Sathyaprayan

CENTENARY CELEBRATIONS

100 Years of Excellence

Centenary Celebrations of St. Thomas' College (Autonomous), Thrissur

1919 Sathyaprayan 2019

Proceedings of

NATIONAL SEMINAR ON

STATISTICAL APPROACHES IN DATA SCIENCE

6-7 February 2019

ISBN: 978-81-935819-2-6

VERITAS VOS LIBERABIT

Estd. 1889

PG & RESEARCH DEPARTMENT OF STATISTICS

ST.THOMAS' COLLEGE

(Autonomous), Thrissur, Kerala - 680 001

Funded by IQAC, St.Thomas College, Thrissur

ISBN: 978-81-935819-2-6

**PROCEEDINGS OF NATIONAL SEMINAR
ON**

STATISTICAL APPROACHES
IN
DATA SCIENCE
6-7 FEB 2019

Publication Division
St.Thomas College (Autonomous), Thrissur
Kerala-680001, India
2019

EDITORIAL BOARD

Language: English, Year: 2019, ISBN 978-81-935819-2-6

Publisher: Publication Division, St. Thomas College (Autonomous), Thrissur

Contact- Email: headstatstct@gmail.com, Ph.No.91 9961 335938, 91 9656 335938

ISBN: 978-81-935819-2-6

ADVISORY COMMITTEE

His Grace Mar Andrews Thazhath, Archbishop, Thrissur Archdiocese, ***Patron***

His Excellency Mar Tony Neelamkavil, Aux.Bishop, Thrissur Archdiocese, ***Manager***

Dr. M. Manoharan, Professor (University of Calicut)

Dr. Davis Antony M (Christ College, Irinjalakkuda)

Dr. S M Sunoj (Cochin University of Science &Technology)

Dr. P. O Jenson (Sahridaya College, Kodakara)

ORGANIZING COMMITTEE

Dr. Ignatius Antony, Principal (Chairman)

Dr.V.M.Chacko, Assistant Professor & Head (Convener)

Prof. Jeena Joseph, Assistant Professor (Coordinator)

Dr. T.A.Sajesh, Assistant Professor & Coordinator (Data Science)

Dr. Rani Sebastian, Assistant Professor

Dr. Nicy Sebastian, Assistant Professor

Prof. R. S. Rasin, Assistant Professor

Prof. Dhanya Sunny Guest Lecturer

Prof. Haritha P U, Guest Lecturer

Prof. Reshma T S., Guest Lecturer

Prof. Rejin Varghese, Assistant Professor (Data Science)

Miss. Sajna O K, Research Scholar

Mrs. Deepthi K S, Research Scholar

Miss. Beenu Thomas, Research Scholar

Miss. Rajitha C, Research Scholar

WELCOME

St. Thomas' College Thrissur is one of the leading academic institutions in the higher education sector of Kerala since 1919. It has a long and proud tradition of excellence in training, teaching and research in many academic disciplines of Science, Arts, Commerce and Humanities.

Post graduate Department of Statistics and Research Centre, St.Thomas' College, Thrissur provides robust computational support to meet the needs of students and researchers. Collection, analysis and interpretation of numerical data have significant role in every research programme. The rapid development in information technology and expansion of the means of communication have made the data collection an easy task upto a certain extent.

However, the processing and organizing of data pose multifarious problems to researchers and experimenters. Here comes the role of Statistics, as a science to provide powerful tools for the analysis and interpretation at every step of research. **Data science is an emerging discipline that draws upon knowledge in statistical methodology and computer science to create impactful predictions and insights for a wide range of traditional scholarly fields.** Statistical ideas are having a profound impact in science, industry and public policy. The Statistics Department at St.Thomas has pioneered many of the tools and ideas behind the research and applications often classified as "data science," where statistics and computer science join together. The Department sees an even brighter future for data science as it harnesses a wider set of ideas to build a new more subtle and powerful science of data. As well as being interested in prediction and statistical computation, our Department puts equal weight on designing experiments, modeling and trying to understand and quantify causal mechanisms, not simply averages and associations, with large data sets. These views are reflected in our curriculum targeted to data science specialists, our faculty's research and the work of our research students.

The Department of Statistics decided to conduct a **National Seminar on Statistical Approaches in Data Science,** during 6 - 7 February 2019. The seminar is intended to impart an understanding of the statistical ideas and methods involved in carrying out research for researchers and teachers from various disciplines of learning.

We are pleased to publish the proceedings of the seminar which containing the abstracts, full papers and programme schedule.

We hope that all of you will enjoy the seminar at large.

Wish you all the best………….

Dr. Ignatius Antony	**Dr. V. M. Chacko**	**Prof. Jeena Joseph**
Principal	HOD& Convener	Coordinator

About St.Thomas College (Autonomous), Thrissur

St. Thomas' College Thrissur is one of the leading academic institutions in the higher education sector of Kerala since 1919. It has a long and proud tradition of excellence in training, teaching and research in many academic disciplines of Science, Arts, Commerce and Humanities. It attracts the brightest minds from all over Kerala and from different parts of the country and Asia. Its alumni have made outstanding contributions to academics, governance and industry. The *University Grants Commission* has granted **Autonomous Status** in 2014 to the college and recognized as **College with Potential for Excellence** in 2016. The college has 14 PG courses and 20 UG courses. All the 9 aided PG departments are Research centers having more than 100 research scholars and 40 research guides.

About the PG and Research Department of Statistics

The Degree Course in B.A. Statistics was started in the year 1955 under the Dept. of Mathematics and Statistics affiliated to the University of Madras. In the year 1958 the B.A. Course was converted to three year B.Sc. course in Statistics under the University of Kerala. The Department of Statistics was established in the year 1984 and the M.Sc. course in Statistics commenced in the same year with Operations Research, Numerical Mathematics and Computer Programming as optional subjects. Department was elevated as Research centre of the Calicut university in the year 2013.Prof.Sebastian J Kulathinal (1984 – 1994), Prof. V.D. Johny (1994 – 2000), Prof A P Jose(2000 – 2003), Dr. T.B. Ramkumar(2003 – 2014) served as the Heads of Department. Prof. A. P. Jose was the Vice principal (2001 – 2003) of the college. Eminant teachers Prof . Krishnakumar (NAAC coordinator 2013), Prof. M.K.Jose, Prof. A. S. Raffy, Prof. P. K. Sasidharan and T. D. Xavier served the department for several years. Dr. P. O. Jenson, the Principal (Rtd) was the member of the department and will be retiring on 2018 April 30. Currently 6 permanent and 3 guest faculty members are working in the department. The department has four research guides and four research scholars. The department offers *Certificate course in Statistical Computing using SPSS* for UG students and *Certificate course in Statistical Computing using R* for PG students. Moreover department provides UGC coaching, JAM coaching etc. The department started B.Voc Data Science course jointly with Department of Computer Science with the UGC assistance in 2018.

Programme Schedule

06 February 2019		
Time	**Session**	**Name**
9.00-9.30 am	Registration	
09.30-10.30 am	Inauguration	
	Welcome	**Dr. V. M. Chacko,** Head of the Department
	Presidential Address	**Dr. Ignatius Antony**, Principal
	Inauguration	**His Grace Mar Tony Neelamkavil** Aux. Bishop of Thrissur Archdiocese
	Benedictory Speech	**Rev. Fr. Varghese Kuthur** Executive Manager
	Felicitation	**Rev. Dr. Martin Kolombrath,** Vice Principal
		Dr. Thomas Paul Kattookkaran Vice Principal
		Dr. K. L. Joy, Vice Principal
	Vote of Thanks	**Prof. Jeena Joseph**, Coordinator
10.30-10.45 am Tea Break		
10.45-11.45 am	**Technical Session I**	**Chair:** Dr.P.O.Jenson, Principal, Sahridaya College, Kodakara [Principal (Rtd), Department of Statistics, St.Thomas College, Thrissur]
	Invited Talk 1	**Dr. A. Krishnamoorthy**, Emeritus Fellow (UGC), Centre for Research in Mathematics, CMS College, Kottayam 686000 [Former Professor, Dept. of Mathematics Cochin University of Science and Technology]
	Topic	*Queueing theory: perspectives from a social angle*
11.45-12.45 pm	**Technical Session II**	**Chair:** Dr. T. D. Xavier, Asso. Professor(Rtd) Department of Statistics, St.Thomas College, Thrissur
	Invited Talk 2	**Dr. P. G. Sankaran,** Professor, Dept. of Statistics, [Former Pro-Vice Chancellor, Cochin University of Science and Technology]
	Topic	*Modeling and analysis of current status lifetime data*
12.45-1.30 pm Lunch Break		
1.30-3.30 pm	**Technical Session III**	**Chair:** Prof. A. P. Jose, HOD Department of Statistics, St.Thomas College, Thrissur (Rtd)
	Invited Talk 3	**Rakesh Poduval,** Senior Statistician, Hedonic Products, Zurich, Switzerland

	Topic	*Introduction to python for data science*
	Invited Talk 4	**Dr. V. Bhuvaneswari,** Asst. Professor, Bharathiar University, Coimbatore
	Topic	*Data Science – A Road Map*
3.30-3.45 pm Tea Break		
3.45-4.30 pm	**Contributory Session I**	**Dr. Mariamma Antony,** Assistant Professor & Head, Department of Statistics, Little Flower College, Guruvayoor *Title: On Some Tailed Distributions And Related Time Series Models* **Dr. DIVYA P R**, Assistant Professor & Head, Dept. of Statistics, Vimala College (Autonomous), Thrissur *Title: Construction and Selection of Single Sampling Variables Plan Through Decision Region*
07 February 2019		
9.30-11.00 am	**Technical Session IV**	**Chair:** Prof. A.S.Raffy, St. Aloysius College, Elthuruth, Thrissur [Associate Professor (Rtd), Department of Statistics, St.Thomas College, Thrissur]
	Invited Talk 5	**M. Syluvai Anthony,** Assistant Professor, Dept. of Statistics, Loyola College, Chennai
11.00-11.15 am	**Tea Break**	
11.15-12.45 pm	**Invited Talk 6**	**Ranjith Kalyana Sundaram,** Data Scientist, Maersk Group Analytics, Chennai, Tamil Nadu, India
Lunch Break 12.45-1.30 pm		
1.30-3.00 pm	**Contributory Session V**	**Dr. Bindu Punathumparambath,** Assistant Professor, Dept. of Statistics, Govt. Arts & Science College, Calicut, Kerala, India *Title: Stress-Strength Reliability of Asymmetric Double Lomax Distribution* **Martin Joseph,** Department of Statistics, Loyola College, Chennai *Title: Advocating Box-Cox transformations to model count response data: A case study* **Reshma T. S.**, Dept. of Statistics, St.Thomas College (Autonomous), Thrissur *Title: Bayesian analysis of generalized maxwell - boltzmann distribution under different loss functions and prior distributions.*

		Rimsha H, Catholicate College, Pathanamthitta Title: *Some Results Related to Location Scale family of Esscher Transformed Laplace Distribution* **Beenu Thomas,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur *Title: Mixed Distribution of Exponential and Gamma* **Deepthi K S,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur *Title: Estimation of Stress-Strength model using Three Parameter Generalized Lindley Distribution* **Rajitha C,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur *Title: A Queueing Network model for the performance analysis of Eye Care Clinic* **Dr.V.M.Chacko**, Assistant Professor & Head**,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur *Title: On Joint Risk Importance Measure*
Tea Break 3.00-3.15 am		
3.15 -3.15pm	Tea Break	
3.15-4.00 pm	**Valedictory Session**	
	Welcome	**Dr.T.A.Sajesh** Coordinator
	Presidential Address	**Dr. V. M. Chacko** Head of the Department
	Inauguration	**Dr. Ignatius Antony** Principal
	Felicitation	**Rev. Dr. Martin Kolambrath** Vice-Principal
		Dr. Thomas Paul Kattookkaran Vice-Principal
		Dr.K.L.Joy, Vice-Principal
	Vote of Thanks	**Dr. Rani Sebastian**, Assistant Professor
4.00-4.30 pm	**Cultural Programmes**	

AUTHERS

INVITED TALKS:

- **Dr. A. Krishnamoorthy**, Emeritus Professor, Dept. of Mathematics, Cochin University of Science and Technology, Cochin
- **Dr. P. G. Sankaran, Professor,** Former Pro-Vice Chancellor Cochin University of Science and Technology, Cochin
- **Dr.V.Bhuvaneswari,** Department of Computer Applciations, Bharathiar University, Coimbatore
- **Rakesh Poduval,** Senior Statistician, Hedonic Products, Zurich, Switzerland
- **M. Syluvai Anthony,** Assistant Professor, Dept. of Statistics, Loyola College, Chennai
- **Ranjith Kalyana Sundaram,** Data Scientist at Maersk Group Analytics, Chennai, Tamil Nadu, India

CONTRIBUTORY PAPERS:

- **Dr.V.M.Chacko,** Assistant Professor & Head, Dept. of Statistics, St.Thomas College (Autonomous), Thrissur
- **Prof. Jeena Joseph,** Assistant Professor, Dept. of Statistics, St.Thomas College (Autonomous), Thrissur
- **Dr. Bindu Punathumparambath,** *Assistant Professor, Dept. of Statistics, Govt. Arts & Science College, Calicut, Kerala, India*
- **Dr.Mariamma Antony,** Assistant Professor & Head, Department of Statistics, Little Flower College, Guruvayoor
- **Dr.Divya P R**, Assistant Professor & Head, Dept. of Statistics, Vimala College (Autonomous), Thrissur
- **Martin Joseph,** Department of Statistics, Loyola College, Chennai
- **Reshma T. S.**, Dept. of Statistics, St.Thomas College (Autonomous), Thrissur
- **Rimsha H,** Catholicate College, Pathanamthitta
- **Beenu Thomas,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur
- **Deepthi K S,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur
- **Rajitha C,** Dept. of Statistics, St.Thomas College (Autonomous), Thrissur

 ISBN: 978-81-935819-2-6

Queueing Theory: Perspectives From A Social Angle

A Krishnamoorthy
Emeritus Fellow (UGC), Centre for Research in Mathematics
CMS College, Kottayam 686000

Abstract

Queueing theory was developed initially for tackling problems arising in telecommunication. However, subsequently it pervaded several branches of Engineering, Technology, Medical field and so on. The purpose of this presentation is to give a social flavour to queueing theory-- individual, social and system optimization.

Modeling And Analysis Of Current Status Lifetime Data

P.G.Sankaran
Cochin University of Science and Technology
Cochin 682022
sankaran.p.g@gmail.com

Abstract

Researchers working with survival data are handling issues associated with incomplete data, particular those associated with various forms of censoring. An extreme form of interval censoring, known as current status observation, refers to situations where the only available information on a survival random variable is whether or not lifetime exceeds a random independent monitoring time . In this talk, a brief review of the extensive literature on the analysis of current status data is presented. Recent extensions of these ideas to more complex forms of survival data including, competing risks, multivariate survival data, and general counting processes are also discussed. Non parametric inference procedures for estimation of survival function are given. In addition, modern theory of efficient estimation in semi-parametric models hasis discussed. Finally, the models are applied to real life situations.

Data Science – A Road Map

V.Bhuvaneswari
Department of Computer Applciations ,Bharathiar University
bhuvanes_v@yahoo.com

Abstract

Data Science has emerged as an important field in the current data era to analyze explosive amount of data generated by machines, business and human. Big Data Analytics is used as a platform to analyze Data insights from huge volume of data. This paper provides a detailed overview on data evolution, data characteristics, Life cycle for analyzing data and techniques. An overview of various open source and commercial tools is presented. The paper provides a Road map of Data Science.

Keywords: Data Analytics, Big Data

I. Introduction

Data Evolution

Data has become the buzz word in the current technological era with new evolving areas related to data management like Data Science, Data Analytics and Big Data. Tremendous increase in data growth happened as social networks like Facebook, Twitter, Whatsup, Orkut, made data sharing easier connecting millions of users through digital tags. The evolution of data transformed, data measurement from Gigabytes to Terabytes, Petabytes, Exabytes, and Yottabytes. The exponential growth of rate of data is expected to be larger than the Physical Universe in the era of Internet of Things creating new universe the "Digital Universe".

All the entities in the digital universe are revolving around "data" creating data deluge. The prediction of increase in size of data in digital universe from 2015 to 2020 is estimated to grow by a factor of 300 exabytes to 40,000 exabytes [1] with a prediction that the growths of data will double every year after 2020. Every entities in digital space like "business, individuals, machines" produce and consume data in multiple forms creating a digital ecosystem coining data with a new term "Big Data".

Data Science has evolved as new a discipline which integrates multiple disciplines like Mathematics, Statistics, Visualization, Databases, Machine Learning, Artificial Intelligence, Visualization and other domains under a common umbrella to analyze huge volume of data. Data Analytics has evolved as important tool to explore and examine insights from the huge data accumulated across domains. Big Data platforms such as Hadoop, SPARK are used for processing storing large volumes of data in conventional computing architectures. This paper provides a detailed road map of data science and analytic tools.

Data Characteristics

The volume of data measured in units greater than terabyte is called as Big Data. The data which do not fit in the commodity hardware for processing is also called as Big Data. The characteristics that transform simple data to Big Data with three defining dimensions are called data volume, velocity and variety. They are collectively termed as 3V's proposed by D.Laney in the year 2001[2]. The three V"s are described below:

***Volume*:** In accordance with the survey organized by IBM in 2012, if the data size exceeds beyond one terabyte, then it could be treated as big data [3]. Snapshots of huge volume of data generated are as follows: Hundreds of petabytes of data roughly equivalent to 100 million gigabytes are processed by Google per month. Amazon maintains a big bank of 152 million customer accounts [4]. Nearly 750 million pictures are uploaded to Facebook on a monthly basis.

***Velocity*:** It refers to rapid and timely collection of data which in turn enhance the commercial value of big data. Few instances are APPLE receives about 47,000 APP downloads every minute. Every 60 seconds, consumers spend $272070 on web shopping.

***Variety*:** It refers to data representation formats, since the data gets generated from multitude of sources. The different data formats are structured, semi-structured and unstructured. According to Cuckier(2010), only 5% of structured data exists in the form of spreadsheets or relational databases. The rest 95% of data which encompasses documents, audio, video, images, graphs and social media text messages are in unstructured form [5].

The other dimensions of data include *Veracity*, *Value*, *Virality* where veracity in data represents the inconsistencies that prevail in data. The Value of data characteristics is analyzing the data which is trustful to drive the business process. Virality of data characteristics represents the rate of data spread. The various types of data that coexist with big data is shown in Fig.1 with description given in Table1.

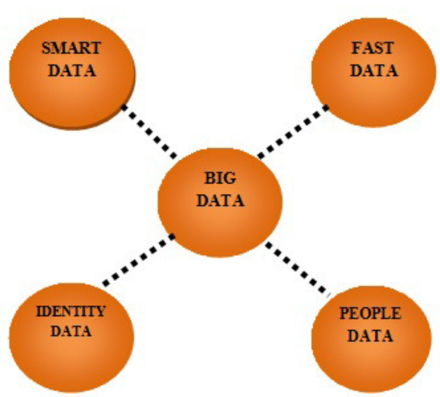


Table 1 Data: Big Data

S.No	Types of Data	Description	Instances
1	Smart Data	Actionable, customized and segmented big data, that can be visualized according to the need of the business[6] .	Brand Sentiment Analysis
2	Fast Data	Instant decisions, real time analysis, crucial for modern enterprises.	Disease outbreaks can be analyzed at the moment so that action can be taken to curb the spread.
3	Identity Data	Future of Big data, used for predictive modeling and machine learning.	Credit card numbers associated with name and address
4	People Data	Customized to analyze the behavior of the customer. Coexist with on-site analytics.	Every activity of the humans on the internet are tracked.

II. Data Analytics

Data is proliferating at an exponential rate, there should be some techniques available to mine and refine it into distilled product information. Data analytics referred as "Actionable Intelligence" is used to convert data into useful and actionable items. The primary objective of data analytics is to control future outcomes.
Data Analytics is important as data is viewed as asset and business organization invest huge budget for storage and maintenance of data and except Return of Investment(ROI) from data. Data Analytics when used as tool in business, data can be turned into a competitive advantage delivering back ROI by providing deep insights of their own business. Analytics also helps to understand the trends in their business process and also helps to drive business for future.

Usecase : Data Analytics Traffic

To control traffic on the road requires analysis of number of vehicles on the road, with a count on maximum, minimum and the average vehicle flow rate. The data analytics when used as a tool help the traffic police to regulate daily traffic, help public to take optimized roots, plan for traffic controlling for events for maintenance of roads, drainage, etc., The trends and patterns in traffic data can also be predicted to understand future traffic in terms of increase of vehicles, congestion in roads which helps for planning construction of roads, bridges and lanes. Data Analytics also exists as solution for complex problems for natural disasters like prediction of frequency and the magnitude of the occurrence of earthquake in a specific area.

 ISBN: 978-81-935819-2-6

2.1 Data Analytics - Process

Data analytics is classified into various types to analyze data in a methodological way the basic hierarchy of Analytics is given in Fig.2

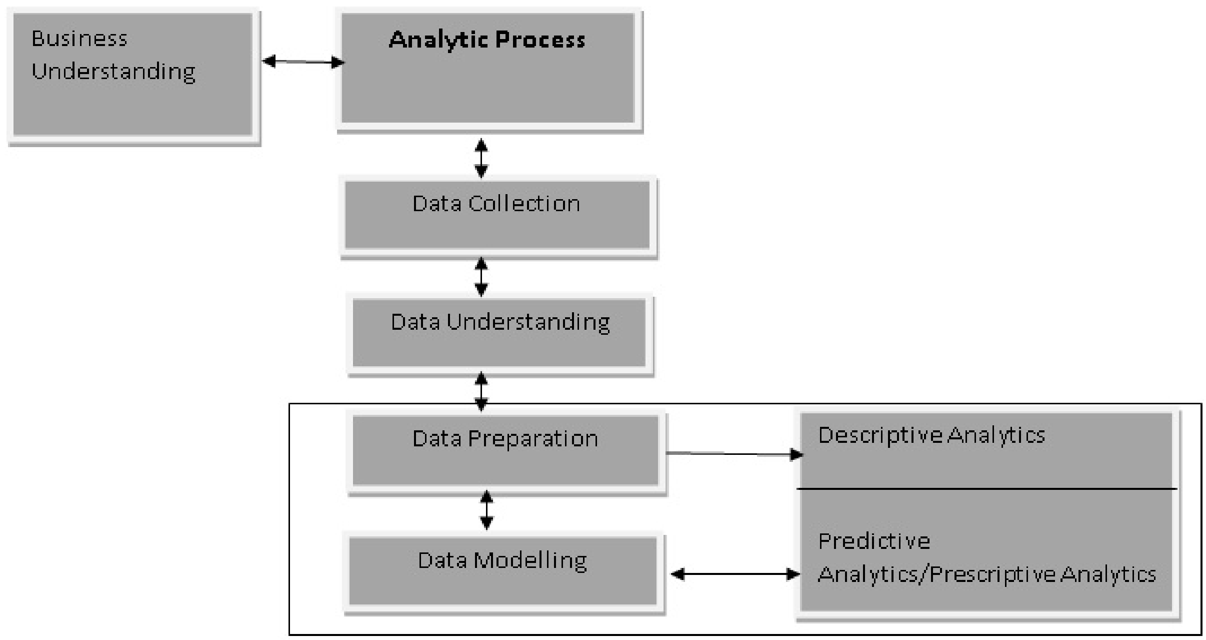


Fig 2. Data Analytics Steps

There are four different types of data analytics as shown in Fig 3. , but choosing the right analytics at the right time make the big data to deliver richer insights , since the data emerges out from multitude of sources[7]. A View of commonly used techniques in data analytics are given in Table 2.

Descriptive Analytics: The most basic form of analytics which learn from the past behaviors to observe their impact on future outcomes. This type of analytics is applied when the company really wants to measure the overall performance of the company in its entirety.

Diagnostic Analytics: An advance form of analytics which is mainly used to discover or to identify the root cause of the problem about why something happened. Analytic Dashboards are used to display the results.

 ISBN: 978-81-935819-2-6

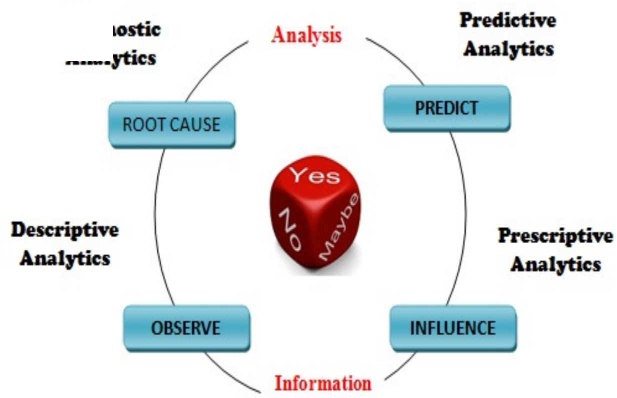


Fig 3: Data Analytics - Types

Predictive Analytics: It helps to forecast the future outcomes by connecting data to effective action. Predictive analytics can be used for the entire sales process, analyzing lead source, number of communications, types of communications, social media etc.,

Prescriptive Analytics: The most vital analysis used by the organizations to reveal the type of action that should be taken in the future. It usually results in the rules and the recommendations for the next steps. For example, in the health care industry, the patient population can be supervised by using prescriptive analytics to measure the number of patients who are clinically obese, then add filters for factors like diabetes and LDL cholesterol levels to determine the type of treatment.

Table 2 Techniques for Data Analytics

Data Analytics	**Techniques and Methods**
Exploratory Data Analytics (EDA)	Statistical Measures, Probability Distributions Correlation, Regression Reporting Visualization
Descriptive Analytics	Cluster and Factor Analysis Multiregression, Bayesian Clustering KNN, Self Organizing Maps Principal Component Analysis Affinity Analysis
Predictive Analytics Algorithms with Re-Enforcement Learning	Decision Tree Ensemble, Bosting, Support Vector Machine Neural Networks Bayesian Classifier Time Series Analysis
Prescriptive Analytics Requires Deep Machine Learning Algorithms	Linear and Non Linear Programming Monte Carlo Simulation Sensitivity Analysis

ISBN: 978-81-935819-2-6

III. Tools for Data Analytics

The various open source and commercial tools available to interpret data are presented below.

3.1 Open Source Analytical Tools

R - It is the most popular and also robust analytical tool widely used now-a-days. 1800 new packages were introduced in R between April 2015 and April 2016. The total number of R packages is now over 8000[8]. R also integrates with Big data platform which in turn helps us to perform the big data analytics in a versatile manner.

Python – The other powerful analytical tool, favorite for data scientists. It includes many analytical and statistical libraries like numpy, scipy etc which makes it as an effective analytical tool.

Tableau Public – It is one of the simple tool used to democratizes visualization. Although there are great alternatives to data visualization, Tableau Public's million row limit acts as a great playground for personal use. With Tableau's visuals, one can quickly investigate a hypothesis, explore the data, and check your intuitions.

Open Refine - A data cleaning software , formerly known as Google Refine helps to get everything ready for analysis. It operates on a row of data which have cells under columns, which is very similar to relational database tables. Open Refine could be used for geocoding addresses to geographic coordinates.

KNIME - One of the best data analytics tools that allow you to manipulate, analyze, and modeling data in an intuitive way via visual programming. KNIME is used to integrate various components for data mining and machine learning via its modular data pipelining concept.

Google Fusion Tables - A much cooler, larger, and nerdier version of Google Spreadsheets. An incredible tool for data analysis, mapping, and large dataset visualization.

NodeXL - It is a visualization and analysis software of relationships and networks. NodeXL provides exact calculations. It is a free and open-source network analysis and visualization software. NodeXL is one of the best statistical tools for data analysis which includes advanced network metrics, access to social media network data importers, and automation[9].

Apache Spark – Spark is another open source processing engine that is built with a focus on analytics, especially on unstructured data or enormous volumes of data. Spark has become tremendously popular in the last couple of years. This is because of various reasons – easy integration with the Hadoop ecosystem being one of them. Spark has its own machine learning library which makes it ideal for analytics as well.

 ISBN: 978-81-935819-2-6

PIG and HIVE – Pig and Hive are integral tools in the Hadoop ecosystem that reduce the complexity of writing MapReduce queries. Both these languages are like SQL (Hive more so than Pig). Most companies that work with Big Data and leverage the Hadoop platform use Pig and/or Hive.

3.2 Commercial Analytics Tools

SAS – SAS continues to be widely used in the industry. It is a robust, versatile and easy to learn tool. SAS has added tons of new modules. Some of the specialized modules that have been added in the recent past are – SAS analytics for IOT, SAS Anti-money Laundering, and SAS Analytics Pro for Midsize Business.

QlikView – Qlikview and Tableau are essentially vying for the top spot amongst the data visualization giants. Qlikview is supposed to be slightly faster than Tableau and gives experienced users a bit more flexibility. Tableau has a more intuitive GUI and is easier to learn.

Excel – Excel is of course the most widely used analytics tool in the world. Whether you are an expert in R or Tableau, you will still use Excel for the grunt work. Excel becomes vital when the analytics team interfaces with the business steam.

References

1 J. Gantz , D. Reinsel. The Digital Universe In 2020: Big Data, Bigger Digital Shadows, And Biggest Growth In The Far East, proceedings of IDC iView,IDC Anal. Future, 2012.
2 Doug Laney, "3d data management : controlling data volume, velocity and variety", Appl. Delivery Strategies Meta Group, 2001.
3 IBM Data Growth and Standards : http://www.ibm.com/developerworks/xml/library/x-dtagrowth/index.html?ca=drs , 2012.
4 K.Cukier, "Data,data everywhere", Economist,2010, pp. 3-16.
5 Amir Gandomi, Murtaza HaiderTed, "Beyond the hype : big data concepts, methods and analytics," International Journal of Information Management, pp. 137-144, 2015.
6 https://www.wired.com/insights/2013/04/big-data-fast-data-smart-data/
7 http://www.revelwood.com/uploads/whitepapers/PA/WP_Real-World-Predictive-Analytics_IBM_SPSS.pdf.
8 http://analyticstraining.com/2011/10-most-popular-analytic-tools-in-business/
9 http://www.digitalvidya.com/blog/data-analytics-tools/

On Joint Risk Importance Measure

V M Chacko
Department of Statistics
St.Thomas College (Autonomous), Thrissur
chackovm@gmail.com

Abstract

Importance measures for the multistate systems mainly concern reliability importance of an individual component. This paper introduces risk importance measure for components in the multistate system with n components. Joint risk importance measures for two and three components are defined. A relationship between Schur-convexity property and joint risk importance is proved. Examples are given.

Keywords- *Importance measures, risk, multistate system, reliability, Schur-convexity*

I. Introduction

Reliability and risk plays a very important role in design, manufacturing and production process. Most of the real life systems are made up of components having two or more levels of performance. A power generation system can produce electric power 100MW, 75MW, 50 MW according to its capacity at the time of performance. An oil transpiration system can supply oil using all its parallel pipes with specified capacity. An amplifier in a radio system is found to be more important component. The identification of most important component in Nuclear power generation system is vital in terms of risk rather than reliability. Importance measures (*IM*s) quantify the criticality of a particular component within a system design. They have been widely used as tools for identifying system weakness, and to prioritize reliability improvement activities. They can also provide valuable information for the safety and operation of a system. In a binary system, reliability optimization mainly deals with maximizing the system reliability under constraints such as cost, weight, and/or size, or on minimizing the cost under reliability constraints. But optimization using risk is still unsolved. However, the extend to which a group of component and its state affect the system is a major concern to the system designer and system controller. From reliability perspective, at the design stage system designers are highly interested in knowing the impact of each potential system component on the system design.

When considering multistate system(*MSS*) with multistate components, research efforts have been focused on generalizing frequently used binary important measures to accommodate the multistate behavior. These approaches characterize, for a given component, the most important component state with regard to its impact on system reliability/risk. For multistate system with multistate components, the problem related to multistate system reliability improvement and risk reduction is still evolving. To solve this problem, methods dependent on the information obtained from multistate *IM*s can be developed for efficient resource allocation. The knowledge about the *IM* can be used as a guide to provide redundancy so that system reliability is increased. Thus, measures that can differentiate such an impact are highly desirable. *IM*s are used to measure the effect of the reliability of individual components on the system reliability. From the design point of view, it is crucial to identify the weakness of the system and how failure of each individual component affects proper functioning of the system; so that efforts can be spent properly to improve the system reliability or reduce risk. In general, there are two ways to improve the reliability or reduce risk of a binary system, 1) increase the reliability of individual components, and/or 2) add redundant components to the system. The Risk Reduction Worth is defined as the decrease in risk when a component is assumed to be optimized or be made perfectly reliable. The Risk Reduction Worth can either be defined as a ratio or as an interval. The FV-importance is introduced by Fussell and Vesely which is based on the cut set and path set. Risk Achievement Worth (RAW) has been introduced for risk analysis terminology. Risk Achievement Worth mainly used as a risk importance measure in probabilistic safety assessments of nuclear power stations. Also Risk Reduction Worth (RRW) is using risk analysis terminology. RRW of component i at time t is denoted by $Ii\ RRW(t)$ and defined as the ratio of the actual system unreliability to the system unreliability with component i replaced by a perfect component i.e. $pi\ (t) \equiv 1$.

Now consider the joint importance. Joint reliability importance*(JRI*) of two or more components is a quantitative measure of the interactions of two or more components or states of two or more components. It is investigated to provide information on the type and degree of interactions between two or more components by identifying the sign and size of it. The value of *JRI* represents the degree of interactions between two or more components with respect to system reliability. *JRI* indicates how components interact in system reliability [2]. Joint structural importance (*JSI*) is used when the component reliabilities are not available. A wide range of *IM*s have been introduced [2], [13]; for binary systems since Birnbaum's work [7]. For example, the Structural importance(*SI*) is used to measure the topological importance of components in the systems. The Birnbaum reliability importance measures the effect of an improvement in component reliability on system reliability; and the joint importance was introduced to measure how two components in a system interact in contribution to the system [2], [13]. The various importance measures can be categorized into either reliability importance or structural importance. A binary system is formed from 2-value logic, e.g., functioning/not-functioning. Although such a system has many practical applications, a model based on dichotomizing the system states is often oversimplified and insufficient for describing many commonly encountered situations in real life. As a result, multistate systems are frequently required. In multistate systems, the components and/or the system performance have more than 2 states. There are numerous examples of multistate systems, with more than 2 ordered or unordered states at the system level, the subsystem level, or the component level [16]. Research on the importance measure for multistate systems has been conducted by several authors [8], [10], [12], [20], who mainly focus on how to extend the reliability importance of individual components from the binary system case to the multistate system case. Ref. [19], extended the joint importance concepts in the multistate system setting with the introduction of *JRI* and *JSI*

 ISBN: 978-81-935819-2-6

for two components. Little research has been done on investigating the joint importance of components (or states) for multistate systems in terms of risk. However, the joint importance provides additional information, which the traditional marginal reliability importance cannot provide, to system designers [14]. Investigating the joint importance in multistate systems is therefore important and helpful in practice. In this paper, the joint risk importance measures two and three components, for multistate systems are studied. The joint risk importance measures how components (or states) in the system interact in contribution to the system performance considering component risk.

The paper is structured as follows. Section II introduces the method of finding Joint risk importance measure (*JRkI)* for two or more components. Section III introduces a relationship between the joint risk importance and Schur-convexity property for the multistate systems. Discussion on this work is given in section IV. An example of an offshore electrical power generation system is given in section V to identify the situation where the requirement of joint risk importance arises. Conclusions are given in section VI.

II. Joint Risk Importance Measures

Assume the n-component multistate system under study is monotone, i.e., the improvement of any component does not degrade the system's state, and the components are mutually state independent. The following definitions are also used in this paper. If for, $\vec{X} = (X_1,\dots,X_n)$, the system structure function is $\phi(.)$,

$$\phi(\underrightarrow{X}) = \min(X_1\dots,X_n) \; or \; \phi(\underrightarrow{X}) = \max(X_1,\dots,X_n)$$

then the system is a multistate series (or parallel) system with state 0,1,…,M, where $\phi(.)$ is a non-decreasing and X_i is state of component i, $0 \le X_i \le M$ [9].

For binary state systems, the pivotal decomposition of structure function is

$$\varphi(X) = x_i\varphi(1_i, X) + (1 - x_i)\varphi(0_i, X)$$

$$1 - \varphi(X) = 1 - x_i\varphi(1_i, X) - (1 - x_i)\varphi(0_i, X)$$

$$1 - E\varphi(X) = 1 - p_i h(1_i, X) - (1 - p_i)h(0_i, X)$$

$$= 1 - (1 - q_i)h(1_i, X) - q_i h(0_i, X)$$

$Let\ F = 1 - (1 - q_i)h(1_i, X) - q_i h(0_i, X)$ where $F = 1 - E\varphi(X)$

$$\frac{\partial}{\partial q_i} F = \square(1_i, X) - h(0_i, X)$$

Clearly $\frac{\partial}{\partial q_i} F = E(\varphi(1_i, X) - \varphi(0_i, X)) = \frac{\partial}{\partial p_i} R$

$$\frac{\partial}{\partial q_i} F = \big(1 - \square(0_i, X)\big) - (1 - \square(1_i, X))$$

$$\frac{\partial}{\partial q_i} F = \big(F(0_i, X)\big) - (F(1_i, X))$$

$$\frac{\partial}{\partial q_i}F = (1-q_j)h(1_i,1_j,X) + q_j h(1_i,0_j,X) - (1-q_j)h(0_i,1_j,X) - q_j h(0_i,0_j,X)$$

$$\frac{\partial^2}{\partial q_j \partial q_i}F = -(1_i,1_j,X) + \square(1_i,0_j,X) + \square(0_i,1_j,X) - h(0_i,0_j,X)$$

$$\frac{\partial^2}{\partial q_j \partial q_i}F = -[h(1_i,1_j,X) - h(1_i,0_j,X)] + [\square(0_i,1_j,X) - h(0_i,0_j,X)]$$

$$\frac{\partial^2}{\partial q_j \partial q_i}F = \frac{\partial}{\partial q_j}F(0_i,X)) - \frac{\partial}{\partial q_j}F(1_i,X) = \frac{\partial}{\partial p_j}h(0_i,X)) - \frac{\partial}{\partial p_j}h(1_i,X)$$

$$\frac{\partial^3}{\partial q_k \partial q_j \partial q_i}F = \frac{\partial}{\partial q_k}[-((1-q_k)h(1_k,1_j,1_i,X) + q_k h(0_k,1_j,1_i,X) - (1-q_k)h(1_k,0_j,1_i,X) - q_k h(0_k,0_j,1_i,X)) + (1-q_k)h(1_k,1_j,0_i,X) + q_k h(0_k,1_j,0_i,X) - (1-q_k)h(1_k,0_j,0_i,X) - q_k h(0_k,0_j,0_i,X)]$$

$$\frac{\partial^3}{\partial q_k \partial q_j \partial q_i}F = [(\square(1_k,1_j,1_i,X) - \square(0_k,1_j,1_i,X) + h(1_k,0_j,1_i,X) + \square(0_k,0_j,1_i,X)) - h(1_k,1_j,0_i,X) + \square(0_k,1_j,0_i,X) + \square(1_k,0_j,0_i,X) - \square(0_k,0_j,0_i,X)$$

The marginal reliability importance and risk importance of a component is same $I_{G(i)} = \frac{\partial R}{\partial p_i} = \frac{\partial F}{\partial q_i}$ and joint risk importance of two components is $I_{G(i,j)} = \frac{\partial^2 F}{\partial q_i \partial q_j}$ where F=1-R, $R(G) = E(\phi(\vec{X}))$ and, q_i, q_j are unreliability/risks of the components i and j of the system G with structure function ϕ respectively, i.e.,

$$JRkI(i,j) = \frac{\partial^2 F}{\partial q_i \partial q_j} = -(h(1_i,1_j,\vec{p}) + h(1_i,0_j,\vec{p}) + h(0_i,1_j,\vec{p}) + h(0_i,0_j,\vec{p})).$$

In order to generalize this equation for more than two components, i.e., to measure the risk importance of the system with respect to the interactive effect of more than two components**,** at first we shall calculate change in the joint risk importance of two components with respect to the change of risk of third component. If there is any change in the joint risk importance due to change in state of third component we can say that there is an interactive effect for three component for the system risk reduction. That is, in the binary setup, the change in the joint risk importance is found to be as follows

$$JRI(i,j,k) = JRI(i,j \mid k=0) - JRI(i,j \mid k=1)$$

where $JRI(i,j \mid k=q) = -h(0_i,0_j,q_k,\vec{p}) + h(0_i,1_j,q_k,\vec{p}) + h(1_i,0_j,q_k,\vec{p}) - h(1_i,1_j,q_k,\vec{p}),$

III. Schur-Convexity And Joint Importance

The characterization help to identify the sign of joint risk importance measures. In this section we shall prove some characterization results for the joint risk importance using Schur-convexity property. [11] discussed Schur-convexity for multistate systems.

A vector $\vec{a} = (a_1, a_2, ..., a_n)$ is said to be majorize the vector $\vec{b} = (b_1, b_2, ..., b_n)$, i.e. $\vec{a} \geq \vec{b}$ if $\sum_{i=j}^{n} a_{[i]} \geq \sum_{i=j}^{n} b_{[i]}$ for $j = 1, 2, ..., n-1$, and $\sum_{i=1}^{j} a_{[i]} \geq \sum_{i=1}^{j} b_{[i]}$ for $j = 2, \ldots, n$, when $a_{[i]}$ and $b_{[i]}$ are components of $\vec{a}$ and $\vec{b}$ arranged in decreasing arrangement.

A real valued function f is said to be Schur-convex (Schur-concave) if $f(\vec{a}) \geq (\leq) f(\vec{b})$ whenever $\vec{a} \geq \vec{b}$.

A characterization of f to be Schur-convex (Schur-concave) is that, for $i \neq j$

$$(a_i - a_j)\left(\frac{\partial f(\vec{a})}{\partial a_i} - \frac{\partial f(\vec{a})}{a_j}\right) \geq (\leq) 0.$$

This characterization is known as Schur-Ostrowaki's condition. If $\vec{a}_1, ..., \vec{a}_n$ denote n vectors each with m components, then f defined on $(R^m)^n$ is said to be a Schur-convex (Schur-concave) if, for $i \neq j$ and for two components k and m,

$$(a_{ki} - a_{mj})\left(\frac{\partial f(\vec{a}_1, ..., \vec{a}_n)}{\partial a_{ki}} - \frac{\partial f(\vec{a}_1, ..., \vec{a}_n)}{a_{mj}}\right) \geq (\leq) 0.$$

For proving Schur-convexity property of performance measure of multistate system, let $\phi_j(\vec{X}) = 1$ if $\phi(\vec{X}) \geq j$ and zero otherwise, we consider

$$(q_{im} - q_{lk})\left(\frac{\partial F}{\partial q_{im}} - \frac{\partial F}{\partial q_{lk}}\right)$$

where

$$F = 1 - R = 1 - E\phi_j(\vec{X}) = 1 - P[\phi(\vec{X}) \geq j],\ q_{im} = P[X_i < m].$$

Let $X_i' = 1$ if $X_i \geq m$ and zero otherwise.

$$R = q_{im}.q_{lk} P[\phi_j(\vec{X}) = 1, X_i' = 0, X_l' = 0] + \ q_{im}(1 - q_{lk}) P[\phi_j(\vec{X}) = 1, X_i' = 0, X_l' = 1] +$$
$$(1 - q_{im}) q_{lk} P[\phi_j(\vec{X}) = 1, X_i' = 1, X_l' = 0] + \ (1 - q_{im})(1 - q_{lk}) P[\phi_j(\vec{X}) = 1, X_i' = 1, X_l' = 1].$$

$$\frac{\partial F}{\partial q_{im}} = -q_{lk} P[\phi_j(\vec{X}) = 1, X_i' = 0, X_l' = 0] - \ (1 - q_{lk}) P[\phi_j(\vec{X}) = 1, X_i' = 0, X_l' = 1] +$$
$$+ q_{lk} P[\phi_j(\vec{X}) = 1, X_i' = 1, X_l' = 0] + \ (1 - q_{lk}) P[\phi_j(\vec{X}) = 1, X_i' = 1, X_l' = 1].$$

$$\frac{\partial F}{\partial q_{lk}} = -q_{im}P[\phi_j(\overrightarrow{X}) = 1, X_i' = 0, X_l' = 0] + \quad q_{im}P[\phi_j(\overrightarrow{X}) = 1, X_i' = 0, X_l' = 1]$$
$$- (1 - q_{lm})P[\phi_j(\overrightarrow{X}) = 1, X_i' = 1, X_l' = 0] \ + (1 - q_{im})P[\phi_j(\overrightarrow{X}) = 1, X_i' = 1, X_l' = 1].$$

Suppose that

$$\phi(X_1,..., x_i,..., y_l,..., X_n) = \phi(X_1,..., y_i,..., x_l,..., X_n).$$

That is, the system is symmetric [21]. We can prove that

$$\left(\frac{\partial F}{\partial q_{im}} - \frac{\partial F}{\partial q_{lk}}\right) = (q_{im} - q_{lk})\big[P[\phi_j(1_i,1_l) = 1] - 2P[\phi_j(1_i,0_l) = 1] + P[\phi_j(0_i,0_l) = 1]\big]$$

and

$$(q_{im} - q_{jk})\left(\frac{\partial F}{\partial q_{im}} - \frac{\partial F}{\partial q_{lk}}\right) = (q_{im} - q_{lk})^2[-JRkI]$$

where $JRkI$ represents the joint risk importance of binary-imaged multistate system, i.e., R is Schur-convex (Schur-concave)implies $LHS>(<)0 \Rightarrow \ JRkI < (>)0$. Thus the sign of the joint importance for two components can be verified using the Schur-convexity property of multistate system.

IV. Discussion

As is the case of reliability, risk plays an important role in system design. Risk reduction individually and jointly require in Nuclear power systems. The basic notion of the joint risk importance measures for two and three of components is studied. The concept of proposed joint risk importance measures can be extended and explained in many engineering applications for risk reduction and risk achievement. Schur-convexity serves as a characterizing criteria for the joint importance measures to be positive or negative. The form and degree of interaction in reducing risk can be obtained using risk importance measures.

V. Example

Eaxmple 1: Ref. [15] considered an offshore electrical power generation system. The purpose of this system is to supply two nearby oilrings with electrical power. Both oilrings have their own main generation, represented by equivalent generators A_1 and A_3 each having capacity of 50MW. In addition the oilring has a standby generator A_2 that is switched into the network in case of outage of A_1 or A_3, or may be used in extreme load situations in either of the oilrings. The A_2 also has capacity 50MW. The control unit, U, continuously supervises the supply from each of the generators with automatic control of the switches. If for instance the supply from A_3 to oilring 2 is not sufficient, whereas the supply from A_1 to oilring 1 is sufficient, U can activate A_2 to supply oilring 2 with electrical power through the subsea cables L.

TABLE 1
MINIMAL PATH VECTORS OF ϕ_2

Levels	U	A_1	L	A_2	A_3
1	4	4	2	2	0
1,2	2	0	0	0	2
2	4	4	2	4	0
2	4	4	4	2	0
3	4	4	2	2	2
3,4	2	0	0	0	4
4	4	4	2	4	2
4	4	4	4	2	2
4	4	4	4	4	0

The components have states {0, 2, 4} and the system have states {0, 1, 2, 3, 4}, where 0, 1, 2, 3 and 4 represents the states of system at capacities 0MW, 12.5MW, 25MW, 37.5MW, and 50MW respectively. The minimal path vectors to the levels of the structure function $\phi_2(U, A_1, L, A_2, A_3) = I(U > 0)\min(A_3 + A_2 I(U = 4)I(A_1 = 4)L/4, 4)$ are given (Table 1).

Example 2: Consider a series system of three components with risk q_1, q_2 and q_3.

The $F=1-(1-q_1)(1-q_2)(1-q_3)$. $\frac{\partial F}{\partial q1} = (1-q_2)(1-q_3)$. $\frac{\partial^2 F}{\partial q2 \partial q1} = -(1-q_3)$.

VI. Conclusions

The joint risk importance measures play an important role in reliability/risk studies. For identifying the most important group of components, we have suggested a new joint risk importance measures for *2 or 3* components to the multistate system. We can use the *JSI* measures when reliabilities are not available. At the reliability design phase, the joint risk importance can improve system designer's understanding of the relationship between components and system and also among the components, which is quite desirable. This paper introduced joint risk importance measures for multistate system on the basis of the performance. Also a characterization result for the joint importance is derived using Schur-convexity property. From the optimization point of view, we need importance of each components, joint risk importance of two components or three or more components. In this respect, the main advantage of our joint risk importance measure is the information provided by them. It gives useful information for safe and efficient operation of the system, where existing importance measures gives information about individual component reliability importance and joint reliability importance of components.

References

[1] A. M. Abouammoh, and Al-kadi, "On measures of importance for components in multistate coherent system", *Microel.Reliab.,* VoL 31, No. 1, pp 109-122, 1991.
[2] M. J. Armstrong, "Joint reliability importance of components", *IEEE Trans.*

*Reliab.,*Vol. 44, No. 3, pp.408-412, Sep. 1995.
[3] T. Aven, "On performance measures for multistate monotone systems*", Reliab. Eng. Syst. Saf.,* Vol. 41(3), pp 259–266, 1993.
[4] R. E. Barlow, and F. Proschan, "Statistical theory of reliability and Life testing", NewYork: Holt, Rinheart & Winston, 1975.
[5] R. E. Barlow, and A. Wu, "Coherent system with multistate components", *Math. Oper. Res.,* Vol. 3, pp. 275-281, Nov. 1978.
[6] H. W. Block and T. Savits, "A decomposition for multistate monotone ystem", *Jr. Appl. Prob*., Vol. 19, pp. 391–402, Jun.1982.
[7] Z. W. Birnbaum, "On the importance of different components in a multicomponent system", in *Multivariate analysis II*, P. R. Krishnaiah, Ed. NewYork: Academic Press, 1969, pp. 591-592
[8] V. C. Bueno, "On the importance of components for multistate monotone systems", *Stat. Prob. Lett*.,Vol. 7, pp. 51-59, Jul. 1989.
[9] V. M. Chacko, "Joint Importance measures for multistate reliability system", Opsearch (Available online) Vol.48, No.3, 257-278, 2011
[10] E. El-Neweihi, F. Proschan, and J. Sethuraman, "Multistate coherent systems", *Jr. Appl. Prob*.,Vol. 15, pp. 675–688, Dec. 1978
[11] G. Gopal, "Schur property of the performance function for the multistate coherent system", *Proc. 3 In. Conf. mathematical meth. Reliab*., Trondheim, Norway, Jun. 2002, pp.255-258.
[12] W. S. Griffith, "Multistate reliability models", *Jr. Appl. Prob*., Vol. 17, pp.735–744, 1989.
[13] J. S. Hong, and C. H. Lie, "Joint reliability importance of two edges in an undirected network*", IEEE Trans. Reliab*.,Vol. 47, No. 1, pp. 97-101, Mar. 1993
[14] J. S. Hong, H. Y. Koo, and C. H. Lie, "Joint reliability importance of *k*-out-*n* systems", *Europ. Jr. Oper. Res*., Vol.42, pp. 539–547, Aug. 2002.
[15] G. Levitin, and A. Lisnianski, "Importance and sensitivity analysis of multistate systems using the universal generating function", *Reliab. Eng. Syst. Saf*, Vol. 65, pp. 271-282, 1999.
[16] B. Natvig, S. Sormo, A. T. Holen, and G. Hogasen, "Multistate reliability theory -a case study", *Adv. Appl. Prob*,Vol. 18, pp. 921-932, 1986.
[17] F. C. Meng, "Component relevancy and characterization results in multistate systems*", IEEE Trans. Reliab*.,Vol. 42, pp.478-483, Sep.1993.
[18] F. C. Meng, "Comparing the importance of system components by some structural characteristics*", IEEE Trans. Reliab.*, Vol. 45, pp. 59-65, Mar. 1996.
[19] S. Wu, "Joint importance of multstate systems", *Comp. Indust. Eng.,* Vol. 49, pp. 63–75, Jul. 2005
[20] S. Wu, and L-Y. Chan, "Performance utility analysis of multistate systems*", IEEE Trans. Reliab*., Vol. 52, pp. 14-21, Mar. 2003.
[21] J. Xue and K. Yang, "Symmetric relations in multistate systems",*IEEE Trans. Reliab*., Vol. 44, pp.689-693, Dec.1995.

Advocating Box-Cox transformations to model count response data: A case study

Martin Joseph[1], *Martin L.William*[2]
Department of Statistics, Loyola College, Chennai
[1]*E-mail :nitram9864@gmail.com*

Abstract

In modelling a count response variable, a commonly followed approach is to assume that it is Poisson and take the square-root transformation to build a linear regression model for the transformed variable or build a 'Generalized Linear Model' with log link function. However, in some situations, the count variable may not be distributed according to Poisson law and, in such cases the above approaches are not justified. In this paper, we advocate the Box-Cox approach which identifies the appropriate transformation to be applied on the response variable to build a regression model. This is performed with a case study on bike-sharing system in the United States for modelling the count data on 'number of bikes taken on rent by customers'.We show that Box-Cox approach leads to more accurate prediction.

REFERENCES

[1] G.E.P. Box and D.R. Cox (1964). Analysis of Transformations, Journal of the Royal Statistical Society (B), 26(2), 211-252

[2] Michael H. Kutner, Christopher J. Nachtsheim, John Neter, William Li. Applied Linear Statistical Models, Fifth Edition, McGraw-Hill Irwin Publications

On a Bathtub Shaped Failure Rate Model

V M Chacko and Beenu Thomas
Department of Statistics, St.Thomas College, Thrissur
Kerala-680001, India
beenuneel.18@gmail.com

Abstract

Many of the real life systems exhibit bathtub shapes for their failure rate functions. This paper considered a generalization of existing model and discusses the failure rate behavior but this proposed distribution allows only bathtub (upside down bathtub) shapes for its hazard rate function. The model is illustrated with example.

Keywords: Bathtub failure rate, Reliability

I. Introduction

There are many distributions for modeling lifetime data. Many of them exhibit bathtub shape for their failure rate functions. Most real life system exhibit bathtub shapes for their failure rate functions.

In this paper we consider generalization of a one parameter bathtub shaped failure rate model discussed in Chacko et.al (2017) but exhibiting bathtub shaped failure rate and discuss the failure rate behavior of the distribution. The inference procedure also becomes simpler than GL, GG and EW distributions.

The generalization of the one parameter bathtub shaped failure rate distribution and their properties are discussed in section II and model using additional scale parameter is discussed in section III respectively. Data analysis is given in section IV. Conclusions are given at the final section.

II. Generalization of Bathtub shaped failure rate model

Consider a life time random variable X having distribution function F(x) and failure rate function h(x),

$$h(x) = \frac{1 + a\lambda x + (\lambda x)^2}{1 + \lambda x + (\lambda x)^2}, x > 0, a > 0, \qquad \lambda > 0,$$

where α is the shape parameter and λ is the scale parameter

$$\int \frac{1 + a\lambda x + (\lambda x)^2}{1 + \lambda x + (\lambda x)^2} dx = \frac{(a-1)}{\lambda}\left[\frac{\ln(1 + \lambda x + (\lambda x)^2)}{2} - \frac{arctan\left(\frac{1+2\lambda x}{\sqrt{3}}\right)}{\sqrt{3}}\right] + x$$

Corresponding distribution function is

$$F(x) = 1 - e^{-(x + \frac{(a-1)}{\lambda}(\frac{\ln(1+\lambda x+(\lambda x)^2)}{2} - \frac{\arctan((1+2\lambda x)/\sqrt{3})}{\sqrt{3}} + \frac{\pi}{6\sqrt{3}})}, x > 0, -\infty < a < \infty,\ \lambda > 0.$$

The probability density function (PDF) is given by

$$f(x)=\frac{1+a\lambda x+(\lambda x)^2}{1+\lambda x+(\lambda x)^2}\,e^{-(x+\frac{(a-1)}{\lambda}(\frac{\ln(1+\lambda x+(\lambda x)^2)}{2}-\frac{\arctan((1+2\lambda x)/\sqrt{3})}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}})},x>0,\ -\infty<a<\infty,\lambda>0.$$

Since distribution function can be computed from failure rate function, we consider failure rate function, for λ=1 at first,

$$h(x)=\frac{1+ax+x^2}{1+x+x^2}\ ,x>0,a>0.$$

$$\int\frac{1+ax+x^2}{1+x+x^2}dx=\int\left(\frac{(a-1)x}{1+x+x^2}+1\right)dx=(a-1)\int\frac{x}{1+x+x^2}dx+\int 1dx$$

$$=(a-1)\int\frac{2x+1}{2(1+x+x^2)}dx-\int\frac{1}{2(1+x+x^2)}dx+\int 1dx$$

$$=\frac{(a-1)}{2}\int\frac{1}{w}dw-\frac{2}{2\sqrt{3}}\int\frac{1}{(1+u^2)}du+\int 1dx$$

(By substituting $w=\frac{1}{1+x+x^2},\ u=\frac{1+2x}{\sqrt{3}}$)

$$=\frac{(a-1)}{2}log(1+x+x^2)-(a-1)\frac{1}{\sqrt{3}}arctan\frac{(1+2x)}{\sqrt{3}}+x$$

Here, we consider a simplified form of distribution function,

$$F(x)=1-e^{-(x+(a-1)(\frac{\log(1+x+x^2)}{2}-\frac{\arctan((1+2x)/\sqrt{3})}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}})},x>0,a>0 \qquad (1)$$

It is an alternative model for GL, GG, EW distributions. Clearly F (0) = 0, F(∞) = 1, *F* is non-decreasing and right continuous. More over F is absolutely continuous. The probability density function (PDF) is given by

$$f(x)=\frac{1+ax+x^2}{1+x+x^2}\,e^{-(x+(a-1)(\frac{\log(1+x+x^2)}{2}-\frac{\arctan((1+2x)/\sqrt{3})}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}})},x>0,\ a>0. \qquad (2)$$

It is a positively skewed distribution. Probability density function is unimodel. Peakedness increases with the increase of value of α

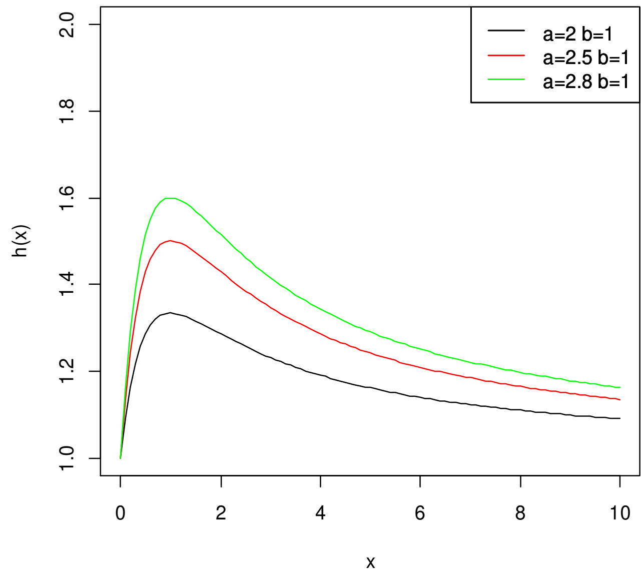


Figure 1. Failure rate function for a=2,2.5,2.8 and b=1

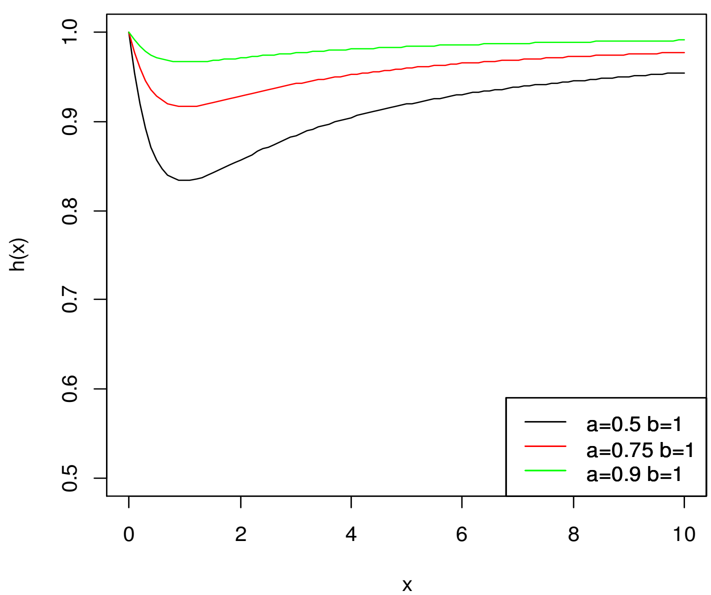


Figure 2. Failure rate function for a=0.5,0.75,0.9 and b=1

Figure 1 shows Upside down bathtub shape in its failure rate function. From Figure 2, the shape of the hazard rate function appears monotonically decreasing or to initially decrease and then increase, a bathtub shape. The proposed distribution allows only bathtub shapes or constant failure rate for its hazard rate function.

ISBN: 978-81-935819-2-6

Proposition 1: The proposed distribution is a generalization of Exponential distribution. When a=1, it becomes exponential distribution $f(x) = e^{-x}, x > 0$. When a = 1, it becomes constant failure rate model.

2.1 Moments

All the raw and central moments, moment generating functions etc exist, since the function f(x) is having countable number of discontinuities, and integrable but the resulting function require more mathematical treatment. It can be done by softwares like Matlab, R etc.

2.2 Quantile and Median

In this section, we determine the explicit formulas of the quantile and the median of this distribution. The quantile x_p of the distribution is given from

$$F(x_p) = p, 0 < p < 1.$$

We obtain the 100 p^{th} percentile,

$$x + (a-1)\left[\frac{\log(1+x+x^2)}{2} - \frac{\left(\tan^{-1}\left(\frac{1+2x}{\sqrt{3}}\right) - \frac{\pi}{6}\right)}{\sqrt{3}}\right] = -\log(1-p)$$

Setting $p = 0.5$ in Eq. (3.1), we get the median of the distribution as follows.

$$x + (a-1)\left[\frac{\log(1+x+x^2)}{2} - \frac{\left(\tan^{-1}\left(\frac{1+2x}{\sqrt{3}}\right) - \frac{\pi}{6}\right)}{\sqrt{3}}\right] = -\log(1-0.5)$$

x_p is the solution of above monotone increasing function. Software can be used to obtain the quantiles/percentiles.

2.3 Distribution of Maximum and Minimum

To analyse series and parallel system reliability, we need to get the distribution of Minimum and Maximum. Let $X_1, X_2, \ldots, X_n$ be a simple random sample from the proposed distribution with CDF and PDF as in (1) and (2), respectively. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics obtained from this sample. In this section we provide the expressions for the PDFs of order statistics for the distribution. The PDF of $X_{(r)}$ is given by,

$$f_{r:n}(x) = \frac{1}{B(r,n-r+1)}[F(x;\alpha)]^{r-1}[1-F(x;\alpha)]^{n-r}f(x;\alpha)$$

where $F(x;\alpha)$, $f(x;\alpha)$ are the CDF and PDF given by (1) and (2), respectively.

$$f_{r:n}(x) = \frac{1}{B(r,n-r+1)}[1-\text{Ex}]^{r-1}\ [\text{Ex}]^{n-r}\frac{1+ax+x^2}{1+x+x^2}\ [\text{Ex}] \qquad (3)$$

where $\text{Ex} = e^{-\left(x+(a-1)\left(\frac{\log(1+x+x^2)}{2}-\frac{\arctan\left(\left(1+\frac{2x}{\sqrt{3}}\right)\right)}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}}\right)\right)}$.

Then the PDF of the smallest and largest order statistics $X_{(1)}$ and $X_{(n)}$ is respectively given by:

$$f_1(x) = \frac{1}{B(1,n)}[\text{Ex}]^{n-1}\frac{1+ax+x^2}{1+x+x^2}[\text{Ex}]$$

$$f_n(x) = \frac{1}{B(n,1)}[1-\text{Ex}]^{n-1}\frac{1+ax+x^2}{1+x+x^2}\ [\text{Ex}]$$

The CDF of $X_{(r)}$ is given by,

$$F_{r:n}(x) = \sum_{j=r}^{n}\binom{n}{j}\ F^j(x)[1-F(x)]^{n-j}$$

$$F_{r:n}(x) = \sum_{j=r}^{n}\binom{n}{j}[1-Ex]^j[Ex]^{n-j} \qquad (4)$$

Then the CDF of the smallest and largest order statistics $X_{(1)}$and $X_{(n)}$is respectively given by:

$$F_1(x) = 1-\left[e^{-\left(x+(a-1)\left(\frac{\log(1+x+x^2)}{2}-\frac{\arctan((1+2x/\sqrt{3}))}{\sqrt{3}}-\frac{\pi}{6\sqrt{3}}\right)\right)}\right]^n$$

$$F_n(x) = \left[1-e^{-\left(x+(a-1)\left(\frac{\log(1+x+x^2)}{2}-\frac{\arctan((1+2x/\sqrt{3}))}{\sqrt{3}}-\frac{\pi}{6\sqrt{3}}\right)\right)}\right]^n$$

Reliability of a series system having n components with n independent and identically distributed (iid) distribution is $F_1(x)$ and reliability of a parallel system having n components with iid distribution is $F_n(x)$.

III. Statistical Inference

Method of maximum likelihood estimation and asymptotic confidence intervels are obtained in this section.

3.1. Maximum likelihood estimation

In this section, we derive the maximum likelihood estimates of the unknown parameters α and λ of the proposed distribution based on a complete sample. Let us assume that we have a simple random sample X_1, X_2, …, X_n from this proposed distribution. The likelihood function of this sample is

$$L = \prod_{i=1}^{n} f(x_i;a,\lambda)$$

The log-likelihood function becomes

$$\log L=\sum_{i=1}^{n}[\log(1+a\lambda x_i+(\lambda x_i)^2)-\log(1+\lambda x_i+(\lambda x_i)^2)]-\sum_{i=1}^{n}\{x_i+\frac{(a-1)}{\lambda}[\frac{\log(1+\lambda x_i+(\lambda x_i)^2)}{2}-\frac{\arctan(\frac{1+2\lambda x_i}{\sqrt{3}})}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}}]\}$$

The first partial derivatives of the log-likelihood function with respect to the two-parameters are

$$\frac{\partial\log L}{\partial a}=\sum_{i=1}^{n}\frac{1}{\frac{1}{\lambda x_i}+a+\lambda x_i}-\sum_{i=1}^{n}\left\{\frac{1}{\lambda}\left[\frac{\log(1+\lambda x_i+(\lambda x_i)^2)}{2}-\frac{\arctan(\frac{1+2\lambda x_i}{\sqrt{3}})}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}}\right]\right\}$$

and

$$\frac{\partial\log L}{\partial\lambda}=\sum_{i=1}^{n}[\frac{(ax_i+2\lambda x_i^2)}{1+a\lambda x_i+(\lambda x_i)^2}-\frac{(x_i+2\lambda x_i^2)}{1+\lambda x_i+(\lambda x_i)^2}]-\sum_{i=1}^{n}\{[\frac{(a-1)}{2\lambda^2}(\lambda\frac{2\lambda x_i^2+x_i}{1+\lambda x_i+(\lambda x_i)^2}$$

$$-\log(1+\lambda x_i+(\lambda x_i)^2))]-\frac{(a-1)}{\lambda^2\sqrt{3}}[\frac{2\lambda x_i}{\sqrt{3}(1+(\frac{1+2\lambda x_i}{\sqrt{3}})^2)}-\arctan(\frac{1+2\lambda x_i}{\sqrt{3}})+\frac{\pi}{6}]\}$$

Setting the left side of the above two equations to zero, we get the likelihood equations as a system of two nonlinear equations in a and λ. Solving this system in a and λ gives the maximum likelihood estimates of a and λ. It can be obtain estimates using R software by numerical methods.

When scale parameter is 1, the likelihood equations becomes,

$$L(a,x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}f(x_i)=\prod_{i=1}^{n}\frac{1+ax_i+x_i^2}{1+x_i+x_i^2}e^{-\left(x_i+(a-1)\left(\frac{\log\left(1+x_i+x_i^2\right)}{2}-\frac{\tan^{-1}\left(\frac{1+2x_i}{\sqrt{3}}\right)}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}}\right)\right)}$$

$$\log L=\sum_{i=1}^{n}\left[\log\left(1+ax_i+x_i^2\right)-\log\left(1+x_i+x_i^2\right)\right]-\sum_{i=1}^{n}\left[x_i+(a-1)\left(\frac{\log\left(1+x_i+x_i^2\right)}{2}-\frac{\tan^{-1}\left(\frac{1+2x_i}{\sqrt{3}}\right)}{\sqrt{3}}+\frac{\pi}{6\sqrt{3}}\right)\right]$$

 ISBN: 978-81-935819-2-6

$$\log L = \sum_{i=1}^{n} \log\left(1 + ax_i + x_i^2\right) - \sum_{i=1}^{n} \log\left(1 + x_i + x_i^2\right) - \sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n}\left(\frac{\log\left(1 + x_i + x_i^2\right)}{2}\right)$$

$$-(a-1)\sum_{i=1}^{n}\left(\frac{\arctan\left(\frac{1+2x_i}{\sqrt{3}}\right)}{\sqrt{3}} + \frac{\pi}{6\sqrt{3}}\right)$$

$$\frac{\partial \log L}{\partial a} = 0 \Rightarrow \sum_{i=1}^{n} \frac{x_i}{\left(1 + ax_i + x_i^2\right)} - \sum_{i=1}^{n}\left[\frac{\log\left(1 + x_i + x_i^2\right)}{2} - \frac{\arctan\left(\frac{1+2x_i}{\sqrt{3}}\right)}{\sqrt{3}} + \frac{\pi}{6\sqrt{3}}\right] = 0$$

$$\frac{\partial \log L}{\partial a} = 0 \Rightarrow \sum_{i=1}^{n} \frac{1}{\left(\frac{1}{x_i} + a + x_i\right)} - \sum_{i=1}^{n}\left[\frac{\log\left(1 + x_i + x_i^2\right)}{2} - \frac{\arctan\left(\frac{1+2x_i}{\sqrt{3}}\right)}{\sqrt{3}} + \frac{\pi}{6\sqrt{3}}\right] = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{1}{\left(\frac{1}{x_i} + a + x_i\right)} = \sum_{i=1}^{n}\left[\frac{\log\left(1 + x_i + x_i^2\right)}{2} - \frac{\arctan\left(\frac{1+2x_i}{\sqrt{3}}\right)}{\sqrt{3}} + \frac{\pi}{6\sqrt{3}}\right] = 0$$

Numerical methods require to solve the equations.

3.2 Asymptotic Confidence Bounds

In this section, we derive the asymptotic confidence intervals of these parameters when

$a > 0$ and $\lambda > 0$ as the MLEs of the unknown parameters $a > 0$ and $\lambda > 0$ cannot be obtained in closed forms, by using variance covariance matrix I^{-1}, where I^{-1} is the inverse of the observed information matrix which defined as follows

$$I^{-1} = \begin{pmatrix} -\frac{\partial^2 L}{\partial a^2} & -\frac{\partial^2 L}{\partial a \partial \lambda} \\ -\frac{\partial^2 L}{\partial \lambda \partial a} & -\frac{\partial^2 L}{\partial \lambda^2} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} var(\hat{a}) & cov(\hat{a}, \hat{\lambda}) \\ cov(\hat{\lambda}, \hat{a}) & var(\hat{\lambda}) \end{pmatrix}$$

The second partial derivatives as follows

$$\frac{\partial^2 \log L}{\partial a^2} = -\sum_{i=1}^{n} \frac{(\lambda x_i)^1}{(1 + a\lambda x_i + (\lambda x_i)^2)^2}$$

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \sum_{i=1}^{n} \left[\frac{\left(1 + a\lambda x_i + (\lambda x_i)^2\right)2x_i^2 - (ax_i + 2\lambda x_i^2)^2}{\left(1 + a\lambda x_i + (\lambda x_i)^2\right)^2} \right] - \sum_{i=1}^{n} \left[\frac{\left(1 + \lambda x_i + (\lambda x_i)^2\right)2x_i^2 - (x_i + 2\lambda x_i^2)^2}{\left(1 + \lambda x_i + (\lambda x_i)^2\right)^2} \right] -$$

$$\sum_{i=1}^{n} \left\{ \left[\frac{(a-1)}{2} \left(\frac{\lambda(1 + \lambda x_i + (\lambda x_i)^2)2x_i^2 - (2\lambda x_i^2 + x_i)^2}{\lambda^2 \left(1 + \lambda x_i + (\lambda x_i)^2\right)^2} \right) - \left(\frac{\frac{\lambda^2 (x_i + 2\lambda x_i^2)}{\left(1 + \lambda x_i + (\lambda x_i)^2\right)} - 2\lambda \log\left(1 + \lambda x_i + (\lambda x_i)^2\right)}{\lambda^4} \right) \right] \right\}$$

$$+ \sum_{i=1}^{n} \frac{(a-1)}{\sqrt{3}} \left\{ \left(\frac{2x_i}{\sqrt{3}} \left(\frac{-1}{\lambda^2 \left(1 + \left(\frac{1 + 2\lambda x_i}{\sqrt{3}} \right)^2 \right)^2} \left(1 + \left[\frac{1 + 2\lambda x_i}{\sqrt{3}} \right] \left[1 + 6\lambda x_i \right] \right) \right) \right) + \left(\frac{\frac{2x_i \lambda^2}{\sqrt{3}} \frac{1}{1 + \left(\frac{1 + 2\lambda x_i}{\sqrt{3}} \right)^2} - 2\lambda \tan^{-1}\left(\frac{1 + 2\lambda x_i}{\sqrt{3}} \right)}{\lambda^4} \right) + \frac{2(a-1)\pi}{\lambda^3 6\sqrt{3}} \right\}$$

$$\frac{\partial^2 \log L}{\partial a \partial \lambda} = \sum_{i=1}^{n} \left[\frac{\left(1 + a\lambda x_i + (\lambda x_i)^2\right)x_i - (ax_i + 2\lambda x_i^2)\lambda x_i}{\left(1 + a\lambda x_i + (\lambda x_i)^2\right)^2} \right] - \sum_{i=1}^{n} \frac{1}{2\lambda^2} \left[\lambda \left(\frac{2\lambda x_i^2 + x_i}{1 + \lambda x_i + (\lambda x_i)^2} \right) - \log(1 + \lambda x_i + (\lambda x_i)^2) \right]$$

$$\frac{\partial^2 \log L}{\partial \lambda \partial a} = \sum_{i=1}^{n} \frac{\left(1 + a\lambda x_i + (\lambda x_i)^2\right)x_i - (ax_i + 2\lambda x_i^2)\lambda x_i}{\left(1 + a\lambda x_i + (\lambda x_i)^2\right)^2} - \sum_{i=1}^{n} \frac{2\lambda \left(\frac{x_i + 2\lambda x_i^2}{1 + \lambda x_i + (\lambda x_i)^2} \right) - 2\log(1 + \lambda x_i + (\lambda x_i)^2)}{4\lambda^2}$$

$$+ \sum_{i=1}^{n} \left\{ \frac{\frac{2x_i \lambda}{1 + \left(\frac{1 + 2\lambda x_i}{\sqrt{3}} \right)^2} - \sqrt{3} \tan^{-1}\left(\frac{1 + 2\lambda x_i}{\sqrt{3}} \right)}{3\lambda^2} + \frac{\pi}{6\sqrt{3}\lambda^2} \right\}$$

We can derive the $(1-\delta)100\%$ confidence intervals of the parameters a and λ by using variance matrix as in the following forms

 ISBN: 978-81-935819-2-6

$$\hat{a} \pm Z_{\frac{\delta}{2}}\sqrt{var(\hat{a})},\ \hat{\lambda} \pm Z_{\frac{\delta}{2}}\sqrt{var(\hat{\lambda})}$$

where $Z_{\frac{\delta}{2}}$ is the upper$\left(\frac{\delta}{2}\right)^{th}$ percentile of the standard normal distribution.

Numerical procedures requires for estimating parameter.

IV. Data Analysis

We considered the data sets are obtained strengths of 1.5 cm glass fibres data [6] to estimate the parameter values.

Data Set : 1

The data are the strengths of 1.5 cm glass fibres, measured at the National Physical Laboratory, England. Unfortunately, the units of measurement are not given in the paper. The data set 1 is in Table 1.

Table 1: Streangth of glass fibers

0.55, 0.93, 1.25, 1.36, 1.49, 1.52, 1.58, 1.61, 1.64, 1.68, 1.73, 1.81, 2, 0.74, 1.04, 1.27, 1.39, 1.49, 1.53, 1.59, 1.61, 1.66, 1.68, 1.76, 1.82, 2.01, 0.77, 1.11, 1.28, 1.42, 1.5, 1.54, 1.6, 1.62, 1.66, 1.69, 1.76, 1.84, 2.24, 0.81, 1.13, 1.29, 1.48, 1.5, 1.55, 1.61, 1.62, 1.66, 1.7, 1.77, 1.84, 0.84, 1.24, 1.3, 1.48, 1.51, 1.55, 1.61, 1.63, 1.67, 1.7, 1.78 and 1.89.

Hence by Non Linear method we get the estimate $\hat{a}$ = -2.7503138 and $\hat{b}$ = 0.2124807

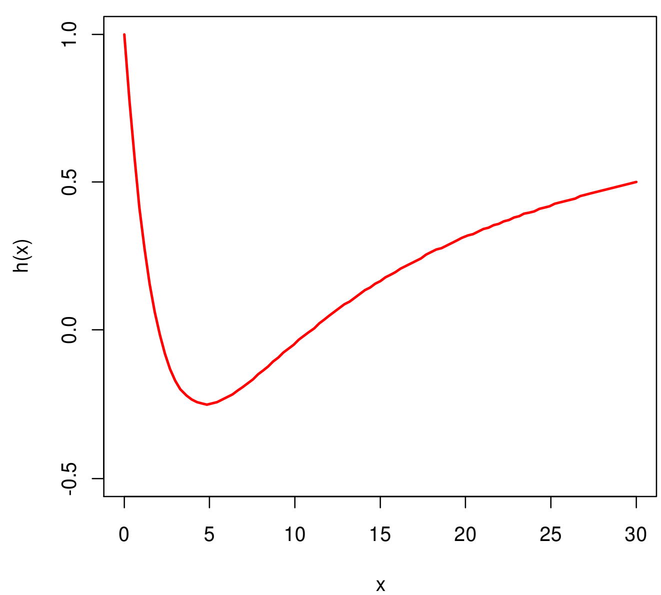


Figure 5. Failure rate function a=-2.7503138 and b=0.2124807

We obtained bathtub shaped curve for the given dataset as in figure 5.

V. Discussion

There are many distributions in reliability which exhibit Bathtub shaped failure rate model, but most of them are complicated in finding estimators or may not be suitable for given data. More over the search for distribution only with Bathtub/Upside down Bathtub shaped failure rate model resulted in the proposed distribution. Identifying the failure rate model is crucial to the maintenance and replacement policies. The optimal burn in time can be computed for the Bathtub shaped failure rate models. The model suggested here provides Bathtub shaped failure rate distributions. Any way the problem of computing Moments, characteristic functions etc still remains.

References

[1]V.M.Chacko, T. Beenu and Deepthi K S**,** A 'One Parameter' Bathtub Shaped Failure Rate Distribution, *Reliability: Theory and Applications, No.3*, Vol 12(46), 38-43, 2017
[2] Nadarajah, S., Bakouch, H.S., and R. Tahmasbi. 2011. *A generalized Lindley distribution*.Technical Report, School of Mathematics, University of Manchester, UK.
[3] Nadarajah, S., and A.K. Gupta. 2007. The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin* 59:233–234.
[4] Pal, M. M. Ali, J.Woo- Exponentiated Weibull Distribution, Statistica, anno LXVI, n.2,2006.
[5] Aarset M V, How to identify a bathtub hazard rate. *IEEE Transactions on Reliability* 36, 106-108. 1987.
[6] Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. Applied Statistics,,Vol. 36, pp. 358–369.

Construction and Selection of Single Sampling Variables Plan Through Decision Region

P.R Divya
Department of Statistics
Vimala College(Autonomous), Thrissur

Abstract

This paper proposes a new procedure for designing Variables Single Sampling Plan through Quality Decision Region (QDR) and Probabilistic Quality Region (PQR). Usually Sampling Plans are designed with levels like AQL, LQL and MAPD. A new method for designing Variable Single Sampling plan based on range of quality instead of point-wise description of quality is studied. This method can be adopted in the elementary production process where the stipulated quality level is advisable to fix at any later stages. The advantages of this new variables plan over conventional sampling plans are well discussed. Tables are constructed for the selection of parameters of this plan indexed through acceptable quality level and limiting quality level, for standard deviation is known or unknown. Further it is shown that the proposed procedure yields reduction in sample size.

Key Words: *Acceptable quality level, Limiting Quality level, Operating Characteristic Curve, Probabilistic Quality Region, Quality Decision Region, Variables Single Sampling Plan.*

I. Introduction

Acceptance sampling is a statistical tool used for making decisions on a lot of products, which should be released for the use of consumer. Variables sampling plan constitute one of the major areas on the theory and practice of acceptance sampling. Variables sampling plans involves a statistic, mean with an acceptance limit which compared with attributes plans. Whenever the quality characteristic of interest is a measurable, variables sampling plan can be applied. The main advantage of the variables sampling plan is the same operating characteristic (OC) curve can be attained with smaller sample size than required by attributes sampling plan. Generally variable sampling plan would require less sample size. Also, when destructive testing is employed, the variables sampling is particularly useful with reduced cost of inspection than any other attribute sampling plan. Another advantage is that measurement data usually provide more information about the manufacturing process than attribute measurements. Generally, measured quality characteristics are more useful than simple classification of the item as conforming or nonconforming. It is emphasized that when acceptable quality levels of a process are very small, the sample size required through attributes sampling plan is very large. Under such conditions, there will be significant advantage towards switching to variables inspection.

Already various authors have studied selection of plan parameters using different acceptable quality levels (AQL) under MIL-STD 414 scheme. Owen (1967) has developed variables sampling plans based on the normal distribution when standard deviation of the process is unknown. Hamaker (1979) has given a procedure for finding

 ISBN: 978-81-935819-2-6

the parameters with unknown sigma variable sampling plans. Bravo and Wetherill (1980) have developed a method for designing double sampling variable plans with OC curves matching with equivalent single sampling plans. Schilling (1982) has studied exclusively acceptance sampling which deals with conventional variable sampling plans. Muthuraj (1988) has given expression for finding inflection point on the OC curve for Single Sampling Variable Plans, for both standard deviation known and unknown. Baillie (1992) has developed tables for variable double sampling plans when the process standard deviation is unknown. Suresh (1993) has constructed tables for designing Single Sampling Variable Plan indexed through AQL and LQL with their relative slopes as a measure for sharpness of inspection. Further Suresh (1993) has also constructed tables for designing Single Sampling Variables Plan indexed through (p_1, K_1) and (p_2, K_2) along with the relative efficiency of variables plan over attributes plan considering filter and incentive effects. Balamurali et.al (2005) have proposed a procedure for designing variables repetitive group sampling plan through minimum average sample number. Balamurali and Jun (2007) have also studied the multiple dependent state sampling plans by variables for lot acceptance based on measurement data.

This paper proposes a new procedure for selection of variable single sampling plan. Also introduces a new concept for designing variable single sampling plan through Quality Decision Region (QDR) and Probabilistic Quality Region (PQR). Certain numerical illustrations are also provided for shop- floor application to industry.

II. Variable single sampling plan with known standard deviation

When the standard deviation σ is known, single sampling variable plan is specified with sample size n_σ and acceptability constant k_σ. A sample of size n_σ is drawn at random from the lot inspected and sentenced by the decision rule

Accept the lot if $\bar{x}+k_\sigma\sigma \le U$ .... (2.1)

Reject the lot if $\bar{x}+k_\sigma\sigma > U$

Where $\bar{x}$ the average quality characteristic derived from the sample and U, upper specification limit.

For a variable single sampling plan (n_σ, k_σ) with known standard deviation (σ) and upper specification limit (U), the proportion non-conforming is given as

$$p = 1 - F(\nu) = F(-\nu) \quad \text{......................... (2.2)}$$

where $\nu = \frac{(U-\mu)}{\sigma}$

and μ = mean of the characteristic under study, then

$$F(y) = \int_{-\infty}^{y} \left(\frac{1}{\sqrt{2\pi}}\right) e^{-\left(\frac{z^2}{2}\right)} dz, \qquad Z \sim N(0,1) \quad \text{...................... (2.2)}$$

The OC function of the plan is Pa(p)= F(w) ....(2.3)

where $w = \sqrt{n_\sigma}(\nu - k_\sigma)$ .... (2.4)

III. Variable single sampling plan with unknown standard deviation

When the standard deviation σ is unknown, single sampling variable plan is specified with its sample size n_s and acceptability constant k_s. A sample of size n_s is drawn at random from the lot inspected and sentenced by the decision rule

Accept the lot if $\bar{x} + k_s s \leq U$ ………………………… (3.1)

Reject the lot if $\bar{x} + k_s s > U$

Where $\bar{x}$ the average quality characteristic derived from the sample and U, upper specification limit.
Under the basic assumptions, the Probability of acceptance P, of a lot is given as
P = F(w) …………………………………………….. (3.2)

with $w = (\nu - k_s)\sqrt{\dfrac{n_s}{1 + \dfrac{k_s^2}{2}}}$. ………………………….. (3.3)

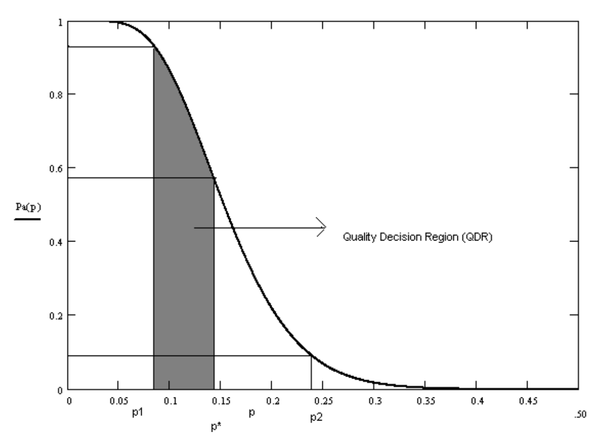
IV. Single Sampling Attribute plan with Quality Decision Region (QDR)

It is an interval of quality $(p_1 < p < p_*)$ in which product is accepted at Engineer's quality average. The quality is reliably maintained up to p_* (MAPD) and sudden decline in quality is expected. This region is called as Reliable Quality Region (RQR).
Quality Decision Region which is denoted as $d_1 = (p_* - p_1)$ …………… (4.1)
is derived from probability of acceptance

$P_a(p_1 < p < p_*) = \sum_{r=0}^{c} \frac{e^{-np}(np)^r}{r!}$ $for\ p_1 < p < p_*$ ………………… (4.2)

where $p < 0.1$ and the number of defects assumed to follow Poisson distribution for the quality characterisitic.

Figure 1 : OC Curve of the SSP (50,7) with Quality Decision Region (d$_1$)



From Figure 1 d_1 represent the Quality Decision Region (QDR). A general result is that $p_1 < p_* < p_2$ for any Single Sampling Attribute plan. When p_* approaching p_1, then np_* also approaching np_1.Then d_1 is small. Therefore straighter OC can be performed. When p_* approaching p_2 then np_* approaching np_2. That is when p_* approaching p_2, at lower quality level acceptance is more. After MAPD, at lower quality level if acceptance is more, then the product is not at all useful. Therefore When np_* approaching np_2, OC will be relatively less discriminating about the quality. Thus Quality Decision Region (QDR), d_1 is a good measure for defining the quality of a product.

4. 1. Variables single sampling plan with Quality Decision Region (QDR) - case of known standard deviation:

It is an interval of quality $(p_1 < p < p_*)$ in which product is accepted at Engineer's quality average. The quality is reliably maintained up to p_* (MAPD) and sudden decline in quality is expected.

Quality Decision Range is denoted as $d_{1\sigma} = n_\sigma(p_* - p_1)$ ………………………… (4.1.1)

is derived from probability of acceptance

$$P_a(p_1 < p < p_*) = F(w), \text{ with } \quad w = \sqrt{n_\sigma}(\nu - k_\sigma) \text{ ……………..for } \quad p_1 < p < p_*$$

Where p is the fraction non- conforming in a lot.

Values of the parameters (n_σ, k_σ) and QDR $(d_{1\sigma})$ when sigma is known, are tabulated in Table1.This is done for several combinations of (AQL, LQL).

Suppose for example that AQL=0.0010, LQL=0.0050 and MAPD=0.00194. Then from Table 1, using equation (4.1.1) the Quality Decision Region $QDR(d_{1\sigma}) = 0.0310$. Thus the Parameters of known sigma variables single sampling plan can be determined as $n_\sigma = 33, k_\sigma = 2.8000, QDR(d_{1\sigma}) = 0.0310.$

4. 2. Variables single sampling plan with Quality Decision Region (QDR) - case of unknown standard deviation:

It is an interval of quality $(p_1 < p < p_*)$ in which product is accepted at Engineer's quality average. The quality is reliably maintained up to p_* (MAPD) and sudden decline in quality is expected.

Quality Decision Range is denoted as $d_{1s} = n_s(p_* - p_1)$ ………………….. (4.2.1)

is derived from probability of acceptance

$$P_a(p_1 < p < p_*) = F(w), \text{ with } \quad w = (\nu - k_s)\sqrt{\frac{n_s}{1 + \frac{k_s^2}{2}}} \text{ ……………..for } \quad p_1 < p < p_*$$

Where p is the fraction non- conforming in a lot.

Values of the parameters (n_s, k_s) and QDR (d_{1s}) when sigma is unknown, are tabulated in Table1.This is done for several combinations of (AQL, LQL).

Suppose for example that AQL=0.0010, LQL=0.0050 and MAPD=0.00194. Then using Table 1, with equation (4.2.1) the Quality Decision Region $QDR(d_{1s}) = 0.1504$. Thus the Parameters of unknown sigma variables single sampling plan can be determined as

ISBN: 978-81-935819-2-6

$n_s = 160, k_s = 2.8000, QDR(d_{1s}) = 0.1504.$

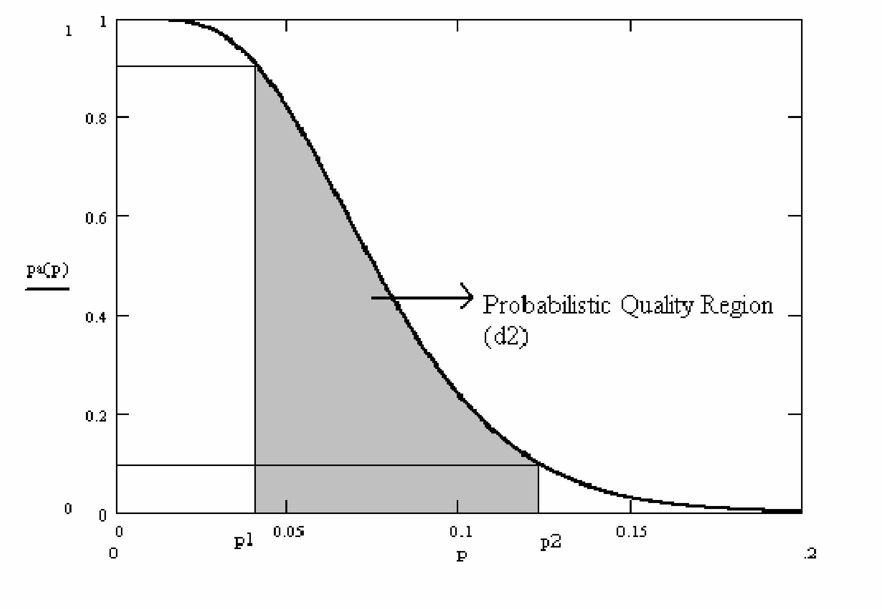
V. Single Sampling Attribute Plan with Probabilistic Quality Region (PQR)

It is an interval of quality $(p_1 < p < p_2)$ in which product is accepted with a minimum probability 0.10 and maximum probability 0.95. Probabilistic Quality Region denoted by $d_2 = (p_2 - p_1)$ is derived from probability of acceptance

$$P_a(p_1 < p < p_2) = \sum_{r=0}^{c} \frac{e^{-np}(np)^r}{r!} \; for \; p_1 < p < p_2 \; \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.1)$$

Where $p < 0.1$ and the number of defects assumed to follow Poisson distribution.
Figure 2 represents the OC curve indicating the Probabilistic Quality Region (PQR). From Figure 2 d_2 represent the Probabilistic Quality Region (PQR).

Figure 2: OC Curve of the SSP (75,5) with Probabilistic Quality Region (d$_2$)



5.1. Variable single sampling plan with Probabilistic Quality Region (PQR) - case of known standard deviation:

It is an interval of quality ($p_1 < p < p_2$) in which product is accepted with a minimum probability 0.10 and maximum probability 0.95.
Probabilistic Quality Region denoted by $d_{2\sigma} = n_\sigma(p_2 - p_1)$ ……………………….. (5.1.1)
is derived from probability of acceptance
$P_a(p_1 < p < p_2) = F(w), \text{with}\; w = \sqrt{n_\sigma}(\nu - k_\sigma)$ for $p_1 < p < p_2$

 ISBN: 978-81-935819-2-6

Where p is the fraction non- conforming in a lot. Values of the parameters (n_σ, k_σ) and PQR $(d_{2\sigma})$ when sigma is known, are tabulated in Table1.This is done for several combinations of (AQL, LQL). Suppose for example that AQL=0.0010, LQL=0.0050 and MAPD=0.00194. Then from Table 1, using equation (5.1.1) the Probabilistic Quality Region, $PQR(d_{2\sigma}) = 0.1320.$ Thus the Parameters of known sigma variables single sampling plan can be determined as $n_\sigma = 33, k_\sigma = 2.8000, PQR(d_{2\sigma}) = 0.1320.$

5.2. Variable single sampling plan with Probabilistic Quality Region (PQR) - case of unknown standard deviation:

It is an interval of quality ($p_1 < p < p_2$) in which product is accepted with minimum probability 0.10 and maximum probability 0.95.

Probabilistic Quality Region denoted as $d_{2s} = n_s(p_2 - p_1)$ ………………… (5.2.1)

is derived from probability of acceptance

$$P_a(p_1 < p < p_2) = F(w), \text{with } w = (v - k_s)\sqrt{\frac{n_s}{1 + \frac{k_s^2}{2}}} \text{ for } p_1 < p < p_2$$

Where p is the fraction non- conforming in a lot.

Values of the parameters (n_s, k_s) and PQR (d_{2s}) when sigma is unknown, are tabulated in Table1.This is done for several combinations of (AQL, LQL). Suppose for example that AQL=0.0010, LQL=0.0050 and MAPD=0.00194. Then from Table 1, using equation (5.2.1) the Probabilistic Quality Region $PQR(d_{2s}) = 0.640.$ Thus the Parameters of unknown sigma variables single sampling plan can be determined as $n_s = 160, k_s = 2.8000, PQR(d_{2s}) = 0.640.$

VI. Comparison of sampling plans

It is observed that, variables single sampling plan yield a reduction in sample size than the attribute single sampling plan at the specified AQL and LQL. It indicates that the variables single sampling plan can achieve the same operating characteristics with smaller sample size as compared with attribute single sampling plan. It is also observed from Table -2, that the smaller sample size are required for variables single sampling plan as compared with attribute single sampling plan. Therefore variables single sampling plan is economically superior to attribute single sampling plan.

It can also be seen that the variables single sampling plan with smaller value of PQR, that is for smaller values of d_2 or $(d_{2\sigma})$ seems to be closer to the ideal OC curve in that the probability of acceptance at AQL or smaller is increased. From Table 3, it can be seen that when Probabilistic Quality Region (PQR) is decreasing, Maximum Allowable Average Outgoing Quality (MAAOQ) and Average Outgoing Quality Limit (AOQL) is increasing. Therefore PQR method possesses wide potential applicability in industry insuring a higher standard of quality attainment. If the Probabilistic Quality Region is small, straighter OC can be performed. Therefore, it can be seen that the sampling plan

 ISBN: 978-81-935819-2-6

with a smaller value of PQR (d_2) seems to be closer to the ideal OC curve. Compared with attribute single sampling plan, the variables single sampling plan seems to be closer to the ideal OC curve.

Suppose for example that AQL=0.6640, LQL=1.4760. Using Table 2, the parameters of the Attribute single sampling plan can be determined as $n = 5601, c = 13, PQR(d_2) = 1056.412$.

The parameters of known sigma variables single sampling plan can be determined as $n_\sigma = 95, k_\sigma = 2.3078, PQR(d_{2\sigma}) = 77.140$.

The parameters of unknown sigma variables single sampling plan can be determined as $n_s = 350, k_s = 2.3095, PQR(d_{2s}) = 284.200$. Certain parametric values for variable single sampling plan are carried out in Table 5.

VII. Efficiency comparison using Probabilistic Quality Region

The relative efficiencies are given by the ratio of Probabilistic Quality Region (PQR) of the sampling plans viz., $\frac{d_2}{d_{2\sigma}}$, $\frac{d_2}{d_{2s}}$, $\frac{d_{2s}}{d_{2\sigma}}$, $\frac{d_{2R}}{d_{2s}}$, have been computed and given in Table 4. From the computed values one can notice that the ratios are decreasing as AQL and LQL increases.

VIII. Conclusion

The purpose of this paper is to design a variable sampling plan through Quality Decision Region (QDR) and Probabilistic Quality Region (PQR) for accepting lots of products whose quality characteristic follows a normal distribution. The proposed variables single sampling plan through Probabilistic Quality Region (PQR) provides better protection than conventional attribute single sampling plans with smaller sample size. This plan is based on a simple statistic and it is easier to apply than double and multiple sampling plans. It is observed that the sampling plan with smaller value of PQR (d_2) seems to be closer to the ideal shape of OC curve. So compared to attribute single sampling plan through PQR, the variable single sampling plan through PQR seems to be closer to the ideal shape of OC curve. It is seen that the variable single sampling plan through Probabilistic Quality region (PQR) is economically superior to attribute single sampling plans. The Scope of this paper is purely based on a new methodology of designing sampling plan based on range of quality instead of point- wise description of quality. So this method can be adopted in the elementary production process where the stipulated quality level is advisable to fix at a later stage. The Quality Decision Region (QDR) and Probabilistic Quality Region (PQR) idea has been experimented on designing of Single Sampling Variables Plan which results with minimization of sample size. Quality Region concept possesses wider potential applicability in industry ensuring higher standard of quality attainment. The Quality Decision Region 'd_1' is a good measure for defining quality, since the Quality Decision Region'd_1', depends upon the inflection point. MAPD is a key measure for assessing the degree of sharpness of inspection with inflection point which empowers the OC curve to discriminate between good and bad lots. Further successful matching based on MAPD has also been accomplished using the abscissa of the inflection point and tangent at that abscissa. Also translation of the original OC curve from an acceptance to a surveillance criterion can be accomplished through a shift in p_* . This mode of relaxing p_* has been found to be of convenience to the technician or engineer who are unfamiliar with basic statistical theory.

REFERENCES

1. **Baillie, D.H., 1992.** Double sampling plans for inspection by variables when the process standard deviation is unknown. International Journal of Quality and Reliabilility Management 9, 59-70.
2. **Balamurali, S., Park,H., Jun,C.H., Kim,K.-J.,Lee,J.,2005.** Designing of variables repetitive group sampling plan involving minimum average sample number. Communications in Statistics- Simulation and computation 34(3), 799-809.
3. **Balamurali, S., Jun,C.H., 2007.** Multiple dependent state sampling plans for lot acceptance based on measurement data. European Journal of Operational Research 180, 1221-1230.
4. **Bravo,P.C., Wetherill, G.B.,1980.** The matching of sampling plans and the design of double sampling plans. Journal of the Royal Statistical Society A 143, 49-67.
5. **Hamaker, H.C., 1979.** Acceptance sampling for percent defective by variables and by attributes. Journal of Quality Technology, Vol.11, No.3, 2(2), 139-148.
6. **Muthuraj .D., 1988.** "Contributions to the study of Sampling Inspection Plans with special reference to point of control and inflection point of the OC curve", *Ph.D Thesis*, Bharathiar University, Coimbatore, India.
7. **Owen, D.**B., **1967**. Variables sampling plans based on the normal distribution. Technometrics 9, 417-423.
8. **Schilling, E.G., 1982.** Acceptance Sampling in Quality Control. Marcel Dekker, New York.
9. **Suresh, K.K., 1993.** "A study on Acceptance Sampling using Acceptable Quality Levels", *Ph.D Thesis*, Bharathiar University, Coimbatore, Tamilnadu, India.

 ISBN: 978-81-935819-2-6

Table 1: Variables Single Sampling Plans with Quality Decision Region (QDR) and Probabilistic Quality Region (PQR) indexed by AQL and LQL

			Known sigma				unknown sigma			
AQL	LQL	MAPD	n_σ	k_σ	QDR $(d_{1\sigma})$	PQR $(d_{2\sigma})$	n_s	k_s	QDR (d_{1s})	PQR (d_{2s})
0.0010	0.0020	0.00141	191	2.9700	0.0783	0.1910	1032	2.970	0.4231	1.032
0.0010	0.0025	0.00155	107	2.9300	0.0589	0.1605	567	2.930	0.3119	0.851
0.0010	0.0030	0.00164	74	2.9000	0.0474	0.1480	381	2.900	0.2438	0.762
0.0010	0.0040	0.00184	45	2.8400	0.0378	0.1350	226	2.840	0.1898	0.678
0.0010	0.0050	0.00194	33	2.8000	0.0310	0.1320	160	2.800	0.1504	0.640
0.0010	0.0060	0.00198	26	2.7700	0.0255	0.1300	124	2.770	0.1215	0.620
0.0010	0.0070	0.00211	22	2.7300	0.0244	0.1320	102	2.730	0.1132	0.612
0.0010	0.0080	0.00211	19	2.7100	0.0211	0.1330	87	2.710	0.0966	0.609
0.0010	0.0090	0.00220	17	2.6800	0.0204	0.1360	76	2.680	0.0912	0.608
0.0010	0.0100	0.00218	15	2.6600	0.0177	0.1350	67	2.660	0.0791	0.603
0.0010	0.0120	0.00226	13	2.6200	0.0164	0.1430	55	2.620	0.0693	0.605
0.0010	0.0150	0.00235	11	2.5700	0.0149	0.1540	44	2.570	0.0594	0.616
0.0025	0.0050	0.00350	161	2.6800	0.1610	0.4025	736	2.680	0.7360	1.840
0.0025	0.0060	0.00382	99	2.6400	0.1307	0.3465	443	2.640	0.5848	1.551
0.0025	0.0075	0.00411	62	2.6000	0.0998	0.3100	267	2.600	0.4299	1.335
0.0025	0.0100	0.00454	38	2.5400	0.0775	0.2850	157	2.540	0.3203	1.178
0.0025	0.0120	0.00480	29	2.5000	0.0667	0.2755	117	2.500	0.2691	1.112
0.0025	0.0150	0.00513	22	2.4500	0.0579	0.2750	85	2.450	0.2236	1.063
0.0025	0.0200	0.00556	16	2.3800	0.0490	0.2800	59	2.380	0.1805	1.033
0.0025	0.0250	0.00551	12	2.3300	0.0361	0.2700	45	2.330	0.1355	1.013
0.0025	0.0300	0.00547	10	2.2900	0.0297	0.2750	37	2.290	0.1099	1.018
0.0025	0.0350	0.00568	9	2.2500	0.0286	0.2925	31	2.250	0.0986	1.008
0.0050	0.0100	0.00698	138	2.4400	0.2732	0.6900	547	2.440	1.0831	2.735
0.0050	0.0120	0.00757	85	2.4000	0.2185	0.5950	327	2.400	0.8404	2.289
0.0050	0.0150	0.00830	53	2.3500	0.1749	0.5300	196	2.350	0.6468	1.960
0.0050	0.0200	0.00929	32	2.2800	0.1373	0.4800	114	2.280	0.4891	1.710
0.0050	0.0250	0.00986	23	2.2300	0.1118	0.4600	79	2.230	0.3839	1.580
0.0050	0.0300	0.01020	18	2.1900	0.0936	0.4500	61	2.190	0.3172	1.525
0.0050	0.0350	0.01062	15	2.1500	0.0843	0.4500	49	2.150	0.2754	1.470
0.0050	0.0400	0.01113	13	2.1100	0.0797	0.4550	41	2.110	0.2513	1.435
0.0100	0.0200	0.01430	116	2.1700	0.4988	1.1600	388	2.170	1.6684	3.880
0.0100	0.0250	0.01563	64	2.1200	0.3603	0.9600	208	2.120	1.1710	3.120
0.0100	0.0300	0.01665	44	2.0800	0.2926	0.8800	137	2.080	0.9111	2.740
0.0100	0.0350	0.01770	33	2.0400	0.2541	0.8250	100	2.040	0.7700	2.500
0.0100	0.0400	0.01876	26	2.0000	0.2278	0.7800	78	2.000	0.6833	2.340
0.0100	0.0450	0.01951	22	1.9700	0.2092	0.7700	64	1.970	0.6086	2.240
0.0100	0.0500	0.02029	19	1.9400	0.1955	0.7600	54	1.940	0.5557	2.160
0.0100	0.0600	0.02143	15	1.8900	0.1715	0.7500	41	1.890	0.4686	2.050
0.0200	0.0400	0.02870	94	1.8800	0.8178	1.8800	259	1.880	2.2533	5.180
0.0200	0.0450	0.03018	67	1.8500	0.6821	1.6750	182	1.850	1.8528	4.550
0.0200	0.0500	0.03174	52	1.8200	0.6105	1.5600	137	1.820	1.6084	4.110
0.0200	0.0600	0.03422	35	1.7700	0.4977	1.4000	89	1.770	1.2656	3.560
0.0200	0.0700	0.03599	26	1.7300	0.4157	1.3000	64	1.730	1.0234	3.200
0.0200	0.0800	0.03799	21	1.6900	0.3778	1.2600	50	1.690	0.8995	3.000
0.0200	0.0900	0.03979	17	1.6500	0.3364	1.1900	40	1.650	0.7916	2.800
0.0200	0.1000	0.04130	15	1.6200	0.3195	1.2000	34	1.620	0.7242	2.720

ISBN: 978-81-935819-2-6

Table 2: Variable Plans by $\sigma-$ method, $s-$ method, and $R-$ method matched with Attribute plans

AQL%	LQL%	Attributes plan			$\sigma-$ method			$s-$ method			$R-$ method		
		n	c	PQR (d_2)	n_σ	k_σ	PQR $(d_{2\sigma})$	n_s	k_s	PQR (d_{2s})	n_R	k_R	PQR (d_{2R})
0.1680	0.3650	5601	14	**1103.397**	138	2.7920	**27.186**	677	2.7930	**133.369**	880	1.2004	**173.360**
0.3320	0.7400	2601	13	**1061.208**	112	2.5586	**45.696**	478	2.5599	**195.024**	616	1.1000	**251.328**
0.6640	1.4760	1301	13	**1056.412**	95	2.3078	**77.140**	350	2.3095	**284.200**	446	0.9922	**362.152**
1.3270	2.9380	651	13	**1048.761**	79	2.0338	**127.269**	245	2.0358	**394.695**	306	0.8744	**492.966**
1.9870	4.3820	435	13	**1041.825**	71	1.8606	**170.045**	194	1.8630	**464.630**	239	0.7999	**572.405**
2.6540	5.3290	326	13	**872.050**	64	1.7292	**171.200**	161	1.7319	**430.675**	197	0.7434	**526.975**
3.3190	7.2580	261	13	**1028.079**	60	1.6229	**236.340**	139	1.6258	**547.521**	168	0.6977	**661.752**
3.9240	8.3150	201	12	**882.591**	52	1.5306	**228.332**	113	1.5541	**496.183**	135	0.6581	**592.785**
4.5650	10.2110	173	12	**976.758**	49	1.4531	**276.654**	101	1.4567	**570.246**	120	0.6248	**677.520**
5.2380	11.6620	151	12	**970.024**	46	1.3805	**295.504**	91	1.3843	**584.584**	107	0.5935	**687.368**
5.8660	13.0070	135	12	**964.035**	44	1.3188	**314.204**	84	1.3228	**599.844**	97	0.5670	**692.677**
6.5540	14.4680	121	12	**957.594**	42	1.2568	**332.388**	77	1.2609	**609.378**	88	0.5403	**696.432**

Table 3: Variable Single Sampling Plan indexed by AOQL and LQL

AOQL%	AQL%	Attributes plan				Matched Variables Plan							
		n	c	MAAOQ	PQR (d_2)	n_σ	k_σ	MAAOQ	PQR $(d_{2\sigma})$	n_s	k_s	MAAOQ	PQR (d_{2s})
0.1000	0.050	1342	2	0.0009	**67.10**	20	2.8799	0.0009	**1.00**	105	2.8868	0.0009	**5.250**
0.1000	0.025	841	1	0.0007	**79.90**	12	2.8763	0.0007	**1.14**	61	2.8883	0.0007	**5.795**
0.1000	0.075	2544	4	0.0008	**165.36**	11	2.8261	0.0008	**0.72**	61	2.8381	0.0008	**3.965**
0.5000	0.075	169	1	0.0030	**14.37**	9	2.3247	0.0036	**0.77**	32	2.3440	0.0033	**2.720**
0.5000	0.100	169	1	0.0030	**13.52**	9	2.3249	0.0036	**0.72**	33	2.3435	0.0033	**2.640**
0.5000	0.250	275	2	0.0040	**13.75**	15	2.3287	0.0047	**0.75**	56	2.3394	0.0045	**2.800**
1.0000	0.250	85	1	0.0070	**5.10**	8	2.0570	0.0075	**0.48**	25	2.0791	0.0069	**1.500**

Table 4: The Relative Efficiencies of Probabilistic Quality Region (PQR)

AQL%	LQL%	Attributes Plan	$\sigma-$ method	$s-$ method	$R-$ method	Relative Efficiency			
		(d_2)	$(d_{2\sigma})$	(d_{2s})	(d_{2R})	$\frac{d_2}{d_{2\sigma}}$	$\frac{d_2}{d_{2s}}$	$\frac{d_{2s}}{d_{2\sigma}}$	$\frac{d_{2R}}{d_{2s}}$
0.1680	0.365	1103.397	27.186	133.369	173.360	40.5870	8.2733	4.9058	1.2999
0.3320	0.740	1061.208	45.696	195.024	251.328	23.2232	5.4414	4.2679	1.2887
0.6640	1.476	1056.412	77.140	284.200	362.152	13.6947	3.7171	3.6842	1.2743
1.3270	2.938	1048.761	127.269	394.695	492.966	8.2405	2.6571	3.1013	1.2490
1.9870	4.382	1041.825	170.045	464.630	572.405	6.1268	2.2423	2.7324	1.2320
2.6540	5.329	872.050	171.200	430.675	526.975	5.0938	2.0248	2.5156	1.2236
3.3190	7.258	1028.079	236.340	547.521	661.752	4.3500	1.8777	2.3167	1.2086
3.9240	8.315	882.591	228.332	496.183	592.785	3.8654	1.7788	2.1731	1.1947
4.5650	10.211	976.758	276.654	570.246	677.520	3.5306	1.7129	2.0612	1.1881
5.2380	11.662	970.024	295.504	584.584	687.368	3.2826	1.6593	1.9783	1.1758
5.8660	13.007	964.035	314.204	599.844	692.677	3.0682	1.6071	1.9091	1.1548
6.5540	14.468	957.594	332.388	609.378	696.432	2.8810	1.5714	1.8333	1.1429

 ISBN: 978-81-935819-2-6

Table 5: Certain Parametric Values for Variable Single Sampling Plan

		known sigma				
AOQL%	**AQL%**	n_σ	k_σ	**MAAOQ**	$R_1 = \frac{MAAOQ}{MAPD}$	$R_2 = \frac{AOQL}{MAPD}$
0.10	**0.08**	11	2.8261	**0.0008**	0.8257	1.0643
0.25	**0.05**	10	2.5746	**0.0017**	0.8172	1.1827
0.50	**0.10**	9	2.3249	**0.0036**	0.8084	1.1224
0.75	**0.25**	8	2.1727	**0.0053**	0.8100	1.1516
1.00	**0.50**	13	2.0606	**0.0094**	0.7321	0.7814
1.50	**0.50**	11	1.8873	**0.0110**	0.7343	0.7917
2.00	**1.00**	11	1.7586	**0.0191**	0.7201	0.7539
2.50	**1.50**	10	1.6528	**0.0238**	0.7193	0.7542
3.00	**2.00**	13	1.5775	**0.0298**	0.6822	0.6860
3.50	**2.50**	16	1.5159	**0.0348**	0.6570	0.6611
4.00	**3.00**	15	1.4484	**0.0396**	0.6557	0.6628
4.50	**3.50**	15	1.3873	**0.0445**	0.6494	0.6561
5.00	**4.00**	17	1.3117	**0.0487**	0.6323	0.6530
5.50	**4.50**	16	1.2892	**0.0536**	0.6345	0.6506
6.00	**5.00**	19	1.2456	**0.0574**	0.6185	0.6363
6.50	**5.50**	18	1.2060	**0.0623**	0.6183	0.6448
7.00	**6.00**	18	1.1665	**0.0666**	0.6145	0.6458
7.50	**6.50**	17	1.1199	**0.0718**	0.6136	0.6408
8.00	**7.00**	19	1.0913	**0.0753**	0.6042	0.6417
8.50	**7.50**	19	1.0584	**0.0793**	0.6011	0.6442
9.00	**8.00**	18	1.0197	**0.0842**	0.6004	0.6422
9.50	**8.50**	18	0.9808	**0.0892**	0.5967	0.6354
10.00	**9.00**	20	0.9661	**0.0912**	0.5899	0.6469

unknown sigma				
n_s	k_s	**MAAOQ**	$R_1 = \frac{MAAOQ}{MAPD}$	$R_2 = \frac{AOQL}{MAPD}$
61	2.8381	**0.0008**	0.8127	1.0093
43	2.5900	**0.0016**	0.8204	1.2645
33	2.3435	**0.0033**	0.8134	1.2190
29	2.1922	**0.0052**	0.8026	1.1534
41	2.0736	**0.0091**	0.7332	0.8099
21	1.9116	**0.0108**	0.7912	1.1039
28	1.7752	**0.0183**	0.7234	0.7909
25	1.6705	**0.0232**	0.7163	0.7731
30	1.5912	**0.0291**	0.6819	0.7041
35	1.5271	**0.0340**	0.6572	0.6759
25	1.4485	**0.0388**	0.6743	0.6980
23	1.3912	**0.0436**	0.6718	0.7022
34	1.3548	**0.0479**	0.6334	0.6620
31	1.3000	**0.0527**	0.6320	0.6591
35	1.2639	**0.0565**	0.6187	0.6567
32	1.2157	**0.0614**	0.6178	0.6546
31	1.1764	**0.0621**	0.6144	0.6561
29	1.1302	**0.0707**	0.6121	0.6495
31	1.1004	**0.0743**	0.6041	0.6508
30	1.0675	**0.0782**	0.6017	0.6539
28	1.0291	**0.0830**	0.6004	0.6514
27	0.9903	**0.0880**	0.5973	0.6450
30	0.9746	**0.0901**	0.5899	0.6550

On Some Tailed Distributions And Related Time Series Models

Mariamma Antony
Department of Statistics, Little Flower College
Guruvayoor, Kerala
Email: mariammaantony@rediffmail.com

Abstract

We encounter tailed distributions in life testing experiments where an item fails instantaneously. In clinical trials, sometimes a medicine has no response initially and later there may be response, the length of response is described by certain probability distribution. In this paper, we study tailed generalized geometric Linnik distribution. Relation between stable laws and tailed geometric Linnik distribution is established. Time series models with tailed generalized geometric Linnik marginals are developed. The stationarity of the processes is established. The model is extended to p^{th} order. Tailed generalized geometric asymmetric Linnik distribution is introduced and studied.

Keywords: Laplace distribution, Linnik distribution, Stable Laws, Stationarity, Tailed distributions, Time Reversibility.

I. Introduction

Tailed distributions have found applications in various fields and were studied by many authors (see, Littlejohn (1994), Muraleedharan (1999), Muraleedharan and Kale (2002)).We encounter tailed distributions in life testing experiments where an item fails instantaneously. In clinical trials, some times a medicine has no response initially with a certain probability and on a later stage there may be response, the length of the response is described by certain probability distribution.

DEFINITION 1.1

Let the random variable X has distribution function F(x) and characteristic function $\phi_X(t)$. A tailed random variable U with tail probability θ associated with X is defined by the characteristic function

$$\phi_U(t) = \theta + (1-\theta)\phi_X(t) \qquad (1.1$$

Linnik (1963) proved that the function

$$\phi(t) = \frac{1}{1+\lambda|t|^{\alpha}}, 0<\alpha\le 2, \lambda>0 \qquad (1.2)$$

is the characteristic function of a probability distribution. The distribution corresponding to the characteristic function (1.2) is called Linnik (or α- Laplace) distribution. A random variable X with characteristic function ϕ in (1.2) is denoted by $X \stackrel{d}{=} L(\alpha,\lambda)$. Note that the $L(\alpha,\lambda)$ distributions are symmetric and for $\alpha = 2$, it becomes the classical symmetric Laplace distribution. Pillai (1985) introduced a generalization of the Linnik distribution with characteristic function (1.2), namely semi α-Laplace distribution. A random variable X on R has semi α-Laplace distribution if its characteristic function $\phi(t)$ is of the form

$$\phi(t) = \frac{1}{1+|t|^{\alpha}\delta(t)} \qquad (1.3)$$

 ISBN: 978-81-935819-2-6

where $\delta(t)$ satisfies the functional equation

$$\delta(t) = \delta\left(p^{1/\alpha}t\right), 0<p<1, 0<\alpha\le 2. \quad (1.4)$$

DEFINITION 1.2

A random variable X on R is said to have geometric Linnik distribution and write $X \underline{\underline{d}}\, GL(\alpha,\lambda)$ if its characteristic function $\phi(t)$ is

$$\phi(t) = \frac{1}{1+\ln(1+\lambda|t|^{\alpha})}, t\in R, 0<\alpha\le 2, \lambda>0. \quad (1.5)$$

DEFINITION 1.3

A random variable X on R has the generalized Linnik distribution and write $X \underline{\underline{d}}\, GeL(\alpha,\lambda,p)$ if it has the characteristic function

$$\phi(t) = \frac{1}{(1+\lambda|t|^{\alpha})^{p}}, \ p>0, \lambda>0, 0<\alpha\le 2. \quad (1.6)$$

DEFINITION 1.4

A random variable X on R is said to have type I generalized geometric Linnik distribution and write $X \underline{\underline{d}}\, GeGL_1(\alpha,\lambda,p)$ if it has the characteristic function

$$\phi(t) = \frac{1}{1+p\ln(1+\lambda|t|^{\alpha})}, \ 0<\alpha\le 2, \ p>0, \ \lambda>0. \quad (1.7)$$

In Section 2, we introduce tailed distributions associated with type I generalized geometric Linnik distribution and study its properties. Tailed Type II generalized geometric Linnik distribution is also discussed in this Section. Tailed type I generalized geometric asymmetric Linnik distribution is studied in Section 3. In Section 4, a first order autoregressive model with tailed Type I generalized geometric Linnik distribution as marginal is developed.

II. Tailed Generalized Geometric Linnik Distributions

Now we introduce tailed Type I generalized geometric Linnik distribution and obtain a representation of the same.

In (1.1), when $\phi_X(t) = \dfrac{1}{1+\tau\ln\left(1+\lambda|t|^{\alpha}\right)}$, then

$$\phi_U(t) = \frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}\right)}{1+\tau\ln\left(1+\lambda|t|^{\alpha}\right)}. \quad (2.1)$$

The random variable U with characteristic function (2.1) is called tailed Type I generalized geometric Linnik and denoted by $TGeGL_1(\alpha,\lambda,\tau,\theta)$

THEOREM 2.1

Let X and Y be independent random variables such that X has tailed generalized geometric exponential distribution with Laplace transform $\theta+(1-\theta)\dfrac{1}{1+\tau\ln(1+\delta)}$ and Y is stable with characteristic function $e^{-\lambda|t|^{\alpha}}$, $0<\alpha\le 2, \lambda,\delta,\tau>0, 0<\theta<1$. Then $Z = X^{1/\alpha}Y$ has $TGeGL_1(\alpha,\lambda,\tau,\theta)$ distribution.

PROOF

$$\phi_Z(t) = E\left[e^{itX^{1/\alpha}Y}\right]$$

$$= \int_0^\infty \phi_Y\left(tx^{1/\alpha}\right) dF(x)$$

$$= \int_0^\infty e^{-\lambda|t|^\alpha x} dF(x)$$

$$= \theta + (1-\theta)\frac{1}{1+\tau\ln\left(1+\lambda|t|^\alpha\right)}$$

$$= \frac{1+\tau\theta\ln\left(1+\lambda|t|^\alpha\right)}{1+\tau\ln\left(1+\lambda|t|^\alpha\right)}.$$

This completes the proof.

Now we consider the type II generalized geometric Linnik distribution and study the tailed distribution generated by it.

DEFINITION 2.1

A random variable X is said to have tailed type II generalized geometric Linnik distribution and write $X \underline{\underline{d}}\, TGeGL_2(\alpha,\lambda,\tau,\theta)$ distribution if it has thecharacteristic function

$$\phi_X(t) = \frac{\theta\left[1+\ln\left(1+\lambda|t|^\alpha\right)\right]^\tau + (1-\theta)}{\left[1+\ln\left(1+\lambda|t|^\alpha\right)\right]^\tau}, 0<\alpha\le 2, 0<\theta<1, \lambda,\tau>0.$$

The tailed type II generalized geometric Linnik distribution being the tailed form of type II generalized geometric Linnik is infinitely divisible.

As in the case of $TGeGL_1$ distribution, we now obtain a representation of $TGeGL_2$ random variables in terms of tailed geometric gamma and stable random variables.

DEFINITION 2.2

A random variable X is said to have tailed geometric gamma distribution if it has Laplace transform

$$\phi_1(\delta) = \theta + (1-\theta)\frac{1}{\left[1+\ln(1+\delta)\right]^\tau}, \quad \delta,\tau>0, 0<\theta<1.$$

THEOREM 2.2

Let X and Y be independent random variables such that X has Laplace transform $\theta + (1-\theta)\frac{1}{[1+\ln(1+\delta)]^\tau}$ and Y is stable with characteristic function $e^{-\lambda|t|^\alpha}, 0<\alpha\le 2.$

Then $U = X^{1/\alpha}Y$ has distribution $TGeGL_2(\alpha,\lambda,\tau,\theta)$.

Proof follows analogous to the proof of Theorem 2.1.

 ISBN: 978-81-935819-2-6

III. Tailed Generalized Geometric Asymmetric Linnik Distribution

Here we discuss the tailed distributions generated by generalized geometric asymmetric Linnik distributions.

DEFINITION 3.1

A random variable X is said to have tailed type I generalized geometric asymmetric Linnik distribution and write $U \underline{\underline{d}} TGeGAL_1(\alpha,\lambda,\mu,\tau,\theta)$ if it has characteristic function

$$\phi_U(t) = \frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}-i\mu t\right)}{1+\tau\ln\left(1+\lambda|t|^{\alpha}-i\mu t\right)}.$$

The type II generalized geometric asymmetric Linnik distribution can be defined using (1.1) with $\phi_X(t)$ replaced by

$$\frac{1}{\left[1+\ln\left(1+\lambda|t|^{\alpha}-i\mu t\right)\right]^{\tau}}.$$

That is, a random variable U having tailed type II generalized geometric asymmetric Linnik distribution denoted by $TGeGAL_2(\alpha,\lambda,\mu,\tau,\theta)$ has characteristic function

$$\theta+(1-\theta)\frac{1}{\left[1+\ln\left(1+\lambda|t|^{\alpha}-i\mu t\right)\right]^{\tau}}.$$

IV. Time Series Models With Tailed Generalized Geometric Linnik Distribution As Marginals

Now we develop a first order autoregressive model with $TGeGL_1$ distribution as marginal. Consider the model

$$X_n = \begin{cases} \varepsilon_n & w.p. \quad p \\ X_{n-1}+\varepsilon_n & w.p. \quad 1-p \end{cases} \qquad (4.1)$$

$$= I_n X_{n-1} + \varepsilon_n$$

where $\{\varepsilon_n\}$ and $\{I_n\}$ are two sequences of independent and identically distributed random variables with I_n, X_{n-1} and ε_n mutually independent such that

$$P(I_n=0)=p=1-P(I_n=1).$$

We have the model (4.1) in terms of characteristic functions is

$$\phi_{\varepsilon_n}(t) = \frac{\phi_{X_n}(t)}{p+(1-p)\phi_{X_{n-1}}(t)}.$$

In the stationary case,

$$\phi_{\varepsilon_n}(t) = \frac{\phi_X(t)}{p+(1-p)\phi_X(t)}.$$

If X has characteristic function (2.1), then

$$\phi_{\varepsilon_n}(t) = \frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}\right)}{1+\tau c\ln\left(1+\lambda|t|^{\alpha}\right)}, \text{ where } c=p+(1-p)\theta.$$

 ISBN: 978-81-935819-2-6

That is, $$\phi_{\varepsilon_n}(t)=\frac{\tau\theta}{c}+\left(1-\frac{\tau\theta}{c}\right)\frac{1}{1+\tau c\ln\left(1+\lambda|t|^{\alpha}\right)}.$$

Hence, if the model (4.1) is stationary with $TGeGL_1(\alpha,\lambda,\tau,\theta)$ marginal distribution, then the distribution of the innovation sequence $\{\varepsilon_n\}$ is $TGeGL_1\left(\alpha,\lambda,\tau c,\frac{\tau\theta}{c}\right)$.

If $X_0\overset{d}{=}TGeGL_1(\alpha,\lambda,\tau,\theta)$ and $\{\varepsilon_n\}$ are independent and identically distributed as $TGeGL_1\left(\alpha,\lambda,\tau c,\frac{\tau\theta}{c}\right)$, then the characteristic function of X_1 is

$$\begin{aligned}\phi_{X_1}(t)&=\left[p+(1-p)\phi_{X_0}(t)\right]\phi_{\varepsilon_1}(t)\\&=\left[p+(1-p)\frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}\right)}{1+\tau\ln\left(1+\lambda|t|^{\alpha}\right)}\right]\left[\frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}\right)}{1+\tau c\ln\left(1+\lambda|t|^{\alpha}\right)}\right]\\&=\frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}\right)}{1+\tau\ln\left(1+\lambda|t|^{\alpha}\right)}.\end{aligned}$$

That is, $X_1\overset{d}{=}X_0$.

If $X_{n-1}\overset{d}{=}X_0$, we can prove that $X_n\overset{d}{=}X_0$ and hence the process $\{X_n\}$ is stationary.

Based on this, we now define stationary first order autoregressive tailed Type I generalized geometric Linnik process as follows:

Let $X_0\overset{d}{=}TGeGL_1(\alpha,\lambda,\tau,\theta)$

and for $n=1,2,...$

$$X_n=\begin{cases}\varepsilon_n & w.p. \quad p\\ X_{n-1}+\varepsilon_n & w.p. \quad 1-p\end{cases}$$

where $\{\varepsilon_n\}$ is a sequence of independent and identically distributed $TGeGL_1\left(\alpha,\lambda,\tau c,\frac{\tau\theta}{c}\right)$ random variables where $c=p+(1-p)\theta.$

It can be shown that the process $\{X_n\}$ is not time reversible. For this, consider the characteristic function

$$\begin{aligned}\phi_{X_n,X_{n+1}}(t_1,t_2)&=E\left(e^{it_1X_n+it_2X_{n+1}}\right)\\&=\phi_{\varepsilon_n}(t_2)\left[p\phi_{X_n}(t_1)+(1-p)\phi_{X_n}(t_1+t_2)\right]\end{aligned}$$

$$=\frac{1+\tau\theta\ln\left(1+\lambda|t|^{\alpha}\right)}{1+\tau c\ln\left(1+\lambda|t|^{\alpha}\right)}\left[p\frac{1+\tau\theta\ln\left(1+\lambda|t_1|^{\alpha}\right)}{1+\tau\ln\left(1+\lambda|t_2|^{\alpha}\right)}+(1-p)\frac{1+\tau\theta\ln\left(1+\lambda|t_1+t_2|^{\alpha}\right)}{1+\tau\ln\left(1+\lambda|t_1+t_2|^{\alpha}\right)}\right].$$

This expression is not symmetric in t_1 and t_2.

REMARK 4. 1

If $\{\varepsilon_n\}$ is a sequence of independent and identically distributed $TGeGL_1\left(\alpha,\lambda,\tau c,\frac{\tau\theta}{c}\right)$, where $c=p+(1-p)\theta$ then (4.1) is asymptotically stationary with $TGeGL_1(\alpha,\lambda,\tau,\theta)$ marginal distribution.

 ISBN: 978-81-935819-2-6

Consider the kth order autoregressive process

$$X_n = \begin{cases} \varepsilon_n & w.p. \quad p \\ X_{n-1} + \varepsilon_n & w.p. \quad p_1 \\ X_{n-2} + \varepsilon_n & w.p. \quad p_2 \\ \vdots & \\ X_{n-k} + \varepsilon_n & w.p. \quad p_k \end{cases}. \qquad (4.3)$$

If the process $\{X_n\}$ is stationary, then in terms of characteristic function, (4.3) is

$$\phi_{\varepsilon_n}(t) = \frac{\phi_X(t)}{p + (1-p)\phi_X(t)}, \text{where } 1 - p = \sum_{i=1}^{k} p_i, \quad 0 < p_i < 1.$$

Thus a necessary and sufficient condition for the model (4.3) defines a stationary AR(k) process with $TGeGL_1(\alpha, \lambda, \tau, \theta)$ marginal distribution is that $\{\varepsilon_n\}$ is distributed as $TGeGL_1\left(\alpha, \lambda, \tau c, \frac{\tau\theta}{c}\right)$.

As in the case of $TGeGL_1$ random variables, we now consider the first order autoregressive model (4.1). We can prove that if $\{\varepsilon_n\}$ are independent and identically distributed $TGeGAL_1\left(\alpha, \lambda, \mu, \tau c, \frac{\tau\theta}{c}\right)$ random variables and $X_0 \underline{\underline{d}}\, TGeGAL_1(\alpha, \lambda, \mu, \tau, \theta)$, then the model (4.1) is stationary with $TGeGAL_1(\alpha, \lambda, \mu, \tau, \theta)$ marginals. Also if the model (4.1) is stationary with $TGeGAL_1(\alpha, \lambda, \mu, \tau, \theta)$ marginals then it can be easily seen that the distribution of $\{\varepsilon_n\}$ is $TGeGAL_1\left(\alpha, \lambda, \mu, \tau c, \frac{\tau\theta}{c}\right)$ where $c = p + (1-p)\theta$. Based on this we can develop first order autoregressive models with tailed Type I generalized geometric asymmetric Linnik marginal distribution.

The model can be easily extended to higher order autoregressive model as in (4. 3).

References

1. Linnik, Yu.V. (1963) Linear forms and statistical criteria, I, II. *Selected Translations in Math. Statist. Prob.* **3**, 1-90.
2. Littlejohn, R.P. (1994) A reversibility relationship for two Markovian time series models with stationary exponential tailed distribution. *J. Appl. Prob.***31**, 575-581.
3. Muralidharan, K. (1999) Tests for the mixing proportions in the mixture of a degenerate and exponential distribution. *J. Ind. Statist. Assoc.* **37**, 105-119.
4. Muralidharan, K. and Kale, B.K. (2002) Modified gamma distribution with singularity at zero. *Commun.Statist. -Simul. Comput*. **31**, 143-158.
5. Pillai, R.N. (1985). Semi α-Laplace distributions. *Commun. Statist.-Theor. Meth.* **14**, 991-1000.

Estimation of Stress-Strength model using Three Parameter Generalized Lindley Distribution

K. S Deepthi & V M Chacko

Department of Statistics, St. Thomas' College (Autonomous), Thrissur, Kerala - 680 001 (India)
dipthiks@gmail.com, chackovm@gmail.com

Abstract

This paper deals with the reliability of a multicomponent stress-strength model assuming that the components follow three parameter generalized Lindley model. In a Multicomponent system with k components, strength of each component experiencing a random stress: the problem of estimation reliability is considered. The reliability of such a system is obtained when strength and stress variables are given by three parameter Lindley distribution. The system is regarded as alive only if with at least r out of $k(r<k)$ components, the strength of system exceeds the stress. The maximum likelihood estimation of reliability of the multicomponent system is obtained.

***Key Words**: Reliability, Stress-Strength Model*

I. Introduction

The estimation of reliability is a very common problem in statistical literature. The most widely approach applied for reliability estimation is well-known stress-strength model. Consider two independent random variables X and Y. Suppose Y represents the 'stress' and X represents the 'strength'. The reliability parameter $R = P(Y < X)$ is referred to as a stress-strength model, which is used in engineering statistics, quality control and other fields. In a reliability, the stress-strength model describes the life of a component that has a random strength variable X and is subjected to random variable stress Y. The system fails if and only if the stress is greater than strength at any time.

The estimation of a stress-strength model when X and Y are random variables having a specified distribution has been discussed by many authors including Raqab and Kundu (2005) discussed comparison of different estimates of stress-strength model for a scaled Burr type X distribution, Kundu and Gupta (2005) studied the stress-strength reliability based on generalized exponential distribution, Kundu and Gupta (2006) discussed estimation of stress-strength reliability in Weibull distribution, Raqab et al. (2008), Kundu and Raqab (2009)discussed estimation of stress-strength reliability with three parameter weibull distribution, Al-Mutairi et al. (2013) discussed estimation of stress-strength reliability in lindley distribution, Sharma et. al (2015) discussed stress-strength reliability model with application to headand neck cancer data based on inverse lindley distribution.

Let the random samples $Y, X_1, X_2, \cdots, X_k$ be independent, $G(y)$ be the continuous distribution function of Y and $F(x)$ be the common continuous distribution function of $X_1, X_2, \cdots, X_k$.The reliability in a multicomponent stress-strength model developed by Bhattacharyya and Johnson (1974) is given by

$R_{s,k} = P[$ at least s of the $(X_1, X_2, \cdots, X_k)$ exceed $Y]$

$$= \sum_{i=s}^{k} \binom{k}{i} \int_{-\infty}^{\infty} [1 - F(y)]^i [F(y)]^{k-i} \, dG(y)$$

where $X_1, X_2, \cdots, X_k$ are identically independently distributed (iid) with common distribution function $F(x)$ and subjected to the common random stress Y. Recently, Srinivasa Rao (2012) developed the multicomponent stress-strength reliability based on generalized exponential distribution, Srinivasa Rao et. al (2015) discussed the multicomponent stress-strength reliability with the Burr-XII distribution, the multicomponent stress-strength model using Lindley distributionis considered Marwa Khalil (2017).

Lindley (1958) introduced Lindley distribution in the context of fiducial and Baysian statistics and the probability density function (pdf) of one parameter Lindley distribution is given by

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

The probability density function of the Lindley distribution given in (1) is a mixture of Exponential(θ) and Gamma $(2,\theta)$. The Equation (1) can be expressed as

$$f(x;\theta) = pf_1(x) + (1-p) f_2(x) \qquad (2)$$

where $f_1(x)$ and $f_2(x)$ are the density function of Exponential(θ) and Gamma $(2,\theta)$ distribution and p is the mixing proportion. Properties, extension and applicationsof the Lindley distributions have been studied in the context of reliability analysis byGhitanyet. al (2008). Shanker and Mishra (2013) obtained a two parameter Lindley distribution is given by

$$g(y;\alpha,\theta) = \frac{\theta^2}{1+\alpha\theta}(\alpha + y)e^{-\theta y}, \; y > 0, \theta > 0, \alpha\theta > -1. \qquad (3)$$

The power Lindley (PL) distribution proposed by Ghitanyet. al (2013) an extensionof the Lindley distribution. Ekhosuehi and Opone (2018) introduced a three parameter generalized Lindley distribution. Then the probability density function (pdf) of TPLDis given by

$$f(x;\alpha,\beta,\lambda) = \frac{\alpha\lambda^2}{1+\beta\lambda}\left(\beta + x^{\alpha}\right)x^{\alpha-1}e^{-\lambda x^{\alpha}}, \; x > 0, \alpha, \beta, \lambda > 0 \qquad (4)$$

Here β and λ are scale parameters and α is shape parameter. The correspondingcumulative distribution function (cdf) is given by

$$F(x;\alpha,\beta,\lambda) = 1 - \left(\frac{1+\beta\lambda+\lambda x^{\alpha}}{1+\beta\lambda}\right)e^{-\lambda x^{\alpha}}, \; x > 0, \alpha, \beta, \lambda > 0 \qquad (5)$$

and the survival function is given by

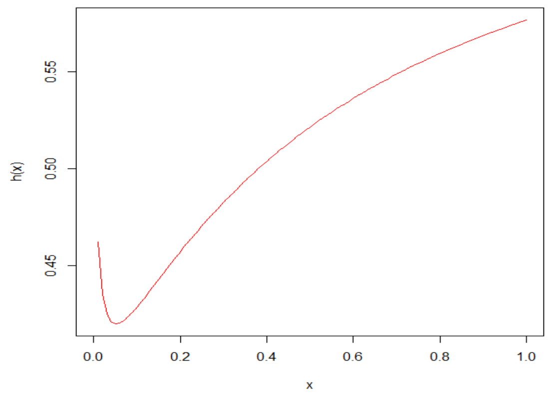
$$S(x;\alpha,\beta,\lambda) = \left(\frac{1+\beta\lambda+\lambda x^{\alpha}}{1+\beta\lambda}\right)e^{-\lambda x^{\alpha}}, \; x > 0, \alpha, \beta, \lambda > 0 \qquad (6)$$

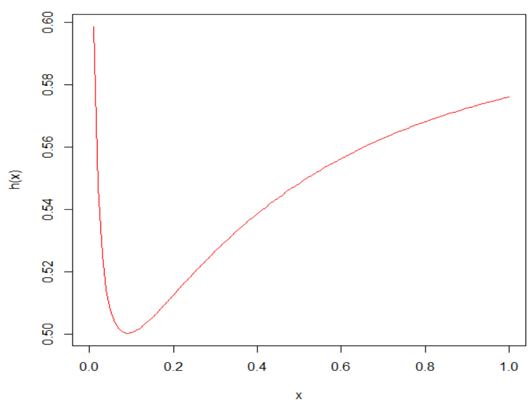
The hazard rate function of TPL distribution is given by

 ISBN: 978-81-935819-2-6

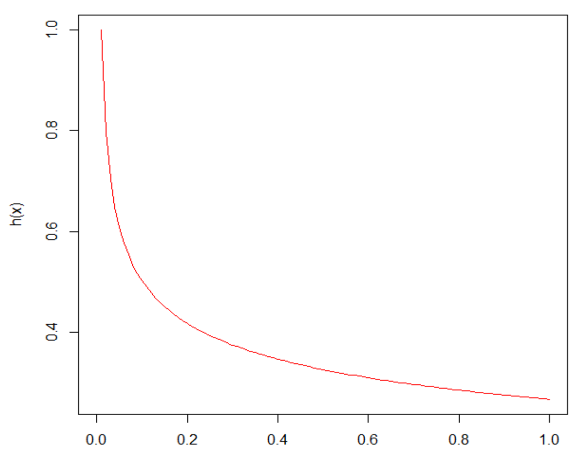
$$h(x;\alpha,\beta,\lambda)=\frac{\alpha\lambda^2\left(\beta+x^\alpha\right)x^{\alpha-1}}{1+\beta\lambda+\lambda x^\alpha},\ x>0,\alpha,\beta,\lambda>0 \qquad (7)$$

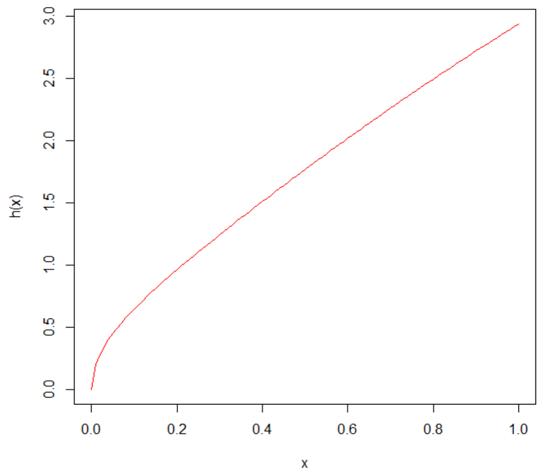
The graph of the hazard rate function of the TPGLD for different value of the parameters is given in Figure 2.





(a) $\alpha=0.85,\beta=0.25,\lambda=1.15$ (b) $\alpha=0.8,\beta=0.25,\lambda=1.2$





(c) $\alpha = 0.5, \beta = 0.15, \lambda = 1$ (d) $\alpha = 1.5, \beta = 0.45, \lambda = 2.5$

Figure 2 – Different parameter values of Hazard rate function of the TPGLD

Clearly from Figure 2, the TPGLD exhibits monotone increasing, decreasing and bathtub shaped failure rate behavior. It increases monotonically when $\alpha > 1$ and, decreasing and bathtub when $\alpha < 1$.

The main objective of this paper is to check the appropriateness of the multicomponent stress-strength reliability of three parameter generalized Lindley distribution. Next we have to examine extent the reliability of parallel case and series case of the multicomponent stress-strength model.

This paper deals with the reliability of a multicomponent stress-strength model assuming that the components follow three parameter Lindley model. Section 2 discussed the multicomponent system reliability in three parameter generalized Lindley distribution. The maximum likelihood estimator of reliability, $R_{r,k}$, is determined in section 3. Real data is used as a practical application of the proposed model in section 4. Conclusions are given in section 5.

II. System Reliability

Let X and Y be two random variables that follows independent identically distributedthree parameter Lindley distribution with common shape parameter $\alpha, (\beta_1, \lambda_1)$ and (β_2, λ_2) respectively. Suppose $X_1, X_2, \ldots, X_{n_2}$ and $Y_1, Y_2, \ldots, Y_{n_1}$ are two independent samples from X and Y, respectively. The stress-strength reliability is

$$R_{r,k} = P\left(Y_{r:n_1} < X_{k:n_2}\right) = \int_0^\infty F_{Y_{r:n_1}}(x) f_{X_{k:n_2}}(x) dx$$

(8)

where, $F_{Y_{r:n_1}}(x)$ and $f_{X_{k:n_2}}(x)$ are the r^{th} cumulative distribution function and k^{th} probability density function of $Y_{r:n_1}$ and $X_{k:n_2}$ respectively.

Then the k^{th} probability density function of $X_{k:n_2}$ is given by

 ISBN: 978-81-935819-2-6

$$f_{X_{k:n_2}}(x)=k\binom{n_2}{k}F^{k-1}(x)[1-F(x)]^{n_2-k}f(x) \quad (9)$$

and the r^{th} cumulative density function of $Y_{r:n_1}$ is given by

$$F_{Y_{r:n_1}}(x)=\sum_{j=r}^{n_1}\binom{n_1}{j}\cdot F^{j}(x)[1-F(x)]^{n_1-j} \quad (10)$$

From (9) and (10) in (8),we obtain

$$R_{r,k}=k\frac{\alpha\lambda_1^2}{1+\lambda_1\beta_1}\binom{n_2}{k}\sum_{j=r}^{n_1}\binom{n_1}{j}\int_0^{\infty}\left(\beta_1+x^{\alpha}\right)\cdot x^{\alpha-1}\cdot e^{-\lambda_1x^{\alpha}}\left[1-\left(\frac{1+\lambda_2\beta_2+\lambda_2x^{\alpha}}{1+\lambda_2\beta_2}\right)\cdot e^{-\lambda_2x^{\alpha}}\right]^{j}$$
$$\times\left[\left(\frac{1+\lambda_2\beta_2+\lambda_2x^{\alpha}}{1+\lambda_2\beta_2}\right)\cdot e^{-\lambda_2x^{\alpha}}\right]^{n_1-j}\left[1-\left(\frac{1+\lambda_1\beta_1+\lambda_1x^{\alpha}}{1+\lambda_1\beta_1}\right)\cdot e^{-\lambda_1x^{\alpha}}\right]^{k-1}\left[\left(\frac{1+\lambda_1\beta_1+\lambda_1x^{\alpha}}{1+\lambda_1\beta_1}\right)\cdot e^{-\lambda_1x^{\alpha}}\right]^{n_2-k}dx$$

$$R_{r,k}=k\frac{\alpha\lambda_1^2}{1+\lambda_1\beta_1}\binom{n_2}{k}\sum_{j=r}^{n_1}\sum_{l_1=0}^{n_1-j}\sum_{l_2=0}^{n_2-k}\sum_{l_3=0}^{k-1}\sum_{l_4=0}^{k-l_3-1}\sum_{l_5=0}^{j}\sum_{l_6=0}^{j-l_5}(-1)^{j+k-l_3-l_5-1}\binom{n_1}{j}\binom{n_1-j}{l_1}\binom{n_2-k}{l_2}\binom{k-1}{l_3}$$
$$\times\binom{k-l_3-1}{l_4}\binom{j}{l_5}\binom{j-l_5}{l_6}\cdot\left(\frac{\lambda_2}{1+\lambda_2\beta_2}\right)^{n_1-l_1-l_5-l_6}\left(\frac{\lambda_1}{1+\lambda_1\beta_1}\right)^{n_2-l_2-l_3-l_4-1}$$
$$\times\left(\frac{\beta_1\Gamma(n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6}}+\frac{\Gamma(n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6+1)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6+1}}\right)$$
(11)

where $\Gamma(\cdot)$ is the gamma function. We now present some special cases of $R_{r,k}$ with a different arrangement of the components.

2.1. Reliability of the Parallel System

For parallel case, the reliability of the system is given by

$$R_1=P\left(\max\left(Y_1,Y_2,\ldots,Y_{n_1}\right)<\max\left(X_1,X_2,\ldots,X_{n_2}\right)\right)$$

Then the stress-strength reliability of parallel system is

$$R_{n_1,n_2}=n_2\frac{\alpha\lambda_1^2}{1+\lambda_1\beta_1}\sum_{l_3=0}^{n_2-1}\sum_{l_4=0}^{n_2-l_3-1}\sum_{l_5=0}^{n_1}\sum_{l_6=0}^{n_1-l_5}(-1)^{n_1+n_2-l_3-l_5-1}\binom{n_2-1}{l_2}\binom{n_1}{l_5}\binom{n_1-l_5}{l_6}\left(\frac{\lambda_2}{1+\lambda_2\beta_2}\right)^{n_1-l_5-l_6}$$
$$\times\left(\frac{\lambda_1}{1+\lambda_1\beta_1}\right)^{n_2-l_3-l_4-1}\left(\frac{\beta_1\Gamma(n_1+n_2-l_3-l_4-l_5-l_6)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_3-l_4-l_5-l_6}}+\frac{\Gamma(n_1+n_2-l_3-l_4-l_5-l_6+1)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_3-l_4-l_5-l_6+1}}\right)$$

2.2. Reliability of the Series System

The reliability of system for the series case is

$$R_2=P\left(\max\left(Y_1,Y_2,\ldots,Y_{n_1}\right)<\min\left(X_1,X_2,\ldots,X_{n_2}\right)\right)$$

Then the stress-strength reliability of series system is

 ISBN: 978-81-935819-2-6

$$R_{n_1,1} = n_2 \frac{\alpha\lambda_1^2}{1+\lambda_1\beta_1}\sum_{l_2=0}^{n-1}\sum_{l_5=0}^{n_1}\sum_{l_6=0}^{n_1-l_5}(-1)^{n_1-l_5}\binom{n_2-1}{l_2}\binom{n_1}{l_5}\binom{n_1-l_5}{l_6}\left(\frac{\lambda_2}{1+\lambda_2\beta_2}\right)^{n_1-l_5-l_6}$$

$$\times\left(\frac{\lambda_1}{1+\lambda_1\beta_1}\right)^{n_2-l_2-1}\left(\frac{\beta_1\Gamma(n_1+n_2-l_2-l_5-l_6)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_2-l_5-l_6}}+\frac{\Gamma(n_1+n_2-l_2-l_5-l_6+1)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_2-l_5-l_6+1}}\right)$$

Now present another special cases of $R_{r,k}$ with a different arrangement of the components.

1. For $r=1$ and $k=1$, the minimum strength component is subjected to the minimum stress component.

$$R_{1,1} = n_1 n_2 \frac{\alpha\lambda_1^2}{1+\lambda_1\beta_1}\sum_{j=1}^{n_1}\sum_{l_1=0}^{n_1-1}\sum_{l_2=0}^{n_2-1}\sum_{l_5=0}^{j}\sum_{l_6=0}^{j-l_5}(-1)^{j-l_5}\binom{n_1}{j}\binom{n_1-1}{l_1}\binom{n_2-1}{l_2}\binom{j}{l_5}\binom{j-l_5}{l_6}$$

$$\times\left(\frac{\lambda_2}{1+\lambda_2\beta_2}\right)^{n_1-l_1-l_5-l_6}\left(\frac{\lambda_1}{1+\lambda_1\beta_1}\right)^{n_2-l_2}\left(\frac{\beta_1\Gamma(n_1+n_2-l_1-l_2-l_5-l_6)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_1-l_2-l_5-l_6}}+\frac{\Gamma(n_1+n_2-l_1-l_2-l_5-l_6+1)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_1-l_1-l_2-l_5-l_6+1}}\right)$$

2. For $r=n_1$ and $k=k$, the k^{th} strength component is subjected to the maximum stress component.

$$R_{n_1,k} = \frac{k\alpha\lambda_1^2}{1+\lambda_1\beta_1}\binom{n_2}{k}\sum_{l_2=0}^{n_2-k}\sum_{l_3=0}^{k-1}\sum_{l_4=0}^{k-l_3-1}\sum_{l_5=0}^{n_1}\sum_{l_6=0}^{n_1-l_5}(-1)^{n_1+k-l_3-l_5-1}\binom{n_2-k}{l_2}\binom{k-1}{l_3}\binom{k-l_3-1}{l_4}$$

$$\times\binom{n_1}{l_5}\binom{n_1-l_5}{l_6}\left(\frac{\lambda_2}{1+\lambda_2\beta_2}\right)^{n_1-l_5-l_6}\left(\frac{\lambda_1}{1+\lambda_1\beta_1}\right)^{n_2-l_2-l_3-l_4-1}$$

$$\times\left(\frac{\beta_1\Gamma(n_1+n_2-l_2-l_3-l_4-l_5-l_6)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_2-l_3-l_4-l_5-l_6}}+\frac{\Gamma(n_1+n_2-l_2-l_3-l_4-l_5-l_6+1)}{\alpha(n_2\lambda_1+n_1\lambda_2)^{n_1+n_2-l_2-l_3-l_4-l_5-l_6+1}}\right)$$

III. Maximum Likelihood Estimator of $R_{r,k}$

Suppose $X_1, X_2, \ldots, X_{n_2}$ be the random sample of the strength of the n_2 systems that are distributed as three parameter generalized Lindley r.v's with parameter λ_1 and β_1. Let $Y_1, Y_2, \ldots, Y_{n_1}$ be the random sample of the stress of the n_1 systems that are distributed as three parameter generalized Lindley r.v's with parameter λ_2 and β_2 and common shape parameter α.

Then the log-likelihood function of the observed sample is

$$\log L(\beta_1,\beta_2,\lambda_1,\lambda_2,\alpha) = (n_1+n_2)\log\alpha + 2n_2\log\lambda_1 - n_2\log(1+\lambda_1\beta_1) + \sum_{i=1}^{n_2}\log(\beta_1 + x_i^{\alpha})$$

$$+(\alpha-1)\sum_{i=1}^{n_2}\log x_i - \lambda_1\sum_{i=1}^{n_2}x_i^{\alpha} + 2n_1\log\lambda_2 - n_1\log(1+\lambda_2\beta_2) + \sum_{j=1}^{n_1}\log(\beta_2 + y_j^{\alpha}) \quad (12)$$

$$+(\alpha-1)\sum_{j=1}^{n_1}\log y_j - \lambda_2\sum_{j=1}^{n_1}y_j^{\alpha}$$

The maximum likelihood estimators of $\beta_1, \beta_2, \lambda_1, \lambda_2$ and α denoted by $\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}_1, \hat{\lambda}_2$ and

ISBN: 978-81-935819-2-6

$\hat{\alpha}$ respectively can be obtained as the solution of the following nonlinear equations

$$\frac{\partial \log L}{\partial \beta_1} = -\frac{n_2 \lambda_1}{1+\lambda_1 \beta_1} + \sum_{i=1}^{n_2} \frac{1}{\beta_1 + x_i^{\alpha}} \tag{13}$$

$$\frac{\partial \log L}{\partial \beta_2} = -\frac{n_1 \lambda_1}{1+\lambda_1 \beta_1} + \sum_{j=i}^{n_1} \frac{1}{\beta_2 + y_j^{\alpha}} \tag{14}$$

$$\frac{\partial \log L}{\partial \lambda_1} = \frac{2n_2}{\lambda_1} - \frac{n_2 \beta_1}{1+\lambda_1 \beta_1} - \sum_{i=1}^{n_2} x_i^{\alpha} \tag{15}$$

$$\frac{\partial \log L}{\partial \lambda_2} = \frac{2n_1}{\lambda_2} - \frac{n_1 \beta_2}{1+\lambda_2 \beta_2} - \sum_{j=1}^{n_1} y_j^{\alpha} \tag{16}$$

$$\log L(\beta_1, \beta_2, \lambda_1, \lambda_2, \alpha) = (n_1 + n_2)\log \alpha + 2n_2 \log \lambda_1 - n_2 \log(1+\lambda_1 \beta_1) + \sum_{i=1}^{n_2} \log(\beta_1 + x_i^{\alpha})$$

$$\text{and} + (\alpha - 1)\sum_{i=1}^{n_2} \log x_i - \lambda_1 \sum_{i=1}^{n_2} x_i^{\alpha} + 2n_1 \log \lambda_2 - n_1 \log(1+\lambda_2 \beta_2) + \sum_{j=1}^{n_1} \log(\beta_2 + y_j^{\alpha})$$

$$+ (\alpha - 1)\sum_{j=1}^{n_1} \log y_j - \lambda_2 \sum_{j=1}^{n_1} y_j^{\alpha}$$

$$\frac{\partial \log L}{\partial \alpha} = \frac{n_1 + n_2}{\alpha} + \sum_{i=1}^{n_2} \frac{x_i^{\alpha} \log x_i}{\beta_1 + x_i^{\alpha}} + \sum_{i=1}^{n_2} \log x_i - \lambda_1 \sum_{i=1}^{n_2} x_i^{\alpha} \log x_i + \sum_{j=1}^{n_1} \frac{y_j^{\alpha} \log y_j}{\beta_2 + y_j} + \sum_{j=1}^{n_2} \log y_j - \lambda_2 \sum_{j=1}^{n_2} y_j^{\alpha} \log y_j \tag{17}$$

By using the invariance property of the maximum likelihood estimator of $R_{r,k}$ denotedby $\hat{R}_{r,k}^{ML}$ can obtained by replacing $\beta_1, \beta_2, \lambda_1, \lambda_2$ and α by their maximum likelihood estimators. Therefore the maximum likelihood of $R_{r,k}$ is obtained by

$$\hat{R}_{r,k}^{ML} = k \frac{\hat{\alpha}\hat{\lambda}_1^2}{1+\hat{\lambda}_1 \hat{\beta}_1} \binom{n_2}{k} \sum_{j=r}^{n_1} \sum_{l_1=0}^{n_1-j} \sum_{l_2=0}^{n_2-k} \sum_{l_3=0}^{k-1} \sum_{l_4=0}^{k-l_3-1} \sum_{l_5=0}^{j} \sum_{l_6=0}^{j-l_5} (-1)^{j+k-l_3-l_5-1} \binom{n_1}{j} \binom{n_1-j}{l_1} \binom{n_2-k}{l_2} \binom{k-1}{l_3}$$
$$\times \binom{k-l_3-1}{l_4} \binom{j}{l_5} \binom{j-l_5}{l_6} \cdot \left(\frac{\lambda_2}{1+\lambda_2 \beta_2}\right)^{n_1-l_1-l_5-l_6} \left(\frac{\hat{\lambda}_1}{1+\hat{\lambda}_1 \hat{\beta}_1}\right)^{n_2-l_2-l_3-l_4-1}$$
$$\times \left(\frac{\hat{\beta}_1 \Gamma(n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6)}{\hat{\alpha}(n_2\hat{\lambda}_1 + n_1\hat{\lambda}_2)^{n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6}} + \frac{\Gamma(n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6+1)}{\hat{\alpha}(n_2\hat{\lambda}_1 + n_1\hat{\lambda}_2)^{n_1+n_2-l_1-l_2-l_3-l_4-l_5-l_6+1}} \right) \tag{18}$$

The asymptotic variance of $\hat{R}_{r,k}^{ML}$ is given by

$$AV(\hat{R}_{r,k}^{ML}) = \sum_{i=1}^{5} \sum_{j=1}^{5} \frac{\partial R_{r,k}}{\partial \theta_i} \frac{\partial R_{r,k}}{\partial \theta_j} I^{-1}(\underline{\theta}) \tag{19}$$

where $\underline{\theta} = (\beta_1, \beta_2, \lambda_1, \lambda_2, \alpha)$ and $I^{-1}(\underline{\theta})$ is the Fisher Information Matrix. Therefore, an asymptotic $100(1-\eta)\%$ confidence interval for Rr;k can obtain as $\hat{R}_{r,k}^{ML} \pm Z_{\frac{\eta}{2}} \sqrt{AV\left(\hat{R}_{r,k}^{ML}\right)}$

where $Z_{\frac{\eta}{2}}$ is the upper $\frac{\eta}{2}$ quantile of standard normal distribution.

IV. Conclusion

The method of reliability estimation of stress strength model is discussed. We obtained an expression of reliability. The non-linear estimation require use of softaware like R, Matlab etc. The distributions of stress and strength are assumed to be bathtub shaped, which is quire desirable since the strength of human life increases at first stage (infant morality period), remains constant for a long period (useful life) and thereafter decreases (wear out period). So modeling stress strength reliability problems using bathtub shaped failure rate model become more realistic.

References

1. Al-Mutairi, D. K., Ghitany, M. E. &Kundu, D. (2013), 'Inferences on stress-strength reliability from lindley distributions', Communications in StatisticsTheory and Methods 42(8), 1443–1463.
2. Bhattacharyya, G. K. & Johnson, R. A. (1974), 'Estimation of reliability in multicomponentstress-strength model', Journal of the American Statistical Association, 69, 966–970.
3. Ghitany, M. E., Atieh, B. &Nadarajah, S. (2008), 'Lindley distribution and it'sapplication', Journal of Mathematics and Computers in Simulation 76, 493–506.
4. Ghitany, M.E., Al-Mutairi, D., Balakrishnan, N. &I.Al-Enezi (2013),'Power Lindleydistribution and associated inference', Computational Statistics and Data Analysis,64, pp. 20-33.
5. Kundu, D. & Gupta, R.D. (2005),'Estimation of $P[Y < X]$ for generalized exponential distribution', Metrika, 61, 291–308.
6. Kundu, D. & Gupta, R.D. (2006),'Estimation of $R = P[Y < X]$ for Weibull distributions', IEEE Transactions on Reliability,55, 270–280.
7. Kundu, D., Raqab, M.Z. (2009),'Estimation of $R = P(Y < X)$ for three-parameter Weibull Distribution',Statistics and Probability Letters, 79, 1839–1846.
8. Lindley, D. V. (1958), 'Fiducial distributions and bayes' theorem', Journal of theRoyal Statistical Society. Series B (Methodological), 102–107.
9. Marwa Khalil (2017), 'Estimation a Stress-Strength Model for Using the Lindley Distribution' , Rev.Colomb.Estad, 40(1), 105 -121.
10. Nosakhare Ekhosuehi & Festus Opone (2018),'A Three Parameter Generalized Lindley Distribution: Properties and Application', Statistica 78(3).
11. Raqab, M.Z., Kundu, D. (2005). 'Comparison of different estimators of $P[Y < X]$ for ascaled Burr type X distribution'. Communications in Statistics-Simulation and Computation, 34, 465–483.
12. Sharma, V. K., Singh, S. K., Singh, U. & Agiwal, V. (2014), 'The inverseLindley distribution: A stress-strength reliability model', arXiv preprint arXiv:1405.6268.
13. Sharma, V. K., Singh, S. K., Singh, U. & Agiwal, V. (2015), 'The inverse lindley distribution: a stress-strength reliability model with application to headand neck cancer data', Journal of Industrial and Production Engineering, 32(3), 162–173.
14. Shanker and Mishra (2013),'A two parameter Lindley distribution',STATISTICS IN TRANSITION-new series, Spring, 14(1), 45–56.
15. Srinivasa Rao, G., Muhammad Aslam & Kundu, D. (2015),'Burr-XII Distribution ParametricEstimation and Estimation of Reliability of Multicomponent Stress-

ISBN: 978-81-935819-2-6

Strength', Communications in Statistics-Simulation and Computation, 44(23),4953-4961.

16. Srinivasa Rao, G.(2012),'Estimation of Reliability in Multicomponent Stress-strength Based on Generalized Exponential Distribution', *Rev.Colomb.Estad.* [online],35(1),67-76.

A Queueing Network model for the performance analysis of Eye Care Clinic.

Rajitha C and Chacko V M
Department of statistics
St. Thomas College (Autonomous), Thrissur-1

Abstract

In this paper we study an open queueing network with Poisson arrival and exponential service times at stations of an eye care clinic. The steady state characteristics of the network is obtained and each station solved independently by using M/M/1/∞ model. Also studied the same with blocking. Blocking occurred when at least one service center has limited queueing space or capacity before it.

Keywords: *Queueing networks, Blocking, Steady state.*

I. Introduction

Collection of interactive queueing systems is known as network of queues. Queueing networks mainly classified as open queueing networks, closed queueing networks and mixed queueing networks. Open queueing networks described as customers can arrive from outside the system at any node and depart from the system from any node.At least one service center has limited waiting space or capacity before it, those are classified in to restricted queueing networks.

When waiting space between the stations are finite then there occur .Hunt [3] used a modified series model to find solution for a two station series queue with finite queueing space between stations. Takahashi et al.[8] , Perros and Atlok[6] suggested approximate analysis for open queueing networks with blocking. Analysis of blocking in open restricted queueing system by decomposition method done Koizumi et al.[5] Recently analysis of restricted queueing networks-a blocking approach with special reference to health care system studied by Sreekala and Manoharan [7].

In this paper first we study an open queueing network of eye care clinic with infinite capacity in each station. Steady state equations and performance parameters are obtained. When waiting space between stations are finite there occurs blocking. The steady state analysis with blocking is done in next section. Analysis of each station using decomposition approach is also presented in the paper.

We consider an eye care system as an open queueing network system. In this model six stations are defined. First node S_1 is for admission, customers arrive according to homogeneous Poisson process. Second station is for vision checking and this node follows M/M/1/∞/FCFS schedule. The third node is for scanning.S_3 also follows M/M/1/K/FCFS. Patients after treatment also joins the queue for further scanning. Fourth node is checking by doctor follows M/M/1/K/FCFS. After checking by doctor some patients leave from the system with probability $\propto_1$ and remaining admitted for treatment with probability 1-$\propto_1$.Last node is for treatment

II. Methodological Framework

We consider an open queueing network of eye care clinic with five single servers of finite and infinite capacity. Stations denoted by S_i (i=1,2,…,5). Customers arrive according to homogeneous Poisson process with rate λ and service rate follows exponential distribution with rate μ_i (i=1,2, ..., 5).There is only finite capacity for nodes 3,4 and 5 and others are of infinite capacity. Diagrammatic representation of which is:

 ISBN: 978-81-935819-2-6

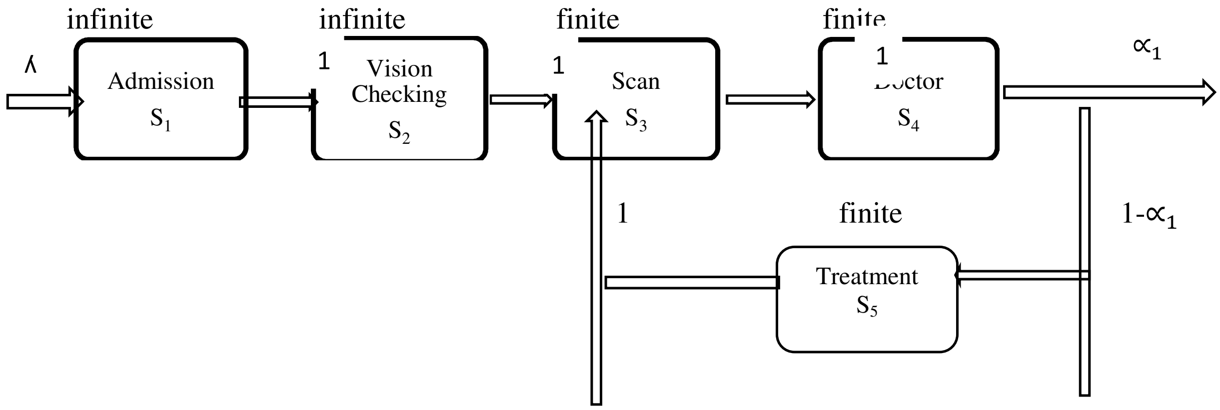


Figure 1. Eye care queueing network with blocking

Waiting space between stations four, five and six are of finite capacity and other stations infinite. Therefore blocking happens only between $S_3 \rightarrow S_4, S_4 \rightarrow S_5$and $S_5 \rightarrow S_3$. We are model the flows$S_2 \rightarrow S_3, S_3 \rightarrow S_4, S_4 \rightarrow S_5$. Arrival to each node is according to Poisson process. The types of congestion listed in table 1.

Table 1: Types of congestions

Flow	Cause of congestion	Facing station	Congestion type
$S_1 \rightarrow S_2$	Not applicable	Not applicable	No congestion
$S_2 \rightarrow S_3$	S_3 is full	S_2	Classic Congestion
$S_3 \rightarrow S_4$	S_3 is full	S_3	Blocking
$S_4 \rightarrow S_5$	S_4 is full	S_4	Blocking
$S_5 \rightarrow S_3$	S_5 is full	S_5	Blocking

III. Steady –state Analysis

In this section we first consider the system without blocking. Here we assumed each station has infinite capacity. The steady state analysis is done by referring Gross and Harris[2] and Bose[1].

3.1 Steady State Analysis without blocking

The routing probability matrix generally defined as

$$R = \begin{bmatrix} r_{00} & r_{01} & r_{02} & r_{03} & r_{04} & r_{05} \\ r_{10} & r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{20} & r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{30} & r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \\ r_{40} & r_{41} & r_{42} & r_{43} & r_{44} & r_{45} \\ r_{50} & r_{51} & r_{52} & r_{53} & r_{54} & r_{55} \end{bmatrix}$$

wherer_{ij} is the routing probability from station i to station j (i,j= 1,2,..,5). The routing probability matrix of our model based on figure 1 is

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \propto_1 & 0 & 0 & 0 & 0 & 1-\propto_1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

We can find $\Lambda_i : i = 1,2, \dots 5$ (Total arrival rates) by solving the traffic equations:

$$\Lambda_1 = \Lambda$$
$$\Lambda_2 = r_{12} \Lambda_1$$
$$\Lambda_3 = r_{23} \Lambda_4 + r_{53} \Lambda_5$$
$$\Lambda_4 = r_{34} \Lambda_3$$
$$\Lambda_5 = r_{45} \Lambda_4$$

By substitution we get,

$$\Lambda_1 = \Lambda_2 = \Lambda$$
$$\Lambda_3 = \frac{\Lambda}{1-(1-\alpha_1)\alpha_1}$$
$$\Lambda_4 = \frac{\alpha_1 \Lambda}{1-(1-\alpha_1)\alpha_1} \quad (1)$$
$$\Lambda_5 = \Lambda$$

We assume there is an infinite buffer between stations. So we can solve each station independently applying M/ M/ 1/∞ queueing model.

3.2 Average queue length and Average queue delay

Average queue length of station i is obtained from the formula

$$L_i^q = \frac{\rho_i^2}{1-\rho_i}, \quad (2)$$

where$\rho_i = \frac{\Lambda_i}{\mu_i} < 1$ (i=1,2,…,5)

By using Little's formula we can obtain the average steady state waiting time,

$$W_i^q = \frac{\rho_i^2}{\Lambda_i(1-\rho_i)}, i = 1,2, \dots 5 \quad (3)$$

3.3 Steady state analysis with blocking

When potential blockingexists between stations (no buffer), congestion at any particular station could potentially affect congestion levels at all upstream stations. To find interactions between stations we modify Jacksons approach with the help of effective service time Takahashi et al. (1980). The effective service time comprised with treatment time and service time. Treatment time equal to service time for networks without blocking. Here we assume effective service times follows exponential distribution, the mean effective service time at station i denoted by$\frac{1}{\tilde{\mu}_i}$. Effective waiting time is defined as the convex combination of waiting times. Defined as

$$\frac{1}{\tilde{\mu}_i} = r_{i0}\left(\frac{1}{\mu_i}\right) + \sum_j r_{ij}\left(\frac{1}{\mu_i} + W_j\right),$$

wherer_{i0} is the routing probability of patients leaving from state i without facing any wait, r_{ij} is the routing probability from station i to j. In our model stations $S_3, S_4, and\ S_5$ face blocking. The effective service time corresponding to $S_3, S_4, and\ S_5$ are

$$\frac{1}{\tilde{\mu}_3} = r_{30}\left(\frac{1}{\mu_3}\right) + r_{34}\left(\frac{1}{\mu_3} + W_4^q\right) \quad (4)$$
$$\frac{1}{\tilde{\mu}_4} = r_{40}\left(\frac{1}{\mu_4}\right) + r_{45}\left(\frac{1}{\mu_4} + W_5^q\right), \quad (5)$$
$$\frac{1}{\tilde{\mu}_5} = r_{50}\left(\frac{1}{\mu_5}\right) + r_{53}\left(\frac{1}{\mu_5} + W_3^q\right) \quad (6)$$

By substituting routing probabilities

$$\frac{1}{\tilde{\mu}_3} = \frac{1}{\mu_3} + W_4^q,$$
$$\frac{1}{\tilde{\mu}_4} = \alpha_1\left(\frac{1}{\mu_4}\right) + 1 - \alpha_1\left(\frac{1}{\mu_4} + W_5^q\right)\ and$$

 ISBN: 978-81-935819-2-6

$$\frac{1}{\tilde{\mu}_5} = \frac{1}{\mu_5} + W_3^q$$

Using equations (2) and (3) obtain steady state queue lengths and waiting times in terms of effective service times.

IV. Analysis of Stations

By using single node decomposition approximation by Takahashi et al.[8] the steady state of every station can solve independently from last station to first station. The steady state of each finite station (M/M/1/∞ $queue$) is analyzed using this approximation.

4.1 Analysis of station five (S_5)

In our model, station five represents who needs recheck and treatment in the clinic. The downstream node S_4 is also finite. So S_5 face blocking if S_4 is full. Corresponding to $S_5$the queue length and queue delay obtained by solving (2) and (3) in terms of effective service times (6). The queue length corresponding to S_5 is

$$L_{45}^q = L_5^q(\Lambda_{45}/\Lambda_5)$$
$$= L_5^q(\mathrm{r}_{45}\,\Lambda_4/\Lambda_5)$$

whereL_{45}^q is the blocked persons at S_4 waiting to enter S_5.

4.2 Analysis of station four (S_4)

Station four represents patients who entered for doctors checking. The downstream node S_3 is also finite. So S_4 face blocking if S_3 is full. Corresponding to $S_4$the queue length and queue delay obtained by solving (2) and (3) in terms of effective service times (5). The queue length corresponding to $S_4$is

$$L_{34}^q = L_4^q(\Lambda_{34}/\Lambda_4)$$
$$= L_4^q(\mathrm{r}_{34}\,\Lambda_3/\Lambda_4)$$

whereL_{34}^q is the blocked persons at S_3 waiting to enter S_4.

4.3 Analysis of station three (S_3)

Station three represents patients who entered for scanning before doctors checking. The downstream node S_2 is infinite. So the effective service time

$$\frac{1}{\tilde{\mu}_3} = \frac{1}{\mu_3}.$$

The queue length corresponding to S_3 is

$$L_{23}^q = L_3^q$$

V. Conclusion

We studied open queueing network with infinite and finite waiting space between stations of eye care clinic. When waiting space is finite blocking occurs between stations. Which is studied by decomposition approach. This approach can apply to many open restricted queueing models of different areas.

References

[1]. S.K Bose, An introduction to queueing Systems, *Kluwer academic/ Plenum publishers*, New York, 2002

[2]. G. Gross and C. Harris, Fundamentals of queueing theory, *John Wiley and sons*, 1998

[3]. G. C. Hunt, Sequential arrays of waiting lines, *Operations Research* 4 (1956), 674-683

[4]. J.R. Jackson, Network of waiting lines, *Operations Research* 5 (1957), 518-521.

[5]. N. Koizumi, E. Kuno and T. E. Smith, Modelling patients flows using a queueing network with blocking, *Health care Manag. Sci*. 8 (2005). 49-60

[6]. H. G. Perros and T. Attiok, Approximate analysis of open networks of queues with blocking: Tandem Configurations, *IEEE Trans. Soft. Eng*. 12 (1986), 450-461

[7]. M. S. Sreekala and M. Manoharan, Analysis of restricted queueing networks- A blocking approach, *Journal of Statistical Science and Application*, Vol.2 (2016), 220-230

[8]. Y. Takahashi, H. Miyahara and T. Hasegawa, An approximation method for open restricted queueing networks, *Operations Research* 28 (1980). 594-602.

ISBN: 978-81-935819-2-6

Some Results Related to Location Scale family of Esscher Transformed Laplace Distribution

Dais George[1] and Rimsha H[2]*
[1]Catholicate College, Pathanamthitta, Kerala, India
daissaji@rediffmail.com
[2] Research Scholar, Bharathiar University
rimshahabeeb@gmail.com

Abstract

Esscher transformed Laplace distribution is an asymmetric and heavy-tailed distribution introduced by Sebastian George and Dais George (2012). It is a tilted version of the classical Laplace distribution. In this paper, we consider some properties and representations of the location scale family of Esscher transformed Laplace distribution. The estimation procedure of the distribution is also considered. Some applications of this distribution are also discussed.

Keywords: *Convolution, Entropy, Hazard Rate Function, Three parameter Esscher transformed Laplace distribution.*

Bayesian Analysis Of Generalized Maxwell - Boltzmann Distribution Under Different Loss Functions And Prior Distributions.

Reshma T. S. and Nicy Sebastian
Department of Statistics, St. Thomas College (Autonomous), Thrisssur
Kerala, India-680001
nicycms@gmail.com, reshmats1996@gmail.com

Abstract

Present paper discussed Bayesian estimation of Maxwell-Boltzmann distribution by considering different combinations of loss functions and prior distributions. Here we assumed that the rate parameter of Maxwell-Boltzmann distribution follows extension of Jeffrey's prior and gamma prior. Square error, precautionary, Stein's and Al-Bayyati's loss function are the loss functions considered in this paper. Under the mutual combinations of considered loss functions and prior distributions, we have derived different estimators of rate parameter. These estimators are compared in terms of mean square error by using the programming language R. We also introduced the corresponding methods by using generalized Maxwell-Boltzmann distribution, which can be obtained by extending the pathway property of Mathai (2005).

Keywords:
Extension of Jeffrey's prior, Gamma prior, squared error loss function, precautionary loss function, Al-Bayyati's loss function, Stein's loss function.

ISBN: 978-81-935819-2-6

Mixed Distribution of Exponential and Gamma Distributions

Beenu Thomas
beenuneel.18@gmail.com
Department of Statistics
St.Thomas College (Autonomous), Thrissur, Kerala, India

Abstract

Lifetime distributions for many components usually have a bathtub-shaped failure rate in practice. However, there are very few practical models having bathtub shaped failure rate function. Models with bathtub-shaped failure rate function are useful in reliability analysis, and particularly in reliability related decision making and cost analysis. This paper introduces a new distribution, which is a mixture of Exponential and Gamma distribution, which may have bathtub shaped failure rate function. Some of its statistical properties are obtained. Parameter estimation methods are given for this new distribution. Application is illustrated using real data.

Key Words*: Reliability, Bathtub shaped failure rate, Mixture, Gamma distribution, Exponential distribution.*