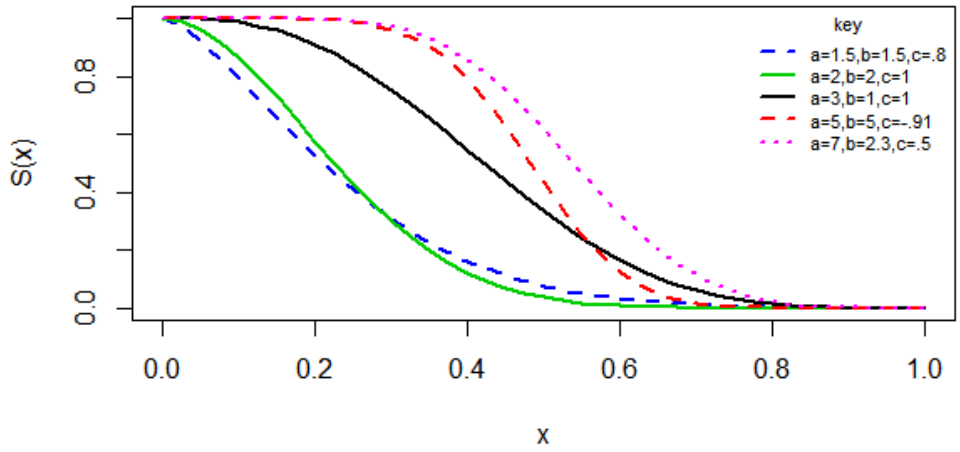


Figure 3: *Survival plot of TE-TLD*

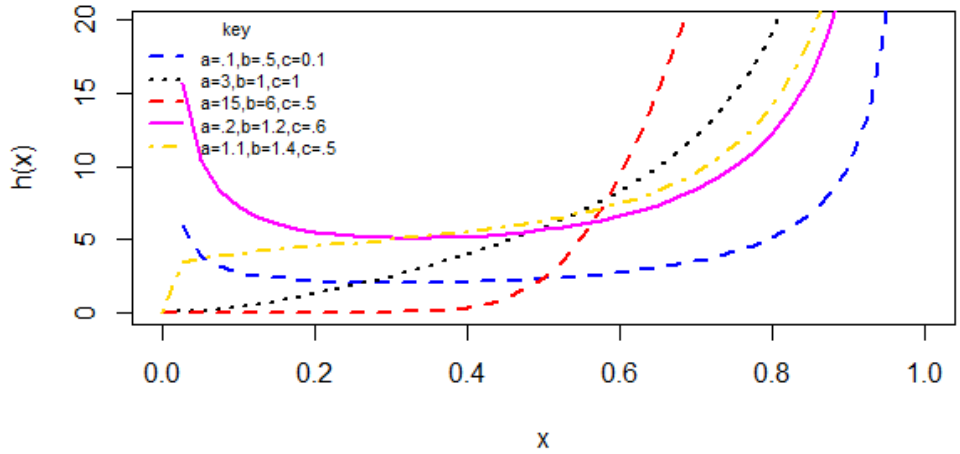


Figure 4: *Hazard plot of TE-TLD*

Figure 4 reveals that the hazard function of the TE-TLD possesses not only monotonic and non-monotonic failure rate shapes but also includes modified increasing failure rates. The monotonicity of the TE-TLD implies that the distribution may be a better choice when modelling age-dependent events where the risk increases with age. The resulting bathtub curve describes not only the behaviour of engineering components but also the lifetimes of human populations. Furthermore, the non-monotonicity of the TE-TLD implies that early failure or "infant mortality" is called the first stage of the bathtub curve and it is characterized by a decreasing component of the hazard rate. The weak members of the population are failing during this period. This section of the curve is based on the widely used testing practice of obviously defective components as well as weak ones with high failure potential. Products must survive some sort of initial stress during screening processes (e.g., burn-in at high temperature, application of electrical overstress, temperature cycling). Furthermore, the second stage is a roughly flat part called the intrinsic failure period. The hazard rate here is approximately constant and the failures occur at random in this area and most of the useful life of a component or system is spent here. The last stage on the curve is called the wear-out failure period and the hazard rate increase in this phase.

3.3. Quantile function of TE-TLD

For a non-negative continuous random variable X, that follows the TE-TLD, the quantile function is given by;

$$Q(u) = 1 - \sqrt{1 - \left(1 - \left(\frac{\theta - 1 + \sqrt{(\theta-1)^2 + 4\theta(1-u)}}{2\theta}\right)^{1/\lambda}\right)^{1/\alpha}} \tag{13}$$

3.3.1 Simulation Study

Numerical results are obtained by generating N=1000 random samples of size n=200 from $TE - TL(\alpha, 0.4, 0.8)$, $TE - TL(0.5, \lambda, 0.8)$ & $TE - TL(0.5, 0.4, \theta)$, where $\alpha = 0.5, 1, 1.5, 2, 3$ & 10; $\lambda = 0.2, 0.5, 1, 1.5, 2$& 2.5; $\theta = -1, -0.5, -0.2, 0.2, 0.5$ & 1. From the numerical results in Table 1, it was observed that the mean increases while variance, skewness and kurtosis decrease when α increases. While from Table 2, it was observed that for a constant value of α *and* θ, the mean and variance decrease both skewness and kurtosis increase as λ increases. Additionally, from Table 3, it was observed that as θ increases, both mean and variance decrease while at some certain point the variance increases and the skewness changes direction from negative to positive and the kurtosis decreases as well as increases at some point.

Table 1: *Mean, variance, skewness and kurtosis of TE-TLD for* $\lambda = 0.4$, $\theta = 0.8$ *and different values of* α

	Parameter (α)					
	0.5	1	1.5	2	3	10
Mean	0.3039 (0.0198)	0.4192 (0.019)	0.4910 (0.0178)	0.5414 (0.0167)	0.6091 (0.0149)	0.7710 (0.0094)
Variance	0.0805 (0.0069)	0.0755 (0.0053)	0.0668 (0.0046)	0.0590 (0.0041)	0.0472 (0.0035)	0.019 (0.0017)
Skewness	0.8090 (0.1226)	0.3622 (0.1042)	0.1389 (0.1005)	0.0011 (0.1008)	-0.1623 (0.1050)	0.4556 (0.1282)
Kurtosis	2.4942 (0.2688)	2.0439 (0.1382)	2.0151 (0.1067)	2.0665 (0.1063)	2.2004 (0.1331)	2.6575 (0.2797)

Table 2: *Mean, variance, skewness and kurtosis of TE-TLD for* $\alpha = 0.5$, $\theta = 0.8$ *and different values of* λ

	Parameter (λ)					
	0.2	0.5	1	1.5	2	2.5
Mean	0.4899 (0.0231)	0.2506 (0.0179)	0.1214 (0.0114)	0.0729 (0.0079)	0.0490 (0.0059)	0.0353 (0.0046)
Variance	0.1128 (0.0060)	0.0660 (0.0068)	0.0265 (0.0047)	0.0128 (0.0031)	0.0070 (0.0021)	0.0042 (0.0015)
Skewness	0.0556 (0.0060)	1.0672 (0.1363)	1.9809 (0.2547)	2.6254 (0.4628)	3.1268 (0.7152)	3.5226 (0.9585)
Kurtosis	1.6151 (0.0674)	3.1405 (0.3954)	7.0270 (1.5151)	11.4423 (3.9941)	15.8680 (7.5188)	19.9265 (11.2741)

Table 3: *Mean, variance, skewness and kurtosis of TE-TLD for* $\alpha = 0.5$, $\lambda = 0.4$ *and different values of* θ

	Parameter (θ)					
	-1	-0.5	-0.2	0.2	0.5	1
Mean	0.6416 (0.0.0187)	0.5469 (0.0216)	0.4901 (0.0223)	0.4151 (0.0223)	0.3594 (0.0215)	0.2667 (0.0063)
Variance	0.0757 (0.0055)	0.0994 (0.0058)	0.1054 (0.0058)	0.1040 (0.0063)	0.0955 (0.0068)	0.0666 (0.0063)
Skewness	-0.5251 (0.1064)	-0.2082 (0.1028)	0.0232 (0.1042)	0.3386 (0.1095)	0.5780 (0.1151)	0.9236 (0.3261)
Kurtosis	2.1826 (0.1715)	1.7621 (0.3954)	1.6593 (0.0691)	1.7660 (0.1167)	2.0456 (0.1837)	2.7921 (0.3261)

4. Maximum likelihood estimation

Let x_1, x_2, ..., x_n be a sample of size (n) from $TE - TL(\alpha, \lambda, \theta)$ distribution. Then, the log-likelihood function (LL) for the parameter vector $\Omega = (\alpha, \lambda, \theta)^T$ is given as;

$$LL(\Omega) = n\log 2 + n\log\alpha + n\log\lambda + (\alpha-1)\sum_{i=1}^{n}\log x_i + \sum_{i=1}^{n}\log(1-x_i) + (\alpha-1)\sum_{i=1}^{n}\log(2-x_i)$$
$$+ (\lambda-1)\sum_{i=1}^{n}\log(1-x_i^{\alpha}(2-x_i)^{\alpha}) + \sum_{i=1}^{n}\log\left[1-\theta+2\theta\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda}\right] \quad (14)$$

To get the MLE of the unknown parameters, find the first partial derivative with respect to each of the parameter of the distribution.

$$\frac{\delta LL(\Omega)}{\delta\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\log x_i + \sum_{i=1}^{n}\log(2-x_i) + (\lambda-1)\sum_{i=1}^{n}\frac{x_i^{\alpha}(2-x_i)^{\alpha}\ln(x_i(2-x_i))}{(1-x_i^{\alpha}(2-x_i)^{\alpha})}$$
$$-2\lambda\theta\sum_{i=1}^{n}\frac{\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda-1}x_i^{\alpha}(2-x_i)^{\alpha}\ln(x_i(2-x_i))}{\left[1-\theta+2\theta\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda}\right]}$$
$$\frac{\delta LL(\Omega)}{\delta\lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n}\log(1-x_i^{\alpha}(2-x_i)^{\alpha}) + 2\theta\sum_{i=1}^{n}\frac{\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda}\ln\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)}{\left[1-\theta+2\theta\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda}\right]}$$
$$\frac{\delta LL(\Omega)}{\delta\theta} = \sum_{i=1}^{n}\frac{2\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda}-1}{\left[1-\theta+2\theta\left(1-x_i^{\alpha}(2-x_i)^{\alpha}\right)^{\lambda}\right]}$$

Finally, setting this system of non-linear equations to zero and solving them simultaneously gives the MLE $\widehat{\Omega} = (\widehat{\alpha}, \widehat{\lambda}, \widehat{\theta})^T$. Furthermore, these non-linear equations cannot be solved analytically and as such a numerical method of optimization should be employed.

5. APPLICATIONS

In this section, we demonstrate empirically the flexibility of the TE-TLD using an application to both real and simulated datasets and provide a comparison with other competing distributions based on some goodness-of-fit statistics.

5.1. Real-Life Datasets

The first data is about the total milk production in the first birth of 107 cows from the SINDI race. These cows are property of the Carnauba farm which belongs to the Agropecuaria Manoel Dantas Ltda (AMDA), located in Taperoa City, Paraiba (Brazil) [12]. The second data set was used by [13] and more recently by [14] which consist of n=50 observations on burr (in millimeter), with hole diameter and sheet thickness 12 mm and 3.15 mm respectively. The competing distributions are transmuted Topp Leon (TTL), Topp Leon Exponential (TLE), and Topp Leon (TL) distribution.

Table 4: *Goodness-of-fit statistics for dataset I*

Distribution	-LL	AIC	CAIC	HQIC	W*	A*	KS
TE-TL	-25.9741	-45.9481	-45.7151	-42.6975	0.1371	0.9035	0.0701
TTL	-22.3121	-40.6241	-40.5087	-38.4570	0.1542	1.0103	0.1086
TLE	-5.0388	-6.0775	-5.9621	-3.9104	0.7291	4.4011	0.1477
TL	-21.5262	-41.0524	-41.0143	-39.9689	0.2333	1.4792	0.0972

Table 5: *Likelihood ratio test statistic for Dataset I*

Distribution	Hypotheses	LRT	P-value
TE-TL vs TTL	$H_0 : \lambda = 1 \; vs \; H_1 : H_0 \; is \; false$	7.324	0.0068
TE-TL vs TLE	$H_0 : \theta = 0 \; vs \; H_1 : H_0 \; is \; false$	41.871	0.00001
TE-TL vs TL	$H_0 : \lambda = 1 \; and \; \theta = 0 \; vs \; H_1 : H_0 \; is \; false$	8.896	0.011

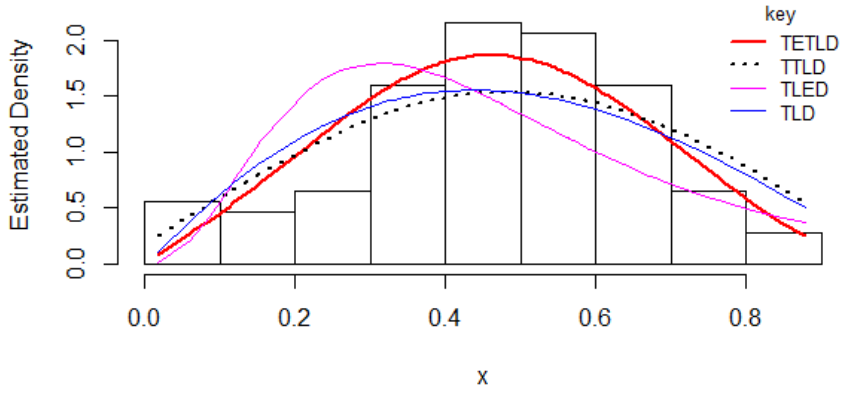


Figure 5: *The plot of the estimated densities for dataset I*

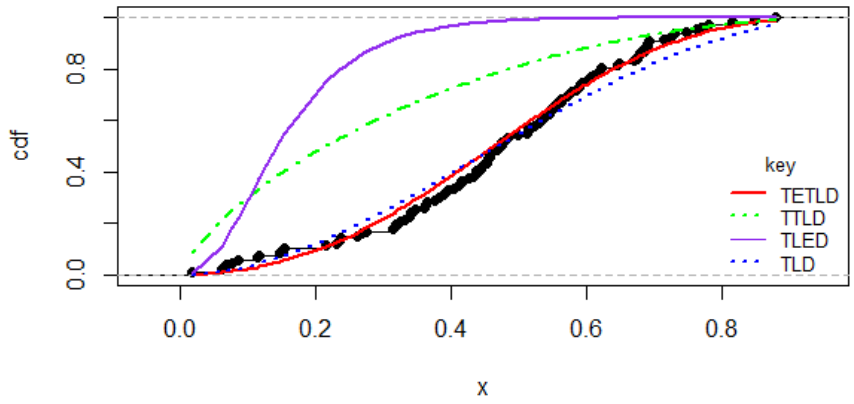


Figure 6: *The plot of the ecdf for dataset I*

The estimated densities and the ecdf for dataset I are respectively displayed in Figures 5 and 6 and the important aspect of these figures is to provide illustration and explanation on the flexibility of the competing distributions in the study in terms of capturing the sensitive parts of the dataset and draw a possible conclusion regarding the performance of the distributions. From Figures 5 and 6, we can observe that the TE-TLD shows a greater performance in capturing the sensitive part of the dataset as compared to other distributions in the study.

Table 6: *Goodness-of-fit statistics for dataset II*

Distribution	-LL	AIC	CAIC	HQIC	W*	A*	KS
TE-TL	-55.9914	-105.9827	-105.4610	-103.7984	0.1015	0.6246	0.1068
TTL	-28.3701	-52.7403	-52.4850	-51.2841	0.1653	0.9917	0.3676
TLE	-52.2863	-100.5725	-100.3172	-99.1163	0.2121	1.2634	0.1653
TL	-28.4078	-54.8156	-54.7323	-54.0875	0.1654	0.9919	0.3623

Table 7: *Likelihood ratio test statistic for Dataset II*

Distribution	Hypotheses	LRT	P-value
TE-TL vs TTL	$H_0 : \lambda = 1 \; vs \; H_1 : H_0 \; is \; false$	55.243	0.00001
TE-TL vs TLE	$H_0 : \theta = 0 \; vs \; H_1 : H_0 \; is \; false$	7.410	0.0065
TE-TL vs TL	$H_0 : \lambda = 1 \; and \; \theta = 0 \; vs \; H_1 : H_0 \; is \; false$	55.167	0.00001

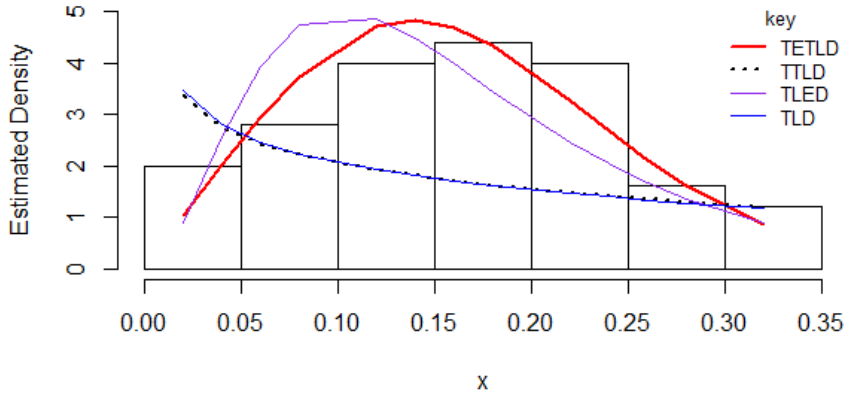


Figure 7: *The plot of the estimated densities for dataset II*

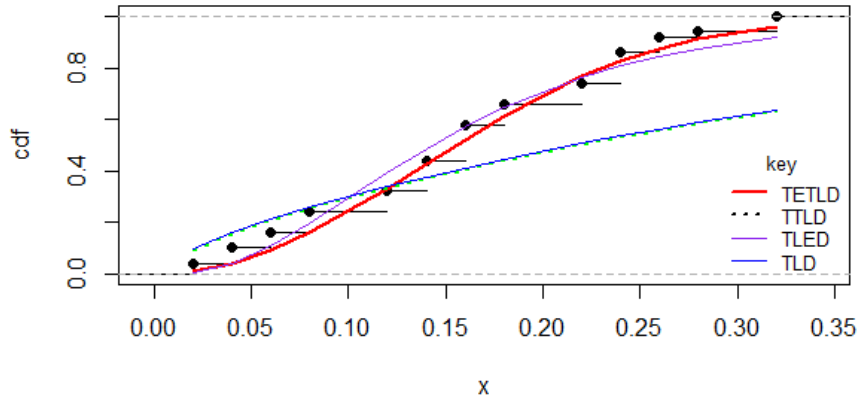


Figure 8: *The plot of the ecdf for dataset II*

The estimated densities and the ecdf for dataset II are respectively displayed in Figures 7 and 8 and the important aspect of these figures is to provide illustration and explanation on the flexibility of the competing distributions in the study in terms of capturing the sensitive parts of the dataset and draw a possible conclusion regarding the performance of the distributions. From Figures 7 and 8, we can observe that the TE-TLD shows a greater performance in capturing the sensitive part of the dataset as compared to other distributions in the study.

Some adequacy measures for the distributions are presented in Tables 4 and 6 for data sets I, II respectively. Hence, it is observed that the proposed distribution has the lowest values of the goodness-of-fit statistics and therefore, outperformed the other competing distributions in the study. Figures 5 and 6 are respectively displayed the estimated densities and the ecdf for datasets I, Figures 7 and 8 are respectively displayed the estimated densities and the ecdf for datasets II.

Likelihood ratio test is carried out to assess the significance of the additional parameter(s) of the TE-G family of distributions. For datasets I and II, since all the P-values are less than $\alpha = 0.05$, we therefore reject H_0 and conclude that the additional parameter(s) are significant which further prove the robustness of the TE-TLD over the nested distributions in the study.

6. Conclusion

In this paper, we proposed a new probability distribution by inducing Topp Leon distribution into transmuted Exponential-G family of distributions. The proposed distribution named transmuted Exponential- Topp Leon distribution in short (TE-TLD) which possessed different density shapes. Some properties of the distribution were presented in an explicit form and the parameters of the distribution are estimated by the method of maximum likelihood. The hazard function of the TE-TLD can be monotonic or non-monotonic failure rate which makes it more robust in terms of studying failure rates. The proposed distribution was found to be more robust as compared to other competing distributions with the same underlying baseline distribution when applied to real datasets in the study.

Disclosure statement
On behalf of The authors, I declare that no potential conflict of interest was reported.
Funding
No funding was provided for the work.

References

[1] Ghitany, M. E., Kotz, S., & Xie, M. (2005). On some reliability measures and their stochastic orderings for the Topp-Leone distribution. *Journal of Applied Statistics*, 32(7), 715-722.
[2] Nadarajah, S., & Kotz, S. (2006). The beta exponential distribution. *Reliability engineering & system safety*, 91(6), 689-697.
[3] Topp, C. W., & Leone, F. C. (1955). A family of J-shaped frequency functions. *Journal of the American Statistical Association*, 50(269), 209-219.
[4] Nadarajah, S., & Kotz, S. (2003). Moments of some J-shaped distributions. *Journal of Applied Statistics*, 30(3), 311-317.
[5] Kotz, S., & Seier, E. (2007). Kurtosis of the Topp-Leone distributions. *Interstat*, 1, 1-15.
[6] Zghoul, A. A. (2011). Record values from a family of J-shaped distributions. *Statistica*, 71, 355-365.
[7] Genc, A. I. (2012). Moments of order statistics of Top-Leone distribution. *Statistical Papers*, 53(1), 117-131.
[8] Al-Zahrani, B. (2012). Goodness-of-fit for the Topp-Leone distribution with unknown parameters. *Applied Mathematical Sciences*, 6(128), 6355-6363.
[9] Al-Shomrani, A., Arif, O., Shawky, A., Hanif, S., & Shahbaz, M. Q. (2016). Topp-Leone family of distributions: some properties and application. *Pakistan Journal of Statistics and Operation Research*, 12(3), 443-451.
[10] Shehla, R., & Khan, A. A. (2016). Reliability analysis using an exponential power model with bathtub-shaped failure rate function: A Bayes study. *SpringerPlus*, 5(1), 1-22.
[11] Mohammed, A. S., & Ugwuowo, F. I. (2020). A new family of distributions for generating skewed models: Properties and applications. *Pakistan Journal of Statistics*, 36(2), 149-168.
[12] Yousof, H. M., Alizadeh, M., Jahanshahi, S. M. A., Ramires, T. G., Ghosh, I., & Hamedani, G. G. (2017). The transmuted Topp-Leone G family of distributions: Theory, characterizations and applications. *Journal of Data Science*, 15(4), 723-740.
[13] Dasgupta, R. (2011). On the distribution of burr with applications. *Journal of Data Science*, 73(1), 1-19.
[14] ZeinEldin, R. A., Hashmi, S., Elsehety, M., & Elgarhy, M. (2020). Type II half logistic Kumaraswamy distribution with applications. *Journal of Function Spaces*, 1-15. https://doi.org/10.1155/2020/1343596

CURVATURE TENSORS IN SP-KENMOTSU MANIFOLDS WITH RESPECT TO QUARTER-SYMMETRIC METRIC CONNECTION

S. Sunitha Devi[1], T. Satyanarayana[2], K. L. Sai Prasad[3,*]

•

Department of Mathematics

[1] Vignan's Institute of Information Technology, Visakhapatnam, 530 049, INDIA
[2] Pragati Engineering College, Surampalem, Near Peddapuram, Andhra Pradesh, India
[3,*] Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam, 530 048, INDIA
sunithamallakula@yahoo.com[1] tsn9talluri@gmail.com[2] klsprasad@yahoo.com[3,*]

Abstract

A conformal curvature tensor and con-circular curvature tensor in an SP-Kenmotsu manifold are derived in this article which admits a quarter-symmetric metric connection. Conclusively, we verified our results by considering a case of 3-D SP-Kenmotsu manifold.

Keywords: η-Einstein manifold,SP-Kenmotsu manifold, con-circular curvature tensor, Quarter-symmetric metric connection, Ricci tensor, conformal curvature tensor.
2010 Mathematics Subject Classification: 53C07, 53C25

I. Introduction

A M_n (Riemannian manifold) is symmetrical locally if $\nabla.R = 0$ and symmetric if $R(X,Y).R = 0$ where $R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$ appears as a derivation. If $R(X,Y).R = 0$, then M_n is turns to be the pseudo symmetric space that is defined with the criteria $R.R = L(g,R)$. A manifold M_n is conformally symmetric if $\nabla.C = 0$ and if $R.C = 0$, it is said to be Weyl semi symmetric which are characterised by the condition $R.C = L_C Q(g,C)$.

Schouten & Friedman proposed the concept of semi-symmetric linear connection on a differentiable manifold. Some of the semi-symmetric curvature criteria in Riemannian manifolds are given by Yano [12].

Semi symmetric metric connection plays a very significant part in the geometry of Riemannian manifolds. For instance, a semi-symmetric metric is the displacement of the earth's surface after a fixed point. A quarter-symmetric connection is a linear connection $\widetilde{\nabla}$ on an n-dimensional Riemannian manifold (M_n, g) if $\widetilde{T}$ is $\widetilde{T}(X,Y) = \eta(Y)\phi X - \eta(X)\phi Y$.

Sato [8] proposed concepts of almost para contact Riemannian manifold. In 1977, Matsumoto and Adati [1] characterized special para-Sasakian as well as para-Sasakian manifolds as a particular type of almost contact Riemannian manifolds. Before Sato, Kenmotsu [6] characterized a type of this manifold. In 1995, Sinha and Sai Prasad [9] characterized a type of almost para contact metric manifolds mainly para-Kenmotsu and special para-Kenmotsu manifolds. For the literature, on Para-Kenmotsu manifolds one can refer to Balga [2], Srivastava and Srivastava [10], Olszak [7] .

On the other hand, various geometers of Riemannian manifolds and specifically, SP-Sasakian

manifolds were widely explored for the quarter-symmetric metric connections [3, 4, 5]. Inspired by these studies, in this work, we explore a class of special para-Kenmotsu manifolds that allowing the quarter-symmetric metric connection.

The current study is arranged as follows: Section 2 has certain prerequisites. In relation to the quarter symmetric metric connection in an *SP*-Kenmotsu manifold, we derive the equations for the Ricci tensor $\widetilde{S}$ & Riemannian curvature tensor $\widetilde{R}$ in Section 3. The equations in relation to quarter symmetric metric connection are also derived in an *SP*-Kenmotsu manifold M_n for concircular curvature tensor $\widetilde{Z}$ in Section 4. It is illustrated that the manifold M_n is η-Einstein given the concircular curvature tensor $\widetilde{Z}$ meets either of these conditions $\widetilde{R}(\xi, U).\widetilde{Z} = 0$, $\widetilde{Z}(\xi, U).\widetilde{R} = 0$, $\widetilde{Z}(\xi, U).\widetilde{Z} = 0$, $\widetilde{Z}(X, Y).\widetilde{S} = 0$. Section 5 is intended to define and analyse the curvature properties in the quarter-symmetric metric connection of the Weyl-conformal curvature tensor $\widetilde{C}$, of form $(0, 4)$, of *SP*-Kenmotsu manifold M_n. Finally, an illustration of a 3d *SP*-Kenmotsu manifold is considered in Section 6.

II. Preliminaries

Suppose M_n be an n-dimensional differentiable manifold provided with structure tensors (Φ, ξ, η) such that

$$\begin{aligned} &(a)\ \eta(\xi) = 1 \\ &(b)\ \Phi^2(X) = X - \eta(X)\xi;\ \overline{X} = \Phi X. \end{aligned} \tag{1}$$

M_n is called an almost para contact manifold.

Suppose that g be a Riemannian metric such that, for all vector fields X and Y on M_n

$$\begin{aligned} &(a)\ g(X, \xi) = \eta(X) \\ &(b)\ \Phi\xi = 0,\ \eta(\Phi X) = 0,\ \text{rank}\ \Phi = n - 1 \\ &(c)\ g(\Phi X, \Phi Y) = g(X, Y) - \eta(X)\eta(Y). \end{aligned} \tag{2}$$

Then it is stated that the manifold [8] M_n accepts an almost para contact structure of Riemannian (Φ, ξ, η, g).

Furthermore, if (Φ, ξ, η, g) fulfils the equations

$$\begin{aligned} &(a)\ (\nabla_X\eta)Y - (\nabla_Y\eta)X = 0; \\ &(b)\ (\nabla_X\nabla_Y\eta)Z = [-g(X, Z) + \eta(X)\eta(Z)]\eta(Y) + [-g(X, Y) + \eta(X)\eta(Y)]\eta(Z); \\ &(c)\ \nabla_X\xi = \Phi^2X = X - \eta(X)\xi; \\ &(d)\ (\nabla_X\Phi)Y = -g(X, \Phi Y)\xi - \eta(Y)\Phi X; \end{aligned} \tag{3}$$

then M_n is termed a para-Kenmotsu manifold or simply a *P*-Kenmotsu manifold [9].

A *P*-Kenmotsu manifold M_n permitting a 1-form η fulfilling

$$\begin{aligned} &(a)\ \ (\nabla_X\eta)Y = g(X, Y) - \eta(X)\eta(Y); \\ &(b)\ (\nabla_X\eta)Y = \varphi(\overline{X}, Y); \end{aligned} \tag{4}$$

here φ signifies Φ associate, is termed a special para-Kenmotsu manifold or shortly *SP*-Kenmotsu manifold [9].

Suppose (M_n, g) be an n-dimensional, $n \geq 3$, differentiable manifold of class C^∞ and let ∇

be its connection Levi-Civita. Then curvature tensor R of class (1, 3) of the the Riemannian Christoffel is provided by:

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z. \tag{5}$$

The (0,2)-tensor S^2 and the Ricci operator S are described as follows

$$g(SX,Y) = S(X,Y), \tag{6}$$

$$\text{and } S^2(X,Y) = S(SX,Y). \tag{7}$$

It is known [9] that the following relationship exist in the P-Kenmotsu manifold:

$$\begin{aligned} &(a)\ S(X,\xi) = -(n-1)\eta(X), \\ &(b)\ g[R(X,Y)Z,\xi] = \eta[R(X,Y,Z)] = g(X,Z)\eta(Y) - g(Y,Z)\eta(X), \\ &(c)\ R(\xi,X)Y = g(X,Y)\xi - \eta(Y)X, \\ &(d)\ R(X,Y)\xi = \eta(Y)X - \eta(X)Y;\ \text{when } X \text{ is orthogonal to } \xi. \end{aligned} \tag{8}$$

Almost para-contact Riemannian manifold M_n is termed to be η-Einstein and form of its Ricci tensor

$$S(X,Y) = a\ g(X,Y) + b\ \eta(X)\ \eta(Y) \tag{9}$$

Fields X and Y for any vector; a and b are a few scalars on M_n. In specific, if $b = 0$ thus M_n is considered to be an Einstein manifold.

III. Curvature tensor

A linear connection $\widetilde{\nabla}$ in a Riemannian manifold M_n is called a quarter-symmetric metric connection [4] if their torsion tensor $T(X,Y)$ meets

$$T(X,Y) = \eta(Y)\ \Phi X - \eta(X)\ \Phi Y, \tag{10}$$

and

$$(\widetilde{\nabla}_X g)(Y,Z) = 0; \tag{11}$$

where Φ iindicates a tensor field of the form (1, 1) and η is a 1-form.

A quarter-symmetric metric connection $\widetilde{\nabla}$ with torsion tensor (10) is given by

$$\widetilde{\nabla}_X Y = \nabla_X Y + \eta(Y)\ \Phi X\ - \varphi(X,Y)\xi \tag{12}$$

here, ∇ indicates Riemannian connection.

Suppose manifold M_n to be an SP-Kenmotsu manifold and $\Phi(X)$ as $\Phi X = \overline{X}$. Therefore the (10) and (11) may be represented as:

$$T(X,Y) = \eta(Y)\overline{X} - \eta(X)\overline{Y} \tag{13}$$

$$(\widetilde{\nabla}_X g)(Y,Z) = 0. \tag{14}$$

Let us choose the linear and Riemannian connection as $\widetilde{\nabla}$ and ∇, respectively

$$\widetilde{\nabla}_X Y = \nabla_X Y + U(X,Y), U \textit{ is a tensor of type } (1,2) \tag{15}$$

We have [12], for $\widetilde{\nabla}$ to be a quarter symmetric metric connection in M_n,

$$U(X,Y) = 1/2[T(X,Y) + T'(X,Y) + T'(Y,X)], \tag{16}$$

where

$$g(T'(X,Y),Z) = g(T(Z,X),Y)]. \tag{17}$$

Using (13) and (17), we get

$$T'(X,Y) = \eta(X)\overline{Y} - {}'F(X,Y)\xi; \tag{18}$$

here ${}'F(X,Y) = g(\overline{X},Y)$, η signifies a 1-form and ξ indicates the associated vector field.

From (13) and (16), in (18), we have

$$U(X,Y) = \eta(Y)\overline{X} - {}'F(X,Y)\xi, \tag{19}$$

and then (15) becomes

$$\widetilde{\nabla}_X Y = \nabla_X Y + \eta(Y)\overline{X} - {}'F(X,Y)\xi; \tag{20}$$

which indicates $\widetilde{\nabla}$ in an *SP*-Kenmotsu manifold.

Suppose $\widetilde{R}$ and R be the curvature tensors of the connections $\widetilde{\nabla}$ and ∇ correspondingly, we get

$$\widetilde{R}(X,Y)Z = \widetilde{\nabla}_X\widetilde{\nabla}_Y Z - \widetilde{\nabla}_Y\widetilde{\nabla}_X Z - \widetilde{\nabla}_{[}X,Y]Z \tag{21}$$

Using (20) and (5) in (21), we have

$$\widetilde{R}(X,Y)Z = R(X,Y)Z + g(Y,Z)X - g(X,Z)Y. \tag{22}$$

If we describe $\widetilde{R}(X,Y,Z,U)$ as $g(\widetilde{R}(X,Y)Z,U)$ and $R(X,Y,Z,U)$ as $g(R(X,Y)Z,U)$; then (22) becomes

$$\widetilde{R}(X,Y,Z,U) = R(X,Y,Z,U) + g(Y,Z)g(X,U) - g(X,Z)g(Y,U). \tag{23}$$

The above expression (23) denotes the relation between $\widetilde{R}(X,Y)Z$ of M_n w.r.t. $\widetilde{\nabla}$ and $R(X,Y)Z$ w..r.t. ∇.

Put X = U = e_i in (23), where e_i be an orthonormal basis of the tangent space at any point of the manifold and taking summation over i ($1 \leq i \leq$n), we get

$$\widetilde{S}(Y,Z) = S(Y,Z) + n\, g(Y,Z) - \eta(Y)\eta(Z); \tag{24}$$

here $\widetilde{S}$ and S signifies the Ricci tensors of $\widetilde{\nabla}$ and ∇.

From (24), by using Y = Z = e_i, we obtain

$$\widetilde{r} = r + n^2 - 1; \tag{25}$$

here $\widetilde{r}$ and r indicates the scalar curvatures of $\widetilde{\nabla}$ and ∇ correspondingly.

Theorem 3.1: Suppose that $\widetilde{S}$ be the Ricci tensor & $\widetilde{R}$ be the curvature tensor in an *SP*-Kenmotsu manifold M_n w.r.t. $\widetilde{\nabla}$, then

(a) $\widetilde{R}(X,Y)Z + \widetilde{R}(Y,Z)X + \widetilde{R}(Z,X)Y = 0$,

(b) $\widetilde{R}(X,Y,Z,U) + \widetilde{R}(X,Y,U,Z) = 0$,

(c) $\widetilde{R}(X,Y,Z,U) - \widetilde{R}(Z,U,X,Y) = 0$,

(d) $\widetilde{R}(X,Y,Z,\xi) = 2R(X,Y,Z,\xi)$,

(e) $\widetilde{S}(X,\xi) = 2S(X,\xi)$.

Proof: Using first Bianchi identity and eq.(22) w.r.t. the Riemannian connection, we obtain (a).

From eq. (23), we obtain (b) & (c). By putting U = ξ in (23) and by using (8) we have (d).

By using Y = Z = e_i in equation (d) as well as summation with i, we obtain (e).

Theorem 3.2: The Ricci tensor $\widetilde{S}$ in an *SP*-Kenmotsu manifold M_n w.r.t. the connection for the quarter-symmetric metric is symmetrical.

Proof: The theorem-proof is based on the eq. provided in (24).

IV. Concircular curvature tensor

The n-dimensional Riemannian manifold M_n is provided by the concircular curvature tensor $Z(X,Y)$ [11, 13]:

$$Z(X,Y)U = R(X,Y)U - \frac{r}{n(n-1)}[g(Y,U)X - g(X,U)Y] \tag{26}$$

for all X, Y, U $\in TM$.

The concircular curvature tensor w.r.t. $\widetilde{\nabla}$ in an *SP*-Kenmotsu manifold is $\widetilde{Z}$.

Therefore, using the equations (22) and (26), we get

$$\widetilde{Z}(X,Y)U = Z(X,Y)U - \frac{1}{n}[g(Y,U)X - g(X,U)Y], \tag{27}$$

which denotes the relation between the concircular curvature tensors w.r.t. $\widetilde{\nabla}$ and ∇.

Theorem 4.1: If $\widetilde{Z}$ w.r.t. $\widetilde{\nabla}$ in an *SP*-Kenmotsu manifold satisfies $\widetilde{R}(\xi,U).\widetilde{Z} = 0$, the manifold is η-Einstein.

Proof: Suppose $\widetilde{R}(\xi,U).\widetilde{Z}(X,Y)\xi = 0$, in an *SP*-Kenmotsu manifold.

Then

$$(\widetilde{R}(\xi,U).\widetilde{Z}(X,Y)\xi) - \widetilde{Z}(\widetilde{R}(\xi,U)X,Y)\xi - \widetilde{Z}(X,\widetilde{R}(\xi,U)Y)\xi - \widetilde{Z}(X,Y).\widetilde{R}(\xi,U)\xi = 0. \tag{28}$$

Also, from (8) and (22), we get

$$\widetilde{R}(X,Y)\xi = 2[\eta(Y)X - \eta(X)Y] \text{ and} \tag{29}$$

$$\widetilde{R}(\xi,X)U = 2[g(X,U)\xi - \eta(U)X]. \tag{30}$$

Then, by using (28), (29) and (30), we get

$$\widetilde{Z}(X,Y)U = 0. \tag{31}$$

Now, using the equations (26) and (27), the equation (31) reduces to

$$\mathbb{R}(X,Y,U) = \frac{r+n-1}{n(n-1)}[g(Y,U)X - g(X,U)Y]. \tag{32}$$

We obtain with the above equation w.r.t. X,

$$S(Y,U) = \frac{r+n-1}{n(n-1)}[ng(Y,U)X - \eta(Y)\eta(U)], \tag{33}$$

which on further contracting, we get

$$r = 1 - n^2. \tag{34}$$

Using (34), the expression (33) becomes

$$S(Y,U) = \eta(Y)\eta(U) - ng(Y,U); \tag{35}$$

which proves η-Einstein manifold.

Theorem 4.2: If $\widetilde{Z}$ with respect to $\widetilde{\nabla}$ in an *SP*-Kenmotsu manifold satisfies $\widetilde{Z}(\xi,U).\widetilde{R} = 0$, the manifold is an η-Einstein.

Proof: Suppose that $\widetilde{Z}(\xi,U).\widetilde{R}(X,Y)\xi = 0$, in an *SP*-Kenmotsu manifold.

Then

$$(\widetilde{Z}(\xi,U).\widetilde{R}(X,Y)\xi) - \widetilde{R}(\widetilde{Z}(\xi,U)X,Y)\xi - \widetilde{R}(X,\widetilde{Z}(\xi,U)Y)\xi - \widetilde{R}(X,Y).\widetilde{Z}(\xi,U)\xi = 0 \tag{36}$$

Also, from (8), (26) and (27), we have

$$\widetilde{Z}(\xi,U)Y = \Big[\frac{r}{n(n-1)} + \frac{1}{n} - 1\Big]\Big[g(U,Y)\xi - \eta(Y)U\Big] \tag{37}$$

and

$$\widetilde{Z}(X,Y)\xi = \Big[\frac{r}{n(n-1)} + \frac{1}{n} - 1\Big]\Big[\eta(X)Y - \eta(Y)X\Big]. \tag{38}$$

By substituting the values from (29), (30), (37) and (38) in the expression (36), we obtain

$$\widetilde{R}(X,Y)U = g(U,Y)X - g(U,X)Y + \eta(U)[1-\eta(X)]Y. \tag{39}$$

Using (22), the above eq. becomes

$$R(X,Y)U = \eta(U)[1-\eta(X)]Y; \tag{40}$$

and it proves.

Theorem 4.3: If the $\widetilde{Z}$ w.r.t. $\widetilde{\nabla}$ in an *SP*-Kenmotsu manifold meets $\widetilde{Z}(\xi,U).\widetilde{Z} = 0$, the manifold is η-Einstein.

Proof: The theorem-proof is trivial by the use of the the fact that $\widetilde{Z}(\xi,U).\widetilde{Z}$ indicates $\widetilde{Z}(\xi,U)$ was acting on $\widetilde{Z}$ as a derivation.

Theorem 4.4: If $\widetilde{Z}$ (concircular curvature tensor) with respect to $\widetilde{\nabla}$(quarter symmetric metric connection) in an *SP*-Kenmotsu manifold fulfills $\widetilde{Z}(X,Y).\widetilde{S} = 0$, the manifold signifies η-Einstein.

Proof: Let $\widetilde{Z}(X,Y).\widetilde{S}(U,V) = 0$ in an *SP*-Kenmotsu manifold.

Then it means

$$\widetilde{S}(\widetilde{Z}(X,Y)U,V) + \widetilde{S}(U,\widetilde{Z}(X,Y)V) = 0. \tag{41}$$

By choosing X = ξ in (41) and on using the equations (37) and (24), we obtain

$$\begin{aligned}\Big[\frac{r}{n(n-1)} + \frac{1}{n} - 1\Big]\Big[&-\eta(U)S(Y,V) - n\eta(U)g(Y,V) + 2\eta(U)\eta(V)\eta(Y)\\ &-\eta(V)S(U,Y) - n\eta(V)g(U,Y)\Big] = 0.\end{aligned} \tag{42}$$

Again by using $U = \xi$ in the eq. (42), we get

$$S(Y,V) = \eta(Y)\eta(V) - ng(Y,V); \tag{43}$$

which provides the required result.

V. Conformal curvature tensor

The Weyl conformal curvature tensor C of the type $(0,4)$ of a manifold M_n w.r.t. a Riemannian connection provided by [12, 13]:

$$\begin{aligned} C(X,Y,Z,U) = \; & R(X,Y,Z,U) - \frac{1}{n-2}[S(Y,Z)g(X,U) - S(X,Z)g(Y,U) \\ & + g(Y,Z)S(X,U) - g(X,Z)S(Y,U)] \\ & + \frac{r}{(n-1)(n-2)}[g(Y,Z)g(X,U) - g(X,Z)g(Y,U)]. \end{aligned} \tag{44}$$

Analogous to this, we define $\widetilde{C}$ i.e.Weyl conformal curvature tensor of the type $(0,4)$, of an SP-Kenmotsu manifold w.r.t. the quarter-symmetric metric connection as:

$$\begin{aligned} \widetilde{C}(X,Y,Z,U) = & \widetilde{R}(X,Y,Z,U) - \frac{1}{n-2}[\widetilde{S}(Y,Z)g(X,U) - \widetilde{S}(X,Z)g(Y,U) \\ & + g(Y,Z)\widetilde{S}(X,U) - g(X,Z)\widetilde{S}(Y,U)] \\ & + \frac{\widetilde{r}}{(n-1)(n-2)}[g(Y,Z)g(X,U) - g(X,Z)g(Y,U)]. \end{aligned} \tag{45}$$

Then, using the equations (23), (24), (25), (44) and (45), we get

$$\widetilde{C}(X,Y,Z,U) = C(X,Y,Z,U), \tag{46}$$

which implies the following statement:

Theorem 5.1: The conformal curvature tensors of $\widetilde{\nabla}$ and ∇ are equal in an SP-Kenmotsu manifold.

Suppose that $\widetilde{R} = 0$. Then $\widetilde{S} = 0$ and $\widetilde{r} = 0$.

From (45) we get that $\widetilde{C} = 0$ and hence using (46), we get $C = 0$.

Therefore, we provide the following theorem.

Theorem 5.2: The manifold is conformally flat in an SP-Kenmotsu manifold if the conformal curvature tensor $\widetilde{C}$ of $\widetilde{\nabla}$ vanishes.

Let $\widetilde{S} = 0$. Then $\widetilde{r} = 0$. Hence from (24) and (25), we get

$$S(Y,Z) = \eta(Y)\eta(Z) - n\, g(Y,Z) \tag{47}$$

and

$$r = 1 - n^2. \tag{48}$$

Then by using (23), (44), (47) and (48), we obtain

$$\widetilde{R}(X,Y,Z,U) = C(X,Y,Z,U). \tag{49}$$

From (49), we state that

Theorem 5.3: Conformal curvature tensor C of the manifold is identical in an SP-Kenmotsu manifold if $\widetilde{S}$ (Ricci tensor) of $\widetilde{\nabla}$ i.e quarter-symmetric metric connection vanishes, then $\widetilde{R}$ i.e. curvature tensor of $\widetilde{\nabla}$.

Using theorem (5.2) and (5.3), we state that

Theorem 5.4: If $\widetilde{S}$ of $\widetilde{\nabla}$ in an SP-Kenmotsu manifold disappears, then the manifold is conformally flat if $\widetilde{R}$ of $\widetilde{\nabla}$ vanishes.

VI. Example of a 3d SP-Kenmotsu manifold admitting the quarter-symmetric metric connection

Example 6.1: Suppose that 3d manifold $M = \{(x, y, u) \in R^3\}$, where (x, y, u) indicates "standard coordinates" in R^3. Considering e_1, e_2 & e_3 be fields of vector in M as

$$e_1 = e^{-u}\frac{\partial}{\partial x}, \;\; e_2 = e^{-u}\frac{\partial}{\partial y}, \;\; e_3 = \frac{\partial}{\partial u}. \tag{50}$$

for each point of M are linearly independent vectors and constitute a basis of $\chi(M)$.

Riemannian metric $g(X, Y)$ is

$$g(e_i, e_j) = \begin{cases} 1, & if \;\; i = j \\ \\ 0, & if \;\; i \neq j; \; i, \; j = 1, 2, 3, 4, 5. \end{cases}$$

$$Let \eta(Z) = g(Z, e_3), for \; any \; Z \; \in \; \chi(M)$$

Let η be a 1-form & (1, 1)-tensor field on M expressed by Φ defined as

$$\Phi^2(e_1) = e_1, \Phi^2(e_2) = e_2, \Phi^2(e_3) = 0.$$

The $g(X, Y)$ and linearity of Φ yields that

$$\begin{aligned} \eta(e_3) = 1, \; \Phi^2(X) = X - \eta(X)e_3 \; ; and \\ g(\Phi X, \Phi Y) = g(X, Y) - \eta(X)\eta(Y) \end{aligned}$$

for all vector fields $X, \; Y \; \in \; \chi(M)$.

Thus for $e_3 = \xi$, (Φ, ξ, η, g) describes an almost para-contact structure in M.

Let ∇ be a Riemannian connection in regard to the Riemannian metric g.

$$\left[e_1, e_2\right] = 0, [e_1, e_3] = e_1, [e_2, e_3] = e_2.$$

The formula of Koszul's is

$$\begin{aligned} 2g(\nabla_X Y, Z) = Xg(Y, Z) + Yg(Z, X) - Zg(X, Y) \\ - g(X, [Y, Z]) - g(Y, [X, Z]) + g(Z, [X, Y]). \end{aligned} \tag{51}$$

By taking $e_3 = \xi$ in (51), one can get

$$\begin{aligned} &\nabla_{e_1} e_1 = -e_3, \nabla_{e_1} e_2 = 0, \nabla_{e_1} e_3 = e_1; \\ &\nabla_{e_2} e_1 = 0, \nabla_{e_2} e_2 = -e_3, \nabla_{e_2} e_3 = e_2; \\ &\nabla_{e_3} e_1 = 0, \nabla_{e_3} e_2 = 0, \nabla_{e_3} e_3 = 0. \end{aligned}$$

Therefore manifold under consideration satisfies $\nabla_X \xi = \Phi^2 X = X - \eta(X)\xi$, $\eta(\xi)$ = 1 and the expression (3)d.

The above expressions satisfy all the properties of SP-Kenmotsu manifold with (Φ, ξ, η, g) . Thus $M(\Phi, \xi, \eta, g)$ is a 3-dimensional manifold.

Further from (20), we get

$$\begin{aligned} &\widetilde{\nabla}_{e_1} e_1 = -2e_3, \widetilde{\nabla}_{e_1} e_2 = 0, \widetilde{\nabla}_{e_1} e_3 = 2e_1; \\ &\widetilde{\nabla}_{e_2} e_1 = 0, \widetilde{\nabla}_{e_2} e_2 = -2e_3, \widetilde{\nabla}_{e_2} e_3 = 2e_2; \\ &\widetilde{\nabla}_{e_3} e_1 = 0, \widetilde{\nabla}_{e_3} e_2 = 0, \widetilde{\nabla}_{e_3} e_3 = 0; \end{aligned}$$

Therefore $T(X,Y)$ of $\widetilde{\nabla}$ can be expressed as:

$$T(e_i,e_i)=0,\ for\ i=1,2,3;\ and$$
$$T(e_1,e_2)=0, T(e_1,e_3)=e_1, T(e_2,e_3)=e_2.$$

Also, we get

$$(\widetilde{\nabla}_{e_1}g)(e_2,e_3)=0, (\widetilde{\nabla}_{e_2}g)(e_3,e_1)=0, (\widetilde{\nabla}_{e_3}g)(e_1,e_2)=0,$$

which proves that the manifold M under consideration admits $\widetilde{\nabla}$.

Thus it proves that M under consideration is an SP-Kenmotsu manifold and allows $\widetilde{\nabla}$.

Acknowledgements: The authors are grateful to Dr. B. Satyanarayana, Assistant Professor of Nagarjuna University for his important ideas in preparation of the article.

Declarations of interest: none

References

[1] Adati, T. and Matsumoto, K. On conformally recurrent and conformally symmetric P-Sasakian manifolds. TRU Math., 13 (1977), 25-32.

[2] Balga, A. M. η-Ricci solutions on para-Kenmotsu manifolds, Balkan Journal of Geometry and Its Applications, 20(1) (2015), 1-13.

[3] Biswas, S. C. and De, U. C. Quarter-symmetric metric connection in an sp-Sasakian manifold, Commun. Fac. Sci. Uni. Ank. Series AI, 46 (1997), 49-56.

[4] Golab, S. On semi-symmetric and quarter-symmetric linear connections, Tensor (N.S.), 29 (1975), 249-254.

[5] Kalpana and Priti Srivastava, Some curvature properties of Quarter-symmetric metric connection in an sp-Sasakian manifold, International Mathematical Forum, 50(5) (2010), 2477-2484.

[6] Kenmotsu, K. A class of almost contact Riemannian manifolds. Tohoku Math. Journal, 24 (1972), 93-103.

[7] Olszak, Z. The Schouten-van Kampen affine connection adapted to an almost (para) contact metric structure. Publications De L'Institut Mathematique, 94(108) (2013), 31-42.

[8] Sato, I. On a structure similar to the almost contact structure, Tensor (N.S.), 30 (1976), 219-224.

[9] Sinha, B. B. and Sai Prasad, K. L. A class of almost para contact metric Manifold. Bulletin of the Calcutta Mathematical Society, 87 (1995), 307-312.

[10] Srivastava, K. and Srivastava, S. K. On a class of α-para Kenmotsu manifolds, Mediterranean Journal of Mathematics, 13(1) (2016), 391-399.

[11] Yano, K. Concircular Geometry, I. Concircular transformations, Proc. Imp. Acad., Tokyo, 16 (1940), 195-200.

[12] Yano, K. On semi-symmetric metric connection, Revue Roumanine de Mathematiques Pures et Appliques, 15 (1970), 1579-1581.

[13] Yano, K. and Boschner, S. Curvature and Betti numbers, Annals of Mathematical Studies 32, Princeton University Press, 1953, pp. 187-191.

Twin-Piston Pressure Balance For Measurement And Uncertainty Evaluation Of Differential Pressure Digital Transducer

Chanchal[1], Renu Singh[1] , Deepika Garg[1]
•
[1]School of Engineering and Sciences
GD Goenka University, Gurugram (India)
Corresponding authors: renusingh@gdgoenka.ac.in[1]
deepika.garg@gdgoenka.ac.in[1]

Ajay Kumar[2]
•
[2]Department of Applied Sciences
Maharaja Surajmal Institute of Technology, New Delhi (India)

Abstract

Pressure measurement plays significant role in development of various instruments and in industry. Pressure measurement, its control and accuracy are always attraction of scientist. There are many devices for the pressure measurement like U-tube manometer, Bourdon tube/Dial gauge, Dead weight tester. The present study focused on the precise generation of differential pressures with static pressure range in 0 MPa to 50 MPa using twin pressure balance in hydraulic mode. The metrological characteristics of a differential pressure digital transducer were evaluated.

Keywords: Metrology, Uncertainty, Dead Weight Tester (DWT), Digital transducer, Twin pressure balance

I. Introduction

Pressure and its measurement are quite complex. A reliable instrument is required to measure pressure precisely and accurately [1, 2]. Depending on mode of measurement, there are different kind of pressure. Pressure which exists in air free space is known as absolute pressure or actual pressure at a point. When pressure exerted by fluid on the wall of the container with respect to the pressure of surrounding medium is gauge pressure. Example: Air plans, cars, weather instrumentation. Pressure which is measured related to atmospheric pressure known as differential pressure, Reference pressure may have any value except zero. When gauge, absolute and differential pressure are measured then they are said to be in gauge, absolute and differential mode respectively. For the calibration of devices and maintain the primary standard directly or from the basic fundamental units, pressure is derived from length mass and time. Now a days, there are many devices to measure pressure such as barometer, manometer, gauge, dead weight tester [3, 4]. Dead weight tester (DWT) has brought a revolutionary change in the calibration of devices. DWT is piston cylinder type primary standard measuring device. It is used to measure pressure generated by gas or liquid and for calibration of pressure gauge over a wide range of pressure.

The pressure measurement is in terms of fundamental unit, force and area. A piston is fitted within a cylinder. A force is applied in terms of mass in a gravitational field on piston and fluid under the piston get pressurized in equilibrium. It is generally used to calibrate pressure gauges, sensors, transmitters and transducers. On the basis of their applications, DWT is divided as hydraulic and pneumatic mode for the calibration of pressure instruments. In hydraulic mode, oil is used as fluid while in pneumatic mode air is used. It measures pressure nearly equal to 10,000 bars with accuracy of 250 ppm. DWT has many advantages such as simple in construction, easy to use, widely used for the calibration, testing and adjustment of huge range of pressure measurement instrument.

In the present study the combination of two dead weight tester (Twin pressure balance) is used for the measurement of differential pressure. Twin Pressure balance increases the pressure measurement and calibration range of instrument. Pressure with larger diameter create low pressure while with smaller diameter generate higher pressure. Thus, it provides two different cross-sectional area in single piston cylinder arrangement and hence it provide the flexibility to generate pressure in wide range (low to high) with single mass load. The calibration of DWT is accredited to international system of units through National metrology institutes (NMI) [5,6]. The NMI plays an important role for sustainability of existing devices and also for the development of new devices.

II. Working principle

Dead weight tester is based on the principle of Pascal's law. In an experiment, twin pressure balance (Model 55614, Desgranges & Huot, France) available at NPL, Delhi, India is used for calibration of test pressure gauge shown in figure 2. The whole arrangement consists of Oil reservoir, pipeline (through which oil flows), pressurization chamber and the gauge under test is fix on the top of pressurization Chamber. Sebacate oil is used as fluid in dead weight tester. The oil flows from the reservoir to the pressurization chamber and air is removed with the help of vacuum pump. The presence of air will create non-uniform pressure which results in inaccurate results. When the system is consisting of oil and is air free then the pressure gradually increases in pressurization chamber. The pressure in piston cylinder arrangement is balanced with an equal amount of force is exerted by the weights which is mounted on the cylinder. The sum of the pressure values mention on the weights is operated on the gauge which is under test and the corrections can be done by using small weights. The schematic diagram of experimental setup is shown in figure 1 [7–10].
The pressure (in Pa) generated by Dead weight tester is obtained by equation (1) [11,12].

$$P = \frac{\Sigma_i m_i .g\,(1-\rho_a/\rho_i) + \gamma C}{A_0\,(1+b_1 p_n + b_2 p_n^2)\,[\,(\alpha_c + \alpha_p)\,(T - T_r)]} \pm \Delta p \quad (1)$$

m_i=Mass of the weight, ϱ_a = Density of air at laboratory condition, ϱ_i = ith weight Density, γ = Surface tension of fluid, C = Circumference of the piston emerging out from the fluid ,A_0= Piston – cylinder's effective area at Zero pressure, α_c& α_p = Thermal expansion coefficients of cylinder's and piston's material ,T = Assembly temperature, T_r = Temperature at which A_0 is referred b= Effective area's pressure Coefficient , Δp = It is head correction in term of pressure (where $\Delta p = [(\varrho_f - \varrho_a)$ .g. H],In this equation, H depicts height difference between two dead weight tester's reference level and (ϱ_f) is transmitted fluid density.

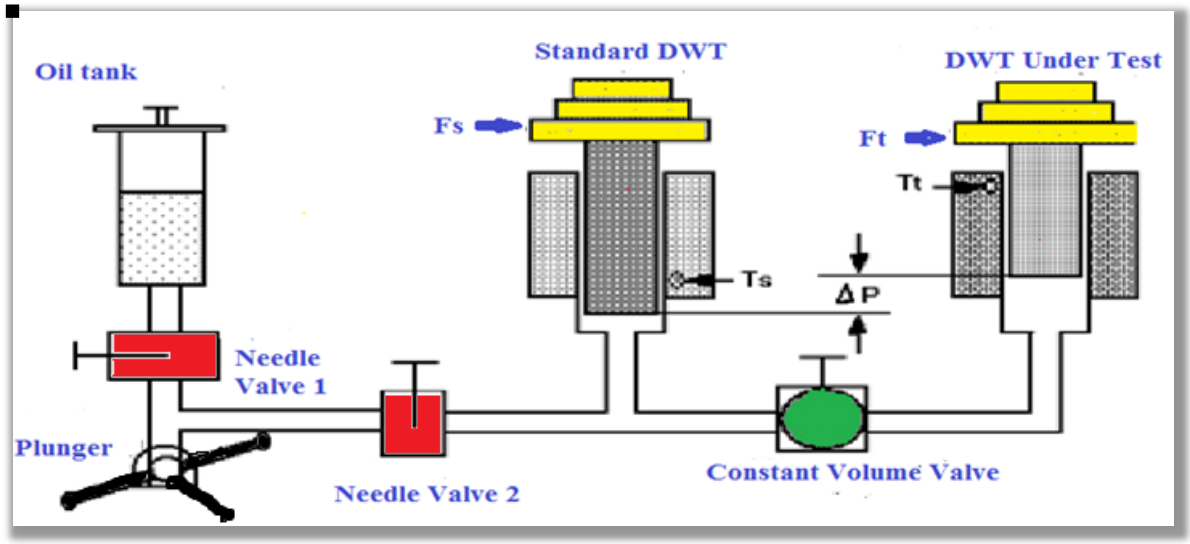


Figure 1: *Experimental setup for the calibration of DWT*

In the experiment the combination of two dead weight testers (Twin-Pressure balance) to calibrate the digital transducer. After connecting the digital gauge to the electric network, warm up time of 30 minutes is provided to it. Leakage testing is done by applying the maximum pressure (50 MPa) for 10 minutes before the experiment starts. The reading of differential pressure by the gauge is taken in increasing and decreasing order at different pressure points [13-16].

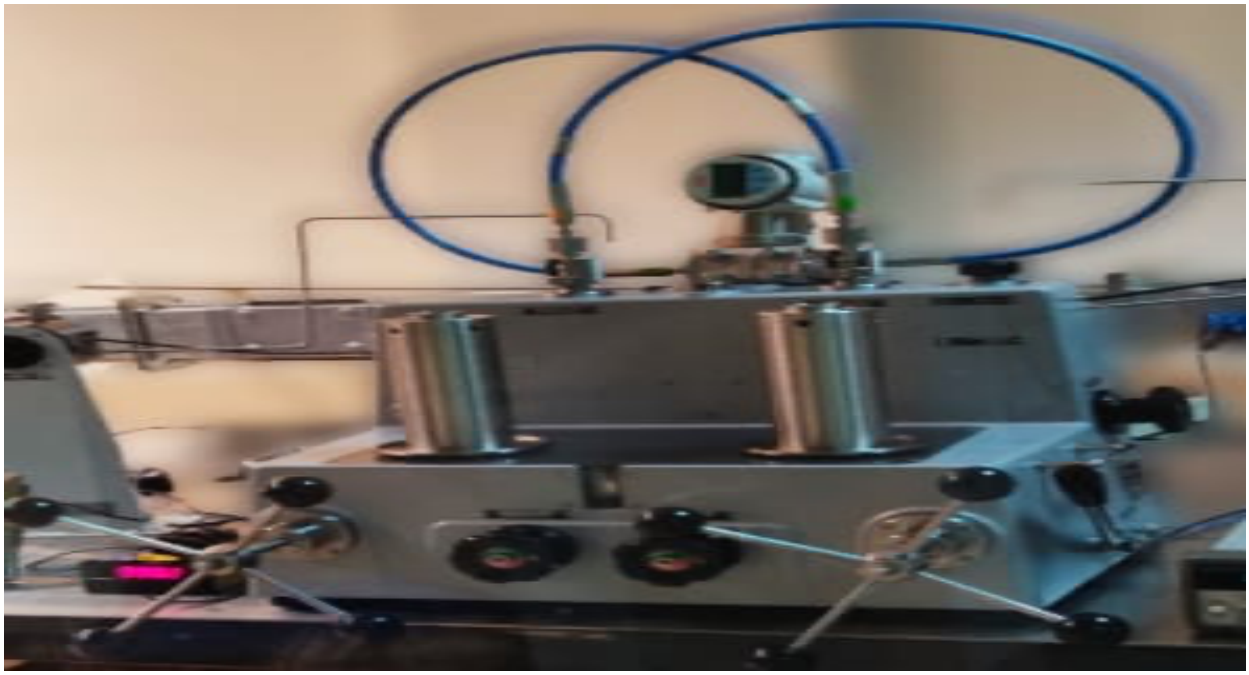


Figure 2: *Pictorial view of experimental setup for the calibration of digital transducer*

III. Calculation

Data is recorded for differential pressure against different values of static pressure and constant line pressure (10 MPa). The nominal differential pressure by twin pressure balance is nearly same as shown by digital transducer. The differential pressure output of twin pressure balance is given by equation

$$\Delta P_r = P_{\text{reference 1}} - P_{\text{reference 2}} \quad (2)$$

During the calibration of gauge by twin - pressure balance, the measurement uncertainty is established in accordance to "JCGM 100: 2008 - GUM 1995 with some small corrections - Guide to the expression of uncertainty in measurement –measured data evaluation - First edition September 2008".
The digital gauge error is evaluated by subtracting the differential pressure recorded by twin pressure balance and differential pressure shown by transducer. For digital gauge calibration, the error is given by the expression

$$E(P) = \Delta P_g - \Delta P_r \quad (3)$$

E (P) = Digital gauge error.
ΔP_g = magnitude value depicted by gauge., ΔP_r = magnitude value measured by twin pressure balance.
The error values obtained with the help of equation 3 are depicted in table 1 at different static pressure of 1, 30 and 49 MPa.in increasing and decreasing cycle.

Table.1: *Instrument errors*

ΔP (MPa)	Error (in MPa)		
	For Static Pr. 1 MPa	For Static Pr. 30 MPa	For Static Pr. 49 MPa
Increasing Cycle			
0	0.000108	0.001368	2.60E-05
0.5	0.000861	0.001742	7.34E-05
1	0.00021	0.003318	0.001719141
2	0.000829	0.004774	0.003363619
Decreasing Cycle			
2	0.000932	0.002379	0.003632561
1	0.000737	0.003219	0.001299314
0.5	0.000228	0.001984	0.000983811
0	0.000508	0.001241	5.46E-05

IV. Results and Discussion

The figure 3 shows the error as the function of pressure. From the figure concluded that the errors are contained within the interval 0.000025959 MPa-0.004774 MPa. This value shows the resolution of the digital gauge.

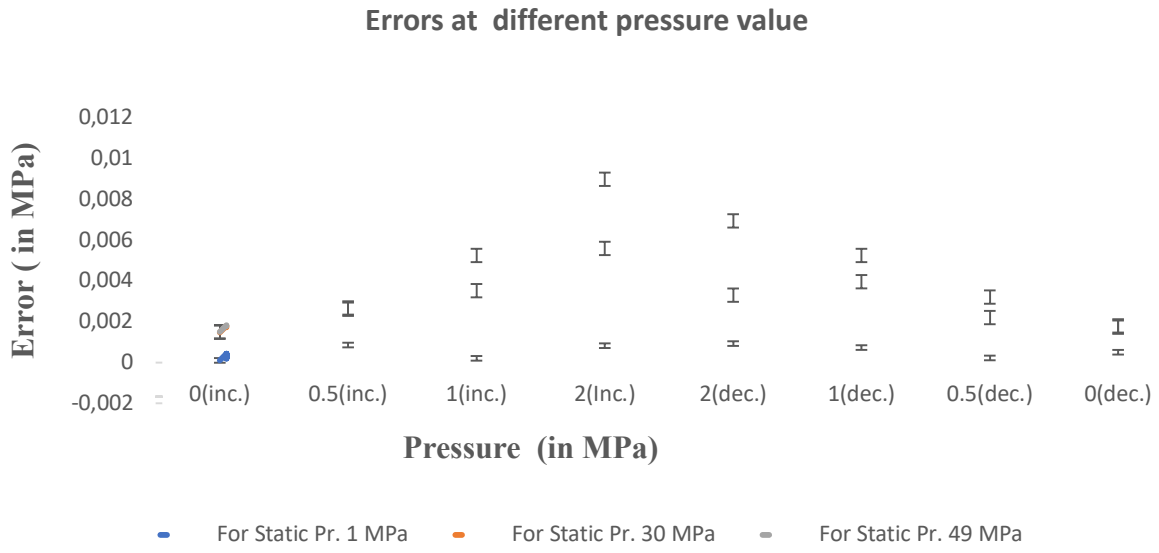


Figure 3: *Plot for digital gauge error*

The above graph between error and pressure plays an important role for the evaluation of calibration quality. The calibration is called as control calibration if the error lies within the minimum acceptable limit.

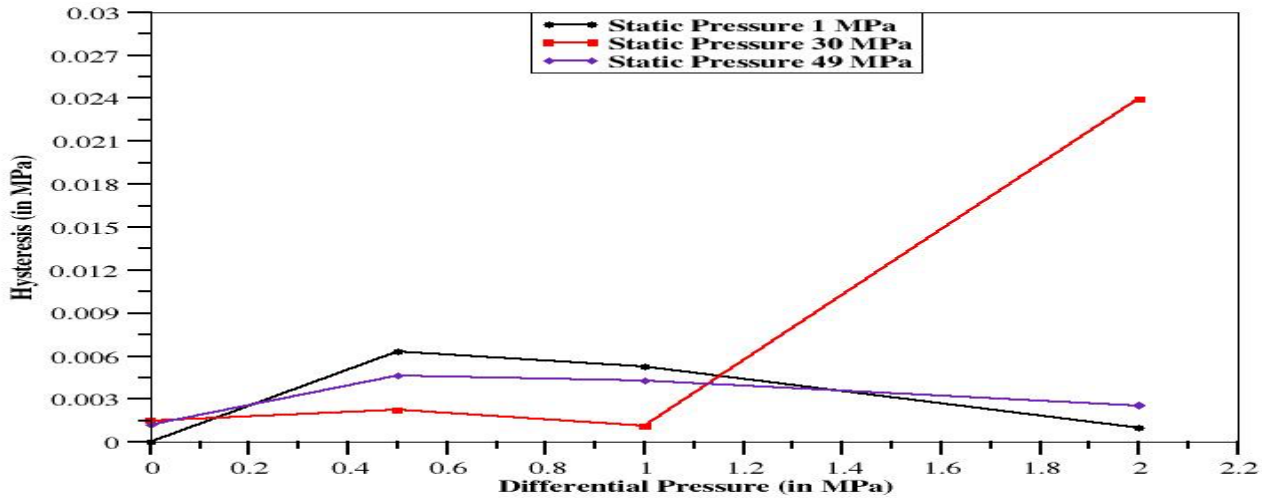


Figure 4: *Plot for the hysteresis of the transducer*

Hysteresis in the measurement is defined as the difference between corresponding values of pressure in increasing and decreasing orders in the pressure cycle. The hysteresis is plotted for the three static pressure points operating up to full range of 49 MPa (shown in Figure 4). For more precise measurements, the transducer may be used either in increasing or decreasing order of pressures. The maximum hysteresis error

is 0.02395 MPa at static pressure 30 MPa i.e. 0.079 % of the full scale which is very minimal in this pressure range.

Reproducibility defines as the closeness of results which is obtained by following the same procedures but under different experimental conditions.

Table 2 shows metrological characteristics of twin- pressure balance: maximum percentage error, hysteresis and reproducibility. The values in the table 2 are in the relation with the amplitude of the measuring range of digital gauge.

Table.2: *Maximum % error, hysteresis and reproducibility*

Maximum % error	0.019
Hysteresis %	0.079
Reproducibility %	0.0667

Table 3 shows the expanded uncertainty (U) of the transducer with the respective coverage factors 2 for the 95% confidence level. The value of uncertainty comes out to be the same for the different values of static pressure at the same value of pressure.

Table.3: *Expanded uncertainty*

Pressure (in MPa)	Coverage factor (k)	Uncertainty (in MPa)
0	2	9.67304E-06
0.5	2	1.10346E-05
1	2	1.22581E-05
2	2	1.42707E-05

V. Conclusion

- A new methodology to be applied for differential pressure measurement using twin pressure balances is proposed. Error values obtained by transducer are lies within the range 0.000025959 MPa - 0.004774 MPa. Which is quite small and shows the best results. The calibration uncertainty varied from 9.67304E-06 MPa to 1.42707E-05 MPa.
- During performance evaluation and calibration process, it is found that the hysteresis loss is very low i.e., 0.079 % of the full scale and reproducibility is also minimal 0.0667% of the full scale. Therefore, the transducer works well within reasonably good accuracy for high pressure range which is less than 1% of the full scale.
- This study concludes that the Twin -Pressure balance can be used as a Primary standard for differential pressure measurement.

VI. Acknowledgement

Present work is carried out at CSIR-NPL, India. Authors are thankful to CSIR-NPL team members, Dr. Sanjay Yadav and Afaqul Zafer for their help in the experimental work at various stages. Director, CSIR-NPL is also thanked for the permission to carry out this work.

References

[1] Dahiya, T.,Garg, D., Devi, S., & Kumar, R., (2021). Reliability Optimization Using Heuristic Algorithm in Pharmaceutical Plant. *Reliability: Theory & Applications, 16(3), 195-205*.

[2] Agarwal, A., Garg, D., Kumar, A., &Kumar, R. (2021). Performance Analysis of the Water Treatment Reverse Osmosis Plant. *Reliability: Theory & Applications*, 16(3), 16-25.

[3] Willink, R. (2013). *Measurement uncertainty and probability*. Cambridge University Press.

[4] Grossman, J. (2009). The Electronic Deadweight Tester--A Modern Replacement for the Conventional Deadweight Tester. *Cal Lab*, *16*(3), 36.

[5] Kobata, T., Kojima, M., Saitou, K., Fitzgerald, M., Jack, D., & Sutton, C. (2007). Final report on key comparison APMP. MP-K5 in differential pressure from 1 Pa to 5000 Pa. *Metrologia*, *44*(1A), 07001.

[6] Bean, V. E. (1994). *NIST pressure calibration service* (p. 98). US Department of Commerce, Technology Administration, National Institute of Standards and Technology.

[7] Yadav, S., Prakash, O., Gupta, V. K., & Bandyopadhyay, A. K. (2007). The effect of pressure-transmitting fluids in the characterization of a controlled clearance piston gauge up to 1 GPa. *Metrologia*, *44*(3), 222.0

[8] Woo, S. Y., Choi, I. M., & Song, H. W. (2009). A low differential pressure standard in the range of 1 Pa to 31 kPa at KRISS. *Metrologia*, *46*(1), 125.

[9] Dilawar, N., Varandani, D., Bandyopadhyay, A. K., & Gupta, A. C. (2003). Characterization of a pneumatic differential pressure transfer standard. *Metrologia*, *40*(2), 74.

[10] Chauhan, J., Vijayalakshmi, V., Muralidharan, V., & Sreedhar, S. (2020, February). Automation of Hydraulic Dead Weight Tester. In *2020 International Conference on Electrical and Electronics Engineering (ICE3)* (pp. 236-239). IEEE.

[11] Rosendahl, M., Nazareth, R. S., Magalhães, M. R., Silva, W. S., Ferreira, P. L. S., Gouveia, J. M., ... & Couto, P. R. G. (2018, June). New calibration procedure for differential pressure using twin pressure balances for flowrate measurement. In *Journal of Physics: Conference Series* (Vol. 1044, No. 1, p. 012053). IOP Publishing.

[12] Yadav, S., Gupta, V. K., & Bandyopadhyay, A. K. (2010). Standardization of pressure calibration (7-70 MPa) using digital pressure calibrator.

[13] Zafer, A., & Yadav, S. (2018). Design and development of strain gauge pressure transducer working in high pressure range of 500 MPa using autofrettage and finite element method. *International Journal of Precision Engineering and Manufacturing*, *19*(6), 793-800.

[14] Yadav S, Bandyopadhyay A K, Dilawar N and Gupta A C 2002 Intercomparison of national hydraulic pressure standards up to 500 M Pa *Measurement and Control* **35** 47–51

[15] Abdalla, M. E., Abdollaah, A. T., & Barakat, T. M. (2019). Pressure Measurement and Calibration Setup (TH2).

[16] Bich, W., Cox, M. G., & Harris, P. M. (2006). Evolution of the 'Guide to the Expression of Uncertainty in Measurement'. *Metrologia*, *43*(4), S161

Performance of a Single Server Batch Queueing Model with Second Optional Service under Transient and Steady State Domain

P. Vijaya Laxmi, Andwilile Abrahamu George and E. Girija Bhavani

•

Department of Applied Mathematics, Andhra University, Visakhapatnam, India.
Email: vijayalaxmiau@gmail.com, gandwilile@gmail.com, girijabhavaniedadasari@gmail.com

Abstract

The aim of this paper is to investigate the performance of a single server batch queueing model with second optional service under transient and steady state domain. It is assumed that the customers arrive in groups as per compound Poisson process and the server gives two types of services, First Essential Service (FES), which is mandatory for all arriving customers and Second Optional Service (SOS), which is given to some customers those who request it. Both FES and SOS are provided in batches of maximum b capacity. The transient and steady state probabilities of the model are obtained by using probability generating function and Laplace transform techniques. Finally, some numerical examples are presented to study the effect of the parameters on the system performance measures.

Keywords: Batch Queueing Model, First Essential Service, Second Optional Service, Transient State, Steady State

I. Introduction

In real-life situations, one encounter numerous examples of queueing models wherein a server gives FES to all arriving customers, and a few of them may only demand the auxiliary service after the completion of the essential service. For instance, all arriving ships at a harbor may need unloading service on arrival but only a few of them may demand re-loading service immediately after the unloading. The concept of SOS was first introduced by [8] where numerous practical applications of SOS were given. [8] presented an $M/G/1$ queue with SOS, whereby the service time distribution of the FES is general and the SOS is exponentially distributed. Later on, [9] generalized the concept of [8] in which the service time for both FES and SOS are independent having a general distribution. [16] studied the SOS in correlated reneging with working vacations. They use matric geometric method to obtain the steady state probabilities distribution of the queueing system size.

Queueing models with bulk input have broad applications in manufacturing, computer networks, communication systems, etc., where the arrivals at a service point (e.g., a switch) may occur in bunches of distinctive sizes. The notation of batch arrival appeared in the queueing theory in the work of [10] who considered the single server queue with fixed size batch Poisson arrivals in transient domain. Similar work of batch arrival has been carried out in [19]. They presented a bulk

input queueing model with single working vacation and they obtained the stationary queue length distribution using the matrix analysis method and probability generating function. [14] and [15] analyzed a bulk arrival queueing model with variant working vacations. The probability generating functions are derived in the stationary state and achieved the expressions of the model when the server is operating in various states. Related studies on the analysis of queueing model of bulk arrival are found in [3], [7], [12], [17], etc.,

Batch service queues have a motivation on numerous applications such as in group testing of blood samples for detecting corona/HIV viruses, in mobile crowd-sourcing app for smart cities, eliminate defective items in manufacturing system, etc. The batch service queueing models has been analyzed by many authors. [11] investigated the batch service queueing model with servers' variant vacations and obtained the steady state solutions using shifting operator and recursive technique. [6] discussed a single server queue with additional optional service in batches and server vacation. They have applied probability generating function method to obtain the queue length in stationary state. The analysis of bulk service queueing system with two heterogeneous servers in a discrete time has been presented in [5] with the help of displacement operator method and obtained closed form expressions for the limiting probabilities at arbitrary epoch.

In this model, we consider the transient state due to its importance especially in manufacturing system with regular beginning up periods and transportation frameworks with time fluctuating interest; for instance, airport terminal runway activities in major airports [4]. The analytical solutions of the transient behavior of queueing systems are very rare due to the complexity of getting analytical solutions. However, there are few works carried out in transient states such as [10], [2], [1], etc.

At the moment, most of the studies including [3], [7], [12], [13], [18] and many other are devoted to a single server batch queueing model with SOS in steady state, whereby customers arrive in groups as per Poisson process and served with general service distribution for both FES and SOS. However, in this paper, we consider a batch queueing model by involving the concept of SOS and investigated in both transient and steady state domains. We computed the probabilities and expected queue lengths when the server is busy in FES or SOS using probability generating function with the help of Laplace transform techniques. The advantage of expressions in Laplace transform is that it can be easily used for numerically transforming into time domain.

The remainder of this paper is structured as follows. In section 2, we present the model description and mathematical formulation. In section 3, we discuss the transient state equations and solving using probability generating function on the Laplace transforms equations. The steady state analysis is obtained by applying the Tauberian property in section 4. Measures of performance are discussed in section 5. Numerical analysis and discussions are presented in section 6 and in section 7, we conclude the paper.

II. Model Description and Mathematical Formula

We consider an $M^X/M^{[b]}/1$ queueing model with FES and SOS. Customers arrive in batches with rate $\lambda > 0$ conforming to a compound Poisson process. Let X be a batch size random variable and $X_1, X_2, \ldots$, are corresponding batch sizes of arriving customers which are independently and identically distributed (i.i.d.) random variables, with probability mass function $P\{X_i = k\} = C_k, k = 1, 2, 3, \ldots$. The service time distribution of both FES and SOS are exponential with rate μ_1 and μ_2, respectively and the services are given in batches of size not more than b such that if the server finds the customers less or equal to b in the waiting queue, the server takes all of them in the batch for service, but if the server finds the customers more than b waiting in the queue, then she or he takes a batch of size b while others remain waiting in the queue. The FES is required by all arriving

customers and after completing FES, they may opt SOS with probability r or may depart from the system with probability $1-r$. Figure 1 below shows the transition rate diagram of various transition states of the model.

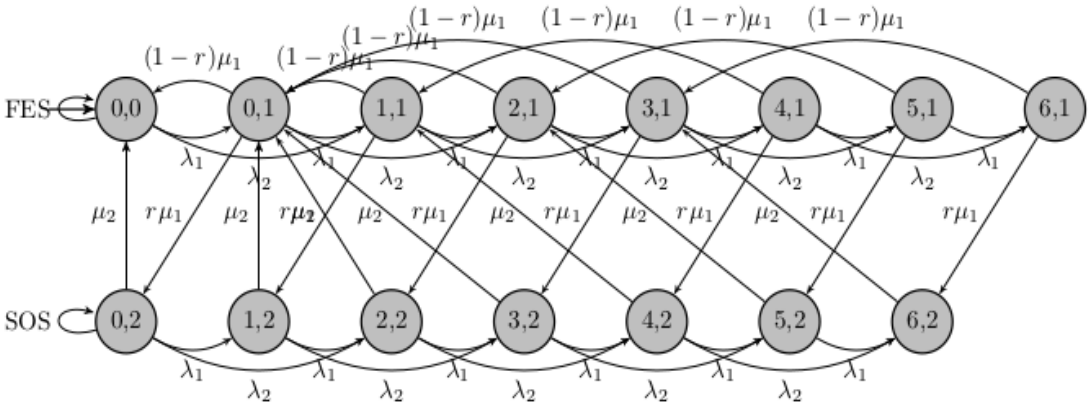


Figure 1: *Transition rate diagram for* $n = 6,\ b = 3,\ k = 2.$

I. Formulation of Mathematical Model

Suppose $L(t)$ be the length of the queue at time t, and $J(t)$ is the server state with

$$J(t) = \begin{cases} 1, & if\ the\ server\ is\ providing\ FES \\ 2, & if\ the\ server\ is\ providing\ SOS. \end{cases}$$

The stochastic process $\{(L(t),\ J(t));\ t \geq 0\}$ is a two-dimensional Markov Chain with the state space:

$$\Omega = \{(n, i);\ n \geq 0;\ i = 1, 2\}.$$

Further, let the transient probabilities are defined as

$$P_{n,i}(t) = Pr\{L(t) = n, \quad J(t) = i\};\ n \geq 0;\ i = 1, 2.$$

Here, $P_{n,i}(t)$ is the transient probability that there are n units in the queue at time t and the server is providing FES and SOS service, and $Q(t)$ is the probability when the queue is empty and the server is idle at time t. Using Markov theory, the differential-difference equations of the model are as follows:

$$Q'(t) = -\lambda Q(t) + (1-r)\mu_1 P_{0,1}(t) + \mu_2 P_{0,2}(t), \tag{1}$$

$$P'_{0,1}(t) = -(\lambda + \mu_1)P_{0,1}(t) + \lambda Q(t) + (1-r)\mu_1 \sum_{i=1}^{b} P_{i,1}(t) + \mu_2 \sum_{i=1}^{b} P_{i,2}(t), \tag{2}$$

$$P'_{n,1}(t) = -(\lambda + \mu_1)P_{n,1}(t) + \lambda \sum_{i=1}^{n} P_{n-k,1}(t)\, C_k + (1-r)\mu_1 P_{n+b,1}(t) + \mu_2 P_{n+b,2}(t), \quad n \geq 1, \tag{3}$$

$$P'_{0,2}(t) = -(\lambda + \mu_2)P_{0,2}(t) + r\mu_1 P_{0,1}(t), \tag{4}$$

$$P'_{n,2}(t) = -(\lambda + \mu_2)P_{n,2}(t) + \lambda \sum_{i=1}^{n} P_{n-k,2}(t)\, C_k + r\mu_1 P_{n,1}(t), \quad n \geq 1. \tag{5}$$

III. Transient Solution of the Model

In this section, the transient system size probability of the expected queue length when the server is idle and busy are presented by using Laplace transform (L.T) and probability generating functions. Let us assume that time is figured from the moment the server has taken a batch for

service, leaving none in the queue. i.e., $P_{0,1}(0) = 1$. Let $Q^*(s)$, $P^*_{n,i}(s)$ denote the L.T of $Q(t)$, $P_{n,i}(t)$, $i = 1, 2$, respectively. Taking L.T of equations from equation (1) to (5), we get

$$(s+\lambda)Q^*(s) = (1-r)\mu_1 P^*_{0,1}(s) + \mu_2 P^*_{0,2}(s), \tag{6}$$

$$(s+\lambda+\mu_1)P^*_{0,1}(s) = 1 + \lambda Q^*(s) + (1-r)\mu_1 \sum_{i=1}^{b} P^*_{i,1}(s) + \mu_2 \sum_{i=1}^{b} P^*_{i,2}(s), \tag{7}$$

$$(s+\lambda+\mu_1)P^*_{n,1}(s) = \lambda \sum_{i=1}^{n} P^*_{n-k,1}(s)C_k + (1-r)\mu_1 P^*_{n+b,1}(s) + \mu_2 P^*_{n+b,2}(s), \quad n \geq 1, \tag{8}$$

$$(s+\lambda+\mu_2)P^*_{0,2}(s) = r\mu_1 P^*_{0,1}(s), \tag{9}$$

$$(s+\lambda+\mu_2)P^*_{n,2}(s) = \lambda \sum_{i=1}^{n} P^*_{n-k,2}(s)C_k + r\mu_1 P^*_{n,1}(s), \quad n \geq 1. \tag{10}$$

Let us define the probability generating functions as:

$$P_1(s,z) = \sum_{n=0}^{\infty} P^*_{n,1}(s)z^n, \qquad P_2(s,z) = \sum_{n=0}^{\infty} P^*_{n,2}(s)z^n$$

The probability generating function of arrival batch size X is defined as:

$$C(z) = \sum_{k=1}^{n} C_k z^k \; ; \; |z| \leq 1; \; k = 1,2,3 \ldots \tag{11}$$

Multiplying equations (7) and (8) by z^n and taking summation from $n = 0$ to $n = \infty$ then, adding to (6) and after simplification, we have

$$P_1(s,z) = \frac{z^b(sQ^*(s)-1) + (1-z^b)A(s,z) - \mu_2 P_2(s,z)}{\lambda C(z)z^b - (s+\lambda+\mu_1)z^b + (1-r)\mu_1}, \tag{12}$$

where

$$A(s,z) = \left((1-r)\mu_1 \sum_{n=0}^{b-1} P^*_{n,1}(s)z^n + \mu_2 \sum_{n=0}^{b-1} P^*_{n,2}(s)z^n \right).$$

Similarly, from equation (9) and (10), we get

$$P_2(s,z) = \frac{-r\mu_1 P_1(s,z)}{\lambda C(z) - (s+\lambda+\mu_2)}. \tag{13}$$

Substituting equation (13) in (12), we obtain

$$P_1(s,z) = \frac{(\lambda C(z) - (s+\lambda+\mu_2))[z^b(sQ^*(s)-1) + (1-z^b)A(s,z)]}{(\lambda C(z))^2 z^b - \lambda C(z)(2s+2\lambda+\mu_1+\mu_2)z^b + B}, \tag{14}$$

where

$$B = (s+\lambda+\mu_1)(s+\lambda+\mu_2)z^b + \lambda C(z)(1-r)\mu_1 z - (s+\lambda+\mu_2)(1-r)\mu_1 - r\mu_1\mu_2.$$

We assume that arrival batch size X follows a geometric distribution with parameter q as given by.

$$P(X = k) = C_k = (1-q)^{k-1}q; \;\; 0 \le q \le 1; \quad k = 1,2,3 \ldots \tag{15}$$

Using (11) and (15), we obtain

$$C(\mathrm{z}) = \frac{qz}{1-z+qz}. \tag{16}$$

Substitute (16) into (14), we obtain

$$P_1(\mathrm{s},\mathrm{z}) = \frac{B_1(1-z+qz)[z^b(sQ^*(s)-1)+(1-z^b)A(s,z)]}{(\lambda q)^2 z^{b+2} - \lambda q(1-z+qz)(2s+2\lambda+\mu_1+\mu_2)z^{b+1}+B_2}, \tag{17}$$

where

$$B_1 = (\lambda qz - (s+\lambda+\mu_2)(1-z+qz),$$
$$B_2 = (1-z+qz)^2(s+\lambda+\mu_1)\,(s+\lambda+\mu_2)z^b + \lambda qz(1-z+qz)(1-r)\mu_1$$
$$-(1-z+qz)^2\,(s+\lambda+\mu_2)(1-r)\mu_1 - (1-z+qz)^2\, r\mu_1\mu_2.$$

We notice that the denominator of $P_1(s,z)$ has $b+2$ zeros. Using Rouche's theorem to the denominator, it follows that b of these roots lie on or inside the unit circle. One zero of the denominator is $z = 1$ and other $b-1$ zeros lie within and should harmonize with those of numerator for $P_1(s,z)$ to converge, so that when a zero shows up in the denominator, it is dropped by one in the numerator. The remaining two zeros of the denominator lie outside the unit circle. Let the roots be z_0 and z_1, we have

$$P_1(\mathrm{s},\mathrm{z}) = \frac{(1-z+qz)[\lambda qz - (s+\lambda+\mu_2)(1-z+qz)](1-z^b)D(s)}{(z-1)(z-z_0)(z-z_1)}, \tag{18}$$

where $D(s)$ is a function independent of z.
For $z = 1$ in (13) and using L'Hospital rule at $z = 1$ in (18), we get

$$P_1(\mathrm{s},1) = \frac{q^2(s+\mu_2)bD(s)}{(1-z_0)(1-z_1)}, \tag{19}$$

$$P_2(\mathrm{s},1) = \frac{r\mu_1 P_1(s,1)}{(s+\mu_2)}. \tag{20}$$

Using the normalization condition $P_1(s,1) + P_2(s,1) + Q^*(s) = \frac{1}{s}$, we have

$$P_1(\mathrm{s},1) = \frac{\big(1-sQ^*(s)\big)(s+\mu_2)}{s(s+r\mu_1+\mu_2)}. \tag{21}$$

Using (19) and (21) one can determine the function of $D(s)$ as

$$P_1(\mathrm{s},1) = \frac{\big(1-sQ^*(s)\big)(1-z_0)(1-z_1)}{s(s+r\mu_1+\mu_2)q^2b}. \tag{22}$$

Substitute (22) into (18), we get

$$P_1(\mathrm{s},\mathrm{z}) = \frac{(1-z+qz)B_1(1-z^b)\big(1-sQ^*(s)\big)(1-z_0)(1-z_1)}{s(s+r\mu_1+\mu_2)q^2b(z-1)(z-z_0)(z-z_1)}. \tag{23}$$

When $z = 0$, equation (23) and (13), respectively becomes

$$P_{0,1}^{*}(s) = \frac{\big(1 - sQ^{*}(s)\big)(s + \lambda + \mu_2)(r_0 - 1)(r_1 - 1)}{s(s + r\mu_1 + \mu_2)q^2 b}, \tag{24}$$

$$P_{0,2}^{*}(s) = \frac{r\mu_1\big(1 - sQ^{*}(s)\big)(r_0 - 1)(r_1 - 1)}{s(s + r\mu_1 + \mu_2)q^2 b}, \tag{25}$$

where $z_0 = 1/r_0$, $z_1 = 1/r_1$.
From equation (6), we can determine the value of $Q^{*}(s)$ by using (24) and (25), we have

$$Q^{*}(s) = \frac{B_3\,(s)}{s[(s + \lambda)(s + r\mu_1 + \mu_2)q^2 b + B_3\,(s)]}, \tag{26}$$

where

$$B_3(s) = [(1 - r)\mu_1(s + \lambda + \mu_2) + r\mu_1\mu_2](r_0 - 1)(r_1 - 1).$$

Equation (26) represents the L.T of the state probability that the queue is empty and the server is idle. It is obtained from the equation (6) by using the equations (24) and (25). In the following section, we obtain the stationary probabilities by using the Tauberian property.

IV. Steady State Solution of the Model

In this part, we obtain the closed form solutions of the limiting state probabilities for the length of the queue size when the server is idle or busy in FES and SOS by using the Tauberian property as defined below:

$$Q = \lim_{t\to\infty} Q(t) = \lim_{s\to 0} sQ^{*}(s), \tag{27}$$

$$P_{n,1} = \lim_{t\to\infty} P_{n,1}(t) = \lim_{s\to 0} sP_{n,1}^{*}(s), \tag{28}$$

$$P_{n,2} = \lim_{t\to\infty} P_{n,2}(t) = \lim_{s\to 0} sP_{n,2}^{*}(s). \tag{29}$$

If the limit exists, the steady state probabilities of (24), (25) and (26) are:

$$P_{0,1} = \frac{(1 - Q)(\lambda + \mu_2)(r_0 - 1)(r_1 - 1)}{(r\mu_1 + \mu_2)q^2 b}. \tag{30}$$

$$P_{0,2} = \frac{r\mu_1(1 - Q)(r_0 - 1)(r_1 - 1)}{(r\mu_1 + \mu_2)q^2 b}, \tag{31}$$

$$Q = \frac{B_3}{\lambda(r\mu_1 + \mu_2)q^2 b + B_3}, \tag{32}$$

where

$$B_3 = [(1 - r)\mu_1(\lambda + \mu_2) + r\mu_1\mu_2](r_0 - 1)(r_1 - 1).$$

V. Performance Measures

Practical applicability of any mathematical model can be accessed in terms of its measures of

performance. In this paper different execution measures of the queue are calculated such as probability that the server is active and the expected queue size when the server is active in FES or SOS. The performance measures are carried out in both transient and steady state as follows:

I. Performance Measures in Transient State

The busy probability in FES is given by:

$$P[FES](s) = \sum_{n=0}^{\infty} P_{n,1}^{*}(s).$$

The busy probability of the server in FES is obtained by setting $z = 1$ in equation (23) and applying L'Hospital rule, we get

$$P[FES](s) = \sum_{n=0}^{\infty} P_{n,1}^{*}(s) = \frac{\big(1 - sQ^{*}(s)\big)(s + \mu_2)}{s(s + r\mu_1 + \mu_2)}. \tag{33}$$

The busy probability in SOS is given by

$$P[SOS](s) = \sum_{n=0}^{\infty} P_{n,2}^{*}(s).$$

The busy probability in SOS is obtained by setting $z = 1$ in equation (13) and using (33), we get

$$P[SOS](s) = \sum_{n=0}^{\infty} P_{n,2}^{*}(s) = \frac{r\mu_1\big(1 - sQ^{*}(s)\big)}{s(s + r\mu_1 + \mu_2)}. \tag{34}$$

The anticipated length of the queue size when the server is busy in FES

$$L[FES](s) = \sum_{n=0}^{\infty} nP_{n,1}^{*}(s).$$

This is obtained by taking derivative of equation (23) with respect to z, setting $z = 1$ and using L'Hospital rule. Thus we get

$$\sum_{n=0}^{\infty} nP_{n,1}^{*}(s) = \frac{\big(1 - sQ^{*}(s)\big)\big[(r_0 - 1)(r_1 - 1)B_4(s) - [q(s + \mu_2)(4r_0r_1 - 2(r_0 + r_1))]\big]}{2q(s(s + r\mu_1 + \mu_2)(r_0 - 1)(r_1 - 1))}, \tag{35}$$

where $B_4(s) = \big[q(s + \mu_2)(b - 1) - 2[\lambda + (s + \mu_2)(2 - 2q)]\big]$.

The anticipated length of the queue size when the server is busy in SOS

$$L[SOS](s) = \sum_{n=0}^{\infty} nP_{n,2}^{*}(s).$$

This is obtained by taking derivative of equation (13) with respect to z and using (35) by setting $z = 1$, we get

$$\begin{aligned}\sum_{n=0}^{\infty} nP_{n,2}^{*}(s) = {} & \frac{r\mu_1\big(1 - sQ^{*}(s)\big)\big[(r_0 - 1)(r_1 - 1)B_4(s) - [q(s + \mu_2)(4r_0r_1 - 2(r_0 + r_1))]\big]}{2q(s(s + r\mu_1 + \mu_2)(s + \mu_2)(r_0 - 1)(r_1 - 1))} \\ & + \frac{\lambda r\mu_1\big(1 - sQ^{*}(s)\big)}{qs(s + \mu_2)(s + r\mu_1 + \mu_2)}.\end{aligned} \tag{36}$$

The overall queue length is

$$L_q(s) = \sum_{n=0}^{\infty} nP_{n,1}^*(s) + \sum_{n=0}^{\infty} nP_{n,2}^*(s). \tag{37}$$

The anticipated waiting time in the queue is

$$W_q(s) = \frac{q \times L_q(s)}{\lambda}. \tag{38}$$

II. Performance Measures in Steady State

Assuming that the limit of the equations (27), (28) and (29) exist, the steady state equations corresponding to the equations (33) to (38), respectively are given by

$$P[FES] = \sum_{n=0}^{\infty} P_{n,1} = \frac{(1-Q)\mu_2}{r\mu_1 + \mu_2},$$
$$P[SOS] = \sum_{n=0}^{\infty} P_{n,2} = \frac{r\mu_1(1-Q)}{(r\mu_1 + \mu_2)},$$

$$\sum_{n=0}^{\infty} nP_{n,1} = \frac{(1-Q)\left[(r_0-1)(r_1-1)B_4 - [q\mu_2(4r_0r_1 - 2(r_0+r_1))]\right]}{2q(r\mu_1+\mu_2)(r_0-1)(r_1-1)},$$
$$\sum_{n=0}^{\infty} nP_{n,2} = \frac{r\mu_1(1-Q)\left[(r_0-1)(r_1-1)B_4 - [q\mu_2(4r_0r_1 - 2(r_0+r_1))]\right]}{2q(r\mu_1+\mu_2)\mu_2(r_0-1)(r_1-1)} + \frac{\lambda r\mu_1(1-Q)}{q\mu_2(r\mu_1+\mu_2)},$$
$$L_q = \sum_{n=0}^{\infty} nP_{n,1} + \sum_{n=0}^{\infty} nP_{n,2}$$

and

$$W_q = \frac{q \times L_q}{\lambda},$$

where $B_4 = \left[q\mu_2(b-1) - 2[\lambda + \mu_2(2-2q)]\right]$.

VI. Numerical Investigation

In this part, we perform the transient and steady state numerical analysis of the model. In transient state, the Laplace transform expressions given in section 5.1 are inverted into time domain using a software package of Mathematica. Furthermore, we study the parameters impact on the model performance and discussion on numerical results by taking the model parameters as: $b = 5$, $\lambda = 3$, $\mu_1 = 3.5$, $\mu_2 = 3$, $r = 0.45$, $q = 0.4$, $r_0 = 0.9039$, and $r_1 = 0.7686$, unless their values are mentioned in the respective places.

Figures 2 and 3 show the time dependent probability of FES and SOS with variation of time points. We observe that the probability values in FES (Figure 2) decrease rapidly in the beginning from point one up to a certain value where it reaches the steady state with increasing of time while the

probability values in SOS (Figure 3) increase progressively from zero initially up to a certain value and it attains the steady state with increasing of time. In addition, it is noticed that the probabilities of both FES and SOS increase as the arrival rate λ increases. Figure 4 plots the transient state probability of empty queue and idle server versus time for different values of arrival rate. In this graph, we observe that the idleness probability decreases as the rate of arrivals increases.

Figure 5 demonstrates the variation of arrival rate λ on the expected queue size L_q with respect to time. It is noticed that expected queue size increases when arrival rate increases. This is due to the fact that when arrival rate increases, more customers join the queue and leads to an increase in the length of the queue. Figure 6 shows the impact of r on the expected waiting time in queue (W_q) and it is observed that as r increases, both $W[FES]$ and $W[SOS]$ increase. In addition, it reaches a point where the waiting time in SOS is more compared to FES as r increases. This is coherent with the fact that the service rate in FES is greater than that in SOS $i.e.$, $\mu_1 > \mu_2$.

Figures 7 and 8 show the effect of arrival rate λ on the expected queue size L_q for different batch size parameter q (Figure 7) and different batch service size b (Figure 8). It is obvious that the anticipated queue length increases with the increase in arrival rate λ (Figure 7). For a particular λ L_q increases as q decreases, this is on the grounds that the mean batch size $(1/q)$ positively influences the number of customers in the queue. Hence, the mean queue size increases. While in Figure 8 we observe that the expected queue size decreases with increase of batch service size b and increases with increasing of arrival rate λ.

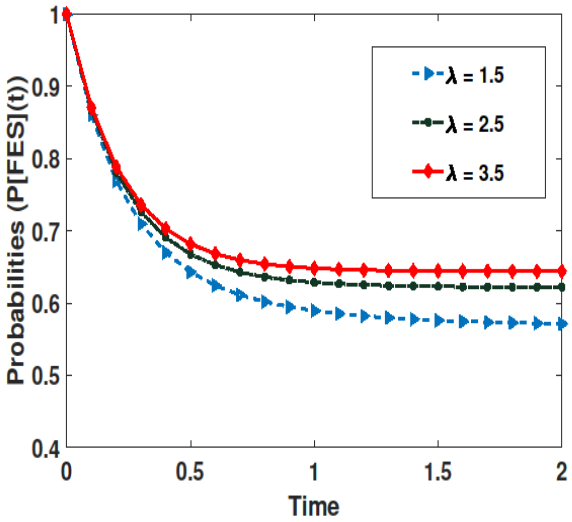


Figure 2: *The probability that the server is busy in FES versus time*

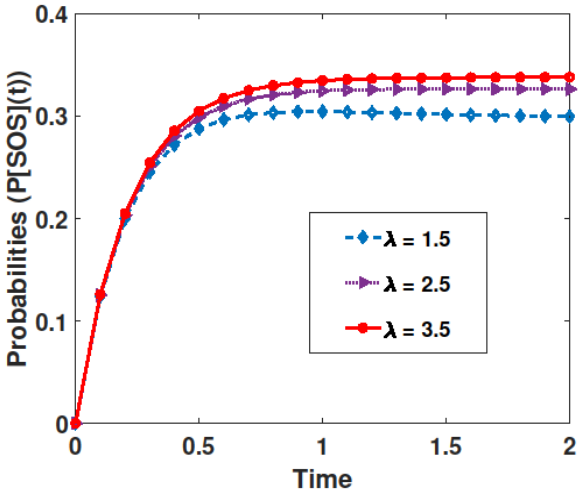


Figure 3: *The probability that the server is busy in SOS versus time*

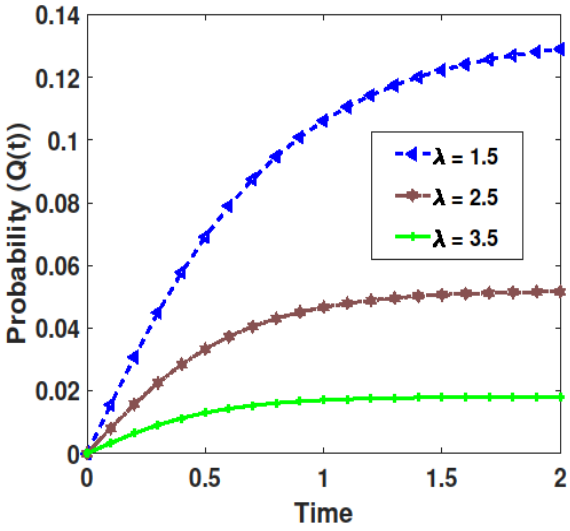


Figure 4: *The transient state probability of empty queue and idle server versus time*

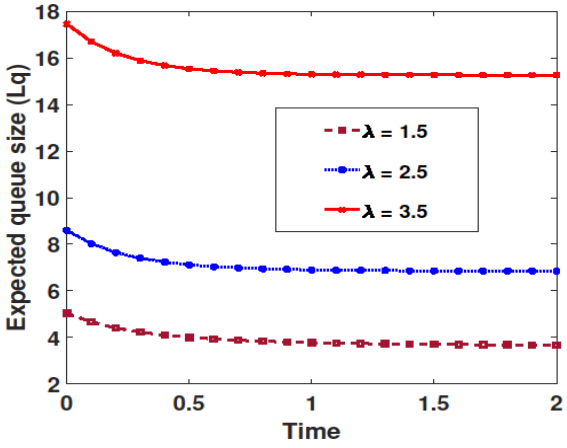


Figure 5: *Effect of variation of λ on L_q with respect to time*

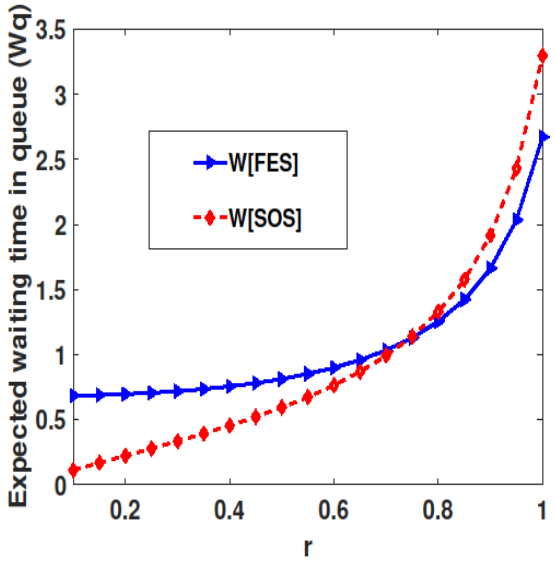


Figure 6: *Effect of r on the expected waiting time in queue (W_q)*

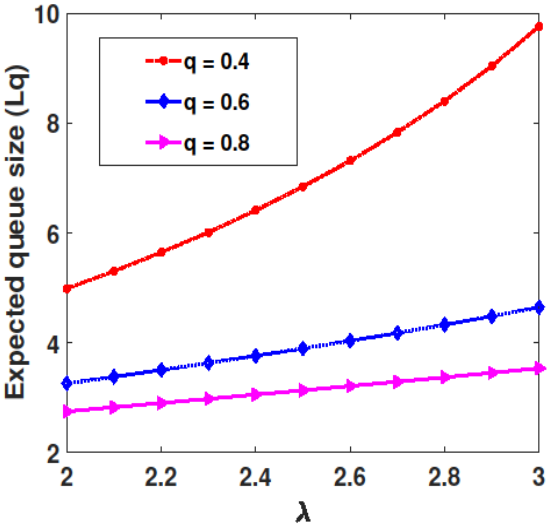


Figure 7: *Effect of variation of λ and q on the expected queue length (L_q)*

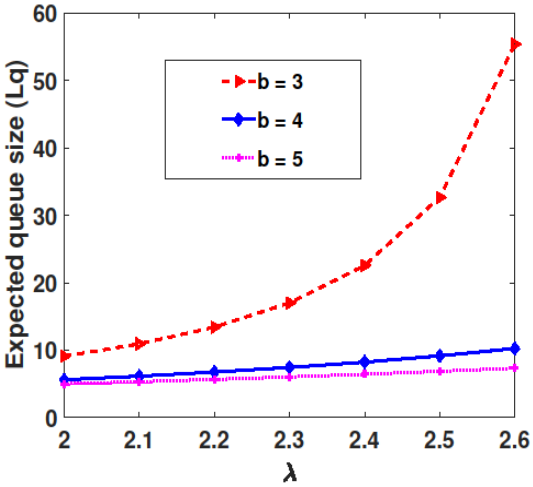


Figure 8: *Effect of variation in λ and b on the expected queue length (L_q)*

VII. Conclusion

In this article, we studied a single server batch queueing model with SOS under transient and steady state domain. We derived the transient and steady state probabilities when the server is busy in FES or SOS. Furthermore, we have studied the impact of various parameters on the performance measures of the model and discussed the results in the form of graphs. In addition, the analysis of the model will motivate a useful performance evaluation tool in practical applications such as telecommunication network through packet switching, in group testing of blood samples for detecting Corona / HIV viruses, package delivery, etc. Finally, the present work might be extended to multi-server multi-arrival system with reneging and vacations.

Acknowledgment

The authors would like to thank the editors and the referees for their valuable comments and suggestions which have helped in improving the quality and presentation of the paper. The authors would also like to thank the Department of Science and Technology, Government of India, for providing the Lab facility in the department under the DST-FIST Project grant No. SR/FST/MS-I/2017/3(c).

References

[1] Ayyappan, G., Devipriya, G. and Subramanian, A. M. G. (2013). Transient analysis of single server queueing system with batch service under catastrophe. *International Journal of Mathematical Archive,* 4(5): 26– 32.

[2] Chandrasekaran, V. M. and Saravanarajan, M. C. (2012). Transient and reliability analysis of M/M/1 feedback queue subject to catastrophes, server failures and repairs. *International Journal of Pure and Applied Mathematics,* 77(5): 605– 625.

[3] Chandrika, U. K. and Kalaiselvi, C. (2013). Batch arrival feedback queue with additional multi optional service and multiple vacation. *International Journal of Scientific Research Publications,* 3(3): 1 − 8.

[4] Dimitris, B. J. and Daisuke, N. (1992). Transient and busy period analysis of the G1/G/1 queue: The method of stages. *Queueing Systems,* 10: 153– 184.

[5] Goswami, V. and Samanta, S. K. (2009). Discrete time bulk service queue with two heterogeneous servers. *Computer and Industrial Engineering,* 56: 1348– 1356.

[6] Kalyanaraman, R. and Marugan, S. P. B. (2008). A single server queue with additional optional service in batches and server vacation. *Applied Mathematical Sciences*, 12(56): 2765– 2776.

[7] Kirupa, K. and Chandrika, K. U. (2015). Batch arrival retrial queue with negative customers, multi optional service and feedback. *Communications on Applied Electronics*, 2(4): 14– 18.

[8] Madan, K. C. (2000). An $M/G/1$ queue with second optional service. *Queueing Systems*, 34: 37– 46.

[9] Medhi, J. (2002). A single server Poisson input queue with a second optional channel. *Queueing Systems*, 42: 239– 242.

[10] Oduol, V. K and Ardil, C. (2012). Transient analysis of a single server queue with fixed size batch arrivals. *International Journal of Electrical and Computer Engineering,* 6(2): 253– 258.

[11] Sree Parimala, R. and Palaniammal, S. (2015). An analysis of bulk service queueing model with servers' various vacations. *International Journal of Advancements in Research and Technology*, 4(2): 22– 33.

[12] Suganya, S. (2014). A batch arrival feedback queue with M-optional service and multiple vacations subject to random breakdown. *International Journal of Science and Research,* 3(11): 1877– 1881.

[13] Uma, S. and Punniyamoorth, K. (2016). Single server bulk queue with second optional service, balking and compulsory vacation. *International Journal of Engineering Research and Application,* 6(10): 15– 20.

[14] Vijaya Laxmi P. and Rajesh, P. (2016). Analysis of variant working vacations on batch arrival queues. *Opsearch,* 53: 303– 316.

[15] Vijaya Laxmi P., Rajesh, P. and Kassahun, T. W. (2018). Performance measures of variant working vacations on batch queue with server breakdowns. *International Journal of Management Science and Engineering Management*, 14(1): 53– 63.

[16] Vijaya Laxmi P., Bhavani, E. G. and Rakesh, K. (2020). Correlated reneging in an optional service markovian queue with working vacations. *Reliability: Theory and Applications,* 15, 4(59): 102– 116.

[17] Vinnarasi, S., Maria Remona, J. and Julia Rose Mary K. (2016). Unreliable batch arrival queueing system with SWV. *International Journal of Innovative Research in Science, Engineering and Technology,* 5(3): 2884– 2889.

[18] Wang, J. and Li, J. (2010). Analysis of the $M^X/G/1$ queues with second multi-optional service and unreliable server. *Acta Mathematica Applicatae Sinica*, 26(3): 353– 368.

[19] Xu, X., Tian, N. and Zhang, Z. (2009). Analysis for the $M^X/M/1$ working vacation queue. *International Journal of Information and Management Sciences,* 20: 379– 394.

STRATIFIED REMAINDER LINEAR SYSTEMATIC SAMPLING BASED CLUSTERING MODEL FOR LOAN RISK DETECTION IN BIG DATA MINING

[1]Kamlesh Kumar Pandey, [2]Diwakar Shukla

•

[1,2]Dept. of Computer Science & Applications,
Dr. Hari singh Gour Vishwavidyalaya, Sagar, M.P., India
[1]kamleshamk@gmail.com,[2]diwakarshukla@rediffmail.com

Abstract

Nowadays, large volumes of data generate by numerous business organizations due to digital communications, web applications, social media, internet of things, cloud and mobile computing. Such has turned the nature of classical data into big data. Loan risk analysis is one of the most importance financial tasks, where financial organizations predict loan risk through customer financial history and behavioral data. Financial institutions face loan risk related issues when they make a loan to a bad customer. As a result, financial institutions divide loan applications into loan risk and non-risk clusters before making a loan for avoiding the loan risk challenges. Clustering approach is a data mining technique that uses data behavior and nature to discover the unexpected loan without any external information. Clustering algorithms face efficiency and effectiveness challenges as a result of the primary characteristics of big data. Sampling is of the data reduction technique that reduces computation time and improves cluster quality, scalability and speed of clustering algorithm. This study suggests a Stratified Remainder linear Systematic Sampling Extension (SRSE) approach for loan risk analysis in big data clustering using a single machine execution. The SRSE sampling plan enhances the effectiveness and efficiency of the clustering algorithm by employing maximum variance stratum formulation, remainder linear systematic sampling and extending sampling results into final result through centroid distance metric. The performance of the SRSE-based clustering algorithm has been compared to existing K-means and K-means++ algorithms using Davies Bouldin score, Silhouette coefficient, SD Validity, Ray-Turi index and CPU time validation metric on risk datasets.

Keywords: Loan Risk Clustering, Big Data Clustering, Stratified Sampling, Remainder linear Systematic Sampling, Sample Extension, K-means, SRSE-K-means, SRSE- K-means++.

I. Introduction

The volume of data has increased rapidly as a result of the development of the internet of things, cloud computing, web applications, communication technologies and social networks. Big data mining is analysis and process dealing the massive amounts of data for an organization's decision-making system [1]. The major characteristics of big data are volume (large scale of data), variety (various categories of data), and velocity (speed of data, stay motion). These three Vs are referred to as core features of big data, whereas the remaining Vs are referred to as supportable characteristics. Veracity (quality of processed data), variability (inconsistency of data), value (importance of data) and visualization (imagining the data) are other characteristics. Volume is a key attribute of big data and is represented in the scale of Terabytes and Petabytes. Variety handles a wide range of heterogeneous data sources, formats and their types. Velocity represents the rate of data creation, generation, delivery and updates in batch time, real-time and streaming across the

heterogeneous sources [2–4]. Veracity determines the quality, trustworthiness and accuracy of the data during the mining process because some heterogeneous sources generate inconsistent, incomplete, imprecise and ambiguous data [3, 4] . Value related to attributes (importance) of data for decision-making during the analysis process. It is described as valuable information on a massive volume and heterogeneous data that does not impair business decisions [2, 5] . Variability indicates the nature of data across the time and is fragments used in big data sentiment analysis. It refers to data whose structure, meaning, and behavior constantly change over time due to rapid data growth [5, 6]. Visualization pictures the raw and analyzed data as per user expectation and understandable in the form of figure or graphical presentation such as a table, graph, picture, chart and so on [7].

Big data mining is the discovery of knowledge, unknown correlations, actionable information and hidden patterns in big data sources useful for decision-making [2]. The objective of data mining is to predict the unknown insight and provide a description of predicate values that users easily can interpret. A data relation approach is another way of big data mining that identifies the relationship between attributes of a dataset. Big data mining research necessitates transparency because the large volume of data provides valuable knowledge, relationships and hidden patterns. The variety of data types and data sources leads to a diversification of mining results, and data velocity defines real-time mining [8]. Big data mining utilizes stability, high-efficiency, low computational cost and better risk management capability [9]. The combination of statistics and data mining techniques is known as intelligent big data mining and addresses the process and management challenges in mining framework [2]. Big data mining under risk reduction is classified as association rule learning, clustering, classification, and regression prediction.

Clustering is a technique used risk reduction for unsupervised predictive data mining that predicts class label based on homogeneity, similarity, or characteristics. Each risk cluster has a high degree of resemblance and a significant separation degree among them. The distance between data points with the shortest distance within-cluster is defined as having a high similarity within-cluster. A high separation results in the maximum distance between clusters. [10]. The application of clustering is in the fields of pattern recognition, image segmentation, artificial intelligence, wireless sensor networks, text analysis, bioinformatics, financial analysis, vector quantization and so on [11, 12].

Clustering is used in risk analysis applications such as supplier risk assessment [13], probabilistic risk assessment [14], project interdependent risk [15], financial risk analysis [16], insurance risk analysis, dynamic rockfall risk analysis [17], fall risk assessment [18] etc. Credit and debit risk concentrations are managed by banks and financial departments. The clustering technique allows customers to spend less time processing loan applications, and financial organizations predict loan risk in terms of good and bad customer for loan repayment. Borrowers' loan repayment capacity and loan risk are determined by their liabilities, reliance on family members, loans from other sources, individual age, increase in future income, etc. The identification of loan risk factors improves organizational safety and performance [14].

Kara et al. [13] assessed the 17 qualitative and quantitative supplier risks using the K-means clustering algorithm. The data points within the cluster indicate the specific risk, and their interpretation facilitates risk management and reduces supplier risk. It used the supplier risk-related dataset to identify the most reliable supplier by minimizing risk. Mandelli et al. [14] used principal component analysis and mean-shift methodology to identify similar behavioral risk events using the clustering algorithm for probabilistic risk assessment. Marle et al. [15] used interaction-based clustering to categorize the risks. The proposed methodology used the clustering objective for prioritization and resource allocation during risk grouping. Kou et al. [16] used real-life credit and bankruptcy risk datasets to evaluate a clustering algorithm based on multiple criteria decision making (MCDM) problems for financial risk analysis.

Fahad et al. [19] outlined the volume, variety, and velocity evolution criteria of the

conventional risk clustering approach for big data. The volume of the conventional clustering technique is recognized as the dataset size, high dimensionality and outlier detection. The variety is recognized as a dataset type and the clustering shape of the conventional clustering algorithm. The velocity is considered in the complexity and execution time of the conventional clustering algorithm. The existing risk clustering algorithms are unsuitable for big data mining due to these characteristics in terms of scalability, performance, quality and speedup. Volume is a dominant attribute of big data that reason data mining algorithms to pose storage and processing challenges. Data Volume necessitates a large amount of hardware and takes a long time to execute algorithms. The most common big data clustering methods are incremental, divide and conquer, data summarization, sampling, efficient nearest neighbor, dimension reduction, parallel computing, condensation, granular computing and so on [11, 20–22].

Nowadays, sampling and distributed/parallelization systems are two major strategies to solve big data mining-related issues. Sampling is a widely scientific method in the context of big data because it accurately reduces the data amount to a manageable size, increases scalability and speeds up algorithm execution with data processing [23, 24]. The execution of risk clustering is divided into single and multiple machines categories under big data mining, where single machine clustering use single machine resources and multiple machines used distributed execution. Parallel/distributed computation and data reduction are two common approaches to large-scale data clustering [22].

Sampling is a data reduction strategy that is useful for improving efficiency and performance when dealing with various types of problems related to data mining and database systems [25–27]. Sampling process minimizes data size and saves computation time and memory, while establishing a balance between the computational cost of high volume data and approximation results [24, 28]. The sampling-based data mining technique reduces the amount of data for mining and is known as an approximation approach [22]. It achieves approximate results within a specific time with query optimization for the decision support system. It is used in high-volume data applications such as risk analysis, database sampling, online aggregation, correlation discovery, stream-sampling, and so on [29, 30].

The analysis of big data necessitates the use of highly scalable clustering techniques. The computational complexity of the classical clustering algorithms is high on large scale data set that reason it cannot be straight applied to large-scale. The computational efficiency and cluster quality are the major challenges in the large scale data clustering. The objective of this study is to improve computational efficiency in terms of scalability, resources utilization, computational cost, and speed-up of big data clustering utilizing stratified remainder linear systematic sampling extension (SRSE) approach in the application of loan risk analysis on single machine execution. This study is organized into five sections. The second section examines sampling-based clustering algorithms and their applications in data mining. The third section introduces the stratified remainder linear systematic sampling extension approach and provides a sampling strategy for big data clustering. Section four contains the proposed work implementation using the K-means and K-means++ algorithms and provides as well as their validation on loan risk datasets using internal measures. The final section of the work wraps up the work and explores new possibilities. The final section of the work concludes the work and explores additional possibilities.

II. Literature review

This section presents sampling-based works on data mining based on existing research perspectives and investigates the advantages of stratified and systematic sampling over other sampling methods. Most of the data mining algorithms use uniform random sampling, systematic sampling, progressive sampling, stratified sampling and reservoir sampling. Uniform random sampling selects data from large data sets using a random number generator [31]. In systematic sampling, the first data point of the sample is selected in random order and the remaining sample

data points are selected at fixed intervals from the dataset [32]. Progressive sampling starts with a small sample size and gradually increases the sample size until a satisfactory performance measure is obtained [25]. Stratified sampling splits the dataset into homogeneous sample data, which is known as strata, then uses random sampling to collect samples from the strata for processing [22]. Reservoir sampling is used for data stream mining for both homogeneous and heterogeneous data sources [33].

Buddhakulsomsiri et al. [34] used a stratified random sampling approach in the application of health care systems to bill processing accuracy. The sampling plan used the rectangular method for strata construction to utilize sample resources, and measured the accuracy by percent and dollar accuracy. Silva et al. [35] proposed the CLUSMASTER (CLUStering on MASTER) algorithm through sampling for data streams in the application of sensor networks. The sampling procedure shortens the execution time and allocates fewer resources to the MASTER algorithm. The CLUSMASTER selects the best samples from each sensor in a network while minimizing the sum of square errors of the cluster. Rajasekaran et al. [36] proposed the DSC (Deterministic Sampling-based Clustering) algorithm for hierarchical and partitional clustering. The DSC algorithm improved the speed and accuracy as compared to the random sampling.

Jaiswal et al. [37] proposed a PTAS method based on D2-Sampling and K-means clustering. The PTAS shortened the time required for exhaustive search and optimized the objective function of clustering. Parker et al. [36] introduced geometric progressive fuzzy c-means (GOFCM) and minimum sample estimate random fuzzy c-means (MSERFCM) accelerated algorithms. Both clustering methods used novel stopping criteria and sampling for subsample size identification to speed up the initialization process. The GOFCM algorithm combines single-pass fuzzy c-means (SPFCM) and progressive sampling, whereas the MSERFCM algorithm combines random sampling and fuzzy c-means extension.

Xu et al. [38] proposed the Summation-bAsed Incremental Learning (SAIL) algorithm to avoid effectiveness and efficiency issues associated with text clustering on a large scale of text documents. The SAIL algorithm employs random sampling to address data scalability issues using an approximate approach. The use of random sampling significantly reduces computation costs and controls sampling error. Luchi et al. [39] use K-means to cluster a large data set using random sampling and a genetic approach. This approach guides better sample selection through genetic operations and reasonable computing time.

Jing et al. [40] combined a stratified sampling method and an ensemble clustering algorithm on a high dimensional dataset. The stratified sampling is used to generate the subspace component of the dataset. The proposed method achieves a better clustering structure and more accurate results than random sampling and random projection methods without sacrificing cluster diversity. Li et al. [41] proposed a Distributed Stratified Sampling approach for big data. The stratified sampling extracts the subsample size from each partition of the data distribution in parallel order. The DSS algorithm achieved higher sample representativeness, accuracy, scalability, and efficiency with low data-transmission costs than state-of-the-art methods.

Zhan et al. [42] solved eigenfunction problems for spectral clustering algorithms in image segmentation applications using the Nyström sampling method. The Nyström technique is used to reduce the time and space complexity. The proposed method is effective for solving high-resolution image-related problems such as high dimensionality, small sample sizes, feasibility, and overall clustering solution. Aloise et al. [43] used an iterative sampling algorithm to solve the strongly NP-hard minimax diameter clustering problem (MMDCP). The proposed algorithm used the heuristic procedure to select the optimal solution across the sample.

Reddy et al. [44] proposed an optimal stratification design for data mining algorithms using Weibull-distributed auxiliary information in the context of a health population. The auxiliary information is used for strata construction in the absence of study variables. This study states that the combination of data mining and a well-designed sampling plan enhances the accuracy of mining results. Sainil et al. [45] compared the performance of stratified random sampling and

stratified ranked set sampling in terms of bias and mean square error. These evaluations show that stratified ranked set sampling is more efficient than stratified random sampling.

Li et al. [46] dveloped the clustering ensemble algorithm through sample stability, which divided the dataset into cluster core and cluster halo for the underlying cluster structure of the data set. The cluster core discovers the cluster structure through samples, and the cluster halo assigns the sample data into cluster construction. Zhao et al. [22] proposed the Stratified Sampling plus Extension FCM (abbr. SSEFCM) algorithm for large-scale datasets by combining stratified sampling and fuzzy c-means clustering. The SSEFCM improves computational efficiency and cluster quality while diminishing computational complexity.

Goshu et al. [47] proposed the Systematic Sampling Evolutionary (SSE) method, which is a derivative-free meta-heuristic type algorithm that combines a systematic sampling procedure and nature-inspired particle swarm optimization algorithm. Systematic sampling is used to determine the leader decision of the evolutionary algorithm, which searches for the action decision at each iteration. Prasad et al. [48] address the solution of the bigVAT algorithm through sampling and crisp partitions. The bigVAT is used for cluster tendency detection of big data clusters using the K-means algorithm on synthetic and real-life datasets. The sampling process selects a sample from inter-cluster data objects, and the crisp partitions technique predicts the cluster labels of sample objects.

Nguyen et al. [49] proposed the S-VOILA (Streaming Variance OptImaL Allocation) algorithm for streaming and non-streaming data using stratified random sampling and mini-batch processing. The S-VOILA algorithm reduces the variance of sample data through locally variance-optimal allocation and maintains the stratum via weighted sampling.

Larson et al. [50] investigated systematic and random sample designs and discovered that systematic sampling outperforms random sampling in terms of variance estimator, sample size, and data range. Stratified sampling outperforms simple random sampling in terms of statistical precision and sampling error. To achieve better accuracy, performance, and computing resource utilization, stratified sampling used a smaller sample size than random sampling [24]. According to the literature [41], stratified sampling can achieve higher statistical precision and improve representativeness by reducing sampling error than simple random sampling, because variability within subgroups with similar properties is lower than that of the entire population. Stratified sampling also extracts better samples from the dataset in terms of size and representativeness, which saves time and costs associated with the data processing algorithm.

The literature [32] states that sampled data from systematic sampling is more accurate and has spatial autocorrelation than random sampling. The results of the experiments [32] show that systematic sampling has variance-related issues that can be resolved by combining systematic and stratified sampling because each stratum has an optimal variance sample. The results of a comparison of uniform random sampling, progressive sampling, biased sampling, and stratified sampling show that stratified sampling achieves higher computational efficiency and quality for the clustering process [22].

III. Proposed Work

The practical approach of the sample plan for clustering across several domains is determined by existing research [22, 51, 52] and literature. The stratification technique reduces sample variance, whereas clustering reduces variance within a cluster. As a result, combining stratification and clustering improves the effectiveness and efficiency of clustering algorithm. Uniform random sampling is entirely dependent on sampling design, data structure, and sampling strategy. The random sampling does not cover the entire dataset; therefore the sample representativeness quality is reduced. To avoid this issue, systematic sampling is preferable because it sample data covers the entire dataset. This section describes the clustering objective, sampling contents, and presents the stratified remainder linear systematic sampling extension approach (SRSE) for loan

risk clustering on big data mining using single machine execution. The proposed method reduces computation costs and improves computational efficiency while maintaining cluster quality during risk clustering.

I. Objective function for loan risk clustering

Let the X loan risk based dataset $N = \{x_1, x_2, \ldots\ldots x_N\}$ to be clustered C into $K = \{C_1, C_2, \ldots\ldots C_K\}$ on the basis of predefined similarity function in d dimension space of loan risk attribute set. The considered clustering approach minimizes the within-cluster Sum of Squared Error (WSSE) and maximizes the between-cluster Sum of Squared Error (BSSE). The objective criterion defined described in Eq. 1 [20].

$$WSSE\ (X, C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (1)$$

where x_i is the data point and μ_K is the centroid of C_K cluster. The content of C_K to the minimum SSE problem is defined by as under [53].

$$C_k = \left\{ x_i \in X | k = \underset{j \in \{1,2,\ldots.K\}}{\arg\min} \left\| x_i - \mu_j \right\|^2 \right\} \quad (2)$$

$$\mu_k = \frac{\sum_{x_i \in C_k} x_i}{|C_i|} \quad (3)$$

II. Sampling content

The presented clustering approach uses the stratification, remainder linear systematic sampling and sample extension process for loan risk group detection.

A. Stratification

The sampling-frame is divided into non-overlapping strata in stratified sampling according to data behaviors, types, location, attributes, variance, correction, regression, characteristics, format and so on. The strata are internally homogeneous with respect to the study variable that maximizes the precision of sampling results. Stratified sampling divides the N heterogeneous data points of loan risk dataset into $L = \{S_1, S_2, \ldots\ldots S_L\}$ homogeneous strata, where each stratum h consists of N_h data units and used the $\{S_1 \cup S_2 \cup \ldots\ldots \cup S_h\} = N$ and $\{S_1 \cap S_2 \cap \ldots\ldots \cap S_h\} = \theta$ conditions, where $h = 1,2,3, \ldots\ldots L$ and $\sum_{h=1}^{L} N_h = N = \{x_1, x_2, \ldots\ldots x_N\}$ [54, 55]. The stratum is derived from the loan risk data set by the stratification process. The maximum variance attribute and ascending sorting heuristics have used in this study to employ novel stratification methods. Algorithm 1 and Figure 1 illustrate the conceptual stratification representation.

This study used a remainder linear systematic sampling approach; therefore the stratification process formed the dataset into two strata. The stratification process first extracts the study variable based on maximum variance and then arranges the entire loan risk dataset based on the selected variable. The remainder linear systematic sampling method is used to determine the number of data points in strata.

B. Remainder Linear Systematic Sampling

Chang et al. proposed the Remainder Linear Systematic Sampling (RLSS) method to overcome the limitations of linear systematic sampling [56] in terms of *N≠nl* and linear sample size. Where N is

the size of the dataset, n is the sample size and l is the sample interval. The RLSS resolved the systematic sampling issue through a combination of stratification and linear systematic sampling. The number of data points in the loan risk dataset is represented in the RLSS approach by $N=nl+r$, $0\leq r\leq n$. The RLSS approach is more efficient when the sample size is not a multiple of the dataset size and in $N\neq nl$ situations. The n, l and r are the integer numbers, and r is the reminder data points of the sampling process [56][57] .

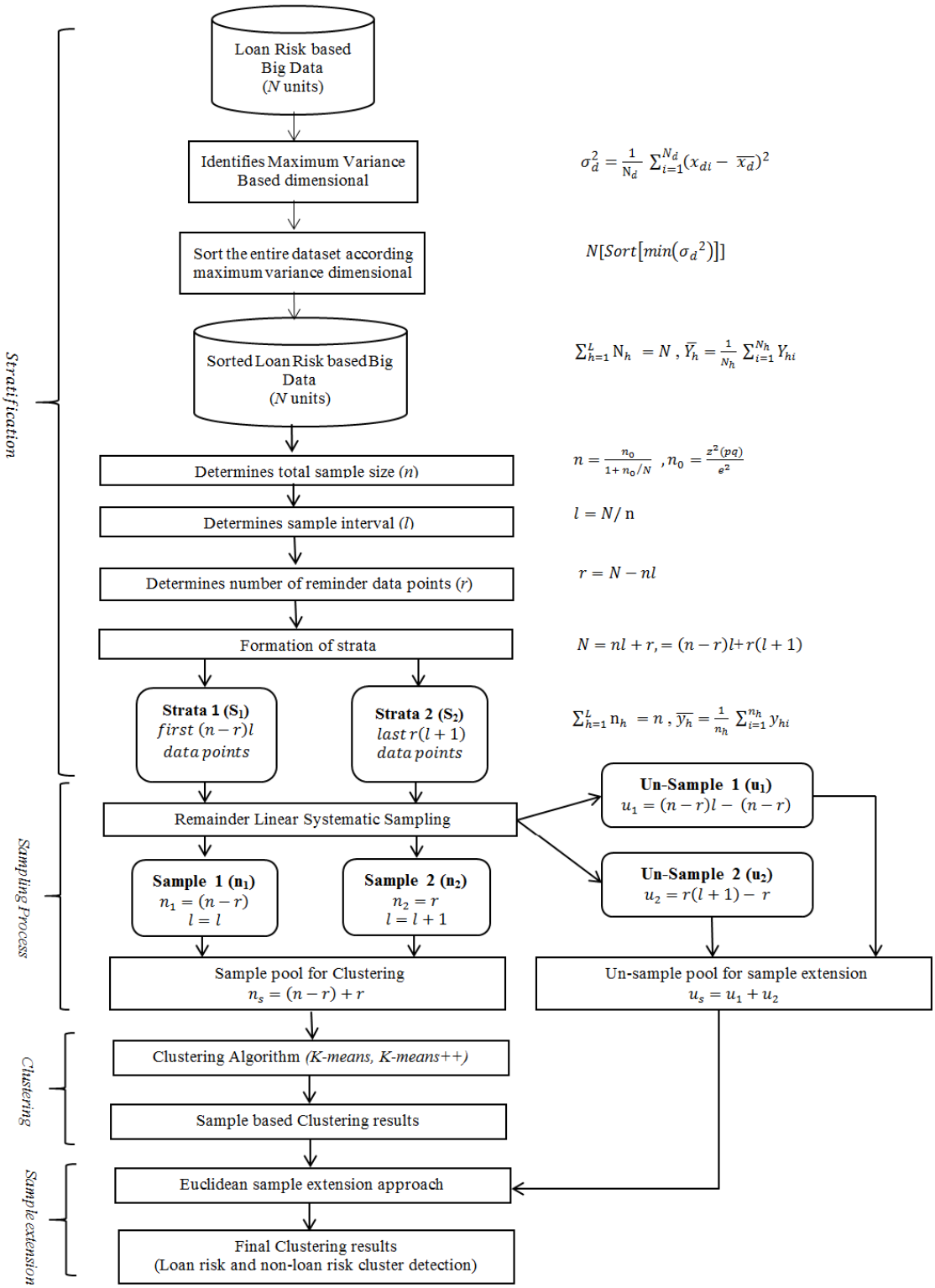


Figure 1. Conceptual Representation of proposed Stratification and Sampling Plan (SRSE)

The number of data points in strata is determined by the integer constraints *n, l,* and *r*. This study has adopted the Cochran formula for sample size identification, which is shown in Eq. 4 and Eq. 5. [58].

$$n_0 = \frac{z^2(pq)}{e^2} \tag{4}$$

where z denotes the standard error, p indicates the variability of the dataset, q signifies the *(p-1)*, and e represents the acceptable sample error. In this study, the standard error is set at 99% for the confidence interval, so the z value is set at 2.576, the variability value p was set at 0.5, and the acceptable sample error e is set at 1% for the 99 confidence interval. To obtain the total sample size, the sample size is normalized by the total number of data points in the loan risk dataset [58].

$$n = \frac{n_0}{1+ n_0/N} \tag{5}$$

The sampling interval is determines every n^{th} data point of stratum is chosen for clustering. Eq. 6 is determined the sampling interval.

$$l = N/n \tag{6}$$

The reminder data points *r* refer to the un-sample data points after the sampling procedure. Eq. 7 describes the identification of the number of reminder data points *r*.

$$r = N - nl \tag{7}$$

The number of data points in the strata is determined by the values of *n, l,* and *r*. The first stratum is made up of the first *(n-r)l* data points from the sorted loan risk dataset. The second strata is represented by the remaining *r(l+1)* data points of the sorted loan risk dataset. These scenarios are described in Eq. 8.

$$N = nl + r = (n - r)l + r(l + 1) \tag{8}$$

After stratification, the RLSS is used to define the sample size and sample selection interval for each stratum. The *(n - r)* data points of the first strata are selected for a sample pool/clustering with a *l* linear systematic sampling interval, while r data points of the second strata are selected for a sample pool/clustering with a *(l +1)* linear systematic sampling interval. As a result, Eq. 9 determines the total number of sample sizes n.

$$n = (n - r) + r \tag{9}$$

C. Sample extension

The Sample extension approach uses centroid-based distance to convert sample-based clustering results into final clustering results. The centroid-based distance used the Euclidean distance approach, which assigns un-sample data to its closed cluster using a centroid of the sample based cluster. Eq. 10 describes the sample extension function, where A_i is the data point of the un-sample data pool and B_i is the mean of the cluster centroid [22].

$$dis_{\text{euclidean}}(A, B) = \sqrt{\sum_{i=1}^{n} |A_i - B_i|^2} \tag{10}$$

III. Algorithm Description

This section describes the stratified remainder linear systematic sampling extension (SRSE) approach for loan risk analysis through stratification, remainder linear systematic sampling, and sample extension. The standard sampling plan first selects the dimension with the highest variance of the risk-based dataset and then sorts the entire dataset based on the selected dimension. The data points are then assigned to both strata using remainder linear systematic sampling rules. After the stratification process, it collects the required number of sample data points into a sample pool for clustering. The strata sample size and sample interval determined the sample data. The sample based clustering results are merged into the final results with the help of the sample extension method. The sample unit is used for clustering, and the resulting value is merged with an un-sample unit via sample extension. The proposed SRSE sampling plan is detailed in Algorithm 1 and the sampling flowchart is shown in Figure 1.

Algorithm 1 Stratified Remainder linear Systematic sampling Extension (SRSE) Big Data Clustering Approach

Input:

1. $N = \{x_1, x_2, \ldots \ldots x_N\}$ *is the data points of the loan risk based D dataset.*
2. *K = Required number of clusters.*

Output:

1. $C_K = \{ C_1, C_2 \ldots \ldots C_k \}$ *of the clustering results.*

Methods:

Stratification

1. Identify the maximum variance dimension of the dataset.
- $v_d = max(\sigma_{1d}{}^2, \sigma_{2d}{}^2, \ldots \ldots \ldots \ldots \ldots, \sigma_{Nd}{}^2,)$

2. Sort the entire data of the dataset according to the v_d dimension in ascending order.

3. Determine the total sample size *n* for the clustering process through Eq. 5.

4. Extracts sample interval *l* from the entire dataset by Eq. 6.

5. Define the number of remainder sample data points *n* through Eq. 7.

6. Determine the number of data points for each stratum with the help of *n, l* and *r*.
- $S1 = (n - r)l$
- $S2 = r(l + 1)$

7. Extract two strata from the entire dataset based on the sorted dataset .
- *S1=N[0:* $(n - r)$]
- *S1=N[*$(n - r): len(N)$]

Sample size identification

8. Determine the number of data point of sample size for both strata according to Eq. 9.
- $n_1 = (n - r)$
- $n_2 = r$

Sample allocation

9. Extract every l^{th} data points for *S1* and every $(l+1)^{th}$ data point for *S2* strata through linear systematic sampling.

10. Combine all *n1* and *n2* sample data points into the n_S sample pool and all un-sample data points into the u_S un-sample pool.

Clustering algorithm

11. Apply necessary clustering algorithms in n_S and achieved approximate clustering results such as *K-means* (n_S *, K), K-means++* (n_S *, K),* etc.

Sample extension

12. According to Eq. 10, assign u_S un-sampled pool data to approximate clustering results based on nearest Euclidean distance.

13. Achieved final clustering results in the loan risk and non-loan risk clusters and Exit.

IV. Experimental Analysis over Loan Risk Data

The experimental study evaluates the research effort based on the computing environment, datasets, existing algorithms, evaluation criteria, and outcomes. This section discusses the experimental environment, loan risk dataset characteristics, and validation criteria. The effectiveness and efficiency-related assessment criteria are used to evaluate the performance of the SRSE-based clustering approach.

I. Experiment Environments and Loan Risk Dataset

The computing environment of the SRSE-based clustering approach used in the Jupyter Notebook computing environment. The experimental environment is configured with an Intel I3 processor, CPU M350@2.27 GHz, 320 GB hard disk, 4(+64) GB DDR3 RAM, Windows 7 operating system, and Python tools. The experimental analysis was performed on four loan risk datasets within a single machine execution environment. Table 1 illustrates the characteristics and sources of the experimental loan risk datasets.

Table 1 Description of the Loan Risk Datasets.

ID	Datasets (DB)	Objects	Attributes	Class	Data Source
LRDB1	Bondora Peer to Peer Lending Loan Data	1,79,235	112	2	https://www.kaggle.com/
LRDB2	Vehicle Loan Default Prediction	3,45,546	41	2	https://www.kaggle.com/
LRDB3	XYZ_Corp Lending Data	8,55,969	70	2	https://www.kaggle.com/
LRDB4	Loan Data for Dummy Bank	8,87,379	30	2	https://www.kaggle.com/

The clustering of the LRDB1 loan risk-related dataset is divided into two classes: default risk and non-default risk. Default risk is a significant risk factor used to evaluate borrowers' behavior in peer-to-peer (P2P) lending. Lenders want to minimize the default risk on each lending decision in order to make rational decisions and to realize a return that compensates for the risk.

The loan risk-related dataset LRDB2 is clustered in order to estimate the determinants of vehicle loan default risk and non-default risk. The clustering process predicts the likelihood of a loanee/borrower defaulting on a vehicle loan during the first EMI (Equated Monthly Instalments) due date. This ensures that clients who are capable of repayment are not turned down. The important determinants are identified, which are used to reduce default rates.

The LRDB3 loan risk dataset clustering manages credit risk by using historical data to determine who to lend to in the future based on default, payment information, credit history, and other factors. The clustering process categorizes the data as capable of loan repayment or incapable of determining loan eligibility.

The LRDB4 clustering divides the data into loan default risk and non-loan risk. Data grouping provides funds for potential borrowers, and banks earn a profit based on the risk they take (the borrower’s credit score).

II. Selected Algorithms for Comparison

The proposed SRSE-based clustering approach is compared to partitional K-means (KM) [20], K-means++ (KM++) [59, 60] clustering algorithms. The goal of both clustering algorithms is to recognize loan risk in terms of default risk by minimizing within-cluster Sum of Squared Error (WSSE) and maximizing between-cluster (BSSE) distance. Except for the initial centroid selection approach, the cluster formulation process of both methods is similar. The KM method chooses the initial centroid at random, whereas the KM++ method chooses the initial centroid based on distance and probability.

III. Evaluation Criteria

Cluster validation is achieved through the application of both internal and external measures. The internal measure is used to compare the cluster's objective to its internal structures. The external measure is used to validate the cluster using outside knowledge. This study employs the Davies Bouldin score (DB), Silhouette coefficient (SC), SD Validity (SD), and Ray-Turi index (RT) as internal validation tools for effectiveness [61–63], with CPU time (CT) serving as an efficiency validation metric [64–66]. The strongest clustering method always maximizes intra-class similarity while decreasing inter-class similarity. As a result, the clustering method maximizes the SC metric value while reducing the DB, SD, RT, and CT metric values.

- Davies Bouldin score (DB): The Davies Bouldin validates within-cluster dispersion and between cluster similarity independently number of cluster. In the DB formulation, $|C_j|$ defines the total number of data point x_i inside of C_j cluster and C_i is another cluster.

$$DB = \frac{1}{K}\sum_{i=1}^{K} max_{i\neq j} \frac{within_i + within_j}{between_{ij}} \tag{11}$$

$$within_j = \frac{1}{|C_j|}\sum_{i=1}^{|c_j|} || x_i - C_j ||^2 \tag{12}$$

$$between_{ij} = || C_i - C_j ||^2 \tag{13}$$

- Silhouette coefficient (SC): The Silhouette coefficient validates cluster similarity by accepting the pairwise difference of cluster distances within (compactness) and between (separation) the clusters. In SC formulation *a(x)* is the average distance of *x* to all other data points in the same cluster *C*, *b(x)* is the average distance of *x* to all other data points in the all *Ci* cluster.

$$S = \left\{\sum_{x\in Ci} \frac{b(x)-a(x)}{\max[b(x),\ a(x)]}\right\} \tag{14}$$

- SD Validity (SD): The SD Validity metric assesses the effectiveness of clustering by averaging dispersion and total separation between clusters with variance. In SD Validity formulation, $\propto$ is constant value equal to 1, S_a is average scattering in term of variance and S_t is the total separation of cluster, $\sigma(C_i)$ is defines variance of C_i cluster, $\sigma(X)$ is represents variance of dataset.

$$\text{SD} = \propto S_a - S_t \tag{15}$$

$$S_a = \frac{1}{k}\sum_{i=1}^{k}\frac{\|\sigma(C_i)\|}{\|\sigma(X)\|} \quad (16)$$

$$S_t = \frac{D_{max}}{D_{min}}\sum_{i=1}^{k}\left(\sum_{j=1}^{k}\| C_i - C_j \|\right)^{-1} \quad (17)$$

$$D_{max} = max_{1\le i,j\le k} \| C_i - C_j \| \quad (18)$$

$$D_{min} = min_{1\le i,j\le k} \| C_i - C_j \| \quad (19)$$

- Ray-Turi index (RT) : The Ray-Turi index measures the mean of the squared distances of the all data points respect to k cluster centroid and minimum squared distance $\Delta_{kk'}^2$ between all cluster centroid. In the RT formulation, N is total length of dataset, M_i^k is the data points of particular cluster k and G^k is the centroid of that cluster. $G^{k'}$is the centroid of remainder cluster.

$$\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in I_k}\|M_i^k - G^k\|^2 = \frac{1}{N}\sum_{k=1}^{K} WGSS^k = \frac{1}{N} WGSS \quad (20)$$

$$\min_{k<k'} \Delta_{kk'}^2 = \min_{k<k'} d(G^k, G^{k'})^2 = \min_{k<k'}\|G^k - G^{k'}\|^2 \quad (21)$$

$$RT = \frac{1}{N}\frac{WGSS}{\min_{k<k'} \Delta_{kk'}^2} \quad (22)$$

- CPU time (CT): CPU time computes the total execution times of any algorithm inside the CPU between the entry ENT and exit EXT times of the clustering algorithm.

$$CT = EX_T - EN_T \quad (23)$$

IV. Experimental Results and Discussion

On the basis of effectiveness and efficiency indices, the performance of SRSE-based clustering algorithms such as SRSE-KM and SRSE-KM++ has been compared to that of the classical KM and KM++ algorithms. Tables 2-3 show the average comparative efficiency and effectiveness results from four loan risk data sets using ten trials. This study used pre-defined Python library functions for DB, SC and SD, as well as technical code for RT, WSSE, BSSE and CT for cluster evaluation. Tables 2-3 highlight the optimal value of each reported result in bold face, where the optimal value of SC is required for maximization and the optimal values of DB, SD, RT, and CT are required for minimization.

Table 2 shows that the proposed SRSE clustering strategy outperformed the KM and KM++ algorithms in terms of WSSE, compaction, separation, similarity, dissimilarity, variance, and density. Table 3 demonstrates that the proposed SRSE strategy is faster than the KM and KM++ algorithms and uses the least amount of CPU time.

The experimental results of the LRDB1, LRDB2, LRDB3, and LRDB4 loan risk datasets illustrate that SRSE-KM and SRSE-KM++ clustering strategies outperform KM and KM++ algorithms in DB, SC, SD, and RT. In terms of clustering quality, the observed DB, SC, SD, and RT values show that the SRSE-KM and SRSE-KM++ algorithms outperform the KM and KM++ algorithms. Inside the LRDB1 risk dataset, the SRSE-KM diminishes the CT to 73.82% as compared to KM, and the SRSE-KM++ decreases the CT to 84.71% than KM++. Over the LRDB2 risk dataset, the SRSE-KM reduces the CT by up to 79.33% compared to the KM, whereas the SRSE-KM++

minimizes the CT to 90.48% than the KM++. For the LRDB3 risk dataset, the SRSE-KM alleviates the CT by 88.06% then KM while the SRSE-KM++ depletes the CT up to 93.79% in the context of the KM++. In LRDB4 risk dataset efficiency observations, the SRSE-KM minimizes the CT to 78.63% with reference to KM, and the SRSE-KM++ reduces the CT to 95.44% than KM++.

Table 2– Comparative average analysis of effectiveness measures ($\boldsymbol{means \pm std}$) over 10 trials

DB	Criteria	KM	**SRSE-KM**	KM++	**SRSE-KM++**
LRDB1	DB	1.93012 ± 0.06722	**1.91441 ± 0.069**	1.95983 ± 0.04012	**1.93705 ± 0.06083**
	SC	0.20382 ± 0.01777	**0.20817 ± 0.01887**	0.1938 ± 0.00344	**0.2002 ± 0.01529**
	SD	1.64826 ± 0.0283	**1.64815 ± 0.02832**	1.66252 ± 0.03457	**1.6553 ± 0.0324**
	RT	0.94748 ± 0.09631	**0.92626 ± 0.10121**	0.99837 ± 0.03155	**0.96234 ± 0.08407**
LRDB2	DB	1.52436 ± 0.42002	**1.47528 ± 0.37001**	1.68744 ± 0.41411	**1.63843 ± 0.38459**
	SC	0.33395 ± 0.10959	**0.34382 ± 0.10014**	0.28805 ± 0.1093	**0.29857 ± 0.1039**
	SD	2.84239 ± 0.63412	**2.76349 ± 0.55905**	3.09039 ± 0.6263	**3.01145 ± 0.58457**
	RT	0.69375 ± 0.47179	**0.62644 ± 0.41411**	0.86921 ± 0.46626	**0.80205 ± 0.43891**
LRDB3	DB	2.94263 ± 0.12489	**2.93696 ± 0.12788**	3.0114 ± 0.22726	**2.97292 ± 0.17355**
	SC	0.10263 ± 0.00735	**0.10265 ± 0.00686**	0.10017 ± 0.01064	**0.10042 ± 0.00948**
	SD	2.32874 ± 0.07231	**2.32833 ± 0.07218**	2.30435 ± 0.15084	**2.30381 ± 0.09601**
	RT	2.20612 ± 0.20542	**2.19989 ± 0.20849**	2.31537 ± 0.38257	**2.25861 ± 0.28156**
LRDB4	DB	1.94253 ± 0.24398	**1.89247 ± 0.18771**	1.99679 ± 0.22767	**1.89885 ± 0.27487**
	SC	0.21184 ± 0.03164	**0.21962 ± 0.02426**	0.2048 ± 0.02778	**0.2205 ± 0.02898**
	SD	2.02541 ± 0.18196	**2.01309 ± 0.158**	2.1079 ± 0.18926	**1.94093 ± 0.13944**
	RT	1.00316 ± 0.2711	**0.93779 ± 0.20157**	1.05623 ± 0.25264	**0.94993 ± 0.31184**

Table 3– Comparative average analysis of efficiency CT measure ($\boldsymbol{means \pm std}$) over 10 trials

DS	KM	**SRSE-KM**	KM++	**SRSE-KM++**
LRDB1	8.83084 ± 2.35755	**2.31235 ± 0.24165**	12.60682 ± 1.69874	**1.92721 ± 0.57942**
LRDB2	7.7521 ± 2.25847	**1.60213 ± 0.09076**	**16.63186 ± 3.9364**	**1.58328 ± 0.64967**
LRDB3	11.26206 ± 2.51383	**1.34438 ± 0.03677**	24.22219 ± 4.92122	**1.50246 ± 0.05333**
LRDB4	6.29301 ± 3.01334	**3.34744 ± 5.10325**	43.21801 ± 6.65378	**1.96903 ± 0.42415**

Figure 2-5 depicts a comparative analysis of clustering objective and efficiency-related measures, with the resulting values ordered ascending to identify minimum to maximum values. The WSSE clustering objective for the KM and SRSE-KM algorithms is depicted in Figure 2, whereas the WSSE clustering objective for the KM++ and SRSE-KM++ algorithms is depicted in Figure 3. The minimum WSSE result shows that the proposed sampling plan consistently achieves the excellence WSSE in each trial on the experimental loan risk data sets. The observation of Figures 2-3 indicates that the SRSE based clustering algorithm achieves better compaction and separation of the cluster with the clustering objective.

Figure 4 demonstrates the computing time efficiency measurements for the KM and SRSE-KM algorithms, while Figure 5 reveals the computing time efficiency measurements for the KM++ and SRSE-KM++ algorithms. The proposed sampling plan minimizes the computation cost, iterations, number of distances, and data comparisons in each trial on the experimental risk data sets, implying that the proposed sampling plan minimizes the computation cost, iterations, number of distances, and data comparisons in each trial on the experimental risk data sets. Figures 4-5 illustrate that the SRSE-based clustering algorithm outperforms the KM and KM++ algorithms in terms of speed and resilience.

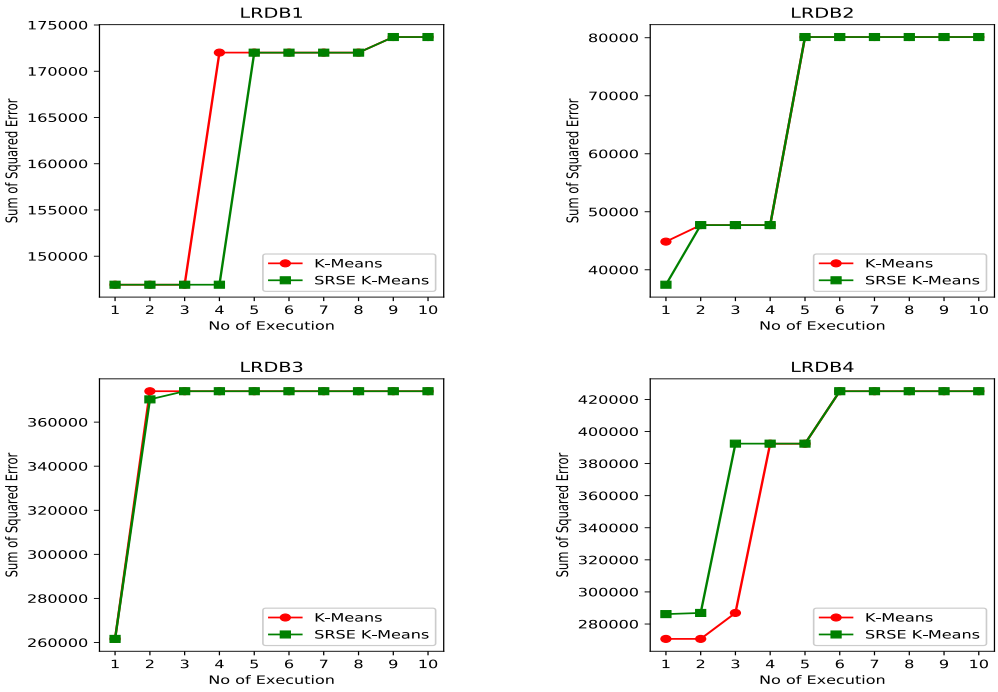


Figure 2 Analysis of total-WSSE between KM and SRSE-KM on each trial

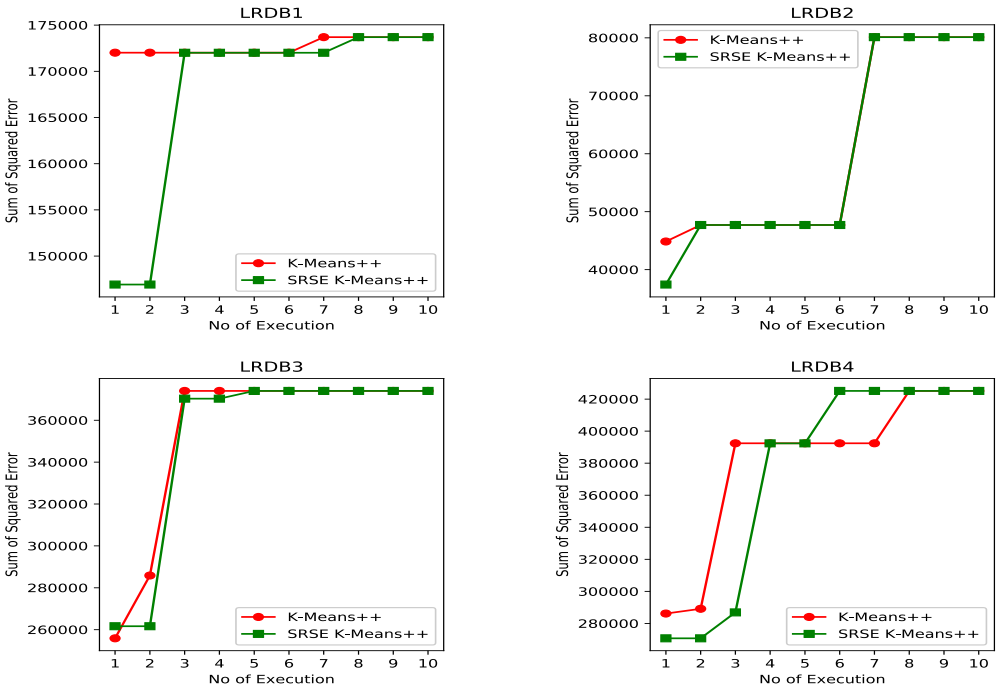


Figure 3 Analysis of total-WSSE between KM++ and SRSE-KM++ on each trial

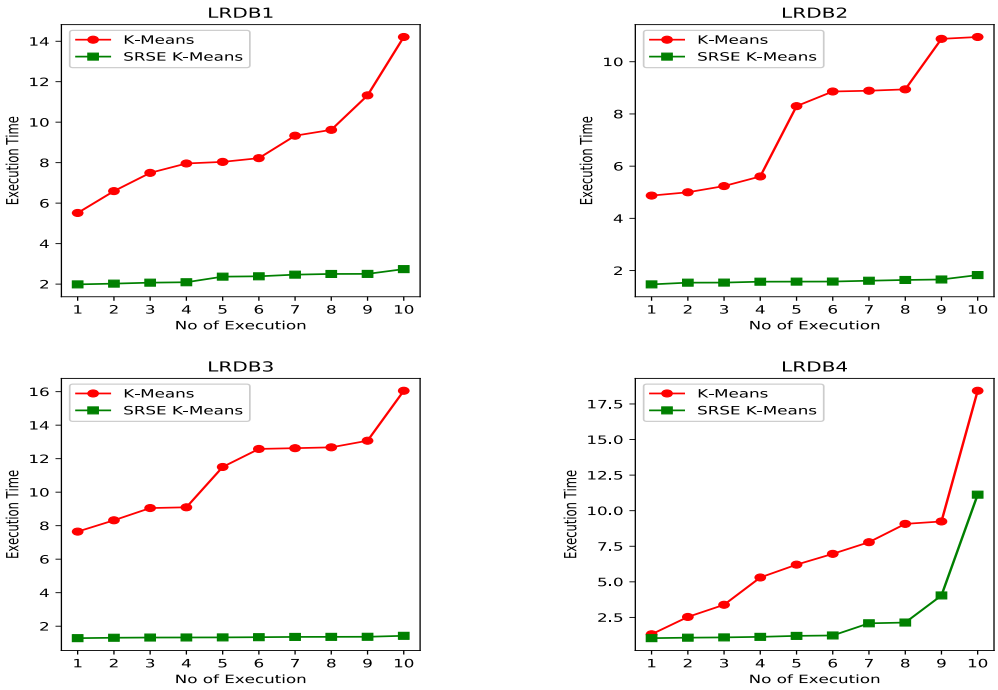


Figure 4 Analysis of computing time between KM and SRSE-KM on each trial

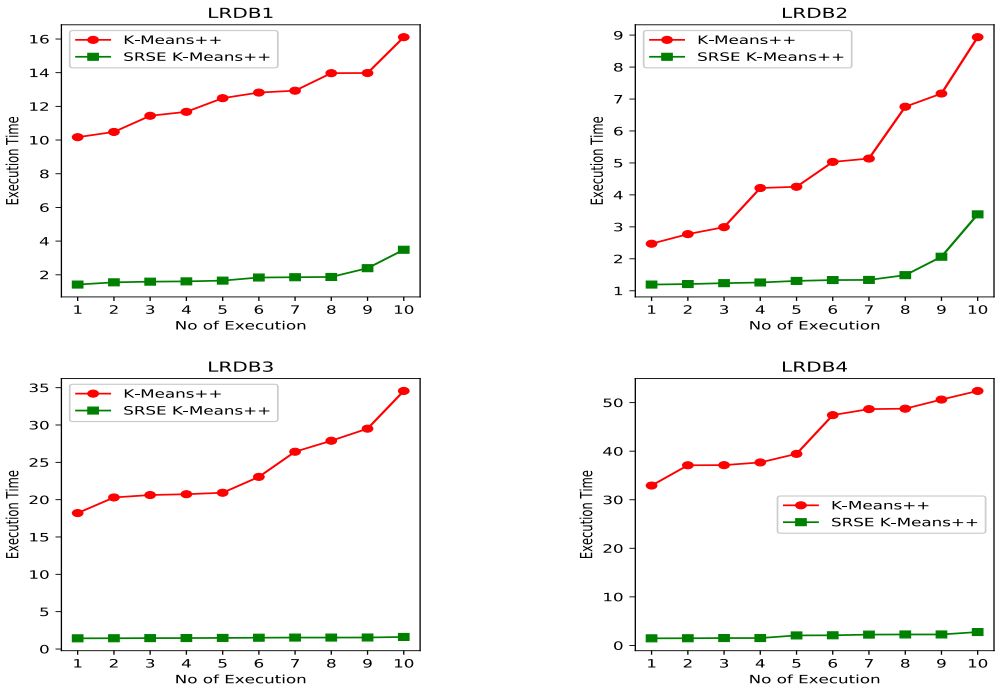


Figure 5 Analysis of computing time between KM++ and SRSE-KM++ on each trial

The proposed sampling-based clustering algorithm improves cluster quality and clustering objective while reducing data and distance comparisons with execution times, as shown in Table 2-3 and Figure 2-5. The presented sampling-based clustering algorithm also outperforms previous KM and KM++ algorithms in terms of speed and scalability on loan risk-based big data. On loan risk-based big data, the SRSE-based clustering algorithm eliminates the worst-case scenario of the KM and KM++ algorithms. The analysis shows that the proposed SRSE-KM and SRSE-KM++ algorithms are more robust for big data clustering than the KM and KM++ algorithms.

V. Conclusion

This study presents, a stratified remainder linear systematic sampling extension (SRSE) based clustering approach for loan risk analysis on big data using the KM and KM++ clustering algorithms. The proposed clustering SRSE-KM and SRSE-KM++ algorithms employ five stages to reduce computing time while improving cluster quality. The first step is to sort the entire dataset in order to create a stratification using the maximum variance attribute approach. The second stage determines the total sample size, sample interval, and reminder sample values in order to calculate the total number of data items and sample size in stratum. The third stage divides the data points into strata and uses a liner systematic sampling procedure to extract sample data from each stratum. The fourth stage clusters the sample data according to the selected clustering algorithm. The final stage uses a centroid-based sample extension approach to merge the sample data results to an un-sample data unit. The final results demonstrate that the unknown loan uncertainty of risk belongs to one cluster and non-risky data belongs to another cluster. Experiment results show that the SRSE-based clustering algorithm never degrades cluster performance and achieves better cluster compaction, separation, variance, density, computing cost, and execution time than classical clustering algorithms. The proposed SRSE-KM algorithm reduces average computing time by up to 75.25% when compared to KM, and the SRSE-KM++ algorithm reduces average computing time by up to 92.78% when compared to KM++. Despite the fact that the SRSE-based clustering algorithm significantly reduces clustering time, but it suffers local optima issues due to randomization. The study's further scope is to open up to resolve local optima concerns on multiple machine-based technologies such as Hadoop and Spark via other internal and external validation indexes using various loan risk-related data sets.

References

[1] Lozada N, Arias-Pérez J, Perdomo-Charry G (2019) Big data analytics capability and co-innovation: An empirical study. Heliyon 5:10. https://doi.org/10.1016/j.heliyon.2019.e02541

[2] Hariri RH, Fredericks EM, Bowers KM (2019) Uncertainty in big data analytics: survey, opportunities, and challenges. J Big Data 6:44. https://doi.org/10.1186/s40537-019-0206-3

[3] Tabesh P, Mousavidin E, Hasani S (2019) Implementing big data strategies: A managerial perspective. Bus Horiz 62:347–358. https://doi.org/10.1016/j.bushor.2019.02.001

[4] Elgendy N, Elragal A (2014) Big Data Analytics: A Literature Review Paper. In: Perner P (ed) ICDM 2014, LNAI 8557. Springer International Publishing Switzerland, pp 214–227

[5] Sivarajah U, Kamal MM, Irani Z, Weerakkody V (2017) Critical analysis of big data challenges and analytical methods. J Bus Res 70:263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

[6] Gandomi A, Haider M (2015) Beyond the hype: big data concepts methods and analytics. Int J Inf Manage 35:137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

[7] Pandey KK, Shukla D (2019) Challenges of big data to big data mining with their processing framework. In: 2018 8th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, pp 89–94

[8] Kacfah Emani C, Cullot N, Nicolle C (2015) Understandable big data: a survey. Comput Sci Rev 17:70–81. https://doi.org/10.1016/j.cosrev.2015.05.002

[9] Moharm K (2019) State of the art in big data applications in microgrid: A review. Adv Eng Informatics 42:. https://doi.org/10.1016/j.aei.2019.100945

[10] Khondoker MR (2018) Big data clustering. In: Wiley StatsRef: Statistics Reference Online.

John Wiley & Sons, Ltd, Chichester, UK, pp 1–10

[11] Pandove D, Goel S, Rani R (2018) Systematic review of clustering high-dimensional and large datasets. ACM Trans Knowl Discov Data 12:1–68. https://doi.org/10.1145/3132088

[12] Xie H, Zhang L, Lim CP, et al (2019) Improving K-means clustering with enhanced Firefly Algorithms. Appl Soft Comput J 84:105763. https://doi.org/10.1016/j.asoc.2019.105763

[13] Kara ME (2018) Supplier Risk Assessment Based on Best-Worst Method and K-Means Clustering: A Case Study. Sustainability 10:1066. https://doi.org/10.3390/su10041066

14. Mandelli D, Yilmaz A, Aldemir T, et al (2013) Scenario clustering and dynamic probabilistic risk assessment. Reliab Eng Syst Saf 115:146–160. https://doi.org/10.1016/j.ress.2013.02.013

[15] Marle F, Vidal L, Bocquet J (2013) Interactions-based risk clustering methodologies and algorithms for complex project management. Int J Prod Econ 142:225–234. https://doi.org/10.1016/j.ijpe.2010.11.022

[16] Kou G, Peng Y, Wang G (2014) Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Inf Sci (Ny) 275:1–12. https://doi.org/10.1016/j.ins.2014.02.137

[17] Wang X, Frattini P, Stead D, et al (2020) Dynamic rockfall risk analysis. Eng Geol 272:105622. https://doi.org/10.1016/j.enggeo.2020.105622

[18] Caicedo PE, Rengifo CF, Rodriguez LE, et al (2020) Dataset for gait analysis and assessment of fall risk for older adults. Data Br 33:106550. https://doi.org/10.1016/j.dib.2020.106550

[19] Fahad A, Alshatri N, Tari Z, et al (2014) A survey of clustering algorithms for big data: taxonomy and empirical analysis. IEEE Trans Emerg Top Comput 2:267–279. https://doi.org/10.1109/TETC.2014.2330519

[20] Jain AK (2010) Data clustering: 50 years beyond k-means. Pattern Recognit Lett 31:651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[21] Wang X, He Y (2016) Learning from Uncertainty for Big Data: Future Analytical Challenges and Strategies. IEEE Syst Man, Cybern Mag 2:26–31. https://doi.org/10.1109/msmc.2016.2557479

[22] Zhao X, Liang J, Dang C (2019) A stratified sampling based clustering algorithm for large-scale data. Knowledge-Based Syst 163:416–428. https://doi.org/10.1016/j.knosys.2018.09.007

[23] Liu Z, Zhang A (2020) A Survey on Sampling and Profiling over Big Data (Technical Report). 1–17

[24] Mahmud MS, Huang JZ, Salloum S, et al (2020) A survey of data partitioning and sampling methods to support big data analysis. Big Data Min Anal 3:85–101. https://doi.org/10.26599/BDMA.2019.9020015

[25] Umarani V, Punithavalli M (2011) Analysis of the progressive sampling-based approach using real life datasets. Open Comput Sci 1:221–242. https://doi.org/10.2478/s13537-011-0016-y

[26] Chen B, Haas P, Scheuermann P (2002) A new two-phase sampling based algorithm for discovering association rules. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Digital Library, pp 462–468

[27] Furht B, Villanustre F (2016) Big Data Technologies and Applications. Springer International Publishing, Cham

[28] Ramasubramanian K, Singh A (2016) Sampling and Resampling Techniques. In: Machine Learning Using R. pp 67–127

[29] Haas PJ (2016) Data-Stream Sampling: Basic Techniques and Results. Springer-Verlag Berlin Heidelberg

[30] Kim JK, Wang Z (2019) Sampling Techniques for Big Data Analysis. Int Stat Rev 87:S177–S191. https://doi.org/10.1111/insr.12290

[31] Brus DJ (2019) Sampling for digital soil mapping: A tutorial supported by R scripts. Geoderma 338:464–480. https://doi.org/10.1016/j.geoderma.2018.07.036

[32] Aune-Lundberg L, Strand G (2014) Comparison of variance estimation methods for use with two-dimensional systematic sampling of land use/land cover data. Environ Model Softw 61:87–97. https://doi.org/10.1016/j.envsoft.2014.07.001

[33] Satyanarayana A (2014) Intelligent sampling for big data using bootstrap sampling and chebyshev inequality. Can Conf Electr Comput Eng. https://doi.org/10.1109/CCECE.2014.6901029

[34] Buddhakulsomsiri J, Parthanadee P (2008) Stratified random sampling for estimating billing accuracy in health care systems. Health Care Manag Sci 11:41–54. https://doi.org/10.1007/s10729-007-9023-x

[35] da Silva A, Chiky R, Hébrail G (2012) A clustering approach for sampling data streams in sensor networks. Knowl Inf Syst 32:1–23. https://doi.org/10.1007/s10115-011-0448-7

[36] Rajasekaran S, Saha S (2013) A novel deterministic sampling technique to speedup clustering algorithms. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 8347 LNAI:34–46. https://doi.org/10.1007/978-3-642-53917-6_4

[37] Jaiswal R, Kumar A, Sen S (2014) A Simple D 2 -Sampling Based PTAS for k -Means. 22–46. https://doi.org/10.1007/s00453-013-9833-9

[38] Xu Z, Wu Z, Cao J, Xuan H (2015) Scaling Information-Theoretic Text Clustering: A Sampling-based Approximate Method. In: Proceedings - 2014 2nd International Conference on Advanced Cloud and Big Data, CBD 2014. pp 18–25

[39] Luchi D, Santos W, Rodrigues A, Varejao FM (2015) Genetic Sampling k -means for Clustering Large Data Sets. In: CIARP 2015, LNCS 9423. pp 691–698

[40] Jing L, Tian K, Huang JZ (2015) Stratified feature sampling method for ensemble clustering of high dimensional data. Pattern Recognit 48:3688–3702. https://doi.org/10.1016/j.patcog.2015.05.006

[41] Li M, Li D, Shen S, et al (2016) DSS: A Scalable and Efficient Stratified Sampling Algorithm for Large-Scale Datasets. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp 133–146

[42] Zhan Q (2017) Improved spectral clustering based on Nyström method. Multimed Tools Appl 76:20149–20165. https://doi.org/10.1007/s11042-017-4566-4

[43] Aloise D, Contardo C (2018) A sampling-based exact algorithm for the solution of the minimax diameter clustering problem. J Glob Optim 71:613–630. https://doi.org/10.1007/s10898-018-0634-1

[44] Reddy KG, Khan MGM (2019) Optimal stratification in stratified designs using weibull-distributed auxiliary information. Commun Stat - Theory Methods 48:3136–3152. https://doi.org/10.1080/03610926.2018.1473609

[45] Saini M, Kumar A (2018) Ratio estimators using stratified random sampling and stratified ranked set sampling. Life Cycle Reliab Saf Eng. https://doi.org/10.1007/s41872-018-0046-8

[46] Li F, Qian Y, Wang J, et al (2019) Clustering ensemble based on sample's stability. Artif Intell 273:37–55. https://doi.org/10.1016/j.artint.2018.12.007

[47] Goshu NN, Kassa SM (2020) A Systematic Sampling Evolutionary (SSE) Method for Stochastic Bilevel Programming Prob-. Comput Oper Res 104942. https://doi.org/10.1016/j.cor.2020.104942

[48] Rajendra Prasad K, Mohammed M, Narasimha Prasad L V., Anguraj DK (2021) An efficient sampling-based visualization technique for big data clustering with crisp partitions. Distrib Parallel Databases 39:813–832. https://doi.org/10.1007/s10619-021-07324-3

[49] Nguyen TD, Shih MH, Srivastava D, et al (2021) Stratified random sampling from streaming and stored data. Springer US

[50] Larson L, Larson P, Johnson DE (2019) Differences in Stubble Height Estimates Resulting from Systematic and Random Sample Designs. Rangel Ecol Manag 72:586–589.

https://doi.org/10.1016/j.rama.2019.03.007

[51] Pandey kamlesh kumar, Shukla D (2020) Stratified Sampling-Based Data Reduction and Categorization Model for Big Data Mining. In: Gupta JC, Kumar BM, Sharma H, Agarwal B (eds) Communication and Intelligent Systems

[52] Pandey KK, Shukla D (2019) Optimized sampling strategy for big data mining through stratified sampling. Int J Sci Technol Res 8:3696–3702

[53] Xiao Y, Yu J (2012) Partitive clustering (k -means family). Wiley Interdiscip Rev Data Min Knowl Discov 2:209–225. https://doi.org/10.1002/widm.1049

[54] Rice JA (2007) Mathematical statistics and metastatistical analysis, Third Edit. Thomson Higher Education

[55] Singh S (2003) Advanced sampling theory with applications, volume 1. Springer Netherlands, Dordrecht

[56] Chang H-J, Huang K-C (2000) Remainder Linear Systematic Sampling. Indian J Stat Ser B 62:249–56

[57] Mostafa SA, Ahmad IA (2018) Recent developments in systematic sampling: a review. J Stat Theory Pract 12:290–310. https://doi.org/10.1080/15598608.2017.1353456

[58] Cochran WG (1962) Samling Techniques. Asia Publishing House, Bombay

[59] Arthur D, Vassilvitskii S (2007) K-means++: The advantages of careful seeding. In: SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. ACM Digital Library, pp 1027–1035

[60] Fränti P, Sieranoja S (2019) How much can k-means be improved by using better initialization and repeats? Pattern Recognit 93:95–112. https://doi.org/10.1016/j.patcog.2019.04.014

[61] HajKacem MA Ben, N'Cir C-E Ben, Essoussi N (2019) Overview of scalable partitional methods for big data clustering. In: Nasraoui O, N'Cir C-E Ben (eds) Clustering Methods for Big Data Analytics, Unsupervised and Semi-Supervised Learning. Springer Nature, Switzerland, pp 1–23

[62] Aggarwal CC, Reddy CK (2014) Data custering algorithms and applications. CRC Press, Boca Raton, United States

[63] Sid R, H TR (1999) Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation. In: 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99). Narosa Publishing House, pp 137–143

[64] Peña J., Lozano J., Larrañaga P (1999) An empirical comparison of four initialization methods for the k-means algorithm. Pattern Recognit Lett 20:1027–1040. https://doi.org/10.1016/S0167-8655(99)00069-0

[65] Celebi ME, Kingravi HA, Vela PA (2013) A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 40:200–210. https://doi.org/10.1016/j.eswa.2012.07.021

[66] Zahra S, Ghazanfar MA, Khalid A, et al (2015) Novel centroid selection approaches for k-means-clustering based recommender systems. Inf Sci (Ny) 320:156–189. https://doi.org/10.1016/j.ins.2015.03.062

Investigation of Effects of Uncertain Weather Conditions on the Power Generation Ability of Wind Turbines

Endalew Ayenew Haile[1], Getachew Biru Worku[2], Asrat Mulatu Beyene[3], and Milkias Berhanu Tuka[4]

•

[1,4] Department of Electrical Power and Control Engineering, School of Electrical Engineering and Computing, Adama Science and Technology University, Adama, Ethiopia
[2,] School of Electrical and Computer Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia.
[3,] Sustainable Energy Center of Excellence, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
Email: [1]endalew.ayenew@astu.edu.et; [2]*gbiru@yahoo.co.uk*;
[3]*asrat.mulatu@aastu.edu.et;* [4]milkiasber@gmail.com

Abstract

Wind energy is one of the abundant and renewable energy sources that can be harvested using wind turbines. Many factors affect the energy-yielding ability of wind turbines. The goal of this paper is to investigate the effects of uncertain weather conditions on the power generation ability of wind turbines. Uniquely, it presented the influence of the uncertain weather conditions and uncertain aerodynamic parameters of wind turbine on wind energy harvesting. The mathematical model of these factors and statistical analysis of their effects on the performance of wind turbines are presented using real-time data. It is found that the impact of uncertain weather conditions on annual average air density, and hence on the performance of wind turbines, is 1.33%. Whereas, the impact of variations of yearly average wind speed on the performance of wind turbines is found to be substantial. In particular, the annual uncertainty output power of wind turbines is found to be 32%. This investigation helps to find the mitigation mechanism and improve power generation efficiency from wind.

Keywords: wind turbine aerodynamics, uncertain weather conditions, wind energy conversion system, wind turbine energy conversion factor.

I. Introduction

Wind turbines convert wind energy to electric energy using generators. In 2019, 651 GW has been harvested globally [1]. The power harvesting ability of wind turbines is one of the key performance measures. Wind turbines must produce desired output power under stated conditions. Technically, variable-speed and variable-pitch regulated wind turbines have good power harvesting ability. However, it is easily affected by the unpredictability of weather conditions.

Most of the related researches consider only the reliability of facilities and physical components of wind turbines, variation in weather conditions are not considered as in [2], [3], and [4]. However, weather parameters variation reduce the reliability of wind turbines [5] by causing failure to components of wind turbines [6] or unable to drive the turbines, because weather is inconsistent.

Weather conditions are stochastic due to the unequal hotness of air on the surface of the earth. The Equator is hotter than polar areas. This varying surface temperature causes variations in atmospheric pressure. As temperature varies, pressure varies proportionally. Hence, around tropical regions, there is higher pressure than the polar. This initiates air blows from equatorial regions towards the poles. That means wind speed is higher around the equator than the polar region on the earth's surface. This indicates more wind energy is available around the equator than in the polar areas. Additionally, variations in temperature cause air humidity variations. This phenomenon gave rise to variations in air density. These fluctuations are called uncertain weather conditions. The variation in wind speed leads into variation aerodynamic parameters of wind turbine tip speed ratio, rotor speed, and power conversion coefficient. Fluctuations in wind speed, air density and aerodynamic parameters of wind turbine create fluctuations in output power of wind turbine.

The flow of atmospheric wind, and hence wind resource assessment, is affected by variations in the dispersion of solar energy, spatial inequalities in heat transfer on the earth's surface, and the earth's rotation [7]. The uncertain characteristic of wind speed is a vital factor in wind power harvesting [8]. Interesting facts like historic climate data accuracy of 1.5% to 4%, future variability accuracy of 1% to 3%, spatial variability accuracy of 1% to 4%, and energy loss accuracy of 1% to 3% are presented in [9]. Interfaces within wind turbines or wake effects create wind speed uncertainties [10], [11], and [12]. The amount of these uncertainties and their causes are presented in [13] and [14]. Every 10-meter vertical extrapolation of wind speed data results in a 1% uncertainty [15].

Wind speed is a stochastic variable that supplies energy and, at the same time, acts as a disturbance in wind energy harvesting systems. Wind speed models have four components; namely base, gust, ramp, and noise that characterize the variation in wind speed [16]. Another factor that affects the power harvesting ability of wind turbines is air thickness. The effect of temperature, pressure and humidity on air density is presented in [17]. These parameters also affect the power conversion coefficient of wind turbines. Varieties of uncertainties present in annual energy production from wind are presented in [18] and [19]. These studies indicated that wind speed uncertainty highly affects energy production from wind. For a 2.6% deviation in a 5 m/s annual average wind speed, there is 9.9% total uncertainty in annual energy production [20].

Power curve variability between the cut-in and the rated values of wind speed is another source of uncertainty in wind energy production [21]. Total wind power production can be varied seasonally or timely. For example, Simon Watson quantified hourly maximum change in wind power production using data derived from 1500 turbines in Germany. Accordingly, there is ± 50% variability in wind power production only within a four-hour duration [22]. Warren K. et.al concluded that there is 75–85% fluctuation in day-to-day maximum power produced by a wind plant in the USA in Texas [23]. The annual production of energy from the wind is varied by ± 40% in the USA at Lake Benton [24]. IEC 614400-12-1 stated there could be a 15.67% error in wind speed due to an error in site calibration [25].

Wind speed uncertainty reduces the output power of wind turbines. For instance, the uncertainty in tip speed ratio has a considerable effect on wind energy harvesting. For 5% uncertainty in the tip speed ratio of blades, there is a 1–3% energy loss while wind turbines run in a region below the rated wind speed [26], [27], and [28]. That means, a single-unit wind turbine of a 1.5 MW rating operates at a 32% capacity factor and produces 4.208 GWh energy annually. Suppose, the cheapest cost of energy is $0.09/kWh, a 1–3% loss of energy is equivalent to a $3787–$11361 loss annually.

This indicates how much money could be lost due to uncertainty alone due to the tip speed ratio of wind turbines. As a result, the power harvesting ability of wind turbines is degraded. Electric loads do not uniform throughout the day. Wind turbine-connected electric loads are another source of uncertainty in power generation [29].

Gaps and contributions: In the aforementioned literature, the influence of the uncertainty of air density and wind turbine aerodynamic on wind energy harvesting is not considered. Moreover, there is no comprehensive mathematical model relating these uncertainty parameters with wind turbine power harvesting capacity. Therefore, the major contributions of this study are

- Comprehensive and concise mathematical models that relate uncertain weather parameters and wind turbine power harvesting ability are formulated.
- The effects of these uncertain parameters on wind turbine performance are investigated.
- The impact of combined uncertainty is investigated by introducing a scaling factor.

This study, therefore, aims to extensively address these points using real-time data of a specific wind farm site. The next parts of this paper include; the research method, analytical model of uncertainty in wind power harvesting, results and discussion, and conclusions.

II. Methods

Real-time annual wind speed data and related weather parameters (temperature, air pressure, and air humidity) from June 2019 to May 2020 were collected at 10 meters above the surface of the earth in a Tropical Zone at the Adama II wind farm site of Adama, Ethiopia using the 10-channel logger, METRO-32. The data is logged every 10 minutes. Daily, socket 3 of the METRO-32 data logger stores 144 samples of wind speed. The METRO-32 data logger 4^{th} socket recorded weather parameters.

Uncertainty models of wind power harvesting are formulated considering a 1.5 MW wind turbine, whose technical specification is given in Table 1. The wind speed data is extrapolated to the 70 m hub height of the same model. Variations in the recorded daily average values of the weather parameters for the duration of June 2019 to May 2020 are depicted in Figure 1. The daily average temperature, pressure, relative humidity, wind speed, and wind turbine rotor speed variations are found to be 11–26 ^{0}c, around 1020 mb, 23% to 85%, 3 to 16 m/s, and 10 to19 rpm, respectively.

Analysis of the power conversion coefficient of the 1.5 MW wind turbine, the numerical computation of the uncertainties in the aforementioned parameters, and their effects on the wind power harvesting ability of the turbine are carried out.

Table 1: *Technical Parameters of the 1.5 MW Wind Turbine Rotor*

Parameter/Description	Value
Rated power	1500 kW
wind speed	Cut-in 3.5 m/s, Rated 12 m/s , Cut-out 25 m/s
Height of wheel hub center	77 m
Rotor radius	37.8 m
Rated rotating speed	19 rpm
Number of blades	3 blades with independent variable pitch controls

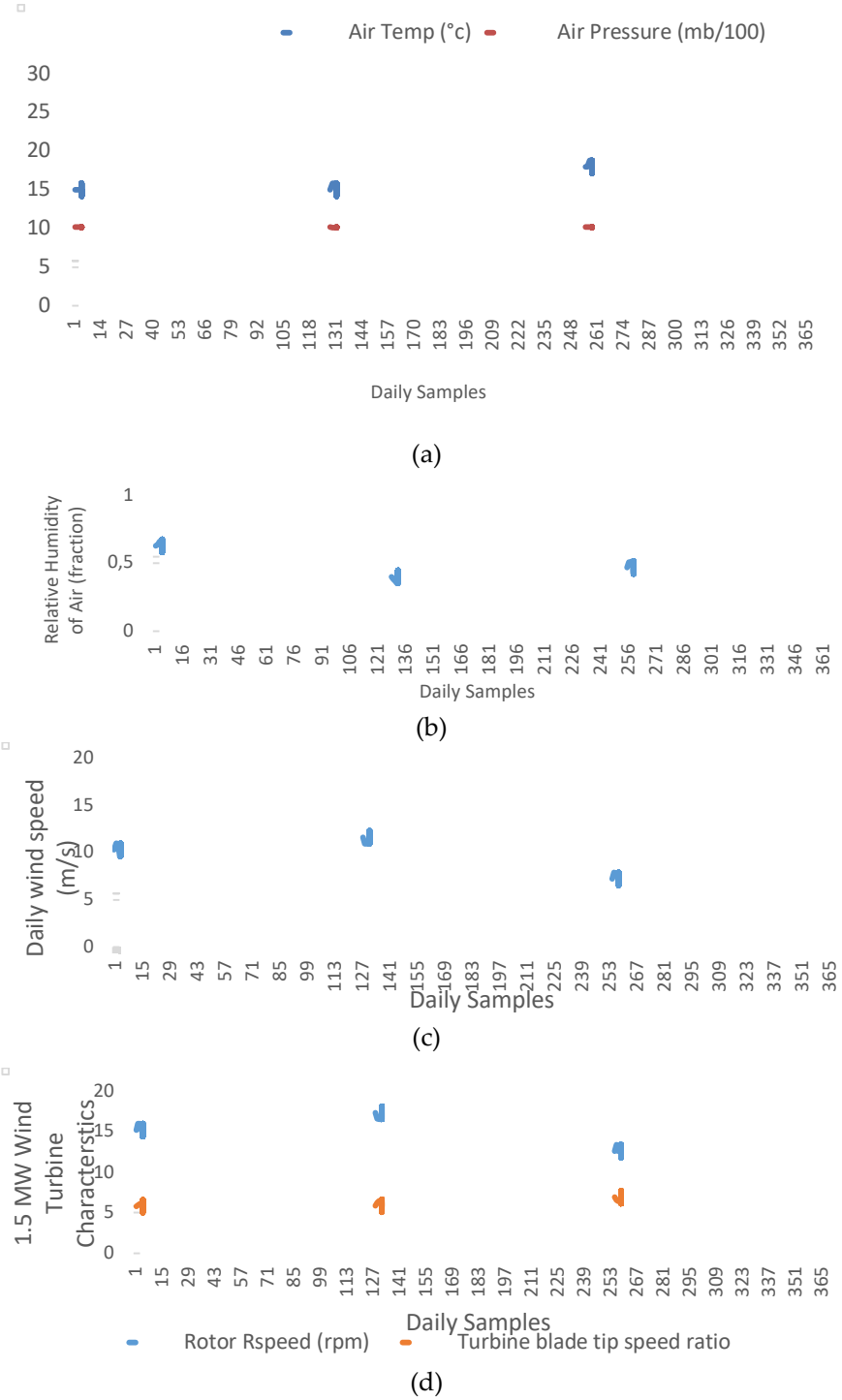


Figure 1: *Adama-II Windfarm annual measured data of (a) air temperature, (b) air pressure, and (c) air relative humidity (d) wind speed*

III. Analytical Model of Uncertainty in Wind Power Harvesting

The components of uncertainty affecting the performance of electric power-producing wind turbines

are categorized as Type-A and Type-B uncertainties [28]. These categories of uncertainties are computed employing statistical methods. As stated in ISO/IEC Guide 98-3: 2008(E), estimation of the combined standard uncertainty of type-A and type-B errors is possible for each can be expressed in terms of standard deviation [30]. The basic mathematical model [7] for the conversion of wind power to mechanical power by a horizontal axis wind turbine is modified with a scale factor ($|\delta| \leq 1$) of the uncertainty in output power which is introduced and expressed in equation (1) for a wind turbine whose specifications are given in Table 1.

$$\mathrm{P}\big(v(t), \rho, C_p\big) = \begin{cases} 0; \text{for } v(\mathrm{t}) < 3\ m/s \\ 0.5\mathrm{A}\rho C_p(\lambda, \beta)v^3 \pm \Delta\mathrm{p}\big(v, \rho, C_p\big)|\delta|;\ \text{for } 3\ v(\mathrm{t}) \leq 12\ \mathrm{m/s} \\ \mathrm{P}_{\mathrm{rated}}\ ; \text{for } 12\ \leq v(\mathrm{t}) < 25\mathrm{m/s} \\ 0; \text{for } v(\mathrm{t}) > 25\ m/s \end{cases} \tag{1}$$

Where A is the swept area of the rotor. A functional model between the turbine output power and measured variables such as wind speed-$v(t)$, air density-ρ and power conversion coefficient ($C_p(\lambda, \beta) = C_p$) is given in equation (2)

$$\mathrm{P}\big(v(t), \rho, C_p\big) = g(v, \rho, C_p) \tag{2}$$

Recognizing variations in the measured or computed variables (v, ρ, C_p), it is important to formulate a model of uncertainties in the functional relationship. For N samples of measured variables $(v_1, v_{2,\ldots}, v_{N;}\, \rho_1, \rho_{2,\ldots}, \rho_{N;}\, C_{p_1}, C_{p_{2,\ldots}}, C_{p_N})$, the averages of each measured variable is as in equation (3)

$$\overline{v} = \frac{1}{N}\sum_{i=1}^{N} v_i;\ \overline{\rho} = \frac{1}{N}\sum_{i=1}^{N} \rho_i;\ \overline{C_p} = \frac{1}{N}\sum_{i=1}^{N} C_{p_i} \tag{3}$$

Wind turbine mechanical output power (P_i) can be computed at any wind speed (v_i), air density (ρ_i), and power conversion coefficient (C_{p_i}) which is $g\big(\overline{v}, \overline{\rho}, \overline{C_p}\big)$ at average values of the measured variables. Using the Taylor series, P_i can be expanded by centering the average values as in equation (4).

$$\begin{aligned} P_i = {} & g\big(\overline{v}, \overline{\rho}, \overline{C_p}\big) + (v_i - \overline{v})\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\Big|_{\overline{v}} + (\rho_i - \overline{\rho})\frac{\partial \mathrm{P}}{\partial \rho}\Big|_{\overline{\rho}} \\ & + \Big(C_{p_i} - \overline{C_p}\Big)\frac{\partial \mathrm{P}}{\partial C_p}\Big|_{\overline{C_p}} + \text{higher} - \text{order terms} \end{aligned} \tag{4}$$

Averaging the measured values, the higher-order terms can be dropped as expressed in equation (5).

$$P_i - \mathrm{P}\big(v(t), \rho, C_p\big) = (v_i - \overline{v})\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\Big|_{\overline{v}} + (\rho_i - \overline{\rho})\frac{\partial \mathrm{P}}{\partial \rho}\Big|_{\overline{\rho}} + \Big(C_{p_i} - \overline{C_p}\Big)\frac{\partial \mathrm{P}}{\partial C_p}\Big|_{\overline{C_p}} \tag{5}$$

Now, we can define the uncertainty (variation or standard deviation) in the output power of wind turbines as

$$
\begin{aligned}
[\Delta \mathrm{p}(v,\rho,C_p)]^2 = [\Delta_P]^2 &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}[P_i - \mathrm{P}(v(t),\rho,C_p)]^2 \\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left[(v_i-\overline{v})\left.\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\right|_{v,\overline{\rho},\overline{C_p}} + \left(C_{p_i}-\overline{C_p}\right)\left.\frac{\partial \mathrm{P}}{C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}} + (\rho_i-\overline{\rho})\left.\frac{\partial \mathrm{P}}{\partial \rho}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right]^2 \\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(v_i-\overline{v})^2\left(\left.\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)^2 + \tfrac{1}{N}\textstyle\sum_{i=1}^{N}\left(C_{p_i}-\overline{C_p}\right)^2\left(\left.\frac{\partial \mathrm{P}}{\partial C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)^2 \\
&\quad + \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(\rho_i-\overline{\rho})^2\left(\left.\frac{\partial \mathrm{P}}{\partial \rho}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)^2 \\
&\quad + \tfrac{2}{N}\textstyle\sum_{i=1}^{N}(v_i-\overline{v})(\rho_i-\overline{\rho})\left(\left.\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial \rho}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right) \\
&\quad + \tfrac{2}{N}\textstyle\sum_{i=1}^{N}(v_i-\overline{v})\left(C_{p_i}-\overline{C_p}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right) \\
&\quad + \tfrac{2}{N}\textstyle\sum_{i=1}^{N}(\rho_i-\overline{\rho})\left(C_{p_i}-\overline{C_p}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial v}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right) \\
&= [\Delta_v]^2\left(\left.\frac{\partial \mathrm{P}}{\partial v}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)^2 + [\Delta_\rho]^2\left(\left.\frac{\partial \mathrm{P}}{\partial \rho}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)^2 + [\Delta_{\mathrm{Cp}}]^2\left(\left.\frac{\partial \mathrm{P}}{\partial C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)^2 \\
&\quad + 2\left(\left.\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial \rho}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\Delta_{v(t)\rho} \\
&\quad + 2\left(\left.\frac{\partial \mathrm{P}}{\partial \mathrm{v}}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\Delta_{v(\mathrm{t})\mathrm{Cp}} + 2\left(\left.\frac{\partial \mathrm{P}}{\partial \rho}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\left(\left.\frac{\partial \mathrm{P}}{\partial C_p}\right|_{\overline{v},\overline{\rho},\overline{C_p}}\right)\Delta_{\rho\mathrm{Cp}}
\end{aligned}
\tag{6}
$$

Equation (6) combines the effects of individual uncertainties of wind speed, air density, and wind energy conversion coefficient. The variances (standard deviations) of wind speed, air density, and power conversion coefficient are expressed as in equation (7a).

$$
\begin{cases}
[\Delta_v]^2 = \frac{1}{N}\sum_{i=1}^{N}(v_i(t)-\overline{v})^2 \\
[\Delta_{\mathrm{Cp}}]^2 = \frac{1}{N}\sum_{i=1}^{N}\left(C_{p_i}-\overline{C_p}\right)^2 \\
[\Delta_{\rho}]^2 = \frac{1}{N}\sum_{i=1}^{N}(\rho_i-\overline{\rho})^2
\end{cases}
\tag{7a}
$$

For correlated measured variables, co-variances among the wind speed, air density, and power conversion coefficient are expressed as in equation (7b).

$$
\begin{cases}
\Delta_{\mathrm{v}\rho} = \frac{1}{N}\sum_{i=1}^{N}(v_i-\overline{v})(\rho_i-\overline{\rho}) \\
\Delta_{\rho\mathrm{Cp}} = \frac{1}{N}\sum_{i=1}^{N}(\rho_i-\overline{\rho})\left(C_{p_i}-\overline{C_p}\right) \\
\Delta_{v\mathrm{Cp}} = \frac{1}{N}\sum_{i=1}^{N}(v_i-\overline{v})\left(C_{p_i}-\overline{C_p}\right)
\end{cases}
\tag{7b}
$$

In the case of uncorrelated variables, equation (7b) shows the modifications. In this study, real-time data of wind speed and air density are independently recorded. The power conversion coefficient is a function of wind turbine blade tip speed ratio and pitch angle. According to IEC 61400-12-1 [23], no air density normalization to its actual average value is required as the average of the recorded data is in the range of 1.225 ± 0.05 kg/m^3. That is confirmed in section IV. Thus, the combined uncertainty in the wind power conversion system expressed in equation (6) can be rewritten as in equation (8).

$$
[\Delta_P]^2 = \left[\Delta_{v(\mathrm{t})}\left.\frac{\partial \mathrm{P}(v(t))}{\partial v(t)}\right|_{\overline{v},\overline{\rho},\overline{\mathrm{Cp}}(\lambda,\beta)}\right]^2 + \left[\Delta_{\mathrm{Cp}(\lambda,\beta)}\left.\frac{\partial \mathrm{P}(\mathrm{v}(t))}{\partial \mathrm{Cp}(\lambda,\beta)}\right|_{v\overline{\rho},\overline{\mathrm{Cp}}(\lambda,\beta)}\right]^2 + \left[\Delta_{\rho}\left.\frac{\partial \mathrm{P}(\mathrm{v}(t))}{\partial \rho(t)}\right|_{\overline{v},\overline{\rho},\overline{\mathrm{Cp}}(\lambda,\beta)}\right]^2
$$

$$+ 2\Delta_{\rho Cp(\lambda,\beta)} \frac{\partial P(v(t))}{\partial \rho} \frac{\partial P(v(t))}{\partial Cp(\lambda,\beta)}\Big|_{\overline{v},\overline{\rho},\overline{Cp}(\lambda,\beta)} + 2\Delta_{v(t)Cp(\lambda,\beta)} \frac{\partial P(v(t))}{\partial v(t)} \frac{\partial P(v(t))}{\partial Cp(\lambda,\beta)}\Big|_{\overline{v},\overline{\rho},\overline{Cp}(\lambda,\beta)} \quad (8)$$

It is shown that to compute the uncertainty in the output power of wind turbines, computation of partial derivatives (sensitivities) of the turbine output power concerning wind speed, air density, and power conversion coefficient is a must. These sensitivities are expressed in equation (9).

$$\begin{cases} \frac{\partial P(v(t))}{\partial v(t)} = \frac{3A\rho v^2 C_p(\lambda,\beta)}{2} \\ \frac{\partial P(v(t))}{\partial \rho} = \frac{A v^3 C_p(\lambda,\beta)}{2} \\ \frac{\partial P(v(t))}{\partial C_p(\lambda,\beta)} = \frac{\rho A v^3}{2} \\ \frac{\partial P(v(t))}{\partial v(t)} \frac{\partial P(v(t))}{\partial Cp(\lambda,\beta)} = \frac{3A^2\rho^2 v^5 C_p(\lambda,\beta)}{4} \\ \frac{\partial P(v(t))}{\partial \rho} \frac{\partial P(v(t))}{\partial Cp(\lambda,\beta)} = \frac{A^2\rho v^6 C_p(\lambda,\beta)}{4} \end{cases} \quad (9)$$

The major causes of uncertainty in the output power of wind turbines are uncertainties of weather parameters (temperature, pressure, and humidity that affect air density and wind velocity) and uncertainties of wind turbine aerodynamic parameters (wind turbine blade tip speed ratio and pitch angle that affect blade lift and drag coefficients and power conversion coefficient). Therefore, to compute the uncertainty in the wind turbine output power, first, computation of individual uncertainties in weather and aerodynamic parameters is essential as described below.

I. Computation of Uncertainties in Weather Parameters Related to Wind Energy Harvesting

Real-time annual temperature, air pressure, air humidity, and wind speed collected during the aforementioned duration are used to compute the effect of uncertainties in air density and wind speed on energy harvesting from the wind.

i. Computation of Uncertainty in Wind Speed

As discussed earlier in this paper at least three categories of uncertainties exist in this data. These are measurement uncertainty, inter-annual wind speed uncertainty, and wind shear model uncertainty. Whatever the type of uncertainty ($\Delta_{v(t)}$) it's computed using the statistical model of equation (10) on the real-time data shown in Figure 1 (c). Where $\bar{v}$ is the yearly average wind speed of i = 1, 2… N (N = 365) and daily wind speed measurement ($v_i(t)$.

$$[\Delta_{v(t)}]^2 = \frac{1}{N}\sum_i^N (v_i(t) - \bar{v})^2; \text{ for } \bar{v} = \frac{1}{N}\sum_i^N v_i(t) \quad (10)$$

ii. Computation of Uncertainty in Air Density

Air density is another main parameter that affects power harvesting from the wind. Air is composed of dry air and water steam. As the atmospheric temperature of the considered location varies, water vapor in the air, i.e. air humidity, varies too. This affects air density. According to the IEC-Wind

Turbines part-12-2 documents [13], [31], and [32] considering the effect of air humidity variations, air density is expressed as

$$\rho(T_{em}, P_a, H) = \frac{1}{T_{em}}\left[\frac{P_a}{\mathcal{R}} - \varphi_w\left(\frac{1}{\mathcal{R}} - \frac{1}{R_w}\right)H\right] \tag{11}$$

Where T_{em} is absolute temp (K) = ^{0}C+273.15, P_a is barometric pressure (Pascal), H is relative humidity between 0 and 1, $\mathcal{R}$ is the gas constant of dry air which equals 287.01 J/kg.K, and R_w is the gas constant of water vapor which equals to 461.5 J/kg.K, the vapor pressure in Pascal (φ_w) which equals to 0.0000205exp (0.0631846T). Substituting these constants and equation (11), we get

$$\rho = \rho(T_{em}, P_a, H) \quad = \frac{1}{T_{em}}\left[\frac{P_a}{287.01} - 2.6995 * 10^{-8} H.\exp(0.0632 T_{em})\right] \tag{12}$$

Due to variations in temperature, air pressure, and air humidity air density varies. There is some contribution of the measuring instruments (thermometer, barometer, and hygrometer) error in calibration or resolution. In a similar fashion in equation (6), the variation in air density can be computed as

$$[\Delta_\rho]^2 = \left[\frac{\partial\rho}{\partial T_{em}}\Big|_{(\bar{T}_{em},\overline{Pa},\overline{\mathrm{H}})}\Delta_{T_{em}}\right]^2 + \left[\frac{\partial\rho}{\partial P_a}\Big|_{(\bar{T}_{em},\overline{Pa},\overline{\mathrm{H}})}\Delta_{P_a}\right]^2 + \left[\frac{\partial\rho}{\partial H}\Big|_{(\bar{T}_{em},\overline{Pa},\overline{\mathrm{H}})}\Delta_{H}\right]^2 + 2\frac{\partial\rho}{\partial T_{em}}\frac{\partial\rho}{\partial P_a}\Big|_{(\bar{T}_{em},\overline{Pa},\overline{\mathrm{H}})}\Delta_{T_{em}P_a}$$
$$+ 2\frac{\partial\rho}{\partial T_{em}}\frac{\partial\rho}{\partial H}\Big|_{(\bar{T}_{em},\overline{Pa},\overline{\mathrm{H}})}\Delta_{T_{em}H} + 2\frac{\partial\rho}{\partial P_a}\frac{\partial\rho}{\partial H}\Big|_{(\bar{T}_{em},\overline{Pa},\overline{\mathrm{H}})}\Delta_{P_aH} \tag{13}$$

In equation (12), putting $2.6995 * 10^{-8} = \gamma_1$ and $0.0632 = \gamma_2$, the sensitivities of air density concerning temperature, pressure, and humidity is presented as in equation (14).

$$\begin{cases} \dfrac{\partial\rho}{\partial P_a} = \dfrac{1}{\mathcal{R}T_{em}} \\ \dfrac{\partial\rho}{\partial H} = \dfrac{-\gamma_1}{T_{em}} exp(\gamma_2 T_{em}) \\ \dfrac{\partial\rho}{\partial T_{em}} = \dfrac{-P_a}{\mathcal{R}{T_{em}}^2} + (1-\gamma_2 T_{em})\dfrac{\gamma_1 H}{{T_{em}}^2} exp(\gamma_2 T_{em}) \\ \dfrac{\partial\rho}{\partial T_{em}}\dfrac{\partial\rho}{\partial P_a} = \dfrac{-P_a}{\mathcal{R}^2{T_{em}}^3} + (1-\gamma_2 T_{em})\dfrac{\gamma_1 H}{\mathcal{R}{T_{em}}^3} exp(\gamma_2 T_{em}) \\ \dfrac{\partial\rho}{\partial T_{em}}\dfrac{\partial\rho}{\partial H} = \dfrac{P_a\gamma_1}{\mathcal{R}{T_{em}}^3}\exp(\gamma_2 T_{em}) - (1-\gamma_2 T_{em})\dfrac{{\gamma_1}^2 H}{{T_{em}}^3}(exp(\gamma_2 T_{em}))^2 \\ \dfrac{\partial\rho}{\partial P_a}\dfrac{\partial\rho}{\partial H} = \dfrac{-\gamma_1}{\mathcal{R}{T_{em}}^2} exp(\gamma_2 T_{em}) \end{cases} \tag{14}$$

The uncertainties in temperature, pressure and humidity in (13) are found from their measured data employing the statistical relations in (15).

$$\begin{cases} [\Delta_{\mathrm{Tem}}]^2 = \frac{1}{N}\sum_i^N (T_{em_i} - \overline{\mathrm{T}}_{em})^2;\ \text{for } \overline{\mathrm{T}}_{em} = \frac{1}{N}\sum_i^N \mathrm{T}_{\mathrm{em_i}} \\ [\Delta_{\mathrm{Pa}}]^2 = \frac{1}{N}\sum_i^N (\mathrm{Pa}_i - \overline{\mathrm{Pa}})^2;\ \text{for } \overline{\mathrm{Pa}} = \frac{1}{N}\sum_i^N \mathrm{Pa}_i \\ [\Delta_{\mathrm{H}}]^2 = \frac{1}{N}\sum_i^N (\mathrm{H}_i - \overline{\mathrm{H}})^2;\ \text{for } \overline{\mathrm{H}} = \frac{1}{N}\sum_i^N \mathrm{H}_i \end{cases} \tag{15}$$

Where $\overline{\mathrm{T}}_{em}$, $\overline{\mathrm{Pa}}$ and $\overline{\mathrm{H}}$ are averages of N measurements T_{em_i}, Pa_i and H_i which are temperature,

pressure, and humidity, respectively. Naturally, the meteorological parameters are correlated with each other. Hence the effects of one parameter on others (co-variances) are found from their measured data employing the statistical models of equation (16).

$$\begin{cases} \Delta_{\mathrm{T_{em}P_a}} = \frac{1}{N}\sum_i^N (T_{em_i} - \overline{\mathrm{T}}_{em})\,(\mathrm{Pa_i} - \overline{\mathrm{Pa}}) \\ \Delta_{\mathrm{T_{em}H}} = \frac{1}{N}\sum_i^N (T_{em_i} - \overline{\mathrm{T}}_{em})\,(\mathrm{H_i} - \overline{\mathrm{H}}) \\ \Delta_{\mathrm{P_aH}} = \frac{1}{N}\sum_i^N (H_i - \overline{\mathrm{H}})\,(\mathrm{P_{a_i}} - \overline{\mathrm{P}}_{\mathrm{a}}) \end{cases} \quad (16)$$

Where $\Delta_{\mathrm{T_{em}P_a}}, \Delta_{\mathrm{T_{em}H}}$ and $\Delta_{\mathrm{P_aH}}$ are co-variances between temperature and pressure, temperature and humidity, and pressure and humidity, respectively.

II. Computation of Uncertainties in Aerodynamic Parameters of Wind Turbine Related to Wind Energy Harvesting

The ratio of wind turbine mechanical output power to its input wind power is the power conversion coefficient. The empirical models of this coefficient ($C_P(\lambda,\beta)$) are deliberated in [33], [34], [35], and [36] in terms of turbine tip-speed ratio (λ) and rotor blade pitch angle (β). One of such models is

$$\mathrm{C_p}(\lambda,\beta) = 0.73\left(\left(\frac{151}{\lambda_i}\right) - 0.58\beta - 0.002\beta^{2.14}\right) \exp\left(-\frac{18.4}{\lambda_i}\right) \quad (17a)$$

$$\frac{1}{\lambda_i} = \frac{1}{\lambda + 0.02\beta} - \frac{0.03}{1+\beta^3} \quad (17b)$$

The variation of $C_P(\lambda,\beta)$ with λ and β is depicted [37]. As indicated in equation (8), uncertainty ($\Delta_{\mathrm{C_p(\lambda,\beta)}}$) in power conversion coefficient results in uncertainty of wind turbines' output power. It combines uncertainties in turbine blade tip speed ratio and pitch angle. This is shown in (17a). The uncertainty in $\mathrm{C_P}(\lambda, \beta)$ is expressed in (18) where $\bar{\lambda}$, and $\bar{\beta}$ are the averages of N measurements of λ_i and β_i of turbine blade tip speed ratio and pitch angle, correspondingly. Δ_λ is uncertainty in λ, Δ_β is uncertainty in β, and $\Delta_{\lambda\beta}$ is the covariance between λ and β.

$$[\Delta_{\mathrm{C_p(\lambda,\beta)}}]^2 = \left[\frac{\partial \mathrm{C_p(\lambda,\beta)}}{\partial\lambda}\Big|_{(\bar{\lambda},\bar{\beta})}\Delta_\lambda\right]^2 + \left[\frac{\partial \mathrm{C_p(\lambda,\beta)}}{\partial\beta}\Big|_{(\bar{\lambda},\bar{\beta})}\Delta_\beta\right]^2 + 2\,\frac{\partial \mathrm{C_p(\lambda,\beta)}}{\partial\lambda}\frac{\partial \mathrm{C_p(\lambda,\beta)}}{\partial\beta}\Big|_{(\bar{\lambda},\bar{\beta})}\Delta_{\lambda\beta} \quad (18)$$

While wind speed is lower than its rated value, the blade pitch angle is set to zero degrees to harvest more power from wind. In this case Δ_β and $\Delta_{\lambda\beta}$ are zero. Thus, (17a) was replaced by (19).

$$\begin{aligned} \mathrm{C_p}(\lambda, 0) &= 0.73((151/\lambda_i) - 13.2).\exp(-18.4/\lambda_i) \\ &= 0.73\left(\frac{151 - 17.73\lambda}{\lambda}\right)\exp\left(\frac{-18.4}{\lambda} + 0.552\right) \end{aligned} \quad (19)$$

As a result, (18) is replaced by (20).

$$[\Delta_{\mathrm{Cp(\lambda,0)}}]^2 = \left[\frac{\partial \mathrm{C_p(\lambda,0)}}{\partial\lambda}\Big|_{(\bar{\lambda},0)}\Delta_\lambda\right]^2 \quad (20)$$

The sensitivity of $\mathrm{C_P}(\lambda, \beta)$ of the blade tip speed ratio is

$$\frac{\partial C_p(\lambda,0)}{\partial\lambda}\Big|_{\overline{(\lambda},0)} = 0.73\left(\frac{2778.4-477.232\lambda}{\lambda^3}\right)\exp\left(\frac{-18.4}{\lambda}+0.552\right) \quad (21)$$

To compute (20), the computation of uncertainty in tip speed ratio is expressed as in (22).

$$[\Delta_\lambda]^2 = \left[\frac{\partial\lambda}{\partial\omega}\Big|_{(\bar{\omega},\bar{v})}\Delta_\omega\right]^2 + \left[\frac{\partial\lambda}{\partial v}\Big|_{(\bar{\omega},\bar{v})}\Delta_v\right]^2 + 2\frac{\partial\lambda}{\partial\omega}\frac{\partial\lambda}{\partial v}\Big|_{(\bar{\omega},\bar{v})}\Delta_{\omega v} \quad (22)$$

Where $\bar{\omega}$, and $\bar{v}$ are the averages of N measurements of ω_i and v_i of wind turbine rotor speed ($\omega(t)$) and wind speed ($v(t)$), respectively. Δ_ω and Δ_v are the variances of $\omega(t)$ and $v(t)$, respectively, and $\Delta_{\omega v}$ is co-variance of $\omega(t)$ and $v(t)$.

The mathematical model discussed in [38] that relates blade tip speed ratio to wind speed and turbine rotor speed is depicted in equation (23). The sensitivities of the tip speed ratio of wind speed and turbine rotor speed are derived as in equation (24).

$$\lambda = \omega(t)R/v(t) \quad (23)$$

$$\begin{cases} \frac{\partial\lambda}{\partial\omega(t)} = \frac{R}{v(t)} \\ \frac{\partial\lambda}{\partial v(t)} = \frac{-\omega(t)R}{v^2(t)} \\ \frac{\partial\lambda}{\partial\omega(t)}\frac{\partial\lambda}{\partial v(t)} = \frac{-\omega(t)R^2}{v^3(t)} \end{cases} \quad (24)$$

Variations in wind speed and turbine rotor speed are evaluated employing statistical tools on real-time measured data. These are

$$\begin{cases} [\Delta_\omega]^2 = \frac{1}{N}\sum_i^N(\omega_i - \bar{\omega})^2; \text{ for } \bar{\omega} = \frac{1}{N}\sum_i^N\omega_i \\ \Delta_{\omega v} = \frac{1}{N}\sum_i^N(\omega_i - \bar{\omega})(v_i - \bar{v}) \end{cases} \quad (25)$$

The co-variances among wind speed, air density, and power conversion coefficient in the model shown in equation (8) are evaluated using real-time data and computed data employing the statistical relations of equation (7a), which are rewritten as in equation (26) as $\overline{C_P}(\lambda,0) = \frac{1}{N}\sum_i^N C_P(\lambda,0)_i$.

$$\begin{cases} \Delta_{v(t)Cp(\lambda,0)} = \frac{1}{N}\sum_i^N(v_i(t) - \bar{v})(C_P(\lambda,0) - \overline{C_P}(\lambda,0)) \\ \Delta_{\rho Cp(\lambda,0)} = \frac{1}{N}\sum_{i=1}^N(\rho_i - \bar{\rho})(C_P(\lambda,0) - \overline{C_P}(\lambda,0)); \end{cases} \quad (26)$$

IV. Results

To compute uncertainties in the output power of wind turbines, sensitivities and individual uncertainties in wind energy associated with weather conditions and wind turbine aerodynamic variables are investigated and presented. Using the real-time data shown in Figure 1, first, air density was computed employing equation (12). The variation in air density because of variations in air temperature, pressure, and humidity is depicted in Figure 2. Accordingly, air density at the Adama-

II wind farm site varies between 1.18 and 1.24 kg/m^3. To compute the uncertainty in it and its sensitivities to the correlated parameters, the variance, and covariance of the measured data in Figure 1 of the weather condition parameters are required. These are computed by employing equation (15) and tabulated in Table 2. From Table 2, variation in temperature is more influential.

Table 2: *Variations in the air temperature, pressure, and humidity*

$\Delta_{T_{em}}$	Δ_{Pa}	Δ_{H}	$\Delta_{T_{em}P_a}$	$\Delta_{T_{em}H}$	Δ_{P_aH}
3.815	2.147	0.129	-0.340	-0.0116	0.119

Where temperature (T_{em}), pressure (P_a), and humidity (H) are in ^{0}C, mb, and % (a fraction), respectively. Using equation (14), the sensitivities of air density concerning temperature, pressure, and humidity are evaluated at average values ($\bar{T}_{em}$,$\bar{P}_a$, $\bar{H}$) = (18.1778 ^{0}C, 1016.333 mb, 0.4849) of the real-time data which are tabulated in Table 3. From Table 3, air density is more sensitivities to temperature. Inserting the values of Table 2 and Table 3 into equation (13), the uncertainty in air density (Δ_ρ) is equal to 0.0161.

Table 3: *Air density sensitivities to temperature, pressure, and humidity at*($\bar{T}_{em}$,$\bar{P}_a$, $\bar{h}$)

$\frac{\partial\rho}{\partial T_{em}}$	$\frac{\partial\rho}{\partial P_a}$	$\frac{\partial\rho}{\partial H}$	$\frac{\partial\rho}{\partial T_{em}}\frac{\partial\rho}{\partial P_a}$	$\frac{\partial\rho}{\partial T_{em}}\frac{\partial\rho}{\partial H}$	$\frac{\partial\rho}{\partial P_a}\frac{\partial\rho}{\partial H}$
4.44E-3	1.2E-5	-0.0092	-5.32E-8	-3.59E-5	-1.1E-07

The uncertainty in $C_P(\lambda,0)$ is the second factor that affects energy harvesting from the wind. $C_P(\lambda,0)$ of the 1.5 MW wind turbine is analytically computed using its tip speed ratio and equation (19). The result is presented in Figure 2. This figure shows the computed $C_P(\lambda,0)$ varies in the range of 0.0265 – 0.4412. The uncertainty in $C_P(\lambda,0)$ is computed employing equation (20). It depends on variations of tip speed ratio. The tip speed ratio is attuned with wind speed status to maintain rotor speed within the limits. To compute the uncertainty in tip speed ratio, first, the uncertainties in wind speed and rotor speed, and the required sensitivities were computed using real-time data as depicted in Figure 1 (c) and (d). The average values and uncertainties in wind speed, turbine rotor speed, and the covariance between these two variables are tabulated in Table 4.

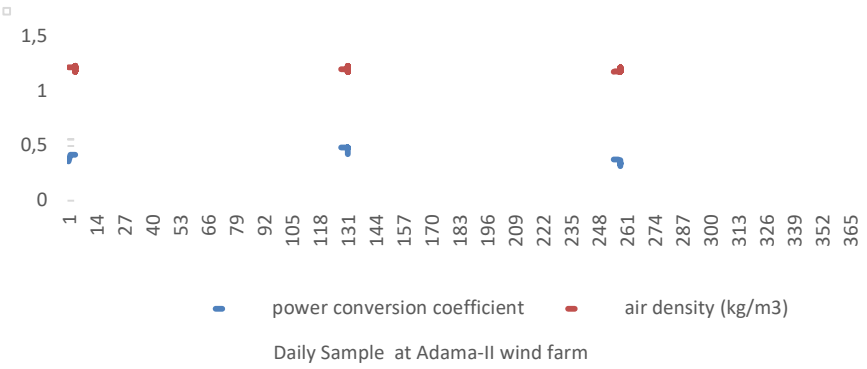


Figure 2: *Annual air density at Adama II Windfarm and power conversion coefficient of the 1.5 MW wind turbine.*

Table 4: *The average value and uncertainty in wind speed and rotor speed.*

$\bar{v}$	$\bar{\omega}$	$\Delta_{v(t)}$	Δ_{ω}	$\Delta_{\omega v}$
9.7335	1.5792	2.3895	0.3473	0.8122

Next, employing equation (22), the uncertainty and the sensitivities of tip speed ratio concerning wind speed and rotor speed are carried at average values of $(\bar{v}, \bar{\omega})$ is equal to (9.7335 m/s, 1.57922 rad/s) based on the wind speed data and the 1.5 MW wind turbine rotor speed characteristics shown in Figure 1 (c) and (d). The results are tabulated in Table 5. The results in Table 5 indicates tip speed ratio is more sensitive to turbine rotor speed.

Table 5: *Blade tip speed ratio sensitivities to rotor speed and wind speed at $(\bar{v}, \bar{\omega})$.*

$\frac{\partial\lambda}{\partial\omega}$	$\frac{\partial\lambda}{\partial v(\mathrm{t})}$	$\frac{\partial\lambda}{\partial \mathrm{v}(\mathrm{t})}\frac{\partial\lambda}{\partial\omega}$
3.8853	-0.63068	-2.45044

From the results in Tables 4 and 5 and applying equation (22), the uncertainty in turbine rotor blade tip speed ratio(Δ_λ) is equal to 0.330804. Via equation (21a), the sensitivity of $C_P(\lambda,0)$ concerning blade tip speed ratio at its average values is $\partial C_p(\lambda, 0)/\partial\lambda\big|_{(5.063,0^0)}$ which equals 0.0938. Therefore, inserting these two values in equation (20), the uncertainty in $C_P(\lambda,0)$ becomes

$$\Delta_{C_p(\lambda,0)} = ((0.330804 * 0.0938)^2)^{1/2} = 0.0325.$$

The co-variances between wind speed and power conversion coefficient, and between air density and power conversion coefficient are evaluated using data in Figure 1 and Figure 2 and the models in equation (26). The summary of the investigated uncertainties is depicted in Table 6.

Table 6: *Summary of uncertainties in wind speed, air density, power conversion coefficient, and covariance.*

$\Delta_{v(t)}$	$\Delta_{Cp(\lambda,0)}$	Δ_ρ	$\Delta_{v(t)Cp(\lambda,0)}$	$\Delta_{\rho Cp(\lambda,0)}$
2.3882	0.0325	0.0161	0.1199	9.51E-05

The sensitivities in equation (8) of wind turbines output power concerning wind speed, air density, and power conversion coefficient are evaluated at average values of $(\bar{v}, \overline{\rho}, \bar{C}_p)$ = (9.7335 m/s, 1.2111 kg/m^3, 0.4156) of the real-time data where the result is tabulated in Table 7. These are indicators of changes in output power of wind turbines at the average values of the associated parameters.

Table 7: *The sensitivities of power concerning wind speed, air density, and power conversion coefficient at $(\bar{v}, \overline{\rho}, \bar{c}_p)$.*

$\frac{\partial \mathrm{P}(v(t))}{\partial v(t)}$	$\frac{\partial \mathrm{P}(v(t))}{\partial \mathrm{C_p}(\lambda,\beta)}$	$\frac{\partial \mathrm{P}(v(t))}{\partial\rho}$	$\frac{\partial \mathrm{P}(v(t))}{\partial v(t)}\frac{\partial \mathrm{P}(v(t))}{\partial \mathrm{C_p}(\lambda,\beta)}$	$\frac{\partial \mathrm{P}(v(t))}{\partial\rho}\frac{\partial \mathrm{P}(v(t))}{\partial \mathrm{C_p}(\lambda,\beta)}$
253907	1073675	937212	6.94191E+11	8.42E+11

From Table 7, output power of wind turbines is more sensitivities to power conversion coefficient. Finally, inserting the results of Tables 6 and 7 into equation (8), the uncertain components ($\Delta P(v(t),\varrho,C_P)$) and ($P(v(t),\varrho, C_P)$) of the output power of the 1.5 MW wind turbine are evaluated at $(\bar{v}, \overline{\rho}, \bar{C}_p)$ using equation (1). The results are presented in Table 8.

Table 8: *Uncertain and certain output powers of the 1.5 MW wind turbine rotor at $(\bar{v}, \overline{\rho}, \bar{c}_p)$.*

$P(v(t),\varrho,C_P)$	1097007 W
$\Delta_P = \Delta P(v(t),\varrho,C_P)$	351042.24 W

At any values of wind speed (v(t)), air density (ρ), and wind power to mechanical power conversion coefficient (C_p), the uncertainty in output power of wind turbine can be expressed in percentage which is maximum at scale factor $|\delta| = 1$. For instance, at the average values of wind speed, air density, and power conversion coefficient ($\bar{v}, \overline{\rho}, \bar{C}_p$), it is

$$\% \text{ uncertainty in power} = \left.\frac{\Delta P(v(t), \rho, Cp)}{P(v(t), \rho, Cp)}\right|_{(\bar{v}, \overline{\rho}, \bar{C}_p)} * 100\ \%$$
$$= \frac{351042.24}{1097007} * 100\ \% = 32\ \%\ .$$

This can be rewritten as 1097007*(1 ± 0.32|δ|) watt. For δ = 0.1, the uncertainty could be 32%*0.1 which is 3.2 % above or below the $P(\bar{v}, \overline{\rho}, \bar{C}_p)$ value. In the case of δ =1, the value 32 % will be included in the annual output power and hence the output power fluctuation interval is defined as

$$745965 \le P(\bar{v}, \overline{\rho}, \bar{C}_p) \le 1448049 \text{ Watt.}$$

Similarly, the uncertainties in wind speed, air density, and power conversion coefficient represented in Table 6 are in percentage and/or relatively at the annual average values of ($\bar{v}, \overline{\rho}, \bar{C}_p$) as 24.55% or $7.3440 \le v(t) \le 12.1230$ m/s, 1.33% or $1.1950 \le \rho \le 1.2272$ kg/m^3 and 7.82% or $0.3831 \le Cp(\lambda,0) \le 0.4471$, respectively.

V. Discussion

The results obtained in this study can be compared to the IEC61400-12-1 standard [20]. Accordingly, air density uncertainty is ± 0.05 kg/m^3 around 1.225kg/m^3. That is $1.175 \le \rho \le 1.275$ kg/m^3. The air density variation result of this study is within the international standard range. Moreover, the wind turbine power conversion coefficient was indicated by the aforementioned standard as $0.03 \le Cp \le 0.45$ when the air density is 1.225kg/m^3. This shows the investigated variation of the power conversion coefficient is within the standard range. The same standard describes that a 1 MW rated wind turbine produces 396.5 kW with an uncertainty of 224.8 kW at a wind speed of 21.5 m/s, air density of 1.225 kg/m^3, and Cp of 0.03. This is equivalent to a maximum of 56.69% uncertainty in power production. Also, according to a study presented in references [21] – [23], the variability of annual total energy production from the wind is more than +/- 40%. These indicated the maximum investigated uncertainty of the power produced from the wind at the Adama wind farm site is acceptable.

The real-time wind speed data shown in Figure 1(c) is arranged in ascending order with corresponding air density and power conversion coefficients shown in Figure 2. Using these data and for δ equal to 0.1 and 1, the certain and uncertain output powers of the 1.5 MW wind turbine are computed and presented in Figure 3 and Figure 4. Figure 3 describes 3.2% uncertainty in the power captured by the 1.5 MW wind turbine at δ is 0.1. The rated output power of the wind turbine is 1.5 MW at nominal input variables (wind speed and air density). Whereas, in the case of uncertain input variables, the output power is also uncertain. For instance, at 3.2% uncertainty in the output power, it is 48 kW above or below the rated value at nominal inputs. This causes the wind energy conversion system to be overloaded or stressed. That is for a negative value of $\Delta P(v(t), \rho, Cp)|\delta|$, the output power of the wind turbine is smaller than the rated power of the turbine. Thus, the turbine output power cannot cover the demand overloading the wind turbine. This forces it to operate only under partial load or even it may shut down. This indicates the degraded output power of the wind turbine and resulting in a low return of the system. At higher wind speeds where $\Delta P(v(t), \rho, Cp)|\delta|$ is positive, it causes more power generation stressing the wind turbine. In this case, the pitch control system of the wind turbine can regulate the power to the rated value safeguarding the turbine from

damage. The worst-case uncertainty is depicted in Figure 4. That is at $\delta = 1$, the annual output power of the wind turbine is 32% uncertain. At rated inputs, it varies between 1.02 and 1.98 Mega Watts. The effect of the positive value of the uncertain component is regulated by the turbine blade pitch mechanism, but the negative value of the uncertain component results in the same effect as discussed before.

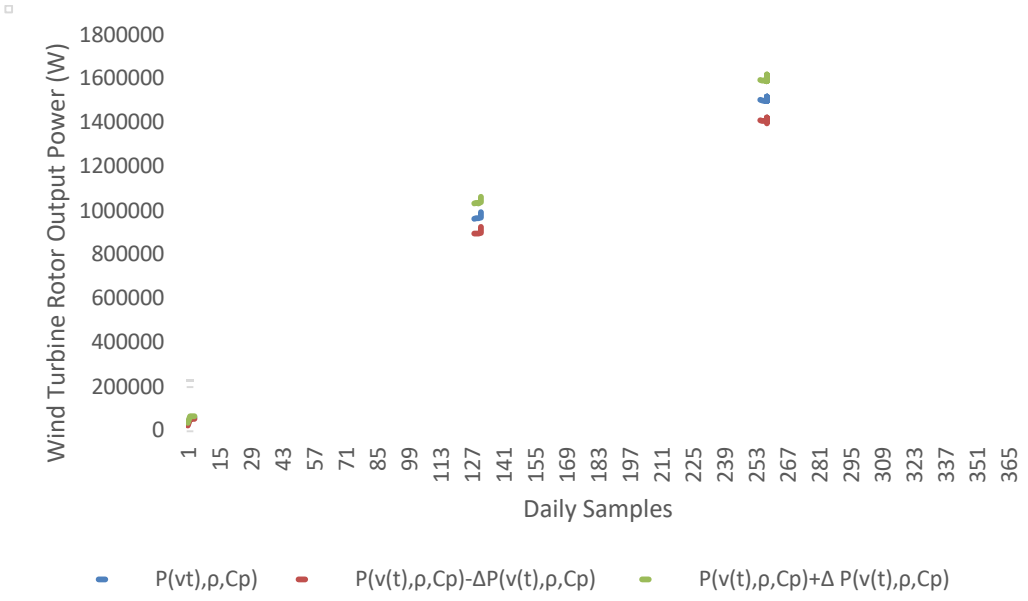


Figure 3: *Uncertainty in the 1.5 MW Wind Turbine Rotor Output Power for δ is equal to 0.1.*

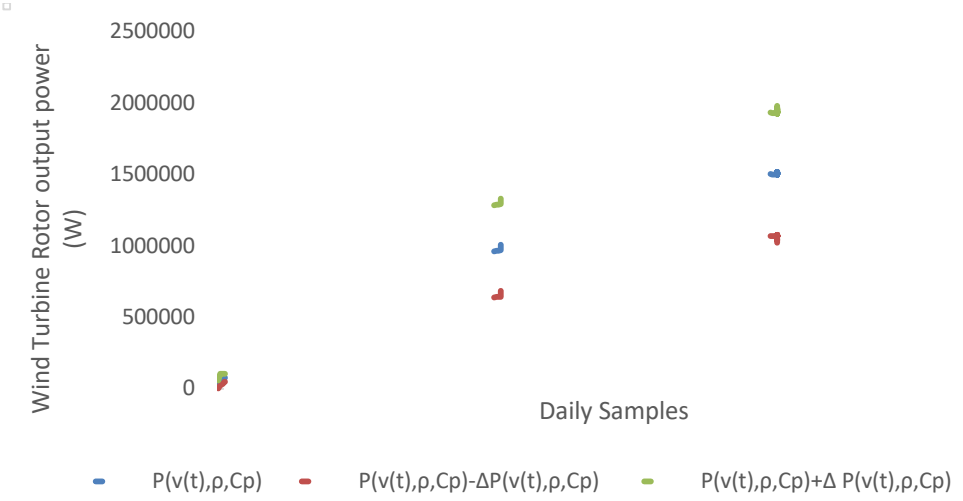


Figure 4: *Uncertainty in the 1.5 MW Wind Turbine Rotor Output Power for δ is equal to 1.*

In conclusion, mathematical modelling of wind turbine rotor output power uncertainties is done. The effect of uncertain weather parameters at Adama-II wind farm, found at Adam, Ethiopia on the wind turbine is examined. Real-time data of weather parameters of the mentioned site and the turbine rotor output power are recorded. Uncertainties in weather parameters and output power are computed according to the formulated models. The annual variation in air temperature and pressure

causes 24.55% uncertainty in the annual wind speed, 1.33% uncertainty in air density, and 7.82% uncertainty in power conversion coefficient of the wind turbine. The output power of wind turbine is more sensitive to varying power conversion coefficient. The annual uncertainty in the turbine rotor output power is 32%. The computed uncertain power is compared to the expected output power of the turbine. This uncertainty percentage in power is either added or subtracted from the expected output power. The additive case is regulated by the wind turbine blade pitch system. Due to the uncertainty effects, the turbine output power becomes unreliable and economically impacts the system's return. The effects of air density and wind speed uncertainties are natural, and hence technical regulation of fluctuation in air density may not be possible. The fluctuation in wind speed and power conversion coefficient of wind turbines can be reduced by advancing the wind turbine blade pitch system and its control techniques. This study results helps as base and guide for wind energy assessing and forecasting, wind turbine control designers and operation professionals to maintain performances of wind turbines well consistent.

References

[1] Murdock, HE. Gibb, D. André, T. Sawin, JL. Brown, A. Appavou, F. Ellis, G. Epp, B. Guerra, F. Joubert, F. Kamara, R. (2020). *Renewables 2020-Global status report, Renewable Energy Policy Network.*

[2] Shuangwen, S. and Ryan, O'. (2017). Reliability of Wind Turbines Chapter | 15. *Wind Energy Engineering*: 299-327.

[3] Samet, O. Vasilis, F. and Stefan, F. (2018). Article Failure Modes, Effects and Criticality Analysis for Wind Turbines Considering Climatic Regions and Comparing Geared and Direct Drive Wind Turbines. *Energies*, 11, 2317: 1-18.

[4] Samet, O. (2019). FORECASTING WIND TURBINE FAILURES AND ASSOCIATED COSTS. *PhD. Dissertation, COLUMBIA UNIVERSITY*: 11, 37.

[5] Simon, E. (2016). Wind Turbine Reliability Modelling. *MSc. Thesis, Reykjavik University*: 19.

[6] Maik, R. and Julio, J., M. (2016). Assessing Wind Speed Effects on Wind Turbine Reliability. *Research Centre for Energy Resources and Consumption Poster*.

[7] J.F., Manwell. J.G., McGowan and A.L., Rogers. (2010). Wind Energy Explained: Theory, Design, and Application. *John Wiley & Sons, University of Massachusetts, Amherst, USA*.

[8] G., Kosmadakis. S., Karellas. and E., Kakaras. (2013). Renewable and Conventional Electricity Generation Systems: Technologies and Diversity of Energy Systems. *In: E. Michalena, J. M. Hills, (Eds.), Renewable Energy Governance:- Complexities and Challenges, London: Springer London*: 9–30.

[9] D., Kema. (2014). Reducing Uncertainty in Wind Project Energy Estimates. *Second Wind,* www.vaisala.com

[10] Tássia, P., Pereira. Stephen, E.-O. João, P., D. Nicholas, J., W. and Americo, C., Jr. (2019). Uncertainty Quantification of Wind Turbine Wakes under Random Wind Conditions. *Proceedings of the ASME, 83501, V013T13A022.*

[11] Sten F. Rebecca, B. Sara, Pryor. Ole, Rathmann. Søren, Larsen. and Jørgen, Høj,strup. (2006). Analytical Modelling of Wind Speed Deficit in Large Offshore Wind Farms. *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology, 9,(1-2):* 39–53.

[12] Yan, Wu. Shuai, Zhang. Ruiqi, Wang. Yufei, Wang. Xiao, Feng. (2020). A design methodology for wind farm layout considering cable routing and economic benefit based on genetic algorithm and GeoSteiner. *Renewable Energy, 146.*

[13] Samira Louassa. Ouahiba, Guerri. Abdelhamid, Kaabeche. and Noureddine, Yassaa. (2019).

Effects of Local Ambient Air Temperatures on Wind Park Performance: The Case of the Kaberten Wind Park. *Taylor & Francis:* 1-14.

[14] A., Lackner. Anthony, L., Rogers. and James, F., Manwell. (2007). Uncertainty Analysis in Wind Resource Assessment and Wind Energy Production Estimation. *American Institute of Aeronautics and Astronautics:* 1222.

[15] Wiebke, Langreder. Madalina, M., Jogararu. and Sergio, A., Casta. (2016). Uncertainty of Vertical Wind Speed Extrapolation. *Brazil Wind Power 2016 Conference and Exhibition, 30.*

[16] Ronilson, Rocha. Gilmar, Alves, Coutinho. Alexandre, J., Ferreiza. and Flavio, A., Torga. (2010). Multivariable H_2 and H_∞ Control for a Wind Energy Conversion System – A Comparison. *The Brazilian Society of Mechanical Sciences and Engineering, 32*(4): 510-518.

[17] Weifei, Hu. Zhenyu, Liu. and Jianrong, Tan. (2019). Thermodynamic Analysis of Wind Energy Systems, in: Kenneth Eloghene Okedu (ed.). *Wind Solar Hybrid Renewable Energy System, DOI: 10.5772/intechopen.85067.*

[18] Akgün, Kalkan. (2015). Uncertainty in Wind Energy Assessment (Emphasis on Model Uncertainty). *INORES*.

[19] Pascal, Richter. Jannick, Wolters. and Martin, Frank. (2020). Uncertainty Quantification for the Planning of Offshore Wind Farms Using Monte Carlo and Sparse Grid. *Preprint Version.*

[20] U., Bunse. H., Mellinghoff. and O., Haack. (2007). Uncertainty of Annual Energy Production for a Specific Turbine Model Based on a Set of IEC 61400-12 Measurements. *DEWI*.

[21] Tongdan, Jin. and Zhigang, Tian. (2010). Uncertainty Analysis for Wind Energy Production with Dynamic Power Curves. *PMAPS, IEEE 11th International Conference*: 745-750. doi.10.1109/PMAPS.2010.5528405.

[22] Y.H., Wan. (2012). Long-Term Wind Power Variability. *Technical Report NREL/TP-5500-53637*.

[23] Simon, Watson. (2013). Quantifying the variability of wind energy. *Wiley Interdisciplinary Reviews: Energy and Environment*.

[24] Warren, Katzenstein. Emily, Fertig. Jay, Apt. (2010). The variability of interconnected wind plants. *Energy Policy, 38:* 4400–4410.

[25] IEC 61400-12-1, (2005). Wind turbines– Part 12-1: Power performance measurements of electricity producing wind turbines. *First Edition*.

[26] Kathryn, E., Johnson. Lucy, Y., Pao. Mark, J., Balas. and Lee, J., Fingersh. (2006). Control of Variable speed wind turbines: Standard and Adaptive Techniques for Maximizing Energy Capture. *IEEE Control System, 26 (3)*: 70-81. doi:10.1109/mcs.2006.1636311.

[27] Lucy, Y., Pao. and Kathryn, E., Johnson. (2009). A Tutorial on the Dynamics and Control of Wind Turbines and Wind Farms. *American Control Conference:* 2076-2089.

[28] Lee, Jay, Fingersh. and Palmer, W., Carlin. (1998). Results from the NREL Variable-speed Testbed. *ASME 17th Wind Energy Symposium, Reno, Nevada*: 1-7.

[29] Hamed, Bakhtiari. Jin, Zhong. and Manuel, Alvarez. (2021). Predicting the stochastic behavior of uncertainty sources in planning a stand-alone renewable energy-based microgrid using Metropolis–coupled Markov chain Monte Carlo simulation. *Applied Energy 290 (2021) 116719*.

[30] ISO/IEC GUIDE 98-3:2008(E). (2008). Uncertainty of measurement - Part 3: Guide to the expression of uncertainty in measurement (GUM: 1995), First edition.

[31] IEC 61400-12-2. (2013). Wind turbines–Part 12-2: Power Performance of Electricity-producing Wind Turbines based on Nacelle Anemometry, First edition.

[32] Rogier, Floors. and Morten, Nielsen**.** (2019). Estimating Air Density Using Observations and Re-Analysis Outputs for Wind Energy Purposes. *Energies,* 12(11). http:/**doi.org/10.3390/en12112038.**

[33] M., Sarvi. Sh., Abdi. and S., Ahmadi. (2013). A New Method for Rapid Maximum Power Point

Tracking of PMSG Wind Generator using PSO-Fuzzy Logic. *Tech Journal of Engg & App Sci, 3(17):* 1984-1995.

[34] Ranjan, Vepa. (2013). Dynamic Modeling, Simulation and Control of Energy Generation. *Lecture Notes in Energy, 20, Springer: Verlag London.*

[35] Ramji, Tiwari. and N., Ramesh, Babu. (2016). Recent Developments of Control Strategies for Wind Energy Conversion System. *Renewable and Sustainable Energy Reviews*, 66: 268-285. doi:10.1016/j.rser.2016.08.005.

[36] X., Jing. (2012). Modeling and Control of a Doubly-Fed Induction Generator for Wind Turbine-Generator Systems. *MSc Thesis, Marquette University*: 29-30.

[37] Sang, LQ,, Li, Q'a. Cai, C. Takao, M. Kamada, Y. Wang, X. Zhou, S. and Zhang, F. (2021). Wind tunnel and numerical study of a floating offshore wind turbine based on the cyclic pitch control. *Renewable Energy*, https://doi.org/10.1016/j.renene.2021.03.027.

[38] Muhammed, Y. Worku. M., A., Abido. and R., Iravani. (2017). PMSG based Wind System for Real-Time Maximum Power Generation and Low Voltage Ride Through. *Journal of Renewable Sustainable Energy*, 9 (1). DOI:10.1063/1.4976141.

RT&A, No 4 (65)
Volume 16, December 2021

A Case Study to Analyze Ageing Phenomenon in Reliability Theory

Pulak Swain[1]*, Subarna Bhattacharjee[2]†, Satya Kr. Misra[3]

•

[1] *School of Basic Sciences, IIT Bhubaneswar, Argul-752050, Odisha, India*

[2] *Department of Mathematics, Ravenshaw University, Cuttack-753003, Odisha, India*

[3] *Department of Mathematics, KIIT University, Bhubaneswar-751024, Odisha, India*

[1]pulakswain1994@gmail.com, [2]subarna.bhatt@gmail.com, [3]satyamisra05@gmail.com

Abstract

Hazard rate, and ageing intensity (AI) are measures or functions required for qualitative and quantitative analysis of ageing phenomena of a system with a well defined statistical distribution respectively. In this paper, we reiterate upon the fact that in a few cases hazard rate and ageing intensity do not depict the same pattern as far as monotonicity is concerned. So, a question naturally arises which among hazard rate, and ageing intensity is a preferable measure for characterizing ageing phenomena of a system. As a consequence, an example involving two design systems are analyzed and is illustrated to answer the aforementioned question.

Keywords: Ageing phenomenon, hazard rate, ageing intensity function.

AMS 2020 Subject Classification: Primary 60E15, Secondary 62N05, 60E05

1. Introduction

The notion of ageing phenomena and its mathematical counterpart are established by Barlow and Proschan (1975), Shaked and Shanthikumar (2007), Deshpande and Purohit (2005), Nanda et al. (2010) to name a few. The measures (or functions) usually used in this context are many, namely, survival function, hazard rate function, reversed hazard rate function, mean residual function, reversed mean residual function (cf. Block et al. (1998), Nanda et al. (2003,2005)).

Jiang et al. (2003) came forward with ageing intensity function relevant in reliability analysis. He established that the quantitative analysis of ageing phenomena for a system can be done using ageing intensity (AI) function, whereas hazard rate does the qualitative analysis.

The ageing intensity function (AI), denoted by $L_X(t)$ of a random variable X at time $t > 0$, with probability density function $f_X(t)$, survival function $\bar{F}_X(t)$ and failure rate $\lambda_X(t) =$

*The work was jointly done with the first author when he was in Ravenshaw University, Cuttack-753003, Odisha, India

†Corresponding author : E-mail: subarna.bhatt@gmail.com

$f_X(t)/\bar{F}_X(t)$ is given by (cf. Jiang et al. (2003)),

$$\begin{aligned} L_X(t) &= \frac{-tf_X(t)}{\bar{F}_X(t)\ln\bar{F}_X(t)}, \text{ where defined,} \\ &= \frac{t\lambda_X(t)}{\int_0^t \lambda_X(u)du}. \end{aligned} \tag{1.1}$$

Nanda et al. (2007) and Bhattacharjee et al. (2013), Giri et al. (2021) derive the AI function of a few distributions. Sunoj and Rasin (2017) introduce quantile-based ageing intensity function and study its various ageing properties. To learn more on ageing intensity function, one can refer to Misra and Bhattacharjee (2018), Szymkowiak (2018a,b) to name a few.

Stochastic orders play an important role in the theory of reliability as it helps in comparison of systems based on the functions, discussed in this section, namely survival function $\bar{F}(t)$, hazard rate function $\lambda(t)$, reversed hazard rate function $\mu(t)$, mean residual function $m(t)$, ageing intensity function $L(t)$ etc. giving rise to usual stochastic order (ST order), hazard rate order (HR order), reversed hazard rate order (RHR order), mean residual order (MRL order) and ageing intensity order(AI order) respectively. The stochastic orders are mathematically represented as given in the next definition.

Definition 1.1. *A random variable X is said to be smaller than another random variable Y in*

(i) usual stochastic order (denoted by $X \leq_{ST} Y$) if $\bar{F}_X(t) \leq \bar{F}_Y(t)$, for all $t \geq 0$.

(ii) hazard rate order (denoted by $X \leq_{HR} Y$) if $\lambda_X(t) \geq \lambda_Y(t)$, for all $t \geq 0$.

(iii) reversed hazard rate order (denoted by $X \leq_{RHR} Y$) if $\mu_X(t) \leq \mu_Y(t)$, for all $t \geq 0$.

(iv) mean residual life order (denoted by $X \leq_{MRL} Y$) if $m_X(t) \leq m_Y(t)$, for all $t \geq 0$.

(v) AI order (denoted by $X \leq_{AI} Y$) if $L_X(t) \geq L_Y(t)$, for all $t > 0$.

Based on the hazard rate function, an ageing class has been defined in the literature as follows.

Definition 1.2. *A random variable X is said to have increasing (decreasing) hazard rate function, denoted by $IFR(DFR)$, if $\lambda_X(t)$ is increasing (decreasing) in $t \geq 0$.*

The words 'failure rate' and 'hazard rate' have been synonymously used in this article. Throughout the article, the words increasing (decreasing) and non-decreasing (non-increasing) are used interchangeably.

Section 2 discuss the monotonic properties of failure rate and ageing intensity functions in a few statistical distributions. Section 3 simply highlights the estimator of functions appearing in this paper. Section 4 cites an example to illustrate the study of ageing phenomena through reliability function, hazard rate, reversed hazard rate and ageing intensity functions. Section 5 demonstrates the concluding remarks of the work.

2. Monotonicity of failure rate and ageing intensity functions

On the basis of the monotonicity of the AI function, Nanda et al. (2007) define ageing classes, namely increasing ageing intensity class (IAI) (decreasing ageing intensity class (DAI)) if

the corresponding AI function $L(t)$ is increasing(decreasing) in $t \geq 0$. It was pointed out that the monotonic behavior of the failure rate function is not, in general, transmitted to the monotonicity of the AI function, which is established by the following examples.

Example 2.1. *(cf. Nanda et al. (2007)) Let X has Erlang distribution, with density function with* $f_X(t) = \lambda^2 t e^{-\lambda t}, t \geq 0$. *Clearly,its failure rate function is* $r_X(t) = \lambda^2 t/(1+\lambda t)$ *which increases for* $t \geq 0$, *i.e., X has increasing failure rate (IFR). On the other hand,* $L_X(t) = \lambda^2 t^2/(1+\lambda t)(\lambda t - \ln(1+\lambda t))$, *decreases for* $t > 0$, *i.e., X is* DAI. *So, X is IFR but* DAI.

Example 2.2. *(cf. Nanda et al. (2007)) Let X be a random variable having uniform distribution over* $[a, b], 0 \leq a < b < \infty$, *i.e., Then, its failure rate* $r_X(t) = 1/(b-t), a < t < b$ *is increasing in* $t \in (a, b)$, *i.e., X is IFR. However,* $L_X(t) = t/(b-t)/\ln(b/b-t)$, *for* $a < t < b$, *is increasing in* $t, a < t < b$. *So, X is IFR and* IAI.

In the next example, we find that a random variable is DFR and DAI.

Example 2.3. *Let X be a random variable having Pareto distribution with density function or* $f_X(t) = ak^a/t^{a+1}$, *for* $t \geq k > 0$, *so that its failure rate* $r_X(t) = a/t$, *is decreasing in* $t \in (k, \infty)$. *i.e., X is DFR. However,* $L_X(t) = 1/(\ln t - \ln k)$, *is increasing in* $t \in (k, \infty)$. *Thus, X is* DFR *and* IAI.

Through these aforementioned examples, one concludes that an IFR random variable could be IAI or DAI. So, does a DFR random variable. The non-monotonic nature are also observed for some statistical distributions (cf. Nanda et al. (2007, 2013)).

Reliability analysts can obviously strive for a question, if a system (or a random variable) depicts different characteristics in terms of failure rate and ageing intensity function then which function should be used in the final conclusion of knowing the behavior of the system in terms of ageing phenomena. In this paper, we try to answer this question by giving a case study mentioned in Section 4 and analyzing it.

3. Estimator of functions

Nanda et al. (2013) gives the logical estimates of survival function $\bar{F}_X(t)$, probability density function $\bar{f}_X(t)$, hazard rate function $\lambda_X(t)$, reversed hazard rate $\mu_X(t)$ and ageing intensity function $L_X(t)$. Let n units be put to test at $t = 0$. Further, let the number of units having survived at ordered times t_j be $n_s(t_j)$. Then logical estimates of $\bar{F}_X(t), f_X(t), \lambda_X(t), \mu_X(t)$ and $L_X(t)$ for $t_j < t < t_j + \Delta t_j$, are respectively given by

$$\hat{\bar{F}}_X(t) = \frac{n_s(t_j)}{n},$$

$$\hat{f}_X(t) = \frac{n_s(t_j) - n_s(t_j + \Delta t_j)}{n\Delta t_j},$$

$$\hat{\lambda}_X(t) = \frac{\left\{n_s(t_j) - n_s(t_j + \Delta t_j)\right\}}{n_s(t_j)\Delta t_j},$$

$$\hat{\mu}_X(t) = \frac{\left\{n_s(t_j) - n_s(t_j + \Delta t_j)\right\}}{(n - n_s(t_j))\Delta t_j}.$$

Thus, logical estimate of $L_X(t)$ is

$$\hat{L}_X(t) = \frac{-t\Big\{n_s(t_j) - n_s(t_j + \Delta t_j)\Big\}}{n_s(t_j)\Delta t_j \ln \frac{n_s(t_j)}{n}},$$

for $t_j < t < t_j + \Delta t_j$.

4. An Example to illustrate the study of ageing phenomena through reliability function, hazard rate, reversed hazard rate and ageing intensity functions

A good number of life testing data can be found for analysis in Shooman (1968), Ebeling (1997) and others.

Example 4.1. *(cf. Ebeling (1997)) Fifteen units each of two different deadbolt locking mechanisms were tested under accelerated conditions until 10 failures of each were observed. The following failure times in thousands of cycles were recorded as in Table 1. Which design appears to provide the best function?*

Note that, estimator of probability density function for $t_i \leq t \leq t_{i+1}$ is

$$\begin{aligned} \hat{f}(t) &= -\frac{\hat{R}(t_{i+1}) - \hat{R}(t_i)}{(t_{i+1} - t_i)} \\ &= \frac{1}{(t_{i+1} - t_i)(n+1)} \end{aligned} \tag{4.2}$$

that of failure rate function is

$$\begin{aligned} \hat{\lambda}(t) &= \frac{\hat{f}(t)}{\hat{R}(t)} \\ &= \frac{1}{(t_{i+1} - t_i)(n+1-i))}. \end{aligned} \tag{4.3}$$

The estimator of reversed hazard rate is given by,

$$\begin{aligned} \hat{\mu}(t) &= (\hat{f}(t))/(\hat{F}(t)) \\ &= \frac{1/(t_{i+1} - t_i)(n+1)}{i/(n+1)} \\ &= \frac{1}{i(t_{i+1} - t_i)} \end{aligned} \tag{4.4}$$

Now, for the ageing intensity, it is given by,

$$\begin{aligned} \hat{L}(t) &= \frac{-t\hat{f}(t)}{\hat{\bar{F}}(t)\ln\hat{\bar{F}}(t)} \\ &= \frac{-t/(t_{i+1} - t_i)(n+1)}{\{(n+1-i)/(n+1)\}\ln\{(n+1-i)/(n+1)\}} \\ &= \frac{-t}{(t_{i+1} - t_i)(n+1-i)\ln(n+1-i)/(n+1)} \end{aligned} \tag{4.5}$$

The detailed analysis of the example considered in this Section are given in Table 2, Table 3, Table 4, Table 5 and Table 6. The Plots are also displayed in Figure 1, Figure 2, Figure 3 and Figure 4.

5. Conclusion

According to the literature on stochastic orders, we know that any system, say, here, Design-A is said to be better than design-B, if design-A has less ageing intensity, less hazard rate and higher reliability than that of design-B. The concluding remarks as noted in Table 6 at a certain interval of time are summarized as follows:

(i) Design A is better than design B in terms of the function being doubly underlined in a time interval.

(ii) Design B is better than design A on the basis of the function being singly underlined during a certain time interval.

(iii) However, the function being starred in a time interval denotes the fact that we cannot specify which among A or B is of the better design.

(iv) For example, in the interval (56.8,77], design B is better in terms of ageing intensity, whereas according to hazard rate, design A is better during (56.8,63] and design B is better in the interval (63,77]. Also, the analyzing the systems in terms of reliability reveal that, both the designs A and B have equal reliabilities during (56.8,63] but design-A is better on (63,77].

(v) It is evident that Table 6 contains more singly underlined cells than than that of doubly underlined cells.

(vi) In a nutshell, design B is more efficient than that of design A.

(vii) We attempt to identify the function which should be preferred in determining the ageing behaviour of a system.

In Table 6, one can observe that if at some interval of time the ageing intensity, hazard rate and the reliability have the same nature (either single underlined or doubly underlined) or (doubly underlined with starred) or (singly underlined with starred),then all the three measures give the same conclusion in choosing the best system design. But if one function is doubly underlined and another is singly underlined, then it gives different conclusion with regard to the performance of the systems.

(viii) For example, on the interval (56.8,63], the ageing intensity and the hazard rate show different behaviour, whereas on the interval (63,77] hazard rate and reliability show different behaviour. And on (897.8,1043.6], all the three measures show same behaviour.

(ix) Clearly, from Table 6 we can see that, hazard rate doesn't have opposite behaviour with the other two measures simultaneously. For example, on the interval (56.8,63], hazard rate shows opposite behaviour to ageing intensity function only, but not to reliability. Also, it shows opposite behaviour to reliability on (63,77], but not to the ageing intensity function in that interval. We note that, hazard rate doesn't have any doubtful situations ($\lambda_1 = \lambda_2$), which are in the case of ageing intensity or reliability at some intervals. (as, the equality sign doesn't say anything about which design is better, so these are the doubtful situations.)

Therefore, we conclude that, hazard rate should be preferred as a measure of ageing phenomena, while comparing the two systems in the problem concerned.

Table 1: *Failure Times*

Design A	44	77	218	251	317	380	438	739	758	1115
Design B	32	63	211	248	327	404	476	877	903	1416

Table 2: *Analysis of Design A*

i	t_i	$R_1(t)$	$\lambda_1(t)$	$\mu_1(t)$	$L_1(t)$
0	0	1	0.002066		
1	44	0.909	0.00303	0.022727	0.3179t
2	77	0.8182	0.000788	0.015152	0.00393t
3	218	0.7273	0.003788	0.002364	0.01189t
4	251	0.6364	0.002165	0.007576	0.00479t
5	317	0.5455	0.002646	0.00303	0.00436t
6	380	0.4546	0.003448	0.002646	0.00437t
7	438	0.3636	0.000831	0.002463	0.00082t
8	739	0.2727	0.017544	0.000415	0.0135t
9	758	0.1818	0.001401	0.005848	0.00082t
10	1115	0.0909		0.00028	

Table 3: *Analysis of Design B*

i	t_i	$R_2(t)$	$\lambda_2(t)$	$\mu_2(t)$	$L_2(t)$
0	0	1	0.002841		
1	32	0.909	0.002933	0.03125	0.0339t
2	63	0.8182	0.000614	0.016129	0.00374t
3	211	0.7273	0.002457	0.002252	0.0106t
4	248	0.6364	0.001151	0.006757	0.004t
5	327	0.5455	0.001181	0.002532	0.00357t
6	404	0.4546	0.001263	0.002165	0.00352t
7	476	0.3636	0.000227	0.001984	0.00062t
8	877	0.2727	0.003497	0.000312	0.00987t
9	903	0.1818	0.000177	0.004274	0.00057t
10	1416	0.0909		0.000195	

Table 4: *Comparison of* $R(t), \lambda(t), \mu(t)$

Time	$R_1(t)$	$R_2(t)$	Order $R(t)$	$\lambda_1(t)$	$\lambda_2(t)$	Order $\lambda(t)$	$\mu_1(t)$	$\mu_2(t)$	Order $\mu(t)$
$(0, 32]$	1	1	$R_1 = R_2$	0.002066	0.002841	$\lambda_1 < \lambda_2$	0.022727	0.03125	$\mu_1 < \mu_2$
$(32, 44]$	1	0.909	$R_1 > R_2$	0.002066	0.003226	$\lambda_1 < \lambda_2$	0.022727	0.016129	$\mu_1 > \mu_2$
$(44, 63]$	0.909	0.909	$R_1 = R_2$	0.00303	0.003226	$\lambda_1 < \lambda_2$	0.015152	0.016129	$\mu_1 > \mu_2$
$(63, 77]$	0.909	0.8182	$R_1 > R_2$	0.00303	0.000751	$\lambda_1 < \lambda_2$	0.015152	0.002252	$\mu_1 > \mu_2$
$(77, 211]$	0.8182	0.8182	$R_1 = R_2$	0.000788	0.000751	$\lambda_1 < \lambda_2$	0.002364	0.002252	$\mu_1 > \mu_2$
$(211, 218]$	0.8182	0.7273	$R_1 > R_2$	0.000788	0.003378	$\lambda_1 < \lambda_2$	0.002364	0.006757	$\mu_1 > \mu_2$
$(218, 248]$	0.7273	0.7273	$R_1 = R_2$	0.003788	0.003378	$\lambda_1 < \lambda_2$	0.007576	0.006757	$\mu_1 > \mu_2$
$(248, 251]$	0.7273	0.6364	$R_1 > R_2$	0.003788	0.001808	$\lambda_1 < \lambda_2$	0.007576	0.002532	$\mu_1 > \mu_2$
$(251, 317]$	0.6364	0.6364	$R_1 = R_2$	0.002165	0.001808	$\lambda_1 < \lambda_2$	0.00303	0.002532	$\mu_1 > \mu_2$
$(317, 327]$	0.5455	0.6364	$R_1 < R_2$	0.002646	0.001808	$\lambda_1 < \lambda_2$	0.002646	0.002532	$\mu_1 > \mu_2$
$(327, 380]$	0.5455	0.5455	$R_1 = R_2$	0.002646	0.002165	$\lambda_1 < \lambda_2$	0.002646	0.002165	$\mu_1 > \mu_2$
$(380, 404]$	0.4546	0.5455	$R_1 < R_2$	0.003448	0.002165	$\lambda_1 < \lambda_2$	0.002463	0.002165	$\mu_1 > \mu_2$
$(404, 438]$	0.4546	0.4546	$R_1 = R_2$	0.003448	0.002778	$\lambda_1 < \lambda_2$	0.002463	0.001984	$\mu_1 > \mu_2$
$(438, 476]$	0.3636	0.4546	$R_1 < R_2$	0.000831	0.002778	$\lambda_1 < \lambda_2$	0.000415	0.001984	$\mu_1 > \mu_2$
$(476, 739]$	0.3636	0.3636	$R_1 = R_2$	0.000831	0.000623	$\lambda_1 < \lambda_2$	0.000415	0.000312	$\mu_1 > \mu_2$
$(739, 758]$	0.2727	0.3636	$R_1 < R_2$	0.017544	0.000623	$\lambda_1 < \lambda_2$	0.005848	0.000312	$\mu_1 > \mu_2$
$(758, 877]$	0.1818	0.3636	$R_1 < R_2$	0.001401	0.000623	$\lambda_1 < \lambda_2$	0.00028	0.000312	$\mu_1 > \mu_2$
$(877, 903]$	0.1818	0.2727	$R_1 < R_2$	0.001401	0.012821	$\lambda_1 < \lambda_2$	0.00028	0.004274	$\mu_1 > \mu_2$
$(903, 1115]$	0.1818	0.1818	$R_1 = R_2$	0.001401	0.000975	$\lambda_1 < \lambda_2$	0.00028	0.000195	$\mu_1 > \mu_2$
$(1115, 1416]$	0.0909	0.1818	$R_1 < R_2$		0.000975				

Figure 1: *Plot of R_1 and R_2 versus time t.*

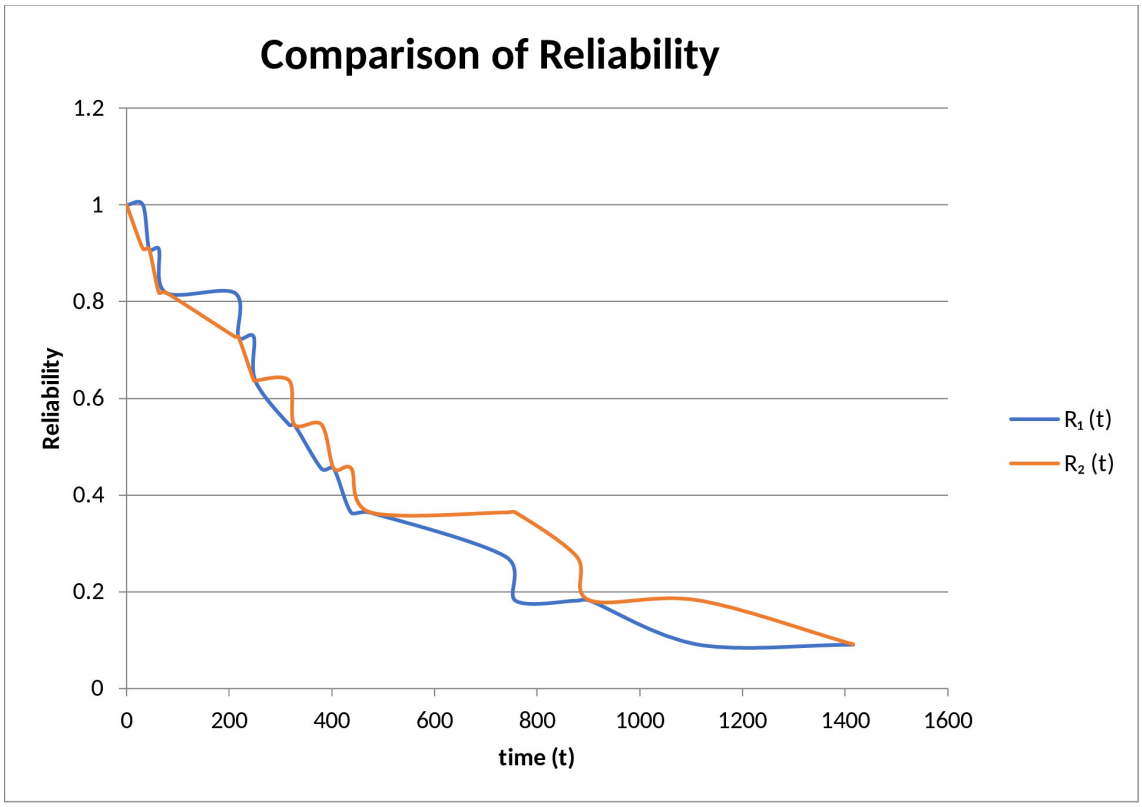


Figure 2: *Plot of HR_1 and HR_2 versus time t*

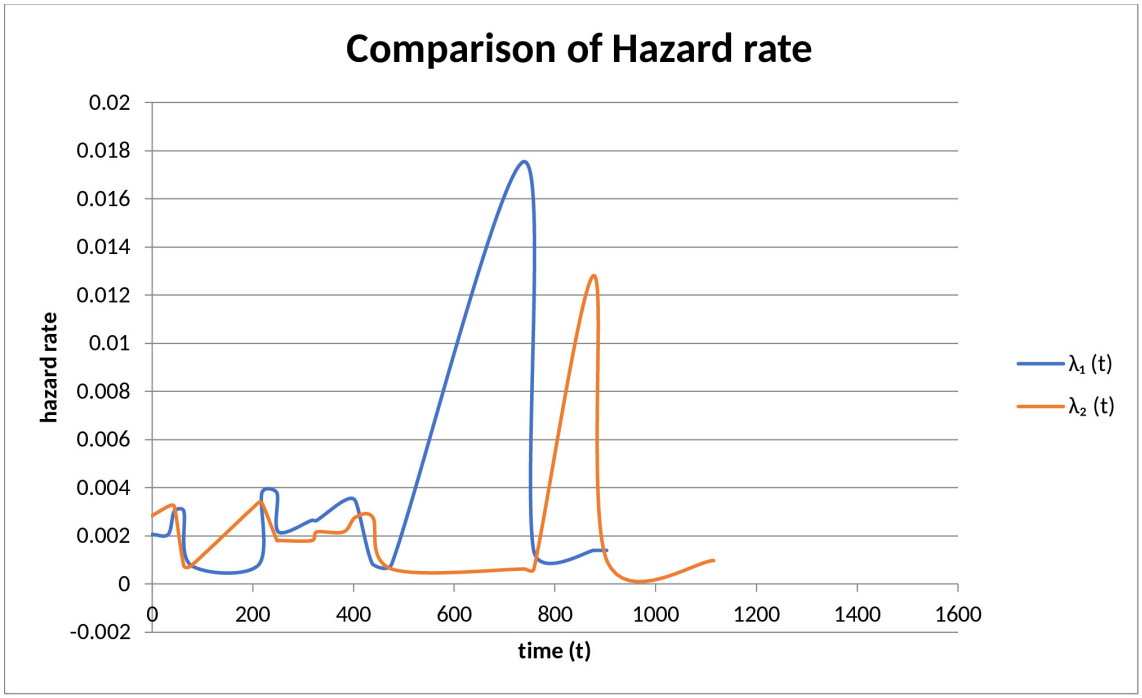


Table 5: $L_1(t)$ *and* $L_2(t)$

Design A		Design B	
t	$L_1(t)$	t	$L_2(t)$
32	1.0848	44	1.39876
38.2	1.29498	50.6	1.608574
44.4	1.50516	57.2	1.818388
50.6	1.71534	63.8	2.028202
56.8	1.92552	70.4	2.238016
63	0.23562	77	0.30261
92.6	0.346324	105.2	0.413436
122.2	0.457028	133.4	0.524262
151.8	0.567732	161.6	0.635088
181.4	0.678436	189.8	0.745914
211	2.2366	218	2.59202
218.4	2.31504	224.6	2.670494
225.8	2.39348	231.2	2.748968
233.2	2.47192	237.8	2.827442
240.6	2.55036	244.4	2.905916
248	0.992	251	1.20229
263.8	1.0552	264.2	1.265518
279.6	1.1184	277.4	1.328746
295.4	1.1816	290.6	1.391974
311.2	1.2448	303.8	1.455202
327	1.16739	317	1.38212
342.4	1.222368	329.6	1.437056
357.8	1.277346	342.2	1.491992
373.2	1.332324	354.8	1.546928
388.6	1.387302	367.4	1.601864
404	1.42208	380	1.6606
418.4	1.472768	391.6	1.711292
432.8	1.523456	403.2	1.761984
447.2	1.574144	414.8	1.812676
461.6	1.624832	426.4	1.863368
476	0.29512	438	0.35916
556.2	0.344844	498.2	0.408524
636.4	0.394568	558.4	0.457888
716.6	0.444292	618.6	0.507252
796.8	0.494016	678.8	0.556616
877	8.65599	739	9.9765
882.2	8.707314	742.8	10.0278
887.4	8.758638	746.6	10.0791
892.6	8.809962	750.4	10.1304
897.8	8.861286	754.2	10.1817
903	0.51471	758	0.62156
1005.6	0.573192	829.4	0.680108
1108.2	0.631674	900.8	0.738656
1210.8	0.690156	972.2	0.797204
1313.4	0.748638	1043.6	0.855752
1416		1115	

Table 6: *Interval-wise Study*

Interval	Compare $L(t)$	Interval	Compare $\lambda(t)$	Interval	Compare $R(t)$
$(56.8, 77]$	$\underline{L_1 > L_2}$	$(56.8, 63]$	$\underline{\underline{\lambda_1 < \lambda_2}}$	$(56.8, 63]$	$R_1 = R_2^*$
		$(63, 77]$	$\underline{\lambda_1 > \lambda_2}$	$(63, 77]$	$\underline{\underline{R_1 > R_2}}$
$(77, 211]$	$L_1 = L_2^*$	$(77, 211]$	$\underline{\lambda_1 > \lambda_2}$	$(77, 211]$	$R_1 = R_2^*$
$(211, 240.6]$	$\underline{L_1 > L_2}$	$(211, 218]$	$\underline{\underline{\lambda_1 < \lambda_2}}$	$(211, 218]$	$\underline{\underline{R_1 > R_2}}$
		$(218, 240.6]$	$\underline{\lambda_1 > \lambda_2}$	$(218, 240.6]$	$R_1 = R_2^*$
$(240.6, 248]$	$L_1 = L_2^*$	$(240.6, 248]$	$\underline{\lambda_1 > \lambda_2}$	$(240.6, 248]$	$R_1 = R_2^*$
$(248, 418.4]$	$\underline{L_1 > L_2}$	$(248, 418.4]$	$\underline{\lambda_1 > \lambda_2}$	$(248, 251]$	$\underline{\underline{R_1 > R_2}}$
				$(251, 317]$	$R_1 = R_2^*$
				$(317, 327]$	$\underline{R_1 < R_2}$
				$(327, 380]$	$R_1 = R_2^*$
				$(380, 404]$	$\underline{R_1 < R_2}$
				$(404, 418.4]$	$R_1 = R_2^*$
$(418.4, 476]$	$\underline{\underline{L_1 < L_2}}$	$(418.4, 438]$	$\underline{\lambda_1 > \lambda_2}$	$(418.4, 438]$	$R_1 = R_2^*$
		$(438, 476]$	$\underline{\underline{\lambda_1 < \lambda_2}}$	$(438, 476]$	$R_1 = R_2^*$
$(476, 636.4]$	$L_1 = L_2^*$	$(476, 636.4]$	$\underline{\lambda_1 > \lambda_2}$	$(476, 636.4]$	$R_1 = R_2^*$
$(636.4, 796.8]$	$\underline{L_1 > L_2}$	$(636.4, 796.8]$	$\underline{\lambda_1 > \lambda_2}$	$(636.4, 739]$	$R_1 = R_2^*$
				$(739, 758]$	$\underline{R_1 < R_2}$
				$(758, 796.8]$	$\underline{R_1 < R_2}$
$(796.8, 897.8]$	$\underline{\underline{L_1 < L_2}}$	$(796.8, 877]$	$\underline{\lambda_1 > \lambda_2}$	$(796.8, 877]$	$\underline{R_1 < R_2}$
		$(877, 897.8]$	$\underline{\underline{\lambda_1 < \lambda_2}}$	$(877, 897.8]$	$\underline{R_1 < R_2}$
$(897.8, 1043.6]$	$\underline{L_1 > L_2}$	$(897.8, 1043.6]$	$\underline{\lambda_1 > \lambda_2}$	$(897.8, 903]$	$\underline{R_1 < R_2}$
				$(903, 1043.6]$	$R_1 = R_2^*$

Figure 3: *Plot of RHR_1 and RHR_2 versus time t*

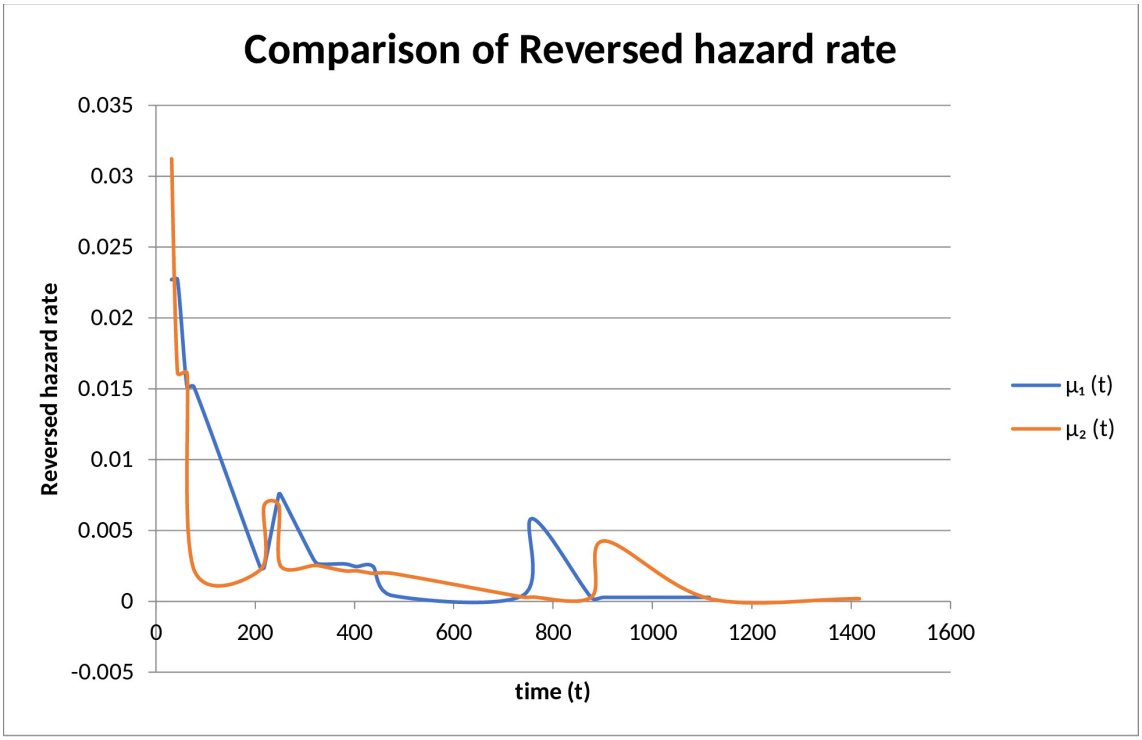


Figure 4: *Plot of AI_1 and AI_2*

Acknowledgements

The authors would like to thank the editor and the anonymous reviewers for their useful comments.

References

[1] Barlow, R.E. and Proschan, F. (1975), *Statistical Theory of Reliability and Life Testing, Probability Models.* Holt Rinehart and Winston Inc. New York.

[2] Bhattacharjee, S., Nanda A.K. and Misra, S.K. (2013), Reliability Analysis using ageing intensity function. *Statistics and Probability Letters,* 83, 1364–1371.

[3] Block, H.W., Savits, T. and Singh, H. (1998), Reversed hazard rate functions. *Probability in the Engineering and Informational Sciences*, 12, 69–90.

[4] Ebeling C.E. (1997), An Introduction to Reliability and Maintainability Engineering Michigan: Mcgraw Hill.

[5] Deshpande, J.V. and Purohit, S.G. (2005), *Life Time Data: Statistical Models and Methods.* Series in Quality, Reliability and Engineering Statistics, Vol 11, World scientific Publishing Co. Pte Ltd.

[6] Giri, R. L., Nanda, A. K., Dasgupta, M., Misra, S. Kr. and Bhattacharjee, S., On ageing intensity function of some Weibull models, *Communications in Statistics-Theory and Methods,* DOI: 10.1080/03610926.2021.1910845, Published online on 26 April 2021.

[7] Jiang, R., Ji, P., and Xiao, X. (2003), Ageing property of univariate failure rate models, *Reliability Engineering and System Safety,* 79, 113-116.

[8] Misra, S.K. and Bhattacharjee, S. (2018), A case study of ageing intensity funtion on censored data, *Alexendria Engineering Journal,* 57, 3931–3952.

[9] Nanda, A. K., Bhattacharjee, S. and Balakrishnan, N. (2010), Mean residual life fuction, associated orederings and properties, *IEEE Transactions on Reliability,* 59(1), 55-65.

[10] Nanda, A.K., Singh, H., Misra, N. and Paul, P. (2003), Reliability properties of reversed residual lifetime, *Communications in Statistics-Theory and Methods,* 32(10), 2031-2042. [Addendum: *Communications in Statistics-Theory and Methods*, (2004), 33(4), 991–992].

[11] Nanda, A.K. and Sengupta, D. (2005), Discrete life distributions with decreasing reversed hazard. *Sankhyā* 67, Part 1, 1-19.

[12] Nanda, A. K., Bhattacharjee, S., Alam, S.S. (2007), Properties of ageing intensity function, *Statistics and Probability Letters,* 77, 365–373.

[13] Nanda, A.K., Bhattacharjee, S. and Balakrishnan,N. (2010), Mean Residual Life Function, Associated Orderings and Properties, *IEEE Transactions on Reliability,* 59(1),55–65.

[14] Nanda, A.K., Bhattacharjee, S. and Alam, S.S. (2006), On Up Shifted Reversed Mean Residual Life Order, *Communications in Statistics: Theory and Methods,* 35(8), 1513–1523.

[15] Shaked, M., and Shanthikumar, J.G. (2007), Stochastic Orders. New York: Springer.

[16] Shooman, M. L. (1968), Probabilistic Reliability: An Engineering Approach. McGraw-Hill, Inc.

[17] Sunoj, S. M., and Rasin, R. S. (2018). A quantile-based study on ageing intensity function, *Communications in Statistics: Theory and Methods,* 47 (22), 5474–5484.

[18] Szymkowiak, M. (2018a), Characterizations of Distributions Through Ageing Intensity, *IEEE Transactions on Reliability,* 67, 446–458.

[19] Szymkowiak, M. (2018 b), Lifetime Analysis by Ageing Intensity Functions Studies in Systems, Decision and Control Volume 196 Series editor, Springer.

ON JOINT IMPORTANCE MEASURES FOR MULTISTATE RELIABILITY SYSTEMS

Chacko V M
•
St.Thomas College(Autonomous), Thrissur
Kerala, India
chackovm@gmail.com

Abstract

The use of importance and joint importance measures to identify the weak areas of a system and signify the roles of components in either causing or contributing to proper functioning of the system, is explained by several researchers in system engineering. But a few research outputs are available in literature for finding joint importance measures for two or more components. This paper introduces, new Joint Reliability Achievement Worth (JRAW), Joint Reliability Reduction Worth (JRRW) and Joint Reliability Fussell-Vesely measure (JRFV) for two components, of a multistate system. This is a new approach to find out the joint effect of group of components in improving system reliability. A steady state performance level distribution with restriction to the component's states is used to evaluate the proposed measures. Universal generating function (UGF) technique is applied for the evaluation of proposed joint importance measures. An illustrative example is provided

Keywords: Multistate system, reliability, joint importance measure, universal generating function.

I. Introduction

Importance and joint importance measures provides useful information to understand the system and apply reliability improvement activities. There are several importance measures available in literature, [1], [2], [3]. Interaction importance of groups of components, with respect to output performance measure(OPM)s, reliability and expected output performance is more helpful to the designers, engineers and managers to arrive at a decision, [4].

The joint importance measures of components for MSS with respect to various OPMs like reliability and expected output performance with reference to the existing measures of importance are discussed in literature in the Birnbaum sense, [5]. Research on joint importance measures for multistate systems is very useful for the researchers, [9]. But, measuring the role of interaction of components in a group consisting two components, in performance measure achievement, reduction and fractional contribution sense, is an unexplored one. In this paper, for two components of binary and MSS, Joint Reliability Achievement Worth (JRAW), Joint Reliability Reduction Worth (JRRW), and Joint Reliability Fussel-Vesely (JRFV) importance measures are introduced by considering groups with two components. JRAW measures the reliability achievement when interaction effect of two components changes from lower level to higher level, JRRW measures the reliability reduction of system when interaction effect of two components changes from higher level

to lower level and JRFV measures the fractional contribution of interaction effect of two components in improving reliability of system. These measures are generalized to the expected output performance.

A steady state performance level distribution for the system is considered for obtaining the proposed measures, [6]. The information derived by these joint importance measures allows the analyst to judge, based on their interaction effect of two components for system OPM improvement: how to give reliability operations?.

Let the components i and j are restricted with respect to performance thresholds α, and β respectively. Let $OPM_{i,j}^{\leq\alpha,\leq\beta}$, $OPM_{i,j}^{>\alpha,\leq\beta}$, $OPM_{i,j}^{\leq\alpha,>\beta}$ and $OPM_{i,j}^{>\alpha,>\beta}$ are state space restricted OPMs. If the performance measure of series system is sum performance measure of components, UGF method is found to be useful to evaluate system performance. Power generation, oil transportation systems etc are such systems.

The paper is arranged as follows. The performance measures of the MSS and new joint importance measures of two components of the binary and MSS are introduced in section II. Discussion is given in section III. Illustrative example is given in section IV. Conclusion is given in section V.

II. New Joint Importance Measures

The performance measures used for the present study are discussed below. Using the performance measure Reliability and expected output performance measure, the new joint importance measures are introduced.

I. Performance Measures of a Multistate system

A multistate system with multistate components is considered. Let the structure function of a MSS at time t be denoted by $\varphi(X(t)) = i$, $i \in \{0,1,2,\dots,M\}$, where $X(t) = (X_1(t), X_2(t), \dots, X_n(t))$, $X_i(t) \in \{0,1,2,\dots,M_i\}$, and $M = \underset{1\leq i\leq n}{max}\{M_i\}$. Let the output performance of the MSS at time t, $W(t)$, where $W(t) \in \{w_i, i = 0,1,\dots,M\}$ corresponds to the system state $\varphi(X(t)) = i$. Let

$$p_i = \lim_{t\to\infty} \Pr\{W(t) = w_i\} = \lim_{t\to\infty} \Pr\{\varphi(X(t)) = i\},\ 0\leq i \leq M.$$

Then the steady state performance distribution of the output performance of system, $\mathbf{w}=\{w_i, 0\leq i \leq M\}$ is represented by $\mathbf{p}=\{p_i, 0\leq i \leq M\}$. Steady state expected performance is

$$E(W) = \sum_{i=0}^{M} p_i w_i. \quad (1)$$

and expected system state is

$$E_s(\varphi(X)) = \sum_{i=0}^{M} i p_i. \quad (2)$$

For constant demand D_k, to state k of the multistate system, reliability is

$$R(t) = Pr\{W(t) \geq D_k\} = Pr\{\varphi(t) \geq k\}. \quad (3)$$

The stationary reliability is

$$R(D_k) = \sum_{i=0}^{M} p_i 1(w_i - D_k). \quad (4)$$

These performance measures are commonly used for reliability importance analysis, [6].

II. New Joint Importance Measures for two components in the MSS

Suppose now the components are statistically independent and reliabilities are known. In order to understand the interaction effect of two components in reliability achievement, reliability reduction and fractional contribution to reliability improvement, three joint importance measures are proposed.

Joint Reliability Achievement Worth (JRAW)

Reliability achievement worth is a measure to understand the improvement in system reliability. Consider a group of two components, with reference to the interaction, the groups having highest reliability achievement worth will be most important to improve the existing level of reliability. In order to assess the change in reliability by the presence or functioning or switching to functioning states of a group, the Joint Reliability Achievement Worth (JRAW) has to be measured.
The role of interaction of components in a group consisting *2* components, in increasing reliability of system, define the following:
ci_j^{+} , indicate i_jth component is in functioning states or up states
ci_j^{-}: indicate i_jth component is in unreliable states or down states
$I_{12} = (c1^{+} - c1^{-})(c2^{+} - c2^{-}) = (c1^{+} - c1^{-})c2^{+} - (c1^{+} - c1^{-})c2^{-} = I_{12}^{+} - I_{12}^{-}$, the contrast of interaction of the component 1 and 2, while they switch from reliable states to down states, where $I_{12}^{+} = (c1^{+} - c1^{-})c2^{+}$ is the high level interaction contrast of component 1 and 2 and $I_{12}^{-} = (c1^{+} - c1^{-})c2^{-}$ is the low level interaction contrast of component 1 and 2.
Let $\partial R_{\varphi}(i) = P\big(\varphi\big(X(t)\big) = 1, I_i^{+}\big) - P\big(\varphi\big(X(t)\big) = 1, I_i^{-}\big) = P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 1\big) - P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 0\big)$ i=1,2,…,n, the Birnbaum importance of component i, and
$\partial R_{\varphi}(i,j) = \partial R\big(\varphi\big(X(t)\big) = 1, I_{ij}^{+}\big) - \partial R\big(\varphi\big(X(t)\big) = 1, I_{ij}^{-}\big) = \big[P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 1, X_j(t) = 1\big) - P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 0, X_j(t) = 1\big)\big] - [P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 1, X_j(t) = 0\big) - P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 0, X_j(t) = 0\big)]$, joint Birnbaum joint importance of components *i* and *j.*

Now define JRAW of two components.

Let
$R_{\{i+,j+\}} = P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 1, X_j(t) = 1\big)$,
$R_{\{i-,j+\}} = P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 0, X_j(t) = 1\big)$,
$R_{\{i+,j-\}} = P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 1, X_j(t) = 0\big), and$
$R_{\{i-,j-\}} = P\big(\varphi\big(X(t)\big) = 1, X_i(t) = 0, X_j(t) = 0\big)$.

$$JRAW = \frac{Maximum\ Reliability\ due\ to\ high\ level\ interaction\ effect\ of\ two\ components}{The\ present\ reliability\ level.}$$

$$JRAW_{i,j} = \frac{\left[R_{\{i+,j+\}} - R_{\{i-,j+\}}\right]}{R}$$

The $JRAW_{i,j}$ measure quantifies the maximum possible achievement of reliability due to interaction effect of component *i* and, *j* which switches from lower level to higher level. For *i*th multistate component with performance threshold α, let $k_{i\alpha}$ be the state in the ordered set of states of component *i* such that $x_{ik_{i\alpha}} \leq \alpha < x_{ik_{i\alpha}+1,}$, [6]. For a constant demand D_k, to define Multistate Joint

Reliability Achievement Worth (MJRAW) of components i and j, let,

$R_{\{i\geq\alpha,j\geq\beta\}} = P\left(\varphi(X(t)) \geq k, X_i(t) \geq x_{ik_{i\alpha}}, X_j(t) \geq x_{jk_{j\beta}}\right)$,

$R_{\{i<\alpha,j\geq\beta\}} = P\left(\varphi(X(t)) \geq k, X_i(t) < x_{ik_{i\alpha}}, X_j(t) \geq x_{jk_{j\beta}}\right)$,

$R_{\{i\geq\alpha,j<\beta\}} = P\left(\varphi(X(t)) \geq k, X_i(t) \geq x_{ik_{i\alpha}}, X_j(t) < x_{jk_{j\beta}}\right)$,

and

$R_{\{i<\alpha,j<\beta\}} = P\left(\varphi(X(t)) \geq k, X_i(t) < x_{ik_{i\alpha}}, X_j(t) < x_{jk_{j\beta}}\right)$

where α is the performance threshold and $x_{ik_{i\alpha}}$ performance of component i in state $k_{i\alpha}$, β is the performance threshold and $x_{jk_{j\beta}}$is the performance of component j in the state $k_{j\beta,}$ $i, j, =1,2,\ldots,n$. Thus, MJRAW of two components i and j can be defined as,

$$MJRAW_{i,j} = \frac{\left[R_{\{i\geq,j\geq\}} - R_{\{i<,j\geq\}}\right]}{R}$$

MJRAW measures the reliability achievement worth of interaction effect of two components.

Joint Reliability Reduction Worth (JRRW)

To measure the role of interaction effect of two components in reducing the present reliability, Joint Reliability Reduction Worth (JRRW) is introduced in this section. To examine how the decrease in reliability happens by the decreased level or low level of interaction effect of two components, JRRW can be defined as follows.

Let

$$R_G^- = The\ decreased\ reliability\ level\ by\ the\ low\ level\ \ interaction\ of\ two\ omponents$$

and $R_0 = Present\ reliability\ level.$ The JRRW of a module is defined as:

$$JRRW = \frac{R_0}{R_G^-}$$

JRRW of two binary components i and j is

$$JRRW = \frac{Present\ Reliability\ Level}{Reliability\ when\ interaction\ of\ two\ components\ is\ at\ low\ level}$$

$$JRRW_{i,j} = \frac{R}{\left[R_{\{i+,j-\}} - R_{\{i-,j-\}}\right]}$$

The $JRRW_{i,j}$ measure of two components i and j, quantifies the maximum possible reduction of reliability due to low level of interaction effect of component i and j. For a constant demand D_k, Multistate Joint Reliability Reduction Worth (MJRRW) of a module consisting of two components i and j is defined as,

$$MJRRW_{i,j} = \frac{R}{\left[R_{\{i\geq\alpha,j<\beta\}} - R_{\{i<\alpha,j<\beta\}}\right]}$$

MJRRW measures the reliability reduction worth of interaction effect of two components *i and j.*

Joint Reliability Fussel-Vesely (JRFV) Measure

To measure the fractional contribution of interaction effect of components to the increase of reliability, Joint Reliability Fussel-Vesly (JRFV) measure can be defined. JRFV measure can be expressed as, $JRFV = \frac{R_0 - R_G^-}{R_0}$.

$$JRFV = \frac{Present\ Reliability\ Level - Reliability\ when\ interaction\ of\ two\ components\ is\ in\ low\ level}{Present\ Reliability\ Level}$$

$$JRFV_{i,j} = \frac{R - \left[R_{\{i+,j-\}} - R_{\{i-,j-\}}\right]}{R}$$

The $JRFV_{i,j}$ measure of two components i and j, quantifies the maximum fractional contribution of reliability due to high level of interaction effect of component i and j. For a constant demand D_k,

Multistate Joint Fussel-Vesly (MJRFV) of two components i and j is defined as,

$$MJRFV_{i,j} = \frac{R - [R_{\{i^{\geq\alpha},j<\beta\}} - R_{\{i<\alpha,j<\beta\}}]}{R}.$$

MJRFV measures the reliability FV of a module consisting of two components.

For the expected output performance measure, define Multistate Joint Output Performance Measure Achievement Worth (MJOPMAW), Multistate Joint Output Performance Measure Reduction Worth (MJOPMRW) and Multistate Joint Output Performance Measure Fussel-Vesely (MJOPFV) measures as below. For two components *i* and *j*,

$$MJOPMAW_{i,j} = \frac{[OPM_{\{i\geq,j\geq\}} - OPM_{\{i<,j\geq\}}]}{OPM}$$

$$MJOPMRW_{i,j} = \frac{OPM}{[OPM_{\{i^{\geq\alpha},j<\beta\}} - OPM_{\{i<\alpha,j<\beta\}}]}$$

$$MJOPMFV_{i,j} = \frac{OPM - [OPM_{\{i^{\geq\alpha},j<\beta\}} - OPM_{\{i<\alpha,j<\beta\}}]}{OPM}$$

A component's performance restriction approach can be adopted for the computation of the joint importance measures and UGF method can be adopted for the evaluation procedure, [6], [7]. The coefficients of UGFs are used for the evaluation of joint importance measures, [8].

III. Discussion

In binary and multistate context, the proposed measures quantify the RAW, RRW and FV measures of interaction effect of two components. Many of the complex systems are made up of two or more components. MJRAW measures the reliability achievement when interaction effect of two components changes from lower level to higher level, MJRRW measures the reliability reduction of system when interaction effect of two components changes from higher level to lower level and MJRFV measures the fractional contribution of interaction effect of two components. Using the information of MJRAW, it is easy to understand and identify the pair of components with highest contribution to system reliability improvement. MJRRW provides the information regarding the group which induce lowest reduction in system reliability with lower level of group performance. The fractional contribution in reliability improvement of a pair of components can be measured using MJRFV. MJOPMAW, MJOPMRW and MJOMPFV measures are useful when a researcher uses output performance measure, expected output performance measure.

IV. Illustrative Example

Consider a system made up of n = 3 multi-state components in series logic. Component states are 0, 1, 2, 3 and 4, with corresponding values of performance x_{j0}=0, x_{j1}=25, x_{j2}=50, x_{j3}=75, x_{j4}=100, j=1, 2, 3,4 (see Figure 1).

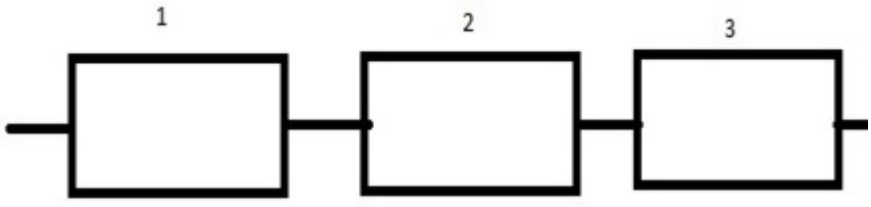


Figure 1: Series system

The probability distribution of component j in state k, p_{jk}, is given in Table 1. Let 0, 1 and 2 are unreliable states for $< \alpha$ or $< \beta$ and 3 and 4 ate reliable states for $\geq \alpha$ or $\geq \beta$.

Table 1. Probability distributions of components 1, 2 and 3.

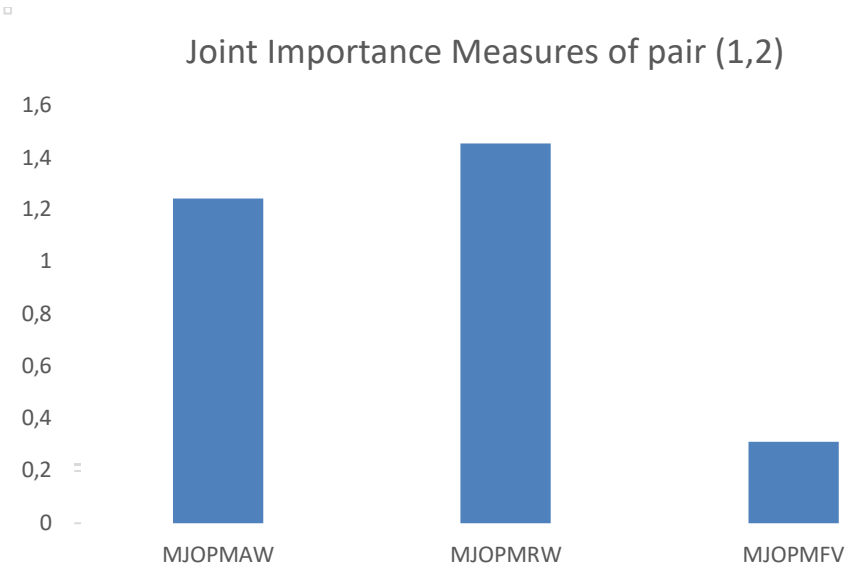
Probability Distribution	1		2		3	
P(Xi0=0)	p10	=0.1	p20	=0.15	p30	=0.4
P(Xi1=25)	p11	=0.1	p21	=0.2	p31	=0.1
P(Xi2=50)	p12	=0.5	p22	=0.3	p32	=0
P(Xi3=75)	p13	=0.2	p23	=0.2	p33	=0.1
P(Xi4=100)	p14	=0.1	p24	=0.15	p34	=0.4

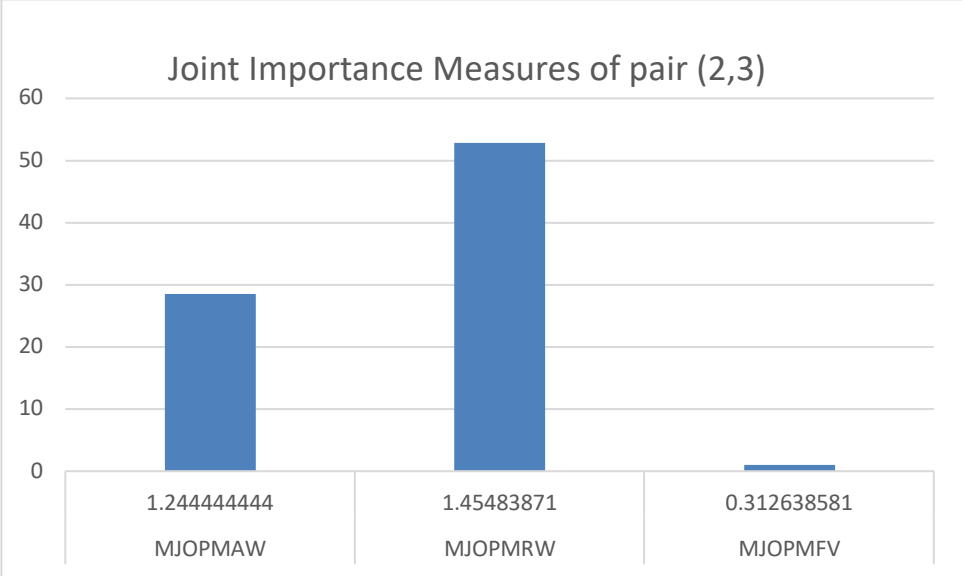
Table 2. Multistate joint importance measures

For components 1, 2	For components 2, 3
MJOPMAW= 1.244444444	MJOPMAW=28.52525253
MJOPMRW= 1.45483871	MJOPMRW= 52.8
MJOPMFV =0.312638581	MJOPMFV =0.981060606

Multistate joint importance measures are given in Table 2 and plotted in Figure 2. The sign and size of the value of joint importance measure with regard to their impact on expected system output performance are found to be different. So, a numerical comparison can be made.

Consider two groups, Group 1 with components 1 and 2 and Group 2 with components 2, and 3. Highest values for MJOPMAW, MJOPMRW and MJOPMFV are attained for pair of components 2 &3. Highest values of joint importance measures are due to highest influence of those groups in change of system reliability.





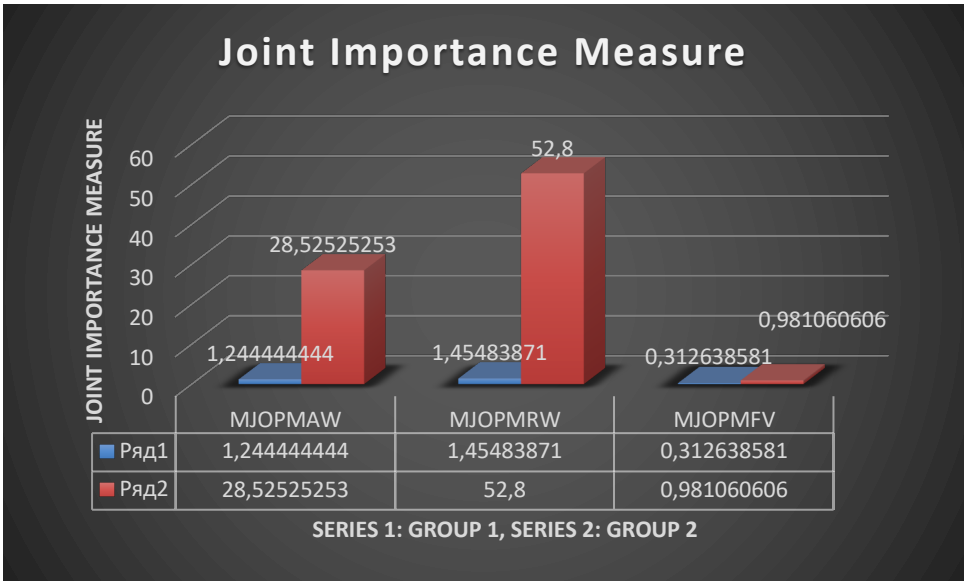


Figure 2: Multistate joint importance measures of Group 1 and Group 2

This information can be used to provide more reliability operations for different pair of components. Highest values in various importance measures indicates the need of highest care in reliability operations. To understand the dynamics of system reliability change, one can use the proposed importance measures.

V. Conclusion

This paper introduced three module joint importance measures for MSSs with reference to the OPMs reliability and expected system output performance. The joint importance measures MJRAW, MJRRW, and MJRFV for two components are introduced and generalized to expected output performance measure. The new joint importance measures are useful for giving priority for reliability improvement activities. The UGF method is used to evaluate the joint importance measures, in which the system performance is measured in terms of productivity or capacity. Joint importance measure values provide useful information for reliability improvement activities. The value and size of the importance measure can be used to make a comparison between different groups.

Acknowledgement: The research is supported by 'Santhome Research Grant(Seed Money)'

References

[1] Figueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330.
[2] Joachims J. Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms, Kluwer, 2002.
[1] Barlow, R. E. and Proschan, F. (1975), Statistical theory of reliability and Life testing, NewYork: Holt, Rinheart & Winston.
[2] Barlow, R. E., and Wu, A. (1978), Coherent system with multistate components, *Mathematics and Operations Research*, 3:275-281.
[3] Bueno, V. C. (1989), On the importance of components for multistate monotone systems, *Statistics and Probability Letters*, 7:51-59.
[4] Chacko, V. M. and Manoharan, M. (2011), Joint Importance measures for multistate reliability system, *Opsearch*, 48:257-278.
[5] Chacko, V. M. (2020), New Joint Importance Measures for Multistate Systems, *International Journal of Statistics and Reliability Engineering*, 7:140-148.
[6] Chacko, V. M. (2021), On Birnbaum Type Joint Importance Measures for Multistate Reliability Systems, *Communications in Statistics - Theory and Methods*, doi.10.1080/03610926.2021.1961000
[7] Chacko, V. M. (2022), On Joint Importance Measures for Multistate System's Reliability, *Operations Research: Methods, Techniques, and Advancements*, Chapter 9, CRC Press. Taylor & Francis Group (To Appear)
[8] Ushakov, I. (1986), A universal generating function,, *Soviet J. Comp. Syst. Sci*., 24:37-39.
[9] Wu, S. (2005), Joint importance measures of multistate systems, *Computers and Industrial Engineering*, 49:63–75.

RT&A, No 4 (65)
Volume 16, December 2021

POWER - EXPONENTIAL GEOMETRIC QUANTILE FUNCTION

Jeena Joseph[1] and Asisha A P[2]

•

Department of Statistics
St. Thomas' Collage (Autonomous), Thrissur, India

sony.jeena@gmail.com, ashapavithran28@gmail.com

Abstract

In this article, we introduced a new quantile function which is the sum of quantile functions of Power and Exponential geometric distributions. Different distributional charactaristics and reliability properties are discussed and also simulation study is conducted by using R software. Finally the new model is applied to a real data set.

Keywords: Exponential geometric distribution; Hazard quantile function; L - moments; Mean residual quantile function; Percentile residual quantile function; Power distribution; Reversed hazard quantile function; Reversed mean residual quantile function.

1. Introduction

Reliability analysis can be done by using distribution functions or by using quantile functions, although both convey the same information about the distribution with different interpretations. In reliability analysis, quantile based methods are particularly useful. For a nonnegative random variable X with distribution function F(x), the quantile function Q(u) is defined by (see Nair and Sankaran(2009))

$$Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\},\ 0 \leq u \leq 1 \tag{1}$$

If f(x) is the probability density function of X, then f(Q(u)) is called the density quantile function. If F(x) is right continuous and strictly increasing, we have,

$$F(Q(u)) = u$$

The derivative of Q(u) is known as the quantile density function of X and is denoted by q(u).i.e.,

$$q(u) = Q'(u) \tag{2}$$

When f(x) is the probability density function (pdf) of X, then by taking the derivative of F(Q(u))=u we get,

$$q(u)f(Q(u)) = 1$$

Quantile functions have several properties that are not shared by distribution functions. For example, the sum of two quantile functions is again a quantile function. Further, the product of two positive quantile functions is again a quantile function in the nonnegative setup. There are explicit general distribution forms for the quantile function of order statistics. It is easier to generate random numbers from the quantile function. A major development in portraying quantile functions to model statistical data is given by Hastings et al. (1947), who introduced a family of distributions by a quantile function. This was refined later by Tukey (1962) to form a symmetric

distribution, called the Tukey lambda distribution. This model was generalized in different ways, referred as lambda distributions which include various forms of quantile functions discussed by Ramberg and Schmeiser (1972), Ramberg (1975), Ramberg et al. (1979) and Freimer et al. (1988).

Govindarajulu (1977) introduced a new quantile function by taking the weighted sum of quantile functions of two power distributions. Hankin and Lee (2006) presented a new Power - Pareto distribution by taking the product of power and Pareto quantile functions. Van Staden and Loots (2009) developed a four - parameter distribution, using a weighted sum of the generalized Pareto and its reflection quantile functions. Sankaran et al. (2016) developed a new quantile function based on the sum of quantile functions of generalized Pareto and Weibull quantile functions. Sankaran and Dileep (2016) introduced a new class of quantile functions by taking the sum of quantile functions of half logistic and exponential geometric distributions which is useful in reliability analysis. Also in (2018), they introduced another class of quantile function by taking the product of quantile functions of Pareto and Weibull distributions. The density and distribution functions for these models are not available in closed forms except for certain special cases. The great advantage of these models is that the simple forms of the quantile functions make it extremely straightforward to simulate random values, which is useful in inference problems.

The power exponential geometric quantile function is derived by taking the sum of quantile functions of power and exponential geometric distributions. The survival function and quantile function of power distribution are respectively given by,

$$S(x) = 1 - \left(\frac{x}{\alpha}\right)^{\beta}, 0 < x < \alpha \text{ and } \alpha > 0, \beta > 0 \tag{3}$$

and

$$Q_1(u) = \alpha u^{\frac{1}{\beta}}, 0 < u < \alpha \text{ and } \alpha > 0, \beta > 0 \tag{4}$$

Adamidis and Loukas (1998) introduced the exponential geometric (EG) distribution with applications to reliability modelling in the context of decreasing failure rate data. The survival function and quantile function of the EG distribution are given by,

$$S(x) = 1 - F(x) = (1-P)e^{-\frac{x}{\alpha}}\left(1 - Pe^{-\frac{x}{\alpha}}\right)^{-1}, 0 < P < 1 \text{ and } \alpha, \lambda > 0 \tag{5}$$

and

$$Q_2(u) = \frac{1}{\lambda}\, log\left(\frac{1-Pu}{1-u}\right), 0 < P < 1 \text{ and } \alpha > 0\ \lambda > 0 \tag{6}$$

The rest of the paper is designed as follows. In section 2, we define the Power Exponential geometric (PEG) Quantile function and the members of this family are discussed in section 3. Distributional characteristics are studied in section 4, L - moments in section 5 and density function of r^{th} order statistic in section 6. In section 7, reliability properties like hazard quantile function, mean residual quantile function, percentile residual quantile function, etc. are studied. A simulation study is conducted in section 8 and concluded in section 9.

2. Power – Exponential Geometric (PEG) Quantile Function

We introduce a new quantile function, which is the sum of quantile functions of power and exponential geometric distributions.

let X and Y be two nonnegative random variables with distribution functions F(x) and G(x) with quantile functions $Q_1(u)$ and $Q_2(u)$, respectively . Then

$$Q(u) = Q_1(u) + Q_2(u) \tag{7}$$

is also a quantile function (see Nair et al. (2013)). We now introduce a new quantile function,

$$Q(u) = \alpha u^{\frac{1}{\beta}} + \frac{1}{\lambda}\, log\left(\frac{1-Pu}{1-u}\right),\ \alpha, \beta,\ \lambda > 0 \text{ and } 0 < P < 1. \tag{8}$$

is the sum of (4) and (6). The support of the new model is $(0, \infty)$. The quantile density function is obtained as,

$$q(u) = \frac{\beta(1-P) + \lambda(1-u-Pu+Pu^2)\alpha u^{\frac{1}{\beta}-1}}{\beta\lambda(1-u-Pu-Pu^2)},\ \alpha, \beta,\ \lambda > 0 \text{ and } 0 < P < 1. \tag{9}$$

For the PEG quantile function, the density function f(x) can be written in terms of the distribtion as,

$$f(x) = \frac{\beta\lambda(1-F(x)-PF(x)+P(F(x))^2)}{\beta-\beta P+\lambda\alpha F(x)^{\frac{1}{\beta}-1}(1-F(x)-PF(x)+P(F(x))^2)}, \alpha > 0, \beta > 0,\ \lambda > 0\ ,\ 0 < P < 1. \tag{10}$$

The quantile function (8) represents a family of distributions which have various shapes for different values of parameters. The shapes of density function for different values of parameters are given below.

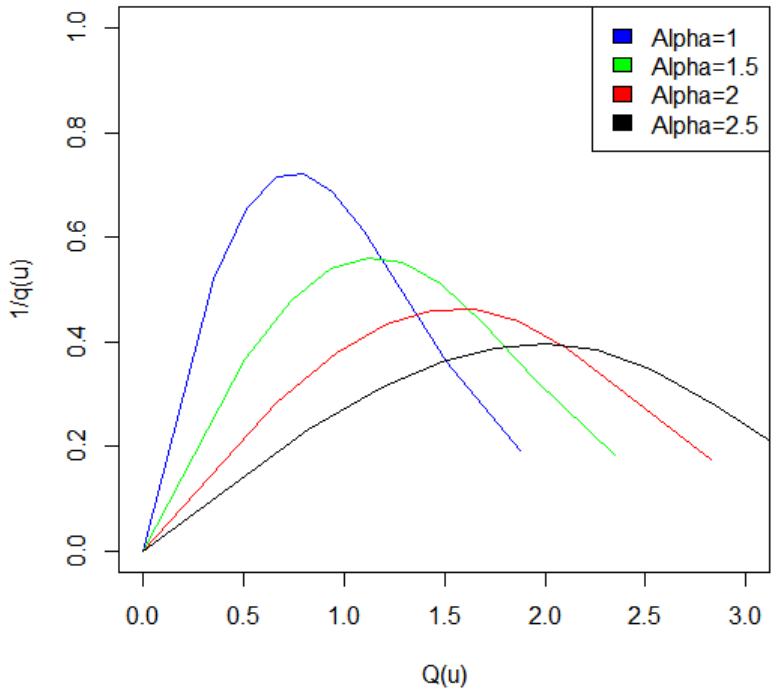


Figure 1: *Plot of density function for different values of α with β= 2, λ=2 and P=0.4.*

.

3. Members of The Family

The PEG quantile function includes several well - known quantile functions for various values of the parameters. We can derive some well - known quantile functions from the proposed model by making use of various transformations.

Case 1 : $\alpha = 0$, $\lambda > 0$ and P = 0

The quantile function of the PEG model reduces to the form

$$Q(u) = \frac{1}{\lambda}\ (-log\,(1-u)) \tag{11}$$

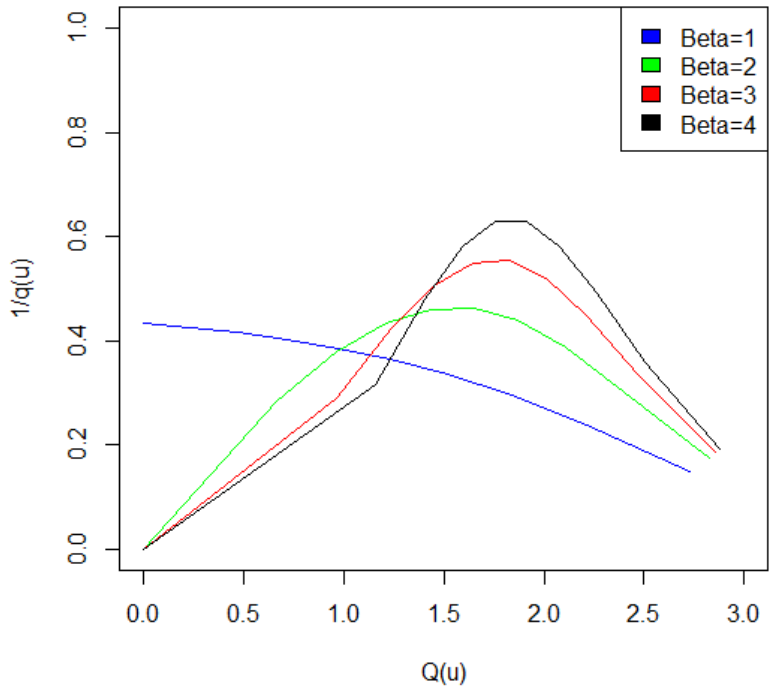


Figure 2: *Plot of density function for different values of β with α= 2, λ=2 and P=0.4.*

.

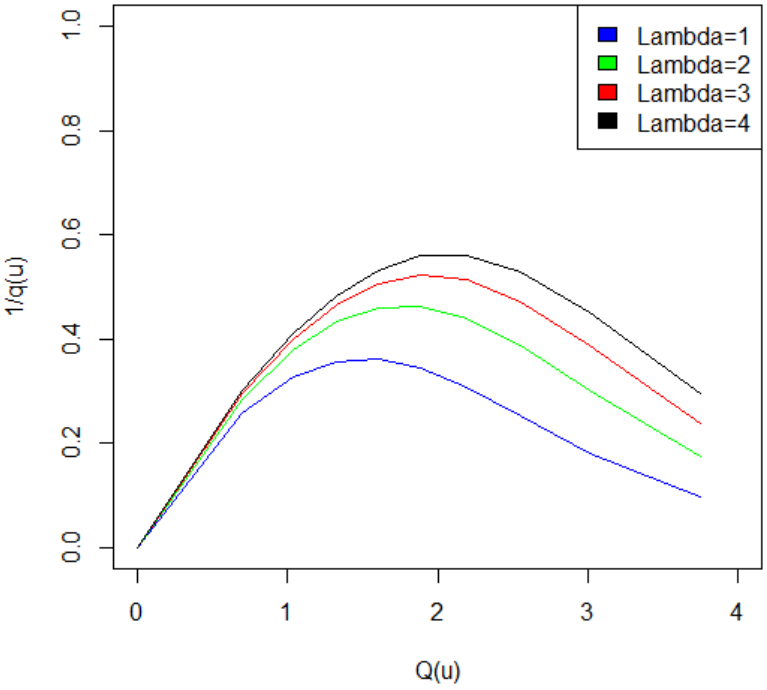


Figure 3: *Plot of density function for different values of λ with α= 2, β=2 and P=0.4.*

.

which is the quantile function of exponential distribution with mean $\frac{1}{\lambda}$.

Case 2 : $\alpha = 0$, $\lambda > 0$ and $0 < P < 1$.
Then the corresponding quantile function is,

$$Q(u) = \frac{1}{\lambda}\ log\left(\frac{1-Pu}{1-u}\right) \tag{12}$$

which belongs to the class of distributions with linear hazard quantile functions defined by Midhu et al. (2014), with quantile function

$$Q(u) = \frac{1}{a(1+\theta)}\ log\left(\frac{1+\theta u}{1-u}\right) \tag{13}$$

where θ = - P, $-1 < \theta < 0$ and λ = a(1 - P)

Case 3 : $\alpha > 0$, β = 1, $0 < P < 1$ and λ tends to ∞
The quantile function of the PEG model is reduced to,

$$Q(u) = \alpha u \tag{14}$$

which is the quantile function of uniform U(0,$\frac{1}{\alpha}$)

Case 4 : We can apply the power transformation on (11) with $\alpha = 0$, $\lambda > 0$ and P = 0 to form the quantile function of Weibull distribution with parameters $\frac{1}{\lambda}$ and K.

$$Q(u) = \frac{1}{\lambda}\ \left(-log\,(1-u)\right)^K \tag{15}$$

where K is the power.
There are some theorems that are applicable in PEG quantile function.

Theorem 1. If X follows Power distribution with distribution function $F_X(x) = \left(\frac{x}{\alpha}\right)^\beta$; $0 \leq$ X $\leq\alpha$, $\beta > 0$, then the random variable Z = X + $\frac{1}{\lambda}log\left(\frac{\alpha^\beta - PZ^\beta}{\alpha^\beta - Z^\beta}\right)$ will follow PEG(α,β,λ,P) distribution.

Proof: Let T and V be the two random variables with $Q_T(u)$ and $Q_V(u)$ be the corresponding quantile functions and $F_T(x)$ and $F_V(x)$ be the corresponding distribution functions respectively. Now suppose Q*(u) is defined by $Q_T(u) + Q_V(u)$.
Then the random variable that corresponds to the quantile function Q*(u) is $T + Q_V(F_T(T))$ or $V + Q_T(F_V(V))$ (Sankaran et al. 2016). Now let Y follow exponential geometric distribution with distribution function $F_Y(x) = (1 - e^{-\lambda x})(1 - Pe^{-\lambda x})^{-1}$ and X follows power distribution with distribution function $F_X(x) = \left(\frac{x}{\alpha}\right)^\beta$ then X + $Q_Y(F_X(X))$ has PEG(α, β, λ, P) distribution. Since $Q_Y(u)$= $\frac{1}{\lambda}log\left(\frac{1-Pu}{1-u}\right)$ and $F_X(x) = \left(\frac{x}{\alpha}\right)^\beta$, we get,

$$X + Q_Y(F_X(X)) = X + \frac{1}{\lambda}log\left(\frac{\alpha^\beta - PZ^\beta}{\alpha^\beta - Z^\beta}\right)$$

Hence the proof.

Theorem 2. Let Z follows EG ($\frac{1}{\lambda}$, P), then a random variable X = $Z + \alpha(1 - e^{-\lambda x})^{\frac{1}{\beta}}(1 - Pe^{-\lambda x})^{\frac{-1}{\beta}}$ will follow PEG(α, β, λ, P) distribution.
The proof is similar to that of Theorem 1.

4. Distributional Characteristics

The quantile based measures of the distributional characteristics of location, dispersion, skewness and kurtosis are popular in statistical analysis. These measures are also useful for estimating parameters of the model by matching population characteristics with corresponding sample characteristics.
Median (M) of the PEG model is,

$$\begin{aligned} M &= Q(0.5) \\ &= \alpha(0.5)^{\frac{1}{\beta}} \end{aligned} \tag{16}$$

Interquartile range (IQR) of the PEG model is,

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= \alpha\left((0.75)^{\frac{1}{\beta}} - (0.25)^{\frac{1}{\beta}}\right) + \frac{1}{\lambda}\log\left(\frac{4-3P}{1.33-0.33P}\right) \end{aligned} \tag{17}$$

Galton's coefficient of skewness (S) of the PEG model is,

$$\begin{aligned} S &= \frac{Q_1 + Q_3 - 2M}{Q_3 - Q_1} \\ &= \frac{\alpha\left((0.25)^{\frac{1}{\beta}} + (0.75)^{\frac{1}{\beta}} - 2(0.5)^{\frac{1}{\beta}}\right) + \left(\log(1.33-0.33P) + \log\left(\frac{4-3P}{2-P}\right)\right)}{\alpha\left((0.75)^{\frac{1}{\beta}} - (0.25)^{\frac{1}{\beta}}\right) + \frac{1}{\lambda}\log\left(\frac{4-3P}{1.33-0.33P}\right)} \end{aligned} \tag{18}$$

Moor's coefficient of kurtosis (T) of the PEG model is,

$$\begin{aligned} T &= \frac{Q(0.875) - Q(0.625) + Q(0.375) - Q(0.125)}{IQR} \\ &= \frac{\alpha\left(0.875^{\frac{1}{\beta}} - 0.625^{\frac{1}{\beta}} + 0.375^{\frac{1}{\beta}} - 0.125^{\frac{1}{\beta}}\right) + \frac{1}{\lambda}\log\left(\frac{0.107P^2-0.41P+0.328}{0.006P^2-0.039P+0.078}\right)}{\alpha\left(0.75^{\frac{1}{\beta}} - 0.25^{\frac{1}{\beta}}\right) + \frac{1}{\lambda}\log\left(\frac{4-3P}{1.33-0.33P}\right)} \end{aligned} \tag{19}$$

5. L - Moments

L- moments are the expected values of linear function of order statistics. The L- moments are often found to be more desirable than the conventional moments in describing the characteristics of the distributions as well as for inference. L-moments can be used as summary measures (statistics) of probability distributions (samples) to identify distributions and to fit models to data. A unified theory and a systematic study on L - moments have been presented by Hosking (1990).

The r^{th} L-moment is given by,

$$L_r = \int_0^1 \sum_{k=0}^{r-1} (-1)^{r-1-k} \binom{r-1}{k} \binom{r-1+k}{k} u^k Q(u) du \tag{20}$$

The first L moment is the mean of the distribution. For the PEG model, L_1 is obtained as,

$$\begin{aligned} L_1 &= \int_0^1 Q(u) du \\ &= \frac{\alpha\beta}{1+\beta} + \frac{(P-1)\log(1-P)}{\lambda P} \end{aligned} \tag{21}$$

The second L moment is obtained as,

$$\begin{aligned} L_2 &= \int_0^1 (2u-1)Q(u)du \\ &= \frac{\alpha\beta}{1+2\beta^2+3\beta} + \frac{(P-1)\,(P+log(1-P))}{\lambda P^2} \end{aligned} \tag{22}$$

The third L moment is,

$$\begin{aligned} L_3 &= \int_0^1 (6u^2-6u+1)Q(u)du \\ &= \frac{\alpha\beta}{\beta+1} - \frac{6\alpha\beta^2}{1+5\beta+6\beta^2} + \frac{2(P-1)}{\lambda P^2} - \frac{(P-1)(P-2)log(1-P)}{\lambda P^3} \end{aligned} \tag{23}$$

Fourth L moment is obtained as,

$$\begin{aligned} L_4 &= \int_0^1 (20u^3-30u^2+12u-1)Q(u)du \\ &= \frac{20\alpha\beta}{1+4\beta} - \frac{30\alpha\beta}{1+3\beta} + \frac{11\alpha\beta+10\alpha\beta^2}{1+3\beta+2\beta^2} + \frac{\alpha(P-1)(P^3-15P^2+30P+6\,((P-5)P+5)\,log(1-P))}{6P^4} \end{aligned} \tag{24}$$

The L - coefficient of variation, analogous to the coefficient of variation based on ordinary moments for model (8) is given by,

$$\tau_2 = \frac{L_2}{L_1} = \frac{\frac{\alpha\beta}{1+\beta} + \frac{(P-1)(P+log(1-P))}{\lambda P^2}}{\frac{\alpha\beta}{1+\beta} + \frac{(P-1)\log(1-P)}{\lambda P}} \tag{25}$$

L - coefficient of skewness is obtained as,

$$\tau_3 = \frac{L_3}{L_2} = \frac{\frac{\alpha\beta}{\beta+1} - \frac{6\alpha\beta^2}{1+5\beta+6\beta^2} + \frac{2(P-1)}{\lambda P^2} - \frac{(P-1)(P-2)log(1-P)}{\lambda P^3}}{\frac{\alpha\beta}{1+2\beta^2+3\beta} + \frac{(P-1)(P+log(1-P))}{\lambda P^2}} \tag{26}$$

L - coefficient of kurtosis of PEG function is,

$$\tau_4 = \frac{L_4}{L_2} = \frac{\frac{20\alpha\beta}{1+4\beta} - \frac{30\alpha\beta}{1+3\beta} + \frac{11\alpha\beta+10\alpha\beta^2}{1+3\beta+2\beta^2} + \frac{\alpha(P-1)(P^3-15P^2+30P+6((P-5)P+5)log(1-P))}{6P^4}}{\frac{\alpha\beta}{1+2\beta^2+3\beta} + \frac{(P-1)(P+log(1-P))}{\lambda P^2}} \tag{27}$$

6. Order Statistics

There are several topics in reliability of analysis in which order statistics appear quite naturally. If $X_{r:n}$ is the r^{th} order statistic in a random sample of size n, then the density function of $X_{r:n}$ can be written as,

$$f_r(x) = \frac{1}{B(r,n-r+1)} f(x)(F(x))^{r-1}(1-F(x))^{n-r} \tag{28}$$

From (10) the equation will be,

$$f_r(x) = \frac{1}{B(r,n-r+1)} \frac{\beta\lambda(1-F(x)-PF(x)+PF(x)^2)(F(x))^{r-1}(1-F(x))^{n-r}}{\beta-\beta P+\lambda\alpha F(x)^{\frac{1}{\beta}-1}(1-F(x)-PF(x)+PF(x)^2)}$$

Hence,

$$E(X_{r:n}) = \frac{1}{B(r,n-r+1)} \int_0^\infty x \frac{\beta\lambda(1-F(x)-PF(x)+PF(x)^2)(F(x))^{r-1}(1-F(x))^{n-r}}{\beta-\beta P+\lambda\alpha F(x)^{\frac{1}{\beta}-1}(1-F(x)-PF(x)+PF(x)^2)} dx$$

In quantile terms, it can be written as,

$$E(X_{r:n}) = \frac{1}{B(r, n-r+1)} \int_0^1 Q(u) \frac{\beta\lambda(1-u-Pu+Pu^2)u^{r-1}(1-u)^{n-r}}{\beta - \beta P + \lambda\alpha u^{\frac{1}{\beta}-1}(1-u-Pu+Pu^2)} dx$$

For the class of distributions (8), the first - order statistic $X_{1:n}$ has the quantile function

$$\begin{aligned} Q_{(1)}(u) &= Q(1-(1-u)^{\frac{1}{n}}) \\ &= \alpha\left(1-(1-u)^{\frac{1}{n}}\right)^{\frac{1}{\beta}} + \frac{1}{\lambda}\log\left(P+(1-P)(1-u)^{\frac{-1}{n}}\right) \end{aligned} \tag{29}$$

and the n^{th} order statistic $X_{n:n}$ has the quantile function

$$\begin{aligned} Q_{(n)}(u) &= Q(u^{\frac{1}{n}}) \\ &= \alpha u^{\frac{1}{\beta n}} + \frac{1}{\lambda}\log\left(\frac{1-Pu^{\frac{1}{n}}}{1-u^{\frac{1}{n}}}\right) \end{aligned} \tag{30}$$

Order statistics have more applications in quantile based reliability analysis as compared to distribution function based reliability analysis.

7. Reliability Properties

Reliabilty properties have an important role in real life situations. Some most relevent quantile based functions used in reliabilty analysis are hazard quantile function, mean residual quantile function, etc.

7.1. Hazard Quantile Function

One of the basic concepts employed for modeling and analysis of lifetime data is the hazard rate. In a quantile setup, Nair and Sankaran (2009) defined the hazard quantile function, which is equivalent to the hazard rate. The hazard quantile function H(u) is defined as

$$H(u) = h(Q(u)) = [(1-u)q(u)]^{-1} \tag{31}$$

Thus, H(u) can be interpreted as the conditional probability of failure of a unit in the next small interval of time given the survival of the unit until 100(1-α)% point of the distribution. Note that H(u) uniquely determines the distribution using the identity,

$$Q(u) = \int_0^u \frac{dp}{(1-p)H(p)} \tag{32}$$

Since the PEG model is the sum of quantile functions of power and exponential geometric quantile functions, (4) and (32) give

$$\frac{1}{H(u)} = \frac{1}{H_1(u)} + \frac{1}{H_2(u)} \tag{33}$$

where H(u) is the hazard quantile function of the PEG model, $H_1(u)$ is the hazard quantile function of Power distribution and $H_2(u)$ is the hazard quantile function of exponential geometric distribution. From (33), the PEG model has hazard quantile function proportional to the harmonic average of the hazard quantile functions of Power and exponential geometric quantile functions. For the class of distributions (8), we have

$$H(u) = \frac{\beta\lambda(1-Pu)}{\alpha\lambda(1-u)(1-Pu)u^{\frac{1}{\beta}-1} + \beta(1-P)} \tag{34}$$

7.1.1 Behavior of Hazard Quantile Function

The shape of the hazard quantile function can explain the behavior of hazard quantile function. It express increasing hazard rate (IHR), decreasing hazard rate (DHR), bathtub shape (BT), upside down bathtub shape (UBT) and constant rate at different values of parameters.

The different shapes of hazard quantile function for various values of parameters are summarized in Table 1 and plots given in Figure (4).

Table 1: *Behavior of hazard quantile function for different regions of parameters.*

No.	Parameter region	Shape of hazard quantile function
1	$\alpha > 1, \beta > 1, P = 0, \lambda > 0$	IHR
2	$\alpha = 1, 0 < \beta < 1, 0 < P < 1, \lambda > 0$	DHR
3	$\alpha = 1, \beta = 1, 0 < P < 1, \lambda > 0$	UBT
4	$\alpha > 1, 0 < \beta < 1, P = 1, \lambda > 0$	BT
5	$\alpha = 0, \beta > 1, P = 0, \lambda > 0$	Constant
6	$\alpha > 1, \beta > 1, 0 < P < 1, \lambda > 0$	UBT
7	$\alpha = 1, \beta > 1, P = 0, \lambda > 0$	IHR
8	$0 < \alpha < 1, 0 < \beta < 1, 0 < P < 1, \lambda > 0$	DHR
9	$\alpha > 1, \beta > 1, P = 1, \lambda > 0$	IHR

7.2. Mean Residual Quantile function

Mean residual function is a well - known measure that has been widely used for modeling lifetime data in reliability and survival analysis. For a nonnegative random variable X, the mean residual life function is defined as,

$$m(x) = \frac{1}{1 - F(x)} \int_x^\infty (1 - F(t))dt \tag{35}$$

In quantile based reliability analysis, the mean residual life function is known as mean residual quantile function, which is the quantile version of the mean residual function (35), defined by Nair and Sankaran (2009), has the expression,

$$M(u) = \frac{1}{1-u} \int_u^1 (Q(p) - Q(u))dp \tag{36}$$

For the PEG model, M(u) has the form,

$$\begin{aligned} M(u) = \frac{1}{1-u} \frac{\alpha\beta(1 - u^{\frac{1}{\beta}+1})}{1+\beta} + \frac{1-P}{P(1-u)}(1 - log(1-P)) - \frac{1-Pu}{P(1-u)}(1 - log(1-Pu)) \\ - (1 - log(1-u)) - \left(\alpha u^{\frac{1}{\beta}} + \frac{1}{\lambda} log(\frac{1-Pu}{1-u})\right) \end{aligned} \tag{37}$$

It is well known that increasing (decreasing) failure rate implies decreasing (increasing) mean residual life (see Lai and Xie 2006). The aging behavior of PEG model based on mean residual quantile function can be defined from Table (2)

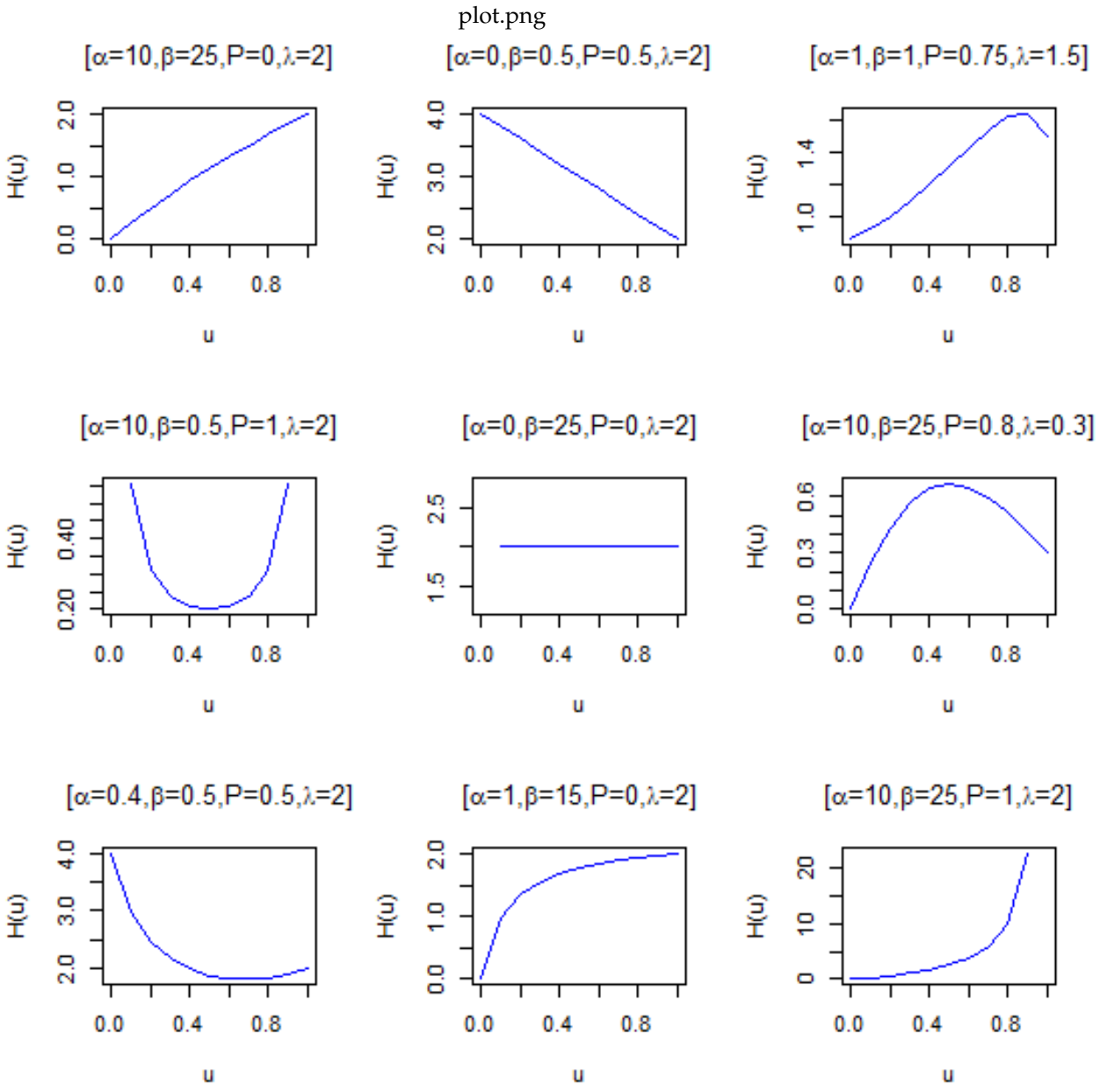


Figure 4: *Behavior of Hazard quantile function.*

7.3. Percentile Residual Quantile Function

Lillo (2005) found that Q(u) can be uniquely determined from the knowledge of $P_k(u)$, where $P_k(u)$ is the kth percentile residual quantile function. We have,

$$\begin{aligned} P_k(u) &= Q(1-(1-k)(1-u)) - Q(u) \\ &= \alpha(1-(1-k)(1-u))^{\frac{1}{\beta}} + \frac{1}{\lambda} log\left(\frac{1-P(1-(1-k)(1-u))}{1-(1-(1-k)(1-u))}\right) - \alpha u^{\frac{1}{\beta}} + \frac{1}{\lambda} log\left(\frac{1-Pu}{1-u}\right) \end{aligned} \tag{38}$$

7.4. Reversed Hazard Quantile Function

Reversed hazard quantile function is,

$$\begin{aligned}\Lambda &= (uq(u))^{-1} \\ &= \frac{\beta\lambda(1-u-Pu-Pu^2)}{u\beta(1-P)+\lambda\alpha u^{\frac{1}{\beta}}(1-u-Pu-Pu^2)}\end{aligned} \tag{39}$$

7.5. Reversed Mean Residual Quantile Function

Reversed mean residual quantile function is represented as R(u). k^{th} reversed mean residual quantile function can be obtained by using the given formula.

$$\begin{aligned}R(u) &= \frac{1}{u}\int_0^u (Q(u)-Q(k))dk \\ &= \alpha u^{\frac{1}{\beta}} + \frac{1}{\lambda}log\left(\frac{1-Pu}{1-u}\right) - \frac{\alpha\beta u^{\frac{1}{\beta}}}{1+\beta} - \frac{1}{\lambda u}\left(\frac{1-Pu}{P}(1-log(1-Pu)) - (1-u)(1-log(1-u))\right)\end{aligned} \tag{40}$$

7.6. Total Time on Test Transform (TTT)

Total time on test transform is represented as T(u), and it is calculated by using

$$T(u) = \int_0^u (1-k)q(k)dk$$

Also there is a relationship between total time on test transform and reversed mean residual quantile function (sankaran (2009)) and it is given by

$$T(u) = Q(u) - uR(u)$$

Using this relation, we can obtain the TTT of PEG model(8) and is given by

$$\begin{aligned}T(u) = (1-u)\left(\alpha u^{\frac{1}{\beta}} + \frac{1}{\lambda}log\left(\frac{1-Pu}{1-u}\right)\right) + \frac{\alpha u^{\frac{1}{\beta}+1}}{1+\beta} \\ + \frac{1}{\lambda}\left((1-Pu)(1-log(1-Pu)) - (1-u)(1-log(1-u))\right)\end{aligned} \tag{41}$$

8. Simulation Study

A simulation study is conducted to examine the performance of PEG quantile function for different sizes n=25, 50, 100, 500 using R package. Here simulate 1000 samples for the parameter values α=5, β=2, λ=3, P=0.87 and for α=0.8, β=0.05, λ=2, P=0.91, the maximum likelihood estimates for α, β, λ and P were determined for each sample, allowing the calculus of mean estimates.
We also evaluate the absolute bias and mean square erro (MSE) defined by,

$$AbsoluteBias = \frac{1}{N}|\sum_{i=1}^{N}(\hat{\epsilon}-\epsilon)|$$

and

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{\epsilon}-\epsilon)^2$$

From the table. we can see that, as the sample size increases, the mean square error decreases for all selected parameter values. Also the bias caused by the estimates decreases as the sample size increases. Thus the estimates tends to the true parameter values with increasing sample size.

Table 2: *Simulation study*

Sample Size	Parameter	Mean	Bias	MSE	Mean	Bias	MES
25	α	0.68112	4.31887	22.05722	16.77764	15.97764	6904.116
	β	0.00028	1.99971	3.99884	0.00779	0.04220	0.00278
	λ	0.24744	2.75255	8.23978	-0.97483	2.97483	40.40615
	P	0.11449	0.75550	0.67065	-2.04215	2.95215	61.96811
50	α	0.85121	4.14879	21.05278	2.03723	1.23723	192.11053
	β	0.20056	1.79943	3.59774	-1.01769	0.06769	0.02431
	λ	0.46151	2.53848	7.57733	-0.29682	2.29682	14.10438
	P	0.13915	0.73084	0.63639	-0.01065	0.9206	1.45400
100	α	1.05596	3.94404	20.04939	0.01600	0.78399	0.62799
	β	0.24022	1.75977	3.51908	0.00054	0.04972	0.00241
	λ	0.57022	2.42977	7.29000	0.03955	1.96003	3.92267
	P	0.18156	0.68874	0.60625	0.01768	0.89245	0.81158
500	α	5.00104	0.00104	0.00055	0.28842	0.51200	0.40922
	β	1.99600	0.00399	0.00797	0.01826	0.03211	0.00165
	λ	3.00001	$3.77e^{-6}$	$7.12e^{-9}$	0.72364	1.28413	2.56012
	P	0.87020	0.00020	$2.16e^{-5}$	0.32690	0.58245	0.52998

9. Data Analysis

This section explains the application of newly proposed quantile function in a real data set. There are many methods to estimate the unknown parameters of the quantile function. Method of maximum likelihood, method of L moments, method of minimum absolute deviation are some of the main mathods to estimate the parameters. In this work, method of maximum likelihood estimation procedure is used to estimate the parameters. Here the PEG function is applied to real data set reported in Zimmer et al. (1998). The data set consist of first failure times of small 20 electric carts used for internal transportation and delivery in a manufacturing company.The estimates so obtained are given by,

$$\hat{\alpha} = 11.2417, \hat{\beta} = 0.0056, \hat{\lambda} = 2.5998, \hat{P} = 0.7101 \tag{42}$$

The model adequacy is checked by using chi-squared goodness of fit test. The test result gives the p-value 0.2358. The significance level is 0.05, hence the test result indicates the adequacy of the PEG model to the data.

10. Summary and Conclusion

In this project work, a new quantile function is introduced, that is the sum of quantile functions of the power and exponential geometric (PEG) quantile functions. And also discovered that some well-known distributions are members of the PEG quantile function. Then plot different shapes of the density model for different values of parameters, also find the distributional characteristics of the model such as median, inter quartile range, skewness and kurtosis. Then derived first four L-moments and order statistics of the PEG quantile function. Important reliability properties are studied such as hazard quantile function, mean residual quantile function, etc. The different shapes of hazard quantile function for different parameter regions are plotted. Simulation study is conducted and it results that the bias and MSE are decreases as sample size increases. Finally the PEG model is applied to a real data set, and parameters are estimated by using maximum likelihood estimation procedure and model adequacy is checked by chi-square goodness of fit test. All these are done by using R software.

Several properties and extensions are possible in this PEG quantile function they are not

considerd in this work, such as stochastic ordering, parameter estimation by L- moment, etc. since the parameters estimated by L moment method will be more efficient.

References

[1] Adamidis, K. and S. Loukas. 1998. A lifetime distribution with decreasing failure rate. Statistics and Probability Letters 39 (1):35–42. doi:10.1016/S0167-7152(98)00012-1.

[2] Freimer, M.G., Kollia, G. S. Mudholkar, and C. T. Lin. 1988. A study of the generalized Tukey lambda family. Communications in Statistics—Theory and Methods 17 (10):3547–67. doi:10.1080/ 03610928808829820.

[3] Govindarajulu, Z. 1977. A class of distributions useful in life testing and reliability. IEEE Transactions on Reliability 26 (1):67–69. doi:10.1109/TR.1977.5215079.

[4] Hankin, R. K. and A. Lee. 2006. A new family of non-negative distributions. Australian and New Zealand Journal of Statistics 48 (1):67–78. doi:10.1111/j.1467-842X.2006.00426.x.

[5] Hastings, C., F. Mosteller, J. W. Tukey, and C. P. Winsor. 1947. Low moments for small samples: A comparative study of order statistics. Annals of Mathematical Statistics 18:413–26. doi:10.1214/ aoms/1177730388.

[6] Hosking, J. R. M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society Series B 52 105–24.

[7] Lai, C. D. and Xie, M. 2006. Stochastic ageing and dependence for reliability. Springer Science and Business Media.

[8] Lillo, R.E.: On the median residual lifetime and its aging properties: A charectarization theorem and application. Nav. Res. Logist. 52, 370-380(2005)

[9] Midhu, N. N., P. G. Sankaran, and N. U. Nair. 2014. A class of distributions with linear hazard quantile function. Communications in Statistics—Theory and Methods 43 (17):3674–89. doi:10.1080/03610926.2012.705211.

[10] Nair, N. U. and P. G. Sankaran. 2009. Quantile-based reliability analysis. Communications in Statistics—Theory and Methods 38 (2):222–32. doi:10.1080/03610920802187430.

[11] Nair, N. U., P. G. Sankaran, and N. Balakrishnan. 2013. Quantile-based reliability analysis. New York, NY: Springer, Birkhauser. doi:10.1007/978-0-8176-8361-0.

[12] Nair, N. Unnikrishnan, Sankaran, P.G., Midhu, N (2016) A New Quantile Function with Applications to Reliability analysis, Communications in statistics - Simulation and Computation, 45:2,566-582

[13] Nair, N. Unnikrishnan, Vineeshkumar, B (2012) Reliability Concepts in Quantile Based Analysis of Life Time Data.

[14] Ramberg, J. S. (1975). A probability distribution with applications to Monte Carlo simulation studies. In Statistical distributions in scientific work: Model building and model selection, eds. G.P. Patil, S. Kotz, J. K. Ord, vol. 2. Dordrecht: D. Reidel.

[15] Ramberg, J. S., E. J. Dudewicz, P. R. Tadikamalla, and E. F. Mykytka. 1979. A probability distribution and its uses in fitting data. Technometrics 21 (2):201–14. doi:10.1080/ 00401706.1979.10489750.

[16] Ramberg, J. S. and B. W. Schmeiser. (1972). An approximate method for generating symmetric random variables. Communications of the ACM 15 (11):987–90. doi:10.1145/355606.361888.

[17] Sankaran P. G and Dileep Kumar M (2018). Journal of Applied Probability and Statistics, vol. 13, N0. 1.pp. 81-95, ISOSS Publications.

[18] Sankaran P. G and Dileep Kumar M (2016) A New Class of Quantile Functions Usefull in Reliability Analysis, Journal of statistical Theory and Practice, 12:3, 615-634

[19] Tukey, J. W. 1962. The future of data analysis. Annals of Mathematical Statistics 33 (1):1–67. doi:10.1214/aoms/1177704711.

[20] Van Staden, P. J. and M. T. Loots. 2009. Method of L-moment estimation for the generalized lambda distribution. Proceedings of the Third Annual ASEARC Conference, New Castle, Australia, December 7–8.

[21] Zimmer, W. J., J. B. Keats, and F. K. Wang. 1998. The Burr XII distribution in reliability analysis. Journal of Quality Technology 30 (4):386–94. doi:10.1080/00224065.1998.11979874.

RELIABILITY SINGLE SAMPLING PLANS UNDER THE ASSUMPTION OF BURR TYPE XII DISTRIBUTION

Vijayaraghavan R

•

Professor & Head Department of Statistics
Bharathiar University Coimbatore 641 046, INDIA
rvijayrn@yahoo.com

Saranya C R*

•

Lecturer Department of Statistics KSMDB College
Sasthamcotta 690 521, Kerala, INDIA
saranyasreekumar17@gmail.com

Sathya Narayana Sharma K

•

Assistant Professor Department of Mathematics
School of Advanced Sciences Vellore Institute of Technology
Vellore 632 014, Tamil Nadu, INDIA
sharma14081992@gmail.com

***Corresponding Author**

Abstract

Acceptance sampling or sampling inspection is an essential quality control technique which describes the rules and procedures for making decisions about the acceptance or rejection of a batch of commodities by the inspection of one or more samples. When quality of an item is evaluated based on the life time of the item, which can be adequately described by a continuous-type probability distribution, the plan is known as life test sampling plan. The application of Burr (XII) distribution in reliability sampling plans is considered in this article. A procedure for selection of the plan parameters to protect the both producer as well as the consumer indexed by the acceptable mean life and operating ratio is evolved. Application of proposed plan is discussed with the help of numerical illustrations. Evaluation of such plans is explained utilising a set of simulated observations from Burr (XII) distribution.

Keywords*:* Acceptable Mean Life, Operating Ratio, Burr Distribution, Reliability Sampling, Type I Censoring.

I. Introduction

Sampling inspection plans are used to determine the acceptability of a lot consisting of the finished products based on the inspection of sampled items. Lifetime of the items which are put under test is considered as an important characteristic in reliability sampling plans. While making the decision on the disposition of the lot based on life testing, the length or duration of the total time spent on the inspection of items would be a major constraint and hence, it would be desirable if a life test is

terminated by specifying a time and observing the number of failures that occur before the pre-assigned time.

The utilization of several continuous probability distributions in the researches pertained to the construction and evaluation of life test sampling plans has been significantly outlined in the literature of product control. While important contributions have been made during the past five decades in the evolution of life test sampling plans employing exponential, Weibull, lognormal, gamma and other lifetime distributions, the literature also provides application of varied distributions for modelling lifetime data. Epstein [1, 2], Handbook H-108 [3], and Goode and Kao [4-6] proposed the construction of life test sampling plans using exponential and Weibull distributions.

The latest advancements in life tests sampling plans includes the works of Gupta [7], Schilling and Neubauer [8], Balakrishnan *et al*., [9], Kalaiselvi and Vijayaraghavan [10], Kalaiselvi *et al*., [11], Loganathan *et al*., [12], Vijayaraghavan *et al*., [13], Vijayaraghavan and Uma [14, 15] and Vijayaraghavan *et al*., [16, 17].

Burr [18] introduced a system of twelve continuous distributions and one among them is termed as the Burr type XII distribution or simply Burr distribution. It is also considered as a generalized log-logistic distribution. Literature in reliability theory advocates the adoptability of the family of Burr-type distributions for modeling lifetime data and for modeling the concept with monotone and unimodal failure rates. The Burr type XII is often considered as a suitable model for failure data. Similar to the log-normal distribution, it has a non-monotone hazard function which can accommodate many shapes of hazard.

Zimmer and Burr [19] considered a wide range of values for the degrees of skewness and kurtosis using a class of Burr distributions and developed the method of deriving variables sampling plans for non-normal populations based on the measures of skewness and kurtosis. Rodriguez [20, 21] has used measures of skewness and kurtosis of Burr distribution to derive the area in the plane based on the Burr type II distribution. Tadikamalla [22] has summarized the relationship between the Burr type II distribution and other distributions such as Lomax, compound Weibull, Weibull-Exponential, logistic, log logistic, Weibull and Kappa family of distributions. Zimmer *et al*., [23] discussed the statistical and probabilistic properties of the Burr type XII distribution and its relationship to other distributions used in reliability analyses.

Lio *et al*., [24] developed single sampling plans based on the percentiles of the Burr type XII distribution percentiles when the life test is truncated at a pre-specified time. Following this, Aslam *et al*., [25] discussed a two-stage group sampling procedure for the Burr type XII distribution percentiles to save sample resource with a truncated censoring scheme. Rao *et al*., [26] attempted to estimate multi-component stress-strength reliability assuming the Burr type XII distribution. Application of Burr distribution in reliability sampling is now considered with a particular reference to single sampling plans.

II. Burr Distribution

Let T be the lifetime of the component, which is considered as a random variable. Assume that T follows the Burr distribution. The probability density function and cumulative distribution function of T are, respectively, given by

$$f(t;\theta,\eta,\delta)=\frac{\eta\delta}{\theta}\left(\frac{t}{\theta}\right)^{\eta-1}\left[1+\left(\frac{t}{\theta}\right)^{\eta}\right]^{-\delta-1},\ t>0,\theta>0,\eta>0,\delta>0 \tag{1}$$

and

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

where η and δ are the shape parameters, and θ is the scale parameter.
The mean life time of Burr distribution is given by

$$\mu = E(t) = \theta \frac{\Gamma(1+1/\eta)\Gamma(\delta - 1/\eta)}{\Gamma(\delta)}, \tag{3}$$

The failure proportion, p, of product before time *t*, is expressed by

$$p = P(T \le t) = F(t;\theta,\eta,\delta) \tag{4}$$

III. Procedure to Determine the Operating Characteristics

The acceptance probabilities of lot of items under a single sampling plan is explained as a function of the failure probability *p* and is expressed by

$$P_a(p) = \sum_{x:0}^{c} p(x). \tag{5}$$

The probability of acceptance under the specified conditions of binomial or Poisson distributions can be obtained utilising the corresponding expressions in (5).

It can be noted that the failure probability, *p*, is a function of *t*, δ and θ, as expressed in (4). Corresponding to a specific value of *p*, a unique value of t/θ would exist and can be derived as a function of *p*, η and δ using (2) and (4) as

$$\frac{t}{\theta} = \left((1-p)^{-1/\delta} - 1\right)^{1/\eta}. \tag{6}$$

Using (3) and (6), the expression for t/μ is obtained as

$$\frac{t}{\mu} = \left[(1-p)^{-1/\delta} - 1)\right]^{1/\eta} \frac{\Gamma(\delta)}{\Gamma\left(1+\frac{1}{\eta}\right)\Gamma\left(\delta - \frac{1}{\eta}\right)}. \tag{7}$$

Every single value of *p* is connected with distinct value t/μ, thus the OC function of RSSP is regarded as a function of t/μ. Plot the acceptance probabilities against the values of t/μ. The resulting figure would be the required OC curve.

IV. Empirical Analysis of Operating Characteristic Curves

It can be noted that the RSSP based on Burr (XII) distribution is specified by the parameters n, c, θ ,η and δ. As the failure probability *p* is associated with the distribution function, which is a function of t/θ, the acceptance probabilities in turn can be computed for given sets of values of *n, c,* η and δ. The acceptance probabilities of the submitted lot under the RSSP are computed against the ratio $E(t)/\mu_0 = \mu/\mu_0$ for different combinations of parameters n, c and δ . Here, μ and μ_0 represent the expected mean life and assumed mean life, respectively. It is to be noted that changes in the values of these parameters will influence the shape of the OC function. In order to explore the impact of the parameters an empirical analysis of the OC curves drawn for various sets of parameters is carried out.

Figure 1 displays the curves for varying values of η, and fixed values of *n, c* and δ, Figure 2 exhibit the curves for different values of δ, and the fixed values of n, c and η. Similarly, Figures 3 and 4 display the sets of curves for varying values of *n* and *c*, respectively, fixing the values of shape

parameters. The curves exhibit the probabilities of acceptance of the lot against the values of μ/μ_0 . From these figures, the following properties are observed: For any specified value of μ/μ_0, $P_a(p)$ increases as η increases; $P_a(p)$ increases as δ increases, $P_a(p)$ decreases as n increases; and $P_a(p)$ increases as c increases.

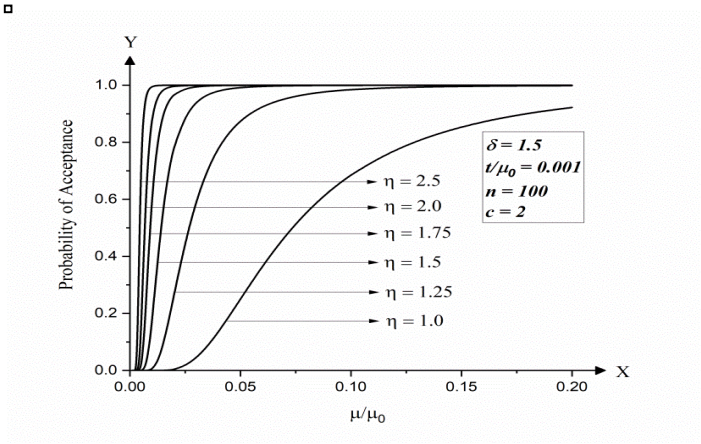


Figure 1: OC Curves of Single Sampling Plans for Life Tests Based on the Burr Distribution for Varying η with Fixed $n = 100, c = 2$ and $\delta = 1.5$

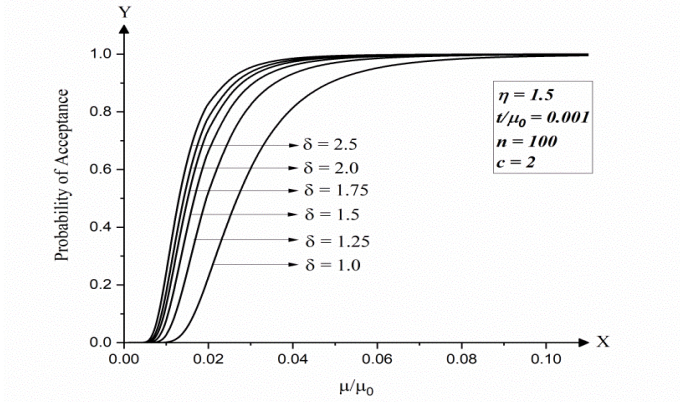


Figure 2: OC Curves of Single Sampling Plans for Life Tests Based on the Burr Distribution for Varying δ with Fixed $n = 100, c = 2$ and $\eta = 1.5$

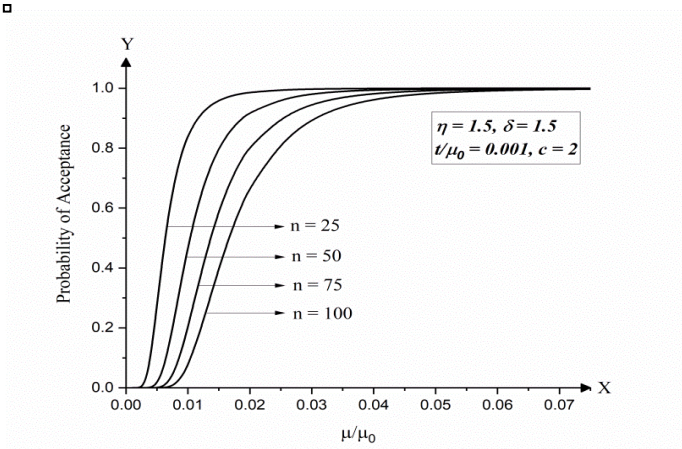


Figure 3: OC Curves of Single Sampling Plans for Life Tests Based on the Burr Distribution for Varying n with Fixed $c = 2, \eta = 1.5$ and $\delta = 1.5$

Hence, for any given value of μ/μ_0, for smaller values of η, the acceptance probabilities are lesser; for smaller values of δ, the acceptance probabilities are lesser; protection to the producer is greater with larger acceptance probabilities for smaller sample sizes as the expected mean life moves towards the assumed mean life; the consumer gets more protection with smaller acceptance probabilities for larger sample sizes when the expected mean life is much smaller than the assumed mean life; smaller the acceptance number, greater is the protection to the consumer; and larger the acceptance number, greater is the protection to the producer.

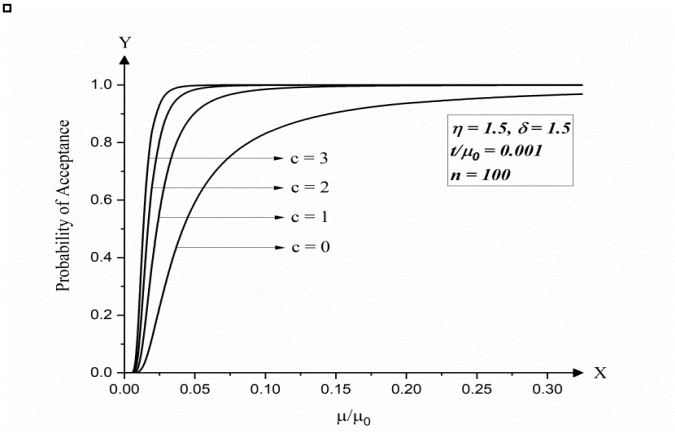


Figure 4: OC Curves of Single Sampling Plans for Life Tests Based on the Burr Distribution for Varying *c* with Fixed $n = 100, \eta = 1.5$ and $\delta = 1.5$

V. Procedure for the Construction of Reliability Single Sampling Plan

Vijayaraghavan and Uma (2016), discussed the procedures for obtaining the values of p_0 and p_1 in association with t/μ_0 and t/μ_1, respectively. In reliability sampling, a specific sampling plan for life tests can be obtained so that the OC curve must pass through two locations, namely, (μ_0, α) and (μ_1, β), which are associated with the risks α and β.

The two conditions specified below must be satisfied, for obtaining the optimum plan parameters with fixed value of α and β, respectively:

$$P_a(\mu_0) \geq 1-\alpha \tag{8}$$

and

$$P_a(\mu_1) \leq \beta. \tag{9}$$

Based on the search procedure, the optimum single sampling plans under Burr (XII) distribution for a range of values of μ_0/μ_1, t/μ_0, η and δ are determined and tabulated in Table 1 associated $\alpha = 0.05$ and $\beta = 0.10$.

VI. Numerical Illustrations

I. Illustration 1

A life test sampling plan is to be instituted under the condition that the life time follows the Burr distribution when the acceptable mean life and unacceptable mean life are prescribed as 6000 *hours* and 3000 *hours*, respectively, with the producer's and consumer's risks fixed as $\alpha = 0.05$ and $\beta = 0.10$. The past history from an industrial process yields the estimates of the shape parameters as $\eta = 1.5$ and $\delta = 1.5$. The experimenter wishes to terminate the life test at *t* = 240 *hours*. For the

given requirements, the values of t/μ_0 and t/μ_1 are obtained as 0.04 and 0.08, respectively and the operating ratio is $R = 2.0$. From Table 1, the optimum single sampling plan is determined as $n = 319$ and $c = 8$.

II. Illustration 2

An industrial practitioner is interested to find out a single sampling plan for its implementation to make a decision about the disposition of a submitted lot of manufactured products whose lifetime follows the Burr distribution. The test termination time for the items to be inspected has been fixed as $t = 325$ *hours*. In order to obtain the required sampling plan, experimental results are observed to estimate the shape parameters. The estimates of δ and η from the experimental results are obtained as 2.0 and 1.5, respectively.

With these values, the acceptable and unacceptable proportions of the lot failing before time, t, are determined as $p_0 = 0.0574$ and $p_1 = 0.18$, respectively, which correspond to the producer's and consumer's risks fixed at the levels $\alpha = 0.05$ and $\beta = 0.10$. Associated with $p_0 = 0.0574$ and $p_1 = 0.18$ are the values of t/μ_0 and t/μ_1, which are obtained as 0.03 and 0.105, respectively. Thus, the value of the operating ratio is obtained as $R = 3.5$.

When entered Table 1 with these values, the parameters of the optimum single sampling plan are obtained as $n = 139$ and $c = 3$. For the specified requirements under the optimum plan, the acceptable mean life and unacceptable mean life are, respectively, obtained as $\mu_0 = t/0.03 = 10833$ *hours* and $\mu_1 = t/0.105 = 3095$ *hours*.

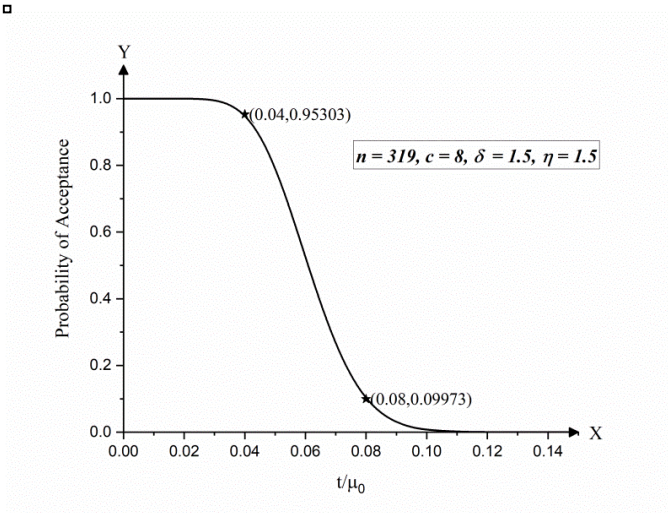


Figure 5: OC Curve of an Optimum Single Sampling Plan for Life Tests Based on the Burr Distribution Having Parameters $n = 319, c = 8, \eta = 1.5$ and $\delta = 1.5$

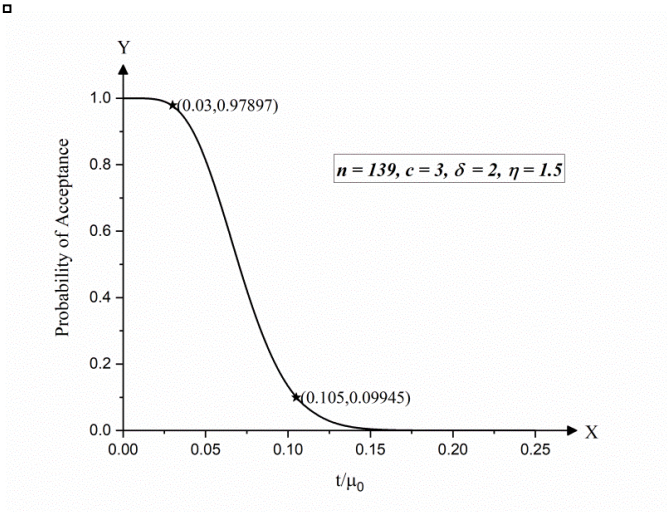


Figure 6: OC Curve of an Optimum Single Sampling Plan for Life Tests Based on the Burr Distribution Having Parameters $n = 139, c = 3, \eta = 1.5$ and $\delta = 2.0$

In order to exhibit the practical performance of the optimum sampling plane determined in the above two illustrations, their operating characteristic curves are drawn, which are displayed in Figures 5 and 6. It can be observed from these figures that the operating characteristic curves pass through the desired points $(p_0, 1-\alpha)$ and (p_1, β).

III. Illustration Based on Simulated Data

A life test sampling plan is to be instituted under the condition that the life time follows the Burr distribution when the acceptable mean life and unacceptable mean life are prescribed as 6000 *hours* and 3000 *hours*, respectively, with the producer's and consumer's risks fixed as $\alpha = 0.05$ and $\beta = 0.10.$

A set of 100 observations is simulated from Burr distribution with shape parameters $\eta = 1.5$ and $\delta = 1.5.$ The life test is decided to terminate at *t* = 240 *hours*. For the given requirements, the values of t/μ_0 and t/μ_1 are obtained as 0.04 and 0.2, respectively and the operating ratio is *R* = 5.0. From Table 1, the optimum single sampling plan is determined as *n* = 35 and *c* = 2.

Simulated observations are: 2398, 1149, 621, 1979, 202, 4859, 133, 99, 655, 1386, 1963, 2132, 13078, 2311, 197, 1734, 13466, 3457, 1077, 4912, 145, 4719, 1833, 1858, 996, 1277, 2450, 18659, 595, 2821, 605, 389, 866, 1366, 1835, 5822, , 33200, 214, 1089, 875, 4660, 660, 466, 1511, 1655, 2126, 1475, 733, 3218, 3439, 1609, 4342, 542, 2709, 4924, 559, 2657, 1373, 2271, 4159, 4829, 636, 437, 668, 2472, 1218, 2278, 258, 2695, 2581, 5282, 2391, 1931, 2293, 1000, 1337, 371, 2201, 896, 115, 6033, 4690, 175, 2602, 2866, 719, 1214, 1629, 3202, 1617, 687, 289, 1357, 56, 4183, 962, 641, 630, 2326 and 3555.

Random sample of 35 observations from the simulated data were 99, 145, 289, 437, 559, 595, 668, 687, 719, 875, 996, 1000, 1157, 1218, 1337, 1511, 1609, 1655, 1734, 1835, 1979, 2132, 2201, 2278, 2311, 2391, 2472, 3202, 4183, 4342, 4719, 4829, 4912, 6033 and 33200

Since the random sample contains two failures before time t=240, the lot is considered as accepted.

VII. Conclusion

Reliability single sampling plans are proposed based on Burr (XII) distribution. The procedures for choosing single sampling plans are developed. Tables are presented for choosing parameters of reliability sampling plans indexed by acceptable mean life and operating ratio for the preassigned time *t* with few specified values of shape parameters.

Table 1: *Optimum Parameters of RSSP Based on Burr (XII) Distribution for Certain Sets of Shape Parameters (Key: n, c)*

$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.01$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	21, 2	34, 1	69, 1	155, 0	29, 2	44, 1	87, 1	188, 0
9.50	22, 2	36, 1	74, 1	171, 0	31, 2	47, 1	94, 1	208, 0
9.00	23, 2	38, 1	80, 1	191, 0	32, 2	50, 1	101, 1	231, 0
8.50	24, 2	41, 1	87, 1	214, 0	34, 2	54, 1	110, 1	259, 0
8.00	25, 2	44, 1	95, 1	241, 0	36, 2	58, 1	121, 1	293, 0
7.50	27, 2	47, 1	105, 1	274, 0	38, 2	62, 1	133, 1	333, 0
7.00	36, 3	51, 1	116, 1	315, 0	41, 2	68, 1	147, 1	382, 0
6.50	38, 3	77, 2	129, 1	616, 1	55, 3	101, 2	164, 1	748, 1
6.00	41, 3	84, 2	145, 1	723, 1	59, 3	112, 2	185, 1	878, 1
5.50	45, 3	94, 2	165, 1	860, 1	64, 3	124, 2	210, 1	1044, 1
5.00	58, 4	105, 2	190, 1	1040, 1	70, 3	140, 2	242, 1	1263, 1
4.75	61, 4	112, 2	281, 2	1152, 1	89, 4	149, 2	358, 2	1399, 1
4.50	64, 4	150, 3	305, 2	1283, 1	93, 4	200, 3	388, 2	1559, 1
4.25	68, 4	161, 3	332, 2	1438, 1	99, 4	214, 3	422, 2	1748, 1
4.00	83, 5	173, 3	363, 2	1623, 1	121, 5	231, 3	462, 2	1973, 1
3.75	89, 5	187, 3	400, 2	1846, 1	129, 5	250, 3	509, 2	2244, 1
3.50	107, 6	244, 4	443, 2	2119, 1	157, 6	326, 4	564, 2	2576, 1
3.25	129, 7	267, 4	621, 3	3363, 2	168, 6	357, 4	791, 3	4087, 2
3.00	153, 8	342, 5	699, 3	3946, 2	203, 7	457, 5	892, 3	4796, 2
2.75	182, 9	432, 6	953, 4	4695, 2	267, 9	578, 6	1215, 4	5707, 2
2.50	233, 11	543, 7	1275, 5	7130, 3	342, 11	727, 7	1626, 5	8668, 3
2.25	312, 14	747, 9	1694, 6	8802, 3	460, 14	1000, 9	2162, 6	10700, 3
2.00	450, 19	1081, 12	2493, 8	15464, 5	664, 19	1448, 12	3182, 8	18800, 5
1.90	536, 22	1304, 14	3191, 10	17134, 5	790, 22	1747, 14	4073, 10	20830, 5
1.80	651, 26	1556, 16	3728, 11	21678, 6	962, 26	2085, 16	4758, 11	26355, 6
1.70	825, 32	2012, 20	4924, 14	29985, 8	1186, 31	2696, 20	6286, 14	36455, 8
1.60	1090, 41	2623, 25	6632, 18	40132, 10	1576, 40	3516, 25	8466, 18	48791, 10
1.50	1487, 54	3715, 34	9319, 24	56185, 13	2200, 54	4981, 34	11897, 24	68308, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.02$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	12, 2	15, 1	26, 1	40, 0	16, 2	20, 1	32, 1	48, 0
9.50	13, 2	16, 1	28, 1	44, 0	17, 2	21, 1	34, 1	53, 0
9.00	13, 2	17, 1	30, 1	49, 0	18, 2	22, 1	37, 1	59, 0
8.50	14, 2	18, 1	32, 1	54, 0	18, 2	24, 1	40, 1	66, 0
8.00	14, 2	20, 1	35, 1	61, 0	19, 2	25, 1	44, 1	74, 0
7.50	19, 3	21, 1	38, 1	69, 0	21, 2	27, 1	48, 1	84, 0
7.00	20, 3	31, 2	42, 1	80, 0	28, 3	41, 2	53, 1	96, 0
6.50	21, 3	34, 2	47, 1	156, 1	29, 3	44, 2	59, 1	189, 1
6.00	23, 3	37, 2	53, 1	182, 1	32, 3	49, 2	67, 1	221, 1
5.50	24, 3	41, 2	60, 1	217, 1	34, 3	54, 2	76, 1	263, 1
5.00	32, 4	46, 2	94, 2	262, 1	44, 4	60, 2	87, 1	317, 1
4.75	33, 4	49, 2	102, 2	290, 1	47, 4	64, 2	128, 2	351, 1
4.50	35, 4	66, 3	110, 2	322, 1	49, 4	86, 3	139, 2	391, 1
4.25	42, 5	70, 3	119, 2	361, 1	60, 5	92, 3	151, 2	438, 1
4.00	45, 5	75, 3	130, 2	407, 1	63, 5	99, 3	165, 2	495, 1
3.75	54, 6	81, 3	143, 2	463, 1	67, 5	107, 3	182, 2	562, 1
3.50	57, 6	106, 4	199, 3	532, 1	81, 6	139, 4	253, 3	645, 1
3.25	68, 7	115, 4	222, 3	843, 2	98, 7	153, 4	282, 3	1024, 2
3.00	81, 8	147, 5	250, 3	989, 2	116, 8	195, 5	318, 3	1201, 2
2.75	96, 9	186, 6	340, 4	1176, 2	138, 9	247, 6	433, 4	1429, 2
2.50	122, 11	233, 7	455, 5	1786, 3	176, 11	310, 7	578, 5	2170, 3
2.25	163, 14	320, 9	604, 6	2204, 3	236, 14	425, 9	768, 6	2678, 3
2.00	245, 20	462, 12	887, 8	3871, 5	340, 19	615, 12	1130, 8	4704, 5
1.90	289, 23	557, 14	1135, 10	4288, 5	405, 22	742, 14	1446, 10	5212, 5

Table 1 (Continued)

$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.02$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
1.80	349, 27	698, 17	1325, 11	5425, 6	492, 26	885, 16	1689, 11	6593, 6
1.70	439, 33	857, 20	1750, 14	7503, 8	623, 32	1144, 20	2230, 14	9119, 8
1.60	576, 42	1155, 26	2356, 18	10041, 10	823, 41	1542, 26	3003, 18	12205, 10
1.50	794, 56	1622, 35	3310, 24	14056, 13	1122, 54	2111, 34	4219, 24	17086, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.03$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	9, 2	10, 1	15, 1	18, 0	12, 2	13, 1	18, 1	22, 0
9.50	9, 2	11, 1	16, 1	20, 0	12, 2	13, 1	20, 1	24, 0
9.00	10, 2	11, 1	17, 1	22, 0	13, 2	14, 1	21, 1	27, 0
8.50	10, 2	12, 1	19, 1	25, 0	13, 2	15, 1	23, 1	30, 0
8.00	13, 3	13, 1	20, 1	28, 0	14, 2	16, 1	25, 1	33, 0
7.50	14, 3	19, 2	22, 1	32, 0	19, 3	17, 1	27, 1	38, 0
7.00	15, 3	20, 2	24, 1	36, 0	20, 3	26, 2	30, 1	43, 0
6.50	16, 3	22, 2	27, 1	70, 1	21, 3	28, 2	33, 1	85, 1
6.00	17, 3	24, 2	30, 1	82, 1	22, 3	30, 2	37, 1	99, 1
5.50	21, 4	26, 2	34, 1	98, 1	24, 3	34, 2	42, 1	118, 1
5.00	23, 4	29, 2	53, 2	118, 1	31, 4	38, 2	66, 2	142, 1
4.75	24, 4	39, 3	57, 2	130, 1	33, 4	40, 2	71, 2	157, 1
4.50	29, 5	41, 3	61, 2	145, 1	34, 4	53, 3	77, 2	175, 1
4.25	30, 5	44, 3	66, 2	162, 1	42, 5	57, 3	84, 2	196, 1
4.00	32, 5	47, 3	72, 2	182, 1	44, 5	61, 3	91, 2	221, 1
3.75	38, 6	51, 3	80, 2	207, 1	53, 6	66, 3	100, 2	251, 1
3.50	40, 6	66, 4	110, 3	238, 1	56, 6	86, 4	139, 3	288, 1
3.25	48, 7	72, 4	123, 3	377, 2	67, 7	94, 4	155, 3	457, 2
3.00	57, 8	91, 5	138, 3	441, 2	80, 8	120, 5	175, 3	535, 2
2.75	73, 10	115, 6	188, 4	525, 2	95, 9	151, 6	238, 4	637, 2
2.50	92, 12	144, 7	250, 5	796, 3	121, 11	189, 7	317, 5	966, 3
2.25	120, 15	196, 9	332, 6	982, 3	162, 14	260, 9	421, 6	1192, 3
2.00	170, 20	283, 12	487, 8	1724, 5	232, 19	375, 12	619, 8	2094, 5
1.90	200, 23	341, 14	623, 10	1909, 5	276, 22	452, 14	791, 10	2319, 5
1.80	249, 28	427, 17	727, 11	2415, 6	335, 26	566, 17	924, 11	2934, 6
1.70	311, 34	524, 20	959, 14	3340, 8	424, 32	696, 20	1220, 14	4057, 8
1.60	406, 43	706, 26	1291, 18	4469, 10	560, 41	938, 26	1642, 18	5430, 10
1.50	555, 57	990, 35	1812, 24	6255, 13	776, 55	1283, 34	2306, 24	7600, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.04$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	8, 2	8, 1	11, 1	11, 0	10, 2	9, 1	13, 1	13, 0
9.50	8, 2	8, 1	11, 1	12, 0	10, 2	10, 1	13, 1	14, 0
9.00	8, 2	9, 1	12, 1	13, 0	10, 2	10, 1	14, 1	15, 0
8.50	8, 2	9, 1	13, 1	15, 0	11, 2	11, 1	16, 1	17, 0
8.00	11, 3	13, 2	14, 1	16, 0	11, 2	12, 1	17, 1	19, 0
7.50	12, 3	14, 2	15, 1	18, 0	15, 3	17, 2	18, 1	22, 0
7.00	12, 3	15, 2	16, 1	21, 0	16, 3	19, 2	20, 1	25, 0
6.50	13, 3	16, 2	18, 1	41, 1	17, 3	20, 2	22, 1	49, 1
6.00	13, 3	18, 2	20, 1	47, 1	18, 3	22, 2	25, 1	57, 1
5.50	17, 4	19, 2	23, 1	56, 1	23, 4	24, 2	28, 1	67, 1
5.00	18, 4	21, 2	35, 2	67, 1	25, 4	27, 2	44, 2	81, 1
4.75	19, 4	28, 3	38, 2	74, 1	26, 4	29, 2	47, 2	89, 1
4.50	23, 5	30, 3	41, 2	82, 1	27, 4	38, 3	51, 2	99, 1
4.25	24, 5	32, 3	44, 2	92, 1	33, 5	41, 3	55, 2	111, 1
4.00	25, 5	34, 3	48, 2	104, 1	34, 5	44, 3	60, 2	125, 1
3.75	30, 6	44, 4	53, 2	118, 1	41, 6	47, 3	66, 2	142, 1
3.50	36, 7	47, 4	73, 3	135, 1	44, 6	61, 4	92, 3	163, 1
3.25	38, 7	51, 4	81, 3	213, 2	52, 7	67, 4	102, 3	258, 2

Table 1 (Continued)

$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.04$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
3.00	49, 9	65, 5	91, 3	250, 2	62, 8	85, 5	115, 3	302, 2
2.75	57, 10	82, 6	124, 4	297, 2	73, 9	107, 6	156, 4	359, 2
2.50	72, 12	103, 7	165, 5	450, 3	93, 11	134, 7	208, 5	545, 3
2.25	94, 15	140, 9	218, 6	554, 3	124, 14	184, 9	276, 6	672, 3
2.00	132, 20	201, 12	319, 8	972, 5	186, 20	265, 12	404, 8	1180, 5
1.90	162, 24	242, 14	408, 10	1077, 5	220, 23	319, 14	517, 10	1307, 5
1.80	194, 28	303, 17	510, 12	1361, 6	266, 27	400, 17	604, 11	1653, 6
1.70	248, 35	387, 21	628, 14	1882, 8	334, 33	491, 20	797, 14	2286, 8
1.60	321, 44	500, 26	845, 18	2518, 10	438, 42	661, 26	1072, 18	3058, 10
1.50	436, 58	701, 35	1185, 24	3524, 13	603, 56	927, 35	1505, 24	4280, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.05$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	9, 3	6, 1	8, 1	7, 0	8, 2	8, 1	10, 1	9, 0
9.50	9, 3	7, 1	9, 1	8, 0	8, 2	8, 1	10, 1	9, 0
9.00	9, 3	7, 1	9, 1	9, 0	9, 2	8, 1	11, 1	10, 0
8.50	9, 3	7, 1	10, 1	10, 0	9, 2	9, 1	12, 1	11, 0
8.00	10, 3	11, 2	11, 1	11, 0	12, 3	9, 1	13, 1	13, 0
7.50	10, 3	11, 2	11, 1	12, 0	13, 3	14, 2	14, 1	14, 0
7.00	11, 3	12, 2	12, 1	23, 1	13, 3	15, 2	15, 1	16, 0
6.50	11, 3	13, 2	14, 1	27, 1	14, 3	16, 2	16, 1	32, 1
6.00	14, 4	14, 2	15, 1	31, 1	15, 3	17, 2	18, 1	37, 1
5.50	15, 4	15, 2	17, 1	37, 1	19, 4	19, 2	21, 1	44, 1
5.00	18, 5	21, 3	26, 2	44, 1	21, 4	21, 2	32, 2	52, 1
4.75	19, 5	22, 3	28, 2	48, 1	21, 4	22, 2	35, 2	58, 1
4.50	20, 5	24, 3	30, 2	54, 1	22, 4	30, 3	37, 2	64, 1
4.25	24, 6	25, 3	33, 2	60, 1	27, 5	32, 3	40, 2	72, 1
4.00	25, 6	27, 3	35, 2	67, 1	29, 5	34, 3	44, 2	81, 1
3.75	26, 6	34, 4	39, 2	76, 1	34, 6	44, 4	48, 2	92, 1
3.50	30, 7	37, 4	54, 3	87, 1	36, 6	47, 4	67, 3	105, 1
3.25	36, 8	40, 4	59, 3	138, 2	43, 7	52, 4	74, 3	166, 2
3.00	41, 9	51, 5	66, 3	161, 2	51, 8	66, 5	83, 3	195, 2
2.75	48, 10	64, 6	90, 4	191, 2	60, 9	83, 6	113, 4	231, 2
2.50	64, 13	80, 7	120, 5	289, 3	82, 12	103, 7	150, 5	350, 3
2.25	83, 16	108, 9	158, 6	356, 3	108, 15	141, 9	199, 6	432, 3
2.00	114, 21	155, 12	231, 8	624, 5	153, 20	203, 12	292, 8	757, 5
1.90	139, 25	198, 15	295, 10	691, 5	180, 23	245, 14	373, 10	838, 5
1.80	165, 29	233, 17	369, 12	874, 6	217, 27	306, 17	435, 11	1060, 6
1.70	210, 36	298, 21	453, 14	1208, 8	273, 33	391, 21	574, 14	1466, 8
1.60	270, 45	397, 27	609, 18	1615, 10	357, 42	506, 26	772, 18	1961, 10
1.50	365, 59	552, 36	854, 24	2260, 13	491, 56	709, 35	1083, 24	2744, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.06$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	8, 3	6, 1	7, 1	5, 0	7, 2	6, 1	8, 1	6, 0
9.50	8, 3	6, 1	7, 1	6, 0	8, 2	7, 1	8, 1	7, 0
9.00	8, 3	6, 1	8, 1	6, 0	8, 2	7, 1	9, 1	7, 0
8.50	9, 3	6, 1	8, 1	7, 0	8, 2	7, 1	9, 1	8, 0
8.00	9, 3	9, 2	9, 1	8, 0	8, 2	11, 2	10, 1	9, 0
7.50	9, 3	10, 2	9, 1	9, 0	11, 3	12, 2	11, 1	10, 0
7.00	12, 4	10, 2	10, 1	17, 1	12, 3	12, 2	12, 1	20, 1
6.50	12, 4	11, 2	11, 1	19, 1	12, 3	13, 2	13, 1	23, 1
6.00	13, 4	12, 2	12, 1	22, 1	13, 3	14, 2	14, 1	26, 1
5.50	13, 4	13, 2	13, 1	26, 1	17, 4	16, 2	16, 1	31, 1
5.00	16, 5	18, 3	21, 2	31, 1	18, 4	17, 2	25, 2	37, 1
4.75	17, 5	19, 3	22, 2	34, 1	22, 5	23, 3	27, 2	41, 1

Table 1 (Continued)

$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.06$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
4.50	18, 5	20, 3	24, 2	38, 1	23, 5	25, 3	29, 2	45, 1
4.25	18, 5	21, 3	26, 2	42, 1	24, 5	26, 3	31, 2	50, 1
4.00	22, 6	22, 3	28, 2	47, 1	25, 5	28, 3	34, 2	57, 1
3.75	25, 7	28, 4	30, 2	54, 1	30, 6	36, 4	37, 2	64, 1
3.50	27, 7	31, 4	42, 3	61, 1	31, 6	39, 4	52, 3	73, 1
3.25	31, 8	38, 5	46, 3	97, 2	37, 7	42, 4	57, 3	116, 2
3.00	36, 9	42, 5	52, 3	113, 2	44, 8	53, 5	64, 3	136, 2
2.75	45, 11	52, 6	70, 4	134, 2	56, 10	67, 6	87, 4	161, 2
2.50	56, 13	65, 7	92, 5	202, 3	70, 12	84, 7	116, 5	244, 3
2.25	72, 16	88, 9	122, 6	249, 3	92, 15	114, 9	153, 6	301, 3
2.00	103, 22	135, 13	178, 8	435, 5	130, 20	164, 12	224, 8	527, 5
1.90	120, 25	160, 15	227, 10	482, 5	153, 23	197, 14	286, 10	584, 5
1.80	142, 29	189, 17	283, 12	609, 6	191, 28	247, 17	357, 12	738, 6
1.70	185, 37	242, 21	348, 14	842, 8	238, 34	315, 21	439, 14	1020, 8
1.60	237, 46	322, 27	468, 18	1125, 10	303, 42	407, 26	591, 18	1364, 10
1.50	323, 61	446, 36	656, 24	1574, 13	424, 57	570, 35	829, 24	1909, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.07$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	7, 3	5, 1	6, 1	4, 0	7, 2	6, 1	7, 1	5, 0
9.50	8, 3	5, 1	6, 1	5, 0	7, 2	6, 1	7, 1	5, 0
9.00	8, 3	5, 1	6, 1	5, 0	7, 2	6, 1	7, 1	6, 0
8.50	8, 3	8, 2	7, 1	6, 0	7, 2	6, 1	8, 1	6, 0
8.00	8, 3	8, 2	7, 1	6, 0	10, 3	10, 2	8, 1	7, 0
7.50	8, 3	9, 2	8, 1	12, 1	10, 3	10, 2	9, 1	8, 0
7.00	11, 4	9, 2	8, 1	13, 1	11, 3	11, 2	10, 1	15, 1
6.50	11, 4	10, 2	9, 1	15, 1	11, 3	12, 2	11, 1	17, 1
6.00	12, 4	10, 2	10, 1	17, 1	14, 4	12, 2	12, 1	20, 1
5.50	12, 4	11, 2	11, 1	20, 1	15, 4	14, 2	13, 1	23, 1
5.00	15, 5	15, 3	17, 2	23, 1	16, 4	15, 2	21, 2	28, 1
4.75	15, 5	16, 3	18, 2	26, 1	19, 5	20, 3	22, 2	30, 1
4.50	18, 6	17, 3	20, 2	28, 1	20, 5	21, 3	24, 2	34, 1
4.25	19, 6	18, 3	21, 2	32, 1	21, 5	22, 3	25, 2	38, 1
4.00	22, 7	19, 3	23, 2	35, 1	25, 6	24, 3	28, 2	42, 1
3.75	23, 7	24, 4	25, 2	40, 1	26, 6	30, 4	30, 2	48, 1
3.50	27, 8	26, 4	34, 3	45, 1	31, 7	33, 4	42, 3	54, 1
3.25	28, 8	33, 5	38, 3	72, 2	33, 7	35, 4	46, 3	86, 2
3.00	32, 9	36, 5	50, 4	84, 2	43, 9	45, 5	52, 3	101, 2
2.75	40, 11	44, 6	56, 4	99, 2	50, 10	56, 6	70, 4	119, 2
2.50	53, 14	61, 8	75, 5	150, 3	62, 12	70, 7	93, 5	180, 3
2.25	67, 17	81, 10	98, 6	184, 3	81, 15	96, 9	123, 6	222, 3
2.00	91, 22	113, 13	156, 9	322, 5	114, 20	137, 12	179, 8	389, 5
1.90	110, 26	135, 15	182, 10	356, 5	134, 23	165, 14	229, 10	430, 5
1.80	134, 31	167, 18	227, 12	449, 6	167, 28	206, 17	286, 12	544, 6
1.70	164, 37	211, 22	279, 14	621, 8	208, 34	263, 21	351, 14	752, 8
1.60	213, 47	270, 27	375, 18	829, 10	271, 43	351, 27	472, 18	1005, 10
1.50	289, 62	383, 37	525, 24	1160, 13	376, 58	475, 35	662, 24	1406, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.08$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	7, 3	6, 2	5, 1	4, 0	6, 2	5, 1	6, 1	4, 0
9.50	7, 3	7, 2	5, 1	4, 0	6, 2	5, 1	6, 1	4, 0
9.00	7, 3	7, 2	6, 1	4, 0	8, 3	8, 2	6, 1	5, 0
8.50	7, 3	7, 2	6, 1	5, 0	9, 3	8, 2	7, 1	5, 0

Table 1 (Continued)

$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.08$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
8.00	9, 4	7, 2	6, 1	5, 0	9, 3	9, 2	7, 1	6, 0
7.50	10, 4	8, 2	7, 1	5, 0	9, 3	9, 2	8, 1	6, 0
7.00	10, 4	8, 2	7, 1	11, 1	10, 3	10, 2	8, 1	12, 1
6.50	10, 4	9, 2	8, 1	12, 1	10, 3	10, 2	9, 1	14, 1
6.00	13, 5	9, 2	9, 1	14, 1	13, 4	11, 2	10, 1	16, 1
5.50	13, 5	10, 2	9, 1	16, 1	14, 4	12, 2	11, 1	18, 1
5.00	14, 5	14, 3	15, 2	18, 1	17, 5	16, 3	17, 2	22, 1
4.75	14, 5	14, 3	16, 2	20, 1	18, 5	17, 3	19, 2	24, 1
4.50	17, 6	15, 3	17, 2	22, 1	18, 5	18, 3	20, 2	26, 1
4.25	17, 6	16, 3	18, 2	25, 1	19, 5	19, 3	21, 2	29, 1
4.00	20, 7	20, 4	19, 2	28, 1	23, 6	21, 3	23, 2	33, 1
3.75	21, 7	21, 4	26, 3	31, 1	24, 6	26, 4	25, 2	37, 1
3.50	24, 8	23, 4	29, 3	35, 1	28, 7	28, 4	35, 3	42, 1
3.25	28, 9	29, 5	31, 3	56, 2	33, 8	36, 5	38, 3	67, 2
3.00	32, 10	31, 5	42, 4	65, 2	38, 9	39, 5	43, 3	78, 2
2.75	39, 12	39, 6	47, 4	77, 2	45, 10	49, 6	58, 4	92, 2
2.50	48, 14	53, 8	62, 5	116, 3	56, 12	61, 7	77, 5	139, 3
2.25	61, 17	70, 10	82, 6	142, 3	77, 16	82, 9	101, 6	171, 3
2.00	86, 23	98, 13	130, 9	248, 5	106, 21	126, 13	148, 8	299, 5
1.90	99, 26	117, 15	151, 10	274, 5	125, 24	150, 15	189, 10	331, 5
1.80	124, 32	144, 18	188, 12	346, 6	149, 28	177, 17	236, 12	418, 6
1.70	151, 38	182, 22	231, 14	477, 8	191, 35	226, 21	290, 14	577, 8
1.60	196, 48	240, 28	310, 18	638, 10	247, 44	301, 27	389, 18	772, 10
1.50	268, 64	330, 37	449, 25	946, 14	340, 59	417, 36	545, 24	1079, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.09$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	7, 3	6, 2	5, 1	3, 0	7, 3	5, 1	5, 1	3, 0
9.50	7, 3	6, 2	5, 1	3, 0	8, 3	5, 1	5, 1	4, 0
9.00	7, 3	6, 2	5, 1	4, 0	8, 3	5, 1	6, 1	4, 0
8.50	7, 3	7, 2	5, 1	4, 0	8, 3	5, 1	6, 1	4, 0
8.00	7, 3	7, 2	6, 1	4, 0	8, 3	8, 2	6, 1	5, 0
7.50	7, 3	7, 2	6, 1	8, 1	9, 3	8, 2	7, 1	5, 0
7.00	9, 4	8, 2	6, 1	9, 1	9, 3	9, 2	7, 1	10, 1
6.50	10, 4	8, 2	7, 1	10, 1	9, 3	9, 2	8, 1	11, 1
6.00	10, 4	8, 2	10, 2	11, 1	12, 4	10, 2	9, 1	13, 1
5.50	12, 5	11, 3	12, 2	13, 1	13, 4	11, 2	13, 2	15, 1
5.00	13, 5	12, 3	13, 2	15, 1	13, 4	15, 3	15, 2	18, 1
4.75	13, 5	13, 3	14, 2	16, 1	16, 5	15, 3	16, 2	19, 1
4.50	16, 6	14, 3	14, 2	18, 1	17, 5	16, 3	17, 2	21, 1
4.25	16, 6	14, 3	15, 2	20, 1	20, 6	17, 3	18, 2	24, 1
4.00	19, 7	18, 4	17, 2	22, 1	21, 6	18, 3	20, 2	26, 1
3.75	22, 8	19, 4	23, 3	25, 1	22, 6	23, 4	22, 2	30, 1
3.50	23, 8	20, 4	25, 3	28, 1	26, 7	25, 4	30, 3	34, 1
3.25	26, 9	26, 5	27, 3	45, 2	30, 8	32, 5	33, 3	53, 2
3.00	30, 10	31, 6	36, 4	52, 2	35, 9	34, 5	44, 4	62, 2
2.75	37, 12	34, 6	40, 4	61, 2	41, 10	43, 6	49, 4	73, 2
2.50	44, 14	47, 8	53, 5	92, 3	54, 13	53, 7	65, 5	111, 3
2.25	59, 18	62, 10	78, 7	113, 3	66, 15	72, 9	86, 6	136, 3
2.00	82, 24	86, 13	110, 9	197, 5	97, 21	110, 13	125, 8	237, 5
1.90	94, 27	103, 15	128, 10	218, 5	113, 24	131, 15	160, 10	262, 5
1.80	114, 32	127, 18	160, 12	307, 7	140, 29	155, 17	199, 12	331, 6
1.70	142, 39	160, 22	208, 15	379, 8	173, 35	197, 21	245, 14	458, 8
1.60	186, 50	211, 28	275, 19	506, 10	223, 44	263, 27	329, 18	612, 10

Table 1 (Continued)

$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.09$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
1.50	248, 65	289, 37	381, 25	750, 14	308, 59	364, 36	460, 24	855, 13
$R = \mu_0 / \mu_1$	$t / \mu_0 = 0.10$							
	$\delta = 1.5$				$\delta = 2.0$			
	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$	$\eta = 1.0$	$\eta = 1.25$	$\eta = 1.5$	$\eta = 2.0$
10.00	6, 3	6, 2	4, 1	3, 0	7, 3	4, 1	5, 1	3, 0
9.50	8, 4	6, 2	4, 1	3, 0	7, 3	4, 1	5, 1	3, 0
9.00	8, 4	6, 2	5, 1	3, 0	7, 3	7, 2	5, 1	3, 0
8.50	8, 4	6, 2	5, 1	3, 0	8, 3	7, 2	5, 1	4, 0
8.00	8, 4	6, 2	5, 1	6, 1	8, 3	7, 2	6, 1	4, 0
7.50	9, 4	7, 2	5, 1	7, 1	8, 3	7, 2	6, 1	4, 0
7.00	9, 4	7, 2	6, 1	8, 1	10, 4	8, 2	7, 1	8, 1
6.50	9, 4	7, 2	6, 1	8, 1	11, 4	8, 2	7, 1	9, 1
6.00	11, 5	10, 3	9, 2	9, 1	11, 4	9, 2	8, 1	11, 1
5.50	12, 5	11, 3	10, 2	11, 1	12, 4	10, 2	12, 2	12, 1
5.00	12, 5	11, 3	11, 2	13, 1	15, 5	13, 3	13, 2	15, 1
4.75	14, 6	12, 3	12, 2	14, 1	15, 5	14, 3	14, 2	16, 1
4.50	15, 6	12, 3	13, 2	15, 1	16, 5	15, 3	15, 2	18, 1
4.25	15, 6	13, 3	14, 2	17, 1	16, 5	16, 3	16, 2	19, 1
4.00	18, 7	17, 4	15, 2	18, 1	19, 6	16, 3	17, 2	22, 1
3.75	18, 7	17, 4	20, 3	21, 1	20, 6	21, 4	19, 2	24, 1
3.50	21, 8	22, 5	22, 3	23, 1	24, 7	23, 4	26, 3	28, 1
3.25	24, 9	23, 5	24, 3	37, 2	28, 8	28, 5	29, 3	44, 2
3.00	30, 11	28, 6	31, 4	43, 2	32, 9	31, 5	38, 4	51, 2
2.75	34, 12	35, 7	35, 4	50, 2	41, 11	38, 6	43, 4	60, 2
2.50	41, 14	42, 8	46, 5	75, 3	50, 13	53, 8	57, 5	90, 3
2.25	55, 18	56, 10	60, 6	92, 3	64, 16	64, 9	74, 6	111, 3
2.00	76, 24	77, 13	96, 9	161, 5	89, 21	98, 13	118, 9	193, 5
1.90	87, 27	97, 16	111, 10	177, 5	108, 25	117, 15	138, 10	214, 5
1.80	108, 33	113, 18	138, 12	250, 7	128, 29	137, 17	172, 12	270, 6
1.70	134, 40	148, 23	179, 15	309, 8	163, 36	175, 21	211, 14	372, 8
1.60	172, 50	194, 29	238, 19	412, 10	209, 45	233, 27	283, 18	497, 10
1.50	236, 67	264, 38	328, 25	611, 14	291, 61	323, 36	396, 24	738, 14

The industrial practitioners can adopt this procedure to the life test and can develop the required plans for other choices of shape parameters. The application of proposed plan is discussed under two real life scenarios. Implementation of proposed plan is discussed with the help of numerical illustrations. Application of proposed plan is detailed with the help of simulated data from Burr distribution. The proposed plan is widely applicable in the manufacturing industries, testing of costly or destructive items, life testing for ball bearing, wind-speed data analysis, low-flow analysis, regional flood frequency, survival data, etc.

References

[1] Epstein, B. (1960b). Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential, Part II. *Technometrics*. **2**:167 - 183.

[2] Epstein, B., (1960a). Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential, Part I. *Technometrics.* **2**:83 - 101.

[3] Handbook H-108. (1960). *Sampling Procedures and Tables for Life and Reliability Testing.* Quality Control and Reliability, Office of the Assistant Secretary of Defense, US Department of Defense, Washington, D.C.

[4] Goode, H. P., and Kao, J. H. K. (1961). Sampling Plans Based on the Weibull Distribution. P*roceedings of the Seventh National Symposium on Reliability and Quality Control*, Philadelphia, PA. 24 - 40.

[5] Goode, H. P., and Kao, J. H. K. (1962). Sampling Procedures and Tables for Life and Reliability Testing Based on the Weibull Distribution (Hazard Rate Criterion). *Proceedings of the Eight National Symposium on Reliability and Quality Control*, Washington, DC. 37 - 58.

[6] Goode, H. P., and Kao, J. H. K., (1964). Hazard Rate Sampling Plans for the Weibull Distribution. *Industrial Quality Control.* 20:30-39.

[7] Gupta, S. S. (1962). Life Test Sampling Plans for Normal and Lognormal Distributions. *Technometrics*. 4:151-175.

[8] Schilling, E. G., and Neubauer, D. V., (2009). *Acceptance Sampling in Quality* Control. Chapman and Hall, New York, NY.

[9] Balakrishnan, N., Leiva, V., and L´opez, J. (2007). Acceptance Sampling Plans from Truncated Life-test Based on the Generalized Birnbaum - Saunders Distribution. *Communications in Statistics - Simulation and Computation*. **36**:643-656.

[10] Kalaiselvi, S., and Vijayaraghavan, R. (2010). Designing of Bayesian Single Sampling Plans for Weibull-Inverted Gamma Distribution. *Recent Trends in Statistical Research*, Publication Division, M. S. University, Tirunelveli. 123 - 132.

[11] Kalaiselvi, S., Loganathan, A., and Vijayaraghavan, R. (2011). Reliability Sampling Plans under the Conditions of Rayleigh – Maxwell Distribution – A Bayesian Approach. *Recent Advances in Statistics and Computer Applications*, Bharathiar University, Coimbatore, 2011. 280 - 283.

[12] Loganathan, A., Vijayaraghavan, R., and Kalaiselvi, S., (2012). Recent Developments in Designing Bayesian Reliability Sampling Plans – An Overview. *New Methodologies in Statistical Research,* Publication Division, M. S. University, Tirunelveli. 61 - 68.

[13] Vijayaraghavan, R., Chandrasekar, K., and Uma, S., (2012). Selection of sampling inspection plans for life test based on Weibull-Poisson mixed distribution. *Proceedings of the International Conference on Frontiers of Statistics and its Applications,* Coimbatore. 225-232.

[14] Vijayaraghavan, R., and Uma, S., (2012). Evaluation of sampling inspection plans for life test based on Exponential-Poisson mixed distribution. *Proceedings of the International Conference on Frontiers of Statistics and its Applications,* Coimbatore. 233-240.

[15] Vijayaraghavan, R., and Uma, S., (2016). Selection of Sampling Inspection Plans for Life Tests Based on Lognormal Distribution. *Journal of Testing and Evaluation*. 44:1960 - 1969.

[16] Vijayaraghavan, R., Saranya, C. R., and Sathya Narayana Sharma, K., (2020b). Reliability Sampling Plans Based on Exponential Distribution, *TEST Engineering and Management*, **83**, 28001-28005.

[17] Vijayaraghavan, R., Sathya Narayana Sharma, K., and Saranya, C. R., (2020a). Reliability Sampling Plans for Life Tests Based on Pareto Distribution, *TEST Engineering and Management*, **83**, 27991-28000.

[18] Burr, Irving W. (1942). Cumulative Frequency Functions. *Annals of Mathematical Statistics*, **13**, No. 2, pp. 215—232.

[19] Zimmer, W. J. and Burr I. W. (1963). Variable Sampling Plans Based on Non-Normal Populations, *Industrial Quality Control*, **3**, pp.18-26.

[20] Rodriguez, R. N. (1976). A Guide to the Burr Type XII Distributions, Institute of Statistics Mimeo Series No. 1064, North Carolina University, Chapel Hill.

[21] Rodriguez, R. N. (1977). A Guide to the Burr Type XII Distributions, *Biometrika*, **64**, No. 1, pp. 129-134.

[22] Tadikamalla, P.R. (1980). A Look at the Burr and Related Distributions, International Statistical Review / Revue Internationale de Statistique, **48**, No. 3, pp. 337-344.

[23] Zimmer, W. J., Wang, Y. and Pathak, P. K. (1998). Log-Odds Rate and Monotone Log-Odds Rate Distributions, *Journal of Quality Technology*, 30, No. 4, pp. 376-385.

[24] Lio, Y. L., Tsai T.-R. and Wu, S.-J. (2010). Acceptance sampling plans from truncated life tests based on the Burr type XII percentiles, *Journal of the Chinese Institute of Industrial Engineers*, **27**, No. 4, pp. 270-280.

[25] Aslam, M., Lio, Y., Azam, M., and Jun C-H. (2012). Two-Stage Improved Group Plans for Burr Type XII Distributions. *American Journal of Mathematics and Statistics*. **2**. pp. 33-39.

[26] Rao, G. S., Aslam, M. and Kundu, D. (2015). Burr-XII Distribution Parametric Estimation and Estimation of Reliability of Multicomponent Stress-Strength, *Communications in Statistics - Theory and Methods*, **44**, No. 23, pp. 4953-4961.

A NOVEL TRANSPORTATION APPROACH TO SOLVING TYPE - 2 TRIANGULAR INTUITIONISTIC FUZZY TRANSPORTATION PROBLEMS

Indira Singuluri
•
Vignan's Institute of Information Technology (A), Duvvada, Visakhapatnam.
indira.singuluri@gmail.com

N. Ravishankar
•
Gitam Deemed to be University, GIS, Visakhapatnam.
Drravi68@gmail.com

Abstract

In this article we propose a new transportation strategy to achieve an ideal answer for triangular intuitionistic fuzzy transportation problem of type – 2 i.e., limits and requests are considered as real numbers and the transportation cost from cause to objective is considered as triangular intuitionistic fuzzy numbers as product cost per unit. The proposed method is solving by using ranking function. The appropriate response system is delineated with a numerical model.

Keywords: IFN, TIFN, IF Optimum solution, TIFTP of type-2.

I. Introduction

In genuine world, there are general complex circumstances in each field, in which specialists and chiefs battle with uncertainty and hesitation. In useful circumstances, assortment of fresh information of different boundaries is troublesome because of absence of precise interchanges, mistake in information, market information and consumer loyalties. The data accessible is some of the time ambiguous and inadequate. The real-life problems, when defined by the decision maker with uncertainty leads to the notion of fuzzy sets. Due to imprecise information, the exact evaluation of participation values is not possible. Moreover, the evaluation of non-participation esteems is consistently impossible. This prompts an in deterministic climate where dithering endures. Managing estimated data while deciding, idea of fuzziness was presented by Bellman and Zadeh [6]. K. T, Atanassov [4] presented idea of Intuitionistic fuzzy set hypothesis, which is more able to manage such issues. B. Chetia and P. K. Das [1] demonstrated a few outcomes on intuitionistic fuzzy delicate network. Intuitionistic fuzzy sets [5], [7], [8] discovered to be exceptionally powerful in managing ambiguity, among a few higher request fuzzy sets. S.K. Singh, S.P. Yadav [9] proposed their strategies to address case 2 sort of intuitionistic fuzzy transportation problem (IFTP) for example IFTP of type-2. G. Gupta and A. Kumara [3] a capable technique was introduced in which limit and request factors are taken as TIFN's utilized in this article to tackle mathematical model. This paper proposes another transportation strategy for tackling TIFTP of type – 2 by applying ranking function found in [2].

The association of this article is as per the following: In Section 2, a review on essentials IFS and IFN's. Segment 3, presents the Ranking function and Comparison of TIFN's. Area 4, briefs the numerical detailing and proposed TP technique. Delineates the mathematical model in Section 5. At long last, Section 6 exposes the conclusion.

II. Preliminaries

In this part a couple of essential definitions and math tasks are examined.

Intuitionistic Fuzzy Set (IFS): An IFS $\tilde{A}^{IFS}$ in X an IFS is described as an object of following design

$$\tilde{A}^{IFS} = \{\langle x, \mu_{\tilde{A}^{IFS}}(x), \nu_{\tilde{A}^{IFS}}(x)\rangle : x \in X\}$$

where, functions $\mu_{\tilde{A}^{IFS}} : X \to [0, 1]$ and $\nu_{\tilde{A}^{IFS}}: X \to [0, 1]$ defines degree of Enrollment work and non-participation element $x \in X$, respectively and $0 \le \mu_{\tilde{A}^{IFS}}(x), \nu_{\tilde{A}^{IFS}}(x) \le 1$, for every $x \in X$.

Intuitionistic Fuzzy Numbers (IFN's): A subset of IFS, $\tilde{A}^{IFS} = \{\langle x, \mu_{\tilde{A}^{IFS}}(x), \nu_{\tilde{A}^{IFS}}(x)\rangle : x \in X\}$, of real line $\Re$ is called an IFN if the following holds:

(i) $\exists m \in \Re, \mu_{\tilde{A}^{IFS}}(m) = 1 \; and \; \nu_{\tilde{A}^{IFS}}(m) = 0$

(ii) $\mu_{\tilde{A}^{IFS}} : \Re \to [0, 1]$ is continuous and for every $x \in \Re, 0 \le \mu_{\tilde{A}^{IFS}}(x), \nu_{\tilde{A}^{IFS}}(x) \le 1$ holds.

Enrollment work and non-participation capacity of $\tilde{A}^{IFS}$ is as follows,

$$\mu_{\tilde{A}^{IFS}}(x) = \begin{cases} f_1(x), & x \in [m-\alpha_1, m) \\ 1, & x = m \\ h_1(x), & x \in (m, m+\beta_1] \\ 0, & otherwise \end{cases} \text{ and } \nu_{\tilde{A}^{IFS}}(x) = \begin{cases} 1, & x \in (-\infty, m-\alpha_2) \\ f_2(x), & x \in [m-\alpha_2, m) \\ 0, & x = m, x \in [m+\beta_2, \infty) \\ h_2(x), & x \in (m, m+\beta_2] \end{cases}$$

Where, $f_i(x)$ and $h_i(x)$; $i = 1,2$ are strictly increasing and decreasing functions in $[m-\alpha_i, m)$ and $(m, m-\beta_i]$ respectively. α_i and β_i are left and right spreads of $\mu_{\tilde{A}^{IFS}}(x)$ and $\nu_{\tilde{A}^{IFS}}(x)$ respectively.

Triangular Intuitionistic Fuzzy Number (TIFN): A TIFN $\tilde{A}^{IFN}$ is an IFS in $\Re$ with the following Enrollment function $\mu_{\tilde{A}^{IFN}}$ and non-participation capacity $\nu_{\tilde{A}^{IFN}}$ defined by

$$\mu_{\tilde{A}^{IFN}}(x) = \begin{cases} \frac{x-a_1}{a_2-a_1}, & a_1 \le x \le a_2 \\ \frac{a_1-x}{a_3-a_2}, & a_2 \le x \le a_3 \\ 0, & otherwise \end{cases} \text{ and } \nu_{\tilde{A}^{IFN}}(x) = \begin{cases} \frac{a'_1-x}{a_2-a'_1}, & a'_1 \le x \le a_2 \\ \frac{x-a_2}{a'_3-a_2}, & a_2 \le x \le a'_3 \\ 0, & otherwise \end{cases}$$

Where $a'_1 \le a_1 \le a_2 \le a_3 \le a'_3$. This TIFN is denoted by $\tilde{A}^{IFN} = (a_1, a_2, a_3; a'_1, a_2, a'_3)$ in Fig 1.

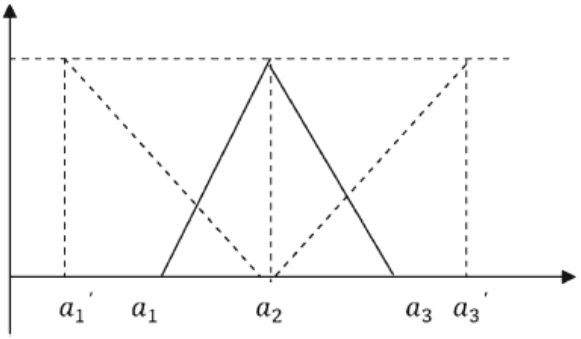


Figure 1: *Participation and non-enrollment elements of TIFN*

Arithmetic operations of TIFN:

For any two TIFN's $\tilde{A}^{IFN} = (a_1, a_2, a_3; a'_1, a_2, a'_3)$ and $\tilde{B}^{IFN} = (b_1, b_2, b_3; b'_1, b_2, b'_3)$, arithmetic operations are as follows,

(i) Addition:

$$\tilde{A}^{IFN} \oplus \tilde{B}^{IFN} = (a_1 + b_1, a_2 + b_{2,}, a_3 + b_3; a'_1 + b'_1, a_2 + b_2, a'_3 + b'_3)$$

(ii) Subtraction:

$$\tilde{A}^{IFN} - \tilde{B}^{IFN} = (a_1 - b_3, a_2 - b_{2,}, a_3 - b_1; a'_1 - b'_3, a_2 - b_2, a'_3 - b'_1)$$

(iii) Multiplication:

$$\tilde{A}^{IFN} \otimes \tilde{B}^{IFN} = (a_1 b_1, a_2 b_{2,}, a_3 b_3; a'_1 b'_1, a_2 b_2, a'_3 b'_3)$$

(iv) Scalar multiplication:

$$k \times \tilde{A}^{IFN} = \begin{cases} (ka_1, ka_2, ka_3; ka'_1, ka_2, ka'_3), k \geq 0 \\ (ka_3, ka_2, ka_1; ka'_3, ka_2, ka'_1), k < 0 \end{cases}$$

III. Ranking Function

Ranking function is taken from [2], i.e., the ranking function is defined for Trapezoidal and triangular Intuitionistic fuzzy number as

$$R\left(\tilde{A}^{IFN}\right) = \left(\frac{a_1 + b_1 + 2(a_2 + b_3) + 5(a_3 + b_2) + (a_4 + b_4)}{18}\right)\left(\frac{4w_1 + 5w_2}{18}\right)$$

$$R\left(\tilde{A}^{IFN}\right) = \left(\frac{(a_1 + b_1) + 14a_2 + (a_4 + b_4)}{18}\right)\left(\frac{4w_1 + 5w_2}{18}\right)$$

Consider $w_1 = w_2 = 1$, we get ranking function is

$$R\left(\tilde{A}^{IFN}\right) = \left(\frac{(a_1 + b_1) + 14a_2 + (a_4 + b_4)}{36}\right)$$

Comparison of TIFN's: To contrast TIFN's and one another, we need to rank them. A function such as R: $F(\Re) \to \Re$, which maps each TIFN's into real line, is called ranking function. Here, $F(\Re)$means the arrangement of all TIFN's.

By using ranking function "R", TIFN's can be compared.

Let $\tilde{A}^{IFN} = (a_1, a_2, a_3; a'_1, a_2, a'_3)$ and $\tilde{B}^{IFN} = (b_1, b_2, b_3; b'_1, b_2, b'_3)$are two TIFN's then

$\mathrm{R}\left(\tilde{A}^{IFN}\right) = \frac{a_1 + 14a_2 + a_3 + a'_1 + a'_3}{36}$ and $\mathrm{R}\left(\tilde{B}^{IFN}\right) = \frac{b_1 + 14b_2 + b_3 + b'_1 + b'_3}{36}$ then the orders are defined as follows

(i) $\tilde{A}^{IFN} > \tilde{B}^{IFN}$ if $\mathrm{R}\left(\tilde{A}^{IFN}\right) > R\left(\tilde{B}^{IFN}\right)$,

(ii) $\tilde{A}^{IFN} < \tilde{B}^{IFN}$ if $\mathrm{R}\left(\tilde{A}^{IFN}\right) < R\left(\tilde{B}^{IFN}\right)$, and

(iii) $\tilde{A}^{IFN} = \tilde{B}^{IFN}$ if $\mathrm{R}\left(\tilde{A}^{IFN}\right) = \mathrm{R}\left(\tilde{B}^{IFN}\right)$

Ranking function R also holds the following properties:

(i) $\mathrm{R}\left(\tilde{A}^{IFN}\right) + \mathrm{R}\left(\tilde{B}^{IFN}\right) = \mathrm{R}\left(\tilde{A}^{IFN} + \tilde{B}^{IFN}\right)$, (ii) $\mathrm{R}\left(\mathrm{k}\tilde{A}^{IFN}\right) = \mathrm{k}\,\mathrm{R}\left(\tilde{A}^{IFN}\right) \forall \mathrm{k} \in \boldsymbol{R}$

IV. Mathematical Formulation of Triangular Intuitionistic Fuzzy transportation problem (TIFTP) and proposed method

I. TIFTP of type - 2:

Consider a transportation with 'm' Intuitionistic Fuzzy (IF) origins and 'n' IF destination.

Let $C_{ij}(i = 1,2, \ldots, m; j = 1,2, \ldots, n)$ be the cost of transporting one unit of the product form i^{th}origin to j^{th}destination.

$\tilde{a}_i^{IFS} = (a_1^i, a_2^i, a_3^i; a_1^{i'}, a_2^i, a_3^{i'})$ be IF extent at i^{th} vendor.

$\tilde{b}_j^{IFS} = (b_1^i, b_2^i, b_3^i; b_1^{i'}, b_2^i, b_3^{i'})$ be IF abundant at j^{th} insistent.

$\tilde{x}_{ij}^{IFS} = (x_1^{ij}, x_2^{ij}, x_3^{ij}; x_1^{i'j'}, x_2^{ij}, x_3^{i'j'})$ be IF quantity transformed fromi^{th}vendor to j^{th} insistent

Then balanced triangular IFTP of type - 2 is given by

$$Min\tilde{Z}^{IFN} = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} \times x_{ij}^{IFN}$$

$$s.t. \sum_{j=1}^{n} \tilde{x}_{ij}^{IFN} = \tilde{a}_i^{IFN} \ , i = 1,2,\ldots,m$$

$$\sum_{i=1}^{m} \tilde{x}_{ij}^{IFN} = \tilde{b}_j^{IFN} \ , j = 1,2,\ldots,n$$

$$\tilde{x}_{ij}^{IFN} \geq \tilde{0}; \ i = 1,2,\ldots,m; j = 1,2,\ldots,n$$

II. Proposed Transportation strategy

Stage 1: Utilizing separation formula, considered in "Comparison of IFTN's" segment, adopt least and greatest IFN from each archive and segmentof intuitionistic fuzzy price matrix of TIFTP of type - 2 and deduct it from each IFN's of their relating line and segment.

Stage 2: Find sum of row difference and column difference and denote row sum by R and column sum by C. Identify Maximum sum of row and column. Select maximum difference in row and column.

Stage 3: Choose the cell having most minimal expense in row and column identified in stage 2.

Stage 4: Make a feasible assignment to the cell picked in stage 5. Delete fulfilled row/column.

Stage 5: Repeat the technique until all the designations has been made.

Stage 6: The Optimum solution and triangular intuitionistic optimum value is attained in step 5, is optimum solution $\{x_{ij}\}$ and triangular intuitionistic fuzzy optimum value is $\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} \otimes x_{ij}$.

V. Numerical Example

In this part, an existing mathematical model ([2]) is solved to illustrate the proposed transportation strategy.

Table 1: *TIFTP of type – 2*

	D_1	D_2	D_3	D_4	Supply (s_i)
S_1	(2,4,5; 1,4,6)	(2,5,7; 1,5,8)	(4,6,8; 3,6,9)	(4,7,8; 3,7,9)	11
S_2	(4,6,8; 3,6,9)	(3,7,12; 2,7,13)	(10,15,20; 8,15,22)	(11,12,13; 10,12,14)	11
S_3	(3,4,6; 1,4,8)	(8,10,13; 5,10,16)	(2,3,5; 1,3,6)	(6,10,14; 5,10,15)	11
S_4	(2,4,6; 1,4,7)	(3,9,10; 2,9,12)	(3,6,10; 2,6,12)	(3,4,5; 2,4,8)	12
Demand (d_j)	16	10	8	11	45

Example 1: An existing TIFTP of type - 2, with four suppliers i.e., S_1, S_2, S_3, S_4 and four destinations i.e.,D_1 , D_2, D_3, D_4, respectively by Table 1, is solved using the proposed method. This problem is solved in the following steps.

Select maximum and minimum TIFN in each row and column take the difference as given in table 2.

Table 2: *Row and Column Difference Table*

	D_1	D_2	D_3	D_4	**Supply** (s_i)	**Row diff**
S_1	(2,4,5; 1,4,6)	(2,5,7; 1,5,8)	(4,6,8; 3,6,9)	(4,7,8; 3,7,9)	11	1.4444
S_2	(4,6,8; 3,6,9)	(3,7,12; 2,7,13)	(10,15,20; 8,15,22)	(11,12,13; 10,12,14)	11	4.5
S_3	(3,4,6; 1,4,8)	(8,10,13; 5,10,16)	(2,3,5; 1,3,6)	(6,10,14; 5,10,15)	11	3.5
S_4	(2,4,6; 1,4,7)	(3,9,10; 2,9,12)	(3,6,10; 2,6,12)	(3,4,5; 2,4,8)	12	2.125
Demand (d_j)	16	10	8	11	45	R=11.56
Column diff	1.0555	2.6111	5.9444	3.9444	C=13.55	

The problem given in Table 2, transformed in Table 3 by using the Stage 2 and assign first allocation using stage 4 of proposed method.

Table 3: *First allocation Table*

	D_1	D_2	D_3	D_4	**Supply**	**Row difference**
S_1	(2,4,5; 1,4,6)	(2,5,7; 1,5,8)	(4,6,8; 3,6,9)	(4,7,8; 3,7,9)	11	1.4444
S_2	(4,6,8; 3,6,9)	(3,7,12; 2,7,13)	(10,15,20; 8,15,22)	(11,12,13; 10,12,14)	11	4.5
S_3	(3,4,6; 1,4,8)	(8,10,13; 5,10,16)	(2,3,5; 1,3,6) [**8**]	(6,10,14; 5,10,15)	11 **3**	3.5
S_4	(2,4,6; 1,4,7)	(3,9,10; 2,9,12)	(3,6,10; 2,6,12)	(3,4,5; 2,4,8)	12	2.125
Demand	16	10	8	11	45	R = 11.5416
Column difference	1.0555	2.6111	5.9444	3.9444	**C = 13.5554**	

Using Stage 4 of proposed method remove D_3 from Table 3. New reduced shown in Table 4 again apply the procedure.

Table 4: *New Reduced Table*

	D_1	D_2	D_4	Supply (s_i)	Row difference
S_1	(2,4,5; 1,4,6)	(2,5,7; 1,5,8)	(4,7,8; 3,7,9)	11	1.4444
S_2	(4,6,8; 3,6,9)	(3,7,12; 2,7,13)	(11,12,13; 10,12,14)	11	3
S_3	(3,4,6; 1,4,8)	(8,10,13; 5,10,16)	(6,10,14; 5,10,15)	3	3
S_4	(2,4,6; 1,4,7)	(3,9,10; 2,9,12)	(3,4,5; 2,4,8)	12	2.125
Demand	16	10	11	45	$R = 9.5694$
(d_j) Column differen	1.0555	2.6111	3.9444	$C = 7.6110$	

Table 5: *Second Allocation table*

	D_1	D_2	D_4	Supply (s_i)	Row diff
S_1	(2,4,5; 1,4,6)	(2,5,7; 1,5,8)	(4,7,8; 3,7,9)	11	1.4444
S_2	(4,6,8; 3,6,9)	(3,7,12; 2,7,13)	(11,12,13; 10,12,14)	11	3
S_3	(3,4,6; 1,4,8) [**3**]	(8,10,13; 5,10,16)	(6,10,14; 5,10,15)	3	3
S_4	(2,4,6; 1,4,7)	(3,9,10; 2,9,12)	(3,4,5; 2,4,8)	12	2.125
Demand	**16** **13**	10	11	45	**$R = 9.5694$**
(d_j) Column diff	1.0555	2.6111	3.9444	$C = 7.6110$	

Again, applying the Stage 5 of the proposed method, all the allocations are made as shown in Table 6.

Table 6: *Final allocation table*

	D_1	D_2	D_3	D_4
S_1	(2,4,5; 1,4,6) [**1**]	(2,5,7; 1,5,8) [**10**]	(4,6,8; 3,6,9)	(4,7,8; 3,7,9)
S_2	(4,6,8; 3,6,9) [**11**]	(3,7,12; 2,7,13)	(10,15,20; 8,15,22)	(11,12,13; 10,12,14)
S_3	(3,4,6; 1,4,8) [**3**]	(8,10,13; 5,10,16)	(2,3,5; 1,3,6) [**8**]	(6,10,14; 5,10,15)
S_4	(2,4,6; 1,4,7) [**1**]	(3,9,10; 2,9,12)	(3,6,10; 2,6,12)	(3,4,5; 2,4,8) [**11**]

Step 6: Optimum solution and IF optimum value

The optimum solution, obtained in Step 5, is $x_{11}=1, x_{12}=10, x_{21}=11, x_{31}=3, x_{33}=8, x_{41}=1$ and $x_{44}=11$. The IF optimum value of IFTP of type - 2, given in Table 1, is

$1\otimes(2,4,5;1,4,6)\oplus 10\otimes(2,5,7;1,5,8)\oplus 11\otimes(4,6,8;3,6,9)\oplus 3\otimes(3,4,6;1,4,8)\oplus 8\otimes(2,3,5;1,3,6)\oplus 1\otimes(2,4,6;1,4,7)\oplus 11\otimes(3,4,5;2,4,8)=(126,204,282;78,204,352).$

VI. Conclusion

Numerical Formulation for IFTP of type – 2 and system for acquiring an IF ideal arrangement is examined with relevant numerical example. The proposed transportation strategy is utilized to get the ideal arrangement of TIFTP of type – 2. The proposed transportation technique gives same outcome, as found by G. Gupta, A. Kumara [3], in single emphasis. Consequently, this might be favored over the current strategies.

References

[1] Chetia. B and Das. P. K., (2012). Some results of intuitionistic fuzzy soft matrix theory, Advances in Applied Science Research, 3, 412-413.

[2] Pardhasaradhi . B, Madhuri.M.V. and Ravi Shankar. N, (2017). Ordering of Intuitionistic fuzzy numbers using centroid of centroids of Intuitionistic fuzzy numbers", International Journal of Mathematics Trends and Technology, Vol. 52, No.5, 276-285.

[3] Gupta.G, Kumara. A (2017). An efficient method for solving intuitionistic fuzzy transportation problem of type-2, International Journal of Applied and Computational Mathematics, 3, 3795-9804.

[4] Atanassov. K. T., (1986). Intuitionstic fuzzy sets", Fuzzy Sets and Systems, 20, 87-96.

[5] Atanassov. K. T., (1989). More on intuitionstic fuzzy sets", Fuzzy Sets and Systems, 33, 37-46.

[6] Bellman.R.E. and Zadeh.L. A., (1970). Decision making in a fuzzy environment", Management Sciences, vol.17, 141-164.

[7] Jahir Hussain. R and Senthil Kumar. P., (2012). Algorithm approach for solving intuitionistic fuzzy transportation problem, Applied mathematical sciences, vol. 6(80), 3981-3989.

[8] SagayaRoseline and Henry Amirth raj, (2015). New Approaches to Find the Solution for the Intuitionistic Fuzzy Transportation Problem with Ranking of Intuitionistic Fuzzy Numbers, International Journal of Innovative Research in Science Engineering and Technology, vol. 4(10), 10222-10230.

[9] Singh.S.K., Yadav.S.P., (2016). A new approach for solving intuitionistic fuzzy transportation problem of type-2, Ann. Oper. Res. 243, 349–363.

THE NEW LENGTH BIASED QUASI LINDLEY DISTRIBUTION AND ITS APPLICATIONS

N. W. Andure (Yawale)[1] and R. B. Ade[2]

[1] Department of Statistics,Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India, neetayawale@gmail.com

[2]Department of Statistics,Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India, rajeshwarb.sc@gmail.com

Abstract

In this paper, Length biased Quasi Lindley (LBQL) distribution is proposed. The different properties of the proposed distribution are derived and discussed. The parameters of the proposed distribution are estimated by using method of maximum likelihood estimation and also the Fisher's Information matrix is obtained. The performance of the proposed distribution is studied using real-life data sets.

Keywords: Length Biased Distribution, Quasi Lindley Distribution, Reliability Analysis, Maximum Likelihood Estimation, Likelihood Ratio test.

I. Introduction

The Quasi Lindley (QL) distribution was introduced by Shanker and Mishra (2013). The QL distribution has two parameters α and θ. The Quasi-Lindley distribution reduces to one parameter Lindley distribution if α =θ. If α =0, it reduces to the gamma distribution with parameter (2, θ). The probability density function of QL distribution is a mixture of exponential (θ) and gamma (2, θ).
The Probability density function of Quasi Lindley distribution (QLD) with parameters α and θ is given by

$$f(x;\alpha,\theta)=\frac{\theta}{1+\alpha}(\alpha+\theta x)e^{-\theta x} \quad ;x>0,\theta>0,\alpha>-1 \qquad (1)$$

and the cumulative distribution function of the two parameter Quasi Lindley distribution is given by

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

II. Length Biased Quasi Lindley Distribution

Length biased distribution is a particular case of weighted distributions that were first introduced by Fisher (1934) to model the ascertainment bias. These weighted distributions were later developed by C R Rao (1965) in a unifying manner. Weighted distributions arise when the observations generated from a stochastic process are not given equal chances of being recorded and moderately, they are recorded in accordance to some weight function. When the weight function depends only on the length of units of interest, the resulting distribution is called as length biased. Length biased concept was firstly given by Cox (1969) and Zelen (1974). The study of weighted distributions helps us to deal with model description and data interpretation problems. In the study of distribution theory, weighted distributions are useful because it provides a new understanding of existing standard probability distributions and also it provides methods for extending existing standard probability distributions for modelling lifetime data due to introduction of additional parameter in the model which creates flexibility in their nature. Much work has been done to characterize the relations between original distributions and their length biased versions.

various researchers have reviewed and studied different weighted distribution and found its applications in different fields such as reliability, biomedicine, ecology, and branching processes [refer Lappi et al. (1987), Mir et al. (2013), Mudassir et al. (2015), Shenbagaraja et al. (2019)].

Definition: The non-negative random variable X is said to have weighted distribution, if the probability density function of weighted random variable X_w is given by

$$f_w(x) = \frac{w(x)f(x)}{E(w(x))} \quad , \quad x > 0$$

Where $w(x)$ be a non - negative weight function and

$$E\big(w(x)\big) = \int w(x)f(x)dx < \infty$$

For different weighted models, different choices of the weight function can be done. When $w(x) = x^c$,the resulting distribution is termed as weighted distribution. In this Paper, the Length biased version of Quasi Lindley distribution is studied, here choice of c = 1 is done as a weight, in order to get the Length biased Quasi Lindley distribution and the probability density function of Length biased Quasi Lindley distribution (LBQLD) is given by

$$f_L(x;\alpha,\theta) = \frac{xf(x,\alpha,\theta)}{E(x)} \tag{3}$$

Where $E(x) = \int_0^\infty x\, f(x;\alpha,\theta)dx$

$$E(x) = \frac{(\alpha+2)}{\theta(\alpha+1)} \tag{4}$$

After substituting from equation (1) and (4) in equation (3), the probability density function of Length biased Quasi Lindley distribution is obtained as

$$f_L(x;\alpha,\theta) = \frac{x\,\theta^2}{(\alpha+2)}(\alpha+\theta x)e^{-\theta x} \qquad ;x>0,\theta>0,\alpha>-2 \tag{5}$$

and the cumulative distribution function (cdf) of LBQL distribution is obtained as

$$F_L(x) = \int_0^x f_L(t;\alpha,\theta)dt$$

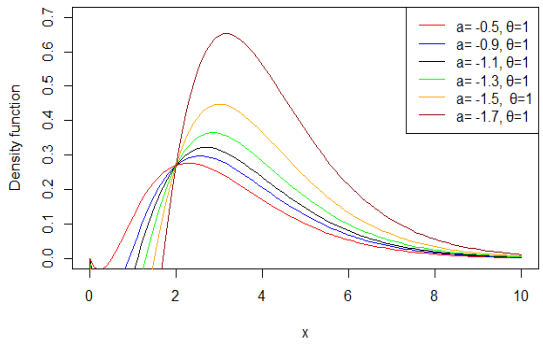
$$F_L(x) = \int_0^x \frac{t\,\theta^2}{(\alpha+2)}(\alpha+\theta t)e^{-\theta t}dt$$

$$F_L(x) = \frac{\theta^2}{(\alpha+2)}\int_0^x t(\alpha+\theta t)e^{-\theta t}dt$$

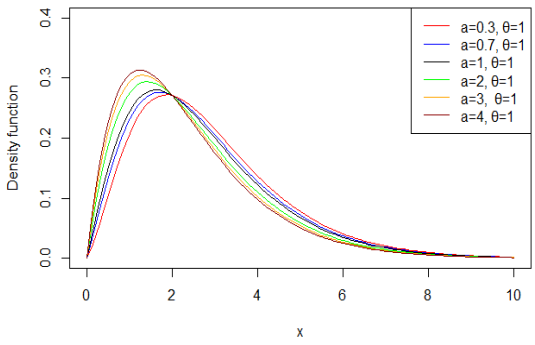
after simplification, the cumulative distribution function of Length biased Quasi Lindley distribution is

$$F_L(x) = \frac{\alpha\gamma(2,\theta x)+\gamma(3,\theta x)}{(\alpha+2)} \quad ; x>0, \theta>0, \alpha>-2 \qquad (6)$$

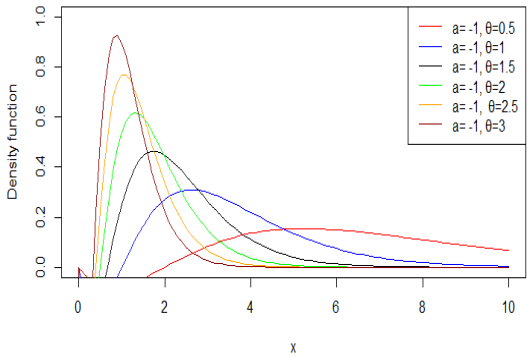
The graph of the probability density function and cumulative distribution function of length biased quasi-Lindley distribution (LBQLD) for different values of parameters, are shown in Figure 1 and 2.



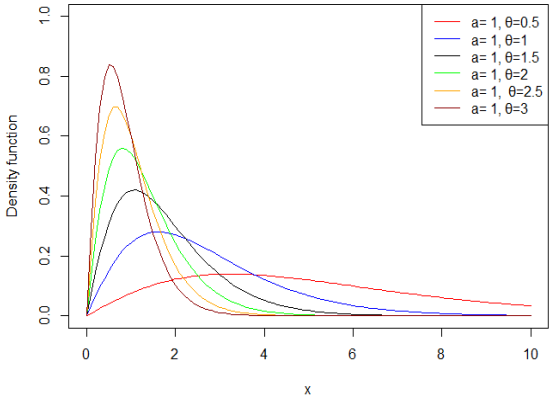
For $-2 < \alpha < 0$ and $\theta = 1$



For $\alpha > 0$ and $\theta = 1$



For $-2 < \alpha < 0$ and $\theta > 0$



For $\alpha > 0$ and $\theta > 0$

Figure 1: *Pdf plot of LBQLD for the different values of* α *and* θ.

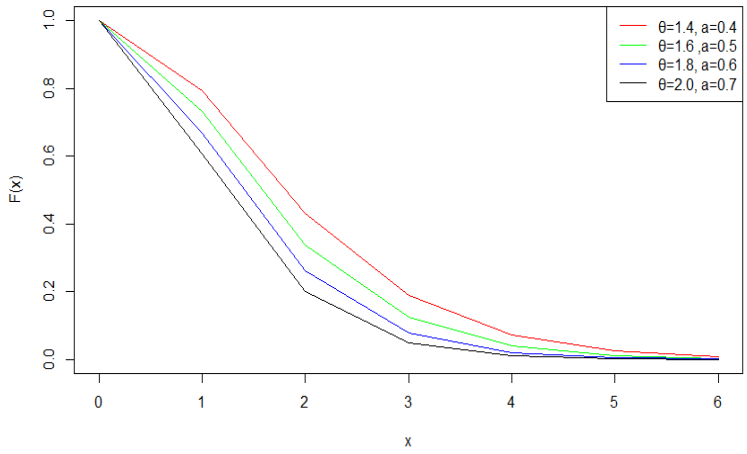


Figure 2: *cdf plot of LBQLD for the different values of α and θ.*

III. Reliability Analysis

In this section, the survival function, hazard rate, reverse hazard rate function and mill's ratio are discussed for the Length biased Quasi Lindley distribution.

The survival function is also known as reliability function and the Survival function of Length biased Quasi Lindley distribution is defined as

$$S(x) = 1 - F_L(x)$$

$$S(x) = 1 - \left(\frac{\alpha\gamma(2,\theta x) + \gamma(3,\theta x)}{(\alpha+2)}\right)$$

The hazard function is also known as hazard rate or instantaneous failure rate or force of mortality and the hazard function of Length biased Quasi Lindley distribution is given by

$$h(x) = \frac{f_L(x;\alpha,\theta)}{S(x)}$$

$$h(x) = \frac{x\theta^2(\alpha+\theta x)e^{-\theta x}}{(\alpha+2) - \left(\alpha\gamma(2,\theta x) + \gamma(3,\theta x)\right)}$$

where $(\alpha+2) - \left(\alpha\gamma(2,\theta x) + \gamma(3,\theta x)\right) > 0$

The reverse hazard function of Length biased Quasi Lindley distribution is given by

$$h_r(x) = \frac{f_L(x;\alpha,\theta)}{F_L(x)}$$

$$h_r(x) = \frac{x\theta^2(\alpha+\theta x)e^{-\theta x}}{\left(\alpha\gamma(2,\theta x) + \gamma(3,\theta x)\right)}$$

The Mills ratio of Length biased Quasi Lindley distribution is given by,

$$Mills\ ratio = \frac{1}{h_r(x)}$$

$$Mills\ ratio = \frac{\left(\alpha\gamma(2,\theta x) + \gamma(3,\theta x)\right)}{x\theta^2(\alpha+\theta x)e^{-\theta x}}$$

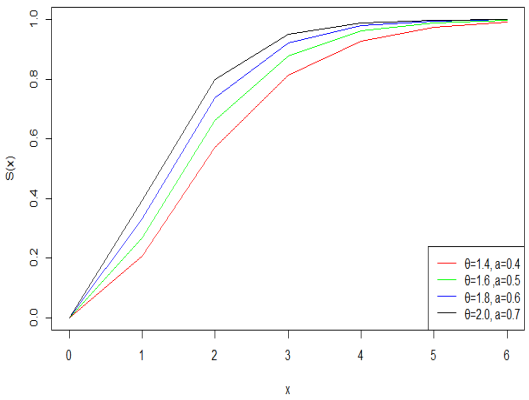


Fig.3: *Graph of survival function of LBQLD.*

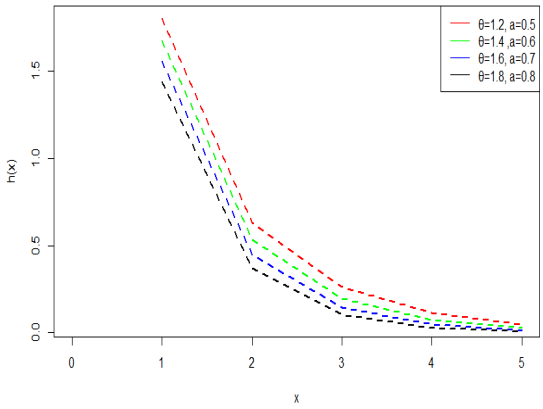


Fig.4: *Graph of Hazard function of LBQLD.*

Figure (4) shows the behavior of hazard function. For different choices of α and θ it shows decreasing failure rate.

IV. Statistical Properties

In this section, the statistical properties of Length biased Quasi Lindley distribution are discussed.

I. Moments

Let X denotes the random variable of LBQL distribution with parameters θ and α, then the r[th] order moment of LBQL distribution is defined as

$$E(X^r) = \mu'_r = \int_0^\infty x^r f_L(x;\alpha,\theta)dx$$

$$= \int_0^\infty x^{r+1}\frac{\theta^2}{(\alpha+2)}(\alpha+\theta x)e^{-\theta x}dx$$

$$= \frac{\theta^2}{(\alpha+2)}\int_0^\infty x^{r+1}(\alpha+\theta x)e^{-\theta x}dx$$

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

after simplification,

$$E(X^r) = \mu'_r = \frac{\theta^2}{(\alpha+2)}\left(\frac{\alpha\Gamma(r+2)+\Gamma(r+3)}{\theta^{r+2}}\right) \qquad (7)$$

putting r = 1 in equation (7), the mean of LBQL distribution is given by

$$\mu_1' = E(X) = \frac{2(\alpha+3)}{\theta(\alpha+2)}$$

After putting r = 2,3 and 4 in equation (7), the second, third and fourth raw moments of Length biased Quasi Lindley distribution are obtained as,

$$\mu_2' = E(X^2) = \frac{6(\alpha+4)}{\theta^2(\alpha+2)}$$

$$\mu_3' = E(X^3) = \frac{24(\alpha+5)}{\theta^3(\alpha+2)}$$

$$\mu_4' = E(X^4) = \frac{120(\alpha+6)}{\theta^4(\alpha+2)}$$

Therefore,

$$Variance = \sigma^2 = \frac{2(\alpha^2+6\alpha+6)}{\theta^2(\alpha+2)^2}$$

$$S.D. = \sigma = \frac{\sqrt{2(\alpha^2+6\alpha+6)}}{\theta(\alpha+2)}$$

$$C.V = \frac{\sigma}{\mu} = \frac{\sqrt{2(\alpha^2+6\alpha+6)}}{2(\alpha+3)}$$

$$C.D.(\gamma) = \frac{\sigma^2}{\mu} = \frac{(\alpha^2+6\alpha+6)}{\theta(\alpha+2)(\alpha+3)}$$

II. Moment generating Function and Characteristic Function LBQLD

Let X follows LBQL distribution, then the moment generating function (MGF) of X is,

$$M_X(t) = \int_0^\infty e^{tx} f_L(x;\alpha,\theta)dx$$

$$= \int_0^\infty (1 + (tx) + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots) f_L(x,\alpha,\theta)dx$$

$$= \int_0^\infty \sum_{r=0}^{\infty} \frac{(tx)^r}{r!} f_L(x,\alpha,\theta)dx$$

$$= \sum_{r=0}^{\infty} \frac{(t)^r}{r!} \int_0^\infty x^r f_L(x,\alpha,\theta)dx$$

$$M_X(t) = \sum_{r=0}^{\infty} \frac{(t)^r}{r!} E(x^r) \quad (8)$$

Substituting value of $E(x^r)$ from equation (7) in equation (8),

$$M_X(t) = \sum_{r=0}^{\infty} \frac{(t)^r}{r!} \left\{ \frac{\theta^2}{(\alpha+2)} \left(\frac{\alpha\Gamma(r+2) + \Gamma(r+3)}{\theta^{r+2}} \right) \right\}$$

Similarly, the characteristic function of LBQL distribution can be obtained as,

$\varphi_X(t) = M_X(it)$

$$= \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \left\{ \frac{\theta^2}{(\alpha+2)} \left(\frac{\alpha\Gamma(r+2) + \Gamma(r+3)}{\theta^{r+2}} \right) \right\}$$

III. Harmonic Mean

Let X follows LBQL distribution then, the harmonic mean is obtained as

$$H = \int_0^{\infty} \frac{1}{x} f_L(x, \alpha, \theta) dx$$

$$= \int_0^{\infty} \frac{\theta^2}{(\alpha+2)} (\alpha + \theta x) e^{-\theta x} dx$$

$$= \frac{\theta^2}{(\alpha+2)} \left[\alpha \int_0^{\infty} e^{-\theta x}\, dx + \theta \int_0^{\infty} x\, e^{-\theta x} dx \right]$$

after simplification,

$$H = \frac{\theta(\alpha+1)}{(\alpha+2)}$$

V. Order Statistics for LBQL Distribution

Order statistics have central role in statistical theory. It deals with the ordered data that is necessary to take for quality control, reliability, hydrological and extreme values analysis. Suppose $X_{(1)}, X_{(2)}, \ldots\ldots\ldots\ldots, X_{(n)}$ be the j[th] order statistic and it is denoted by $X_{(j)}$.
The probability density function of the j^{th} order statistics $X_{(j)}$ for $1 \le j \le n$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} [F(x)]^{j-1} [1 - F(x)]^{n-j} f_1(x) \qquad (9)$$

Substitute the value from equation (5) and (6) in equation (9), the probability density function of j[th] order statistics of LBQL distribution is given as

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} \left[\frac{\alpha\gamma(2,\theta x) + \gamma(3,\theta x)}{(\alpha+2)} \right]^{j-1} \times \left[1 - \frac{\alpha\gamma(2,\theta x) + \gamma(3,\theta x)}{(\alpha+2)} \right]^{n-j} \frac{x\,\theta^2}{(\alpha+2)} (\alpha + \theta x) e^{-\theta x} \qquad (10)$$

For $j = 1$ in equation (10), therefore the probability density function of first order statistics of LBQL distribution is obtained as

$$f_{X_{(1)}}(x) = n \left[1 - \frac{\alpha\gamma(2,\theta x) + \gamma(3,\theta x)}{(\alpha+2)} \right]^{n-1} \times \frac{x\,\theta^2}{(\alpha+2)} (\alpha + \theta x) e^{-\theta x} \qquad (11)$$

Put $j = n$ in equation (10), the probability density function of n^{th} order statistics of LBQL distribution is given by,

$$f_{X_{(n)}}(x) = n\left[\frac{\alpha\gamma(2,\theta x)+\gamma(3,\theta x)}{(\alpha+2)}\right]^{n-1} \times \frac{x\,\theta^2}{(\alpha+2)}(\alpha+\theta x)e^{-\theta x} \quad (12)$$

VI. Entropy

The concept of Entropies points out the diversity, uncertainty, or randomness of a system and the entropies have large application in several fields such as probability & statistics, physics, communication theory and economics. Entropy of a random variable X is a measure of variation of the uncertainty.

I. Renyi Entropy

The entropy termed as Renyi entropy is important in ecology and statistics as index of diversity. Renyi entropy is an extension of Shannon's entropy. Renyi (1961) give an expression of the entropy function is defined by

$$e(\delta) = \frac{1}{1-\delta}\log\left(\int_0^\infty {f_L}^{\delta}(x;\alpha,\theta)dx\right)$$

where $\delta > 0$ and $\delta \neq 1$

$$e(\delta) = \frac{1}{1-\delta}\log\left(\int_0^\infty \left(\frac{\theta^2}{(\alpha+2)}x(\alpha+\theta x)e^{-\theta x}\right)^{\delta} dx\right)$$

$$e(\delta) = \frac{1}{1-\delta}\log\left(\left(\frac{\theta^2}{(\alpha+2)}\right)^{\delta}\int_0^\infty x^{\delta}(\alpha+\theta x)^{\delta}e^{-\theta\delta x}dx\right) \quad (13)$$

Using binomial expansion in equation (13),

$$e(\delta) = \frac{1}{1-\delta}\log\left\{\left(\frac{\theta^2}{(\alpha+2)}\right)^{\delta}\sum_{k=0}^{\infty}\binom{\delta}{k}(\alpha)^{\delta-k}\theta^k\int_0^\infty (x)^{\delta+k}e^{-\theta\delta x}dx\right\}$$

$$e(\delta) = \frac{1}{1-\delta}\log\left\{\left(\frac{\theta^2}{(\alpha+2)}\right)^{\delta}\sum_{k=0}^{\infty}\binom{\delta}{k}(\alpha)^{\delta-k}\theta^k\int_0^\infty (x)^{\delta+k+1-1}e^{-\theta\delta x}dx\right\}$$

$$e(\delta) = \frac{1}{1-\delta}\log\left\{\left(\frac{\theta^2}{(\alpha+2)}\right)^{\delta}\sum_{k=0}^{\infty}\binom{\delta}{k}(\alpha)^{\delta-k}\theta^k\frac{\Gamma(k+\delta+1)}{(\theta\delta)^{(k+\delta+1)}}\right\}$$

II. Tsallis Entropy

Tsallis entropy was introduced by Tsallis (1988) as a basis for generalizing the standard statistical mechanics. For a continuous random variable X, Tsallis entropy is defined as follows.

$$S_\lambda = \frac{1}{1-\lambda}\left(1 - \int_0^\infty {f_L}^\lambda(x;\alpha,\theta)dx\right)$$

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

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

$$S_\lambda = \frac{1}{1-\lambda}\left(1 - \left(\frac{\theta^2}{(\alpha+2)}\right)^\lambda \int_0^\infty x^\lambda\,(\alpha+\theta x)^\lambda e^{-\lambda\theta x}dx\right) \qquad (14)$$

Using binomial expansion in equation (14),

$$S_\lambda = \frac{1}{1-\lambda}\left\{1 - \left(\frac{\theta^2}{(\alpha+2)}\right)^\lambda \sum_{k=0}^{\infty}\binom{\lambda}{k}(\alpha)^{\lambda-k}\theta^k \int_0^\infty (x)^{\lambda+k}\, e^{-\lambda\theta x}dx\right\}$$

$$S_\lambda = \frac{1}{1-\lambda}\left\{1 - \left(\frac{\theta^2}{(\alpha+2)}\right)^\lambda \sum_{k=0}^{\infty}\binom{\lambda}{k}(\alpha)^{\lambda-k}\theta^k \int_0^\infty (x)^{\lambda+k+1-1}e^{-\lambda\theta x}\, dx\right\}$$

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

VII. Likelihood Ratio Test

Let $x_1, x_2, x_3, \ldots\ldots, x_n$ be a random sample from the LBQL distribution. To test the hypothesis

$H_0: f(x) = f(x;\alpha,\theta)$ Against $H_1: f(x) = f_L(x;\alpha,\theta)$

In order to test whether the random sample of length n has been drawn from length biased Quasi Lindley distribution or not the following test statistics is used

$$\Delta = \frac{L_1}{L_0} = \prod_{i=1}^{n}\frac{f_L(x;\alpha,\theta)}{f(x;\alpha,\theta)}$$

$$\Delta = \left(\frac{\theta(\alpha+1)}{(\alpha+2)}\right)^n \prod_{i=1}^{n} x_i$$

Reject the null hypothesis, if

$$\Delta = \left(\frac{\theta(\alpha+1)}{(\alpha+2)}\right)^n \prod_{i=1}^{n} x_i > k$$

$$\Delta = \prod_{i=1}^{n} x_i > k\left(\frac{(\alpha+2)}{\theta(\alpha+1)}\right)^n$$

$$\Delta^* = \prod_{i=1}^{n} x_i > k^* \quad \text{where } k^* = k\left(\frac{(\alpha+2)}{\theta(\alpha+1)}\right)^n > 0$$

For large sample size n, 2logΔ is distributed as chi-square distribution with 1 degree of freedom (df) and also *p*-value is obtained from the chi-square distribution. Thus, reject the null hypothesis, when the probability value is given by

$$P(\Delta^* > \beta^*)$$

Where $\beta^* = \prod_{i=1}^{n} x_i$ is less than specified level of significance and $\prod_{i=1}^{n} x_i$ is observed value of the statistics Δ*.

VIII. Bonferroni and Lorenz curves

The Bonferroni and Lorenz curves are given as

$$B(p) = \frac{1}{p\mu}\int_0^q x\, f_L(x;\alpha,\theta)dx$$

$$L(p) = pB(p) = \frac{1}{\mu}\int_0^q x\, f_L(x,\alpha,\theta)dx$$

Where $E(x) = \mu = \frac{2(\alpha+3)}{\theta(\alpha+2)}$ and $q = F^{-1}(p)$

$$B(p) = \frac{\theta(\alpha+2)}{p2(\alpha+3)}\int_0^q x^2 \frac{\theta^2}{(\alpha+2)}(\alpha+\theta x)e^{-\theta x}dx$$

$$B(p) = \frac{\theta^3}{p2(\alpha+3)}\int_0^q x^2\,(\alpha+\theta x)e^{-\theta x}dx$$

$$B(p) = \frac{\theta^3}{p2(\alpha+3)} \times \left(\alpha\int_0^q x^{3-1}e^{-\theta x}\,dx + \theta\int_0^q x^{4-1}e^{-\theta x}\;dx\right)$$

$$B(p) = \frac{\alpha\gamma(3,\theta q) + \gamma(4,\theta q)}{2(\alpha+3)p}$$

$$L(p) = pB(p) = \frac{\alpha\gamma(3,\theta q) + \gamma(4,\theta q)}{2(\alpha+3)}$$

IX. Maximum Likelihood Estimation

In this section, the maximum likelihood estimation of the parameters of Length biased Quasi Lindley (LBQL) distribution is discussed. Let $x_1, x_2, \dots, x_n$ be a random sample of size n from the LBQL distribution, then the corresponding likelihood function is given by

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

$$L(x;\alpha,\theta) = \left(\frac{\theta^2}{(\alpha+2)}\right)^n \prod_{i=1}^{n} x_i\,(\alpha+\theta x_i)e^{-\theta\sum_{i=1}^{n}x_i}$$

Takin log and solving likelihood function is obtained as follows

$$\frac{\partial \log L}{\partial \alpha} = \frac{-n}{(\alpha+2)} + \sum_{i=1}^{n}\frac{1}{(\alpha+\theta x_i)} = 0 \qquad (15)$$

$$\frac{\partial \log L}{\partial \theta} = \frac{2n}{\theta} + \sum_{i=1}^{n}\frac{x_i}{\alpha+\theta x_i} - \sum_{i=1}^{n} x_i = 0 \qquad (16)$$

The MLE of the parameters cannot be obtain in close form. The exact solution of above equation for unknown parameters is not possible manually. So, we can solve above equations with the help of *R* Software using (optim function, nlminb (), nlm ()).

To obtain confidence interval we use the asymptotic normality tests. If as $\hat{\lambda} = (\hat{\alpha}, \hat{\theta})$ denote the MLE of $\lambda = (\alpha, \theta)$, state the results as follows:

$$\sqrt{n}(\hat{\lambda} - \lambda) \to N(0, I^{-1}(\lambda))$$

Where $I(\lambda)$is Fisher's Information Matrix is

$$I(\lambda) = -\frac{1}{n}\begin{bmatrix} E\left(\frac{\partial^2 \log l}{\partial^2\theta}\right) & E\left(\frac{\partial^2 \log l}{\partial\theta\partial\alpha}\right) \\ E\left(\frac{\partial^2 \log l}{\partial\alpha\partial\theta}\right) & E\left(\frac{\partial^2 \log l}{\partial^2\alpha}\right) \end{bmatrix}$$

Where

$$\frac{\partial^2 \log l}{\partial^2\alpha} = \frac{n}{(\alpha+2)^2} - \sum_{i=1}^{n}\frac{1}{(\alpha+\theta x_i)^2}$$

$$\frac{\partial^2 \log l}{\partial^2\theta} = \frac{-2n}{\theta^2} - \sum_{i=1}^{n}\frac{{x_i}^2}{(\alpha+\theta x_i)^2}$$

$$\frac{\partial^2 \log l}{\partial\theta\partial\alpha} = \frac{\partial^2 log\ l}{\partial\alpha\partial\theta} = -\sum_{i=1}^{n}\frac{x_i}{(\alpha+\theta x_i)^2}$$

Since λ being unknown, $I^{-1}(\lambda)$ is estimated by $I^{-1}(\hat{\lambda})$and this can be used to obtain asymptotic confidence intervals for α and θ.

X. Application

In this section, three real life data set are studied for the purpose of illustration to show the usefulness

and flexibility of the LBQL distribution.

To compare the length biased Quasi Lindley (LBQL) distribution with QL, Power Lindley (PL), Exponential (Exp.) distributions, the criteria like Bayesian information criterion (BIC), Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICC), HQIC are used and parameters are estimated using ML method of estimation.

The real-life data sets are given as follows:

Data set I: The first real life data set represents the breaking stress of carbon fibres (in Gba) observed and reported by Nichols and Padgett (2006) and is executed below in table 1.

Table 1. *Data consists of breaking stress of carbon fibres (in Gba) observed by Nichols and Padgett (2006).*

Data set I								
3.70	2.74	2.73	2.50	3.60	3.11	3.27	2.87	1.47
3.11	3.56	4.42	2.41	3.19	3.22	1.69	3.28	3.09
1.87	3.15	4.90	1.57	2.67	2.93	3.22	3.39	2.81
4.20	3.33	2.55	3.31	3.31	1.25	4.38	1.84	0.39
3.68	2.48	0.85	1.61	2.79	4.70	2.03	1.89	2.88
2.82	2.05	3.65	3.75	2.43	2.95	2.97	3.39	2.96
2.35	2.55	2.59	2.03	1.61	2.12	3.15	1.08	2.56
2.85	1.80	2.53						

Data set 2: The second real life data set represent the fatigue life of some aluminum's coupons cut in specific manner (see, Birnbaum and Saunders, 1969). The dataset (after subtracting 65) is given below in table 2.

Table2. *The fatigue life of some aluminum's coupons cut in specific manner (Birnbaum and Saunders, 1969).*

Data set II								
5	25	31	32	34	35	38	39	39
40	42	43	43	43	44	44	47	47
48	49	49	49	51	54	55	55	55
56	56	56	58	59	59	59	59	59
63	63	64	64	65	65	65	66	66
66	66	67	67	67	68	69	69	69
69	71	71	72	73	73	73	74	74
76	76	77	77	77	77	77	77	79
79	80	81	83	83	84	86	86	87
90	91	92	92	92	92	93	93	94
97	98	98	99	101	101	103	105	109
139	147							

Data set 3:

This data set presented in Murthy et al. (2004) and used by some researchers. This data set present failure times for a particular windshield model including 85 observations that are classified as failure times of windshields.

Table3.*failure times for a particular windshield model including 85 observations that are classified as failure times of windshields.*

Data set III						
0.040	1.866	2.385	3.443	0.301	1.876	2.481
3.467	0.309	1.899	2.610	3.478	0.557	1.911
2.625	3.578	0.943	1.912	2.632	3.595	1.070
1.914	2.646	3.699	1.124	1.981	2.661	3.779
1.248	2.010	2.688	3.924	1.281	2.038	2.822
3.000	4.035	1.281	2.085	2.890	4.121	1.303
2.089	2.902	4.167	1.432	2.097	2.934	4.240
1.480	2.135	2.962	4.255	1.505	2.154	2.964
4.278	1.506	2.190	3.000	4.305	1.568	2.194
3.103	4.376	1.615	2.223	3.114	4.449	1.619
2.224	3.117	4.485	1.652	2.229	3.166	4.570
1.652	2.300	3.344	4.602	1.757	2.324	3.376
4.663						

R software is used for determining the estimation of unknown parameters and is also used for estimating the model comparison criterion values (AIC, BIC, AICC, HQIC) and -2logL. To compare the Length biased Quasi Lindley distribution with Quasi Lindley and Power Lindley, Exponential distributions, the criterion like AIC (Akaike information criterion), AICC (corrected Akaike information criterion), BIC (Bayesian information criterion) and HQIC (Hannen-Quinn information criterion) are used for comparison. The better distribution corresponds to lesser values of AIC, AICC, BIC, HQIC and -2logL.

Table 4. *Estimate and goodness of fit measures under considered distribution based on data set I.*

Distribution	M.L. E		- 2**log**L	AIC	AICC	BIC	HQIC
	$\hat{\alpha}$	$\hat{\theta}$					
LBQL Distribution	**-0.3883002**	**1.1745**	**199.7838**	**203.7838**	**203.9743**	**208.1631**	**205.5142**
QL Distribution	-0.3401128	0.9116	204.4596	208.4596	208.6501	212.8389	210.1901
Exponential Distribution	0.3900	2.36944	245.8762	249.8762	249.9397	258.2555	255.6067
PL Distribution	0.5781	1.1286	490.4955	494.4955	494.5590	498.8748	496.2260

Table 5. *Estimate and goodness of fit measures under considered distribution based on data set II.*

Distribution	M.L. E		- 2**log**L	AIC	AICC	BIC	HQIC
	$\hat{\alpha}$	$\hat{\theta}$					
LBQL Distribution	**-0.16555**	**0.04489**	**899.4582**	**903.4582**	**903.499**	**908.6884**	**905.5756**
QL Distribution	-0.14116	0.03144	982.2110	986.2110	986.2518	991.4412	988.3284
Exponential Distribution	5.000	63.83168	1041.562	1045.562	1045.603	1050.792	1047.679

PL Distribution	0.1447	1.09255	1688.346	1692.346	1692.387	1697.577	1694.463

Table 6. *Estimate and goodness of fit measures under considered distribution based on data set III.*

Distribution	M.L. E		- 2**log**L	AIC	AICC	BIC	HQIC
	$\hat{\alpha}$	$\hat{\theta}$					
LBQL Distribution	**0.1458**	**1.14416**	**276.632**	**280.632**	**280.6808**	**285.5173**	**282.5970**
Exponential Distribution	0.040	2.5626	276.7906	280.7906	280.8394	285.6759	282.7556
QL Distribution	0.00362	0.77904	289.5131	293.5131	293.5619	298.3984	295.4781
PL Distribution	0.58946	1.18093	594.0202	598.0202	598.069	602.9055	599.9852

From table (4), (5) and (6) it can be seen that the value of the statistics -2logL, AIC, BIC, AICC and HQIC of the Length biased Quasi Lindley distribution are comparatively smaller than the other distributions on a real-life data set. Therefore, the result shows that the Length biased Quasi Lindley distribution provides a significantly better fit than other models. So, it can be chosen to model the life testing data.

XI. Conclusion

In this paper, the Length biased Quasi Lindley distribution is proposed as a new extension of Quasi Lindley distribution. The newly introduced distribution is generated by using the Length biased techniques and taking the Quasi-Lindley distribution as the base distribution. The various statistical properties of the proposed distribution have been derived and discussed. Supremacy of the new distribution in real life is established with demonstration of real-life data sets and it is found from the results of data sets that the Length biased Quasi Lindley distribution performs better than the Quasi Lindley, Power Lindley and Exponential distributions.

References

[1] A. Renyi; On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Contributions to the Theory of Statistics, Berkeley, California: University of California Press, (1): 547–561, (1961).

[2] C. R. Rao; On discrete distributions arising out of method of ascertainment in classical and Contagious Discrete, G.P. Patiled; Pergamum Press and Statistical publishing Society, Calcutta.320-332 (1965).

[3] C. Tsallis; Possible generalization of Boltzmann-Gibb's statistics. Journal of statistical physics, 52(1-2), 479-487, (1988).

[4] D. N. P. Murthy; M. Xie; R. Jiang; Weibull models, John Wiley & Sons, New York, (2004).

[5] D. R. Cox; Some sampling problems in technology in New Development in Survey Sampling, Johnson, N. L. and Smith, H., Jr.(eds.) New York Wiley- Interscience, 506- 527 (1969).

[6] J. Lappi; R.L. Bailey; Forest Science, 33 725- 739 (1987).

[7] K. A. Mir; A. Ahmed; J. A. Reshi; Structural properties of length biased beta distribution of first kind, American Journal of engineering Research, 02(02) 01-06 (2013).

[8] M. D. Nichols; W. J. Padgett; "A bootstrap control chart for Weibull percentiles", Quality & Reliability Engineering International, 22(2) 141-151 (2006).

[9] M. Zelen; Problems in cell kinetic and the early detection of disease, in Reliability and Biometry, F. Proschan& R. J. Sering, eds, SIAM, Philadelphia, 701-706 (1974).

[10] O. Jones; R. Maillardet; A. Robinson; Introduction to scientific Programming and Simulation Using R, New York: Taylor and Francis Group, (2009).

[11] R. A. Fisher; The effects of methods of ascertainment upon the estimation of frequencies, Ann. Eugenics, 6,13-25 (1934).

[12] R. Shanker; A. Mishra; A Quasi-Lindley distribution, African Journal of Mathematics and Computer Science Research,6(4),64-71 (2013).

[13] R. Shenbagaraja; A. A. Rather; C. Subramanian; On Some aspects of Length biased technique with Real life data, Science, technology & Development, 8(9) 326-335 (2019)

[14] S. Mudasir; S. P. Ahmad; Structural properties of length biased Nakagami distribution, International Journal of Modern Mathematical Sciences, 13(3) 217-227 (2015).

[15] Z.W. Birnbaum; S. C. Saunders; Estimation for a Family of Life Distributions with Applications to Fatigue, Journal of Applied Probability, 6(2), (1969).

IDENTIFICATION OF SPATIAL RELATIONS IN MATHEMATICAL EXPRESSIONS

Sridevi Ravada
Dept.of Information Technology
Gayatri College of Engineering for Women
Visakhapatnam, India
srideviravada@gvpcew.ac.in

Sudheer Gopinathan
Dept.of Information Technology & Mathematics
Gayatri College of Engineering for Women
Visakhapatnam, India
g.sudheer@gvpcew.ac.in

D.Lalitha Bhaskari
Dept. of Computer Science & System Engg ,
Andhra University
Visakhapatnam, India
lalithabhaskari@yahoo.co.in

Abstract

The automatic recognition of mathematical expressions in digital content is a challenging task due to the complex spatial relationships between the symbols involved in the expression. The accuracy of the recognition is dependent on a variety of factors that includes nature of the input medium. The reliability of the performance of the system is dependent on the identification of the spatial relationships in an expression. Symbol recognition and structural analysis are the two important stages in the recognition process. In the present work these two stages are considered using the concepts of connected components and minimum spanning tree. For our analysis, we have created a database of 500 expression images drawn from standard databases and the experimental results are reported on them.

Keywords: minimum spanning tree, connected, spatial, segmentation.

I. Introduction

Continuous research and development over the last few decades have enabled optical character recognition (OCR) systems to achieve a sufficient level of maturity in the recognition and retrieval of information from digital content. However, recognition and retrieval of images, tables, diagrams and mathematical expressions (ME) are yet to reach a comfortable level of applicability [1]. The amount of work in the literature that is directed towards the extraction and recognition of ME is a mature field of study and the research efforts in this area have been surveyed in the works [2,3].The recognition of ME is a challenging pattern recognition problem that includes segmentation ambiguities, symbol recognition challenges and ambiguity of meaning. The problem finds application in several areas of science and engineering that include document searching / editing , computer algebra systems, tutoring systems, and mathematical information retrieval to name a few[1]. The existing OCR systems have difficulty in converting the ME present in scientific/technical documents into a corresponding digital/electronic form for recognition. Instead of developing OCR systems specifically for scientific documents, the emphasis has been on including mathematical /math OCR module into the existing OCR systems. The digital document analysis in OCR systems typically consists of the pre-processing stage followed by the determination of the physical layout and logical structure of the document. The pre-processing is basically concerned with the removal/correction of noise, artifacts, unwanted variations introduced during the document generation stage and is an essential part of the OCR system. The physical layout of a document is basically its geometric structure and its analysis aims to decompose the

document into a hierarchy of homogenous/similar regions. The logical structure is the actual content of the document and its analysis is aimed at understating the logical/ functional entities in a document along with their interrelationships [4]. The document structure analysis that includes these tasks get complicated due to the variations in the source of the input documents that consists of vector graphics, historical documents, printed documents etc., each having distinct physical and layout structures.

In the development of math- OCR module, the math zones are to be segmented from the input documents either manually or through semi or fully automated segmentation logic. Though research into the segmentation of math zones from documents have been carried out in the past few decades, challenges still remain as the developed methods have not been at a sufficient level to be adopted in realistic application scenarios [5]. It is to be noted that, no one technique for layout analysis completely dominates another and improving these methods in an active area of research[1]. The present work is concerned with the main module of a math OCR which is the math expression recognition module that consists of the two stages: Component character recognition and structural analysis [6]. The recognition of mathematical symbols is a difficult problem due to the presence of a large character set with a variety of font styles and a range of font sizes coupled with a set of symbols having an enormous range of possible scales [7,2]. The symbols occurring in the ME are arranged around different operators that form different layouts, some of which are one dimensional, while others are two dimensional. The 2D structure induced by the operators appear in normal or nested modes increasing the complexity of the expression which is further increased by the number of horizontal lines on which the constituent symbols are arranged[8]. In the structural analysis part, the spatial relationships between symbols are analysed to capture the structure of the expression, together with the logical meaning to aid the recognition process. The complex structure of the ME makes structural analysis a challenging task even when all the symbols have been properly recognized [9].

Document mathematical expression recognition is generally considered to be printed mathematical expression recognition[10].Over the past few decades, many excellent methods have been proposed in the field of ME recognition, however the search for an optimum technique/method is far from being over. In recent years, the focus has shifted to handwritten mathematical expression recognition. Though the techniques for both printed and handwritten ME recognitions are almost similar, handwritten ME recognition is difficult and is more or less dependent on the strokes used by the individual person. In addition to segmentation and character classification, the spatial relation classification is one of the dominant problem associated with ME recognition[11].

The ME can be represented in many formats such as symbol layout tree, operated tree, label graphs, Tex and so on and the output of the ME recognition systems are to be put into these formats for reconstruction. A network representation of the ME has helped researchers to utilize graph theory concepts in tasking symbol segmentation and recognition. The minimum spanning tree (MST) problem is one of the famous mathematical problems in computer science that has been adopted suitably by many workers for the development of mathematical OCR[12]. In unconstrained handwritten documents, separation of text lines is a challenge because of the skewed, curved and non-uniform structure of the text .[13] developed an approach for text line segmentation in unconstrained handwritten Chinese document using an approach based on the minimum spanning tree .A variation of the MST problem was applied to mathematical OCR by [15] utilizing the notion of candidate selection and link-label selection. A structural analysis method for the recognition of online handwritten ME based on a MST construction and symbol dominance was presented in [14]

The present work is concerned with the recognition of ME. Two approaches are utilized in the process. The first approach considers bounding box, labeling matrix and minimum spanning tree to group similar component and study the spatial relationship between them. The second

approach considers the separation of horizontal lines into five regions and allot the components of ME into these regions for the identification of base symbol subscript, superscript and sub expressions. The relationship between these components is then obtained utilizing a minimum spanning tree.

II. Characteristics of Mathematical expressions

Symbol recognition and structural analysis are the two involved activities that are crucial in understanding a mathematical expression. Pre-processing, segmentation, recognition, identification/recognition of spatial and logical relationships and construction of meaning of the symbol involved in the expressions are the processes involved in these two activities [16]. A mathematical expression that is printed/handwritten can be viewed as a collection of symbols with spatial relationships among them. The problem of searching for the most likely Interpretation for a given set of mathematical inputs reduces to the problem of searching for the most likely symbol identities with the likely spatial relationships [17].

Understanding the essential features of the ME aids the processes involved in recognition. The mathematical expression / formulae are represented with various kinds of entities such as

i) Arabic digits, Greek, Latin, Roman, calligraphic letters etc., in addition to the English characters.

ii) Mathematical symbols that include bracket symbols, arrow symbols, miscellaneous symbols etc.,

iii) Mathematical operators – logical, set and relational etc., function names

In a ME, the alphabetic characters occur with a variety of typefaces such as normal, bold, italic etc., and can be touching, broken, overlaid etc.,. In addition, the alphabetic letter can be of type ascender (b,d,h,k…..) or descender (g,j,p,y…) or normal (a,m,r,n....). In the case of operators, factors such as operator range, operator precedence, symbol identity, relative symbol size and case interact in a complex way and understanding these provide a clue for the structural analysis of ME[13]. Some of the features can be further extended as :

1. Relative symbol placement is important in the identification of operators. For e.g.; the commonly used basic spatial relationships: left (L), Right (R), Below(B), Above (A), above right(AR) and above left(AL), below right (BR), below left (BL) and can be visualized as shown in Figure 1. Identification of these relationships is an important aspect of a recognition system.

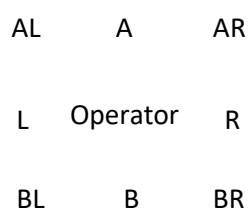


Figure 1: *Relative placement of symbols in an expression*

2.The symbol/expression have geometric complexity which is determined by the horizontal width occupied by the expression. The component symbols in the expression are arranged in a number of horizontal lines that increased the complexity as seen from a typical example shown in Figure 2.

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} i^2}{\sum_{i=1}^{n} i^3}$$

Figure 2: *Illustration of geometry complexity in an expression*

3. Nested structure of the expression as seen in Figure 3a, in the figure, the argument of arc tan, that contains the square root operator and further involves fraction operator is a nested structure. Similarly a matrix image consists of element that may have fractions, square roots etc., as in Figure 3b.

$$Y=\tan^{-1}\sqrt{\frac{1-\cos x}{1+\cos x}} \qquad \begin{pmatrix} 2x & \sqrt{\frac{x}{2}} \\ \frac{x+1}{x^2+1} & 4 \end{pmatrix}$$

(a) **(b)**

Figure3: *Illustration of nested structure in an expression*

4.Links representing the structure of ME. Consider the expression $\overline{x_t} = 5x_{t-k} + p^2$.

The horizontal/base line together with the spatial relationships of the expressions are shown in the Figure 4. Understanding these is a part of structural analysis.

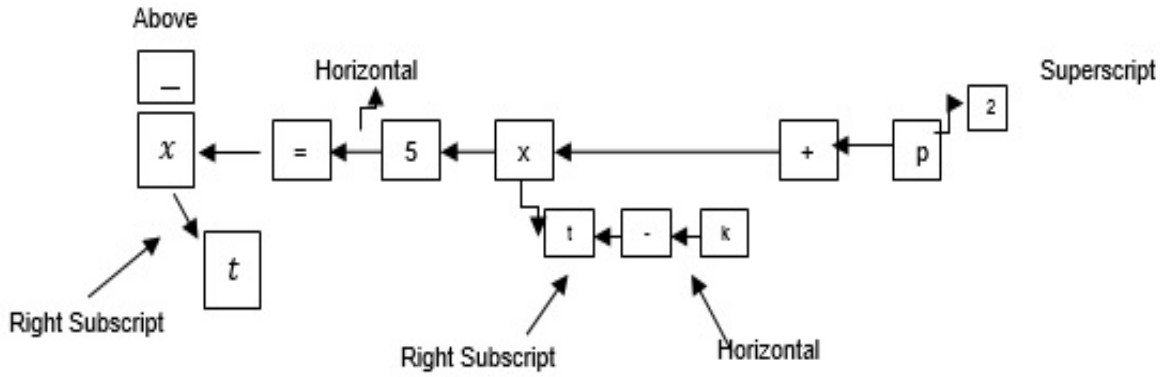


Figure 4: *Spatial relationships in an expression*

.

III. Proposed methodology

The first stage of the proposed method consists in symbol segmentation using connected component labelling method. For a given input image containing a mathematical expression, binarization is carried out. The binary image is scanned pixel by pixel and its pixels are grouped into components based on their connectivity. Once the connected components are extracted, their bounding boxes are obtained.

The bounding box of connected component (symbol) is defined to be the smallest rectangle which circumscribes the connected component (CC) [18]. Labels are then assigned to the CCs. The area, aspect ratio and centroid of the CCs are calculated. The centroid is a point attribute that helps determine the symbols location. Most of the single CC symbols can be segmented in this process but symbols like : , i , % , = , ≥, ≤ , ÷ , etc., which are multi CC are not segmented.

For example division (÷), a multi CC consists of three CC that consist of one horizontal line (-) and two dots (.). to resolve the problem of multi CC symbols the minimum spanning tree is utilized.

Considering the centroid of the CCs as vertices and distances between the centroids as weights a minimum spanning tree is constructed. The minimum distance between the components is used to resolve the ambiguity in the multi CC symbols like || , i , % , = , ≥, ≤ , ÷ etc., .

The resolved multi CCs are labelled as composite symbols and the spanning tree is updated to

group the similar components of remaining symbols if any in the expression. The mathematical functions like min, max, SIN, COS are also resolved through this process and symbol segmentation of ME is thus carried out.

In the second stage, the structural analysis part is considered. Here the spatial relationship

between the symbols is identified. Horizontal profiling is used to split the math block containing the ME into five regions as shown in fig.5. Level three is the main base line of the ME and contains the parent/ dominant symbols. Level 1, level 2 and level 4, level 5 contain the super, super expressions, subscripts, sub expressions. The peak values in the profiles together with some heuristic rules are employed to get a coarse split of the region. The symbol component height and the y-coordinate of the centroid of each component is used to allot the symbol into the 5 regions.

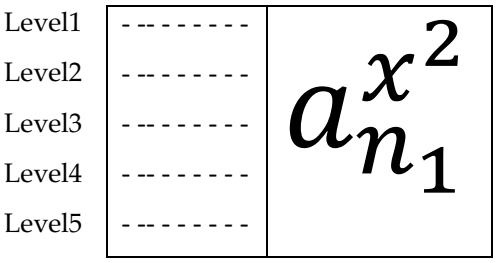


Figure 5: *A sample layout of the decomposition of a ME into five levels*

The statistical properties of the symbols together with the centroid values are utilized in finding the dominant base line. This is followed by the coarse identification of the relationship between two symbols such as above, below, in the same row, super script, sub script, prescript, nested using the relative geometric attributes of the bounding boxes of symbols. The horizontal and vertical projection profiles of CCs aid the process. The relationship tree based on the symbol dominance is then generated by constructing a minimum spanning tree that finds the different relationships among the components of the formulae.

IV. Results and discussion

The proposed methodology is tested on about 500 Mathematical Expressions taken from various mathematical documents collected from the internet including the database of Infty Project. The MEs considered cover various branches of mathematics. For handwritten samples we have considered only 20 expressions. The samples considered contain almost all mathematical expressions. The current algorithm is developed in MATLAB and Python. In the work, the connected components are sorted in the increasing order of x-coordinate and incase two or more (CC) have same x-coordinate, the y-coordinate are considered for sorting. The ordered CCs are scanned from left to right and basing on the spatial features the CCs are combined to form a single composite symbol. For e.g.; we consider some mathematical expressions and as a sample the components of the expressions are listed in Figure 6.The minimum spanning tree is used to find the spatial relations between the symbols of the expressions and the relationship obtained is shown in Figure 7. For the expression on printed documents the spatial relations could be identified correctly to the tune of 95%, however in handwritten cases the accuracy is around 60%. Further the element extraction could be accomplished with great ease.

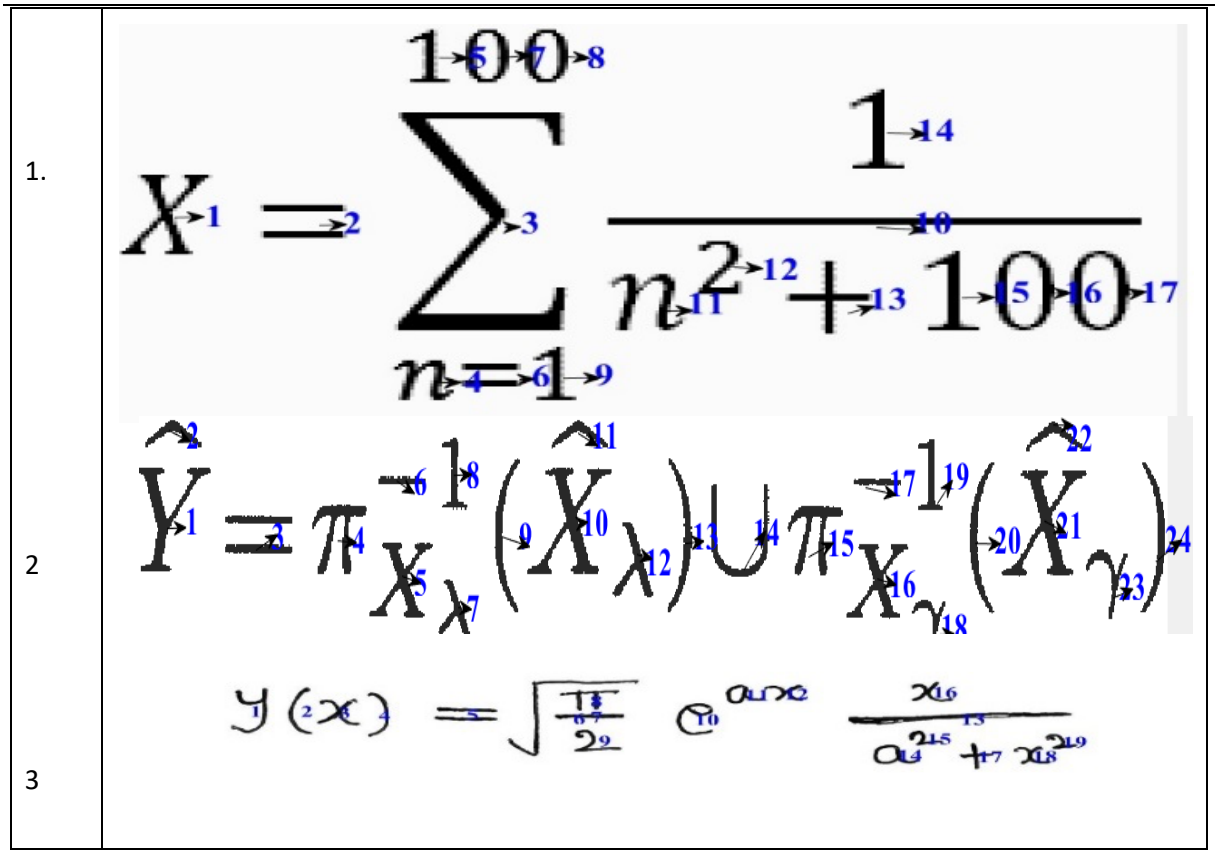


Figure 6: *Sample identification of different components of an expression*

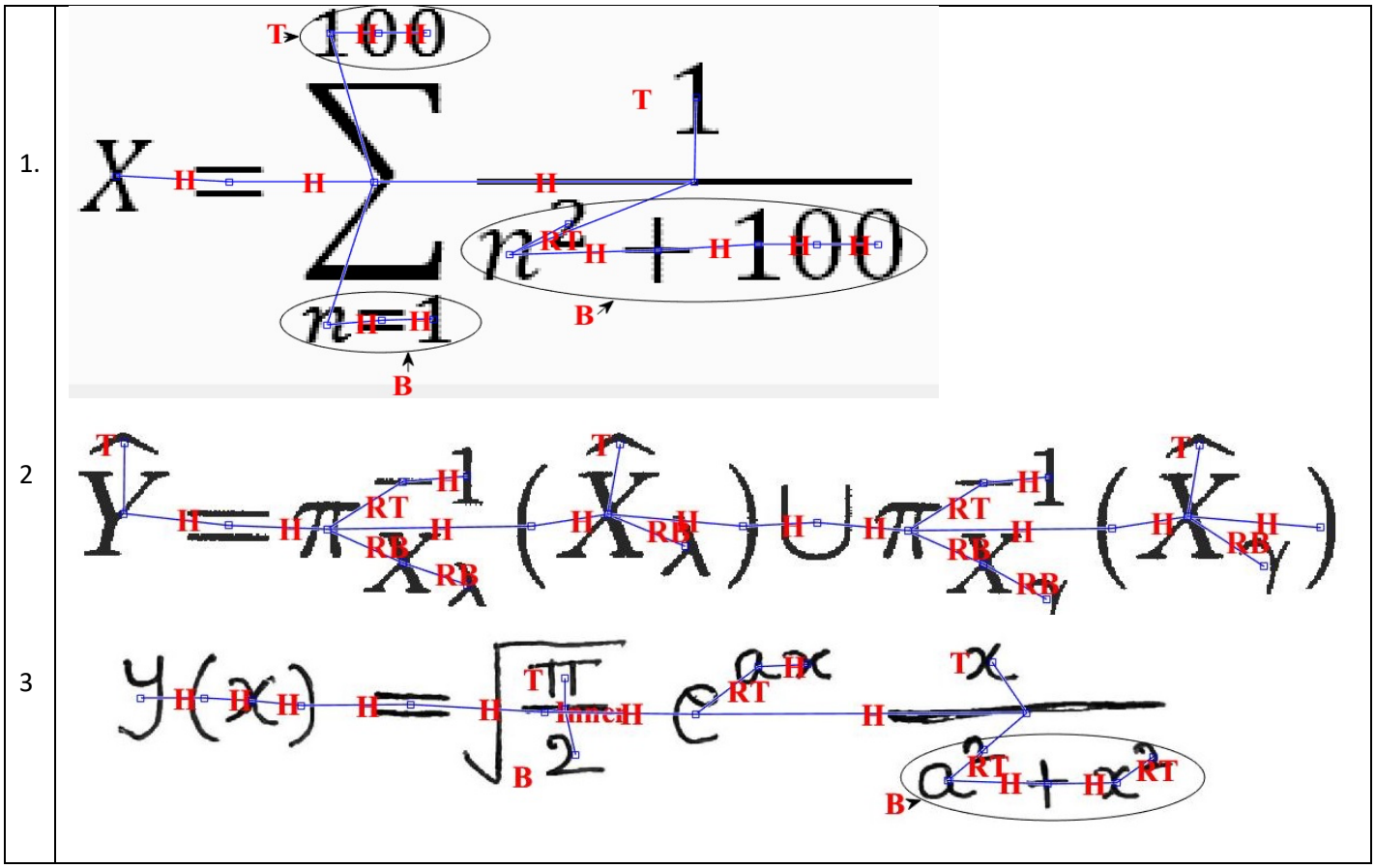


Figure 7: *The spatial relation between symbols in an expression*

V. Conclusion

In this paper, we focus on symbol segmentation and structural analysis of mathematical expressions in both printed and handwritten documents. The multiconnected components and context dependent symbols are resolved using the minimum spanning tree. The approach has been tested on around 500 MEs collected over the internet and the results are reported on them. We hope a good classifier can be added to the existing methodology to increase the accuracy on handwritten MEs.

References

[1] Zanibbi, R. and Blostein, D. (2012). Recognition and retrieval of mathematical expressions, International Journal on Document Analysis and Recognition. 15:331–357.

[2] Zanibbi, R. and Blostein, D. and Cordy, J.R. (2002). Recognizing mathematical expressions using tree transformation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(11):1455-1467.

[3] Chan, K. and Yeung, D. (2000). Mathematical expression recognition: a survey, International Journal on Document Analysis and Recognition,3:3–15.

[4] Namboodiri, A.M. and Jain, A.K. (2007). Document Structure and Layout Analysis In: Chaudhuri B.B. (eds) Digital Document Processing. Advances in Pattern Recognition. Springer, London, 29-48.

[5] Lin, X. Gao, L., Tang, Z. Baker, T. and Sorge, V. (2014). Mathematical formula identification and performance evaluation in PDF documents, International Journal on Document Analysis and Recognition. 17: 239–255.

[6] Aly, W. Uchida, S. and Suzuki, M. (2008). Identifying subscripts and superscripts in mathematical documents, Mathematics in Computer Science, 2:195-209.

[7] Garain, U. Chaudhuri, B.B. and Ghosh, R.P. (2004). A multiple-classifier system for recognition of printed mathematical symbols, Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004 1:380-383.

[8] Garain, U. and Chaudhuri, B.B. (2005). A corpus for OCR research on mathematical expressions, International Journal on Document Analysis and Recognition. 7:241–259.

[9] Pavan Kumar, P. Agarwal, A. and Bhagvati, C. (2018). Isolated structural error analysis of printed mathematical expressions, Pattern Analysis and Applications. 21:1097-1107.

[10] Huang, J. Tan, J. and Bi, N.(2020). Overview of Mathematical Expression Recognition, In book: Pattern Recognition and Artificial Intelligence, Proceedings International Conference on Pattern Recognition and Artificial Intelligence ICPRAI 2020, Y.Lu et al(Eds),LNCS 12068, 41-54.

[11] Zhelezniakov, D. Zaytsev, V. and Radyvonenko, O. (2021). Online Handwritten Mathematical Expression Recognition and Applications: A Survey. IEEE Access, 9:38352-38373.

[12] Fujiyoshi, A. and Suzuki, M. (2011). Minimum Spanning Tree Problem with Label Selection. IEICE Transactions on Information and Systems .E94-D(2).

[13] Yin, F., Liu, C.L. (2009). Handwritten Chinese text line segmentation by clustering with distance metric learning, Pattern Recognition,42: 3146-3157.

[14] Tapia, E., Rojas, R. (2004). Recognition of On-line Handwritten Mathematical Expressions Using a Minimum Spanning Tree Construction and Symbol Dominance. In: Llados, J., Kwon YB. (eds). Graphics Recognition Recent Advances and Perspectives. GREC 2003. LNCS, 3088,329-340.

[15] Fujiyoshi, A. and Suzuki, M. (2010). A variation of the minimum spanning tree problem for the application to mathematical OCR. Journal of Math-for-Industry,2:183-197.

[16] Blostei, D. and Grbavec, A. (1997). Recognition of Mathematical notation, In Handbook of Character Recognition and Document Image Analysis, 557-582.

[17] Rhee, T.H. and Kim, J.H. (2009). Efficient search strategy in structural analysis for handwritten mathematical expression recognition,Pattern Recognition, 42(12):3192-3201,

[18] Ha, J. Haralick, R.M. and I. T. Phillips. (1995). Recursive X-Y cut using bounding boxes of connected components. Proc of 3rd Int Conference on Document Analysis and Recognition, IEEE, 2:952-955.

On Discrete Scheduled Replacement Model of a Single Device Unit

Tijjani A. Waziri
•
School of Continuing Education, Bayero University Kano, Nigeria.
tijjaniw@gmail.com

Abstract

This paper considered a device which is subjected to three types of failures (category I, category II and category III). Also the paper tries to combine discrete age replacement model with minimal repair, where it dealt with a discrete scheduled replacement policy. Category I failure is an un-repairable failure, which occurs, suddenly, and if it occurs, the device is replaced completely, while category II and category III failures are repairable failures, which occurs, due to time and usage, and the two failures are rectified with minimal repair. To investigate the characteristics of the model constructed and determine optimum replacement number (N^*) of the device, a numerical example is provided, where it is assumed that the rate of the three categories of failures follow Weibull distribution.

Keywords: category, discrete, number, optimal, replacement, scheduled

I. Introduction

Most systems deteriorate and subsequently fail due to age and usage. These deficiencies have a detrimental impact on sales, the production of faulty goods and the delay in the provision of customer services. For these reasons, many optimal replacement policies have been built by several researchers to minimize excessively high running costs and prevent sudden failure of systems. For certain purposes, such as shortage of spare units, lack of money or staff, or inconvenience of time needed to complete replacement, an operating unit may often not be replaced at the exact optimum replacement times, but in idle periods, a unit can be replaced instead. Aven and Castro [1] constructed a minimal replacement policy for a system subject to two types of failures, which determined optimal replacement time for the system. Briš *et al.* [2] presented a new approach for optimizing a complex system's maintenance strategy that respects a given reliability constraint. Chang [3] considered a device that faces two types of failures (repairable and non-repairable) based on a random mechanism. Coria *et al*. [4] introduced a method of analytical optimization for preventive maintenance policy with historical failure time data. Enogwe *et al.* [5] used the distribution of the probability of failure times and come up with a replacement model for items that fails un-notice. Fallahnezhad and Najafian [6] investigated the number of spare parts and installations for a unit and parallel systems, so as cut down the average cost per unit time. Jain and Gupta [7] studied optimal replacement policy for a repairable system with multiple vacation and imperfect coverage. Lim *et al.* [8] studied the characteristics of some age substitution policies. Liu *et al.* [9] developed mathematical models of uncertain reliability of some multi-component systems.

Malki *et al.* [10] analyzed age replacement policies of a parallel system with stochastic dependency. Murthy and Hwang [11] presented that, the failures can be reduced through effective maintenance actions (in a probabilistic sense), and such maintenance actions can occur either at discrete time instants or continuously over time. Nakagawa [12] modified the continuous standard age replacement for a unit, and come up with a discrete replacement model for the unit. Nakagawa *et al.* [13] presented the advantages of some replacement policies. Safaei *et al.* [14] studied the optimal preventive maintenance action for a system based on some conditions. Sudheesh *et al.* [15] studied age replacement policy in discrete approach. Tsoukalas and Agrafiotis [16] presented a new replacement policy warrant for a system with correlated failure and usage time. Waziri and Yusuf [17] constructed an age replacement model for a parallel-series system based on some proposed policies. Xie *et al.* [18] assessed the effects of safety barriers on the prevention of cascading failures. Yaun and Xu [19] studies a cold standby repairable system with two different components and one repairman who can take multiple vacations, where they assumed that, if there is a component which fails and the repairman is on vacation, the failed component will wait for repair until the repairman is available. Yusuf and Ali [20] constructed a minimal repair replacement model for two parallel units in which both units operate simultaneously, such that, the two components are two types of failures. Yusuf *et al.* [21] analyzed the characteristics of reliability and availability of certain number of devices. Zhao *et al.* [22] proved that age replacement policy is optimal among all replacement policies.

The key contribution of this study is to come up with a discrete scheduled replacement model for a device that is exposed to three categories of failures, in order to (i) provide opportunity to replace a system at the ideal time (ii) provide the ability to skip such special hours to avoid, and (iii) investigate those aspects of the discrete model of scheduled replacement requiring limited repair.

II. Methods

Reliability measures namely reliability function and failure rates are used to obtain the expressions of discrete scheduled replacement model involving minimal repair based on some assumptions. A numerical example was given for the purpose of investigating the characteristics of the model constructed.

III. Notations

- C_2 : cost of repair due to failure of category II.
- C_3 : cost of repair due to failure of category III.
- C_p : cost of scheduled replacement at NT, for $N = 1, 2, 3$ ....
- C_r : cost of unscheduled replacement due to failure of category I.
- $C(N)$: replacement cost rate in one replacement cycle.
- N^*: the device's optimum discrete scheduled replacement time.
- $r_1(t)$: rate of category I failure.
- $r_2(t)$: rate of category II failure .
- $r_3(t)$: rate of category III failure.
- $R_1(t)$: reliability function due to category I failure.

IV. Description of the System

Consider a device which is subjected to three independent types of failures, which are named as category I, category II and category III, such that, all the three failures arrives according to non-

homogeneous Poisson process. It is assumed that, category I failure is unrepairable one, while category II and category III failures are repairable failures. The device is replace with new one whenever it reaches scheduled replacement time NT $(N = 1, 2, 3, \dots)$ for a fixed T or at category I failure, whichever occurs first.

V. Discrete Scheduled Replacement Model

This section considers a fundamental discrete scheduled replacement model involving minimal repair.

Assumptions for this model:

1. Category I failure is un-repairable one, while category II and category III failures are repairable failure.
2. Category I, Category II and Category III failures arrives according to a non-homogeneous Poisson process with failure intensity $r_1(t)$, $r_2(t)$ and $r_3(t)$, respectively, such that: $r_3(t) \geq r_2(t) \geq r_1(t)$.
3. The cost of replacement/minimal repair follows the order : $C_r > C_p > C_2 > C_3$.
4. All three failures are detected instantaneously.
5. When needed, all the resources required are available.
6. If the device fails with respect to category I failure, the device will be replaced completely.
7. If the device fails with respect to category II or category II failure, the device is minimally restored back to operation.
8. The device is replaced completely at planned time $NT(N = 1, 2, 3 \dots)$ for a fixed T or at category I failure, whichever arrives first.

Based on the assumptions, the probability of the device being replaced before category I failure occurs at the scheduled time T is

$$R_1(NT) = e^{-\int_0^{NT} r_1(t)dt}, \quad (1)$$

where $N = 1, 2, 3 \dots$ and T is fixed.

Based on the assumptions, the cost of unscheduled replacement of the device in one replacement cycle is

$$C_r(NT)\big(1 - R_1(NT)\big), \quad (2)$$

where $N = 1, 2, 3 \dots$ and T is fixed.

Based on the assumptions, the cost of scheduled replacement of the device at time NT in one replacement cycle is

$$C_p R_1(NT), \quad (3)$$

where $N = 1, 2, 3 \dots$ and T is fixed.

Based on the assumptions, the cost of minimal repair of the device due to category II failure in one replacement cycle is

$$\int_0^{NT} C_2 r_2(t) R_1(t) dt, \quad (4)$$

where $N = 1, 2, 3 \dots$ and T is fixed.

Based on the assumptions, the cost of minimal repair of the device due to category III failure in one replacement cycle is

$$\int_0^{NT} C_3 r_3(t) R_1(t) dt, \quad (5)$$

where $N = 1, 2, 3$ ... and T is fixed.

Based on the assumptions, the mean of one replacement cycle is

$$\int_0^{NT} R_1(t) dt, \quad (6)$$

where $N = 1, 2, 3$ ... and T is fixed.

Adding up equation (1) to equation (6), the device's cost rate in one replacement cycle is

$$C(N) = \frac{C_r(1-R_1(NT))+C_p R_1(NT)+\int_0^{NT} R_1(t)K(t)dt}{\int_0^{NT} R_1(t)dt}, \quad (7)$$

where

$$K(t) = C_2 r_2(t) + C_3 r_3(t). \quad (8)$$

Noting the following:

1. If the value of T is taking as one (that is, $T = 1$), then $C(N)$ will be a continuous standard age replacement model with minimal repair.
2. $C(N)$ is adopted as an objective function of an optimization problem, and the main goal is to obtain an optimal discrete scheduled replacement time N^* that minimizes $C(N)$.

VI. Numerical Example

In this section, we will give two numerical example, so as to illustrate the characteristics of the constructed discrete scheduled replacement model.

Let the rate of arrival of category I, category II and category III failures follows the Weibull distribution:

$$r_i(t) = \lambda_i \propto_i t^{\propto_i - 1}, \quad \text{for } i = 1, 2, 3, \quad (9)$$

where $\propto_i > 1$ and $t \geq 0$.

Let the collection of parameters and repair/replacement costs be used in this specific example:

1. $\propto_1 = 2, \propto_2 = 3, \propto_3 = 3.$
2. $\lambda_1 = 0.0002, \lambda_1 = 0.04, \lambda_3 = 0.02.$
3. $C_r = 50, C_p = 40, C_2 = 3, C_3 = 1.5.$

By substituting the parameters in equation (9), the category I, category II and category III failure rates were obtained as follows:

$$r_1(t) = 0.0004t, \quad (10)$$

$$r_2(t) = 0.12t^2, \quad (11)$$

$$r_3(t) = 0.06t^3. \quad (12)$$

Table 1 below is obtained, by substituting the assumed cost of replacement/repair ($C_r = 50$, $C_p = 40$, $C_2 = 3$, $C_3 = 1.5$) and rates of category I, category II and category III failures obtained above (equations (10), (11) and (12)) in equation (7), so as to evaluate the device's optimal discrete scheduled replacement time. When obtaining table 1, the value of $T = 1, T = 2, T = 3, T = 4, T = 5$ and $T = 6$ are considered so as to investigate the properties of the device's optimal discrete scheduled replacement time. Figure 1 is the graph of $C(N)$ against N, as $T = 1$. Figure 2 is the graph of $C(N)$ against N, as $T = 2$. Figure 3 is the graph of $C(N)$ against N, as $T = 3$. Figure 4 is the graph of $C(N)$ against N, as $T = 4$. Figure 5 is the graph of $C(N)$ against N, as $T = 5$. Figure 6 is the graph of $C(N)$ against N, as $T = 6$.

Table 1: *Values of C(N) for T = 1, T = 2, T = 3, T = 4, T = 5 and T =6, versus N* $(1, 2, 3 \ldots)$

N	C(N) as T=1	C(N) as T=2	C(N) as T=3	C(N) as T=4	C(N) as T=5	C(N) as T=6
1	400.82	203.18	140.39	112.46	99.38	94.47
2	203.18	112.46	94.47	99.13	116.34	142.63
3	140.39	94.47	106.45	142.63	195.82	263.13
4	112.46	99.13	142.63	216.79	314.97	433.39
5	99.38	116.34	195.82	314.97	465.78	641.83
6	94.47	142.63	263.13	433.39	641.83	876.92
7	94.86	176.41	342.80	568.80	836.42	1125.87
8	99.13	216.79	433.39	717.84	1042.19	1374.67
9	106.45	263.13	533.53	876.92	1251.20	1608.65
10	116.34	314.97	641.83	1042.19	1455.03	1813.40
11	116.34	371.85	756.81	1209.54	1645.18	1975.90
12	128.47	433.39	876.92	1374.67	1813.40	2085.74
13	142.63	499.18	1000.52	1533.20	1952.23	2136.30
14	158.65	568.80	1125.87	1680.83	2055.48	2125.54
15	176.41	641.83	1251.20	1813.40	2118.67	2056.29

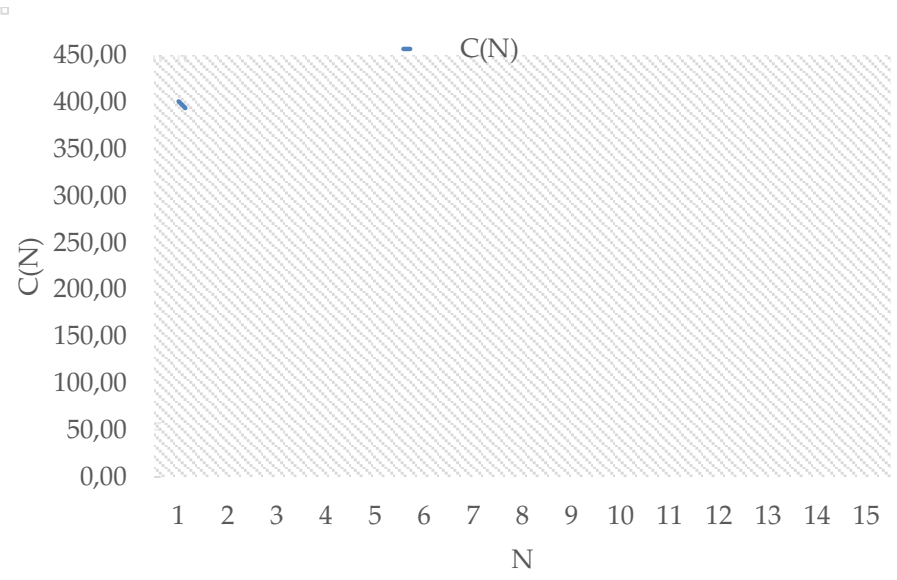


Figure 1: *C(N) against N, as T=*

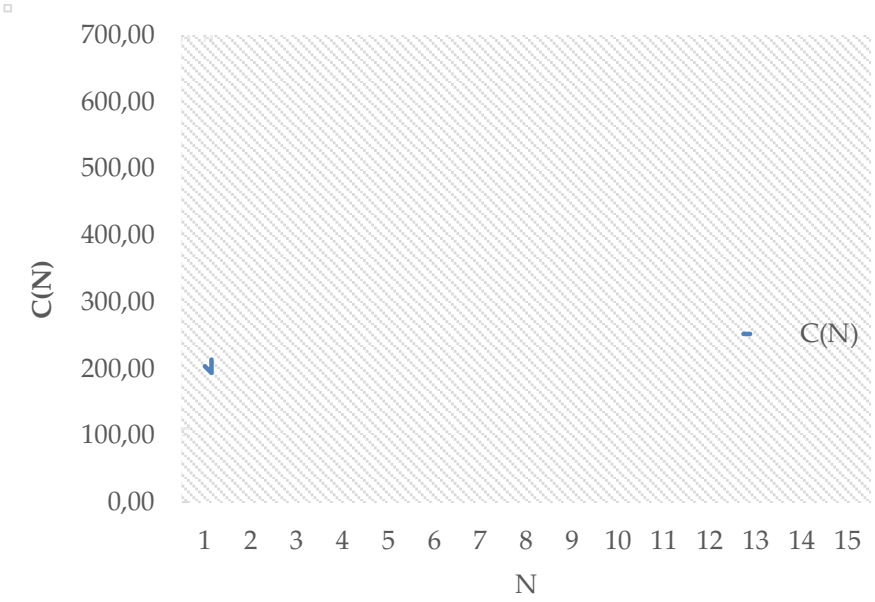


Figure 2: *C(N) against N, as T=2*

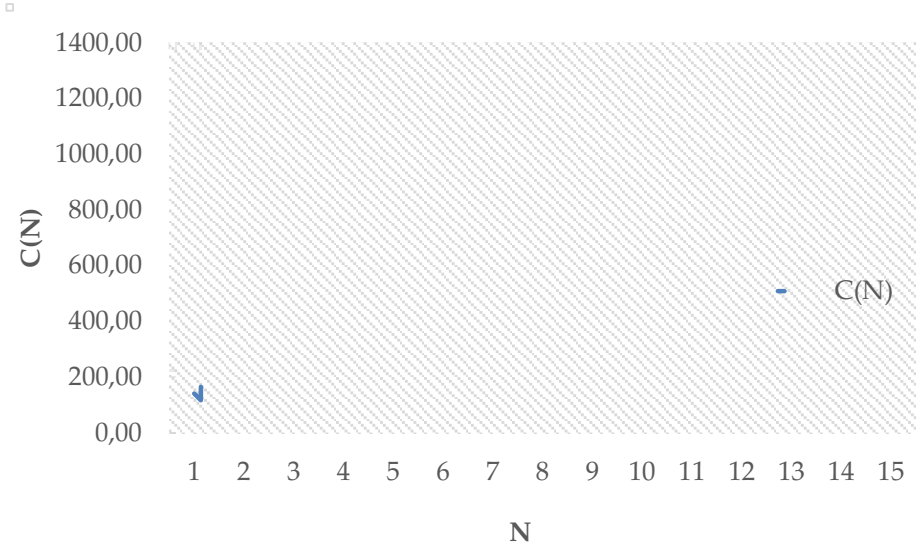


Figure 3: *C(N) against N, as T=3*

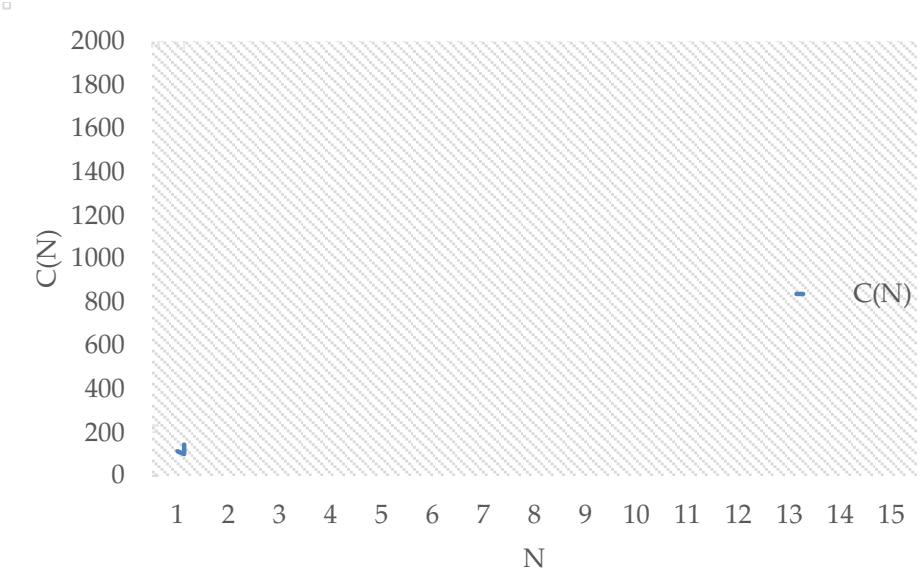


Figure 4: *C(N) against N, as T=4*

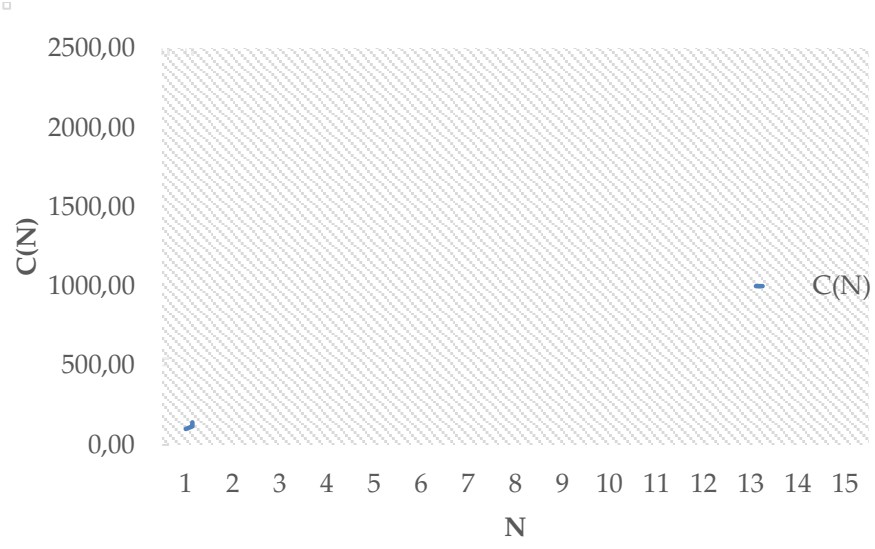


Figure 5: *C(N) against N, as T=5*

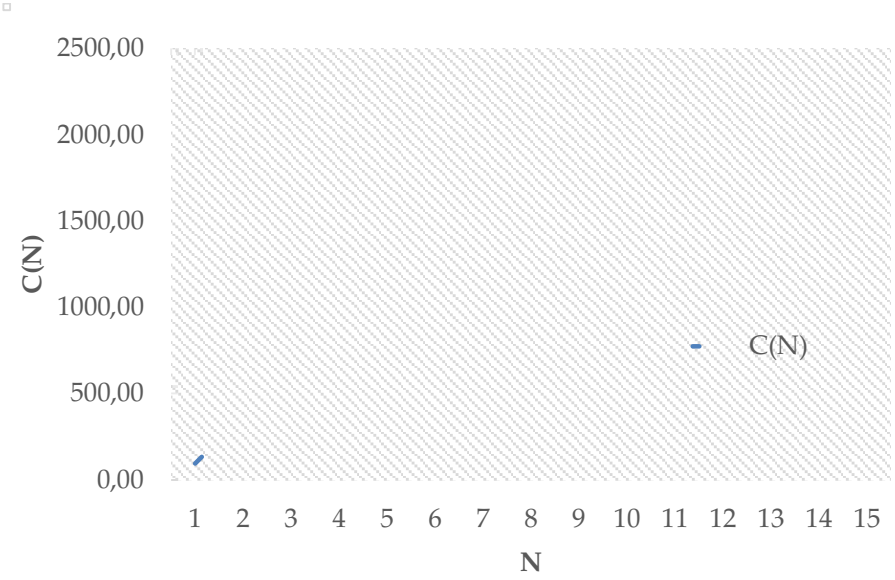


Figure 6: *C(N) against N, as T=6*

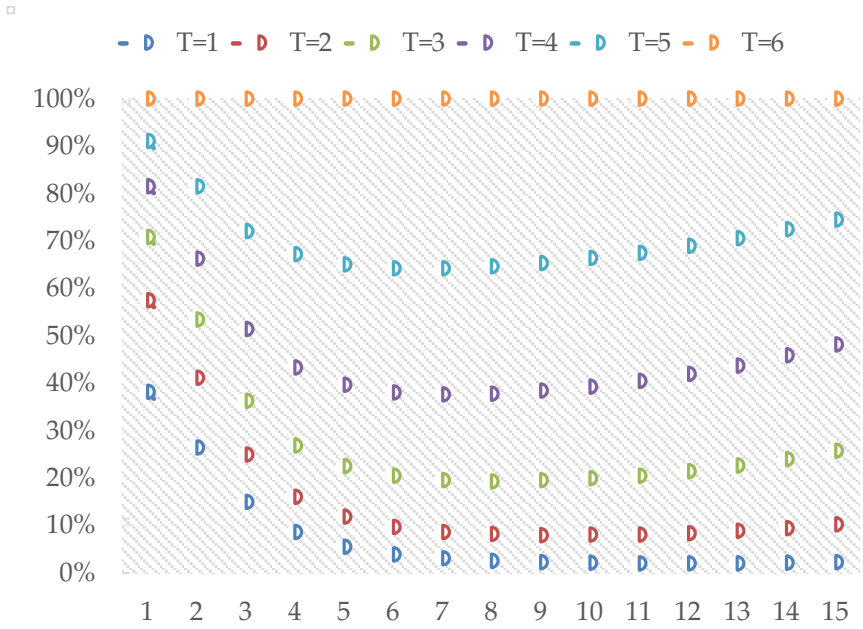


Figure 6: *Comparing C(N) for T=1, 2, 3, 4, 5, 6*

Some observations from the results obtained are as follows:

1. Observe from table 1, the optimum discrete scheduled replacement time is 6, when $T = 1$. That is, $N^* = 6$, with $C(N^* = 6) = 94.47$, when $T = 1$. See figure 1 below for the plot of $C(N)$ against N.
2. Observe from table 1, the optimum discrete scheduled replacement time is 3, when $T = 2$. That is, $N^* = 3$, with $C(N^* = 3) = 94.47$, when $T = 2$. See figure 2 below for the plot of $C(N)$ against N.
3. Observe from table 1, the optimum discrete scheduled replacement time is 2, when $T = 3$. That is, $N^* = 2$, with $C(N^* = 2) = 94.47$, when $T = 3$. See figure 3 below for the plot of $C(N)$ against N.
4. Observe from table 1, the optimum discrete scheduled replacement time is 2, when $T = 4$. That is, $N^* = 2$, with $C(N^* = 2) = 99.13$, when $T = 4$. See figure 4 below for the plot of $C(N)$ against N.
5. Observe from table 1, the optimum discrete scheduled replacement time is 1, when $T = 5$. That is, $N^* = 1$, with $C(N^* = 1) = 99.38$, when $T = 5$. See figure 5 below for the plot of $C(N)$ against N.
6. Observe from table 1, the optimum discrete scheduled replacement time is 1, when $T = 6$. That is, $N^* = 1$, with $C(N^* = 1) = 99.47$, when $T = 6$. See figure 6 below for the plot of $C(N)$ against N.
7. Observe from figure 7, we have : $(C(N),\ T = 1) < (C(N),\ T = 2) < (C(N),\ T = 3) < (C(N),\ T = 4) < (C(N),\ T = 5) < (C(N),\ T = 6)$.
8. Observe from figure 1, figure 2, figure 3 and figure 4 are all in convex shaped, which corresponded to $T = 1, T = 2,\ T = 3$ and $T = 4$, respectively.
9. Observe from figure 5 and figure 6 are in s-shaped, which are corresponded to $T = 5$ and $T = 6$, respectively.
10. Observe from table 1, as the value of T is increasing, the optimum discrete scheduled replacement time decreases.

VII. Conclusion and recommendations

This paper developed a discrete scheduled model for a device that is exposed to three categories of failures. category I failure is an un-repairable one, which occurs suddenly, and if it occurs, the device is replaced completely, while category II and category III failures are repairable failures, which occurs due to time and usage, and the two failures are minimally repaired. A numerical example was provided to test the constructed model so as to investigate the characteristics of the discrete scheduled model constructed and determine the optimum replacement number (N^*) of the device. A numerical example was provided for simple illustrations. From the results obtained, it is discovered or verified that, the value of T have an effect on the discrete scheduled replacement model, because of the following reasons:

1. as the value of T decreases, the optimal discrete replacement time (N^*) increases, while as the value of T increases, the optimal discrete replacement time (N^*) decreases.
2. as the value of T increases, $C(N)$ increases, while as the value of T decreases, $C(N)$ increases decreases.

With such reasons above, it can be easily seen that, continuous scheduled replacement model (continuous age replacement model) is better than discrete scheduled replacement model (discrete age replacement model). This paper is important to engineers, maintenance managers and plant management in maintaining multi-component systems at idle times, such as weekend, month-end or year-end.

References

[1] Aven, T. and Castro, I.T. (2008). A minimal repair replacement model with three types of failures. *European Journal of Operational Research*, 188:506–515.

[2] Briš, R., Byczanski, P., Goňo, R. and Rusek, S. (2017). Discrete maintenance optimization of complex multi-component systems. *Reliability Engineering and System Safety,* 168:80–89.

[3] Chang, C. (2014). Optimum preventive maintenance policies for systems subject to random working times, replacement and minimal repair. *Computers and Industrial Engineering,* 67:185–194.

[4] Coria, V. H., Maximov, S., Rivas-Davalos, F., Melchor, C. L. and Guardado, J. L. (2015). Analytical method for optimization of maintenance policy based on available system failure data. *Reliability Engineering and System Safety,* 135:55–63.

[5] Enogwe, S. U., Oruh, B. I. and Ekpenyong, E. J. (2018). A modified replacement model for items that fail suddenly with variable replacement costs. *American Journal of Operations Research,* 8:457–473.

[6] Fallahnezhad, M. S., Najafian, E. (2017). A model of preventive maintenance for parallel, series and single item replacement systems based on statistical analysis. *Communications in Statistics-Simulation and Computation,* 46:5846–5859.

[7] Jain, M. and Gupta, R. (2013). Optimal replacement policy for a repairable system with multiple vacations and imperfect fault coverage. *Computers and Industrial Engineering,* 66: 710–719.

[8] Lim, J. H., Qu, J. and Zuo, J. M. (2016). Age replacement policy based on imperfect repair with random probability. *Reliability Engineering and System Safety*, 149: 24–33.

[9] Liu, Y., Ma, Y., Qu, Z. and Li, X. (2018). Reliability mathematical models of repairable Systems with uncertain lifetimes and repair time. *IEEE,* 6: 71285–71295.

[10] Malki, Z., Ait, D. A. and Ouali, M. S. (2015). Age replacement policies for two-component systems with stochastic dependence. *Journal of Quality in Maintenance Engineering,* 20(3):346-357.

[11] Murthy, D. N. P. and Hwang, M. C. (2007). Optimal discrete and continuo 'n ıance policy for a complex unreliable machine *International Journal of Systems Science,* 6(1):35–52.

[12] Nakagawa T. Maintenance Theory of Reliability: *Springer-Verlag, London Limited*, 2005.

[13] Nakagawa, T., Chen, M. and Zhao, X. (2018). Note on history of age replacement policies. *International Journal of Mathematical, Engineering and Management Sciences,* 3(2): 151–161.

[14] Safaei, F., Ahmadi, J. and Balakrishnan, N. (2018). A repair and replacement policy for systems based on probability and mean of profits. *Reliability Engineering and System Safety,* 183: 143–152.

[15] Sudheesh, K. K, Asha, G. and Krishna, K. M. J. (2019). On the mean time to failure of an age-replacement model in discrete time. *Communications in Statistics - Theory and Methods,* 50 (11): 2569–2585.

[16] Tsoukalas, M. Z. and Agrafiotis, G. K. (2013). A new replacement warranty policy indexed by the product's correlated failure and usage time. *Computers and Industrial Engineering,* 66:203–211.

[17] Waziri, T. A. and Yusuf, I. (2020). On age replacement policy of system involving minimal repair. *Reliability Theory and Application*, 4(59): 54–62.

[18] Xie, L., Lundteigen, M. A. and Liu, Y. (2020). Reliability and barrier assessment of series-parallel systems subject to cascading failures. *Journal of Risk and Reliability,* 00 (0): 1–15.

[19] Yuan, L. and Xu, J. (2011). An optimal replacement policy for a repairable system based on its repairman having vacations. *Reliability Engineering and System Safety,* 96:868–875.

[20] Yusuf, I. and Ali, U. A. (2012). Structural dependence replacement model for parallel system of two units. *Nigerian Journal of Basic and Applied Science,* 20(4): 324–326.

[21] Yusuf, I., Hussaini, N. and Yakasai, B. M. (2014). Some reliability measures of a deteriorating system. *International Journal of Applied Mathematical Research,* 3(1): 23–29.

[22] Zhao, X., Mizutani, S. and Nakagawa, T. (2014). Which is batter for replacement policies with continuous or discrete scheduled times ? *European Journal of Operational Research,* 000: 1–10.

Inverse Weibull-Rayleigh Distribution Characterisation with Applications Related Cancer Data

Aijaz Ahmad [1], S. Qurat ul Ain[2], Rajnee Tripathi[3] and Afaq Ahmad[4]

[1,2,3] Department of Mathematics, Bhagwant University, Ajmer, Rajasthan, India

[4]Department of Mathematical Sciences, IUST, Awantipora, Kashmir

E-mail: [1]ahmadaijaz4488@gmail.com E-mail: [2]andrabiqurat19@gmail.com

E-mail: [3]rajneetripathi@hotmail.com E-mail: [4]baderaafaq@gmail.com

Abstract

The current study establishes a new three parameter Rayleigh distribution that is based on the inverse Weibull-G family and is an extension of the Rayleigh distribution. The formulation is known as the inverse Weibull-Rayleigh distribution (IWRD). The distinct structural properties of the formulated distribution including moments, moment generating function, order statistics, quantile function, and Renyi entropy have been discussed. In addition expressions for survival function, hazard rate function and reverse hazard rate function are obtained explicitly. The behaviour of probability density function (p.d.f) and cumulative distribution function (c.d.f) are illustrated through different graphs. The estimation of the formulated distribution parameters are performed by maximum likelihood estimation method. A simulation analysis has been carried out to evaluate and compare the effectiveness of estimators in terms of their bias, variance and mean square error (MSE). Eventually, the usefulness of the formulated distribution is illustrated by means of real data sets which are related distinct areas of science.

Keywords: *Inverse Weibull-G family, Rayleigh distribution, moments, Renyi entropy, simulation, maximum likelihood estimation.*

Mathematics classification: *60E05, 62FXX, 62F10, 62G05*

I. Introduction

There is a plethora of univariate distributions in the statistics literature. However, statisticians have found it difficult to find an effective distribution for analysing or modelling complicated real-life data sets. To resolve such challenges, new probability distributions must be formed or fundamental type must be modified. Over recent times, researchers have investigated a plethora of new methods and approaches, and by employing these approaches, generalization or extensions can be accomplished from baseline distributions. The main objective for these modifications is to enhance the accuracy or flexibility of distributions while assessing more complicated real-life data sets.

Waloddi Weibull, a Swedish mathematician, introduced the Weibull distribution in 1951. Because it may be used to analyse real life data with monotone failure rates, this distribution is considered versatile for data sets with bathtub shapes or unimodal. The Weibull distribution, on the other hand,

may not necessarily give a best fit for data sets with a bathtub shape or failure rates that are unimodal.

Let X be a random variable follows the Weibull distribution with parameter β and θ. Then its probability density function (pdf) is defined as

$$\psi(x,\beta,\theta) = \beta\theta^{\beta} x^{\beta-1} e^{-\theta^{\beta} x^{\beta}} \quad ; x > 0, \beta, \theta > 0$$

The inverse of the Weibull distribution is obtained by applying the transformation $T = \frac{1}{X}$.

Thus the probability density function (pdf) of inverse Weibull distribution takes following form.

$$h(t,\beta,\theta) = \beta\theta^{\beta} t^{-\beta-1} e^{-\theta^{\beta} t^{-\beta}} \quad ; t > 0, \beta, \theta > 0 \tag{1}$$

The inverse Weibull distribution is a subclass of the generalised extreme value distribution, which was previously researched by B.V. Gnedenko (1941) and Frechet (1927). In this paper, we develop the inverse Weibull-Rayleigh distribution, which is an extension of the Rayleigh distribution. Rayleigh distributions have a broad array of applications in research to simulate real life data, including reliability analysis, engineering, communication theory, medical science, and applied statistics. Rayleigh distribution has been expanded by researchers to make it more comprehensive and efficient for assessing more diverse factual data, for instance, due to its immense variety of applications. Weibull-Rayleigh distribution by Faton Merovci [11], odd generalized exponential Rayleigh distribution by Albert Luguterah [2], Topp-Leone Rayleigh distribution by Fatoki olayode [12], new generalisation of Rayleigh distribution by A.A Bhat et al [8].The probability density function (pdf) of Rayleigh distribution with scale parameter α is defined by

$$g(y,\alpha) = \alpha y e^{-\frac{\alpha}{2} y^2} \quad ; \; y > 0, \alpha > 0 \tag{2}$$

The associated cumulative distribution function (cdf) is given by

$$G(y,\alpha) = 1 - e^{-\frac{\alpha}{2} y^2} \quad ; y > 0, \alpha > 0 \tag{3}$$

In recent past years researcher have focussed to explore new generators from continuous standard distributions. As a result, the obtained distribution enhances the effectiveness and flexibility of data modelling. Some generated families of distribution are as follows: beta-G family of distribution explored by Eugene et al [10], kumaraswamy-G family by Cordeiro et al [9], transformed-transformer(T-X) by Alzaatrh et al [1], Weibull-G by Bourguignon et al [5], Lindley-G by Frank Gomes-silva et al [12], Topp-Leone odd log-logistic family of distributions by Brito et al [7], inverse Weibull-G by Amal S. Hassan et al [3], among others.

II. T-X Transformation

T-X family of distributions defined by Alzaatreh et al [1] is given by

$$F(y) = \int_0^{W[G(y)]} r(t)\,dt \tag{4}$$

Where $r(t)$ be the probability density function of a random variable T and $W[G(y)]$ be a function of cumulative density function of random variable Y.

Suppose $G(y,\eta)$ denotes the baseline cumulative distribution function, which depends on parameter vector η. Now using T-X approach, the cumulative distribution function $F(y)$ of inverse Weibull generator (IWG) can be derived by replacing $r(t)$ in equation (4) with (1) and $W[G(y)]=\frac{G(y,\eta)}{\overline{G}(y,\eta)}$, where $\overline{G}(y,\eta)=1-G(y,\eta)$ which follows

$$F(y,\beta,\theta,\eta)=\int_0^{\frac{G(y,\eta)}{\overline{G}(y,\eta)}}\beta\theta^{\beta}t^{-\beta-1}e^{-\theta^{\beta}t^{-\beta}}dt$$

$$=e^{-\theta^{\beta}\left(\frac{G(y,\eta)}{\overline{G}(y,\eta)}\right)^{-\beta}};\ y>0,\beta,\theta>0 \tag{5}$$

The corresponding pdf of (5) becomes

$$f(y,\beta,\theta,\eta)=\beta\theta^{\beta}g(y,\eta)\frac{[G(y,\eta)]^{-\beta-1}}{[\overline{G}(y,\eta)]^{-\beta+1}}e^{-\theta^{\beta}\left(\frac{G(y,\eta)}{\overline{G}(y,n)}\right)^{-\beta}};\ y>0,\beta,\theta>0 \tag{6}$$

The survival $S(y)$ and hazard rate function $h(y)$ are respectively given by

$$S(y)=1-F(y,\beta,\theta,\eta)=1-e^{-\theta^{\beta}\left(\frac{G(y,\eta)}{\overline{G}(y,n)}\right)^{-\beta}}$$

$$h(y)=\frac{\beta\theta^{\beta}g(y,\eta)\frac{[G(y,\eta)]^{-\beta-1}}{[\overline{G}(y,\eta)]^{-\beta+1}}e^{-\theta^{\beta}\left(\frac{G(y,\eta)}{\overline{G}(y,n)}\right)^{-\beta}}}{1-e^{-\theta^{\beta}\left(\frac{G(y,\eta)}{\overline{G}(y,n)}\right)^{-\beta}}}$$

III. Useful Expansion

Applying Taylor series expansion to the exponential function of the pdf in equation (6) we have

$$e^{-\theta^{\beta}\left(\frac{G(y,\eta)}{\overline{G}(y,n)}\right)^{-\beta}}=\sum_{i=0}^{\infty}\frac{(-1)^i\theta^{i\beta}}{i!}\left[\frac{G(y,\eta)}{\overline{G}(y,\eta)}\right]^{-\beta i} \tag{7}$$

Substitute equation (7) in (6), we have

$$f(y,\beta,\theta,\eta)=\beta g(y,\eta)\sum_{i=0}^{\infty}\frac{(-1)^i\theta^{-\beta(i+1)}}{i!}\frac{(G(y,\eta))^{-\beta(i+1)-1}}{(\overline{G}(y,\eta))^{-\beta(i+1)+1}} \tag{8}$$

Since $\beta>0$ and $|z|<1$, using generalised binomial theorem, we have

$$(1-z)^{\beta-1}=\sum_{j=0}^{\infty}(-1)^j\binom{\beta-1}{j}z^j$$

$$\left(\overline{G}(y,\eta)\right)^{\beta(i+1)-1}=\left(1-G(y,\eta)\right)^{\beta(i+1)-1}=\sum_{j=0}^{\infty}(-1)^j\binom{\beta(i+1)-1}{j}\left(G(y,\eta)\right)^j \tag{9}$$

Using equation (9) in equation (8), we have

$$f(y,\beta,\theta,\eta)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}}{i!}\beta\theta^{\beta(i+1)}g(y,\eta)\binom{\beta(i+1)-1}{j}\left(G(y,\eta)\right)^{j-\beta(i+1)-1}$$

$$=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\delta_{i,j}g(y,\eta)\left(G(y,\eta)\right)^{j-\beta(i+1)-1} \tag{10}$$

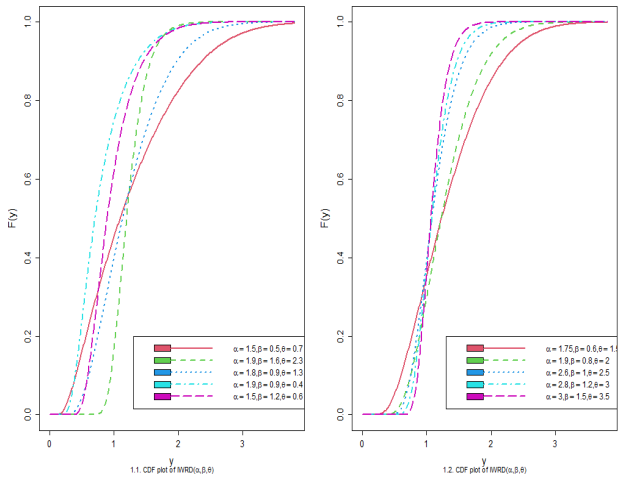
Where $\delta_{i,j}=\frac{(-1)^{i+j}\beta\theta^{\beta(i+1)}}{i!}\binom{\beta(i+1)-1}{j}$

The paper is framed as. In section 2, we derive the cumulative distribution function (cdf), probability density function (pdf). In section 3, we study the reliability measures, survival function, hazard rate function and reverse hazard rate function. In section 4, different statistical properties are studied including, moments, moment generating function, quantile function and random number generation. In section 5, Renyi entropy is discussed. In section 6, order statistics is expressed, in section 7, the estimation of parameters are performed by maximum likelihood estimation. In section 8, simulation study is performed. Finally in section 9 the efficiency of the established distribution is examined through data sets.

IV. The Inverse Weibull-Rayleigh Distribution

In this section we explore the inverse Weibull-Rayleigh distribution and studied its different statistical properties. Using equation (3) in equation (5), we obtain the cumulative distribution function (cdf) of the proposed distribution which follows

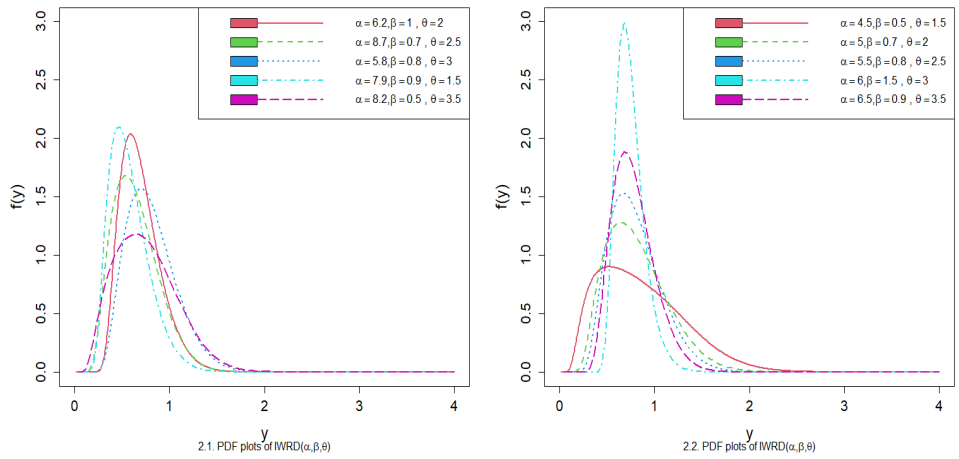
$$F(y,\alpha,\beta,\theta)=e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}},\ y>0,\alpha,\beta,\theta>0 \tag{11}$$



Figures (1.1) and (1.2) illustrates some of possible shapes of the cdf of IWRD for different values α,β and θ

The associated probability density function of inverse Weibull-Rayleigh distribution is given by

$$f(y,\alpha,\beta,\theta)=\alpha\beta\theta^{\beta}ye^{\frac{\alpha}{2}y^2}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta-1}e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}},\ y>0,\alpha,\beta,\theta>0 \tag{12}$$



Figures (2.1) and (2.2) illustrates some of possible shapes of the pdf of IWRD for different values α,β and θ

V. Reliability Measures

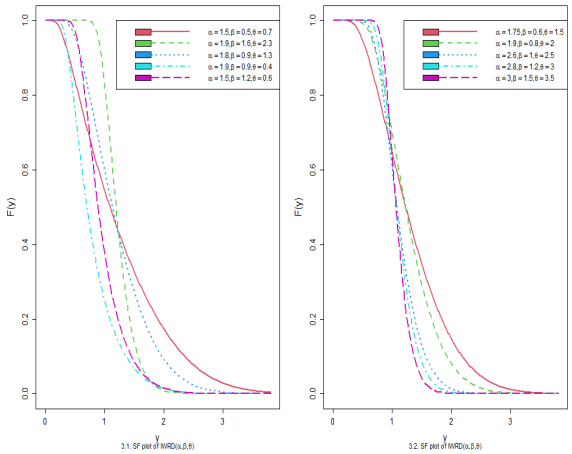
Suppose Y be a continuous random variable with cdf $F(y),\ y\geq 0$. Then its reliability function which is also called survival function is defined as

$$S(y)=p_r(Y>y)=\int_y^{\infty}f(y)dy=1-F(y)$$

The survival function of inverse Weibull-Rayleigh distribution is given as

$$S(y,\alpha,\beta,\theta)=1-F(y,\alpha,\beta,\theta)$$

$$=1-e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}} \tag{13}$$



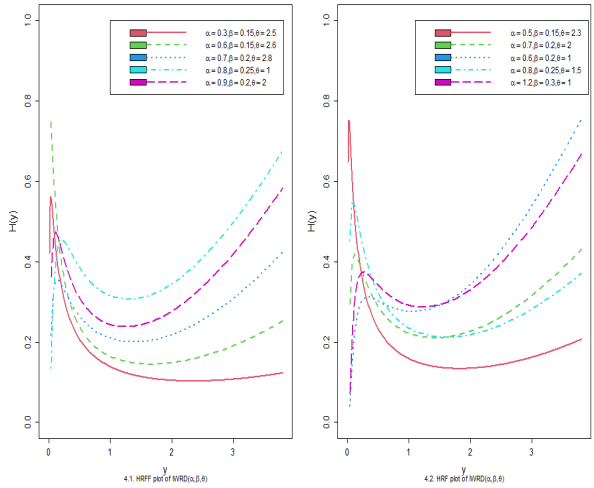
Figures (3.1) and (3.2) illustrates some of possible shapes of the survival function of IWRD for different values α, β and θ

The hazard rate function of inverse Weibull-Rayleigh distribution is given as

$$H(y,\alpha,\beta,\theta)=\frac{f(y,\alpha,\beta,\theta)}{S(y,\alpha,\beta,\theta)} \tag{14}$$

Substituting equations (12) and (13) in equation (14), we have

$$H(y,\alpha,\beta,\theta)=\frac{\alpha\beta\theta^{\beta}ye^{\frac{\alpha}{2}y^2}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta-1}e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}}{1-e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}}$$

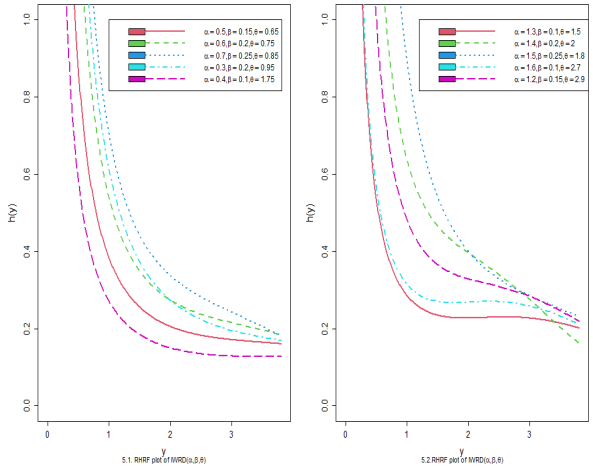


Figures (4.1) and (4.2) illustrates some of possible shapes of the hazard rate function of IWRD for different values α, β and θ

Reverse hazard rate function of inverse Weibull-Rayleigh distribution is given as

$$h(y,\alpha,\beta,\theta)=\frac{f(y,\alpha,\beta,\theta)}{F(y,\alpha,\beta,\theta)}$$

$$=\alpha\beta\theta^{\beta}ye^{\frac{\alpha}{2}y^2}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta-1}$$



Figures (5.1) and (5.2) illustrates some of possible shapes of the reverse hazard rate function of IWRD for different values α,β and θ

VI. Structural properties of inverse Weibull-Rayleigh distribution

Theorem 4.1:- Suppose y denotes a random variable follows IWRD with p.d.f $f(y,\alpha,\beta,\theta)$. Then the r^{th} moment of inverse Weibull-Rayleigh distribution is given by

$$\mu_r{}' = \sum_{i,j=0}^{\infty}\sum_{k=0}^{\infty}\delta_{i,j}(-1)^k\binom{j-\beta(i+1)-1}{k}\frac{2^{\frac{r}{2}}\alpha\Gamma\left(\frac{r}{2}+1\right)}{[\alpha(k+1)]^{\frac{r}{2}+1}}$$

Proof:- Let Y denotes a random variable follows inverse Weibull-Rayleigh distribution. Then r^{th} moment denoted by $\mu_r{}'$ is given as

$$\mu_r{}' = E(Y^r)=\int_0^{\infty} y^r f(y,\alpha,\beta,\theta)dy$$

Substituting equations (1) and (2) in equation (10), we get

$$\mu_r^{'} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\delta_{i,j}\alpha\int_0^{\infty} y^{r+1}e^{-\frac{\alpha}{2}y^2}\left(1-e^{-\frac{\alpha}{2}y^2}\right)^{j-\beta(i+1)-1} dy$$

We know the formulae of generalized binomial expansion, which follows

$$(1-a)^{p-1} = \sum_{j=0}^{\infty}(-1)^j\binom{p-1}{j}a^j$$

Now applying the above formulae, we get

$$\mu_r^{'} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\delta_{i,j}\alpha\int_0^{\infty} y^{r+1}e^{-\frac{\alpha}{2}y^2}\sum_{k=0}^{\infty}(-1)^k\binom{j-\beta(i+1)-1}{k}e^{-\frac{k\alpha}{2}y^2} dy$$

$$\mu_r^{'} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^k\delta_{i,j}\alpha\binom{j-\beta(i+1)-1}{k}\int_0^{\infty} y^{r+1}e^{-(k+1)\frac{\alpha}{2}y^2} dy$$

Making substitution $\frac{\alpha}{2}(k+1)y^2 = z$ so that $ydy = \frac{1}{\alpha(k+1)}dz$

$$\mu_r^{'} = \sum_{i,j=0k=0}^{\infty}\sum^{\infty}(-1)^k\delta_{i,j}\alpha\binom{j-\beta(i+1)-1}{k}\frac{2^{\frac{r}{2}}}{\left[\alpha(k+1)\right]^{\frac{r}{2}+1}}\int_0^{\infty} z^{\left(\frac{r}{2}+1\right)-1}e^{-z}dz$$

$$\mu_r^{'} = \sum_{i,j=0k=0}^{\infty}\sum^{\infty}(-1)^k\delta_{i,j}\binom{j-\beta(i+1)-1}{k}\frac{2^{\frac{r}{2}}\alpha\Gamma\left(\frac{r}{2}+1\right)}{\left[\alpha(k+1)\right]^{\frac{r}{2}+1}}$$

Theorem 4.2:- Suppose y denotes a random variable follows IWRD with pdf $f(y,\alpha,\beta,\theta)$. Then the moment generating function of inverse Weibull-Rayleigh distribution is given by

$$M_Y(t) = \sum_{r=0}^{\infty}\sum_{i,j=0k=0}^{\infty}\sum^{\infty}\frac{(it)^r}{r!}\delta_{i,j}(-1)^k\binom{j-\beta(i+1)-1}{k}\frac{2^{\frac{r}{2}}\alpha\Gamma\left(\frac{r}{2}+1\right)}{\left[\alpha(k+1)\right]^{\frac{r}{2}+1}}$$

Proof:- Let Y be a random variable follows inverse Weibull-Rayleigh distribution. Then the moment generating function of the distribution denoted by $M_Y(t)$ is given

$$M_Y(t) = E\left(e^{ty}\right) = \int_0^{\infty} e^{ty}f(y,\alpha,\beta,\theta)dy$$

Using Taylor's series

$$= \int_0^{\infty} \left(1 + ty + \frac{(ty)^2}{2!} + \frac{(ty)^3}{3!} + \ldots\right) f(y,\alpha,\beta,\theta)dy$$

$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} \int_0^{\infty} y^r f(y,\alpha,\beta,\theta)dy$$

$$= \sum_{r=0}^{\infty} \frac{t^r}{r!} E(Y^r)$$

$$M_Y(t) = \sum_{r=0}^{\infty} \sum_{i,j=0}^{\infty} \sum_{k=0}^{\infty} \frac{t^r}{r!} \delta_{i,j} (-1)^k \binom{j-\beta(i+1)-1}{k} \frac{2^{\frac{r}{2}} \alpha \Gamma\left(\frac{r}{2}+1\right)}{[\alpha(k+1)]^{\frac{r}{2}+1}}$$

The characteristics function of the IWRD denoted as $\phi_Y(t)$ can be obtained by replacing $t = it, i = \sqrt{-1}$ is given by

$$\phi_Y(t) = \sum_{r=0}^{\infty} \sum_{i,j=0}^{\infty} \sum_{k=0}^{\infty} \frac{(it)^r}{r!} \delta_{i,j} (-1)^k \binom{j-\beta(i+1)-1}{k} \frac{2^{\frac{r}{2}} \alpha \Gamma\left(\frac{r}{2}+1\right)}{[\alpha(k+1)]^{\frac{r}{2}+1}}$$

VII. Quantile function of inverse Weibull-Rayleigh distribution

The quantile function of random variable Y, where $Y \sim IWRD(\alpha,\beta,\theta)$, can be obtained by inverting equation (11), we have

$$Q(u) = F^{-1}(u) = \left[\frac{2}{\alpha} \log\left\{\left(\frac{-1}{\theta^{\beta}} \log u\right)^{\frac{-1}{\beta}} + 1\right\}\right]^{\frac{1}{2}}$$

In particular, the median of the distribution can be obtained by setting $u = 0.5$

$$M = \left[\frac{2}{\alpha} \log\left\{\left(\frac{-1}{\theta^{\beta}} \log(0.5)\right)^{\frac{-1}{\beta}} + 1\right\}\right]^{\frac{1}{2}}$$

VIII. Random number generation of inverse Weibull-Rayleigh distribution

Suppose y denotes a random variable with cdf given in equation (11). The random number of inverse Weibull-Rayleigh distribution can be generated as

$$F(y) = u \Rightarrow y = F^{-1}(u)$$

So that

$$y=\left[\frac{2}{\alpha}\log\left\{\left(\frac{-1}{\theta^{\beta}}\log u\right)^{\frac{-1}{\beta}}+1\right\}\right]^{\frac{1}{2}}$$

Where u is the uniform random variable defined in an open interval $(0,1)$.

IX. Renyi entropy of inverse Weibull-Rayleigh distribution

If Y is a continuous random variable having probability density function $f(y,\alpha,\beta,\theta)$. Then Renyi entropy is defined as

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left\{\int_0^{\infty}f^{\gamma}(y)dy\right\},\text{ where }\gamma>0\text{ and }\gamma\neq 1$$

Using equation (6), we have

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left\{\int_0^{\infty}\left[\beta\theta^{\beta}g(y,\eta)\frac{(G(y,\eta))^{-(\beta+1)}}{(\overline{G}(y,\eta))^{(-\beta+1)}}e^{-\theta^{\beta}\left[\frac{G(y,\eta)}{\overline{G}(y,\eta)}\right]^{-\beta}}\right]^{\gamma}dy\right\}$$

$$=\frac{1}{1-\gamma}\log\left\{\int_0^{\infty}\beta^{\gamma}\theta^{\beta\gamma}(g(y,\eta))^{\gamma}\frac{(G(y,\eta))^{-(\beta+1)\gamma}}{(\overline{G}(y,\eta))^{(-\beta+1)\gamma}}e^{-\gamma\theta^{\beta}\left[\frac{G(y,\eta)}{\overline{G}(y,\eta)}\right]^{-\beta}}dy\right\}\qquad(15)$$

Now using the power series expansion for exponential function, we have

$$e^{-\gamma\theta^{\beta}\left[\frac{G(y,\eta)}{\overline{G}(y,\eta)}\right]^{-\beta}}=\sum_{i=0}^{\infty}\frac{(-1)^i\gamma^i\theta^{\beta i}}{i!}\left[\frac{G(y,\eta)}{\overline{G}(y,\eta)}\right]^{-\beta i}\qquad(16)$$

Substituting equation (16) into (17), we obtain

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left\{\sum_{i=0}^{\infty}\frac{(-1)^i\beta^{\gamma}\gamma^i\theta^{\beta(\gamma+i)}}{i!}\int_0^{\infty}(g(y,\eta))^{\gamma}(G(y,\eta))^{-(\beta+1)\gamma-\beta i}(1-G(y,\eta))^{(\beta-1)\gamma+\beta i}dy\right\}$$

$$=\frac{1}{1-\gamma}\log\left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}\beta^{\gamma}\gamma^i\theta^{\beta(\gamma+i)}}{i!}\binom{\beta(\gamma+i)-\gamma}{j}\int_0^{\infty}(g(y,\eta))^{\gamma}(G(y,\eta))^{j-\beta(\gamma+i)-\gamma}dy\right\}$$

Thus, the Renyi entropy for inverse Weibull-Rayleigh distribution, is given by

$$=\frac{1}{1-\gamma}\log\left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^{i+j}\alpha^{\gamma}\beta^{\gamma}\gamma^i\theta^{\beta(\gamma+i)}}{i!}\binom{\beta(\gamma+i)-\gamma}{j}\int_0^{\infty}y^{\gamma}e^{-\frac{\alpha\gamma}{2}y^2}\left(1-e^{-\frac{\alpha}{2}y^2}\right)^{j-\beta(\gamma+i)-\gamma}dy\right\}$$

$$=\frac{1}{1-\gamma}\log\left\{\sum_{i,j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^k\alpha^{\gamma}\delta_{i,j}\binom{j-\beta(\gamma+i)-\gamma}{k}\int_0^{\infty}y^{\gamma}e^{-\frac{\alpha}{2}(\gamma+k)y^2}dy\right\}$$

Where

$$\delta_{i,j}=\frac{(-1)^{i+j}\beta^{\gamma}\gamma^{i}\theta^{\beta(\gamma+i)}}{i!}\binom{\beta(\gamma+i)-\gamma}{j}$$

Making the substitution $\frac{\alpha}{2}(\gamma+k)y^2=z$, we have

$$=\frac{1}{1-\gamma}\log\left\{\sum_{i,j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{k}\alpha^{\gamma}\delta_{i,j}\binom{j-\beta(\gamma+i)-\gamma}{k}\frac{1}{2}\left(\frac{2}{\alpha(\gamma+k)}\right)^{\frac{\gamma+1}{2}}\int_{0}^{\infty}z^{\frac{\gamma+1}{2}-1}e^{-z}dz\right\}$$

After solving above integral, we obtain

$$T_R(\gamma)=\frac{1}{1-\gamma}\log\left\{\sum_{i,j=0}^{\infty}\sum_{k=0}^{\infty}(-1)^{k}\alpha^{\gamma}\delta_{i,j}\binom{j-\beta(\gamma+i)-\gamma}{k}\frac{1}{2}\left(\frac{2}{\alpha(\gamma+k)}\right)^{\frac{\gamma+1}{2}}\Gamma\left(\frac{\gamma+1}{2}\right)\right\}$$

X. Order statistics of inverse Weibull-Rayleigh distribution

Let us suppose $Y_1,Y_2,...,Y_n$ be random samples of size n from IWRD distribution with pdf $f(y)$ and cdf $F(y)$. Then the probability density function of k^{th} order statistics is given as

$$f_{Y_{(k)}}(y)=\frac{n!}{(k-1)!(n-k)!}f(y)[F(y)]^{k-1}[1-F(y)]^{n-k} \tag{17}$$

Now using the equation (11) and (12) in (17). The probability of k^{th} order statistics of inverse Weibull-Rayleigh distribution is given as

$$f_{Y_{(k)}}(y)=\frac{n!}{(k-1)!(n-k)!}\alpha\beta\theta^{\beta}ye^{\frac{\alpha}{2}y^2}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta-1}e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\left[e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\right]^{k-1}$$

$$\left[1-e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\right]^{n-1}$$

Then, the pdf of first order statistics Y_1 inverse Weibull-Rayleigh distribution is given as

$$f_{Y_{(1)}}(y)=n\alpha\beta\theta^{\beta}ye^{\frac{\alpha}{2}y^2}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta-1}e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\left[1-e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\right]^{n-1}$$

Then, the pdf of nth order statistics Y_n inverse Weibull-Rayleigh distribution is given as

$$f_{Y_{(n)}}(y)=n\alpha\beta\theta^{\beta}ye^{\frac{\alpha}{2}y^2}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta-1}e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\left[e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y^2}-1\right)^{-\beta}}\right]^{n-1}$$

XI. Maximum likelihood estimation and Fisher's information matrix of inverse Weibull- Rayleigh distribution

Suppose $Y_1, Y_2, ..., Y_n$ denotes random sample of size n from inverse Weibull-Rayleigh distribution then its likelihood function is given by

$$l=\prod_{i=1}^{n}f(y,\alpha,\beta,\theta)$$

$$=\prod_{i=1}^{n}\alpha\beta\theta^{\beta}y_ie^{\frac{\alpha}{2}y_i^2}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1}e^{-\theta^{\beta}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}}$$

$$=\left(\alpha\beta\theta^{\beta}\right)^n\prod_{i=0}^{n}y_i\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1}e^{\frac{\alpha}{2}\sum_{i=1}^{n}y_i^2}e^{-\theta^{\beta}\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}}$$

The log likelihood function is given by

$$\log l=n\log\alpha+n\log\beta+n\beta\log\theta+\sum_{i=1}^{n}\log y_i+\frac{\alpha}{2}\sum_{i=1}^{n}y_i^2-(\beta+1)\sum_{i=1}^{n}\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)$$
$$-\theta^{\beta}\sum_{i=1}^{n}\left(e^{-\frac{\alpha}{2}y_i^2}-1\right)^{-\beta} \quad (18)$$

Differentiating equation (18) with respect each parameter α, β and θ, we have

$$\frac{\partial\log l}{\partial\alpha}=\frac{n}{\alpha}+\frac{1}{2}\sum_{i=1}^{n}y_i^2-\frac{(\beta+1)}{2}\sum_{i=1}^{n}\frac{y_i^2e^{\frac{\alpha}{2}y_i^2}}{e^{\frac{\alpha}{2}y_i^2}-1}+\frac{\beta\theta^{\beta}}{2}\sum_{i=1}^{n}y_i^2e^{\frac{\alpha}{2}y_i^2}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1} \quad (19)$$

$$\frac{\partial\log l}{\partial\beta}=\frac{n}{\beta}+n\log\theta-\sum_{i=1}^{n}\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)+\theta^{\beta}\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)$$

$$-\theta^{\beta}\log(\theta)\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta} \quad (20)$$

$$\frac{\partial\log l}{\partial\theta}=\frac{n\beta}{\theta}-\beta\theta^{\beta-1}\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta} \quad (21)$$

By setting equations (19), (20) and (21) to zero the MLE of parameters can be obtained. However the above equations are non-linear which cannot be expressed in closed form. So numerical techniques such as Newton-Raphson, Regula-Falsi and bisection methods must be applied to obtain MLE of parameters denoted by $\hat{\varsigma}(\hat{\alpha},\hat{\beta},\hat{\theta})$ of $\varsigma(\alpha,\beta,\theta)$.

Since the MLE of $\hat{\varsigma}$ follows asymptotically normal distribution which is given as

$$\sqrt{n}(\hat{\varsigma}-\varsigma)\to N(0,I^{-1}(\varsigma))$$

Where $I^{-1}(\varsigma)$ is the limiting variance-covariance matrix $\hat{\varsigma}$ and $I(\varsigma)$ is a 3×3 Fisher information matrix

i,e
$$I(\varsigma)=-\frac{1}{n}\begin{bmatrix} E\left(\frac{\partial^2\log l}{\partial\alpha^2}\right) & E\left(\frac{\partial^2\log l}{\partial\alpha\partial\beta}\right) & E\left(\frac{\partial^2\log l}{\partial\alpha\partial\theta}\right) \\ E\left(\frac{\partial^2\log l}{\partial\beta\partial\alpha}\right) & E\left(\frac{\partial^2\log l}{\partial\beta^2}\right) & E\left(\frac{\partial^2\log l}{\partial\beta\partial\theta}\right) \\ E\left(\frac{\partial^2\log l}{\partial\theta\partial\alpha}\right) & E\left(\frac{\partial^2\log l}{\partial\theta\partial\beta}\right) & E\left(\frac{\partial^2\log l}{\partial\theta^2}\right) \end{bmatrix}$$

Where

$$\frac{\partial^2\log l}{\partial\alpha^2}=\frac{-n}{\alpha^2}+\frac{(\beta+1)}{4}\sum_{i=1}^{n}\frac{y_i^4e^{\frac{\alpha}{2}y_i^2}}{\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^2}$$

$$+\frac{\beta\theta^\beta}{4}\sum_{i=1}^{n}y_i^4e^{\frac{\alpha}{2}y_i^2}\left\{\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1}-(\beta+1)\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-2}\right\}$$

$$\frac{\partial^2\log l}{\partial\beta^2}=\frac{-n}{\beta^2}-\theta^\beta\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}\left\{\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)\right\}^2-\theta^\beta(\log\theta)^2\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}$$

$$+2\theta^\beta\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)$$

$$\frac{\partial^2\log l}{\partial\theta^2}=\frac{-n\beta}{\theta^2}-\beta(\beta-1)\theta^{\beta-2}\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}$$

$$\frac{\partial^2\log l}{\partial\alpha\partial\beta}=\frac{\partial^2\log l}{\partial\beta\partial\alpha}=-\sum_{i=1}^{n}\frac{y_i^2e^{\frac{\alpha}{2}y_i^2}}{\left(e^{\frac{\alpha}{2}y_i^2}-1\right)}-\frac{\beta\theta^\beta}{2}\sum_{i=1}^{n}y_i^2e^{\frac{\alpha}{2}y_i^2}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1}\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)$$

$$+\theta^{\beta}\left(1+\beta\log\theta\right)\sum_{i=1}^{n}y_i^2 e^{\frac{\alpha}{2}y_i^2}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1}$$

$$\frac{\partial^2\log l}{\partial\alpha\partial\theta}=\frac{\partial^2\log l}{\partial\theta\partial\alpha}=\frac{\beta^2\theta^{\beta-1}}{2}\sum_{i=1}^{n}y_i^2 e^{\frac{\alpha}{2}y_i^2}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta-1}$$

$$\frac{\partial^2\log l}{\partial\beta\partial\theta}=\frac{\partial^2\log l}{\partial\theta\partial\beta}=\frac{n}{\theta}+\beta\theta^{\beta-1}\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}\log\left(e^{\frac{\alpha}{2}y_i^2}-1\right)$$

$$-\theta^{\beta-2}\left(\theta+\beta^2-\beta\right)\sum_{i=1}^{n}\left(e^{\frac{\alpha}{2}y_i^2}-1\right)^{-\beta}$$

Hence the approximate $100(1-\psi)\%$ confidence interval for α,β and θ are respectively given by

$$\hat{\alpha}\pm Z_{\frac{\psi}{2}}\sqrt{I^{-1}_{\alpha\alpha}(\hat{\varsigma})}\ \ ,\hat{\beta}\pm Z_{\frac{\psi}{2}}\sqrt{I^{-1}_{\beta\beta}(\hat{\varsigma})}\ \text{And}\ \hat{\theta}\pm Z_{\frac{\psi}{2}}\sqrt{I^{-1}_{\theta\theta}(\hat{\varsigma})}$$

Where $Z_{\frac{\psi}{2}}$ denotes the ψ^{th} percentile of the standard normal distribution.

XII. Results

I. Simulation Analysis

In this section we demonstrate the simulation analysis which examines the effectiveness of the M L estimators. The inverse cdf method is employed to generate random samples of size $n=30,50,75,100$ and 150 which is discussed in section (4). This procedure is repeated $N=500$ times for calculation of bias, variance and MSE. Four separate combinations of parameters are selected and it is observed that bias, variance and MSE decrease significantly, when we increase sample size. The efficiency of ML estimators is therefore relatively strong, consistent in case of IWRD.

Table 1: Average bias, variance and MSEs of 5,00 simulations of IWRD for different parameters values.

II. Applications

This section is dedicated to demonstrate the effectiveness of the established distribution by taking into account real data sets taken from medical science. The established distribution is compared with power Erlang distribution (PED), Weighted Gumbel-II distribution (WG-IID), power Gompertz distribution (PGD), inverse Weibull distribution (IWD), Rayleigh distribution (RD), inverse Rayleigh distribution (IRD) and inverse Lindley distribution (ILD). It is revealed that the developed distribution offers an appropriate fit.

To compare the versatility of the explored distribution, we consider the criteria like AIC (Akaike information criterion), CAIC (Consistent Akaike information criterion), BIC (Bayesian information criterion) and HQIC. Distribution having lesser AIC, CAIC, BIC, HQIC and KS values is considered better also having higher probability value (p-value).

Sample Size n	Parameters	$\alpha=0.4$, $\beta=0.7$, $\theta=0.8$			$\alpha=0.4$, $\beta=0.5$, $\theta=0.7$		
		Bias	variance	MSE	Bias	variance	MSE
30	α	0.10667	0.09549	0.10687	0.03396	0.04469	0.04584
	β	0.01601	0.03820	0.03846	0.03240	0.01454	0.01559
	θ	0.99840	6.86551	7.86231	0.46101	2.29695	2.50948
50	α	0.03325	0.04242	0.04353	0.00960	0.02461	0.02471
	β	0.02828	0.01952	0.02032	0.02701	0.00881	0.00954
	θ	0.29418	1.33982	1.42637	0.22597	0.96048	1.01155
75	α	0.03266	0.02648	0.02755	0.01118	0.01587	0.01599
	β	0.01992	0.01395	0.01434	0.01276	0.006403	0.00656
	θ	0.23048	0.54670	0.59982	0.14271	0.26261	0.28298
100	α	0.01522	0.01673	0.01696	0.00267	0.01220	0.01221
	β	-0.0008	0.00980	0.00980	0.01304	0.00462	0.00479
	θ	0.15424	0.34874	0.37253	0.06209	0.12997	0.13383
150	α	0.01240	0.01161	0.01177	0.00537	0.00995	0.00998
	β	0.01080	0.00754	0.00765	0.01761	0.00403	0.00434
	θ	0.09099	0.16775	0.17603	0.06589	0.13020	0.13255
		$\alpha=0.6$, $\beta=0.5$, $\theta=0.7$			$\alpha=0.6$, $\beta=0.8$, $\theta=0.9$		
30	α	0.01382	0.01359	0.01379	-0.0018	0.01282	0.01282
	β	0.00500	0.00398	0.00400	0.01203	0.00447	0.00462
	θ	0.10007	0.19036	0.20038	0.07110	0.19553	0.20058
50	α	0.01956	0.01135	0.01174	0.00314	0.01279	0.01280
	β	0.00397	0.00422	0.00423	0.00645	0.00402	0.00406
	θ	0.11939	0.15628	0.17053	0.07557	0.19296	0.19867
75	α	0.01085	0.01139	0.01151	-0.0052	0.00959	0.00962
	β	0.00533	0.00384	0.00386	0.01512	0.00440	0.00403
	θ	0.08410	0.13345	0.14052	0.03145	0.09593	0.09691
100	α	0.00531	0.01086	0.01089	0.00303	0.00873	0.00874
	β	0.00616	0.00401	0.00305	0.01137	0.00396	0.00402
	θ	0.06256	0.12807	0.13199	0.04911	0.10033	0.00275
150	α	0.01534	0.010668	0.01070	0.02240	0.01451	0.00502
	β	0.00422	0.00389	0.00301	0.00551	0.00436	0.00400
	θ	0.12016	0.17372	0.12816	0.14134	0.10026	0.00084

$$AIC = 2k - 2\ln l\,;\, CAIC = \frac{2kn}{n-k-1} - 2\ln l\,;\, BIC = k\ln n - 2\ln l$$

And $HQIC = 2k\ln(\ln(n)) - 2\ln l$

The descriptive statistics of the data set 1 and 2 are presented in Table 1and 4.The estimates of the parameters are shown in Table 2 and 5 for data set 1and 2 respectively. Log-likelihood, Akaike information criteria (AIC) etc for the data set 1 and 2 are generated and presented in Table 3 and 6 respectively.

Data set 1:- The data was collected from a group of 46 patients, per years, upon the recurrence of leukemia whom received autologous marrow. The data repoted by Jhon H Kersey [14], as follows

0.0301,0.0384,0.063, 0.0849, 0.0877, 0.0959, 0.1397, 0.1616, 0.1699, 0.2137,0.2137, 0.2164, 0.2384, 0.2712, 0.274, 0.3863, 0.4384, 0.4548, 0.5918, 0.6,0.6438, 0.6849, 0.7397, 0.8575, 0.9096, 0.9644, 1.0082, 1.2822, 1.3452, 1.4,1.526, 1.7205, 1.989, 2.2438, 2.5068, 2.6466, 3.0384, 3.1726, 3.4411, 4.4219,4.4356, 4.5863, 4.6904, 4.7808, 4.9863, 5

Table 2: Descriptive statistics of data set first

Min	Q_1	Median	Mean	Q_3	Skew	Kurt.	Max
0.0301	0.221	0.798	1.517	2.441	1.036	2.655	5

Table 3: The ML Estimates and standard error of the unknown parameters

Model		IWRD	PED	WG-IID	PGD	IWD	RD	IRD	ILD
$\hat{\alpha}$		0.6737	1.2685	0.7023	0.6483	……..	0.4075	0.0343	0.4333
$\hat{\beta}$		0.2764	0.6654	0.4651	0.2006	0.7017	……..	……..	…….
$\hat{\theta}$		0.0641	1.4619	0.0010	0.7515	0.3371	…….	…….	…….
S.E	$\hat{\alpha}$	0.2252	2.6342	0.3287	0.1832	0.1154	0.0600	0.0050	0.0462
	$\hat{\beta}$	0.0399	0.6548	0.6414	0.3050	……..	……..	……..	…….
	$\hat{\theta}$	0.0491	2.3917	0.5574	0.1553	……..	……..	……..	…….

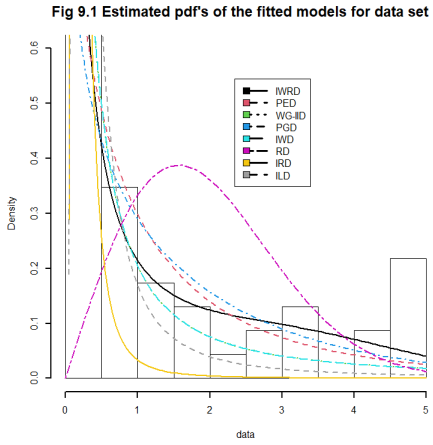
Table 4: Performance of distributions for data set first

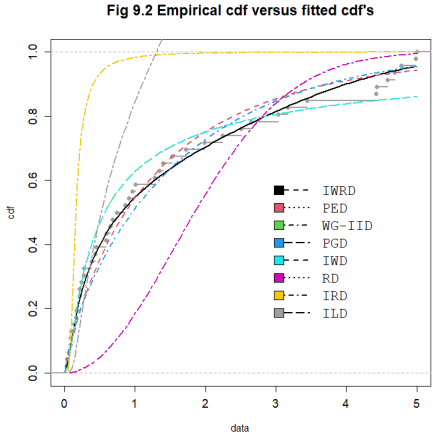
Model	IWRD	PED	WG-IID	PGD	IWD	RD	IRD	ILD
$-2\log l$	118.90	127.91	138.90	127.36	138.89	207.42	303.76	176.43
AIC	124.90	133.91	144.90	133.36	142.89	209.42	305.76	178.43
CAIC	125.47	134.48	145.47	133.93	143.17	209.51	305.85	178.52
BIC	130.39	139.40	150.38	138.84	146.55	211.25	307.59	180.25
HQIC	126.96	135.97	146.95	135.41	144.26	210.10	306.45	179.11
K-S Value	0.079	0.0980	0.1399	0.1013	0.139	0.3998	0.566	1.342
P Value	0.935	0.7686	0.3286	0.7325	0.328	8.1e-07	2.9e-13	2.2e-16

The asymptotic variance-covariance matrix of maximum likelihood estimates under IWRD for data set first is computed as

$$I^{-1}(\omega)=\begin{pmatrix} 0.0507 & -0.0053 & 0.0075 \\ -0.0053 & 0.0015 & -0.0011 \\ 0.0075 & -0.0011 & 0.0024 \end{pmatrix}$$

Therefore, the 95% confidence interval for α, β and θ are given as $(0.2322,1.1152)$, $(0.1982,0.3546)$ and $(-0.0321,0.1604)$, respectively.





Data set 2:- The data set represents the survival times(in years) of a group of patients given chemotherapy treatment reported by Bekker et al.[4]. The data follows

0.047, 0.115, 0.121, 0.132, 0.164, 0.197, 0.203, 0.260, 0.282, 0.296, 0.334, 0.395, 0.458, 0.466, 0.501, 0.507, 0.529, 0.534, 0.540, 0.641, 0.644, 0.696, 0.841, 0.863, 1.099, 1.219, 1.271, 1.326, 1.447, 1.485, 1.553, 1.581, 1.589, 2.178, 2.343, 2.416, 2.444, 2.825, 2.830, 3.578, 3.658, 3.743, 3.978, 4.003, 4.033

Table 5: Descriptive statistics of data set second

Min	Q_1	Median	Mean	Q_3	Skew	Kurt.	Max
0.047	0.39	0.84	1.34	2.17	0.972	2.663	4.03

Table 6: The ML Estimates of the unknown parameters for data set second

Model		IWRD	PED	WG-IID	PGD	IWD	RD	IRD	ILD
$\hat{\alpha}$		0.9335	1.9113	0.8677	0.6504	…….	0.6026	0.1072	0.6584
$\hat{\beta}$		0.3361	0.6756	0.4977	0.1215	0.8671	…….	…….	……..
$\hat{\theta}$		0.1524	2.1173	0.0010	0.9762	0.4482	……..	…….	……...
S.E	$\hat{\alpha}$	0.3257	3.6555	0.3392	0.1719	……..	0.0898	0.0159	0.0723
	$\hat{\beta}$	0.0505	0.6238	0.5701	0.2587	0.0927	…….	…….	……
	$\hat{\theta}$	0.1078	3.4214	0.5884	0.1949	0.0818	……..	……..	…….

Table 7: Performance of distributions for data set second

Model	IWRD	PED	WG-IID	PGD	IWD	RD	IRD	ILD
$-2\log l$	110.24	115.97	127.64	115.96	127.63	155.83	230.17	138.88
AIC	116.24	121.97	133.64	121.96	131.63	157.83	232.17	140.88
CAIC	116.82	122.56	134.23	122.54	131.92	157.92	232.26	140.97
BIC	121.66	127.39	139.06	127.38	135.25	159.63	233.98	142.69
HQIC	118.26	123.99	135.66	123.98	132.98	158.50	232.84	141.56
K-S Value	0.0660	0.9811	0.1383	0.1139	0.138	0.353	0.507	1.252
P Value	0.982	2.2e-16	0.3251	0.5638	0.325	1.5e-05	2.8e-11	2.2e-16

The asymptotic variance-covariance matrix of maximum likelihood estimates under IWRD for data set first is computed a

$$I^{-1}(\omega)=\begin{pmatrix} 0.1061 & -0.0105 & 0.0263 \\ -0.0105 & 0.0025 & -0.0034 \\ 0.0263 & -0.0034 & 0.0116 \end{pmatrix}$$

Therefore, the 95% confidence interval for α,β and θ are given as $(0.2950,1.5720)$, $(0.2370,0.4353)$ and $(-0.0590,0.3638)$, respectively.

Fig 9.3 Estimated pdf's of the fitted models for data set 2

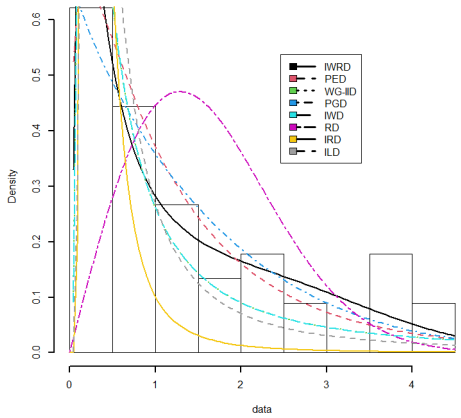
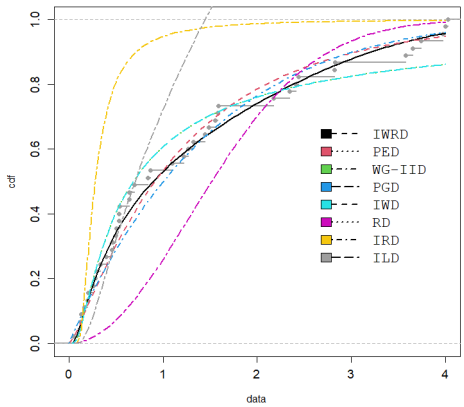


Fig 9.4 Empirical cdf versus fitted cdf's



it is evident from Table (4) and (7) that IWRD has lesser values of AIC, CAIC, BIC, HQIC and K-S statistics along with higher p-value. When it is compared with IWD, RD, IRD and ILD models. Hence we conclude that IWRD provides an adequate fit than compare ones

XIII. Discussion

This paper deals with a new generalisation of Rayleigh distribution called inverse Weibull-Rayleigh distribution. We have added extra two parameters to the Rayleigh distribution by inverse Weibull-G generator, the main purpose for such modification is that the formulated distribution become more richer and flexible in modelling datasets. Several distinct properties of formulated distribution has been studied and discussed. The model parameters of the distribution are estimated by the known method of maximum likelihood estimation. Eventually, the efficiency of the explored distribution is examined through real data sets which reveals that the formulated distribution provides an adequate model fit than competing ones.

Refrences

1. Alzaatreh A, Lee C, Famoye F (2013). A new method for generating families of distributions. Metron, 71, 63-79.
2. Albert L (2016). Odd generalizd exponential Rayleigh distribution. Advances and
3. applications in statistics, vol 48(1), 33-48.
4. Amal H and Said G. N (2018). The inverse Weibull-G familiy. Journal of data science, 723-742.
5. Bekker A, Roux J and Mostert P (2000). A generalisation of the compound Rayleigh distribution using a Bayesian methods on cancer survival times. Communication in statistics theory and methods, 29, 1419-1433.
6. Bourguingnon M, Silva R.B and Corderio G.M (2014). The Weibull-G family of probability distributions. Journal of data science, 12, 53-68.
7. Bhaumik D. K , Kapur K. G and Gibbons R.D (2009). Testing of parameters of a gamma distribution for small samples, Technometrics, 51(3), 326-334.
8. Brito E, Cordeiro G. M, Yousuf H. M, Alizadeh M, Silva G.O (2017). The Topp-Leone odd log-logistic family of distributions. J stat comput simul,87(15), 3040-3058.
9. Bhat A.A and Ahmad S.P (2020). A new generalization of Rayleigh distribution properties and applications. Pakistan journal of statistics, 36(3), 225-250.
10. Corderio R and Pulcini G (2011). A new family of generalized distributions. Journal of statistical computation and simulation, 81, 883-893.
11. Eugene N, Lee C and Famoye F (2002). Beta-normal distribution and its applications. Communication in statistics- theory and methods, 31, 497-512.
12. Faton M and Ibrahim E (2015). Weibull-Rayleigh distribution theory and applications. Applied mathematics and information sciences an international journal, vol 9(5), 1-11.
13. Frank G. S, Ana P, Edleide D.B (2016). The odd Lindley-G family of distributions. Austrian journal of statistics, 10, 1-20.
14. Fatoki O (2019). The Topp-Leone Rayleigh distribution with application. American journal of mathematics and statistics, 9(6), 215-220.
15. Frechet, M. (1927). Sur la loi de probabilite de l'ecart maximum. Annals de la societe poloaise de mathematiques, t. VI, 93.
16. Gnedenko, B. V. (1941). On limit theorems for maximul term of a ramdom series. Proceedings of Akademy of science USSR, v. 32(1), 7-9.
17. Jhon H. K, Daniel W, Mark E.N, Tucker W. L, Waliam G.W, Philips B. McGlave, Tae K, Daniel A. Vallera Anne I. Goldman, Bruce Bostrom, David Hurd, and Norma K.C. Ramsay (1987). Comparison of autologous and allogeneic bone marrow transplantation for treatment of high risk refractory acute lymphoblastic leukemia. New england journal of medicine, 317, 461-467.
18. Morad A, Gauss M. C (2017). The Gompertz-G family of distributions. Journal of statistical theory and practice, 11(1), 179-207.

STOCHASTIC ANALYSIS OF COMPLEX REDUNDANT SYSTEM HAVING PROBLEM OF WAITING LINE IN REPAIR USING COPULA METHODOLOGY

Surabhi Sengar[1], Mangey Ram[2], Yigit Kasancoglu[3]

•

[1]Pimpri Chinchwad College of Engineering, Pune, India.
[2]Department of Mathematics; Computer Science & Engineering,

Graphic Era Deemed to be University, Dehradun, Uttarakhand, India

[3]Department of International Logistics Management, Yasar University, Izmir, Turkey.

[1]sursengar@gmail.com, [2]drmrswami@yahoo.com, [3]yigit.kazancoglu@yasar.edu.tr

Abstract

This paper investigates the stochastic behavior of a redundant system having problem of waiting line in the maintenance section in terms of various aspects such as reliability, availability, sensitivity etc. The system under consideration has three parts X, Y and Z connected in series. Each part has two units. Out of which part X has one main unit and other applied redundant unit. Similarly part Y has one main unit and other cold redundant unit to support the system. When main units of both the part have failed, then redundant units start automatically. While part Z has two units connected in parallel configuration. Here a realistic situation is discussed that when main units and redundant units of part X and Y are failed and arrived for repair then due to unavailability of repair men a line is generated there and its affect on systems reliability. So the focus of the study is to investigate the nature of the system using supplementary variable technique with the application of copula methodology under the condition when all four units are in line for repair.

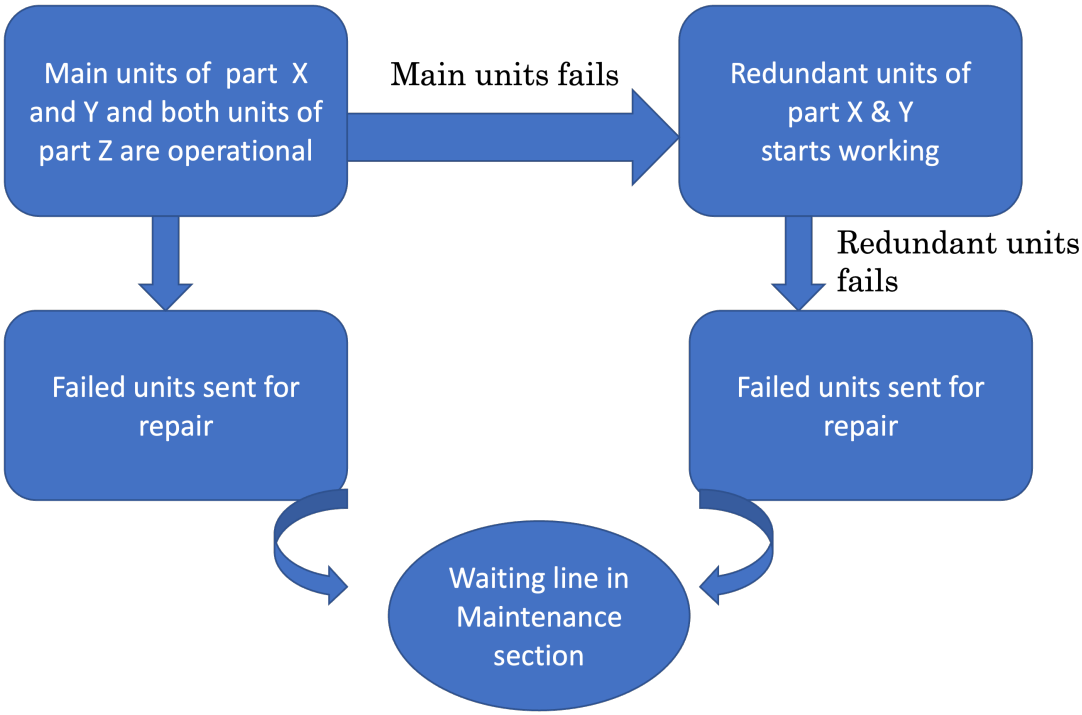
Keywords: Stochastic analysis, Supplementary variable technique, Copula methodology availability analysis etc.

I. Introduction

Redundancy is a very general technique used to improve the performance and reliability of the system. The word redundancy is generally addressed to forecast the replacement of a unit or Part by another unit or Part in case of failure. Redundancy functions in two ways: Applied redundancy and Passive redundancy. Both redundancies prevent system performance from un predicted failure and downfall without human interference. Applied redundancy monitors the performance of individual devices so it reduces the performance decline, when a condition occurs in the system with a number of failures. The system of electricity supply is a good example of applied redundancy as huge number of electrical lines is connected to generate facility with the consumer as well as each electrical line monitors that detect overload and circuit breakers. On the other hand passive redundancy provides extra competence or capacity to avoid or to decrease the impact of system failure. This extra competency permits the failure of some parts without system failure. For example in structural engineering, the additional cables and struts that are used in construction of overpass allow some parts to fail without the whole structure fall down. Also, queuing problem is an area of mathematics which deals with the models and situation that arises due to waiting line in maintenance or service. When we talk about analysis of repair services, queue analysis plays an important role. These models have been used by various repair systems like communication and manufacturing, networking and simulation for calculating the behavior of the system and various

reliability aspects. Previously, many researchers predict the system performance by assuming failure free service. But later different types of failures have been considered while evaluating a systems performance.

On the basis of above study and related facts, here in this paper we have discussed the stochastic behavior of a redundant system having three-Parts X, Y and Z, connected in series. Part X and Y have two main units and at same time two redundant units, one is applied redundant and another is Cold redundant. Part Z has two units which are connected in parallel configuration. Due to failure of any of the part, the system can result in complete failure [3]. It is assumed that when system starts operating, all the main units except redundant units are fully operational. Also, When the main units of X and Y fails, redundant units are switched on automatically and failed units are sent for repair to repairing section. The condition that has been taken into the consideration of the study is when all four units main units as well as supporting units are failed and sent for repair to maintenance section but because of unavailability of repairmen a line is generated in this section [4]. The system working process is described by the figure. Transition state diagram is shown by Figures 1. Table 1 shows the state specification of the system.



II. Assumptions

- In the beginning the system is in good operating state.
- All Parts are connected in series.
- System has two states only good and failed not degraded.
- Catastrophic failure is also responsible for system failure in the study also they require constant and exponential repair. So, copula technique is used for finding probability distribution [5].
- Repair facility which follows general time distribution is there for the service of both the Parts of unit 3 and also failure are exponential in both cases.
- For the Parts 1 and 2 failure and repairs both are exponential.

Table 1: *State specification of the system*

States	Description	System State
S0	The system is in good working state	G
S1	The system is in working state when key unit is failed.	G
S2	The system is in failed state because of failure of superfluous unit.	F
S3	When all four units are in waiting at repair section, system is in failed state.	F
S4	The system is in working state when superfluous unit of Part X is failed.	G
S5	The system is in failed state due to the failure of key unit of Part X.	F
S6	The system is in working condition when key unit Part Y is failed.	G
S7	The system is in failed state when superfluous unit of Part Y is failed.	F
S8	The system is in operable condition when key unit of Part Z failed.	G
S9	The system is in failed state from the state S8 due to failure of superfluous unit of Part Z.	FR
S10	The system is in operable condition when superfluous unit of Part Z is failed.	G
S11	The system is in failed state from the state S10 due to failure of key unit of Part Z.	FR
S12	The system is in failed state from the state S1 due to failure of Part Z.	FR
S13	The system is in failed state from the state S6 due to failure of Part Z.	FR
S14	System is failed state because of catastrophic failure.	FR

G: Good state; F: Failed State; FR= Failed state and under repair.

III. Notations

Pr Probability

$P_0(t)$ Pr (at time t system is in good state S_0)

$P_i(t)$ Pr {the system is in failed state due to the failure of the i^{th} Part at time t}, where i=2, 5, 7, 14.

λ_i Failure rates of Parts, where I = x1, x2, y1, y2, z1, z2, csf.

ψ Arrival rate of all four units of Parts X and Y to the repair section named as x1, x2, y1, y2.

μ Repair rate of unit's x1, x2, y1, y2.

$\phi_i(k)$ General repair rate of ith system in the time interval (k, k+⊚), where i= z1, z2,(names for the units of Part Z) csf and k=v, g, r, l.

$P_3(t)$ Pr (at time t there is a queue (x1, x2, y1, y2) in the maintenance section due to servicing of some other unit and all four machines are waiting for repair.

$P_i(j,k,t)$ Pr (at time t system is in failed state due to the failure of jth unit when kth unit has been already failed, where i=9, 11. j=g, v. and k=v, g.

K1, K2 Profit cost and service cost per unit time respectively.

Let $u_1 = e^l$ and $u_2 = \phi_{csf}(l)$ then the expression for joint probability according to Gumbel-Hougaard family of copula is given as $\phi_{csf}(l) = \exp[l^\theta + (\log \phi_{csf}(l))^\theta)^{1/\theta}]$

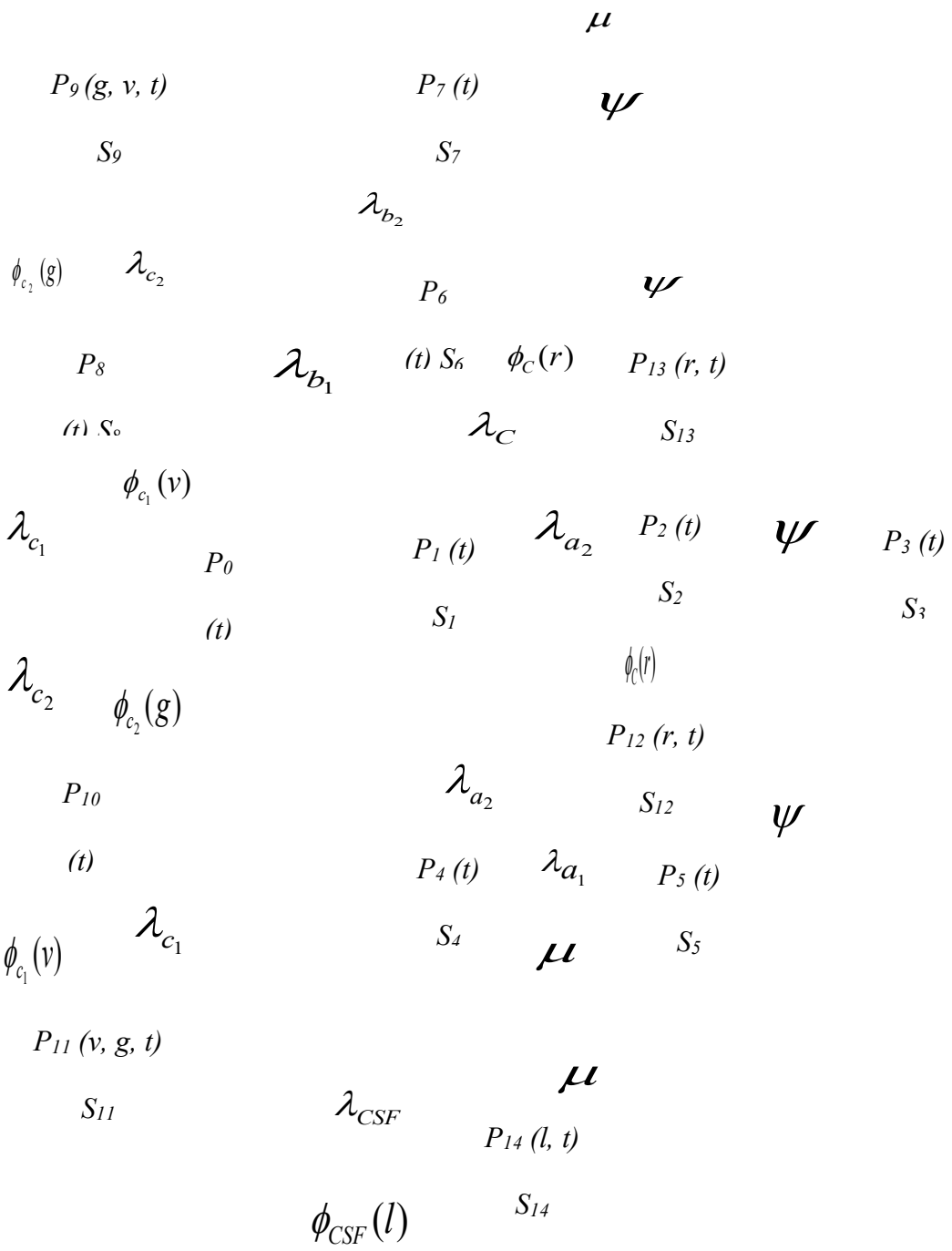


Figure 1: *Transition state diagram*

IV. Formulation of the mathematical model

The following differential equations have been obtained by considering limiting procedures and different probability constraints which satisfying the model:

$$\left[\frac{d}{dt}+\lambda_{x_1}+\lambda_{x_2}+\lambda_{y_1}+\lambda_{y_2}+\lambda_{z_1}+\lambda_{z_2}+\lambda_{csf}\right]P_0(t)=\int_0^\infty \mu\ \ (i)P_3(t)di+\varphi_{z_1}P_9(t)+\varphi_{z_2}P_{10}(t) \quad \ldots (1)$$

$$\left[\frac{\partial}{\partial t}+\lambda_{x_2}+\lambda_z\right]P_1(t)=\lambda_{a_1}P_0(t)+\int_0^\infty \varphi_z(r)P_{12}(r,t)dr \quad \ldots (2)$$

$$\left[\frac{\partial}{\partial t}+\psi\right]P_2(t)=\lambda_{x_2}P_1(t) \quad \ldots (3)$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial i}+(\mu+\psi)\right]P_3(t)=\psi[P_2(t)+P_5(t)+P_6(t)+P_7(t)]+\frac{(\psi t)^3 e^{-\psi t}}{6} \quad \ldots (4)$$

$$\left[\frac{\partial}{\partial t}+\lambda_{x_1}\right]P_4(t)=\lambda_{x_2}P_0(t) \quad \ldots (5)$$

$$\left[\frac{\partial}{\partial t}+\psi\right]P_5(t)=\lambda_{x_1}P_4(t) \quad \ldots (6)$$

$$\left[\frac{\partial}{\partial t}+\lambda_{y_2}+\psi\right]P_6(t)=\lambda_{y_1}P_0(t)+\int_0^{\infty}\varphi_Z(r)P_{13}(r,t)dr \quad \ldots (7)$$

$$\left[\frac{\partial}{\partial t}+\psi\right]P_7(t)=\lambda_{y_2}P_6(t) \quad \ldots (8)$$

$$\left[\frac{\partial}{\partial t}+\varphi_{z_1}(v)+\lambda_{z_2}\right]P_8(t)=\lambda_{z_1}P_0(t)+\int_0^{\infty}\varphi_{z_2}(g)P_{11}(g,v,t)dg \quad \ldots (9)$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial g}+\phi_{z_2}(g)\right]P_9(g,v,t)=0 \quad \ldots (10)$$

$$\left[\frac{\partial}{\partial t}+\phi_{z_2}(g)+\lambda_{z_1}\right]P_{10}(t)=\lambda_{z_2}P_0(t)+\int_0^{\infty}\phi_{z_1}(v)P_{11}(v,g,t)dv \quad \ldots (11)$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial v}+\phi_{z_1}(v)\right]P_{11}(v,g,t)=0 \quad \ldots (12)$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial r}+\phi_Z(r)\right]P_{12}(r,t)=0 \quad \ldots (13)$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial r}+\phi_z(r)\right]P_{13}(r,t)=0 \quad \ldots (14)$$

$$\left[\frac{\partial}{\partial t}+\frac{\partial}{\partial l}+\phi_{csf}(l)\right]P_{14}(l,t)=0 \quad \ldots (15)$$

Boundary Conditions:

$$P_3(i=0,t)=\psi[P_2(t)+P_3(t)+P_6(t)+P_7(t)] \quad \ldots (16)$$

$$P_8(0,t)=\lambda_{z_1}P_0(t) \quad \ldots (17)$$

$$P_9(0,v,t)=\lambda_{z_2}P_8(t) \quad \ldots (18)$$

$$P_{10}(0,t)=\lambda_{z_2}P_0(t) \quad \ldots (19)$$

$$P_{11}(0,g,t)=\lambda_{z_1}P_{10}(t) \quad \ldots (20)$$

$$P_{12}(0,t)=\lambda_z P_1(t) \quad \ldots (21)$$

$$P_{13}(0,t)=\lambda_z P_6(t) \quad \ldots (22)$$

$$P_{14}(0,t)=\lambda_{csf}P_0(t) \quad \ldots (23)$$

Initial condition:

$P_0(0)=1$, otherwise zero.

Solving equations (1) through (15) by taking Laplace transform and by using initial and boundary conditions we obtained following probabilities of system is in up and down states at time t,

$$\overline{Pup}=\overline{P}_0(s)+\overline{P}_1(S)+\overline{P}_4(s)+\overline{P}_6(s)+\overline{P}_8(s)+\overline{P}_{10}(s)$$

$$= \frac{1}{K(s)}[\ _{1+}\frac{\lambda_{x_1}}{[s+\lambda_{x_2}+\lambda_z-\lambda_z\overline{S}_{\phi_z}(s)]}{}_{+}\frac{\lambda_{x_2}}{[s+\lambda_{x_1}]}{}_{+}\frac{\lambda_{x_1}}{[s+\lambda_{y_2}+\psi-\lambda_z\overline{S}_{\phi_z}(s)]}{}_{+}$$

$$\frac{\lambda_{z_1}}{[s+\lambda_{z_2}+\phi_{z_1}(v)-\lambda_{z_2}\overline{S}_{\phi_{z2}}(s)]}{}_{+}\frac{\lambda_{c_2}}{[s+\lambda_{c_1}+\phi_{c_2}(g)-\lambda_{c_1}\overline{S}_{\phi_{c1}}(s)]}\ ] \quad \ldots (24)$$

$$\overline{P}_{down} = \overline{P}2(s)+\overline{P}5(s)+\overline{P}7(s)+\overline{P}9(s)+\overline{P}11(s)+\overline{P}12(s)+\overline{P}13(s)+\overline{P}14(s)$$

$$= \frac{\lambda_{x_1}\lambda_{x_2}}{[s+\psi][s+\lambda_{x_2}+\lambda_z-\lambda_z\overline{S}_{\phi_z}(s)]}\frac{1}{K(s)}+\frac{\lambda_{x_1}\lambda_{x_2}}{[s+\psi][s+\lambda_{x_1}]}\frac{1}{K(s)}+$$

$$\frac{\lambda_{y_1}\lambda_{y_2}}{[s+\psi][s+\lambda_{y_2}+\psi-\lambda_z\overline{S}_{\phi_z}(s)]}\frac{1}{K(s)}+\frac{\lambda_{c_1}\lambda_{c_2}D_{\phi_{c_2}}(s)}{[s+\lambda_{c_2}+\phi_{c_1}(v)-\lambda_{c_2}\overline{S}_{\phi_{c_2}}(s)]}\frac{1}{K(s)}+$$

$$\frac{\lambda_{z_1}\lambda_{z_2}D_{\phi_{z_1}}(s)}{[s+\lambda_{z_1}+\phi_{z_2}(g)-\lambda_{z_1}\overline{S}_{\phi_{z1}}(s)]}\frac{1}{K(s)}+\frac{\lambda_z\lambda_{x_1}D_{\phi_z}(s)}{[s+\lambda_{x_2}+\lambda_z-\lambda_z\overline{S}_{\phi_z}(s)]}\frac{1}{K(s)}+$$

$$\frac{\lambda_z\lambda_{y_1}D_{\phi_z}(s)}{[s+\lambda_{y_2}+\psi-\lambda_z\overline{S}_{\phi_z}(s)]}\frac{1}{K(s)}+\frac{\lambda_{csf}D_{\phi_{csf}}(s)}{K(s)} \quad \ldots (25)$$

where,

$$K(s) = s+\lambda_{x_1}+\lambda_{x_2}+\lambda_{y_1}+\lambda_{y_2}+\lambda_{z_1}+\lambda_{z_2}+\lambda_{csf}-\psi\{[\frac{\lambda_{x_1}\lambda_{x_2}}{[s+\psi][s+\lambda_{x_2}+\lambda_z-\lambda_z\overline{S}_{\phi_z}(s)]}$$

$$+\frac{\lambda_{x_1}\lambda_{x_2}}{[s+\psi][s+\lambda_{x_1}]}+\frac{\lambda_{y_1}}{[s+\lambda_{y_2}+\psi-\lambda_z\overline{S}_{\phi_z}(s)]}+\frac{\lambda_{y_1}\lambda_{y_2}}{[s+\psi][s+\lambda_{y_2}+\psi-\lambda_z\overline{S}_{\phi_z}(s)]}]D_\mu(s)$$

$$+\frac{\psi^3}{(s+\psi)^4}\}-\frac{\lambda_{z_1}\phi_{z_1}(v)}{[s+\lambda_{z_2}+\phi_{z_1}(v)-\lambda_{z_2}\overline{S}_{\phi_{z_2}}(s)]}-\frac{\lambda_{z_2}\phi_{z_2}(g)}{[s+\lambda_{z_1}+\phi_{z_2}(g)-\lambda_{z_1}\overline{S}_{\phi_{z1}}(s)]}$$

$$-\lambda_{csf}\overline{S}_{\phi_{csf}}(s)$$

$$M(s) = \psi\{[\frac{\lambda_{x_1}\lambda_{x_2}}{[s+\psi][s+\lambda_{x_2}+\lambda_z-\lambda_z\overline{S}_{\phi_z}(s)]}+\frac{\lambda_{x_1}\lambda_{x_2}}{[s+\psi][s+\lambda_{x_1}]}+\frac{\lambda_{y_1}}{[s+\lambda_{y_2}+\psi-\lambda_Z\overline{S}_{\phi_Z}(s)]}$$

$$+\frac{\lambda_{y_1}\lambda_{y_2}}{[s+\psi][s+\lambda_{y_2}+\psi-\lambda_z\overline{S}_{\phi_z}(s)]}]D_\mu(s)++\frac{\psi^3}{(s+\psi)^4}\} \quad \ldots (26)$$

$$D_\mu(s) = \frac{1-\overline{S}_\mu(s)}{s+\psi} \quad \ldots (27)$$

$$\phi_{csf}(l) = \exp[l^\theta+(\log\phi_{csf}(l))^\theta)^{1/\theta}] \quad \ldots (28)$$

Also,

$$\overline{P}_{up}(s)+\overline{P}_{down}(s) = \frac{1}{s} \quad \ldots (29)$$

Steady state behavior of the system By Abel's lemma we have,

$$\lim_{s\to 0}\{s\overline{F}(s)\} = \lim_{t\to\infty}F(t)$$

In equations (24) and (25) we get,

$$\overline{P}_{up}(s) = \frac{1}{K(0)}[1 + \frac{\lambda_{a_1}}{\psi\lambda_{a_2}} + \frac{\lambda_{a_2}}{\lambda_{a_1}} + \frac{\lambda_{b_1}}{\lambda_{b_2} + \psi - \lambda_c} + \frac{\lambda_{c_1}}{\phi_{c_1}(v)} + \frac{\lambda_{c_2}}{\phi_{c_2}(g)}] \quad \ldots (30)$$

$$\overline{P}_{down} = \frac{1}{K(0)}[\frac{\lambda_{a_1}}{\psi} + M(0) + \frac{\lambda_{a_2}}{\psi} + \frac{\lambda_{b_1}\lambda_{b_2}}{\psi(\lambda_{b_2} + \psi - \lambda_C)} + \frac{\lambda_{c_1}\lambda_{c_2}M_{\phi_{c_2}}}{\phi_{c_1}(v)} + \frac{\lambda_{c_1}\lambda_{c_2}M_{\phi_{c_1}}}{\phi_{c_2}(g)} + \frac{\lambda_c\lambda_{a_1}M_{\phi_C}}{\lambda_{a_2}} + \frac{\lambda_C\lambda_{b_1}M_{\phi_C}}{\lambda_{b_2} + \psi - \lambda_C} + \lambda_{CSF}M_{\phi_{CSF}}] \quad \ldots (31)$$

where,

$$M(0) = \lim_{s\to 0} M(s) \quad \ldots (32)$$

$$M_{\phi_i} = \lim_{s\to 0} \frac{1 - \overline{S}_{\phi_i}(s)}{s} \quad \ldots (33)$$

$$S_{\phi_i}(s) = \frac{\phi_1}{s + \phi_1} \quad \ldots (34)$$

IV. Discussion

In this paper, A common problem of waiting line, which generally occurs in the manufacturing industries is discussed through the reliability, availability, Mean time to failure and cost analysis of the considered system by using supplementary variable technique and copula methodology. Also, we have analyzed the steady state behaviour to improve the practical utility of the system. One can easily observe from figure 2 that reliability of the system decreases rapidly with the transitions in time when all failures follow exponential time distribution. The reason for this decrease is waiting line in the repair section because of which the system is in non operational state for a long time. Figure 3 gives an idea about the availability of the system that decreases approximately in a constant manner with the increment in time.

Here, we have also done the analysis of effect of various parameters on mean time to failure of the system. Figures 4 represents decreases in MTTF with the increases failure rate (λ_{x_1}) of main unit of part X. Similarly, figures 5 and 6 shows the MTTF decreases of the system with the increase of failure rates λ_{z_1}, λ_{z_2} and λ_{csf}. A common phenomenon can be observed from the graphs of all the parameters that initially because of waiting line in the repair section the MTTF is negative and gradually it becomes positive.

At last cost function G(t) analysis, for different values of K_1 and K_2 with respect to time is done in figure 7. This analysis reveals that expected profit decreases as the service cost of the system increases.

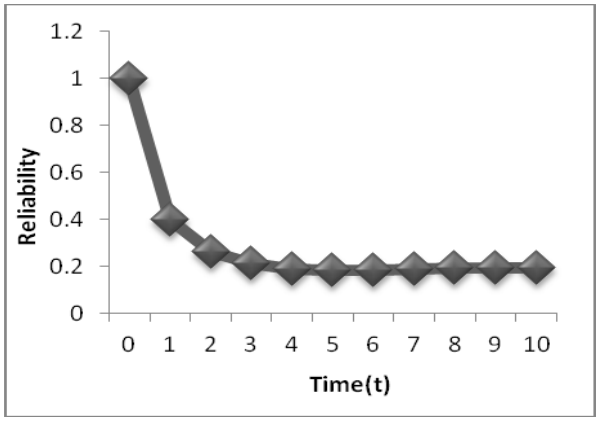


Figure 2: *Reliability against time*

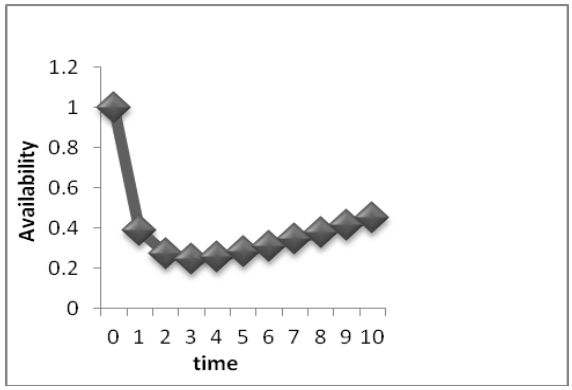


Figure 3: *Availability against time*

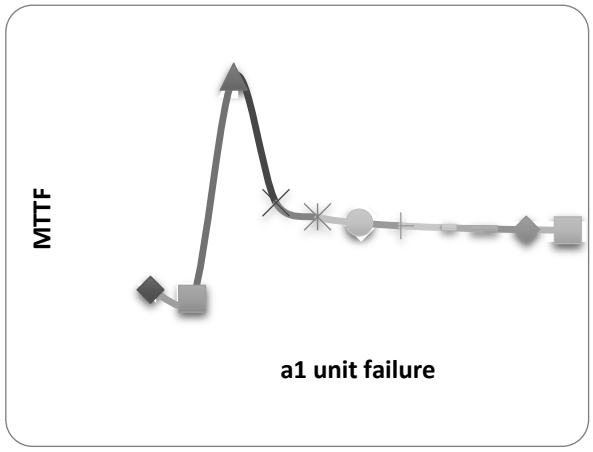


Figure 4: *MTTF Vs* λ_a

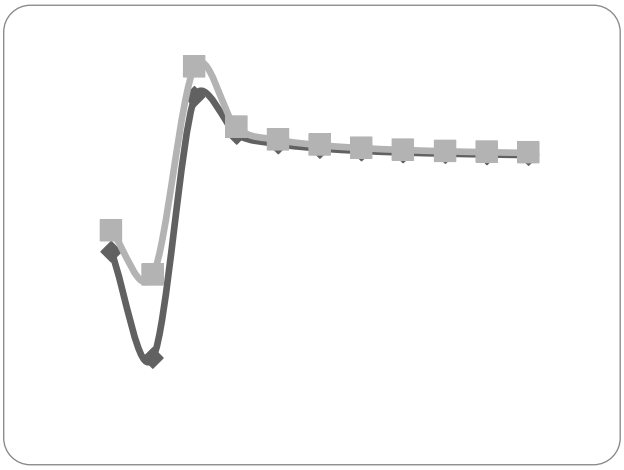


Figure 5: *MTTF Vs* λ_{c_1} *and* λ_{c_2}

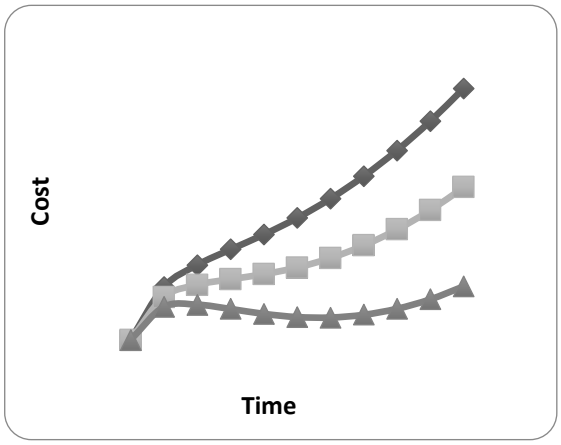


Figure 6: *MTTF Vs* λ_{CSF}

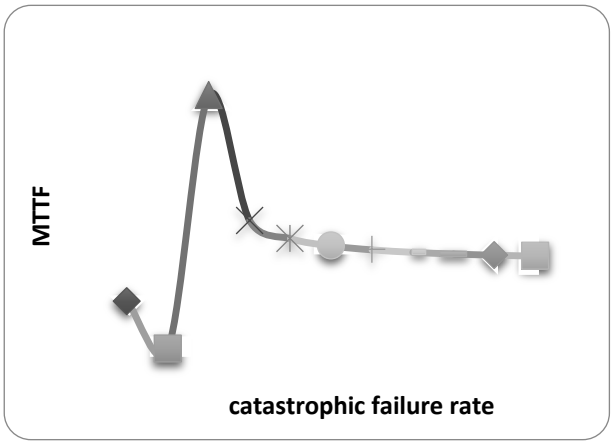


Figure 8: *Cost Vs time*

References

[1] Barlow, R. E. & Proschan, F. (1975), Statistical Theory of Reliability and Life Testing: Probability models, New York: Holt, Rinehart and Winston.

[2] Brown, M. and Proschan, F. (1983), Imperfect repair. Journal of Applied Probability: 20: 851-859.

[3] Cox, D.R. (1995) ' The analysis of non-Markov processes by inclusion of supplementary variables', Proc. Camb. Phil. Soc. (Math. Phys. Sci.), Vol. 51, pp.433-441.

[4] Liebowitz, B. R. (1966), Reliability considerations for a two element redundant system with generalized repair times", Operation Research : 14, 233-241.

[5] Ram, M., SIngh, S.B. and Singh, V.V. (2013) 'Stochastic analysis of a standby complex system with waiting repair strategy', IEEE Transactions on system, Man, and Cybernetics Part A: System and Humans, Vol.43, No. 3, pp. 698-707.

[6] Sengar, Surabhi and Singh (2014), Reliability Analysis of an Engine Assembly Process of Automobiles with Inspection Facility. Mathematical Theory and Modelling: 4(6):153-164.

[7] Sengar, Surabhi and Singh, S. B. (2012), Operational Behaviour and Reliability Measures of a Viscose Staple Fibre Plant Including Deliberate Failures, International journal of reliability and applications: Vol. 13, No.1, pp.1-17.

[8] Sengar, Surabhi and Singh, S. B. (2013), Operational behaviour of major part assembly system of an automobile incorporating human error in maintenance, Applied Mathematics, Vol. 55A pp.13179-13186.

ISSN 1932-2321