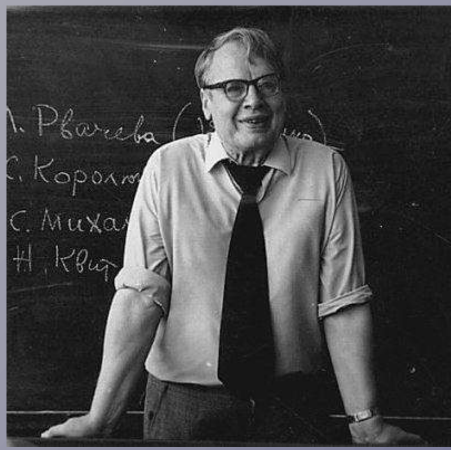
ELECTRONIC JOURNAL
OF INTERNATIONAL
GROUP ON RELIABILITY

Gnedenko Forum Publications



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

•

ISSN 1932-2321

•

VOL.14 NO.2 (53)
JUNE, 2019

RELIABILITY:
THEORY&APPLICATIONS

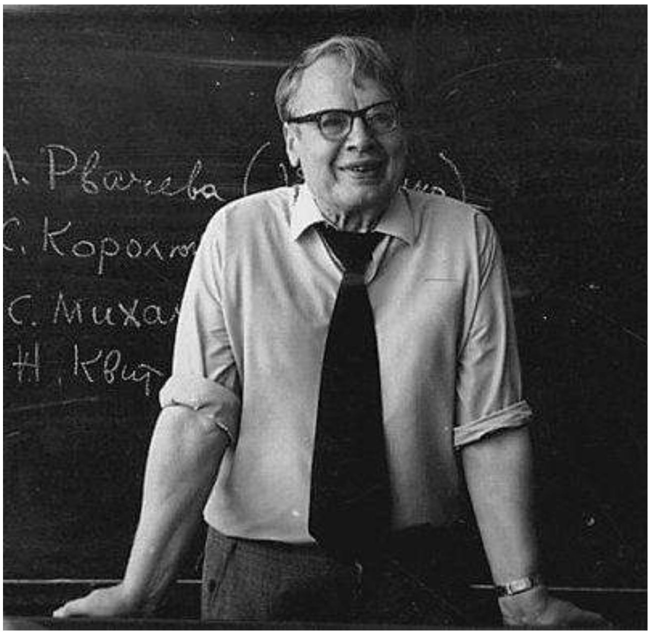
RELIABILITY
RISK ANALYSIS
MAINTENANCE
SAFETY

Vol.14
No.2 (53)
•
June
2019

San Diego

ISSN 1932-2321

Journal of International Group on Reliability



RELIABILITY:

THEORY & APPLICATIONS

Vol.14 No.2 (53),
June, 2019

San Diego
2019

Editorial Board

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

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

Kovalenko, Igor (Ukraine)
Doctor of Sci., Professor, Academician of Academy of Sciences Ukraine, Head of the Mathematical Reliability Dpt. of the V.M. Glushkov Institute of Cybernetics of the Nat. Acad. Scis. Ukraine, Kiev (since July, 1971).
e-mail: kovigo@yandex.ru

Korolyuk, Vladimir (Ukraine)
Doctor of Sci., Professor, Academician of Academy of Sciences Ukraine, Institute of Mathematics, Ukrainian National Academy of Science, Kiev, Ukraine
e-mail: vskorol@yahoo.com

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

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

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

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

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

Nikulin, Mikhail (France)
Doctor of Sci., Professor of statistics, Université Victor Segalen Bordeaux 2, France (Bordeaux, France)
e-mail: mikhail.nikouline@u-bordeaux2.fr

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

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

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

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

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

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

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

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

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

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

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

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

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

Table of Contents

Computer Technology of Formation Control Samples of Fire Safety Rules of Objects Electro Power Systems 10

Farhadzadeh E.M., Muradaliyev A.Z., Ismailova S.M., Yusifli R.F.

The safety of electrical power system facilities is one of the most important characteristics of operational efficiency. The importance of safety is shown, first of all, that their discrepancy to shown requirements results not only in the big material damage, but also to infringement of ecology, a traumatism and destruction of the personnel serving object. Increase of efficiency of the control and the analysis of execution of Safety rules can be reached on the basis of computer technologies, by transition from the qualitative characteristic of safety to quantitative. The initial stage of increase of safety of objects of electro power systems is transition to modelling control samples Fire Safety rules, to a substantiation of volume samples and to documentary acknowledgement of execution of Safety rules.

Events Dependence Used in Reliability Speaks More to Know. Modeling Competing Risks 19

Boyan Dimitrov

In this article we show how some known to us measures of dependence between random events can be easily transferred into measures of local dependence between random variables. This enables everyone to see and visually evaluate the local dependence between uncertain units on every region of their particular values. We believe that the true value of the use of such dependences is in applications on non-numeric variables, as well as in finances and risk studies. We also trust that our approach may give a serious push into the microscopic analysis of the pictures of dependences offered in big data. Numeric and graphical examples should confirm the beauty, simplicity and the utility of this approach, especially in reliability models.

Defuzzification of Healthcare Related Problem by Using $(\boldsymbol{\lambda}, \boldsymbol{\rho})$ Intervalued Trapezoidal Fuzzy Numbers and Their Functional 33

Rajesh Dangwal, Kapil Naithani

In healthcare related problems there exist some medical error due to involvement of human errors and technologies error. In general the available data is not sufficient to assess the clinical process. This study uses level (λ, ρ) intervalued trapezoidal fuzzy numbers and their functional to evaluate the estimate reliability of system in the fuzzy sense, In this paper we uses fault tree diagram of the mixed system (series and parallel system) and we change their crisp probability to trapezoidal fuzzy number and a new approach of functional of fuzzy number has developed and Computed results have been compared with results obtained from other existing techniques.

On Warranty Cost Analysis For a Software Reliability Model Via Phase Type Distribution 46

Y. Sarada, R. Shenbagam

This research work investigates an optimal software release problem via phase type distribution, warranty and risk cost analysis. The inter arrival time of software failure is assumed to be a phase type distribution. The PH-SRM is one of the most flexible models, which overarches the existing non-homogeneous Poisson process (NHPP) models, and can approximate any type of NHPP-based models with high accuracy. Based on the phase type Non- homogeneous Poisson Process (PHNHPP) formulation and using the renewal reward theorem, the long run average cost rate is obtained. As model parameter estimation is an important issue in developing software reliability models, the software failure parameter has been estimated by the moment matching method. Finally, a numerical example is provided to illustrate the theoretical results therein.

A Generalization Of Weibull Distribution 57

Rama Shanker, Kamlesh Kumar Shukla

In this paper, a generalization of Weibull distribution (GWD), which includes Weibull and exponential distributions as special cases, has been proposed and investigated. Its moments, hazard rate function and stochastic ordering have been discussed. The method of maximum likelihood estimation has been discussed for estimating its parameters. The goodness of fit of the proposed distribution has been discussed with a real lifetime dataset and the fit has been found quite satisfactory over some well-known lifetime distributions.

Modelling of an Offline and Online Software for Normalization of Microarray Data of Gene Expression by Perl, Bioperl and PerlTk and Perl-CGI 72

Gaurav Kumar Srivastava, Dr. Santosh Kumar, Dr. Himanshu Pandey

B-Chip Reverence is an online database which isfreely accessible for microarray redundancy removal & normalizationand various data analysis techniques are applied on the data. This software accurately handle the massive amount of data.The growing use of DNA microarrays in biomedical research has led to the proliferation of analysis tools. These software programs address different aspects of analysis (e.g. normalization and clustering within and across individual arrays) as well as extended analysis methods (e.g. clustering, annotation and mining of multiple datasets).After studying all the terms and problems related to Microarray technique, we tried to make an-open and user friendly software to deal with all the problems and to run all the steps of this technique, so that we used Perl & Perl-cgi. perl-cgi stands for Common Gateway Interface, is a standard programming interface between Web servers and external programs. perl-cgi executes external programs on the webserver.

Moving Video Camera Vigilance Using DBSCAN 80

Jagdatta Singh, Gaurav Kumar Srivastava, Himanshu Pandey

The author is trying to develop a model for dynamic or moving video camera vigilance using Density Based Clustering and location sensors. The authors try to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates. The author uses DBSCAN algorithm along with sensor enabled test ground area that keeps the X and Y co-ordinates of the moving objects. The idea here is to capture continuous video of the densest cluster of objects moving together. One practical usage of such system is a wild landscape where groups of animals are moving together to some destination. There will be a somewhat unorganized haphazard movement but we intend to capture only those animals that are greater in number as a group and the camera should move picturing them. This can be achieved by the DBSCAN algorithm

Analysis of MAP/PH/1 retrial queue with constant retrial rate, Bernoulli schedule vacation, Bernoulli feedback, breakdown and Repair 86

G.Ayyappan, R.Gowthami

A retrial queueing model in which the inter arrival times follow Markovian Arrival Process (MAP) and the service times follow phase type distribution is studied. At the end of receiving the service, the customer has two options namely, either he may go to orbit with probability q_1 to get the service again, if he is not satisfied or with probability p_1, he may depart the system. Similarly, at the end of providing service, the server can either opt to take vacation with probability p_2 or be idle with probability q_2. During the busy period, the server may experience breakdown. Both the breakdown times and repair times of the server follow exponential distribution with parameter σ and δ respectively. The resulting QBD process is analysed in the steady state by employing matrix analytic method. The busy period analysis of our model has also been done. Finally, the numerical and graphical illustration of our model has been given.

Computer Technology of Formation Control Samples of Fire Safety Rules of Objects Electro Power Systems

Farhadzadeh E.M., Muradaliyev A.Z., Ismailova S.M., Yusifli R.F.
•

Azerbaijan Scientific-Research and Design-Prospecting
Institute of Energetic AZ1012, Ave. H.Zardabi-94,
e-mail: elmeht@rambler.ru

Abstract

The safety of electrical power system facilities is one of the most important characteristics of operational efficiency. The importance of safety is shown, first of all, that their discrepancy to shown requirements results not only in the big material damage, but also to infringement of ecology, a traumatism and destruction of the personnel serving object. Increase of efficiency of the control and the analysis of execution of Safety rules can be reached on the basis of computer technologies, by transition from the qualitative characteristic of safety to quantitative. The initial stage of increase of safety of objects of electro power systems is transition to modelling control samples Fire Safety rules, to a substantiation of volume samples and to documentary acknowledgement of execution of Safety rules.

Keywords: Safety, object, rules, sample, modelling, the control

I. Introduction

1. Statement of the problem. Along with reliability (dependability, maintainability, durability and persistence) [1] and efficiency, safety is the most important characteristic of the performance of electric power systems (EES).is safety of work. Consequences of absence of safety show not only in the big damage, but also, unfortunately, in infringement of ecology, a traumatism and destruction of the personnel. According to [2] with 2005 on 2015 on objects of an electro power complex of Russia there were 11485 fires on which 242 persons were lost, have was traumatized 472 persons, and the direct material damage has made nearby 3 billion rbl. Of them on transformer substations there were 36,5% of fires.

The control of a condition of safety of objects EPS carried out by a number of the organizations [2] including, annually inspectors of corresponding departments of Management EPS. As a result of check certificates in which the revealed infringements are resulted are made. On the basis of certificates the plan of measures on elimination of these of infringement is made. This system has proved the working capacity. However, in due course, preservation of efficiency of the control becomes more and more problematic. Influence on this system renders:

– unfairly intensive operation of the equipment and devices (further ED) objects service life, which exceeds settlement. In other words, the fact of their ageing insufficiently full considered. In many power supply systems, the relative number of such objects exceeds 50%. Their reliability regularly decreases owing to growth of speed of deterioration;

– in system of maintenance service and scheduled repair, presence at ED residual deterioration is insufficiently considered. That brings mistakes in an estimation of their loading ability and residual service life;
– absence of recommendations on operation ED, which service life, exceeds settlement, corresponding supervising and methodical instructions;
– an insufficient system effectiveness of improvement of professional skill of the personnel. The system of remote training at absence of corresponding methodical instructions lost the efficiency, and traditional methods of improvement of professional skill demanded not only greater expenses, but also did not consider absence of the personnel for business trip on rates;
– increase in distinction in knowledge of young experts of bachelors and necessary in EPS engineers;
– discrepancy of an existing material resources for diagnostics of a technical condition growing old ED to shown requirements;
– consequence of reduction of number of the personnel was sharp decrease in quality of examination;
– absence of computer technologies of the analysis possible discrepancy of a condition of safety of objects to shown requirements. It essentially limited opportunities of the joint analysis of safety of objects, the enterprises and Managements EPS, revealing of characteristic infringements, formation system a plan of measures on increase of safety;
– subjective character of the control safety of objects shown requirements.

In these conditions maintenance of execution of Fire safety Rules (FSR) objects EPS becomes in a number of the major problems EPS.

II. The automated system of the selective control of execution of safety Rules.

Distinction of objects EPS cause distinction of features of maintenance of their safety. Various kinds of safety, the monitoring system of execution and the analysis of results cause multidimensionality of a problem. Below the characteristic of the automated system of forming sample of Rules for check of safety of objects EPS on example FSR is resulted.

2.1. Conditions of application of the automated system

* presence of the list of the enterprises of a power supply system with the instruction; names and type of the enterprises; names and types of objects of the enterprise;

* the approved list of consumers of the analysis of results of the control of execution FSR. To them concern:

** department of the chief engineer of a power supply system. Approves decisions on monitoring procedure of execution FSR of the enterprises and actions on reduction of danger of occurrence of fires by objects of enterprises EPS;

** the head of a department of the control over Management EPS, providing conformity of fire safety of enterprises EPS to shown requirements;

** heads of the enterprises of a power supply system;

** chiefs of objects of the enterprises of a power supply system;

* list FSR reflecting features of objects of enterprises EPS. As an example in table 1 fragments of codes FSR are resulted.

The code Rules in [3] include the following information: the first position corresponds to the name of section FSR; the second and the third - to number of chapter FSR; the fourth and the fifth - to number of item of chapter FSR; the sixth and the seventh - to number of the sub item of item of chapter FSR. For example, the first Rule located in section A, in the first chapter and in first item FSR, looks like A010100 (see tabl.1). The Rule located in section, in the fifth chapter, in 32 items and the fourth sub item designated as B053204.

Table 1. Fragment of codes FSR of transformer substations.

№	Code	№	Code	№	Code	№	Code
1	A010100	*48*	A021300	*95*	Б040400	*142*	Б053204
2	A010200	*49*	A021400	*96*	Б040500	*143*	Б053205
3	A010301	*50*	A021500	*97*	Б040600	*144*	Б053206
4	A010302	*51*	A021600	*98*	Б040700	*145*	Б053207
5	A010303	*52*	A030101	*99*	Б040800	*146*	Б053208
6	A010304	*53*	A030102	*100*	Б040900	*147*	Б053300
7	A010305	*54*	A030103	*101*	Б041000	*148*	E140100

The list of results of the control of execution FSR of names of sections used at the analysis and chapters of the collection of Rules in [3] are resulted in table 2.

Table 2. Data on sections and chapters FSR.

Classifiers FPR	The name of sections and chapters	Code	Distribution samples
Section A	General provisions	A	7
Chapter 1	Organizational requirements of fire safety	01	1
Chapter 2	The basic requirements to the organization of preparation of the personnel	02	2
Chapter 3	The basic documentation on fire safety	03	4
Section B	The basic requirements of fire safety at the enterprises of branch	B	5
Chapter 4	The maintenance of territory	04	4
Chapter 5	The maintenance of buildings and constructions	05	1
Section E	Switching centre	E	5
Chapter 14	Switching centre of power stations and substations	14	1
Chapter 15	Cable facilities	15	1
Chapter 16	Power transformers	16	1
Chapter 17	Storage installations	17	2
Section Z	Repair and reconstruction of the equipment	Z	0
Chapter 21	Fire safety at repair and reconstruction of the process equipment	21	0
Chapter 22	Fire safety at carrying out welding and other inflammable works	22	0
Section I	Fire-prevention water supply and fire extinguishing means	I	8
Chapter 23	Fire-prevention water supply	23	6
Chapter 24	Installations of detection and fire extinguishing	24	0
Chapter 25	Fire extinguishing means of fires	25	2
Section K	The order of the organization of suppression of fires on the equipment of power objects energized up to 0,4 кV	K	0
Chapter 26	General provisions	26	0
Chapter 27	Safety requirements at performance of works on fire extinguishing	27	0
Chapter 28	Actions of the personnel at occurrence of a fire	28	0
Section P	Applications	P	0
Numbers	1÷13	29	0

* presence of the approved plan of the control of conformity of fire safety of enterprises EPS to shown requirements;
* the approved list of the documents confirming full executions FSR;
* an opportunity of annual improvement of professional skill of executives of requirements of maintenance of fire safety;
* presence at the enterprises of the complete set of the documents confirming executions FPR.

2.2. A way of elimination of subjective character of sample controllable FSR

As general number FSR of concrete objects together with sub items and appendices is estimated in hundreds [3], the control of execution of all Rules demands a lot of time, is labour-consuming, and considering annual periodicity of the control - and is inexpedient. Labour input of the control caused by a wide spectrum of means of maintenance of fire safety when the inspector in many of them is incompetent. If besides to consider, that in some cases the control is carried out by the expert on suppression of fires which is not familiar with fire-dangerous ED EPS, labour input of the objective control essentially increases. Therefore, in practice the control of execution FSR carried out under preliminary made list of questions. Their number depending on type of object varies within the limits of (15÷25). As an example in table 2 distribution of traditional control sample FSR in volume n_s=18 under chapters of the collection of Rules [3] is resulted also. Asymmetry of distribution pays attention. Certainly, this list has subjective character, as well as process of the control. And use of identical questions for all same objects of the enterprise, and is frequent - and all same enterprises essentially deforms the characteristic of the revealed infringements, does its unpresentable.

Increase of efficiency of the control of performance FSR, as well as the control of many other things of Rules, reached application of computer technologies. In particular, subjective character of the control and its uniformity it is possible to avoid in the known way - modelling casual sample of Rules from set FSR. For what, proceeding from general number FSR for the set object equal K_i^{ob} where $i = 1, m_{ob}$; m_{ob} - the number of the same objects of enterprises EPS, "is played" $K_{s,i}^{ob} << K_i^{ob}$ serial numbers of Rules, where $K_{s,i}^{ob}$- volume of sample (s) Rules i- th type of object, under the formula:

$$N_j^{ob} = \mathrm{Abc}\left[\xi_j \cdot K_i^{ob}\right] + 1 \quad (1)$$

where: $j = 1, K_{s,i}^{ob}$; N_j^{ob} - casual and j- th realization of serial number FSR; ξ_j- j- th realization of a random variable ξ corresponding uniform distribution in an interval [0,1].

Random variables ξ modelled by subroutine RANDU(ξ). At formation of control sample corrected manually, for example, at absence of the automated system of formation of control sample and synthesis of results of the control, it is possible to take advantage of function SLCIK in system Excel or (for an example) the table of random numbers [4]. For example, if $\xi_j = 0,7213$, and $K_i^{ob} = 566$, number next (j- th) Rules will be equal: (0,7213·566 =408,3. Hence, N_j^{ob} =409. It is necessary to note, that at modelling control sample FPR the opportunity of equality of serial numbers of realizations of separate Rules not excluded. Such it is possible, if size $\left|\xi_i - \xi_j\right| \cdot M_i < 1$, where ξ_i and ξ_j - i- th and j- th realizations ξ. As presence of identical serial numbers of control sample is deprived physical sense (it means that execution of some Rule should supervised twice), in algorithm of calculation check of each realization on individuality provided. If the serial number of the Rule in sample repeats, this realization excluded from consideration. As FSR in the received sample in volume $K_{s,i}^{ob}$ place in the any order for convenience of recognition of Rules expediently to range their serial numbers N_j^{ob} in ascending order together with a code of each Rule.

2.3. The importance of scoping of sample

One of the basic questions at modelling serial numbers FSR is the volume of sample. Unfortunately, known methods of calculation of the minimal volume of sample of general set as FSR there is no general set of random variables here cannot used. However, for FSR many features of traditional calculation of volume of sample are characteristic. Here as [5]:
– the increase in volume of sample demands increase in time of the control, and reduction - leads to increase in risk of the erroneous decision;
– heterogeneity even the big sample does not guarantee success;
– extrapolation of results of the control over set FSR on its sections is erroneous;
– as a whole, scoping of sample is a sequence of the big number of compromises.

As an example in table 3 list FSR for control sample $K_{s,i}^{ob} = 20$ is resulted.

Table 3. Results of modelling and ranging FSR.

№	Results of modelling			Results of ranging	
	Random variables	Serial numbers FPR	Coordinates FPR	Serial numbers FPR	Coordinates FPR
1	0,1009	45	A021000	37	A020502
2	0,3754	165	E150300	44	A020900
3	0,0842	37	A020502	45	A021000
4	0,9901	433	П291107	52	A030101
5	0,128	56	A030105	56	A030105
6	0,6606	289	З211401	136	Б052900
7	0,3106	136	Б052900	165	E150300
8	0,8526	373	И240400	278	З210502
9	0,6357	278	З210502	287	З211200
10	0,7379	323	З221500	289	З211401
11	0,9852	431	П291105	306	З220400
12	0,118	52	A030101	323	З221500
13	0,8345	365	И232403	325	З221701
14	0,8868	388	И241900	351	И231200
15	0,9959	436	П291110	365	И232403
16	0,6548	287	З211200	373	И240400
17	0,8012	351	И231200	388	И241900
18	0,7435	325	З221701	431	П291105
19	0,6991	306	З220400	433	П291107
20	0,0989	44	A020900	436	П291110

Data of modelling allow draw the important conclusion: casual character small samples causes non-uniformity of their distribution under chapters. We based on data of table 4, where the number of Rules on sections and chapters of collection FSR is resulted, it is possible to conclude:
– casual samples of Rules insufficiently full reflect the maintenance of chapters FSR;
– take place the chapter of collection FSR, which Rules are not presented absolutely not in sample. It A01, B05, E14, E16, E17, C18, C19, C20, I25.
– number of controllable Rules in each chapter insufficiently full characterizes volume of these chapters. Discrepancy of number of Rules of separate chapters and numbers of controllable Rules under these chapters observed. For example, execution of 46 Rules of the fifth chapter of section B is represented sample of 4 Rules. At the same time, 37 Rules 22 chapters of section 3 are not presented absolutely not in control sample. To lacks this way of a finding of control sample concerns and the appointed volume of sample.

Scoping samples from the sets, which are not concerning a class general, represents a complex and unresolved problem [5]. Now the volume of sample defined proceeding from indirect restrictions. In our case is an opportunity to estimate fire safety of objects in an interval of duration of business trip of the inspector. But also the appointed volume of sample should cover, at least, all of chapter FSR. For example, for samples from general set in sociology a popular belief, that the volume of sample should be equal 10% from general set which is estimated in millions [6].

Table 4. Illustration advantages of block modelling of sample FPR.

Serial numbers		Code of sections and chapters	Number FSR on sections and chapters	Modelling of sample	
Sections	Chapter			On set FSR	Under chapters FSR
	1	01	29	0	3
	2	02	22	3	2
	3	03	40	2	4
1		*A*	*91*	*5*	*9*
	4	04	10	1	1
	5	05	46	0	5
2		*B*	*56*	*1*	*6*
	6	14	15	0	1
	7	15	29	1	3
	8	16	20	0	2
	9	17	10	0	1
3		*E*	*74*	*1*	*7*
	10	18	22	0	2
	11	19	17	0	1
	12	20	14	0	1
4		*C*	*51*	*0*	*4*
	13	21	29	3	3
	14	22	37	3	4
5		*Z*	*66*	*6*	*7*
	15	23	31	2	3
	16	24	20	2	2
	17	25	16	0	2
6		*I*	*67*	*4*	*7*
	18	29	33	3	3
7		*P*	*33*	*3*	*3*
Всего			**438**	**20**	**43**

2.4. A block method of formation of volume of sample

It is offered to spend liquidation of discrepancies of result of modelling of control sample of Rules not under the formula (1) for $K^{ob}_{s,i}$, and under chapters FSR, i.e. to apply a "block" way of modelling. Thus, the minimal number of controllable Rules in each chapter accepted equal to unit. Calculation of volume of sample in each chapter spent as follows:

– we define the chapter (c) with the minimal number of Rules, we establish for this chapter number of controllable Rules equal to unit;

– the number of controllable Rules of other chapters (see table 4), and them for considered objects - 18, is defined by the attitude of number of Rules in each chapter to the minimal number of Rules of one of chapters $K^{ob}_{i,min}$ with the standard order of their rounding off. For example, for 22 chapters $K_i^{ob} = 37$. As $K_{i,min}^{ob} = 10$, the number of realizations of controllable sample from 22 chapters will be equal to four.

Thus it is obvious, that the volume of sample of the control of performance FSR will be approximately equal to the attitude of number of Rules of set for considered (i) object K_i^{ob} and the minimal number of Rules of each of chapters $K_{i,min}^{ob}$.

According to results of the calculations resulted in table 4 size $K_i^{ob} = 438$ $K_{i,min}^{ob} = 10$, and volume of sample $n_i = 43$. Thus, all the lacks of formation noted above samples eliminated. The increase in volume of control sample FSR with 20 up to 43 can be eliminated doubly: transition from one control over year to two, i.e. transition to half year an interval of the control and change of structure of the control.

The block diagram of algorithm of block modelling control samples FSR is resulted on fig.1.

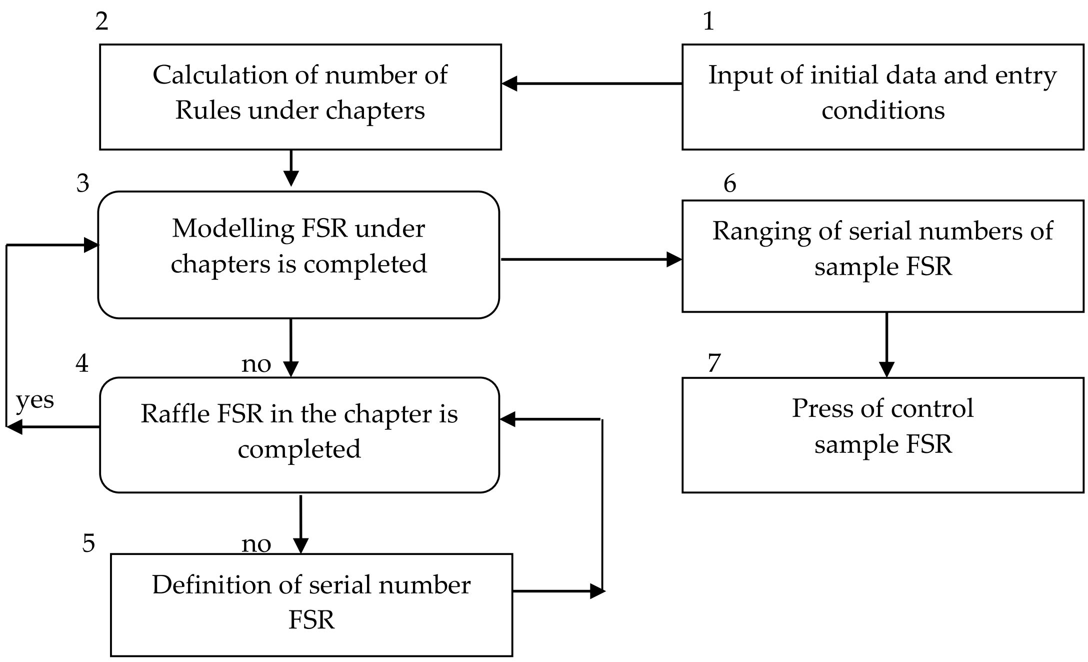


Fig. 1. The integrated block diagram of algorithm of modelling of sample FSR.

III. The organization of the selective control of execution FSR

The following sequence of actions is provided:

3.1. According to the approved schedule of carrying out of the selective control of execution FSR at the enterprises for two weeks up to the appointed term, the department of the control of execution FSR of management EPS for all objects of the enterprise prepares individual list FSR which control of execution is a subject to check. Individual lists FSR affirm the Head of a department, and its copy in an electronic kind and on paper, carriers send to the Director.

3.2. Within two weeks the inspector on fire safety of objects of the enterprise together with Chiefs of objects prepare copies of the documents confirming performance control FSR for each controllable object.

3.3. Results of the control will be coordinated with Director of the enterprise;

3.4. Certificates of the control, together with confirming performance FSR of documents affirm the head of a department on fire safety;

3.5. All these documents placed in archive of a database of the automated monitoring system of execution FSR.

In table 5 results of block modelling control samples FSR for five objects are resulted. Results of the analysis of an opportunity of repetition control FSR at modelling control samples the same objects allow to conclude, that repeated FSR (in tab. 5 are allocated by a font) make about 10%.

Table 5. An example of questions for control of substation

	№	Substation									
		SS № 1		SS № 2		SS № 3		SS № 4		SS № 5	
First half of year	1	37	A020502	5	A010303	4	A010302	23	A010800	90	A031502
	2	***44***	***A020900***	8	A010306	16	A010502	***44***	***A020900***	100	B040900
	3	45	A021000	33	A020302	23	A010800	69	A030503	***154***	***E140700***
	4	***52***	***A030101***	***44***	***A020900***	62	A030402	92	B040100	192	E160101
	5	56	A030105	***52***	***A030101***	78	A030604	99	B040800	196	E160200
	6	136	B052900	131	B052400	79	A030700	108	B050503	228	J180500
	7	165	E150300	141	B053203	104	B050300	111	B050506	230	J180700
	8	278	Z210502	146	B053208	128	B052100	119	B051200	236	J181301
	9	287	Z211200	191	E152700	138	B053100	127	B052000	299	Z211505
	10	289	Z211401	***205***	***E161100***	149	E140200	151	E140400	300	Z211506
	11	306	Z220400	211	E161700	***154***	***E140700***	176	E151400	316	Z220906
	12	323	Z221500	***221***	***E171000***	168	E150600	187	E152300	336	Z221900
	13	***325***	***Z221701***	232	J180900	185	E152100	200	E160600	340	I230200
	14	***351***	***I231200***	252	J190900	203	E160900	***205***	***E161100***	348	I230902
	15	352	I231300	274	Z210200	214	E170300	210	E161600	***351***	***I231200***
	16	365	I232403	283	Z210800	218	E170700	***221***	***E171000***	357	I231800
	17	373	I240400	284	Z210900	237	J181302	222	J180100	373	I240400
	18	388	I241900	302	Z220100	293	Z211405	227	J180400	382	I241300
	19	***400***	***I251100***	304	Z220202	294	Z211406	318	Z221000	***400***	***I251100***
	20	431	P291105	350	I231100	322	Z221400	393	I250400	409	P290202
	21	433	P291107	***400***	***I251100***	*325*	*Z221701*	394	I250500	430	P291104
	22	436	P291110	428	P291102	430	P291104	411	P290204	435	P291109

This information is rather useful, since repeated FSR in control samples allow to judge a degree of identity of the approach to maintenance of performance FSR on all same objects of enterprise EPS. If to consider, that maintenance of performance FSR on the same objects of the enterprise is carried out on-line, existing lacks of execution FSR for all same objects of the enterprise in many respects are similar - it is possible to approve, that set control samples the same objects will be completely characterized fire safety as the enterprises as a whole, and each of objects.

Conclusion

1. Perfection of the monitoring system of execution of Fire prevention rules first of all can be reached by transition to formation of control sample of Rules on the COMPUTER:

2. Modelling of control sample of Rules on their set for the same objects has a number of essential lacks, basic of which are:
   - subjective character of purpose of volume of the control of sample;
   - casual character of sample FSR does not reflect distinction of number of Rules on sections and chapters;
3. The block way of modelling of control sample under chapters is recommended.
4. The probability of recurrence control FSR in the list of Rules of some objects does not exceed units of percent.

References

[1] GOST R.27.002-2009. Reliability in technics. The term and definitions IEC 60050 (191): 1990 and 12. (NEQ.), Standartinform, 2011.
[2] Fires and fire safety in 2005-2015. The statistical collection, statistics of fires and their consequences / under Edition Matjushin A.V., M.: Scientific research institute the Ministry of Emergency Measures of Russia, 2006-2016, 124 p.
[3] LD 153-34.0-03.301-00 Fire prevention rules
[4] Shor Y.B., Kuzmin T.I. Table for the analysis and the control of reliability. M.: «Soviet radio» 1968, 254 p.
[5] Boloev E.V., Dubiskiy M.A. Safety of electro power systems. Bulletin SSTU, 2011, 4 (62), release 4.
[6] Orlov A.S. Eagles statistics. M.: Publishing house "Examination", 2004, 483 p.

Events Dependence Used in Reliability Speaks More to Know. Modeling Competing Risks

Boyan Dimitrov

•

Department of Mathematics, Kettering University,
1700 West University Avenue, Flint, Michigan, USA
email: bdimitro@kettering.edu

Abstract

In this article we show how some known to us measures of dependence between random events can be easily transferred into measures of local dependence between random variables. This enables everyone to see and visually evaluate the local dependence between uncertain units on every region of their particular values. We believe that the true value of the use of such dependences is in applications on non-numeric variables, as well as in finances and risk studies. We also trust that our approach may give a serious push into the microscopic analysis of the pictures of dependences offered in big data. Numeric and graphical examples should confirm the beauty, simplicity and the utility of this approach, especially in reliability models.

Keywords: local measures of dependence, local regression coefficienstins, local correlation, mapping the local dependence, big data tools, microscopic analysis of dependence in reliability models - graphic illustrations

1. Introduction

The big data files contain a number of simultaneous multi-dimensional observations. This fact offers plenty of opportunities for establishing possible dependences between observed variables. Most of these dependences will be of global nature. However, there exist (or can be created) techniques to take a microscopic look on more details into it. In this article we want to show the ideas of these microscopic looks.

The concepts of measuring dependence should start from the very roots of Probability Theory. Independence for random events is introduced simultaneously with conditional probability. Where independence does not hold, events are dependent. Further, the focus in text-books is on the independence. No text-books usually discus what to do if events are dependent. However, there are ways to go deeply in the analysis of dependence, to see some detailed pictures, and use it later in the studies of random variables. This question is discussed in our previous articles (Dimitrov 2010, 2015) and more (Esa-Dimitrov 2013, 2017). Some particular situations are analyzed in Dimitrov and Esa 2014 and Esa, Dimitrov 2017. Applications in study of politics are used in Esa, Dimitrov 2013. We refer to these articles for making a quick passage to the essentials.

First we notice here that the most informative measures of dependence between random events are the two *regression coefficients*. Their definition is given here:

Definition1. Regression coefficient $R_B(A)$ of the event A with respect to the event B is called the difference between the conditional probability for the event A given the event B, and the conditional probability for the event A given the complementary event $\overline{B}$, namely

$$R_B(A) = P(A|B) - P(A|\overline{B}).$$

This *measure of the dependence of the event A on the event B,* is *directed dependence*.

The regression coefficient $R_A(B)$ of the event B with respect to the event A is defined analogously.

From the many interesting properties of the regression coefficients we would like to point out here just few:

(R1) The equality to zero $R_B(A) = R_A(B) = 0$ takes place if and only if the two events are independent.

(R2) The regression coefficients $R_B(A)$ and $R_A(B)$ are numbers with equal signs and this is the sign of their connection $\delta(A,B) = P(A \cap B) - P(A)P(B)$. The relationships

$$R_B(A) = \frac{P(A\cap B) - P(A)P(B)}{P(B)[1-P(B)]}, \textit{ and } R_A(B) = \frac{P(A\cap B) - P(A)P(B)}{P(A)[1-P(A)]}.$$

The numerical values of $R_B(A)$ and $R_A(B)$ may not always be equal. There exists an asymmetry in the dependence between random events, and this reflects the nature of real life.

(R3) The regression coefficients $R_B(A)$ and $R_A(B)$ are numbers between –1 and 1, i.e. they satisfy the inequalities

$$-1 \le R_B(A) \le 1; \qquad -1 \le R_A(B) \le 1.$$

(R4.1) The equality $R_B(A) = 1$ holds only when the random event A coincides with (or is equivalent to) the event B. Then it is also valid the equality $R_A(B) = 1$;

(R4.2) The equality $R_B(A) = -1$ holds only when the random event A coincides with (or is equivalent to) the event $\overline{B}$ - the complement of the event B. Then it is also valid $R_A(B) = -1$, and respectively $\overline{A} = B$.

We interpret the properties (r4) of the regression coefficients in the following way: As closer is the numerical value of $R_B(A)$ to 1, "as denser inside within each other are the events A and B, considered as sets of outcomes of the experiment". In a similar way we interpret also the negative values of the regression coefficient.

There is a symmetric measure of dependence between random events, and this is their coefficient of correlation.

Definition 2. Correlation coefficient between two events A and B we call the number

$$\rho_{A,B} = \pm\sqrt{R_B(A)\cdot R_A(B)},$$

where the sign, plus or minus, is the sign of the either of the two regression coefficients.

Remark. The correlation coefficient $\rho_{A,B}$ between the events A and B equals to the formal correlation coefficient ρ_{I_A,I_B} between the random variables I_A and I_B, the indicators of the two random events A and B.

The correlation coefficient $\rho_{A,B}$ between two random events is symmetric, is located between the numbers $R_B(A)$ and $R_A(B)$.

The following statements hold:

ϱ1. $\rho_{A,B} = 0$ holds if and only if the two events A and B are independent. The use of the numerical values of the correlation coefficient is similar to the use of the two regression coefficients. As closer is $\rho_{A,B}$ located to the zero, as "closer" to the independence are the two events A and B.

For random variables similar statement is not true. The equality to zero of their mutual correlation coefficient does not mean independence

ϱ2. The correlation coefficient $\rho_{A,B}$ always is a number between –1 and +1, i.e.

$$-1 \leq \rho_{A,B} \leq 1.$$

ϱ2.1. The equality $\rho_{A,B}$ = 1 holds if and only if the events A and B are equivalent, i.e. when A = B.

ϱ2.2. The equality $\rho_{A,B}$ = - 1 holds if and only if the events A and $\overline{B}$ are equivalent, i.e. when A = $\overline{B}$.

As closer is $\rho_{A,B}$ to the number 1, as "more dense one within the other" are the events A and B, and when $\rho_{A,B}$ = 1, the two events coincide (are equivalent).

As closer is $\rho_{A,B}$ to the number -1, as "denser one within the other" are the events A and $\overline{B}$, and when $\rho_{A,B}$ = - 1, the two events coincide (are equivalent). Denser one within the other are then the events $\overline{A}$ and B.

2. The transfer rules

The above measures allow studying the behavior of interaction between any pair of numeric r.v.'s (X,Y) throughout the sample space, and better understanding and use of dependence.

Let the joint cumulative distribution function (c.d.f.) of the pair (X,Y) be $F(x,y)=P(X\leq x,\ Y\leq y)$, and marginals $F(x) = P(X \leq x)$, $G(y)=P(Y\leq y)$. Let introduce the events

$$A_x=\{x\leq X\leq x+\Delta_1 x\}; \qquad B_y=\{y\leq Y\leq y+\Delta_2 y\}, \text{ for any } x, y \in (-\infty, \infty).$$

Then the measures of dependence between events A_x and B_y turn into *a measure of local dependence between the pair of r.v.'s X and Y on the rectangle* $D=[x, x+\Delta_1 x]\times[y, y+\Delta_2 y]$. Naturally, they can be named and calculated as follows:

Regression coefficient of X with respect to Y, and of Y with respect to X on the rectangle $[x, x+\Delta_1 x]\times[y, y+\Delta_2 y]$. By the use of Definition 1 we get

$$R_Y((X,Y)\in D)=\frac{\Delta_D F(x,y)-[F(x+\Delta_1 x)-F(x)][G(y+\Delta_2 y)-G(y)]}{[F(x+\Delta_1 x)-F(x)]\{1-[F(x+\Delta_1 x)-F(x)]\}}.$$

Here $\Delta_D F(x,y)$ denotes the two dimensional finite difference for the function $F(x,y)$ on rectangle $D=[x, x+\Delta_1 x]\times[y, y+\Delta_2 y]$. Namely

$$\Delta_D F(x,y) = F(x+\Delta_1 x, y+\Delta_2 y)- F(x+\Delta_1 x, y)- F(x, y+\Delta_2 y)+ F(x, y).$$

In an analogous way is defined $\rho_X((X,Y)\in D)$. Just denominator in the above expression is changed respectively.

Correlation coefficient $\varrho_Y((X,Y)\in D)$ between the r.v.s X and Y on rectangle $D=[x, x+\Delta_1 x]\times[y, y+\Delta_2 y]$ can be presented in similar way by the use of Definition 2. We omit detailed expressions as something obvious.

It seems easier to find out the local dependence at a value $(X=i,\ Y=j)$ for a pair of discretely distributed r.v. (X,Y). Regression coefficient of X with respect to Y, and of Y with respect to X at a value ($X=i$, $Y=j$) is determined by the rule

$$R_Y(X=i,Y=j)=\frac{P(X=i,Y=j)-P(X=i)P(Y=j)}{P(X=i)[1-P(X=i)]}=\frac{p(i,j)-p_{i.}p_{.j}}{p_{i.}(1-p_{i.})}.$$

Similarly, you can get that the local correlation coefficient between the values of the two r.v.'s (X,Y) is given by

$$\rho_{X,Y}(X=i,Y=j)=\frac{p(i,j)-p_{i.}p_{.j}}{\sqrt{p_{i.}(1-p_{i.})}\sqrt{p_{.j}(1-p_{.j})}}.$$

Using these rules one can see and visualize the local dependence between every pair of two r.v.'s with given joint distribution.

This ends our theoretical background of the local dependence structural study. Next we illustrate its application on qualitative and quantitative probability models.

3. Illustrations

3.1 Reliability systems

In this case let us consider the two traditional systems of independent components, the system in series and the system in parallel. We want to study how the regression coefficients of a component with respect to the system, and vise verse, regression coefficient of the system with respect to a component change in time during the work of the system. For simplicity consider system of just two components, since considering one component, everything else can be considered as a second component. Results of the studies are shown next.

3.1. A system in series. Assume both components have live times exponentially distributed with parameters λ_1 and λ_2. Then the reliability function at any time instant t (this is the event B) equals r(t)= $e^{-(\lambda_1+\lambda_2)t}$, and the probability that component 1 functions (this is the event A) is $e^{-\lambda_1 t}$. The regression coefficient of the system with respect to component 1 is then

$$R_1(S)=\frac{r(t)-r(t)e^{-\lambda_1 t}}{e^{-\lambda_1 t}(1-e^{-\lambda_1 t})}=e^{-\lambda_2 t}.$$

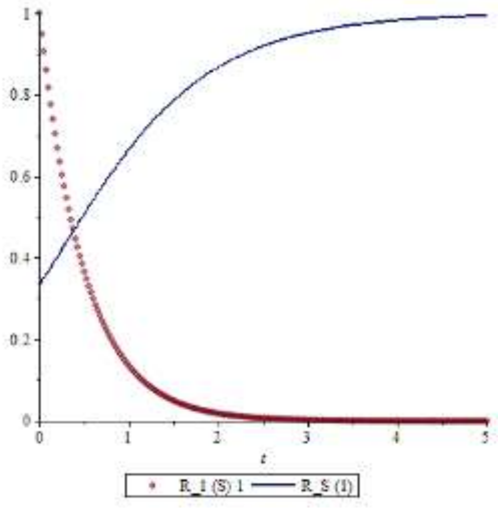
Analogously we evaluate the regression coefficient of the component 1 with respect to the system at time t. It is given by the relation

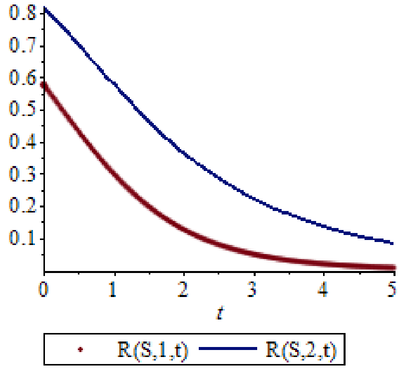
$$R_S(1)=\frac{r(t)-r(t)e^{-\lambda_1 t}}{r(t)[1-r(t)]}=\frac{1-e^{-\lambda_1 t}}{1-e^{-(\lambda_1+\lambda_2)t}}.$$

And the correlation coefficient between system reliability and the component reliability are changing during the time according to the relations

$$\rho_{s,1}(t)=\sqrt{\frac{e^{-\lambda_2 t}(1-e^{-\lambda_1 t})}{1-e^{-(\lambda_1+\lambda_2)t}}};\quad \rho_{s,2}(t)=\sqrt{\frac{e^{-\lambda_1 t}(1-e^{-\lambda_2 t})}{1-e^{-(\lambda_1+\lambda_2)t}}}.$$

Notice that all dependences are positive. Graphs of these functions of local dependence in time for λ_1=1 and λ_2=2 are shown on next figures.





We observe that the system reliability local correlation measures of dependence is decreasing to 0 for both components, but is higher with the weakest component 2, when the time increases. In the same time the regression coefficients between the system and the strongest component behave different: Local dependence $R_1(S)$ approaches 0 with the time (like system becomes independent on component 1with the growth of the time) when the local dependence $R_S(1)$ of strongest component 1 on the system reliability approaches 1 with the growth of the time.

3.2. System in parallel. Assume again both components have live times exponentially distributed with parameters λ_1 and λ_2. Then the reliability function at any time instant t (this is the event B) equals r(t)=$1-(1-e^{-\lambda_1 t})(1-e^{-\lambda_2 t})$, and the probability that component 1 functions (this is the event *A*) is $e^{-\lambda_1 t}$. Applying the rules we obtain:

The regression coefficient of the system with respect to component 1 is then

$$R_1(S) = \frac{r(t) - r(t)e^{-\lambda_1 t}}{e^{-\lambda_1 t}(1-e^{-\lambda_1 t})} = 1 - e^{-\lambda_2 t}.$$

Analogously we evaluate the regression coefficient of the component 1 with respect to the system at time t. It is given by the relation

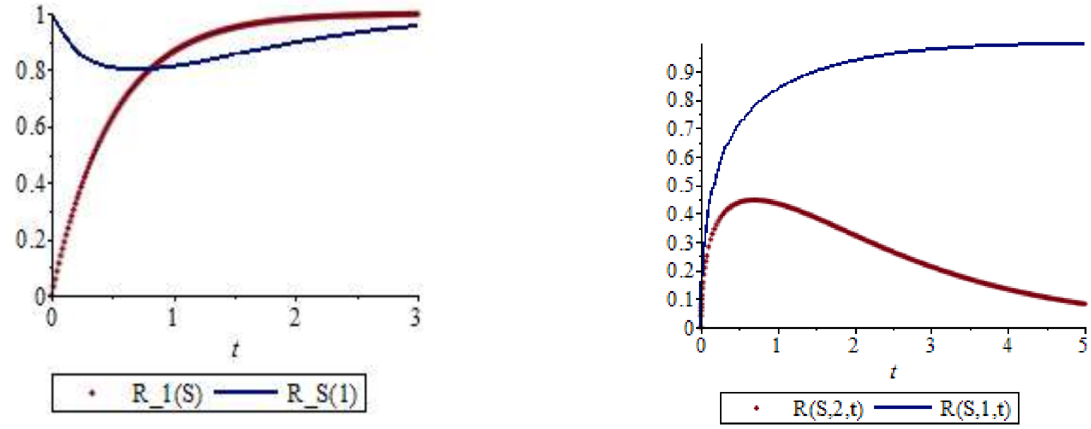
$$R_S(1) = \frac{r(t) - r(t)e^{-\lambda_1 t}}{r(t)[1-r(t)]} = \frac{e^{-\lambda_1 t}}{1-(1-e^{-\lambda_1 t})(1-e^{-\lambda_2 t})}.$$

And the correlation coefficient between system reliability and the component reliability are changing during the time according to the relations

$$\rho_{s,1}(t) = \sqrt{\frac{e^{-\lambda_1 t}(1-e^{-\lambda_2 t})}{1-(1-e^{-\lambda_1 t})(1-e^{-\lambda_2 t})}}; \quad \rho_{s,2}(t) = \sqrt{\frac{e^{-\lambda_2 t}(1-e^{-\lambda_1 t})}{1-(1-e^{-\lambda_1 t})(1-e^{-\lambda_2 t})}}.$$

Notice that all dependences are positive. Graphs of these functions of local dependence in time for λ_1=1 and λ_2=2 are shown on next figures. First one represents the two regression coefficients superimposed on the same graph, and he second represents the two correlation coefficients.

We see that the system reliability local correlation measure of dependence is approaching 1 with the strongest component 1, and approaches 0 with the weakest component when the time increases. In the same time the both regression coefficient between the system and the strongest component approach 1 with the growth of the time.



3.3. Categorical variables

The most interesting and valuable applications in the Big Data analysis we see in the analysis of local dependences between non-numeric vs non numeric variables, as well as between non-numeric vs numeric variables. Since analysis in this kind of studies is (according to us obviously) too similar, we recommend here as an example of local dependence between categories of two non-numeric random variables. It is just an illustration of the proposed measures of dependence between random events. We analyze here an example from the book of Alan Agresti, (2006). You can see this illustration in the work of Dimitrov, 2010.

3.4. A challenging idea in modeling dependent variables

Modeling dependence in multivariate distributions always has been and still is a hot topic in applied probability, statistics and risk studies. One of the most popular approach in modeling dependence is known as Farlie–Gumbel–Morgenstern dependence model. It is using a construction of bivariate distributions as a mixture of two or more marginal distributions. The main disadvantage of this approach is that it produces multivariate distributions with limited magnitude of the correlation coefficient ρ_{XY}. Original construction gives ρ_{XY} within [-.32, .32]. Some generalizations lately (Bekrizadeh et al, 2012) expanded this range to [-.5, +.43]. Other approaches based on copula constructions (Joe, 1997, Nelsen 2006) offer constructions for dependent multivariate distributions with desired marginal. In most of these constructions is used mostly analytical instrumentation where one can get the goal, but loses the meaning.

In this subsection we offer a construction which is based on the dependence between the two components of the random vector due to the presence of a common random component in each. In our opinion, such models are of interest in reliability and risk modeling where competitive risks are presented and have realistic meaning. And each risk is presented by a r.v. kind of independent on the others. We illustrate this approach on a very particular bivariate dependence, where components are indicator variables.

Let U, V and W be independent one dimensional r.v. Consider the following constructions:

A) $X = \min(U, W)$; $Y = \min(V, W)$, and the pair (X, Y);

B) $X = \max(U, W)$; $Y = \max(V, W)$, and the pair (X, Y);

C) $X = \min(U, W)$; $Y = \max(V, W)$, and the pair (X, Y);

D) $X = U + W$; $Y = V + W$, and the pair (X, Y).

Other algebraic operations also may be used in similar constructions. Obviously, the components of each pair are dependent due to the presence of one and the same component W in both. The good thing here is that we see the interaction between *X*, and *Y*. And also, here one may use any distributions or the original risks *U*, *V* and *W*. Our goals here are to find the correlation coefficients in each of the above 4 constructions, and also to investigate the local correlation structure between *X*, and *Y* in the light of the proposed measures for the strength of local dependence explored recently (Dimitrov 2010, 2015). Actually, we will use the measure $\rho_{A,B}$ defined in Definition 2.We start on the grass roots, considering the examples when *U*, *V* and *W* have the simplest Bernoulli distribution i=1,2,3, or have the Uniform distributions on [0, 1].

U, V, W	0	1
$f_i(.)$	$q_i = 1 - p_i$	p_i

Everyone knows that the expected values and standard deviation of a Bernoulli distributed r.v. are E(*U*) = *p*, and $\sigma_U = \sqrt{pq}$.

3.4.1 Minimum-Minimum competing risks

Elementary combinatorial considerations will convince you, that the joint distribution of the random vector (*X*, *Y*) is presented by the table

Table 1

X \ Y	0	1	$f_X(.)$
0	$1 - p_3(p_1+p_2-p_1p_2)$	$q_1p_2p_3$	$1 - p_1p_3$
1	$p_1q_2p_3$	$p_1p_2p_3$	p_1p_3
$f_Y(.)$	$1 - p_2p_3$	p_2p_3	1

On the margins are the marginal distributions of the components *X*, and *Y*, and each of it is also a Bernoulli distributed r.v. This fact will simplify your calculation of the correlation coefficient ρ_{XY}, using the short cut rule

$$\rho_{XY} = [E(XY) - E(X)E(Y)]/ [\sigma_X \, \sigma_Y],$$

and the results above about Bernoulli distributed r.v. After several algebraic manipulations we arrive to the expression

$$\rho_{XY} = q_3 \sqrt{\frac{p_1 p_2}{(1 - p_1 p_3)(1 - p_2 p_3)}} .$$

A brief analysis of this expression shows, that this correlation coefficient can take any value between 0 (when q_3 is close to 0, and p_1, p_2 are small), and 1 (when q_3 is close to 1, and so are p_1, p_2). Hence, pending on probabilities p_1, p_2 and p_3, any correlation between *X* and *Y* is feasible.

In particular if *U*, *V* and *W* are equally distributed, then

$$\rho_{XY} = p/(1+p).$$

But this time the correlation coefficient may take values only between 0 and 0.5. Of course, you may get negative correlations of same size if change second component *Y* by its negative –*Y*.

Local dependence magnitudes.

Now let see the strength of dependence of the event {*Y*=0} with respect to the event {*X*=0}. It means that we may predict the event {*Y*=0} if we know that it occurred {*X*=0}, by making use of relations above. First determine the local regression coefficient using Definition 1 and data in Table 1. We get

$$R_{X=0}(Y=0) = \frac{p_2 q_3}{1-p_1 p_3}.$$

Further considerations show, that if we know individual parameters of variables *U,V* and *W*, and know that event (*X*=0) occurred, then our prediction of event (*Y*=0) will be given by the posterior probability

$$P(Y=0\,|\,X=0) = 1 - q_1 p_2 p_3/[1 - p_1 p_3],$$

Due to the positive dependence, the prediction probability increases with the information of the known value of component *X*.

Not going into detailed explanations, we get

$$R_{X=0}(Y=1) = -\frac{p_2 q_3}{1-p_1 p_3};\ \ P(Y=1\,|\,X=0) = \mathbf{q_1 p_2 p_3/(1-p_1 p_3)}$$

As we expect, we have $P(Y=0\,|\,X=0) + P(Y=1\,|\,X=0) = 1$.

Let see the local strength of dependence of the event {*Y*=0} with respect to the event {*X*=1}. It means that we may predict the event {*Y*=0} if we know that it occurred {*X*=1}, by making use of Definition 2. First determine the local regression coefficient using Definition 1 and data in Table 1. We get

$$Rr_{X=1}(Y=0) = \frac{-p_2 q_3}{1-p_1 p_3}.$$

This negative regression coefficient indicates that chances of the event {*Y*=0} to happen decrease if it is known that event {*X*=1} occurred. And the equivalent to the Bayes posterior probability rule now is valid

$$P(Y=0\,|\,X=1) = 1 - p_2 p_3 - p_2 q_3 = 1 - p_2.$$

Similarly, we determine $R_{X=1}(Y=1)$, and respective posterior probabilities:

$$Rr_{X=1}(Y=1) = \frac{p_2 q_3}{1-p_1 p_3};\ \ P(Y=1\,|\,X=1) = p_2 p_3 + p_2 q_3 = p_2.$$

Compare all four results, we observe complete symmetry in regards of the local dependence strengths: Positive regression coefficients for same results in *Y* as in *X*, and negative (same magnitudes) for opposite result. So to say, *the two risks support each other in the sense that they act in same direction.*

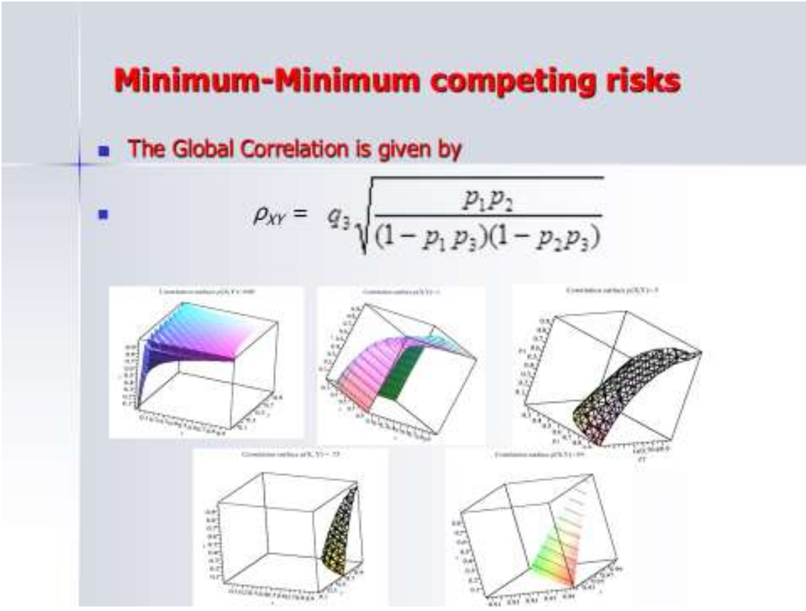
Since symmetric constructions, the regression coefficients $R_{Y=j}(X=i)$, for i , j = 0, 1 in relationships above can be found by same expressions, when keeping p_3 and q_3 as is, and changing indices p_1 and q_1 to p_2 and q_2, and vice versa. We skip details, but give a numeric example for (A): p_1=.3, p_2=.6 and p_3=.9 with calculated correlation coefficient (as measures of global dependence), the regression coefficients (as measures of local dependence) and the posterior probabilities for each variable. For comparison, we also give the same characteristics calculated for the combination of numeric parameters (B): p_1=.3, p_2=.6 and p_3=.1. We have

$$\rho_{XY}(p_1=.3,\ p_2=.6,\ p_3=.9) = 0.0732143;$$

Table 2. Joint distribution (X,Y)

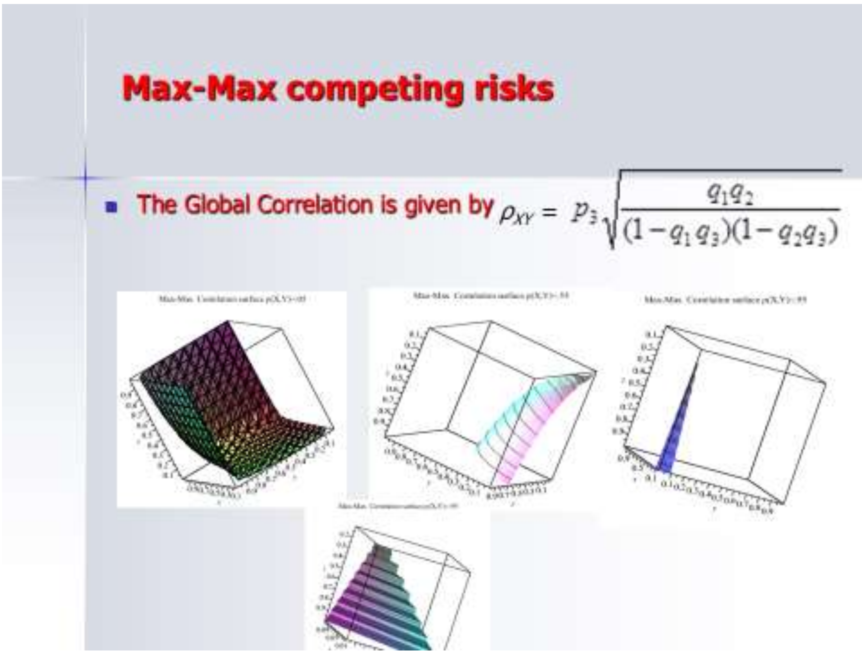
X \ Y	0	1	$f_X(.)$
0	.352	.378	.73
1	.108	.162	.27
$f_Y(.)$	.46	.54	1

The next graphs show surfaces within the cube {0,1]x[0,1]x[0,1], where combination of values p_1, p_2, and p_3 produce correlation coefficient of equal values.

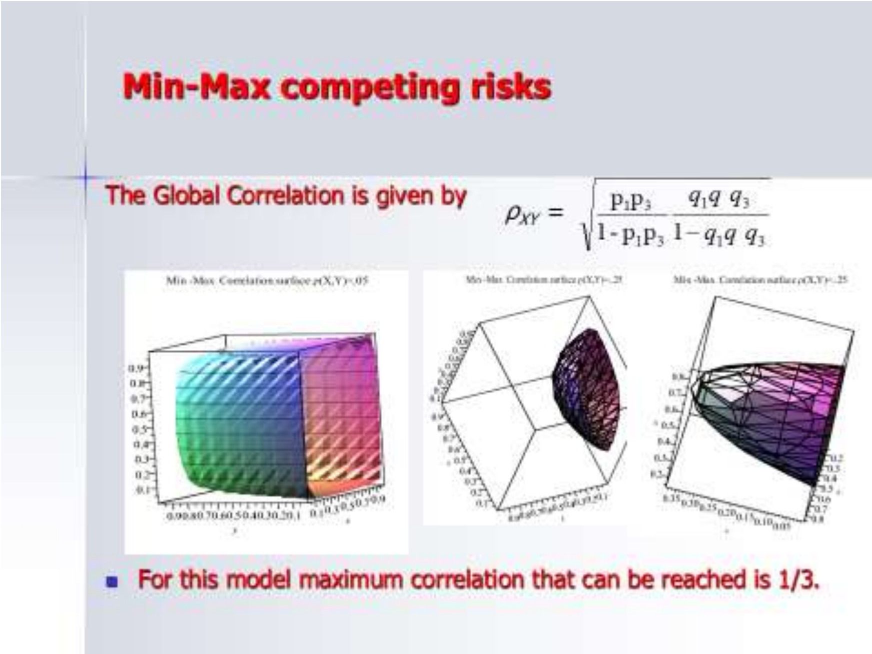


In the next illustrations we will not give detailed numerical analysis, and will show jist the summary graphs similar to this one.

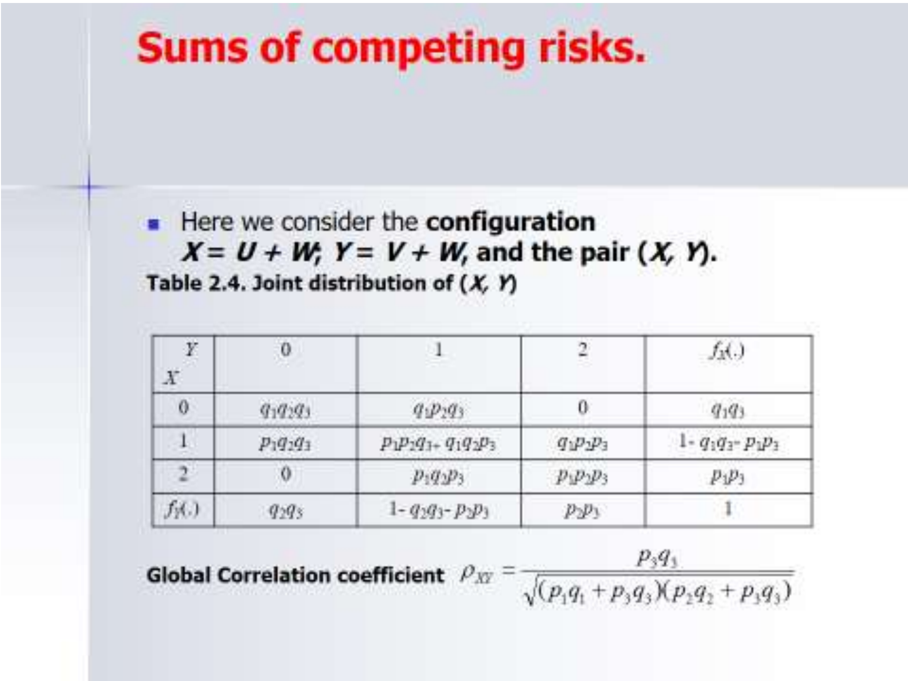
3.4.2 Maximin-Maximum competing risks



3.4.3 Minimum-Maximum competing risks



3.4.4 Sums of competing risks



Sums of competing risks.

- Here we consider the **configuration** $X = U + W$; $Y = V + W$, and the pair (X, Y).

Table 2.4. Joint distribution of (X, Y)

Y / X	0	1	2	$f_X(.)$
0	$q_1q_2q_3$	$q_1p_2q_3$	0	q_1q_3
1	$p_1q_2q_3$	$p_1p_2q_3 + q_1q_2p_3$	$q_1p_2p_3$	$1 - q_1q_3 - p_1p_3$
2	0	$p_1q_2p_3$	$p_1p_2p_3$	p_1p_3
$f_Y(.)$	q_2q_3	$1 - q_2q_3 - p_2p_3$	p_2p_3	1

Global Correlation coefficient $\rho_{XY} = \frac{p_3q_3}{\sqrt{(p_1q_1 + p_3q_3)(p_2q_2 + p_3q_3)}}$

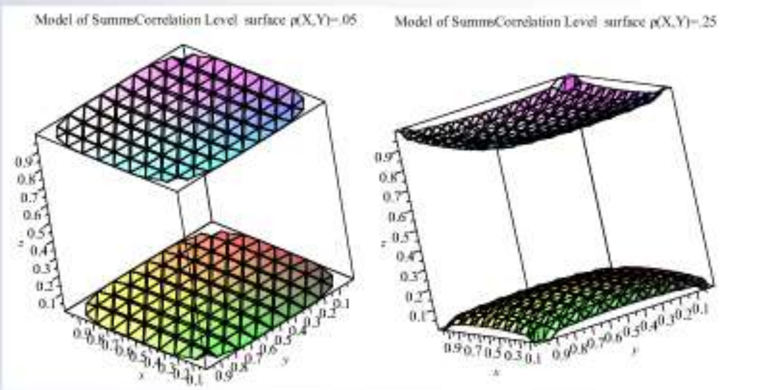
Sums of competing risks.

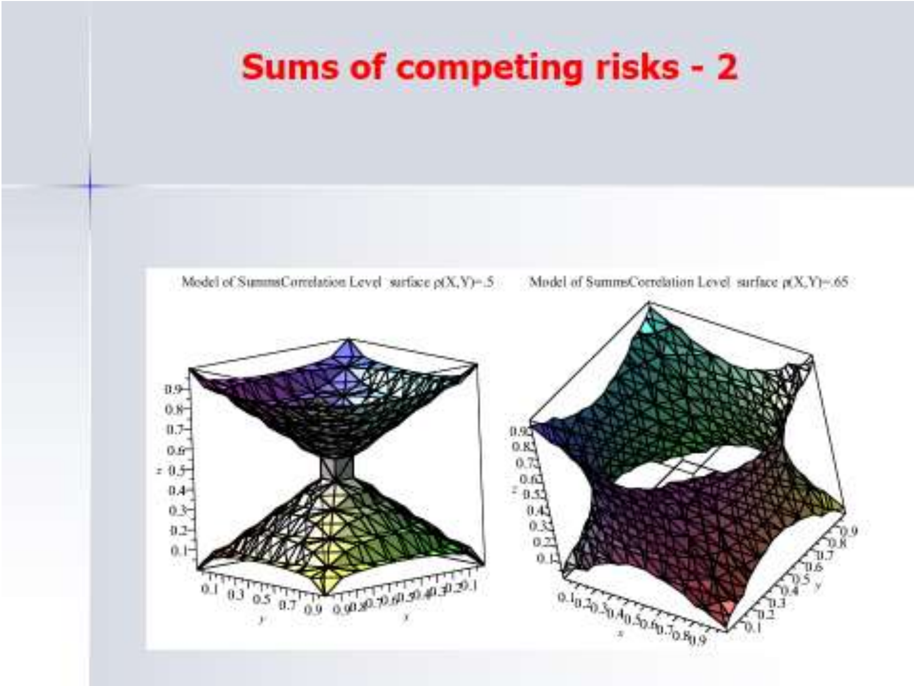
- **Local measures of dependence.**
- These are calculated by use of the rules
- $R_{Y=j}(X=i) = \dfrac{P\{X=i, Y=j) - P\{X=i\}P\{Y=j\}}{P\{X=i\}[1-P\{X=i\}]}$
- In this way we find:

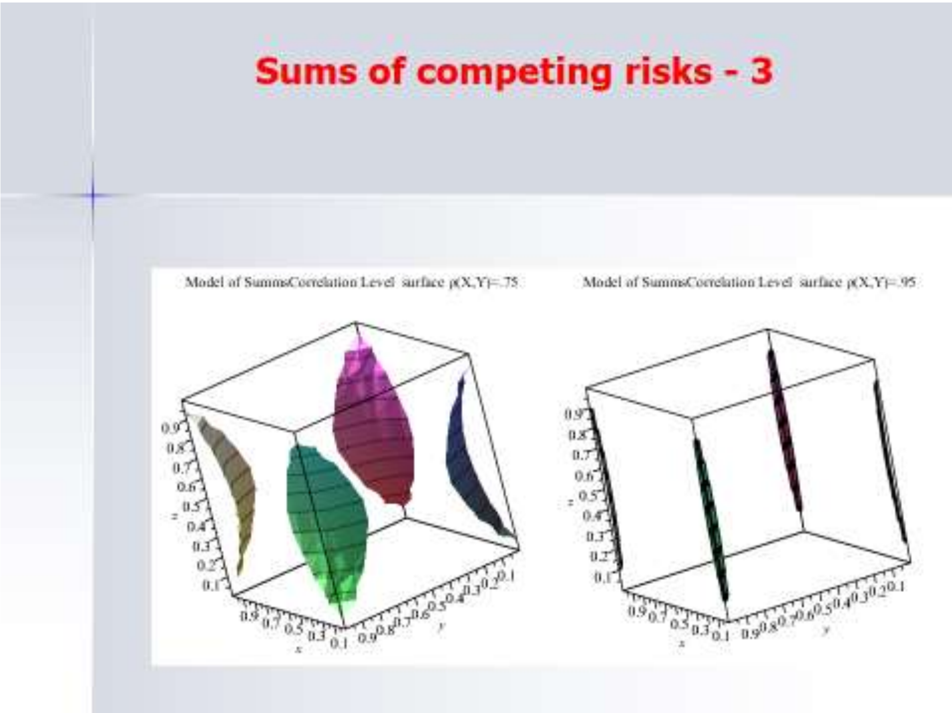
$$R_{Y=0}(X=0) = \frac{q_2 p_3}{1-q_1 q_3} \quad R_{Y=0}(X=1) = \frac{q_2 q_3 (p_1 - q_1)}{(1-q_1 q_3 - p_1 p_3)(q_1 q_3 + p_1 p_3)}$$

$$R_{Y=0}(X=2) = -\frac{q_2 q_3}{1-p_1 p_3}$$

Sums of competing risks -1




Sums of competing risks - 2
Model of SummsCorrelation Level surface ρ(X,Y)=.5
Model of SummsCorrelation Level surface ρ(X,Y)=.65


Sums of competing risks - 3
Model of SummsCorrelation Level surface ρ(X,Y)=.75
Model of SummsCorrelation Level surface ρ(X,Y)=.95

Conclusions

We extend our previous study of local dependence between random events to measures of local dependences between random variables. This turns into a study of the local dependence at a rectangle where interval values of the random variables meet. These local dependences are universally valid and can be continued for higher dimensions. As illustrations, we consider local dependences in reliability systems. The numerical illustrations can be graphically visualized, and show that local dependence is essentially different on different areas in the field. Graphics offer much more comments and further thoughts. Our expectations are that the analysis of Big Data sets will be enriched with the inclusion of our approach into its system tools. An excellent example of this approach can be seen in Dimitrov and Esa (2018).

We also discussed four models for constructing of dependence between two random variables (X,Y) build on 3 independent Bernoulli distributed r.v.'s Um V and W with different parameters.

These models are producing Correlation coefficients in different ranges. These ranges are shown on Correlation Level surfaces in the space of probabilities for success in the used Bernoulli variables in the models. Local dependences between values of X and Y are studied via the correlation coefficient' magnitudes.

Their numerical values serve are presented for particular combinations of parameters, and graphs of some level surfaces are shown.

We are sure that using other particular distributions of the components, different from the Bernoulli ones, may lead to more inetersting and useful results.

References

[1] A. Agresti (2006). *Categorical Data Analysis*, New York, John Wiley & Sons.

[2] I. Bairamov, S., Kotz (2002), Dependence structure and symmetry of Huang-Kotz FGM distributions and their extensions, Metrika 56 , pp 55-72.

[3] N. Blomqvist (1950), On a measure of dependence between two random variables, Annals of Mathematical Statistics 21, pp. 593–600.

[4] J. F. Caari`ere (2004), Copulas, Encyclopedia of Actuarial Science, Vol., 1-3, Wiley, New York, NY.\

[5] B. Dimitrov, (2014) Dependence between Random Events and Its Use, Proceedings of the FISC Conference booklet (2014), pp. 119– 132.

[6] B. Dimitrov (2013). Measures of dependence in r4liability, Proceedings of the 8-th MMR'2013, Stellenbosh, South Africa, July 1-4, pp 65-69.

[7] B. Dimitrov (2010). Some Obreshkov Measures of Dependence and Their Use Compte Rendus de l'Academie Bulgare des Sciences, V. 63, No.1, pp. 15-18.

[8] B. Dimitrov and. S. Esa (2918) Interval dependence structures of two bivariate distributions in riskand reliability, RTA v. 13, No 1, pp. 28 - 38.

[9] D. Drouet-Mari, S., Kotz (2001), *Correlation and dependence*, Imperial College. Press, London,

[10] F. Durante, A new class of symmetric bivariate copulas, Nonparametric Stat. 8(2006), 499–510.

[11] F. Durante, P. Jaworski, A new characterization of bivariate copulas, AMS Subject Classification: 60E05, 62H20, 26A24, 2009.

[12] Esa, S. and Dimitrov, B. (2016). Dependence Structures in Politics and in Reliability. In Proceedings of the Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO'16), I. Frenkel and A. Lisnianski (eds.), Beer Sheva, Israel, February 15-18, pp. 318 - 322, IEEE CPS, 978-1-4673-9941-8/16.

[13] S. Esa and B. Dimitrov. Dependencies in the world of politics, Proceedings of the 8-th MMR’2013, Stellenbosh, South Africa, July 1-4 , pp 70 - 73.

[14] S. Esa and B. Dimitrov. (2013) "Survival Models in Some Political Processes", Risk Analysis and Applications (RT&A), v. 8, # 3, (30) pp 97 - 102.

[15] D. G. J. Farlie, The performance of some correlation coefficients for a general Bivariate distribution, Biometrika 47(1960), 307–323.

[16] H. Joe, Multivariate models and dependence concepts, Chapman & Hall, London, 1997.

[17] J. S. Huang, S., Kotz, Modifcations of the Farlie-Gumbel-Morgenstern distributions, A tough hill to climb. Metrika 49(1999), 135-145.

[18] J. M. Kim and E. A. Sungur, New class of bivariate copulas, In Proceedings for the Spring Conference of Korean Statistical Society, pages 207–212, 2004.

[19] D. Morgenstern, Einfache beispiele zweidimensionaler verteilungen. Mitteilungsblatt für Mathematische Statistik 8(1956), 234–235.

[20] R. B. Nelsen, An introduction to copulas, Springer Series in Statistics. Springer, New York, second edition, 2006.

[21] R. B. Nelsen, Characterization of the Farlie-Gumbel-Morgenstern distribution by the property of the correlation coefficient, SankhyaÂ A, 56(1994), 476-479.

Defuzzification of Healthcare Related Problem by Using (λ, ρ) Intervalued Trapezoidal Fuzzy Numbers and Their Functional

Dr. Rajesh Dangwal

•

Professor, B.G.R. Campus H.N.B. Garhwal
University, Srinagar, Uttarakhand,
E-Mail: dangwalrajeshhnbgu@gmail.com

Kapil Naithani

•

Research Scholar, B.G.R. Campus
H.N.B. Garhwal University, Srinagar, Uttarakhand,
E-Mail: kapilpintu88@gmail.com

Abstract

In healthcare related problems there exist some medical error due to involvement of human errors and technologies error. In general the available data is not sufficient to assess the clinical process. This study uses level (λ, ρ) intervalued trapezoidal fuzzy numbers and their functional to evaluate the estimate reliability of system in the fuzzy sense, In this paper we uses fault tree diagram of the mixed system (series and parallel system) and we change their crisp probability to trapezoidal fuzzy number and a new approach of functional of fuzzy number has developed and Computed results have been compared with results obtained from other existing techniques.

Key Words: Healthcare, Fuzzy sets, inter-valued trapezoidal fuzzy numbers, Functional of fuzzy numbers, Fault tree analysis, Defuzzification.

I. Introduction

In the traditional set theory, an element either belong to the set or not, that is the answer become yes or no rather than more or less. Fuzzy set theory provide means of uncertainty. Probability theory also a primary tool for representing uncertainty but they are random. And all the uncertainty are not random. Fuzzy set theory is marvellous tool for modelling such kind of uncertainty associated with vagueness which are not random.

Healthcare is a series of process for a patient to receive medication..A Joint Health Commission report indicates that medical errors result in the death of between 44,000 and 98,000 patients every year and concludes that healthcare is a high risk [5] There are several examples where reliability analysis methods such as root cause analysis (RCA), failure mode and effect analysis (FMEA), fault tree analysis (FTA) and event tree analysis (ETA) have been applied for patient safety risk modelling in healthcare [6,7–10]. Fault tree analysis has been extensively used as a powerful technique in health related risk analysis from both qualitative and quantitative perspectives [8–10]. . Some of the suggested healthcare areas where FTA can be used are equipment failures and malfunctions, material faults, human errors, environment-related risks, management deficiencies, communication and measurement errors, etc

II. Fuzzy Sets

In real world there exist fuzzyness i.e vagueness , uncertainty. Fuzzy sets were introduced independently by Lotfi A. Jadeh and Dieter Klaua in 1965 as an extension of classical notion of a set .In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent conditions that is an element either belong to set or not, fuzzy set theory permits the gradual assessment of the membership of elements in a set this is described with the aid of a membership function valued belong to the unit interval [0,1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special case of membership functions of fuzzy sets, if the latter only take values 0 or 1. Classical bivalent sets are usually called crisp sets it is used in a wide range of domains in which information is incomplete or imprecise

A fuzzy set is defined by a membership function from the universal set to the interval [0,1], as given below;

$$\mu_A(x) : X \rightarrow [0,1] \tag{1}$$

where $\mu_A(x)$ gives the degree of belongingness of x in the set A. A fuzzy set A can be expressed as follows:

$$\tilde{A} = \{(x, \mu_A(x)) : x \in X\} \tag{2}$$

Fuzziness can be found in many areas of daily life such as in engineering, medicine, manufacturing and others. In all areas in which human judgements, evaluation and decision are important. These are the areas of decision making reasoning, learning and so on.

III. Fuzzification and Inter-Valued Trapezoidal Fuzzy Number

In various situations, the exact values of any parameters of a system are not known due to unavailability of data and complete knowledge about the system, and thus the uncertainty arise. In order to quantify uncertainty, Fuzzification of parameter's value or collected data are done by system experts. In the process of fuzzification, crisp value is transformed into fuzzy value with the help of fuzzy membership functions. For analysing safety and healthcare related problems, inter-valued trapezoidal fuzzy membership functions or more simply inter-valued trapezoidal fuzzy number (IVTPFNs) are often utilized to provide more precise descriptions and to obtain more accurate solutions [27]. In this paper, IVTPFNs are used for quantifying data uncertainty associated with basic events. Mathematically, An interval-valued fuzzy set $\tilde{A}(i-v\ fuzzzyset)$ on R is derived by $\tilde{A} \equiv \{(x, [\mu_{\tilde{A}^L}(x), \mu_{\tilde{A}^u}(x)]) / x \in R\}$ $0 \leq \mu_{\tilde{A}^L}(x) \leq \mu_{\tilde{A}^u}(x) \leq 1 \forall x \in R$, It is denoted by $\mu_{\tilde{A}}(x) = \lfloor \mu_{\tilde{A}^L}(x)\, \mu_{\tilde{A}^U}(x), \rfloor$ $x \in R$ or $\tilde{A} = [\tilde{A}^L, \tilde{A}^U]$

The i-v fuzzy set $\tilde{A}$ indicates that, when the membership grade of x belongs to the interval $\lfloor \mu_{\tilde{A}^L}(x)\, \mu_{\tilde{A}^U}(x), \rfloor$ the largest grade is $\mu_{\tilde{A}^U}(x)$ and the smallest grade is $\mu_{\tilde{A}^L}(x)$

$$\mu_{\tilde{A}^L}(x) = \begin{cases} \dfrac{\lambda(x-a)}{b-a} & a \leq x \leq b \\ \lambda & b \leq x \leq c \\ \dfrac{\lambda(d-x)}{d-c} & c \leq x \leq d \\ 0 & otherwise \end{cases} \tag{3}$$

Therefore, $\tilde{A}^L = (a, b, c, d, \lambda)\ a < b < c < d$
Let

$$\mu_{\tilde{A}^U}(x) = \begin{cases} \dfrac{\rho(x-e)}{b-e} & e \le x \le b \\ \rho & b \le x \le c \\ \dfrac{\rho(f-x)}{(f-c)} & c \le x \le f \\ 0 & otherwise \end{cases} \quad (4)$$

Therefore $\tilde{A}^U = (e, b, c, f, \rho), e < b < c < f,$ Consider the case in which $0 < \lambda \le \rho \le 1$ and $e < a < b < c < d < f$, from (3) and (4) we obtain $\tilde{A} = [\tilde{A}^L, \tilde{A}^U]$ $[(a, b, c, d; \lambda), (e, b, c, f; \rho)]$, Which is called the level (λ, ρ) $i - \text{v}$ fuzzy number.

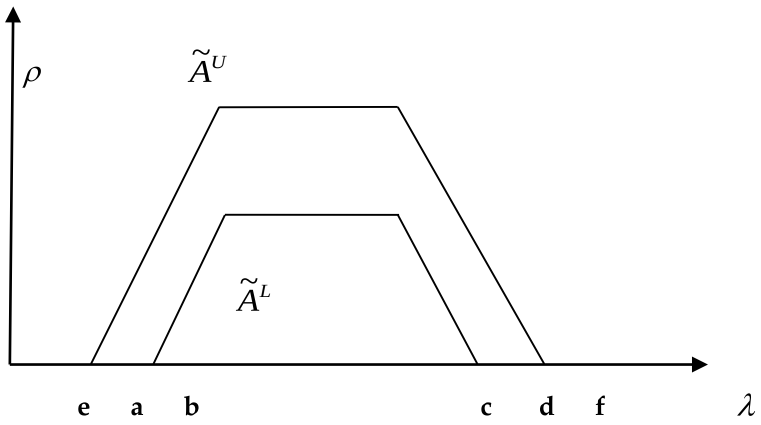


Fig1: i-v trapezoidal fuzzy number $\tilde{A}$

IV. Fault Tree Analysis and Fuzzy probability

Fault-tree analysis (FTA) is a top-down. Deductive failure analysis in which an undesired state of a system is analyzed using Boolean logic to combine a series of lower level basic events it is to evaluate probability of an accident resulting from sequences and combinations of faults and failure events. A fault tree describes an accident-model and interprets the relations between components.

Thus, the fault tree is useful for understanding logically the mode of occurrence of an accident. Furthermore, given the failure probabilities of system components, the probability of the top event can be calculated. In fault tree diagram there are two gates are used one is "AND" and another is "OR" AND (conjunction) means the failure probability is depend on all those event which are associated with AND gate and in OR gate the event associated with OR gate they are work independently to failure next event.

The fuzzy failure probability can be calculated by following arithmetic operation on fuzzy numbers, here when we take probability in fuzzy sense the FTA is called FFTA(fuzzy fault tree analysis)

Table1:

Row. No.	Approach	Gate	Operation	Equation
(1)	**Traditional FTA**	**OR**	**Conjunction**	**P**OR $=1-[(1-q_1)\times(1-q_2)\times\ldots\times(1-q_n)]$
		AND	**Intersection**	**P**AND $= q_1 \times q_2 \times \ldots. \times q_n$
(2)	**Traditional FFTA**	**OR**	**Conjunction**	**P**OR $= \tilde{1}\Theta[(\tilde{1}\Theta\tilde{q}_1)\otimes(\tilde{1}\Theta\tilde{q}_2)\otimes\ldots.\otimes(\tilde{1}\Theta\tilde{q}_n)]$
		AND	**Intersection**	**P**AND $= \tilde{q}_1 \otimes \tilde{q}_2 \otimes \ldots\ldots \otimes \tilde{q}_n$

V. Defuzzification

Defuzzification is the process of producing a quantifiable result in crisp logic. Given fuzzy sets and corresponding memberships degree it is the process to transformed fuzzy numbers to crisp value there are many rule to transform a number of variable into a fuzzy set and then defuzzified. Defuzzification is interpreting the membership degrees of the fuzzy sets into a specific decision or real value the simplest method to defuzzification is choose to set of highest membership function another a common and useful defuzzification technique is centre of gravity. For the interval valued trapezoidal fuzzy number we will take the mean of the COG of upper and lower trapezoidal fuzzy numbers. Also we use the signed distance method and bisector of area method to evaluate the defuzzified the fuzzy numbers.

VI. Steps of the methodology

Step1. Construction of fault-tree

Construct fault-tree for some healthcare related problems (e.g. patient transfer without infection control, etc.) by using fault-tree logical symbols.

Step2. Obtain fundamental events failure probabilities in the form of level (λ, ρ) interval-valued trapezoidal fuzzy numbers

Possible failure probability of each fundamental event is obtained by aggregating experts knowledge and experience, and represented in terms of level (λ, ρ) interval-valued trapezoidal fuzzy numbers.

Step3. Computation of system top event fuzzy failure probability ($\tilde{q}_T$)

Using fault-tree diagram and possible failure of fundamental events represented in terms (λ, ρ) interval-valued trapezoidal fuzzy numbers, the system top event fuzzy failure probability $(\tilde{q}_T)$ can be computed by using operations given in table 4. Also, the defuzzified value of system top event can be easily computed using various defuzzification methods as COG, bisector of area, middle of maxima.

Step4.Compute system top event fuzzy reliability

Compute system top event fuzzy reliability which is equal to one minus the fuzzy failure probability of the top event.

Step5. Find the most and least influential fundamental events of the problem.

Tanaka et. al. V −index will be extended for level (λ, ρ) interval-valued trapezoidal fuzzy numbers, the most and least influential fundamental events of the considered problems will be evaluated by finding $\max\{V(\tilde{q}_T, \tilde{q}_{T_i})\forall\, i\}$ and $\min\{V(\tilde{q}_T, \tilde{q}_{T_i})\forall\, i\}$ values respectively for the system, where $\tilde{q}_{T_i}$ is the system top event fuzzy failure probability after eliminated i^{th} basic event.

Step6. Analyze the results and give suggestions based on it for improving the efficiency of considered healthcare related problems.

Index V, measure the difference between $\tilde{q}_T$ and $\tilde{q}_{T_i}$, and defined as

$$V(\tilde{q}_T, \tilde{q}_{T_i}) = \begin{matrix} (\tilde{q}^L{}_{T_1} - \tilde{q}^L{}_{Ti_1}) + (\tilde{q}^L{}_{T_2} - \tilde{q}^L{}_{Ti_2}) + (\tilde{q}^L{}_{T_3} - \tilde{q}^L{}_{Ti_3}) + (\tilde{q}^L{}_{T_4} - \tilde{q}^L{}_{Ti_4}) + \\ (\tilde{q}^U{}_{T_5} - \tilde{q}^U{}_{Ti_5}) + (\tilde{q}^U{}_{T_2} - \tilde{q}^U{}_{Ti_2}) + (\tilde{q}^U{}_{T_3} - \tilde{q}^U{}_{Ti_3}) + (\tilde{q}^U{}_{T_6} - \tilde{q}^U{}_{Ti_6}) \end{matrix} > 0 \qquad (5)$$

$V(\tilde{q}_T, \tilde{q}_{T_i})$ indicates the extent of improvement in eliminating the failure of the ith component.

If $V(\tilde{q}_T, \tilde{q}_{T_i}) \geq V(\tilde{q}_T, \tilde{q}_{T_j})$ then preventing failure of i-th component is more effective than j-th component

VII. Example

Fault tree analysis of patient transfer without infection control precaution here is an example of study extent and execution of redundant process during in patient transfer to radiology, and their impact on errors during the transfer process; in which there are given some basic events which are fundamental event to reliable whole system. There were four ways to communicate the required infection control precautions to a porter: (1) verbal handover at Radiology; (2) written handover at Radiology using a transfer form; (3) verbal handover at the ward during patient collection; and (4) verification of the transfer form by the ward nurse fig 2 depicts a fault tree for the events leading to inadequate infections control precautions during transfers. The basic events are given as follow.

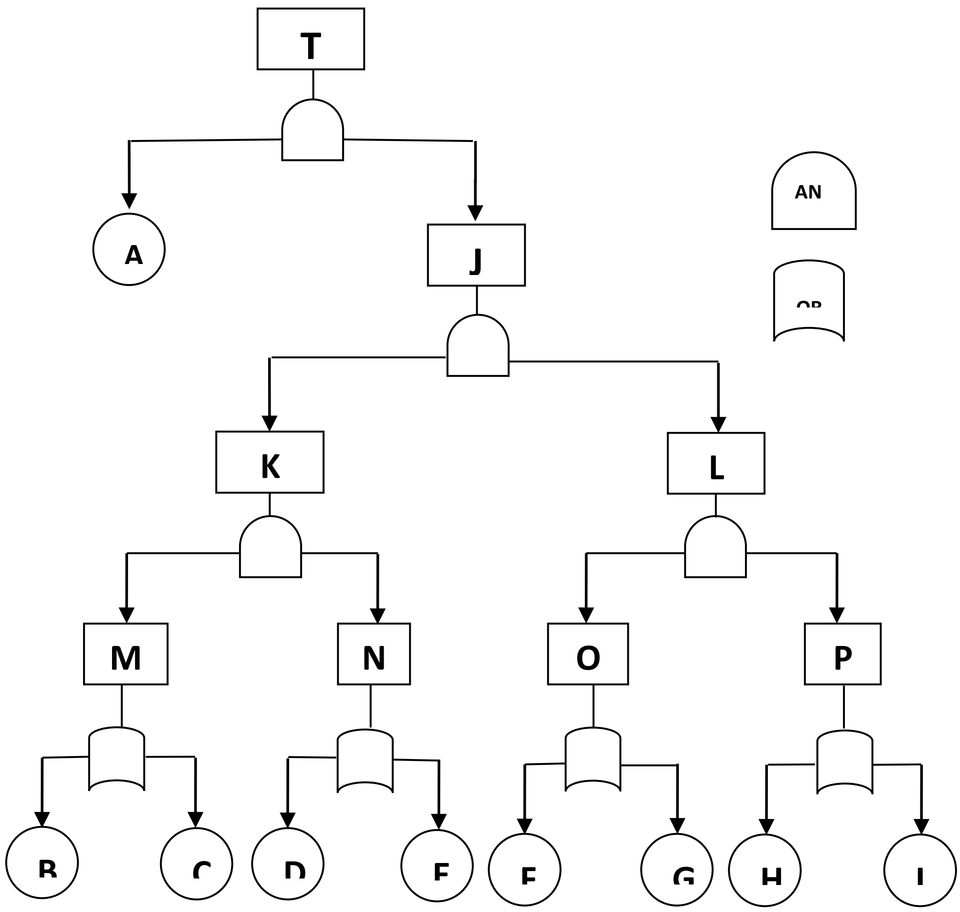


Fig 2: Fault tree of patient transfer without infection control precaution

T=Inadequate infections control precautions

A= Patient was infectious

B= patient infectious status was not communicated through verbal handover at radiology

C= Patient infectious status was communicated through verbal handover at radiology, but error still occurred

D= Patient infectious status was not communicated through written handover at radiology

F= Patient infectious status was not communicated through verbal handover at ward

G= Patient infectious status was not communicated through verbal handover at ward, but error still occurred

H= Patient infectious status was not communicated through written handover at ward.

I= Patient infectious status was not communicated through written handover at ward, but error still occurred

The fault tree diagram of "**Patient transfer without infection control precaution**" is follow, here the Shape show "AND" and for "OR" gate

We express and gate by Intersection $\cap$ and OR gate by conjunction (union) $\cup$. The fault tree can be expressed as

$$\begin{aligned} T &= A\cap J \\ &= A\cap(k\cap L) \\ &= A\cap[(M\cap N)\cap(O\cap P)] \\ &= A\cap[(B\cup C)\cap(D\cup E)\cap(F\cup G)\cap(H\cup I)] \end{aligned} \quad (6)$$

The mathematical formula of failure probability o top event T is given by

$$\begin{aligned} q_{\tilde{T}} &= q_{\tilde{A}}\times q_{\tilde{J}} \\ &= q_{\tilde{A}}\times q_{\tilde{K}}\times q_{\tilde{L}} \\ &= q_{\tilde{A}}\times[(q_{\tilde{M}}\times q_{\tilde{N}})\times(q_{\tilde{O}}\times q_{\tilde{P}})] \\ &= q_{\tilde{A}}\times[\{1-(1-q_{\tilde{B}})(1-q_{\tilde{C}})\}\times\{1-(1-q_{\tilde{D}})(1-q_{\tilde{E}})\}\times\{1-(1-q_{\tilde{F}})(1-q_{\tilde{G}})\}\times\{1-(1-q_{\tilde{H}})(1-q_{\tilde{I}})\}] \end{aligned} \quad (7)$$

Using this formula the failure probability $q_{\tilde{T}}$ by various method can be calculated and the reliability of system calculated by $1-q_{\tilde{T}}$.

TABLE 2: Crisp values and their corresponding trapezoidal fuzzy numbers

S.No.	Failure Probability	Crisp value	IVTPFN
1	$q_{\tilde{A}}$	0.27	$\begin{pmatrix}0.23,0.25,0.29,0.31:0.8\\0.21,0.25,0.29,0.33:1.0\end{pmatrix}$
2	$q_{\tilde{B}}$	0.93	$\begin{pmatrix}0.89,0.91,0.95,0.97:0.8\\0.87,0.91,0.95.0.99:1.0\end{pmatrix}$
3	$q_{\tilde{C}}$	0	$\begin{pmatrix}0.00,0.00,0.00,0.00:0.8\\0.00,0.00,0.00,0.00:1.0\end{pmatrix}$
4	$q_{\tilde{D}}$	0.52	$\begin{pmatrix}0.48,0.50,0.54,0.56:0.8\\0.46,0.50,0.54,0.58:1.0\end{pmatrix}$
5	$q_{\tilde{E}}$	0.22	$\begin{pmatrix}0.18,0.20,0.24.0.26:0.8\\0.16,0.20,0.24,0.28:1.0\end{pmatrix}$
6	$q_{\tilde{F}}$	0.74	$\begin{pmatrix}0.70,0.72,0.76,0.78:0.8\\0.68,0.72,0.76,0.80:1.0\end{pmatrix}$

S.No.	Failure Probability	Crisp value	IVTPFN
7	$q_{\tilde{G}}$	0	$\begin{pmatrix} 0.00,0.00,0.00,0.00 : 0.8 \\ 0.00,0.00,0.00,0.00 : 1.0 \end{pmatrix}$
8	$q_{\tilde{H}}$	0.56	$\begin{pmatrix} 0.52,0.54,0.58,0.60 : 0.8 \\ 0.50,0.54,0.58,0.62 : 1.0 \end{pmatrix}$
9	$q_{\tilde{I}}$	0.30	$\begin{pmatrix} 0.26,0.28,0.32,0.34 : 0.8 \\ 0.24,0.28,0.32,0.36 : 1.0 \end{pmatrix}$

TABLE 3: Fuzzy operation of two intwervalued fuzzy numbers

OPERATION	TRAPEZOIDAL FUZZY INTERVALUED NUMBERS
MULTIPPLICATION	$\begin{pmatrix} a_1, b_1, c_1, d_1 : \lambda \\ e_1, b, c, f_1 : \rho \end{pmatrix} \times \begin{pmatrix} a_2, b_2, c_2, d_2 : \lambda \\ e_2, b_2, c_2, f_2 : \rho \end{pmatrix} = \begin{pmatrix} a_1 . a_2, b_1 b_2, c_1 c_2, d_1 d_2 : \lambda \\ e_1 e_2, b_1 b_2, c_1 c_2, f_1 f_2 : \rho \end{pmatrix}$
COMPLEMENT	$1 - \begin{pmatrix} a, b, c, d : \lambda \\ e, b, c, f : \rho \end{pmatrix} = \begin{pmatrix} 1-d, 1-c, 1-b, 1-a : \lambda \\ 1-f, 1-c, 1-b, 1-a : \rho \end{pmatrix}$

VIII. Failure probability by various method

VIII(I). Traditional Method: The fuzzy failure probability of top event by traditional method taking crisp value of failure probability is given by **0.080441705** and reliability is **0.919558295**

VIII(II). Max Min Method:-It is for crisp value of failure probability, it is introduced by huang.et.al.,in this method we take maximum for union and minimum for intersection. In given example the fuzzy failure probability can be calculated as follow from equation (5)

$$q_M = \max(q_B, q_C) = \max(0.93, 0.00) = 0.93$$
$$q_N = \max(q_D, q_E) = \max(0.52, 0.22) = 0.52$$
$$q_O = \max(q_F, q_G) = \max(0.74, 0.00) = 0.74$$
$$q_P = \max(q_H, q_I) = \max(0.56, 0.30) = 0.56$$
$$q_K = \min(q_M, q_N) = \min(0.93, 0.52) = 0.52$$
$$q_L = \min(q_O, q_P) = \min(0.74, 0.56) = 0.56$$
$$q_J = \min(q_K, q_L) = \min(0.52, 0.56) = 0.52$$
$$q_T = \min(q_A, q_J) = \min(0.27, 0.52) = 0.27$$

That is fuzzy failure probability of top event is 0.27 and fuzzy reliability of top event is 0.73

VIII(III). TANAKA ET AL METHOD: The fuzzy failure probability of top event of fault tree by tanaka. et. al method from table 2 and is given by $\begin{pmatrix} 0.0527192568, 0.0668251584, 0.0986636751, 0.1179379848 : 0.8 \\ 0.0418668562, 0.0668251584, 0.0986636751, 0.1396448571 : 1.0 \end{pmatrix}$ and the fuzzy reliability of top event is $\begin{pmatrix} 0.8820620152, 0.9013363249, 0.9331748416, 0.9472807432 : 0.8 \\ 0.8603551429, 0.9013363249, 0.9331748416, 0.9581331438 : 1.0 \end{pmatrix}$

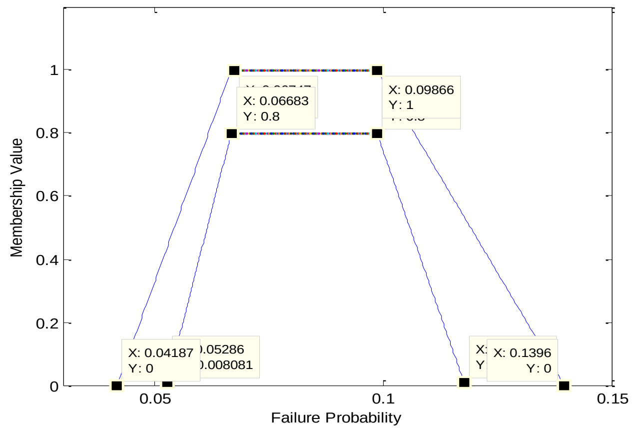


Fig3: Failure Probability Of Top Event In Trapezoidal Fuzzy Number

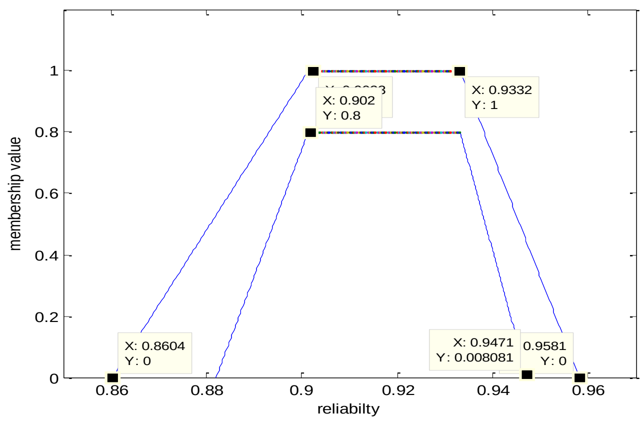


Fig4: Fuzzy Reliability Of Top Event In Trapezoidal Fuzzy Numbers

VIV. DEFUZZIFICATION METHODS

Defuzzificztion is a process when we transformed fuzzy numbers to discrete value which is most probable value among all the probabilities there are some important defuzzification methods, which are following.

VIV(I) . MEAN OF COG METHOD

COG method is one of the most defuzzification method. In this method we take mean value of lower and upper trapezoidal fuzzy numbers, the upper left and right and similarly lower left and right membership function are

$$\mu_{A_L^U}(x) = \rho\left(\frac{x-e}{b-e}\right) \qquad \mu_{A_R^U}(x) = \rho\left(\frac{x-f}{c-f}\right)$$

$$\mu_{A_L^L}(x) = \lambda\left(\frac{x-a}{b-a}\right) \qquad \mu_{A_R^L}(x) = \lambda\left(\frac{x-d}{c-d}\right)$$

Now centre of gravity is given by $\bar{x} = \dfrac{\int x.\mu_A(x)dx}{\int \mu_A(x)dx}$

$$\bar{x}^U = \frac{\int x^U.\mu_{A^U}(x)dx}{\int \mu_{A^U}(x)dx} \qquad \bar{x}^L = \frac{\int x^L.\mu_{A^L}(x)dx}{\int \mu_{A^L}(x)dx}$$

MEAN OF COG = $\dfrac{1}{2}\left(\bar{x}^U + \bar{x}^L\right)$

$$x = \frac{1}{2}\left[\left\{\frac{\int_e^b x.\rho\left(\frac{x-e}{b-e}\right)dx + \int_b^c x\rho dx + \int_c^f x.\rho\left(\frac{x-f}{c-f}\right)dx}{\int_e^b \rho\left(\frac{x-e}{b-e}\right)dx + \int_b^c \rho dx + \int_c^f \rho\left(\frac{x-f}{c-f}\right)dx}\right\} + \left\{\frac{\int_a^b x.\rho\left(\frac{x-a}{b-a}\right)dx + \int_b^c x\lambda dx + \int_c^d x.\lambda\left(\frac{x-d}{c-d}\right)dx}{\int_e^b \rho\left(\frac{x-a}{b-a}\right)dx + \int_b^c \lambda dx + \int_c^d \lambda\left(\frac{x-d}{c-d}\right)dx}\right\}\right]$$

(8)

$$x = \frac{1}{6}\left[\left\{\frac{c^2 + f^2 - b^2 - e^2 - fc - eb}{c + f - e - b}\right\} + \left\{\frac{c^2 + d^2 - b^2 - a^2 + cd - ab}{c + d - a - b}\right\}\right] \tag{9}$$

In this method the fuzzy failure probability obtained by tanaka. et. al. of top event can be calculated and is equal to 0.0504603867 and fuzzy reliability of top event is **0.949539613**

VIV(II). BISECTOR OF AREA METHOD:

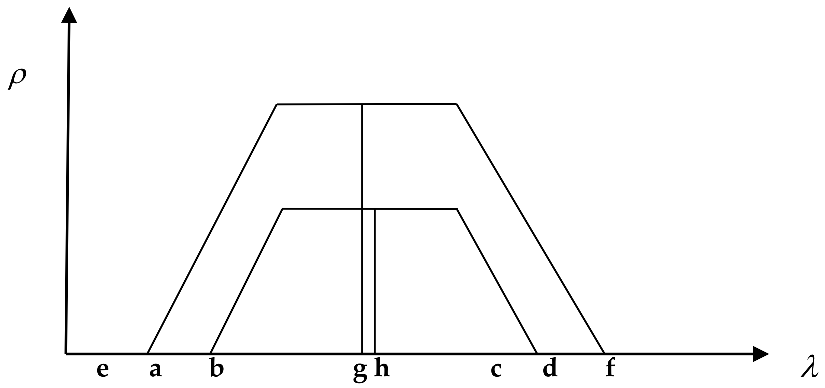


FIG 5: Bisector Of Area

in this method we evaluate the point where the area are bisected let 'g" and "h" be the point where the upper and lower trapezoidal region is bisected for upper trapezoidal area

$$\frac{1}{2}[(g-b)+(g-e)]\times\rho = \frac{1}{2}[(f-g)+(c-g)]\times\rho$$
$$2g-(b+e) = (f+c)-2g$$
$$4g = (f+c)+(b+e)$$
$$g = \frac{1}{4}(f+c+b+e) \tag{10}$$

Similarly for lower trapezoidal area $h = \dfrac{1}{4}(d+c+b+a)$ (11)

Therefore mean of the bisector of upper and lower trapezoidal fuzzy number will be given by

$$\frac{1}{2}(g+h) = \frac{1}{8}(a+e+d+f+2b+2c) \tag{12}$$

In this method the fuzzy failure top event is = 0.0853933277 and the fuzzy reliability of top event is **0.914606672**

VIV(III). MIDDLE OF MAXIMA: –

In this method we get the value of x which is middle of maximum membership function which is given by $x_{MOM} = \frac{1}{2}(b+c)$ (13)

By this method the fuzzy failure probability of top event is =0.0827444168 and the fuzzy reliability is =**0.917255583**

VIV(IV) FUNCTIONAL OF FUZZY NUMBERS:

It is defined as a function of function of x that is if we give the membership value to membership function .it is defined as follow

$\mu_A(x): X \rightarrow [0,1] \quad \mu(\mu_A(x)): \mu_A(x) \rightarrow [0,1]$

$\tilde{A} = \{((x, \mu_A(x)), \mu(\mu_A(x))): x \in X\}$

Let $\mu(\mu_A(x)) = \frac{1}{x+1}$ and $0 \le \frac{1}{x+1} \le 1$ for any non negative value of x then the fuctional form of fuzzy number as following

TABLE 4

S.No.	Failure Probabiliy	Crisp value	Functional value of IVTPFN $((x, \mu_A(x)), \mu(\mu_A(x)))$
1	$q_{\tilde{A}}$	0.27	$\begin{pmatrix} 0.23:0.8:0.813,0.25:0.8:0.8,0.29:0.8:0.775,0.31:0.8:0.763 \\ 0.21:1.0:0.826,0.25:1.0:0.8,0.29:1.0:0.775,0.33:1.0:0.751 \end{pmatrix}$
2	$q_{\tilde{B}}$	0.93	$\begin{pmatrix} 0.89:0.8:0.529,0.91:0.8:0.523,0.95:0.8:0.512,0.97:0.8.0.507 \\ 0.87:1.0:0.534,0.91:1.0:0.523,0.95:1.0:0.512,0.99:1.0:0.502 \end{pmatrix}$
3	$q_{\tilde{C}}$	0	$\begin{pmatrix} 0.00:0.8:1,0.00:0.8:1,0.00:0.8:1,0.00:0.8:1 \\ 0.00:1.0:1,0.00:1.0:1,0.00:1.0:1,0.00:1.0:1 \end{pmatrix}$
4	$q_{\tilde{D}}$	0.52	$\begin{pmatrix} 0.48:0.8:0.675,0.50:0.8:0.666,0.54:0.8:0.649,0.56:0.8:0.641 \\ 0.46:1.0:0.684,0.50:1.0:0.666,0.54:1.0:0.649,0.58:1.0:0.632 \end{pmatrix}$
5	$q_{\tilde{E}}$	0.22	$\begin{pmatrix} 0.18:0.8:0.847,0.20:0.8:0.833,0.24:0.8:0.806,0.26:0.8:0.793 \\ 0.16:1.0:0.862,0.20:1.0:0.833,0.24:1.0:0.806,0.28:1.0:0.781 \end{pmatrix}$
6	$q_{\tilde{F}}$	0.74	$\begin{pmatrix} 0.70:0.8:0.588,0.72:0.8:0.581,0.76:0.8:0.568,0.78:0.8:0.561 \\ 0.68:1.0:0.595,0.72:1.0:0.581,0.76:1.0:0.568,0.80:1.0:0.555 \end{pmatrix}$
7	$q_{\tilde{G}}$	0	$\begin{pmatrix} 0.00:0.8:1,0.00:0.8:1,0.00:0.8:1,0.00:0.8:1 \\ 0.00:1.0:1,0.00:1.0:1,0.00:1.0:1,0.00:1.0:1 \end{pmatrix}$
8	$q_{\tilde{H}}$	0.56	$\begin{pmatrix} 0.52:0.8:0.657,0.54:0.8:0.649,0.58:0.8:0.632,0.60:0.8:0.625 \\ 0.50:1.0:0.666,0.54:1.0:0.649,0.58:1.0:0.632,0.62:1.0:0.612 \end{pmatrix}$
9	$q_{\tilde{I}}$	0.30	$\begin{pmatrix} 0.26:0.8:0.793,0.28:0.8:0.781,0.32:0.8:0.757,0.34:0.8:0.746 \\ 0.24:1.0:0.806,0.28:1.0:0.781,0.32:1.0:0.757,0.36:1.0:0.735 \end{pmatrix}$

The fuzzy functional failure probability of top event is given by from tanaka.et.al. method from table 2 and is given by

$$\begin{pmatrix} 0.0527192568:0.8:0.949,0.0668251584:0.8:0.937,0.0986636751:0.8:0.910,0.1179379848:0.8:0.988 \\ 0.0418668562:1.0:0.959,0.0668251584:1.0:0.937,0.0986636751:1.0:0.910,0.1396448571:1.0:0.877 \end{pmatrix}$$

and reliability of top event given by

$$\begin{pmatrix} 0.8820620152\ 0.8:0.531, 0.9013363249\ 0.8:0.525, 0.9331748416\ 0.8:0.517, 0.9472807432\ 0.8:0.513 \\ 0.8603551429\ 1.0:0.537, 0.9013363249\ 1.0:0.525, 0.9331748416\ 1.0:0.517, 0.9581331438\ 1.0:0.510 \end{pmatrix}$$

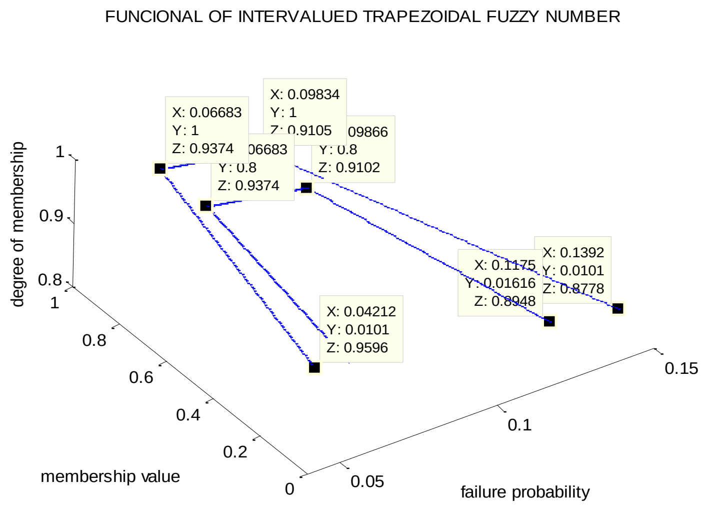


Fig6: Graph Of Functional Value Of Failure Probability F Top Event

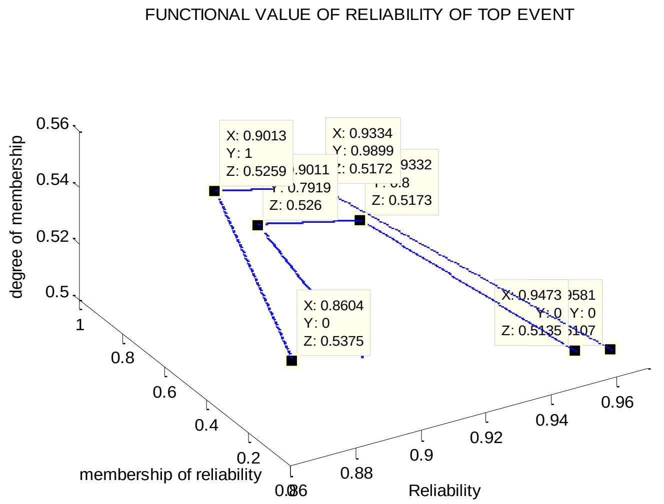


Fig7: Graph Of Functional Value Of Reliabilty Of Top Event

For defuzzification of functional of fuzzy numbers we are using the centre of gravity method(centre of volume)

Now to determine the centre of gravity we can use the formula

$$x^* = \frac{\int x.\mu_A(x).\mu(\mu_A(x))dx}{\int \mu_A(x).\mu(\mu_A(x))dx} \tag{14}$$

We evaluate the defuzzified value of each lower and upper trapezoidal fuzzy numbers and then we take the mean value of both defuzzification value.

$$x^{U^*}=\frac{\int_e^b x.\rho\left(\frac{x-e}{b-e}\right).\frac{1}{x+1}dx+\int_b^c x.\rho.\frac{1}{x+1}dx+\int_c^f x.\rho\left(\frac{f-x}{f-c}\right).\frac{1}{x+1}dx}{\int_e^b \rho\left(\frac{x-e}{b-e}\right).\frac{1}{x+1}dx+\int_b^c \rho.\frac{1}{x+1}dx+\int_c^f \rho\left(\frac{f-x}{f-c}\right).\frac{1}{x+1}dx}$$

$$x^{L^*}=\frac{\int_a^b x.\lambda\left(\frac{x-a}{b-a}\right).\frac{1}{x+1}dx+\int_b^c x.\lambda.\frac{1}{x+1}dx+\int_c^d x.\lambda\left(\frac{d-x}{d-c}\right).\frac{1}{x+1}dx}{\int_a^b \lambda\left(\frac{x-a}{b-a}\right).\frac{1}{x+1}dx+\int_b^c \lambda.\frac{1}{x+1}dx+\int_c^d \lambda\left(\frac{d-x}{d-c}\right).\frac{1}{x+1}dx}$$

$$x^*=\frac{1}{2}\left[x^{U^*}+x^{L^*}\right]$$

$$x^{U^*}=\frac{\left[\frac{f+c-b-e}{2}+\frac{e+1}{b-e}\ln\left(\frac{b+1}{e+1}\right)-\ln\left(\frac{c+1}{b+1}\right)-\frac{f+1}{f-c}\ln\left(\frac{f+1}{c+1}\right)\right]}{\left[\frac{f+1}{f-c}\ln\left(\frac{f+1}{c+1}\right)-\frac{e+1}{b-e}\ln\left(\frac{b+1}{e+1}\right)+\ln\left(\frac{c+1}{b+1}\right)\right]} \quad (15)$$

$$x^{L^*}=\frac{\left[\frac{d+c-b-a}{2}+\frac{a+1}{b-a}\ln\left(\frac{b+1}{a+1}\right)-\ln\left(\frac{c+1}{b+1}\right)-\frac{d+1}{d-c}\ln\left(\frac{d+1}{c+1}\right)\right]}{\left[\frac{d+1}{d-c}\ln\left(\frac{d+1}{c+1}\right)-\frac{a+1}{b-a}\ln\left(\frac{b+1}{a+1}\right)+\ln\left(\frac{c+1}{b+1}\right)\right]} \quad (16)$$

Defuzzified value of failure probability of top event of upper trapezoidal fuzzy number is 0.083528098 and defuzzified value of failure probability of lower trapezoidal fuzzy number is 0.083528098.Similarly defuzzified value of reliabilty of lower trapezoidal fuzzy number is 0.916471902 and defuzzified value of reliability of lower trapezoidal fuzzy number is is 0.912979297 .and the mean value of both values is **0.9147255995**

In traditional method the fuzzy failure probability can be calculated as follow

$$=0.27\times\left[\{1-(1-0.93)(1-0.00)\}\times\{1-(1-0.52)(1-0.22)\}\times\{1-(1-0.74)(1-0.00)\}\times\{1-(1-0.56)(1-0.30)\}\right]$$

=**0.080441705**

And the fuzzy reliability of top event is **0.919558295**

The difference of reliability value from traditional method and the tanaka et al method is 0.004832695 to the left of trapezoidal fuzzy numbers.

Table 5.Ranking of basic event of Example using failure difference

Eliminated event	$\tilde{q}_{Ti_1}$	$V\left(\tilde{q}_{T_1},\tilde{q}_{T_{1i}}\right)$	Rank
$A(i=1)$	$\begin{pmatrix}0.00,0.00,0.00,0.00\\0.00,0.00,0.00,0.00\end{pmatrix}$	**0.6831466219**	**1**
$B(i=2)$	$\begin{pmatrix}0.00,0.00,0.00,0.00\\0.00,0.00,0.00,0.00\end{pmatrix}$	**0.6831466219**	**1**
$C(i=3)$	$\begin{pmatrix}0.0527192568,0.0668251584,0.0986636751,0.1179379848\\0.0418668562,0.0668251584,0.0986636751,0.1396448571\end{pmatrix}$	**0.00**	**6**
$D(i=4)$	$\begin{pmatrix}0.016249086,0.0219098888,0.0358994573,0.0448827226\\0.0120220692,0.0219098888,0.0358994573,0.0553832294\end{pmatrix}$	**0.4389908241**	**2**

Eliminated event	$\tilde{q}_{Ti_1}$	$V\left(\tilde{q}_{T_1}, \tilde{q}_{T_{1i}}\right)$	Rank
$E(i=5)$	$\begin{pmatrix} 0.0433308960, 0.054774720, 0.0807737789, 0.0966704794 \\ 0.0345634491, 0.054774720, 0.0807737789, 0.1147221028 \end{pmatrix}$	**0.1227626968**	**5**
$F(i=6)$	$\begin{pmatrix} 0.00, 0.00, 0.00, 0.00 \\ 0.00, 0.00, 0.00, 0.00 \end{pmatrix}$	**0.6831466219**	**1**
$G(i=7)$	$\begin{pmatrix} 0.0527192568, 0.0668251584, 0.0986636751, 0.1179379848 \\ 0.0418668562, 0.0668251584, 0.0986636751, 0.1396448571 \end{pmatrix}$	**0.00**	**6**
$H(i=8)$	$\begin{pmatrix} 0.0217571536, 0.027977040, 0.0441942554, 0.054482212 \\ 0.0166138318, 0.027977040, 0.0441942554, 0.0664272576 \end{pmatrix}$	**0.3615252583**	**3**
$I(i=9)$	$\begin{pmatrix} 0.0435143026, 0.053955720, 0.0801020878, 0.0961450963 \\ 0.0346121496, 0.053955720, 0.0801020878, 0.1144024992 \end{pmatrix}$	**0.1263569586**	**4**

Conclusion

It is seen that when the membership function of fuzzy failure probability is also in fuzzy sense than reliability of top event can also go to another value**.** Here the error outcome from trapezoidal fuzzy failure probability to left side, and using equation 5 and table 5, the value of index is analyzed that the most critical basic events are A B and F whereas least critical basic events is C and G. The order of all critical basic events are given below in decreasing manner (A, B, F) > D > H > I > E > C > G.

References

[1] Reason J. Achieving a safe culture: theory and practice. Work Stress 1998; 12:293e306.

[2] Hollnagel E. Barrier and accident prevention. Aldershot, UK: Ashgate, 2004:68e80.

[3] Dzik WH. Emily Cooley lecture 2002: transfusion safety in the hospital. Transfusion 2003; 43:1190e8.

[4] McClelland DBL, Phillips P. Errors in blood transfusion in Britain: Survey of hospital haematology departments. BMJ 1994; 308: 1205e6

[5] The Joint Commission, Improving Patient and Worker Safety: Opportunities for Synergy, Collaboration and Innovation, The Joint Commission, Oakbrook Terrace, IL, November 2012

[6] G. Yucel, S. Cebi, B. Hoege, A.F. Ozok, A fuzzy risk assessment model for hospital information system implementation, Expert Syst. Appl. 39 (2011) 1211–1218.

[7] A.C. Cagliano, S. Grimaldi, C. Rafele, A systemic methodology for risk management in healthcare sector, Saf. Sci. 49 (2011) 695–708.

[8] W.A. Hyman, E. Johnson, Fault tree analysis of clinical alarms, J. Clin. Eng. (April/June) (2008) 85–94.

[9] A. Park, S.J. Lee, Fault tree analysis on handwashing for hygiene management, Food Control 20 (2009) 223–229.

[10] Z.A. Abecassis, L.M. McElroy, R.M. Patel, R. Khorzad, C. Carroll IV, S. Mehrotra, Applying fault tree analysis to the prevention of wrong-site surgery, J. Surg. Res. (2014) 1–7.

[11] D. Raheja, M.C. Escano, Reducing patient healthcare safety risks through fault tree analysis, J. Syst. Saf. (September–October) (2009).

[12] R. Ferdous, F. Khan, B. Veitch, P.R. Amyotte, Methodology for computer aided fuzzy fault tree analysis, Process Saf. Environ. Prot. 87 (2009) 217–226.

On Warranty Cost Analysis For a Software Reliability Model Via Phase Type Distribution

Y. Sarada and R. Shenbagam
•
CEGC, Anna University, Chennai, Tamil Nadu, India – 600 025
sarada@ annauniv.edu, senbagam@ymail.com

Abstract

This research work investigates an optimal software release problem via phase type distribution, warranty and risk cost analysis. The inter arrival time of software failure is assumed to be a phase type distribution. The PH-SRM is one of the most flexible models, which overarches the existing non-homogeneous Poisson process (NHPP) models, and can approximate any type of NHPP-based models with high accuracy. Based on the phase type Non- homogeneous Poisson Process (PHNHPP) formulation and using the renewal reward theorem, the long run average cost rate is obtained. As model parameter estimation is an important issue in developing software reliability models, the software failure parameter has been estimated by the moment matching method. Finally, a numerical example is provided to illustrate the theoretical results therein.

Keywords: phase type distribution, software reliability, method of moments, renewal reward theorem, optimal software release policy

I. Introduction

The most significant feature of commercial software is Reliability, since it quantifies software failures during the growth procedure. Software is examined from an assortment of judgment in the testing procedure. For example, functionality, reliability, usability and the defects located in the procedure should be defined before it is released to the society. Decreasing the development cost and improving the quality of software are notable facts in software testing management. Software managers are confronted with many complicated problems in software testing. The plan for software maintenance (patch schedule) and the decision for the software release (Okumoto and Goel (1980), Pham (1996)) are based on reliability. Therefore, software reliability is an excellent tool to estimate the number of bugs (Musa *et al*. (1987), Musa (1999), Pham (2000)). Previous efforts on software reliability models (SRMs) mainly focused on NHPPs owing to their mathematical tractability and varied modeling situations such as imperfect debugging, change points, testing efforts and fault detection process (Xie *et al*. (2007), Okamura *et al*. (2013)). For example, the debugging process is modelled as a counting process which follows Poisson distribution with a time dependent hazard rate function (Goel and Okumoto (1979); Littlewood (1981), Yamada *et al*. (1983), Goel (1985), Langberg and Singpurwalla (1985), Laprie *et al*. (1991), Gokhale and Trivedi (1998)).

One of the major issues addressed in NHPP based software reliability is that of determining the best model whose solution lies in the statistical methods encompassing fitting of the observed bug data and to decide the model parameters (Langberg and Singpurwalla (1985)). Working in this direction, Okamura and Dohi (2006, 2008) introduced the phase type software reliability model (PHSRM) wherein the fault detection time follows a phase type distribution with the underlying counting process following NHPP.

A warranty is an agreement between a buyer and a seller at the time of product sale. It is a detailed study of the reimbursement type for a given product at the time of occurrence of failures. Also, it plays a significant role to safeguard the customer's interest particularly in the case of complicated products such as automobiles or electric devices. Recently attention has been directed towards warranty policies and warranty cost modelling (Nguyen and Murthy (1984), Blischke and Murthy(1995)). In today's market, many goods like mobile phones, electronic items and home devices are sold with extended warranty policy, in which a few choices are available for the consumer at the expiry time of the free warranty period. Extended warranties present extra security in the event of expensive failures after the initial warranty period and thereby safeguard the buyer against inflation. Also, extended warranty has attracted significant attention among practitioners. Lam and Lam (2001) proposed an extended warranty model with options open to customers to obtain an optimal policy for the consumers.

Furthermore, in the software management scheme the most significant calculation is to find an optimum software release time, referred to as an optimal software release problem. An optimal release problem with warranty cost and reliability requirement was studied by Yamada (1994). Also, Jain and Handa (2001) developed a software reliability model by employing a Hybrid warranty policy. Zhang and Pham (1998) studied a software cost model under warranty with a risk cost due to software failure and a cost to remove each error detected in the software. Prince Williams (2007) derived optimal release time policies to predict the optimal release time of software using imperfect debugging phenomena and warranty. The optimal release problem with simulated cost and reliability requirements was further implemented by Okumoto *et al*. (2013).

From a thorough review of the existing literature done, a combination of Phase type distribution and warranty (fixed / extended) has not been employed anywhere in the literature, in the analysis of software reliability. Motivated by this and in order to fill the gap in the literature, following Okamura and Dohi (2006,2008), the two new features attempted in this research article are the parameter estimation through moment matching method and the long run average cost rate analysis under the combination of phase type distribution and extended warranty in the context of software reliability models.

The structure of the paper is organized as follows: Section II furnishes the basics for the related work. Section III gives a detailed problem description and model assumptions. An explicit expression for the long run average cost rate is obtained in Section IV, while parameter estimation is discussed in Section V. Further, Section VI provides the numerical illustrations. Finally Section VII presents the concluding remarks.

II. Basics

Software reliability model based on Non-Homogeneous Poisson process:

The NHPP modelling in the SRMs essentially treats a counting process of software failures / faults / bugs in software system testing. It is virtually similar to a functioning profile of released software (Musa 1999), which provides information regarding the number of software failures in the system testing vis-à-vis the software reliability in the working phase. Further, existing NHPP-based SRMs are categorized into finite and infinite models. In the finite models, the detection slowly reduces with testing time and ultimately becomes zero, while in the infinite model, it does not become zero, that is, the number of software faults infinitely increases with time.

Specifically, if M(t) represents the number of software faults by time t with F(t) as the cumulative distribution function(c.d.f) of the detection times of software faults while the random variable N is the total number of software faults with mean m, then the probability mass function (p.m.f) of M(t) is given by

$$P\{M(t)=k\}=P_k(t)=\frac{(m\,F(t))^k}{k!}e^{-m\,F(t)},k=0,1,2,... \tag{1}$$

(Refer Langberg and Singpurwalla (1985)).

2.1 PH-SRM

Okamura and Dohi (2006, 2008) introduced phase type distribution in SRMs in which the fault detection time is a PH distribution in the NHPP- based model, referred to as PH-SRM.

A PH distribution is defined as the time to absorption in a continuous-time Markov chain (CTMC) with one absorbing state. Let *Q* denote the infinitesimal generator matrix of the CTMC with one absorbing state. Without loss of generality, *Q* is assumed to be partitioned as follows:

$$Q=\begin{pmatrix} U & U^0 \\ 0 & 0 \end{pmatrix}$$

where *U* and U^0 correspond to transition rates of transient states and exit rates from transient states to the absorbing state, respectively and a probability vector (α,α_{m+1}) exists such that $F_{PH}(x)=1-\alpha.\exp(U.x).e$ for x≥0. Here e denotes a column vector of ones with an appropriate dimension (Neuts (1981)). Note that the exit vector U^0 is given by $U^0=-U$e. The transient states of PH distribution are often called phases. PH distribution is proved to be dense, so that it can approximate any probability distribution with any precision as the size of *U* (the number of phases) increases (Asmussen and Koole (1993)).

By substituting the c.d.f. of PH distribution into (1), the p.m.f. of PH-SRM is given by

$$P\{M(t)=k\}=P_k(t)=\frac{(m\,F_{PH}(t))^k}{k!}e^{-m\,F_{PH}(t)},k=0,1,2,... \tag{2}$$

Let

$$\Lambda(t)=mF_{PH}(t)=\left(1-\pi\exp(Ut)e\right)m \tag{3}$$

so that

$$P\{M(t)=k\}=P_k(t)=\frac{(\Lambda(t))^k}{k!}e^{-\Lambda(t)},k=0,1,2,... \tag{4}$$

Thus PH-SRM exactly comprises of NHPP-based SRMs whose fault detection time distributions are exponential (Goel & Okumoto 1979), *k*-stage Erlang (Yamada et al. 1983; Khoshgoftaar 1988; Zhao & Xie 1996), hyperexponential (Laprie et al. 1991) and hypoexponential distributions (Fujiwara & Yamada 2001).

III. Model assumptions

We make the following assumptions about software reliability model for a phase type distribution in the context of warranty (fixed and an extended) modelling.

i. C_1: The set up cost of software development process is a constant.
ii. C_2: The cost to remove errors during debugging period is proportional to the total time of removing all errors detected during this period.
iii. C_3: The cost to remove errors during fixed warranty period (T_W) is proportional to the total time of removing all errors detected in the time interval [T, T+T_W].
iv. C_4: The cost to remove errors during extended warranty period (T_E) is proportional to the total time of removing all errors detected in the time interval [T+T_W, T+ T_W+T_E].
v. C_5: Risk cost due to the software failure after its release.
vi. It takes random time to remove errors and hence it is assumed that the time to remove each error follows a phase type distribution.

IV. Cost analysis

No efforts were made previously to carry out the cost analysis under warranty (fixed / extended) and phase type modeling in SRMs. Thus the goal of cost analysis is to obtain the long-run average cost rate for the proposed warranty model. By applying the standard result based on the renewal reward theorem, the long-run average cost per unit time is given by, (Ross, 1996):

$$\text{C(T)} = \frac{\text{Expected cost incurred in a cycle}}{\text{Expected length of a cycle}} = \frac{E(C)}{E(L)} \tag{5}$$

Here, the expected length of the cycle E(L) is given by,

$$E(L) = T + T_W + T_E \tag{6}$$

Next, the expected cost in a cycle E(C) can be expressed as,

$$E(C) = E(C_1) + E(C_2) + E(C_3) + E(C_4) + E(C_5) \tag{7}$$

In what follows, the calculations pertaining to the cost analysis are presented.

- $E(C_2)$:

$y_i = y_0 + (i-1)Y, \quad \text{i} = 1,2,3,....$

Let y_i be the cost of fixing i^{th} software fault. It consists of a deterministic part y_0 and an incremental random part (i-1) Y, where Y is a phase type random variable with mean $\mu_Y = -\chi S^{-1} e_2$. Note that the cost of fixing a fault is increasing as the number of faults removed is increasing. This is reasonable because it may become difficult to identify and fix a fault that occurs in the later testing phases.

Let N(T) denote the number of detected errors removed by time T, so that:

$$E(C_2) = C_2 \left\{ E\left[\sum_{i=1}^{N(T)} \left(y_0 + (i-1)Y \right) \right] \right\}$$

$$= C_2 \left\{ \sum_{n=1}^{\infty} E\left[\sum_{i=1}^{N(T)} \left(y_0 + (i-1)Y \mid N(T) = n \right) P[N(T) = n] \right] \right\}$$

$$= C_2 \left\{ \frac{1}{2} \left\{ \sum_{n=1}^{\infty} n[2y_0 + \mu_Y] P_n(T) + \sum_{n=1}^{\infty} n^2 \mu_Y P_n(T) \right\} \right\}$$

Employing (3) we have,

$$E(C_2) = C_2\left\{\frac{1}{2}\left[2y_0\Lambda(T) + \mu_Y\left(\Lambda(t)\right)^2\right]\right\} \tag{8}$$

- $E(C_3)$:

$$E(C_3) = C_3\left\{E\left[\sum_{N(T)}^{N(T+T_W)} W_i\right]\right\}.$$

Using $E\left[\sum_{N(T)}^{N(T+T_W)} W_i\right] = \mu_W\left\{\Lambda(T+T_W) - \Lambda(T)\right\}$, we have,

$$E(C_3) = C_3\left\{\mu_W\left[\Lambda(T+T_W) - \Lambda(T)\right]\right\} \tag{9}$$

- $E(C_4)$:

$$E(C_4) = C_4\left\{\mu_E\left[\Lambda(T+T_W+T_E) - \Lambda(T+T_W)\right]\right\} \tag{10}$$

- $E(C_5)$:

$$E(C_5) = C_5\left\{1 - R(x/T)\right\}$$

Here R(x/T) is the software reliability expressed as:

$$R(x/T) = e^{-[\Lambda(T+x)-\Lambda(T)]} = e^{-m[\alpha e^{UT}e_1 - \alpha e^{U(T+x)}e_1]},$$

so that

$$E(C_5) = C_5\left\{1 - e^{-m[\alpha e^{UT}e_1 - \alpha e^{U(T+x)}e_1]}\right\} \tag{11}$$

Now, employing (8) – (11), the expected total software cost E[C] can be expressed as in (12).

$$\begin{aligned} E(C) = {} & C_1 + C_2\left[\frac{1}{2}\left[2y_0\Lambda(T) + \mu_Y\left(\Lambda(t)\right)^2\right]\right] + C_3\left[\mu_W\left(\Lambda(T+T_W) - \Lambda(T)\right]\right. \\ & + C_4\left[\mu_E\left(\Lambda(T+T_W+T_E\right) - \Lambda(T+T_W)\right] \\ & + C_5\left[1 - e^{-m\left[\alpha e^{UT}e_1 - \alpha e^{U(T+x)}e_1\right]}\right] \end{aligned} \tag{12}$$

The optimal T* can be determined from (5), by using numerical or analytical methods.

Optimization:

This section is devoted to obtain the optimal software release time T* using the concept of pseudo-convexity of a function.

Lemma:

Using the convexity property of C(T), the optimal software release time T* is determined by solving the following equation.

$$\left\{[T+T_W+T_E]\left[\begin{array}{l} C_2\left(y_0\Lambda'(T)+\mu_Y\Lambda(T)\Lambda'(T)\right)+C_3\mu_W\left(\Lambda'(T+T_W)-\Lambda'(T)\right) \\ +C_4\left[\mu_E\left(\Lambda'(T+T_W+T_E)-\Lambda'(T+T_W)\right)\right] \\ +C_5 m\left(\left(\alpha e^{UT}Ue_1-\alpha e^{U(T+x)}Ue_1\right)e^{-m[\alpha e^{UT}e_1-\alpha e^{U(T+x)}e_1]}\right) \end{array}\right]\right\}- \tag{13}$$

$$\left\{\begin{array}{l} C_2\left[\frac{1}{2}\left[2y_0\Lambda(T)+\mu_Y(\Lambda(t))^2\right]\right]+C_3\left[\mu_W\left(\Lambda(T+T_W)-\Lambda(T)\right]\right. \\ +C_4\left[\mu_E\left(\Lambda(T+T_W+T_E)-\Lambda(T+T_W)\right]\right. \\ +C_5\left[1-e^{-m\left[\alpha e^{UT}e_1-\alpha e^{U(T+x)}e_1\right]}\right] \end{array}\right\}=C_1+C_5$$

Proof

Consider the derivative of C(T) with respect to T as given below:

$$\left\{[T+T_W+T_E]\left[\begin{array}{l} C_2\left(y_0\Lambda'(T)+\mu_Y\Lambda(T)\Lambda'(T)\right)+C_3\mu_W\left(\Lambda'(T+T_W)-\Lambda'(T)\right) \\ +C_4\left[\mu_E\left(\Lambda'(T+T_W+T_E)-\Lambda'(T+T_W)\right)\right] \\ +C_5 m\left(\left(\alpha e^{UT}Ue_1-\alpha e^{U(T+x)}Ue_1\right)e^{-m[\alpha e^{UT}e_1-\alpha e^{U(T+x)}e_1]}\right) \end{array}\right]\right\}-$$

$$\left\{\begin{array}{l} C_2\left[\frac{1}{2}\left[2y_0\Lambda(T)+\mu_Y(\Lambda(t))^2\right]\right]+C_3\left[\mu_W\left(\Lambda(T+T_W)-\Lambda(T)\right]\right. \\ +C_4\left[\mu_E\left(\Lambda(T+T_W+T_E)-\Lambda(T+T_W)\right]\right. \\ +C_5\left[1-e^{-m\left[\alpha e^{UT}e_1-\alpha e^{U(T+x)}e_1\right]}\right] \end{array}\right\}=C_1+C_5$$

If the cost function given in (5) is pseudo-convex, then it has only one local minimum and thus there will exist a unique global minimum. Now consider (12):

$$\frac{d}{dT}E(C)=\left[\begin{array}{l} C_2\left(y_0\Lambda'(T)+\mu_Y\Lambda(T)\Lambda'(T)\right)+C_3\mu_W\left(\Lambda'(T+T_W)-\Lambda'(T)\right) \\ +C_4\left[\mu_E\left(\Lambda'(T+T_W+T_E)-\Lambda'(T+T_W)\right)\right] \\ +C_5 m\left(\left(\alpha e^{UT}Ue_1-\alpha e^{U(T+x)}Ue_1\right)e^{-m[\alpha e^{UT}e_1-\alpha e^{U(T+x)}e_1]}\right) \end{array}\right]>0$$

$$\frac{d^2}{dT^2}E(C)=\left[\begin{array}{l} C_2\left(y_0\Lambda''(T)+\mu_Y(\Lambda'(T))^2+\mu_Y\Lambda(T)\Lambda''(T)\right)+C_3\mu_W\left(\Lambda''(T+T_W)-\Lambda''(T)\right) \\ +C_4\left[\mu_E\left(\Lambda''(T+T_W+T_E)-\Lambda''(T+T_W)\right)\right] \\ +C_5 m\left(\begin{array}{l}\left(\alpha e^{UT}U^2e_1-\alpha e^{U(T+x)}U^2e_1-m\left(\alpha e^{UT}Ue_1-\alpha e^{U(T+x)}Ue_1\right)^2\right) \\ \left(e^{-m[\alpha e^{UT}e_1-\alpha e^{U(T+x)}e_1]}\right)\end{array}\right) \end{array}\right]$$

$$\frac{d^2}{dT^2}E(C)>0$$

Hence E(C) is positive and convex. Since C(T) is linear in T and positive, C(T) is pseudo- convex in E(C). Hence the lemma.

V. Parameter estimation

In this section statistical estimation of the model parameters is to be performed which will take the modelling closer to the realm of applicability. Methods based on the estimation of the parameters for a phase type distribution, can be classified into three types such as (i) moment matching method (MM), (ii) maximum likelihood (ML) estimation and (iii) Bayes estimation. In this section, the method of moments (MM) is mainly used for the PH distribution. The concept of MM method is to find PH parameters so that the population moments can fit the moments derived from samples or probability density function of the true distribution. As mentioned before, the accuracy of MM method depends on the number of moments used.

Consider the Hyper - exponential distribution of order 2, that can be expressed as, $\Lambda(t) = mF(t) = (1 - \pi \exp(Ut)e)m$, so that

$$\pi = [p \quad 1-p],\ \mathrm{U} = \begin{bmatrix} -\frac{1}{\lambda_1} & 0 \\ 0 & -\frac{1}{\lambda_2} \end{bmatrix}, \mathrm{e} = [1 \quad 1]`.$$

Therefore, the p.d.f and the first four moments are given by:

$$f(t) = \Lambda'(t) = m\left[\frac{p}{\lambda_1} e^{-\frac{t}{\lambda_1}} + \frac{(1-p)}{\lambda_2} e^{-\frac{t}{\lambda_2}}\right], t \geq 0, 0 \leq \mathrm{p} < 1 .$$

$$m_1 = m[p\lambda_1 + (1-p)\lambda_2] \tag{14}$$

$$m_2 = m[2p\lambda_1^2 + 2(1-p)\lambda_2^2] \tag{15}$$

$$m_3 = m[6p\lambda_1^3 + 6(1-p)\lambda_2^3] \tag{16}$$

$$m_4 = m[24p\lambda_1^4 + 24(1-p)\lambda_2^4] \tag{17}$$

From (14) we have,

$$\lambda_2 = \frac{m_1 - pw\lambda_1}{m(1-p)} \tag{18}$$

Substituting (18) in (15), and after some simple calculations,

$$\lambda_1 = \frac{m_1}{w}\left\{1 \pm \sqrt{\frac{(m_2 m - 2m_1^2)(1-p)}{2pm_1^2}}\right\} \tag{19}$$

We have two cases:

Case (i)

When

$$\lambda_1 = \frac{m_1}{m}\left\{1 - \sqrt{\frac{(m_2 m - 2m_1^2)(1-p)}{2pm_1^2}}\right\} \tag{20}$$

Substituting (20) in (18) we have,

$$\lambda_2 = \frac{m_1}{m}\left\{1+\sqrt{\frac{(m_2 m - 2m_1^2)(p)}{2(1-p)m_1^2}}\right\} \tag{21}$$

Since (20) and (21) involve λ_1 and λ_2 expressed in terms of the unknown parameters p and m only. Using (16) and (17) the following system of equations are obtained:

$$m_3 - \frac{6m_1^3}{m^2}\left\{\frac{3m_2 m}{2m_1^2} - 2 + \left(\frac{m_2 m - 2m_1^2}{2m_1^2}\right)^{\frac{3}{2}} \frac{2p-1}{\sqrt{p(1-p)}}\right\} = 0 \tag{22}$$

$$m_4 - \frac{24m_1^4}{m^3}\left\{\frac{m^2 m_3}{3m_1^3} - 1 + \left[\left(\frac{m_2 m - 2m_1^2}{2m_1^2}\right)^{\frac{3}{2}} \frac{2(2p-1)}{\sqrt{p(1-p)}}\right] + \left[\left(\frac{m_2 m - 2m_1^2}{2m_1^2}\right)^2 \left(\frac{3p^2-3p+1}{p(1-p)}\right)\right]\right\} = 0 \tag{23}$$

Thus, (22) and (23) are to be solved to get the parameters m and p. Subsequently the remaining parameters are obtained by the method of back substitution.

Case (ii)

Proceeding similarly as in case (i), we have,

$$\lambda_1 = \frac{m_1}{m}\left\{1+\sqrt{\frac{(m_2 m - 2m_1^2)(1-p)}{2pm_1^2}}\right\} \tag{24}$$

$$\lambda_2 = \frac{m_1}{m}\left\{1-\sqrt{\frac{(m_2 m - 2m_1^2)p}{2(1-p)m_1^2}}\right\} \tag{25}$$

$$m_3 - \frac{6m_1^3}{m^2}\left\{\frac{3m_2 m}{2m_1^2} - 2 + \left(\frac{m_2 m - 2m_1^2}{2m_1^2}\right)^{\frac{3}{2}} \frac{1-2p}{\sqrt{p(1-p)}}\right\} = 0 \tag{26}$$

$$m_4 - \frac{24m_1^4}{m^3}\left\{\frac{m^2 m_3}{3m_1^3} - 1 + \left[\left(\frac{m_2 m - 2m_1^2}{2m_1^2}\right)^{\frac{3}{2}} \frac{2(1-2p)}{\sqrt{p(1-p)}}\right] + \left[\left(\frac{m_2 m - 2m_1^2}{2m_1^2}\right)^2 \left(\frac{3p^2-3p+1}{p(1-p)}\right)\right]\right\} = 0 \tag{27}$$

It is easily seen that both cases lead to the same result, but with the roles of p and 1-p interchanged. Hence, the parameter estimation carried out may help the practitioners in making flexible decisions for the proposed software reliability model via phase type distribution.

Additionally a numerical example is provided to demonstrate the applicability of the proposed software reliability model with warranty in Section III.

VI. Numerical illustration and sensitivity analysis

In this section, a numerical example is given to illustrate the impact of combining the phase type distribution and an extended warranty in the field of software reliability. The first four moments are assumed to be:

$$m_1 = 2.0125\times10^4, m_2 = 1.62125\times10^7, m_3 = 1.9607\times10^{10}, m_4 = 3.1644\times10^{13}.$$

The inter arrival times of software faults are assumed to be hyper exponential. Employing the moment matching method and using the above moments, the estimated values are obtained as:

$$w = 50,\ \pi = [0.95 \quad 0.05],\ \mathrm{U} = \begin{bmatrix} -\frac{1}{400} & 0 \\ 0 & -\frac{1}{450} \end{bmatrix}, \mathrm{e} = [1 \quad 1]`.$$

Also, assume that,

$$C_1 = 5000, C_2 = 50, y_0 = 0.5, \mu_y = 0.9, \mu_W = 0.95, \mu_E = 0.85, T_W = 500, T_E = 400, x = 1.5, C_3 = 360,$$
$$C_4 = 200, C_5 = 500$$

Utilizing the above parameter values in (5), the result has been presented in Table 1. The optimal software release time T*=67 and the corresponding cost 10.0249 is depicted in Table 1. Additionally, the sensitivity analysis of a proposed software reliability model using C(T) is analyzed. Tables 2 and 3 illustrate the sensitiveness of the long run average cost rate. Table 2 shows that as C_1, C_2 , C_3 and C_5 increase respectively, the long run average cost rate increases while an increase in C_4 results in a decrease in C(T). In a similar manner, Table 3 indicates that the long run average cost rate increases as x, μ_Y, μ_W and y_0 increase and decreases as the parameter μ_E increases. Further, Tables 2 and 3 illustrate that the optimal software release time T* increases with an increase in the parameters C_1, C_3, x and μ_W while it decreases as the parameters C_2, C_4, μ_Y, μ_E and y_0 increase. Also, the optimal software release time T* is unchanged as the parameter C_5 increases. Thus, the sensitivity analysis of such parameters may aid the software system manager in making decisions to model the software system testing.

Table 1: *C(T) versus T*

T	C(T)	T	C(T)	T	C(T)	T	C(T)	T	C(T)
20	10.6718	45	10.1554	62	10.0316	67	***10.0249***	80	10.0631
25	10.5325	50	10.1018	63	10.0292	68	10.0250	85	10.0975
30	10.4121	55	10.0629	64	10.0274	69	10.0257	90	10.1420
35	10.3097	60	10.0379	65	10.0261	70	10.0268	95	10.1959
40	10.2244	61	10.0345	66	10.0252	75	10.0393	100	10.2587

Table 2 : *Sensitivity analysis for the parameters C_1, C_2, C_3, C_4 and C_5 in C(T)*

C_1	T*	C(T*)	C_2	T*	C(T*)	C_3	T*	C(T*)	C_4	T*	C(T*)	C_5	T*	C(T*)
4000	65	8.9898	40	88	9.6337	**360**	**67**	**10.024**	150	69	11.882	400	67	10.0098
5000	**67**	**10.024**	**50**	**67**	**10.024**	400	76	11.189	**200**	**67**	**10.024**	**500**	**67**	**10.0240**
6000	69	11.057	60	54	10.288	450	87	12.595	250	65	8.1645	600	67	10.0400
7000	72	12.088	70	45	10.478	500	98	13.947	300	63	6.3018	700	67	10.0550

Table 3: *Sensitivity analysis for the parameters x, μ_y, μ_W, μ_E and y_0 in C(T)*

X	T*	C(T*)	μ_y	T*	C(T*)	μ_W	T*	C(T*)	μ_E	T*	C(T*)	y_0	T*	C(T*)
0.5	67	9.9760	**0.9**	**67**	**10.024**	.75	52	7.7208	.75	68	10.899	**.5**	**67**	**10.024**
1.5	**67**	**10.024**	1.1	54	10.276	.85	59	8.8888	**.85**	**67**	**10.024**	1.5	57	10.3964
2.5	67	10.068	1.3	45	10.451	**.95**	**67**	**10.024**	.95	66	9.1497	2.5	47	10.7144
3.5	68	10.108	1.5	39	10.581	1.05	76	11.129	1.05	65	8.2740	3.5	37	10.9770

VII. Conclusion

In this study, a phase type Non – homogeneous Poisson process software reliability model that incorporates warranty (fixed and an extended) via long run average cost rate in order to determine the optimal software release time is developed. Hence, PH-SRM is a promising tool to reduce the effort to select the best models in software reliability assessment. Also, the parameter estimation was carried out using moment matching method. The graphical illustrations conform to the theoretical observations made earlier. Additionally, a sensitivity analysis has been carried out for all the parameters, to exemplify the optimal software release policy (T*) and the corresponding long run average cost rate C(T*). To the best of authors' knowledge, it is observed that, phase type distribution in the cost analysis has not been studied from the view point of software reliability systems with fixed and extended warranty model. As a final remark, the extended warranty model enables the software manger to decide on whether the software is sufficiently tested to allow its release or unrestricted use. Such predictions provide a quantitative basis for achieving reliability, risk and cost goals.

References

[1] Asmussen, S. and Koole, G. (1993). Marked point processes as limits of Markovian arrival streams. *Journal of Applied Probability,* 30: 365–372.

[2] Blischke, W. and Murthy, D. Product Warranty Handbook, Marcel Dekker, New York, 1995.

[3] Fujiwara, T. and Yamada, S. (2001). Testing - domain dependent software reliability growth models and their comparisons of goodness-of-fit. *In Proceedings of the 7th ISSAT international conference on reliability and quality in design* (pp. 36–40).

[4] Goel, A. L. and Okumoto, K. (1979). Time – dependent error-detection rate model for software reliability and other performance measures. *IEEE Transactions on Reliability, R–28*: 206–211.

[5] Goel, A. L. (1985). Software reliability models: Assumptions, limitations and applicability. *IEEE Transactions on Software Engineering, SE–11*: 1411–1423.

[6] Gokhale, S. S. and Trivedi, K. S. (1998). Log-logistic software reliability growth model. *In Proceedings of the 3rd IEEE international high – assurance systems engineering symposium (HASE 1998)* (pp. 34–41).

[7] Jain, M. and Handa, B. R. (2001). Cost analysis for repairable units under hybrid warranty. Recent developments in Operational Research. (ed. Manju Lata Agarwal and Kanwar Sen), New Delhi, Narosa Publishing House, 149-165.

[8] Khoshgoftaar, T. M. (1988). Non - Homogeneous Poisson processes for software reliability growth. *In Proceedings of the international conference on computational statistics (COMPSTAT)* (pp. 13–14).

[9] Lam, Y. and Lam, P. (2001). An extended warranty policy with options open to consumers. *European Journal of Operational Research,* 131: 514-529.

[10] Langberg, N. and Singpurwalla, N. D. (1985). Unification of some software reliability models. SIAM Journal on Scientific Computing 6 (3): 781–790.

[11] Laprie, J. C., Kanoun, K., Béounes, C. and Kaâniche, M. (1991). The KAT (knowledge-action - transformation) approach to the modelling and evaluation of reliability and availability growth. *IEEE Transactions on Software Engineering,* 17(4): 370–382.

[12] Littlewood, B. (1981). Stochastic reliability growth: A model for fault removal in computer programs and hardware designs. *IEEE Transactions on Reliability, 30* (10):313–320.

[13] Musa, J. D., Iannino, A. and Okumoto, K. *Software reliability, measurement, prediction application*. New York: McGraw-Hill, 1987.

[14] Musa, J. D. *Software reliability engineering*. New York: McGraw-Hill, 1999.

[15] Nguyen, D. G. and Murthy, D. (1984). Cost analysis of warranty policies. *Naval Research Logistics Quaterly*, 31:525-541.

[16] Neuts, M. F. *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*. The John Hopkins University Press, Baltimore, 1981.

[17] Okamura, H. and Dohi, T. (2006). Building phase - type software reliability model. In *Proceedings of the 17th international symposium on software reliability engineering (ISSRE 2006)*, 289–298.

[18] Okamura, H. and Dohi, T. (2008). Hyper-Erlang software reliability model. *Proceedings of the 14th Pacific Rim international symposium on dependable computing (PRDC 2008),* 232–239.

[19] Okumoto, K. and Goel, L. (1980). Optimum release time for software systems based on reliability and cost criteria. *Journal of Systems and Software*, 1: 315–318.

[20] Okamura, H., Dohi, T. and Osaki, S. (2013). Software reliability growth models with normal failure time distributions. *Reliability Engineering and System Safety,* 116:135–141.

[21] Prince Williams, D. R. (2007). Study of the Warranty Cost Model for Software Reliability with an Imperfect Debugging Phenomenon. *Turkish Journal of Electrical Engineering & Computer Sciences,* 15 (3): 369-381.

[22] Pham, H. (1996). A software cost model with imperfect debugging, random life cycle and penalty cost. *International Journal of System Science,* 27(5): 455–463.

[23] Pham, H. *Software Reliability*. Springer, Singapore, 2000.

[24] Ross, S. M. *Stochastic Processes*. Second edition. Wiley, New York, 1996.

[25] Xie, M., Hu, Q. P., Wu, Y. P. and Ng, S. H. (2007). A study of the modeling and analysis of software fault detection and fault - correction processes. *Quality and Reliability Engineering International,* 23:459–470.

[26] Yamada, S., Ohba, M. and Osaki, S. (1983). S-shaped reliability growth modeling for software error detection. *IEEE Transactions on Reliability,* 32: 475–478.

[27] Yamada, S. (1994). Optimal release problems with warranty period based on a software maintenance cost model. *Transactions of Information Processing Society of Japan,* 35 (9), 2197-2202.

[28] Zhao, M. and Xie, M. (1996). On maximum likelihood estimation for a general non-homogeneous Poisson process. *Scandinavian Journal of Statistics, 23:*597–607.

[29] Zhang, X. and Pham, H. (1998). A software cost model with warranty cost, error removal times and risk costs. *IIE Transactions*, 30 (12): 1135-1142.

A Generalization Of Weibull Distribution

Rama Shanker, Kamlesh Kumar Shukla*

•

Department of Statistics
College of Science, Eritrea Institute of Technology, Asmara, Eritrea
Email: shankerrama2009@gmail.com; kkshukla22@gmail.com
*Corresponding Author

Abstract

In this paper, a generalization of Weibull distribution (GWD), which includes Weibull and exponential distributions as special cases, has been proposed and investigated. Its moments, hazard rate function and stochastic ordering have been discussed. The method of maximum likelihood estimation has been discussed for estimating its parameters. The goodness of fit of the proposed distribution has been discussed with a real lifetime dataset and the fit has been found quite satisfactory over some well-known lifetime distributions.

Keywords: Exponential distribution, Weibull distribution, Moments, Hazard rate function, Stochastic ordering, Maximum likelihood estimation, Application.

2000 Mathematics Subject Classification: 62E05, 62E99

1. Introduction

The exponential distribution having scale parameter θ is defined by its probability density function (pdf) and cumulative distribution function (cdf)

$$f(y;\theta)=\theta e^{-\theta y};\quad y>0,\theta>0 \tag{1.1}$$

$$F(y;\theta)=1-e^{-\theta y}\ ;y>0,\theta>0, \tag{1.2}$$

Epstein (1958) has detailed study on exponential distribution and its role in life testing. Sato *et al* (1999) obtained a discrete exponential distribution and discussed its properties and applied the distribution to model defect count distribution in semi-conductor deposition equipment and defect count distribution per chips. Gupta and Kundu (1999) have obtained a two-parameter generalized exponential distribution (GED) and discussed its statistical properties, estimation of parameter and applications. Nekoukhou *et at* (2012) obtained a discrete analogue of the generalized exponential distribution and discussed its properties, estimation of parameters and applications. Most of the research works done by different researchers on exponential distributions are available in Ahsanullah and Hamedani (2010).

It should be noted that the two-parameter gamma distribution is the weighted exponential distribution and are defined by its pdf and cdf

$$f(y;\theta,\alpha)=\frac{\theta^{\alpha}}{\Gamma(\alpha)}e^{-\theta y}y^{\alpha-1};\quad y>0,\theta>0,\alpha>0\ , \tag{1.3}$$

$$F(y;\theta,\alpha)=1-\frac{\Gamma(\alpha,\theta y)}{\Gamma(\alpha)}\;;y>0,\theta>0,\alpha>0\;, \tag{1.4}$$

where θ is a scale parameter, α is a shape parameter and the function $\Gamma(\alpha,z)$ is the upper incomplete gamma function defined as

$$\Gamma(\alpha,z)=\int_z^\infty e^{-y}\,y^{\alpha-1}\,dy;\alpha>0,\;z\ge 0 \tag{1.5}$$

Chakraborty and Chakraborty (2012) obtained a discrete analogue of the two-parameter gamma distribution and discussed its properties, parameters estimation and applications.
A two-parameter Weibull distribution (WD) introduced by Weibull (1951) having scale parameter θ and shape parameter β is defined by its pdf and cdf

$$f(x;\theta,\beta)=\theta\,\beta\,x^{\beta-1}\,e^{-\theta x^{\beta}}\;\;;x>0,\theta>0,\beta>0 \tag{1.6}$$

$$F(x;\theta,\beta)=1-e^{-\theta x^{\beta}}\;\;;x>0,\theta>0,\beta>0 \tag{1.7}$$

It can be easily shown that the Weibull distribution reduces to classical exponential distribution at $\beta=1$. It should be noted that Weibull distribution is nothing but the power exponential distribution. Taking $x=y^{1/\beta}$ and thus $y=w(x)=x^{\beta}$ in (1.1), we have

$$g(x;\theta,\beta)=f\left[w(x)\right]w'(x)=\theta e^{-\theta x^{\beta}}\,\beta\,x^{\beta-1}=\theta\,\beta\,x^{\beta-1}e^{-\theta x^{\beta}}\;, \tag{1.8}$$

which is the pdf of Weibull distribution defined in (1.6). Nakagawa and Osaki (1975) obtained discrete Weibull distribution. Stein and Dattero (1984) introduced a new discrete Weibull distribution. Recently Shanker *et al* (2016) have detailed critical and comparative study on modeling of lifetime data using two-parameter gamma and Weibull distributions and it has been observed that in some datasets gamma gives better fit than Weibull whereas in some datasets Weibull gives better fit than gamma and hence these two distributions are competing each other. Most of the research works done on Weibull distributions are available in Murthy *et al* (2004). The Weibull distribution has been modified and extended to three, four and five parameters by introducing additional parameters to suite to specific set of data by different researchers including Mudholkar and Srivastava (1993), Mudholkar and Kollia 91994), Mudholkar et al (1995), Marshall and Olkin (1997), Lai *et al* (2003), Nadarajah and Kotz (2005), are some among others. Most of the research works done regarding generalizations, extension and modifications of Weibull distributions are available in Lai et al (2011) and Alamalki and Nadarajah (2014). After careful reading of these papers it has been observed that a simple generalization of Weibull distribution can be done which will be different from the previous ones.
The pdf and the cdf of generalized gamma distribution (GGD) introduced by Stacy (1962) having parameters $\theta,\alpha,$ and β is given by

$$f(x;\theta,\alpha,\beta)=\frac{\beta\,\theta^{\alpha}}{\Gamma(\alpha)}x^{\beta\alpha-1}\,e^{-\theta x^{\beta}};\;x>0,\theta>0,\alpha>0,\beta>0 \tag{1.9}$$

$$F(x;\theta,\alpha,\beta,)=1-\frac{\Gamma\left(\alpha,\theta x^{\beta}\right)}{\Gamma(\alpha)};\;\;x>0,\theta>0,\alpha>0,\beta>0 \tag{1.10}$$

where α and β are the shape parameters and θ is the scale parameter, and $\Gamma(\alpha,z)$ is the upper incomplete gamma function defined in (1.5). Note that Stacy (1962) obtained GGD by taking taking $x=y^{1/\beta}$ and thus $y=w(x)=x^{\beta}$ in (1.3), and using the approach of obtaining the pdf of Weibull distribution in (1.6).

Clearly the gamma distribution, the Weibull distribution and the exponential distribution are particular cases of (1.9) for $(\beta=1)$, $(\alpha=1)$ and $(\alpha=\beta=1)$ respectively. Detailed discussion about GGD is available in Stacy (1962) and parametric estimation for the GGD has been discussed by Stacy and Mihram (1965). Chakraborty (2015) has obtained a new discrete gamma distribution corresponding to GGD and discussed some of its properties. Note that the research works done on exponential distribution, Weibull distribution, gamma distribution, GGD, their extensions and applications for life testing and modeling are available in Lee and Wang (2003). Recently Shanker and Shukla (2016) have detailed critical and comparative study on modeling of lifetime data using three-parameter GGD and generalized Lindley distribution (GLD) introduced by Zakerzadeh and Dolati (2009) and it has been observed that GGD gives much closer fit than GLD in majority of datasets.

Since Weibull distribution gives better fit than Gamma distribution, Lindley distribution introduced by Lindley (1958), Lognormal distribution and exponential distribution, it is expected and hoped that the generalization of Weibull distribution (GWD) will be a better model than three-parameter GGD and GLD, two-parameter Weibull distribution, Gamma distribution, Lognormal distribution, weighted Lindley distribution (WLD) introduced by Ghitany *et al* (2011), GED and one parameter Lindley and exponential distributions. Keeping this point in mind, an attempt has been made to obtain GWD which includes Weibull distribution and exponential distribution as special cases. Some of its properties including hazard rate function and stochastic ordering have been discussed. The estimation of its parameters has been discussed using maximum likelihood estimation. A real lifetime data has been presented to test the goodness of fit of GWD over GGD, GLD, WLD, GED, Weibull, Gamma, Lognormal, Lindley and exponential distributions and the fit has been found quite satisfactory over these considered distributions.

2. A Generalization of Weibull Distribution

Taking $x=\frac{1}{\alpha}y^{1/\beta}$ and thus $y=w(x)=(\alpha x)^{\beta}$ in (1.1), and following the approach of obtaining the pdf of Weibull distribution, the pdf of generalization of Weibull distribution (GWD) can be obtained as

$$f(x;\theta,\alpha,\beta)=\alpha\beta\theta(\alpha x)^{\beta-1}e^{-\theta(\alpha x)^{\beta}};\ x>0,\theta>0,\beta>0,\alpha>0 \tag{2.1}$$

where α,β are shape parameters and θ is the scale parameter. It can be easily verified that Weibull distribution of Weibull (1951) and exponential distribution are particular cases of GWD for $(\alpha=1)$ and $(\alpha=1,\beta=1)$ respectively.

The corresponding cdf of GWD can be obtained as

$$F(x;\theta,\alpha,\beta)=1-e^{-\theta(\alpha x)^{\beta}};\ x>0,\theta>0,\beta>0,\alpha>0 \tag{2.2}$$

Graphs of the pdf and the cdf of GWD are shown in figures 1 and 2 for varying values of the parameters θ,α and β. From figure 1, it is observed that pdf of GWD is taking different shapes including momotonically decreasing, positively skewed, platykurtic, leptokurtic,negatively skewed etc for varying values of parameters.

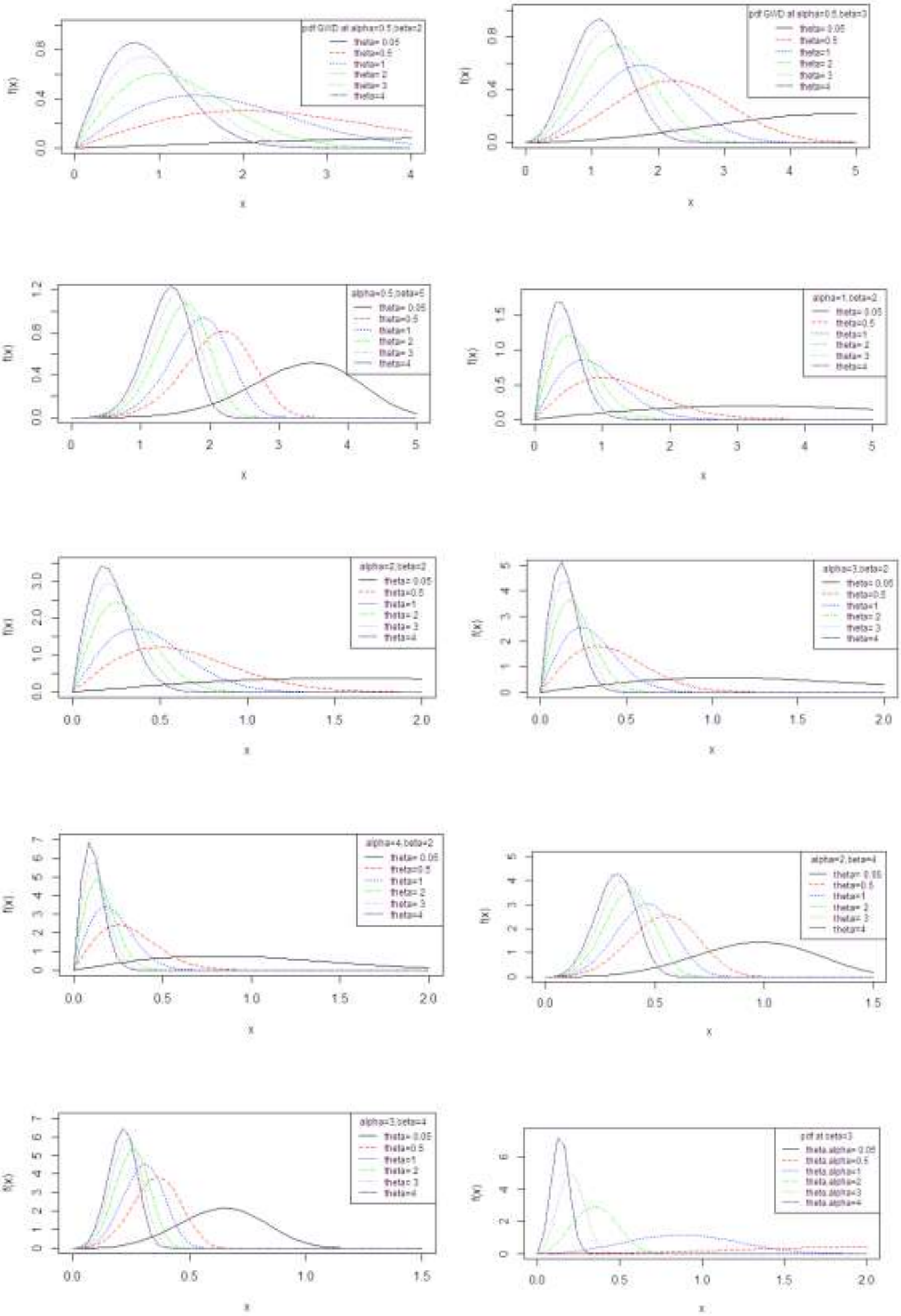


Fig.1. Graphs of pdf of GWD for varying values of parameters θ, α and β

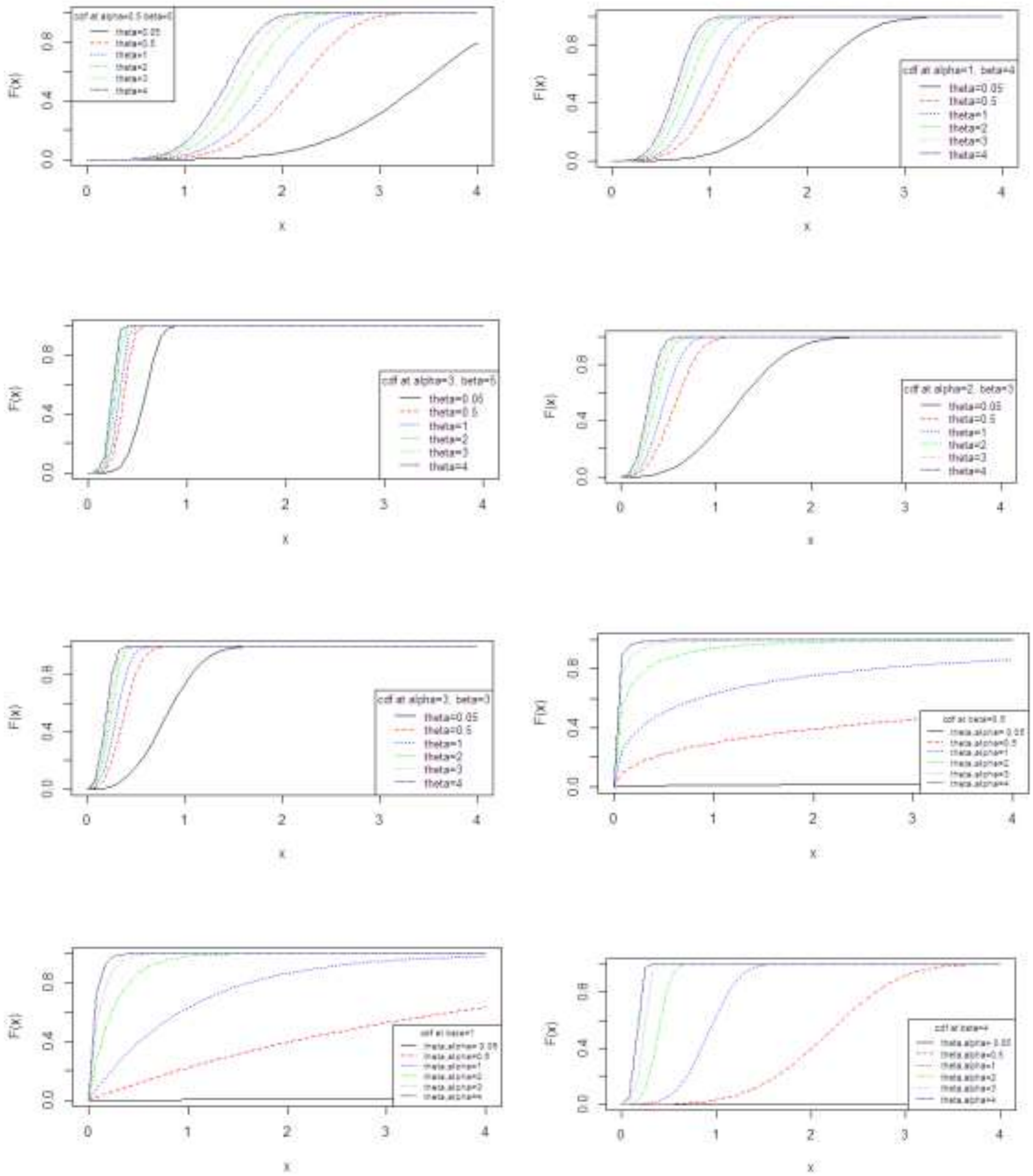


Fig.2. Graphs of cdf of GWD for varying values of parameters θ, α and β.

3. Moments

Using (2.1), the r th moment about origin of GWD can be obtained as

$$\mu_r' = \int_0^\infty x^r \alpha\, \beta\, \theta \left(\alpha x\right)^{\beta-1} e^{-\theta(\alpha x)^\beta} dx$$

Taking $u = \theta\left(\alpha\, x\right)^\beta$ and thus $x = \frac{1}{\alpha}\left(\frac{u}{\theta}\right)^{1/\beta}$ and $du = \alpha\, \theta\, \beta \left(\alpha\, x\right)^{\beta-1} dx$, we get

$$\mu_r' = \int_0^\infty \left\{ \frac{1}{\alpha}\left(\frac{u}{\theta}\right)^{1/\beta} \right\}^r e^{-u} du = \frac{1}{\alpha^r}\left(\frac{1}{\theta}\right)^{r/\beta} \int_0^\infty u^{\frac{r}{\beta}+1-1} e^{-u} du = \frac{\Gamma\left(\frac{r}{\beta}+1\right)}{\alpha^r \theta^{r/\beta}}; r = 1,2,3,... \qquad (3.1)$$

Thus the first two moments about origin and the variance of GWD are obtained as

$$\mu_1' = \frac{\Gamma\left(\frac{1}{\beta}+1\right)}{\alpha\, \theta^{1/\beta}}, \quad \mu_2' = \frac{\Gamma\left(\frac{2}{\beta}+1\right)}{\alpha^2\, \theta^{2/\beta}}$$

and $$\mu_2 = \frac{\Gamma\left(\frac{2}{\beta}+1\right)}{\alpha^2\, \theta^{2/\beta}} - \left(\frac{\Gamma\left(\frac{1}{\beta}+1\right)}{\alpha\, \theta^{1/\beta}} \right)^2 = \frac{1}{\alpha^2\, \theta^{2/\beta}} \left[\Gamma\left(\frac{2}{\beta}+1\right) - \left\{ \Gamma\left(\frac{1}{\beta}+1\right) \right\}^2 \right].$$

Similarly substituting $r = 3$ and 4 in (3.1), second and third moments about origin can be obtained. Then using the relationship between central moments and moments about origin, third and fourth central moments can be obtained. The expressions for third and fourth central moments are complicated and very big and hence they are not being presented here.

4. Hazard Rate Function

Suppose X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$ from 1.2 and 1.3. The hazard rate function (also known as the failure rate function) function of X is defined as

$$h(x) = \lim_{\Delta x \to 0} \frac{P(X < x + \Delta x \mid X > x)}{\Delta x} = \frac{f(x)}{1 - F(x)}$$

The corresponding $h(x)$ of GWD can be obtained as

$$h(x;\theta,\alpha,\beta) = \frac{f(x;\theta,\alpha,\beta)}{1 - F(x;\theta,\alpha,\beta)} = \alpha\, \beta\, \theta(\alpha\, x)^{\beta-1}; x > 0, \theta > 0, \alpha > 0, \beta > 0 \qquad (4.1)$$

Graphs of $h(x)$ of GWD for varying values of the parameters θ, α and β are shown in figure 3. The $h(x)$ of GWD are monotonically decreasing, constant, monotonically increasing or bathtub shape varying for varying values of parameters θ, α and β.

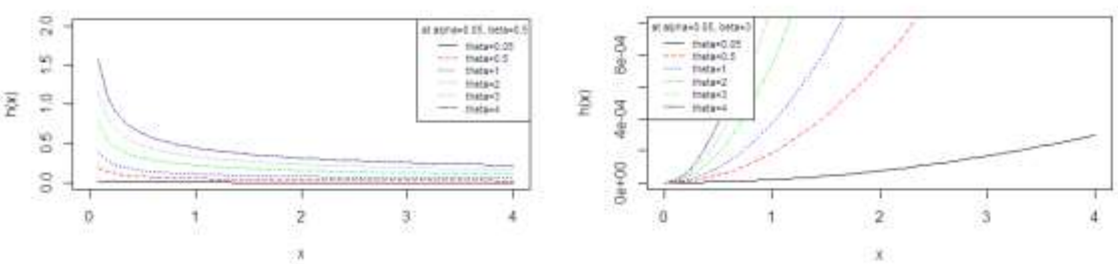


Fig.3. Graphs of $h(x)$ of GWD for varying of values of parameters θ, α and β.

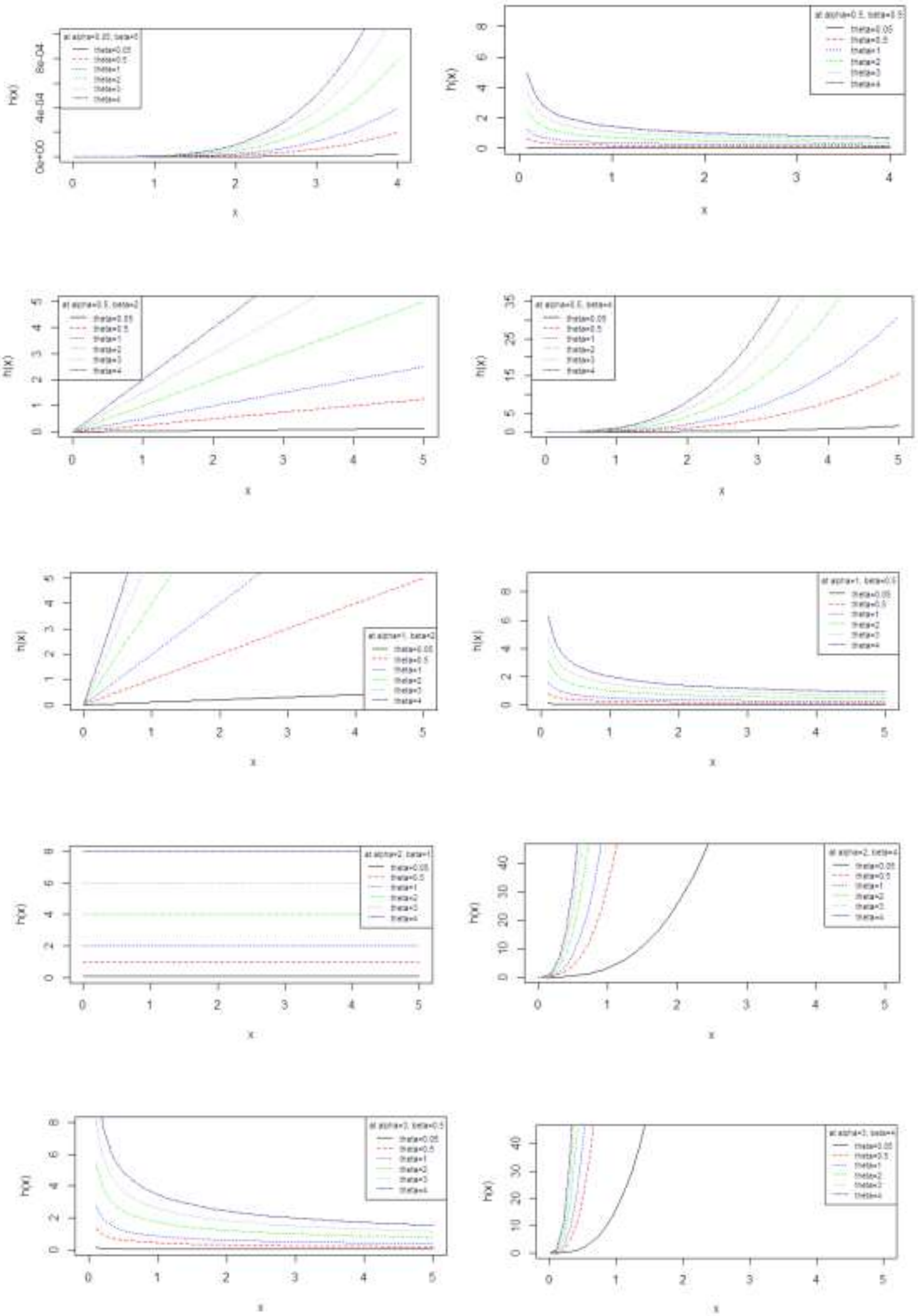


Fig.3. Graphs of $h(x)$ of GWD for varying of values of parameters θ, α and β (continuation)

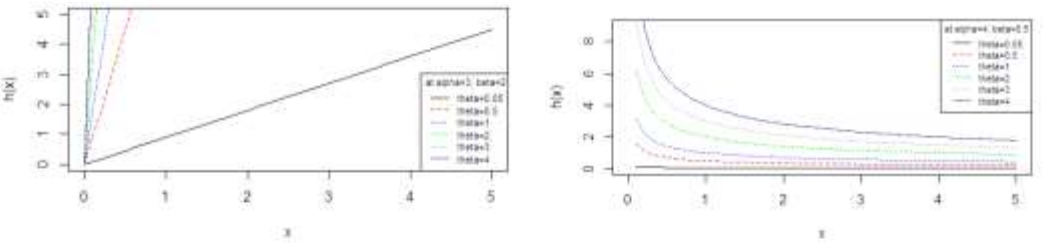


Fig.3. Graphs of $h(x)$ of GWD for varying of values of parameters θ, α and β (continuation)

5. Stochastic Ordering

Stochastic ordering of positive continuous random variables is an important tool for judging their comparative behavior. A random variable X is said to be smaller than a random variable Y in the

(i) stochastic order $\left(X \le_{st} Y\right)$ if $F_X\left(x\right) \ge F_Y\left(x\right)$ for all x

(ii) hazard rate order $\left(X \le_{hr} Y\right)$ if $h_X\left(x\right) \ge h_Y\left(x\right)$ for all x

(iii) mean residual life order $\left(X \le_{mrl} Y\right)$ if $m_X\left(x\right) \le m_Y\left(x\right)$ for all x

(iv) likelihood ratio order $\left(X \le_{lr} Y\right)$ if $\dfrac{f_X\left(x\right)}{f_Y\left(x\right)}$ decreases in x.

The following results due to Shaked and Shanthikumar (1994) are well known for establishing stochastic ordering of distributions

$$X \le_{lr} Y \Rightarrow X \le_{hr} Y \Rightarrow X \le_{mrl} Y$$
$$\Downarrow$$
$$X \le_{st} Y$$

GWD is ordered with respect to the strongest 'likelihood ratio ordering' as shown in the following theorem:

Theorem: Let X ~ GWD$\left(\theta_1, \alpha_1, \beta_1\right)$ and Y ~ GWD$\left(\theta_2, \alpha_2, \beta_2\right)$. If one of the following conditions satisfied

(i) $\theta_1 = \theta_2, \alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$ (ii) $\theta_1 = \theta_2, \beta_1 = \beta_2$ and $\alpha_1 > \alpha_2$

(iii) $\theta_1 > \theta_2, \alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$

then $X \le_{lr} Y$ and hence $X \le_{hr} Y$, $X \le_{mrl} Y$ and $X \le_{st} Y$.

Proof: We have

$$\frac{f_X\left(x;\theta_1,\alpha_1,\beta_1\right)}{f_Y\left(x;\theta_2,\alpha_2,\beta_2\right)} = \left(\frac{\beta_1\,\alpha_1^{\beta_1}\theta_1}{\beta_2\,\alpha_2^{\beta_2}\theta_2}\right) x^{\beta_1-\beta_2}\, e^{-\left\{\theta_1\alpha_1^{\beta_1}x^{\beta_1} - \theta_2\alpha_2^{\beta_2}x^{\beta_2}\right\}} \quad ;\ x > 0$$

Now

$$\ln \frac{f_X\left(x;\theta_1,\alpha_1,\beta_1\right)}{f_Y\left(x;\theta_2,\alpha_2,\beta_2\right)} = \ln\left(\frac{\beta_1\,\alpha_1^{\beta_1}\theta_1}{\beta_2\,\alpha_2^{\beta_2}\theta_2}\right) + \left(\beta_1 - \beta_2\right)\ln x - \left(\theta_1\alpha_1^{\beta_1}x^{\beta_1} - \theta_2\alpha_2^{\beta_2}x^{\beta_2}\right).$$

.

This gives $\frac{d}{dx}\ln\frac{f_X(x;\theta_1,\alpha_1,\beta_1)}{f_Y(x;\theta_2,\alpha_2,\beta_2)} = \frac{\beta_1-\beta_2}{x} - \left(\theta_1\beta_1\alpha_1^{\beta_1}x^{\beta_1-1} - \theta_2\beta_2\alpha_2^{\beta_2}x^{\beta_2-1}\right)$.

Thus under the given conditions in the theorem, $\frac{d}{dx}\ln\frac{f_X(x;\theta_1,\alpha_1,\beta_1)}{f_Y(x;\theta_2,\alpha_2,\beta_2)} < 0$. This means that $X \leq_{lr} Y$ and hence $X \leq_{hr} Y$, $X \leq_{mrl} Y$ and $X \leq_{st} Y$.

6. Maximum Likelihood Estimation of Parameters

Let $(x_1, x_2, x_3, \dots, x_n)$ be a random sample of size n from GWD(θ,α,β). Then natural log likelihood function is given by

$$\ln L = \sum_{i=1}^{n} \ln f(x_i;\theta,\alpha,\beta) = n(\beta\ln\alpha + \ln\beta + \ln\theta) + (\beta-1)\sum_{i=1}^{n}\ln x_i - \theta\alpha^{\beta}\sum_{i=1}^{n} x_i^{\beta}$$

The maximum likelihood estimate (MLE) $(\hat{\theta},\hat{\alpha},\hat{\beta})$ of (θ,α,β) of GWD (2.1) are the solutions of the following log likelihood equations

$$\frac{\partial\ln L}{\partial\theta} = \frac{n}{\theta} - \alpha^{\beta}\sum_{i=1}^{n} x_i^{\beta} = 0$$

$$\frac{\partial\ln L}{\partial\alpha} = \frac{n\beta}{\alpha} - \theta\beta\alpha^{\beta-1}\sum_{i=1}^{n} x_i^{\beta} = 0$$

$$\frac{\partial\ln L}{\partial\beta} = n\ln\alpha + \frac{n}{\beta} + \sum_{i=1}^{n}\ln x_i - \theta\alpha^{\beta}\ln\alpha\sum_{i=1}^{n} x_i^{\beta} - \theta\alpha^{\beta}\sum_{i=1}^{n} x_i^{\beta}\ln x_i = 0$$

These three log likelihood equations do not seem to be solved directly because these cannot be expressed in closed form. However, Fisher's scoring method can be applied to solve these equations iteratively. For, we have

$$\frac{\partial^2\ln L}{\partial\theta^2} = -\frac{n}{\theta^2}$$

$$\frac{\partial^2\ln L}{\partial\alpha^2} = -\frac{n\beta}{\alpha^2} - \theta\beta(\beta-1)\alpha^{\beta-2}\sum_{i=1}^{n} x_i^{\beta}$$

$$\frac{\partial^2\ln L}{\partial\beta^2} = -\left[\frac{n}{\beta^2} + \theta\alpha^{\beta}\sum_{i=1}^{n} x_i^{\beta}(\ln x_i)^2 + \theta\alpha^{\beta}(\ln\alpha)^2\sum_{i=1}^{n} x_i^{\beta} + 2\theta\alpha^{\beta}(\ln\alpha)\sum_{i=1}^{n} x_i^{\beta}\ln x_i\right]$$

$$\frac{\partial^2\ln L}{\partial\theta\,\partial\alpha} = -\beta\alpha^{\beta-1}\sum_{i=1}^{n} x_i^{\beta} = \frac{\partial^2\ln L}{\partial\alpha\,\partial\theta}$$

$$\frac{\partial^2\ln L}{\partial\theta\,\partial\beta} = -\left[\alpha^{\beta}\ln\alpha\sum_{i=1}^{n} x_i^{\beta} + \alpha^{\beta}\sum_{i=1}^{n} x_i^{\beta}\ln x_i\right] = \frac{\partial^2\ln L}{\partial\beta\,\partial\theta}$$

$$\frac{\partial^2\ln L}{\partial\alpha\,\partial\beta} = \frac{n}{\alpha} - \theta\alpha^{\beta-1}\sum_{i=1}^{n} x_i^{\beta} - \theta\beta\alpha^{\beta-1}\ln\alpha\sum_{i=1}^{n} x_i^{\beta} - \theta\beta\alpha^{\beta-1}\sum_{i=1}^{n} x_i^{\beta}\ln x_i = \frac{\partial^2\ln L}{\partial\beta\,\partial\alpha}$$

The MLE $(\hat{\theta},\hat{\alpha},\hat{\beta})$ of (θ,α,β) of GWD (2.1) are the solution of the following equations

$$\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta^2} & \frac{\partial^2 \ln L}{\partial \theta \partial \alpha} & \frac{\partial^2 \ln L}{\partial \theta \partial \beta} \\ \frac{\partial^2 \ln L}{\partial \alpha \partial \theta} & \frac{\partial^2 \ln L}{\partial \alpha^2} & \frac{\partial^2 \ln L}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \ln L}{\partial \beta \partial \theta} & \frac{\partial^2 \ln L}{\partial \beta \partial \alpha} & \frac{\partial^2 \ln L}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\alpha}=\alpha_0 \\ \hat{\beta}=\beta_0}} \begin{bmatrix} \hat{\theta}-\theta_0 \\ \hat{\alpha}-\alpha_0 \\ \hat{\beta}-\beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta} \\ \frac{\partial \ln L}{\partial \alpha} \\ \frac{\partial \ln L}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\alpha}=\alpha_0 \\ \hat{\beta}=\beta_0}}$$

where θ_0, α_0 and β_0 are the initial values of θ, α and β. These equations are solved iteratively till sufficiently close estimates of $\hat{\theta}, \hat{\alpha}$ *and* $\hat{\beta}$ are obtained. In this paper, R-software has been used to estimate the parameters θ, α and β for the considered datasets.

7. Goodness of Fit

In this section, the application and goodness of fit of GWD has been discussed with a real lifetime dataset from engineering and the fit has been compared with one parameter exponential and Lindley distributions, two-parameter Gamma, Weibull, Lognormal, GED and WLD, and three-parameter GGD and GLD. The following dataset has been considered for the goodness of fit of the considered distributions.

The dataset in table 1 is given by Birnbaum and Saunders (1969) on the fatigue life of 6061 – T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second. The dataset consists of 100 observations with maximum stress per cycle 31,000 psi. The data ($\times 10^{-3}$) are presented below (after subtracting 65**).**

Table 1: The data is regarding fatigue life of 6061 – T6 aluminum coupons cut parallel to the direction of rolling and oscillated at 18 cycles per second, available in Birnbaum and Saunders (1969)

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

The pdf and the cdf of the considered distributions are presented in table 2.

Table 2: **pdf and cdf of GLD, WLD, GED, Lognormal and Lindley distributions**

Distributions	pdf/cdf	
GLD	Pdf	$f\left(x;\theta,\alpha,\beta\right)=\frac{\theta^{\alpha+1}}{\left(\beta+\theta\right)}\frac{x^{\alpha-1}}{\Gamma\left(\alpha+1\right)}\left(\alpha+\beta x\right)e^{-\theta x}$
	Cdf	$F\left(x;\theta,\alpha,\beta\right)=1-\frac{\alpha\left(\beta+\theta\right)\Gamma\left(\alpha,\theta x\right)+\beta\left(\theta x\right)^{\alpha}e^{-\theta x}}{\left(\beta+\theta\right)\Gamma\left(\alpha+1\right)}$
WLD	Pdf	$f\left(x;\theta,\alpha\right)=\frac{\theta^{\alpha+1}}{\left(\theta+\alpha\right)}\frac{x^{\alpha-1}}{\Gamma\left(\alpha\right)}\left(1+x\right)e^{-\theta x}$
	Cdf	$F\left(x;\theta,\alpha\right)=1-\frac{\left(\theta+\alpha\right)\Gamma\left(\alpha,\theta x\right)+\left(\theta x\right)^{\alpha}e^{-\theta x}}{\left(\theta+\alpha\right)\Gamma\left(\alpha\right)}$
GED	pdf	$f\left(x;\theta,\alpha\right)=\theta\alpha\left(1-e^{-\theta x}\right)^{\alpha-1}e^{-\theta x}$
	cdf	$F\left(x;\theta,\alpha\right)=\left(1-e^{-\theta x}\right)^{\alpha}$
Lognormal	pdf	$f\left(x;\theta,\alpha\right)=\frac{1}{\sqrt{2\pi}\alpha x}e^{-\frac{1}{2}\left(\frac{\log x-\theta}{\alpha}\right)^{2}}$
	cdf	$F\left(x;\theta,\alpha\right)=\phi\left(\frac{\log x-\theta}{\alpha}\right)$
Lindley	pdf	$f\left(x;\theta\right)=\frac{\theta^{2}}{\theta+1}\left(1+x\right)e^{-\theta x}$
	cdf	$F\left(x;\theta\right)=1-\left[1+\frac{\theta x}{\theta+1}\right]e^{-\theta x}$

Table 3: Summary of the ML estimates of parameters

Model	ML Estimates		
	$\hat{\theta}$	$\hat{\alpha}$	$\hat{\beta}$
GWD	2.6864	0.0097	3.2122
GGD	0.01880	4.8293	1.3071
GLD	0.1152	7.0755	0.4492
WLD	0.1128	6.7274	------
GED	0.4140	9.6988	-------
Weibull	0.0021	-------	1.4573
Gamma	0.1125	17.4395	------
Lognormal	4.1579	0.41112	------
Lindley	0.0288	--------	-------
Exponential	0.0146	-------	--------

In order to compare considered distributions maximum likelihood estimates (MLEs) of parameters have been presented in table 3 and values of $-2\ln L$, AIC (Akaike information criterion), K-S Statistic (Kolmogorov-Smirnov Statistic) and p-value for the considered dataset have been computed and presented in table 4. The formulae for AIC and K-S Statistics are as follows: $AIC = -2\log L + 2k$, and $K\text{-}S = \underset{x}{\text{Sup}}\left|F_n(x) - F_0(x)\right|$, where k being the number of parameters involved in the respective distributions, n is the sample size and $F_n(x)$ is the empirical distribution function. The best distribution corresponds to the lower values of $-2\ln L$, AIC and K-S statistic and higher p- value.

Table 4: Summary of Goodness of fit by K-S Statistic

Model	$-2\log L$	AIC	K-S	p-value
GWD	908.52	914.52	0.071	0.695
GGD	912.44	918.44	0.087	0.429
GLD	914.95	920.44	0.098	0.281
WLD	915.56	919.56	0.099	0.273
GED	927.78	931.78	0.854	0.000
Weibull	982.66	986.66	0.986	0.000
Gamma	915.76	919.76	0.895	0.000
Lognormal	937.59	941.59	0.345	0.000
Lindley	983.11	985.11	0.252	0.000
Exponential	1044.87	1046.87	0.366	0.000

It is obvious from the goodness of fit in table 4 that GWD gives much closer fit than GGD,GLD,WLD, GED, Weibull, Gamma, Lognormal, Lindley and exponential and hence it can be considered an important distribution for modeling lifetime data over these distributions. The variance-covariance matrix of the parameters θ, α and β of GWD for the considered dataset has been presented in table 5.

Table 5: Variance-covariance matrix of the parameters θ, α and β of GWD for dataset

$$\begin{array}{c} \\ \hat{\theta} \\ \hat{\alpha} \\ \hat{\beta} \end{array} \begin{array}{c} \begin{array}{ccc} \hat{\theta} & \hat{\alpha} & \hat{\beta} \end{array} \\ \begin{bmatrix} 0.1690 & -0.0404 & 0.0221 \\ -0.0404 & 0.0097 & -0.0060 \\ 0.0221 & -0.0060 & 0.0597 \end{bmatrix} \end{array}$$

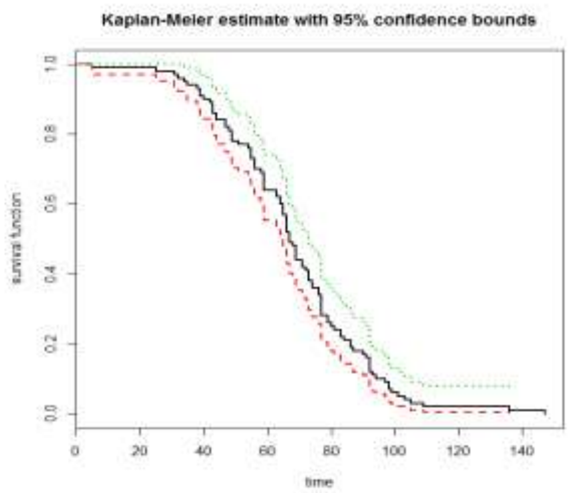


Fig4: Kaplan Meier plot on considered data set

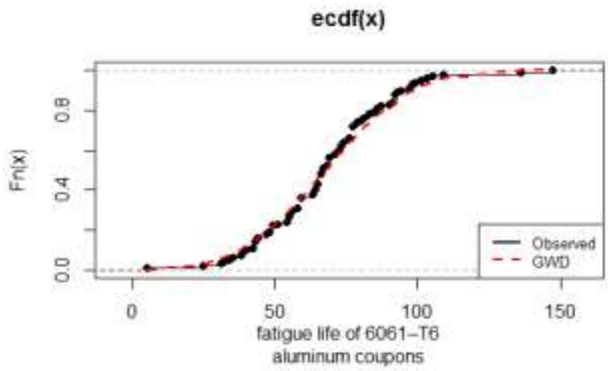


Fig.5: Fitted cdf plot on considered dataset

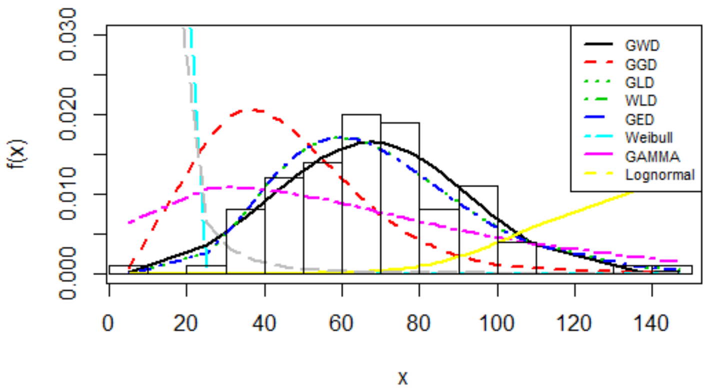


Fig.6: fitted probability plot of distributions on considered data set.

8. Concluding Remarks

A three-parameter generalization of Weibull distribution (GWD) introduced by Weibull (1951), which includes the two-parameter Weibull introduced by Weibull (1951) and exponential distribution as special cases, has been proposed and investigated. Its hazard rate function and stochastic ordering has been discussed and their behaviors have been studied graphically. Maximum likelihood estimation has been discussed for parameter estimates of GWD. The goodness of fit of the proposed distribution has been discussed with a real lifetime dataset. The proposed distribution is useful for the dataset whose hazard rate is monotonically decreasing, constant, monotonically increasing or bathtub shape. Further, it is worth mentioning that, although we have used non-censored data but it can also be used for censored data including left censoring, right censoring and interval censoring

References

1. Ahsanullah, M. &Hamedani, G.G. (2010): Exponential Distribution-Theory and Methods, Nova Science Publishers, Inc., New York
2. Almalki, S.J. and Nadarajah, S. (2014): Modifications of the Weibull distribution-A review, *Reliability Engineering and System Safety*, 124, 32 - 55

3. Birnbaum, Z.W. and Saunders, S.C. (1969): Estimation for a family of life distributions with applications to fatigue, *Journal of Applied Probability*, 6, 328-347.
4. Chakraborty, S. and Chakraborty, D. (2012): Discrete gamma distributions-Properties and Parameter estimation, *Communications in Statistics-Theory & Methods*, 41(18), 3301 – 3324.
5. Chakraborty, S. (2015): A new discrete distribution related to generalized gamma distribution and its properties, *Communications in Statistics-Theory & Methods*, 44(8), 1691 – 1705
6. Epstein, B. (1958): The exponential distribution and its role in life testing, *Industrial Quality control*, 15, 2 – 7.
7. Ghitany ,M.E., Alqallaf ,F., Al-Mutairi, D.K., Husain, H.A. (2011): A two-parameter weighted Lindley distribution and its applications to survival data, *Mathematics and Computers in simulation*, 81, 1190-1201
8. Gupta, R.D. and Kundu, D. (1999): Generalized Exponential Distribution, *Austalian & New Zealand Journal of Statistics*, 41(2), 173 – 188.
9. Lai, C.D. (2014): Generalized Weibull distributions, Springer briefs in Statistics, 23 – 75.
10. Lai, C.D., Xie, M. and Murthy, D,N.P (2003): Modified Weibull Model, *IEEE Transactions on Reliability*, 52, 33- 37
11. Lai, C.D., Murthy, D.N.P, and Xie, M. (2011): Weibull distributions, *Wiley Interdisciplinary Reviews*, 3(3), 282 - 287
12. Lee, E.T. and Wang, J.W. (2003): Statistical Methods for Survival Data Analysis, 3rd edition, Jon Wiley & Sons, Inc., New Jersey.
13. Lindley, D. V., (1958). Fiducial distributions and Bayes' theorem, *Journal of the Royal Statistical Society, Series B*, 20, PP. 102- 107
14. Marshall, A. W. and Olkin, I. (1997): A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families, *Biometrika*, 84, 641 – 652
15. Mudholkar, G.S. and Srivastava, D.K. (1993): Exponentiated Weibull family for analyzing bathtub failure rate data, *IEEE Transactions on Reliability*, 42, 299 - 302
16. Mudholkar,G.S. and Kollia, G.D. (1994): Generalized Weibull family –A structural analysis, *Communications in Statistics-Theory and Methods*, 23, 1149 – 1171
17. Mudholkar, G.S., Srivastava, D.K. and freimer, M. (1995): The exponential Weibull family-A reanalysis of bus-motor-failure data, *Technometrics*, 37, 436 - 445
18. Murthy, D.N.P., Xie, M. and Jiang, R. (2004): Weibull Models, John Wiley
19. Nakagawa, T. and Osaki, S. (1975): The discrete Weibul distribution, *IEEE tans. Reliability*, R-24(5), 300 – 301.
20. Nekoukhou, V.M., Alamatsaw, M.H., and Bidram, H. (2012): A discrete analogue of the generalized exponential distribution, *Communications in Statistics-Theory & Methods*, 41, 2000 – 2013.
21. Sato, H., Ikota, M., Aritoshi, S., and Masuda, H. (1999): A new defect distribution in meteorology with consistent discrete exponential formula and its applications, *IEEE Trans. Semicond. Manufactur*, 12 (4), 409 – 418.
22. Shaked, M. and Shanthikumar, J.G.(1994): Stochastic Orders and Their Applications, Academic Press, New York
23. Shanker, R., Shukla, K.K., Shanker, R. and Tekie, A.L. (2016): On modeling of Lifetime data using two-parameter Gamma and Weibull distributions, *Biometrics & Biostatistics International journal*, 4(5), 1 – 6.
24. Shanker, R. and Shukla, K.K. (2016): On modeling of lifetime data using three-parameter Generalized Lindley and Generalized Gamma distributions, *Biometrics &Biostatistics International Journal*, 4(7), 1 – 7.
25. Stacy, E.W. (1962): A generalization of the gamma distribution, *Annals of Mathematical Statistics*, 33, 1187 – 1192.
26. Stacy, E.W. and Mihram, G.A. (1965): Parametric estimation for a generalized gamma

distribution, *Technometrics*, 7, 349 – 358.
27. Stein, W.E. and Dattero, R. (1984): A new discrete Weibull distribution, *IEEE tans. Reliability*, R-33(2), 196 – 197.
28. Weibull, W. (1951): A statistical distribution of wide applicability, *Journal of Applied Mathematics*, 18, 293 – 297.
29. Zakerzadeh, H. and Dolati, A. (2009): Generalized Lindley distribution, *Journal of Mathematical extension*, 3 (2), 13 – 25.

Modelling of an offline and online software for normalization of microarray data of gene expression by Perl, Bioperl and PerlTk and Perl-CGI

Gaurav Kumar Srivastava[1], Dr. Santosh Kumar[2], Dr. Himanshu Pandey[3]

•

[1]Research Scholar, Maharishi University of Information technology, Lucknow
[2]Assosiate Professor, Maharishi University of Information technology, Lucknow
[3]Assistant Professor, BBDNIIT, Lucknow

Abstract

B-Chip Reverence is an online database which isfreely accessible for microarray redundancy removal & normalizationand various data analysis techniques are applied on the data. This software accurately handle the massive amount of data.The growing use of DNA microarrays in biomedical research has led to the proliferation of analysis tools. These software programs address different aspects of analysis (e.g. normalization and clustering within and across individual arrays) as well as extended analysis methods (e.g. clustering, annotation and mining of multiple datasets).After studying all the terms and problems related to Microarray technique, we tried to make an-open and user friendly software to deal with all the problems and to run all the steps of this technique, so that we used Perl & Perl-cgi. perl-cgi stands for *Common Gateway Interface*, is a standard programming interface between Web servers and external programs. perl-cgi executes external programs on the webserver.

I. Introduction

A microarray database is a repository containing microarray gene expression data. The key uses of a microarray database are to store the measurement data, manage a searchable index, and make the data available to other applications for analysis and interpretation. The concept behind this, a microarray is a pattern of ssDNA probes which are immobilized on surface of chip or a slide. The probe sequences are designed and placed on an array in regular patter of spots. The chip or slide is usually made of glass or nylon and is manufactured using technologies developed for silicon computer chips. Each microarray chip is arranged as a checkerboard of 105 or 106 spots or features, each spot containing millions of copies of a unique DNA probe (often 25 nt long).Microarray technology allows the monitoring of expression levels for thousands of genes simultaneously. Even in replicated experiment, some variations are commonly observed. Normalization is the process of removing some sources of variation which affect the measured gene expression levels. In gene expression microarray data analysis, selecting a small number of discriminative genes from thousands of genes is an important problem for accurate classification of diseases or phenotypes.The ability of microarray chip to capture the expressional level of thousands of genes in one snapshot becomes a major attraction for biologists. By performing parallel microarray experiments under different conditions, biologists seek useful information of the underlying biological process that lies in the hundreds of thousands of data points obtained. The first step of task is classifies data through a single step partition. In such task,cluster the genes into biological meaningful groups according to their pattern of expression, based on the assumption that expressional similarity of genes implies their functional similarity. The clustering methods which are used for this include conventional clustering methods (such as k-means clustering, and self-organizing maps).k-means clustering is a simple and a divisive approach. In this method, data are partitioned into k-clusters, which are prespecified at the outset. Self-Organizing Maps is pattern recognition algorithm employs neural networks and based on the machine-learning method.

B chip reverence database dealing the removing microarray redenduncy and normalization, for that using bioperl (kmean and SOM executed by cpan module) dreamweaver8 (for designing web pages on website) photoshop (for logo designing) PERL CGI (for programs to interface with information servers such as HTTP (web servers.)

II. Aim & Object

- To remove duplicates, repetitive and blank genes from our raw data. After removing redundancy, normalize the datasets using the PERL-CGI.
- To make user friendly, open and easily accessible interactive CGI interface for database and various tools for analysis of clustering (k-mean and SOM) for microarray using PERL-CGI algorithms.

III. Materials & Methods

- **Perl & Perl-cgi**

Perl is a programming language developed by Larry Wall, designed for text processing. ThoughPerl is not officially an acronym but many times it is used as it stands for *Practical Extraction and Report Language*. It runs on platforms like Windows, Mac OS, and UNIX.

Perl CGI is the Common Gateway Interface, a standard for programs to interface with information servers such as HTTP (web) servers. CGI allows the HTTP server to run an executable program or script in response to a user request, and generate output on the fly. This allows web developers to create dynamic and interactive web pages. Perl is a very common language for CGI programming as it is largely platform independent and the language's features make it very easy to write powerful applications.

- **Bio Perl**

BioPerl is a collection of Perl modules that facilitate the development of Perl scripts for bioinformatics applications. It has played an integral role in the Human Genome Project.BioPerl is an active open source software project supported by the Open Bioinformatics Foundation.

- **Self-Organizing Maps (SOM) and K-Means Clustering (KMC)**

As a machine-learning method, a SOM belongs to the category of neural networks. It provides a technique to visualize the HD input data on an output map of neurons. The map is often presented in a 2D grid of neurons.KMC is a simple and widely used partitioning method for data analysis. It's helpfulness in discovering group of co-expressed genes has been demonstrated.

- **Dreamweaver8 and WampServer**

Dreamweaver8 allows to create professional web pages and also quickly add objects and functionality to pages without having to program the HTMLcode manually.WampServer is a windows web development environment for Apachey, MySQL, PHP databases. It's also virtual server for windows platform, allows it user to manage Website and its components.

IV. Techniques/Databases Used

- Westudied information about genes through the method of Gene Ontology by Gene Cards, next we did Microarray data retrieval from NCBI Geo profiles of SMAD7. Data Normalization & Redundancy Removal of Gene Expression do with the help of Microsoft office excel.
- Next we studied Perl elementary and their different algorithms and logics in Perl includes:-
  - **a) Regular Expression**
    A regular expression is a string of characters that defines the pattern or patterns you are viewing. The syntax of regular expressions in Perl is very similar to what you will find within other regular expression.There are three regular expression operators within Perl.
    - Substitute Regular Expression - s///
    - Transliterate Regular Expression - tr///
    - Match Regular Expression - m//
  - **b) File Handling**
    A filehandle is a named internal Perl structure that associates a physical file with a name. All filehandles are capable of read/write access, so you can read from and update any file or device associated with a filehandle. However, when you associate a filehandle, you can specify the mode in which the filehandle is opened. Three basic file handles are - **STDIN**, **STDOUT**, and **STDERR,** which represent standard input, standard output and standard error devices respectively.
  - **c) Sub-Routines**
    A Perl subroutine or function is a group of statements that together performs a task. You can divide up your code into separate subroutines. How you divide up your code among different subroutines is up to you, but logically the division usually is so each function performs a specific task.
  - **d) Parsing**
    parsing is the process of analyzing an input sequence (read from a file or a keyboard, for example) in order to determine its grammatical structure with respect to a given formal grammar. It is formally named syntax analysis. A parser is a computer program that carries out this task.
- And Normalization & Redundancy removal (to make data sorted and fine/removal of repeated genes and blank genes).
- Next we did the major/main part of the Microarray technique i.e. clustering: the clustering problem of microarray data only as an analysis to find genes that behave similarly over the experimental conditions. The first generation of clustering techniques includes hierarchical clustering, K-Means clustering and Self-Organizing Maps, we used K-means clustering (KMC): helpful in discovering group of co-expressed genes,& Self-organization map (SOM): provides a technique to visualize the HD input data on an output map of neurons.
- Beside all the steps which we did, we also studied CGI, HTML, Perl toolkit packages etc. in which we used Perl-cgi to develop the online web server B-Chip reverence software for users to perform Microarray Technique computationally.
- We also used BioPerl, we correlated K-Means Clustering and Self-Organizing Maps (SOM) algorithms to perform clustering with Bio Perl. Bio Perl is a collection of Perl modules and it facilitates the development of Perl scripts for bioinformaticsapplications.

V. Result

Fig 1: Display page of Pre-Clustering

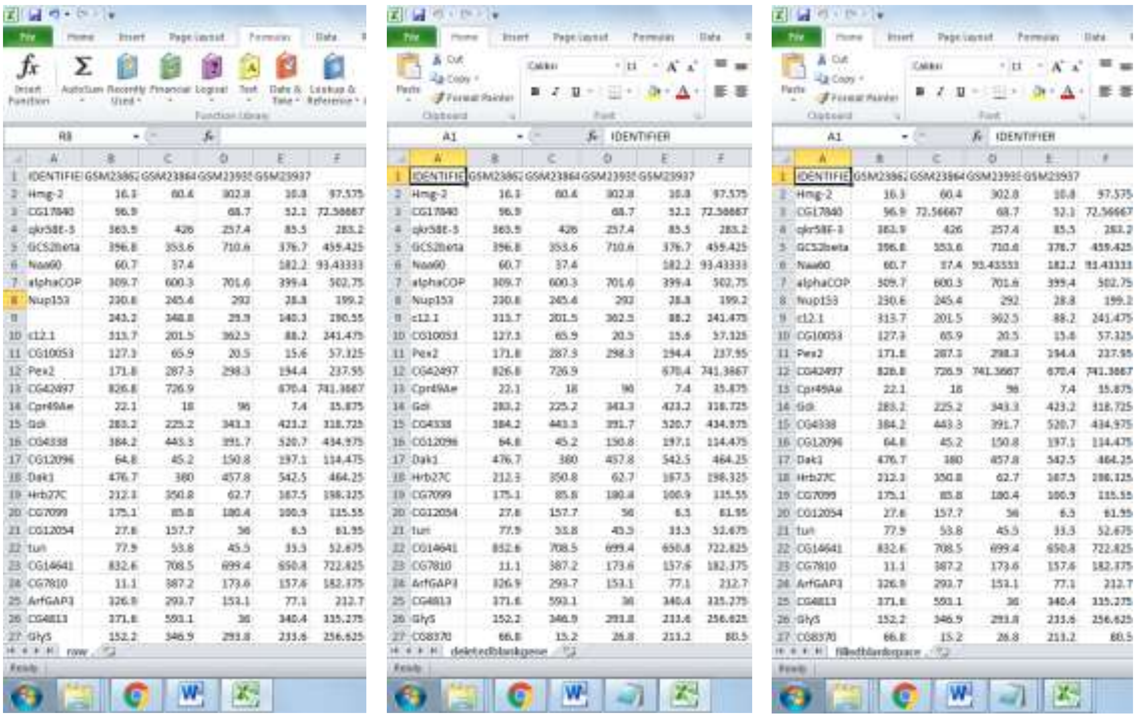


Fig 2: Pre-Clustering Results and analysis example

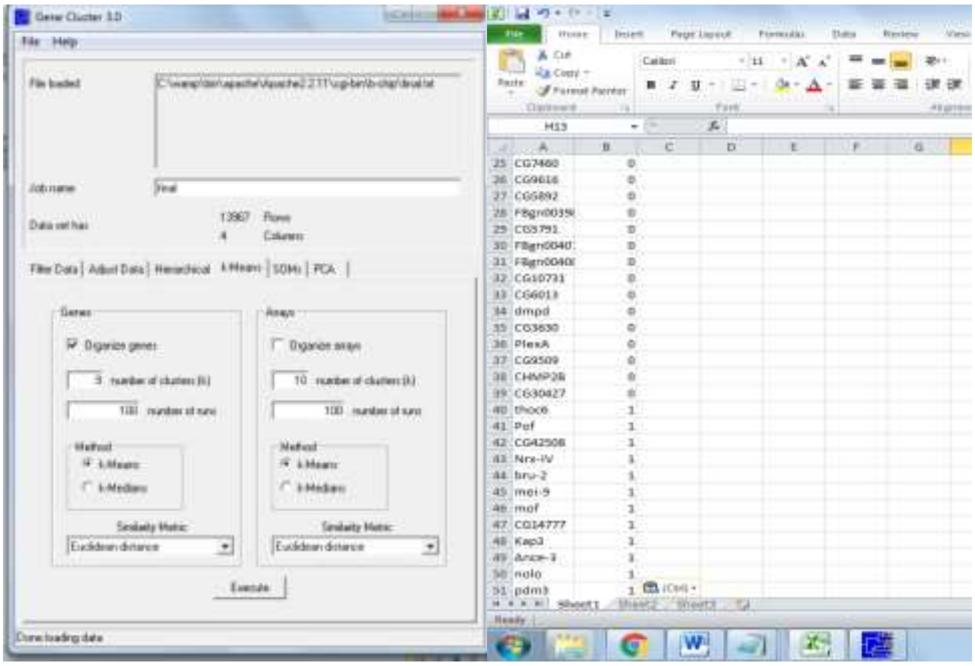


Fig 3: KMeans-Clustering Results and analysis example

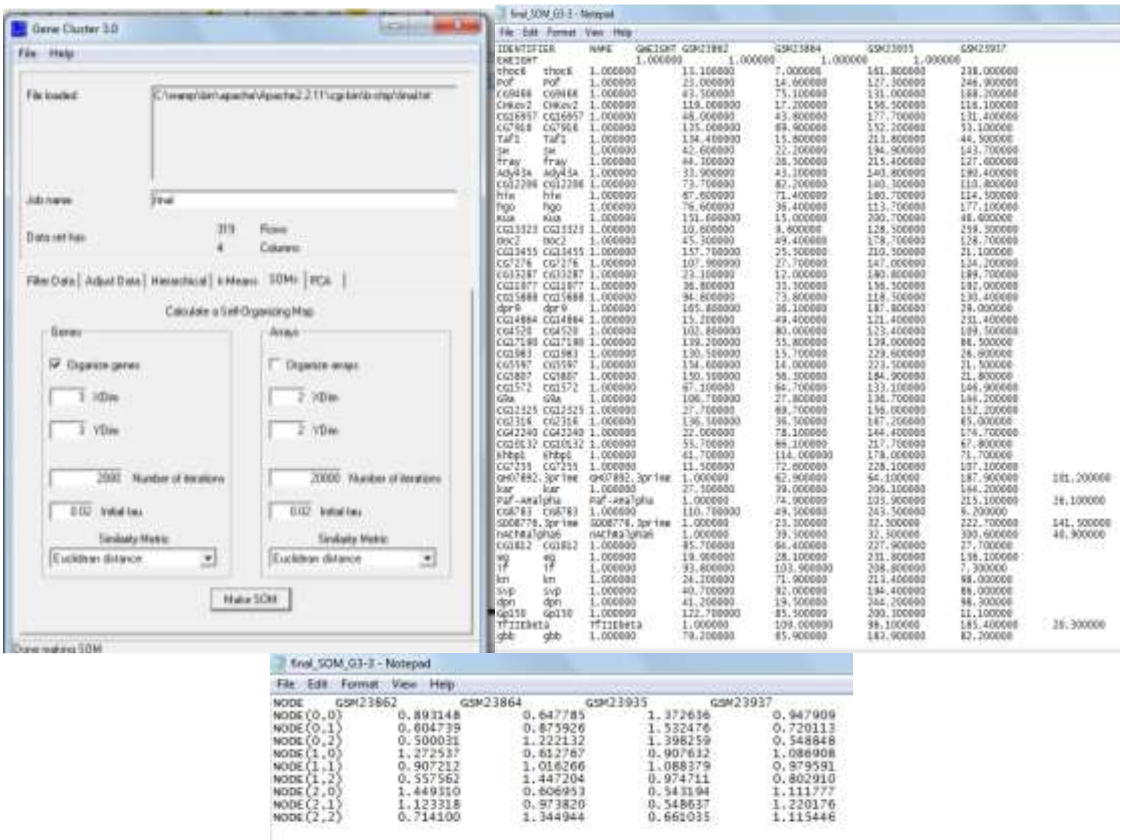


Fig 4: SOM-Clustering Results and analysis example

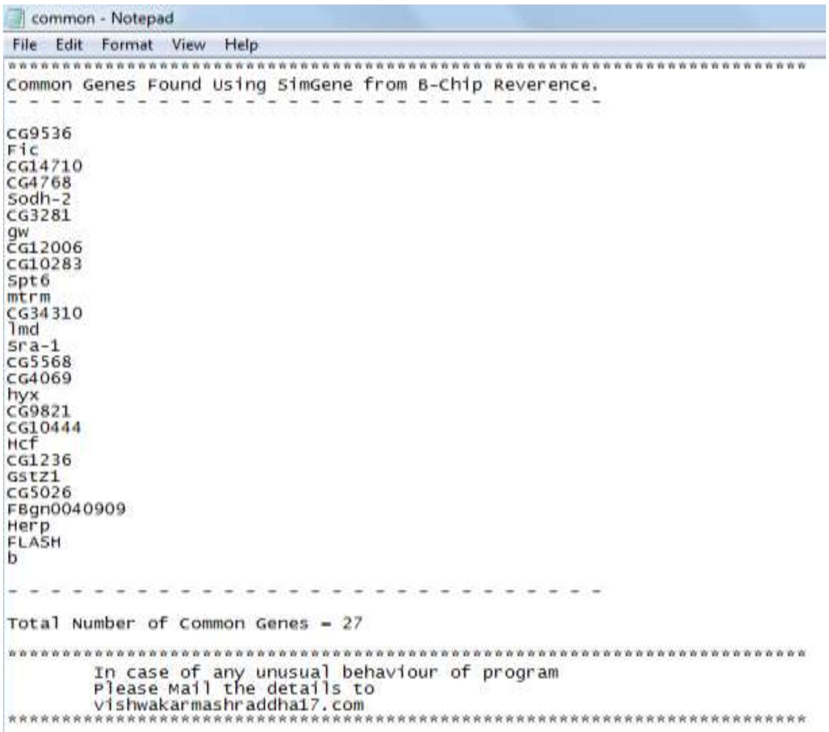


Fig 5: Common Gene of SOM and KMC Clustering

VI. Conclusion

B-Chip Reverence is a database, specially design for in silico analysis of microarray. In the field of Bioinformatics, B-chip Reverence fulfills the basic needs during the online microarray analysis.hence is one of a kind of its innovative database which provides web server for microarray redundancy removal & normalizationand various data analysis techniques are applied on the data.

VII. Summary

B-Chip Reverence is a online database which is freely accessible for microarray redundancy removal & normalization and various data analysis techniques are applied on the data. This software accurately handle the massive amount of data. The growing use of DNA microarrays in biomedical research has led to the proliferation of analysis tools. These software programs address different aspects of analysis (e.g. normalization and clustering within and across individual arrays) as well as extended analysis methods (e.g. clustering, annotation and mining of multiple datasets). After studying all the terms and problems related to Microarray technique, we tried to make an-open and user friendly software to deal with all the problems and to run all the steps of this technique, so that we used Perl & Perl-cgi. perl-cgi stands for *Common Gateway Interface*, is a standard programming interface between Web servers and external programs. perl-cgi executes external programs on the web server. We also used BioPerl, we correlated K-Means Clustering and Self-Organizing Maps (SOM) algorithms to perform clustering with Bio Perl. Bio Perl is a collection of Perl modules and it facilitates the development of Perl scripts for bioinformatics applications. And Dreamweaver8 used to create professional web pages and also quickly add objects and functionality to pages without having to program the HTML code manually.B-Chip Reverence is a database, specially design for in silico analysis of microarray. In the field of Bioinformatics, B-chip Reverence fulfills the basic needs during the online microarray analysis. hence is one of a kind of its innovative database which provides web server for microarray redundancy removal & normalization and various data analysis techniques are applied on the data.

References

1. Vroh Bi I, McMullen MD, Sanchez-Villeda H, Schroeder S, Gardiner J, Polacco M, Soderlund C, Wing R, Fang Z, Coe EH., Jr Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map. Crop Sci. 2006;46:12–21. doi: 10.2135/cropsci2004.0706. [Cross Ref]
2. Wright SI, Vroh Bi I, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. The effects of artificial selection on the maize genome. Science. 2005;308:1310–1314. doi: 10.1126/science.1107891. [PubMed] [Cross Ref]
3. Schwartz AS, Pachter L. Multiple alignment by sequence annealing. Bioinformatics. 2006;23:e24–e29. doi: 10.1093/bioinformatics/btl311. [PubMed] [Cross Ref]
4. Ewing B, Green P. Basecalling of automated sequencer traces using *phred*. II. Error probabilities. Genome Res. 1998;8:186–194. [PubMed]
5. Ewing B, Hillier L, Wendl M, Green P. Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998;8:175–185. [PubMed]
6. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res. 1994;22:4673–4680. doi: 10.1093/nar/22.22.4673. [PMC free article][PubMed] [Cross Ref]
7. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res. 2004;32:1792–1797. doi: 10.1093/nar/gkh340. [PMC free article] [PubMed] [Cross Ref]
8. Page RDM. TREEVIEW: An application to display phylogenetic trees on personal computers. Comp Appl Biosci. 1996;12:357–358. [PubMed]
9. Rozas J, Sánchez-DelBarrio SJ, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 2003;19:2496–2497. doi: 10.1093/bioinformatics/btg359. [PubMed] [Cross Ref]
10. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999;41:95–98.
11. Clamp M, Cuff J, Searle SM, Barton JG. The Jalview Java alignment editor. Bioinformatics. 2004;20:426–427. doi: 10.1093/bioinformatics/btg430. [PubMed] [Cross Ref]
12. Pible O, Imbert G, Pellequer J-L. INTERALIGN: interactive alignment editor for distantly related protein sequences. Bioinformatics. 2005;21:3166–3167. doi: 10.1093/bioinformatics/bti474. [PubMed] [Cross Ref]
13. Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell. 2005;17:2859–2872. doi: 10.1105/tpc.105.037242. [PMC free article] [PubMed] [Cross Ref]
14. Canaran P, Stein L, Ware D. Look-Align: An interactive web-based multiple sequence alignment viewer with polymorphism analysis support. Bioinformatics. 2006;22:885–886. doi: 10.1093/bioinformatics/btl028. [PubMed] [Cross Ref]
15. Skolnick J, Zhang Y, Arakaki AK, Kolinski A, Boniecki M, Szilagyi A, Kihara D, (2003) TOUCHSTONE: A unified approach to protein structure prediction, *Proteins* (53 Suppl 6):469–479.
16. Stark A., Sunyaev S., Russell R.B. (2003) A model for statistical significance of local similarities in structure. J. Mol. Biol. 2003;326:1307–1316.
17. Sonnhammer, E., Eddy, S., and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins 28, 405–420.
18. Tatusov, R., Galperin, M., Natale, D., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28, 33–36.

19. Pandey. H, Darbari. M and Singh. V.K, "Coalescence of Evolutionary Multi-Objective Decision making approach and Genetic Programming for Selection of Software Quality Parameter", International Journal of Applied Information System (IJAIS), Foundation of Computer Science, New York, USA, Volume 7, No. 11, PP. ISSN: 2249-0868, Nov. 2014.
20. Bansal. S and Pandey. H, "Develop Framework for selecting best Software Development Methodology", International Journal of Scientific and Engineering Research, Volume 5, Issue 4, PP. 1067-1070, ISSN: 2229-5518, Apr. 2014.
21. Srivastava. M and Pandey. H, "A Literature Review of E- Learning Model Based on Semantic Web Technology", International Journal of Scientific and Engineering Research" Volume 5, Issue 10, PP. 174-178, ISSN: 2229-5518, Oct. 2014.
22. Pandey. H and Singh. V.K, "A New NFA Reduction Algorithm for State Minimization Problem", International Journal of Applied Information Systems (IJAIS), Foundation of Computer Science FCS, New York, USA, Volume 8, No.3, PP. 27-30, ISSN: 2249-0868, Feb. 2015.
23. Pandey. H and Singh. V.K, "LR Rotation rule for creating Minimal NFA", International Journal of Applied Information Systems (IJAIS), Foundation of Computer Science FCS, New York, USA, Volume 8, No.6, PP. 1-4, ISSN: 2249-0868, Apr. 2015.
24. Pandey. H and Darbari. M, "Estimation of Software Quality Parameters Using Combination of Quality Function Deployment and Messy Genetic Algorithm", Grenze Scientific Society, Associate publisher: **McGraw-Hill Education**, ISBN: 978-93-392-2169-0. Feb. 2015.
25. Darbari. M, Srivastava. G and Pandey. H, "New Assumption of Cognitive Model for Information Foraging on Web", International Journal of Advances in Engineering & Technology, Volume 8, Issue 2, PP. 163-169, ISSN: 22311963, Apr. 2015.
26. Pandey. H and Singh. V.K, "A Fuzzy Logic based Recommender System for E-Learning System with Multi-Agent Framework", International Journal of Computer Applications, Foundation of Computer Science FCS, New York, USA, Volume 122, No.17, PP. 18-21, ISSN: 0975 – 8887, July. 2015.
27. Rai. V and Pandey. H, "Estimation of Maintainability in Object Oriented Design Phase: State of the art", International Journal of Scientific and Engineering Research" Volume 6, Issue 9, PP. 25-35, ISSN: 2229-5518, Sept. 2015.
28. Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y et. al.: E-Cell: software environment for whole-cell simulation. Bioinformatics 1999, 15:72-84.
29. Von Dassow G, Meir E, Munro EM, Odell GM: The segment polarity network is a robust developmental module. Nature 2000, 406:188-92.
30. Wilson CA, Kreychman J, Gerstein M (2000). Assessing annotation transfer for genomics: Quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores, *J MolBiol*297(1):233–249.
31. Wu, C., Huang, H., Yeh, L., and Barker, W. (2003).Protein family classification and functional annotation.CompBiolChem 27, 37–47.

RT&A, No 2 (53)
Volume 14, June 2019

Moving Video Camera Vigilance Using DBSCAN

Jagdatta Singh[1], Gaurav Kumar Srivastava[2], Himanshu Pandey[3]

•

[1]Research Scholar, BBDU, Lucknow
jagdattasingh@gmail.com
[2]Assistant Professor, BBDU, Lucknow
gaurav18hit@gmail.com
[3]Assistant Professor, BBDNIIT, Lucknow
hpandey010@gmail.com

Abstract

The author is trying to develop a model for dynamic or moving video camera vigilance using Density Based Clustering and location sensors. The authors try to exploit the rich functionality exposed by the machine learning paradigm in which the stochastic environment to learn is depicted as a two dimensional graph where the position of an object can be given by its coordinates. The author uses DBSCAN algorithm along with sensor enabled test ground area that keeps the X and Y co-ordinates of the moving objects. The idea here is to capture continuous video of the densest cluster of objects moving together. One practical usage of such system is a wild landscape where groups of animals are moving together to some destination. There will be a somewhat unorganized haphazard movement but we intend to capture only those animals that are greater in number as a group and the camera should move picturing them. This can be achieved by the DBSCAN algorithm

1. Introduction

1.1 Clustering Techniques:

In Clustering we split the data into groups of similar objects. Each group is known as a cluster. The intra-cluster similarity is high while inter-cluster similarity index is low. It is a very important technique in data mining. Traditionally it is seen as part of unsupervised learning. Different types of clusters as reported in the literature [1], [2].

Well Separated Clusters: Every node in this type of cluster is much similar to every other node in the cluster, but different from any other node not in the cluster.

Centre-Based clusters: Every object in the cluster is more similar to the centre also called the centroid than to the centre of any other cluster. Contiguous clusters: A node in a cluster is nearest (or more alike) to one or more other nodes in the cluster as compared to any node that is not in the cluster.

Density based clusters: A cluster is a thick region of points, which is separated by according to the low-density regions, from other regions that is of high density Conceptual clusters: A conceptual cluster shares some common feature, or indicates a particular thought.

1.2. Use of Clustering and Methods

Clustering has wide applications in Image Processing, Document Classification, Pattern Recognition, Spatial Data Analysis, Economic Science and Cluster Web log data to discover similar web access patterns, etc. Various Methods of clustering have been reported in literature [3], [4], [5]:

Partitioning method: In literature different Partitioning methods reported are: K-mean method [3], [4], K-Medoids method (PAM) [5], [6], Farthest First Traversal k-center (FFT) [7], [8], CLARA [9], CLARANS [10], Fuzzy K-Means [11], Fuzzy K-Modes [12], K-Modes [13], Squeezer [14], K-prototypes [15] and COOLCAT [16], etc.

Hierarchical Methods: Agglomerative Nesting (AGNES) [17], Divisive Analysis (DIANA) [18], Clustering using Representatives (CURE), Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) are some of the hierarchical methods.

Grid Based: Some of the Grid based clustering methods are STING, Wave Cluster, CLIQUE [19] and MAFIA [20].

Density Based Methods: Density based clustering methods include DBSCAN, GDBSCANS, OPTICS, DBCLASD and DENCLUE.

Model Based method: Model based methods are divided into two approaches: Statistical approach includes AutoClass method while Neural Network Approach includes Competitive learning and Self-organizing feature maps.

2. DBSCAN

(**D**ensity-**B**ased **S**patial **C**lustering of **A**pplications with **N**oise) is a popular **unsupervised** learning method utilized in model building and machine learning algorithms. Before we go any further, we need to define what an "unsupervised" learning method is. **Unsupervised** learning methods are when there is no clear objective or outcome we are seeking to find. Instead, we are clustering the data together based on the similarity of observations.

DBSCAN is a clustering method that is used in machine learning to separate clusters of high density from clusters of low density. Given that **DBSCAN** is a **density based clustering algorithm**, it does a great job of seeking areas in the data that have a high density of observations, versus areas of the data that are not very dense with observations. DBSCAN can sort data into clusters of varying shapes as well, another strong advantage. DBSCAN works as such:

- Divides the dataset into *n* dimensions
- For each point in the dataset, DBSCAN forms an *n* dimensional shape around that data point, and then counts how many data points fall within that shape.
- DBSCAN counts this shape as a *cluster.* DBSCAN iteratively expands the cluster, by going through each individual point within the cluster, and counting the number of other data points nearby. Take the graphic shown in figure for an example:

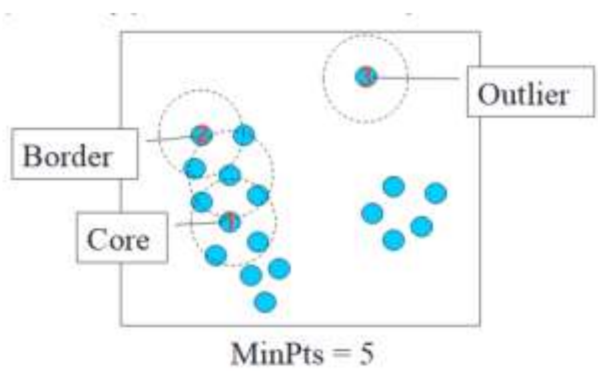


Fig. 1

3. Location Detection and Tracking of Moving Targets

Many applications require information about an object's location for rescue, emergency and security purposes. The approaches that access an object's location are typically divided into two groups: active and passive localization. In the former approach, the object is associated with a mobile station (MS), such as a tag or device in a communication network. The object's location is determined by sharing data between the MS and the base stations (BSs). The Global Positioning System (GPS), cellular networks, Bluetooth and wireless sensor networks (WSNs) are used in active localization. In the latter approach, the object does not communicate with other devices. However, the object's location can be determined by using the reflected signal from the object. Radio detection and ranging (radar), sound navigation and ranging (sonar) and laser detection and ranging (LADAR) are the most common types of passive localization. These methods have both advantages and disadvantages.

However, GPS and long-range radar generate many errors during indoor localization and tracking. Cellular networks and WSNs are limited by their complicated controls and protocols. Sonar and LADAR signals are degraded by environmental interference. Therefore, ultra-wide band (UWB) radar has become an emerging technology that is appropriate for indoor localization and tracking. UWB radar has many advantages, such as a high spatial resolution, the ability to mitigate interference, through-the-wall visibility, a simple transceiver and a low cost.

4. Methodology

The 2 dimensional area, A, assumed square in shape is plotted having X and Y coordinates. A random number of moving objects, here assumed to be small robotic cars with constant movement are left in the aforementioned area. Since the area, A, assumed here is small, location tracing sensors are fitted on the boundary of A. A video camera, C, is also planted which is used to position on the selected target. The ideal position of the camera should be on top, middle of A [21].

The DBSCAN algorithm then determines the cluster of robotic cars with maximum density. The algorithm also returns the center of the cluster which is one of the robotic cars. All the cars have built in emitters that generate a specific signal when they are selected as the center of the densest cluster shown in figure 1, called the core. Once the car is selected as the center of the densest cluster, it emits a signal that is received by the location tracing sensors [22]. As soon as the sensors receive the signal, they generate the X and Y coordinates of the car that emitted the signal. The coordinates are fed to the camera and the movement of the cluster gets recorded. This process is continuous and if the cluster changes then the process is repeated for the new cluster, center of focus being the new selected center of the densest cluster. The moving camera continuously positions its lens on the moving densest cluster and if the density of the cluster reduces then the new densest cluster is located by the DBSCAN algorithm and the camera starts focusing on the new most densely populated cluster.

The system demonstrated above can also find its application in larger areas. As pointed out earlier that the same process with slight modification can be applied to traffic monitoring and even wild life for framing videos on moving animals in groups, etc. With the aforementioned process a traffic accumulation can be reported or even a traffic jam for the traffic controllers.

In order to find the coordinates in bigger areas, we need the geo-locations in the form of the X and Y coordinates of the moving objects and the video camera will be fed with the coordinates as broadcasted by the satellites instead of location tracing sensors [23], [24].

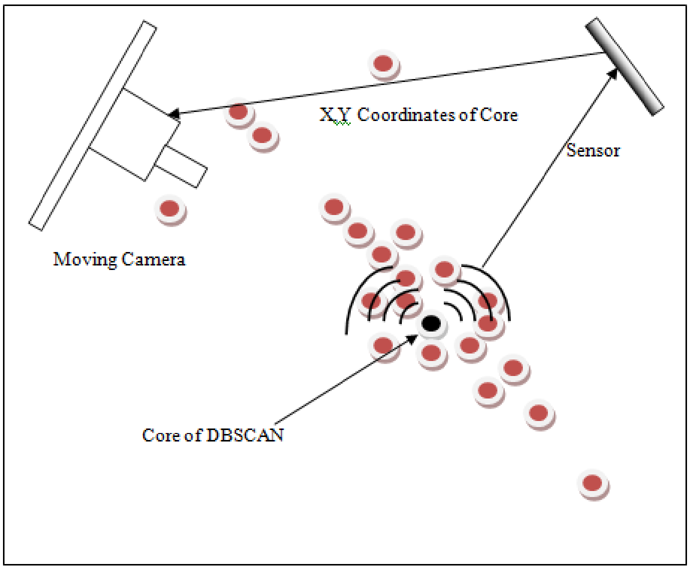


Fig. 2

5. Conclusion

In this paper the author has proposed a model to provide continuous moving camera recording for the most densely populated group of objects. Here, the author has used an unsupervised learning algorithm of the artificial intelligence, called DBSCAN to find out the most intensively crowded orientation of the objects under vigilance. The DBSCAN algorithm reports the densest point, called the core of a population. The crowd is depicted by robotic cars having a facility to emit radio signals. Once a robotic car is selected as the core, it emits radio signals. This signal is received by the sensor installed for this purpose. This sensor calculates the X and Y coordinates of the core robotic car and sends them to the positioning system of the camera. With coordinates at hand, the camera focuses its lens on the selected X and Y coordinates. In this manner, the automatic moving camera is able to keep track of the core. With time, the core is changed and so is the camera's focus. It focuses on the new car selected as the core. This installation facilitates a system where the camera always focuses on the densest part of the moving objects. As a future research, this concept can be applied in controlling the traffic, where the radio signals can be replaced by the geo-location finders.

References

[1] K.Kameshwanran, & Malarvizhi, K. Survey on Clustering Techniques in Data Mining. International Journal of Computer Science and Information Technologies , 5 .2 (2014), 2272-2276.

[2] Junaid, S., & Bhosle, K. Overview of Clustering Techniques. International Journal of Advanced Research in Computer Science and Software Engineering , 4.11 (2014), 621-624.

[3] Wagstaff, Kiri, et al. "Constrained k-means clustering with background knowledge." Icml. Vol. 1. 2001.

[4] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." Journal of the Royal Statistical Society. Series C (Applied Statistics) 28.1 (1979): 100-108.

[5] Park, Hae-Sang, and Chi-Hyuck Jun. "A simple and fast algorithm for K-medoids clustering." Expert systems with applications 36.2 (2009): 3336-3341.

[6] Velmurugan, T., and T. Santhanam. "Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points." Journal of computer science 6.3 (2010): 363.

[7] Panda, Mrutyunjaya, and Manas Ranjan Patra. "A Hybrid clustering approach for network intrusion detection using cobweb and FFT." Journal of Intelligent systems 18.3 (2009): 229-246.

[8] Matsuoka, Hidehiro, and Hiroki Shoki. "Comparison of pre-FFT and post-FFT processing adaptive arrays for OFDM systems in the presence of co-channel interference." 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003.. Vol. 2. IEEE, 2003.

[9] Ng, Raymond T., and Jiawei Han. "E cient and E ective Clustering Methods for Spatial Data Mining." Proceedings of VLDB. 1994.

[10] Ng, Raymond T., and Jiawei Han. "CLARANS: A method for clustering objects for spatial data mining." IEEE Transactions on Knowledge & Data Engineering 5 (2002): 1003-1016.

[11] Gasch, Audrey P., and Michael B. Eisen. "Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering." Genome biology 3.11 (2002): research0059-1.

[12] Huang, Zhexue, and Michael K. Ng. "A fuzzy k-modes algorithm for clustering categorical data." IEEE Transactions on Fuzzy Systems 7.4 (1999): 446-452.

[13] Chaturvedi, Anil, Paul E. Green, and J. Douglas Caroll. "K-modes clustering." Journal of classification 18.1 (2001): 35-55.

[14] He, Zengyou, Xiaofei Xu, and Shengchun Deng. "Squeezer: an efficient algorithm for clustering categorical data." Journal of Computer Science and Technology 17.5 (2002): 611-624.

[15] Huang, Zhexue. "Extensions to the k-means algorithm for clustering large data sets with categorical values." Data mining and knowledge discovery 2.3 (1998): 283-304.

[16] Barbará, Daniel, Yi Li, and Julia Couto. "COOLCAT: an entropy-based algorithm for categorical clustering." Proceedings of the eleventh international conference on Information and knowledge management. ACM, 2002.

[17] Struyf, Anja, Mia Hubert, and Peter Rousseeuw. "Clustering in an object-oriented environment." Journal of Statistical Software 1.4 (1997): 1-30.

[18] Datta, Susmita, and Somnath Datta. "Comparisons and validation of statistical clustering techniques for microarray gene expression data." Bioinformatics 19.4 (2003): 459-466.

[19] De Amorim, Saul G., Jean-Pierre Barthélemy, and Celso C. Ribeiro. "Clustering and clique partitioning: simulated annealing and tabu search approaches." Journal of Classification 9.1 (1992): 17-41.

[20] Goil, Sanjay, Harsha Nagesh, and Alok Choudhary. "MAFIA: Efficient and scalable subspace clustering for very large data sets." Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Vol. 443. ACM, 1999.

[21] Ahiska, Yavuz. "Multiple-view processing in wide-angle video camera." U.S. Patent No. 7,450,165. 11 Nov. 2008.

[22] Diener, Neil R., David S. Kloper, and Anthony T. Collins. "Server and multiple sensor system for monitoring activity in a shared radio frequency band." U.S. Patent No. 7,184,777. 27 Feb. 2007.

[23] Belisle, Timothy, and Thomas Belisle. "Geo-location system, method and apparatus." U.S. Patent Application No. 11/381,097.

[24] Williams, Darin Scot. "Car-finder method and aparatus." U.S. Patent Application No. 12/157,889.

[25]. Pandey, H., Kumar, S., & Darbari, M. (2018). Crowd Sourcing Rules in Agile Software Engineering to Improve Efficiency using Ontological Framework. Reliability: Theory & Applications, 13(2 (49)).

[26]. Pandey, H. (2018). Comparison of usage of crowdsourcing in traditional and agile software development methodologies on the basis of effectiveness. Reliability: Theory & Applications, 13(3 (50)).

[27] Singh. J and **Pandey. H** "Learning and Consensus in Multi-Agent Systems", International Journal of Research in Electronics and Computer Engineering (UGC approved Journal), Volume-6, Issue-4, PP. 1437-1442, ISSN: 2348-2281 (Online), Oct.-Dec. 2018.

[28] Pandey, H., & Singh, V. K. (2015). A fuzzy logic based recommender system for E-learning system with multi-agent framework. International Journal of Computer Applications, 122(17), 0975-8887.

[29] Pandey, H., Singh, V. K., & Verma, N. K. LR Rotation rule for creating Minimal NFA. International Journal of Applied Information Systems (IJAIS-ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA.

[30] Pandey, H., Singh, V. K., & Pandey, A. (2015). A new NFA reduction algorithm for state minimization problem. Int'l Journal of Applied Information System (IJAIS), 8(3), 27-30.

[31] Pandey. H, Darbari, M., Singh, V. K., & Kumar, G. (2014). Coalescence of Evolutionary Multi-Objective Decision Making approach and Genetic Programming for Selection of Software Quality Parameter. Complexity, 7(11).

Analysis of MAP/PH/1 Retrial Queue With Constant Retrial Rate, Bernoulli Schedule Vacation, Bernoulli Feedback, Breakdown and Repair

G. Ayyappan, R. Gowthami

•

Department of Mathematics, Pondicherry Engineering College
ayyappanpec@hotmail.com, gowthami9424@gmail.com

Abstract

A retrial queueing model in which the inter arrival times follow Markovian Arrival Process(MAP) and the service times follow phase type distribution is studied. At the end of receiving the service, the customer has two options namely, either he may go to orbit with probability q_1 to get the service again, if he is not satisfied or with probability p_1, he may depart the system. Similarly, at the end of providing service, the server can either opt to take vacation with probability p_2 or be idle with probability q_2. During the busy period, the server may experience breakdown. Both the breakdown times and repair times of the server follow exponential distribution with parameter σ and δ respectively. The resulting QBD process is analysed in the steady state by employing matrix analytic method. The busy period analysis of our model has also been done. Finally, the numerical and graphical illustration of our model has been given.

Keywords: Markovian arrival process, Phase type distribution, Retrial queues, Bernoulli vacation, Bernoulli feedback, Breakdown and Repair.

I Introduction

On contrary to the normal queueing model, customers arriving to the retrial queueing system join the orbit, if all the servers are busy. They stay in the orbit for some amount of time which is usually exponentially distributed and then they try to know whether there is a possibility to receive service. If the server is free at that moment, they start to receive the service; if not, they come back to the orbit and repeat the process again. In this situation, if the server is available, then there is also a possibility for the new arrival to receive service directly without joining the orbit. Due to this complexity, it is usually tedious to derive analytic results for the retrial models. However, a vast amount of numerical and approximation methods are available to study the retrial queueing systems.

One of the most versatile modelling tools in the theory of point processes is the Markovian Arrival Process(MAP). Neuts(1979) introduced a new concept namely Versatile Markovian PointProcesses(VMPP) to model the arrival processes which are not essentially renewal processes. In order to understand VMPP in a clear and simpler way, Lucantoni et al (1990) coined two new terms, namely MAP and Batch MAP. Chakravarthy (2010) have greatly discussed about MAP in the Encyclopaedia of Operations Research and Management Science. A MAP is usually characterized by the parameter matrices (D_0, D_1), each of which of dimension n in which D_0 governs the change over related to no arrivals whereas D_1 governs the change over related to arrivals. The generator matrix of the resulting continuous time Markov chain is given by $Q = D_0 + D_1$. The point process characterized by the MAP is a peculiar class of semi-Markov process whose transition probability matrix is given by

$$\int_0^t e^{D_0 x} D_1 dx = [I - e^{D_0 t}](-D_0)^{-1} D_1.$$

Let the transition probability vector of the generator matrix $D = D_0 + D_1$ be denoted by π such that $\pi Q = 0$ and $\pi e = 1$. Then, the average arrival count per unit time in the stable version of the MAP, also termed as the fundamental rate is defined as $\lambda = \pi D_1 e_m$. Latouche et al (1999) have deeply discussed about PH-distributions and QBD.

If the duration between consecutive retrials are exponentially distributed with rate nα, where n is the size of the orbit, then such retrial queues are said to follow classical retrial policy and most of the retrial queues follow this policy. But, present situation in communication protocols and local area networks indicates that there are a vast number of queueing situations in which the retrial rate does not depend on the orbit size. This is known as constant retrial policy and it was first studied by Fayolle (1986) while examining the telephone exchange model as a classical single server Markovian retrial queue. He has modelled his system in such a way that the orbital customers make a queue and the request for service can be made by the foremost customer in the waiting line and the retrial times follow exponential distribution with rate α. Artalejo et al (2000) studied the Markovian retrial queueing system with multiserver and constant retrial rate.

If a server gets breakdown, then the customer who is currently receiving service has the choice of being in the system till the repair process gets over or permanently move out of the system or go back to the orbit to resume the service. Such situations are mostly seen in computer and communication networks, at airport with stacked aircraft and in retail shops. Aissani (1998), Kulkarni and Choi (1990) have studied about the effect of unreliable server and the repair process on retrial queues. Reliability of the non-Markovian queueing system when the server is prone to breakdown was analysed by Aissani and Artalejo (1998).

Wang et al (2009) have greatly analyzed the discrete time retrial queue in which server is subject to breakdown and repair and obtained generating function for both system size and orbit size. Retrial queues with batch Markovian Arrival Process, breakdown and repair has been investigated by Li et al (2006) They studied the system by combining matrix-analytic method and the censoring technique with the supplementary variable method. By using Generalized Stochastic Petri Nets(GSPN), Gharbi et al (2011) have proposed an approach for analyzing the finite source retrial systems in which server is prone to breakdown and repair. Efrosin et al (2011) have examined the Markovian retrial queueing model with constant retrial rate and an unreliable server. Dimitriou et al (2010) have investigated the repairable retrial model and derived the stability condition for it.

The concept of Bernoulli schedule vacation is that after providing service to the customer, the server may opt to take vacation with probability p or starts the service to next customer with probability $1 - p$. Keilson et al (1986) have analyzed the non-Markovian vacation queueing model with Bernoulli schedules. The average waiting time for the non-Markovian cyclic service queues with Bernoulli schedules has been derived by Servi (1986).

The concept of Bernoulli feedback is that after receiving the service, the customer can go to orbit with probability p to get the service again or departs the system permanently with probability $q(p + q = 1)$. While investigating the single server non-Markovian model, Tak\`αcs (1963) has introduced the concept of feedback mechanism. Krishnakumar et al (2010) have studied the single server Markovian retrial queueing system with feedback and derived the joint distribution of the server state and the orbit size by using generating function technique. Choudhury et al (2005) have derived the distribution for queue size and busy period for the non-Markovian queueing model with two stages of non-homogeneous services and Bernoulli feedback mechanism. Chen et al (2015) have greatly analysed the concept of Bernoulli vacation policy and Bernoulli feedback for the retrial model in discrete time. Badamchi Zadeh et al (2008) have discussed about the non-Markovian queueing system with Bernoulli feedback and Bernoulli vacation. Retrial queues in which the server is prone to experience breakdown and repair have been studied by Kulkarni and Choi (1990). The steady state probability vectors of our model are computed by employing matrix geometric method which was introduced by M. F. Neuts (1981). The rate matrix is obtained by means of the logarithmic reduction algorithm specified by Latouche et al (1999).

Hunter (1983) has shown that the queue length process examined at every distinct embedding is a Markov renewal process. He has given more emphasis on finding discrete time outputs for the queue length processes. Finally, he has examined the Markovian models with finite/infinite waiting space and the instantaneous Bernoulli feedback. A non-Markovian queueing model with Bernoulli feedback mechanism has been studied by Disney et al. (1980). They have shown that the output of their model is also Markov-renewal. Foley et al. (1983) have studied the non-Markovian queueing system with two servers and delayed feedback mechanism. They have given a choice for the customers who have received service from the lower server, either to depart the system or to go to the upper server to get the re-service, They have shown that their system is uniquely weakly lumpable to a Poisson process.

A non-Markovian batch arrival queueing model with Bernoulli vacation under multiple vacation policy has been investigated by Choudhury et al. (2018). The busy period distribution and the waiting time distribution have been derived for their model under steady state. Chakravarthy (2008) has studied the multi server queueing model in which arrival follows MAP and the vacation times follow phase type distribution. He has obtained the waiting time distribution and has presented various numerical examples. Chang et al. (2018) have studied the multi server Markovian retrial queueing model with feedback customers and unreliable servers. They have calculated the stationary distribution by developing a new recursive algorithm and have derived the cost function. Finally, they have made a comparison to validate the exactness of the approximate optimal solution.

The main motivating factor for our model is from online shopping. Nowadays, most of the people give more preferences to online shopping. Every company is selling different varieties of goods like dresses, electronic gadgets, house hold articles, etc,. These items may be considered as different phases. A customer may get service(booking a product) from any of these phases. If a customer is arriving during the busy period, he/she has to wait for some time in the invisible queue(orbit). After providing service to the customer, the server may either go for other jobs like updating their websites, launching new products, etc., (which may be considered as vacation) or the server may remain idle. During busy period, the server may get breakdown(the server problem). The customer who is receiving service at that moment has to join with the orbital customers and retry to get the service again. Moreover, after booking a particular item, if the customer is not satisfied with it(that is in the sense of price or quality of the item), the customer may again join the orbit and retry to get the service. We have framed our model in such a way that it will match with this situation.

On analysing the literature of queueing theory, we could find many articles that have discussed about the Bernoulli schedule vacation, Bernoulli feedback, breakdown and repair individually and also in different combinations for various arrival and service patterns. In our work, we have modelled our system in such a way that with all these attributes, we have considered MAP for arrival and phase type distribution for service times.

The rest of our work is structured in this manner: Section 2 briefly discusses about the model under study. The infinitesimal generator matrix is obtained in the Section 3. The analysis of this paper has been done in the steady state in the Section 4. The busy period analysis of our model has been done in the Section 5. Some of the performance measures are evaluated in Section 6. Finally, under the Section 7, the behavioural aspects of our queueing model has been analysed with the support of numerical values and graphical representations.

II Model Description

We consider a retrial queueing model with single server in which the customers come to the system as indicated by the Markovian Arrival Process with D_0 and D_1 as its parameter matrices of dimension n. The service times are supposed to follow phase type distribution with representation (α, T) of order m such that $T_0 + Te = 0$. On an arrival of a customer, if the server is available, then he provides service instantly. If not, the customer has to join the orbit of infinite capacity. Irrespective of the orbit size, each unit makes retrial at a constant rate from the orbit. The inter retrial times follow exponential distribution with parameter μ. At the end of providing service to the customer, the server may either go for vacation with probability $\mathfrak{p}_2$ or remains idle with probability $\mathfrak{q}_2$ where $\mathfrak{p}_2 + \mathfrak{q}_2 = 1$. Similarly, after receiving the service, if the customer is satisfied, then he leaves the system with probability $\mathfrak{p}_1$. Otherwise, if the customer is not satisfied, then he joins the orbit with probability $\mathfrak{q}_1$ to get the service, where $\mathfrak{p}_1 + \mathfrak{q}_1 = 1$. But, he has no priority over the orbital customers and he has to compete with them to get the service. During the busy period, the server may get breakdown. As a result, the customer who is receiving service at that time has to join the orbit of infinite capacity. After the completion of repair process, the server becomes idle. The breakdown times, the repair times and the vacation times are all supposed to follow exponential distribution with parameters σ, δ and η respectively.

III The generator matrix

The formulation of the generator matrix for our queueing model has been done in this section. We will begin our study by describing the following notations which are needed.

Notations:

- $N(t)$: Number of customers in the orbit
- I_n: An n-dimensional identity matrix
- **e**: A column vector (of needed dimension) with each of its entries as 1
- $\otimes$: Kronecker multiplication of two matrices
- $\oplus$: Kronecker addition of two matrices
- $Y(t)$ - Nature of the server at time t,

where

$Y(t) = \{0$, the server is in vacation.
1, the server is idle.
2, the server is offering service.
3, the server is in breakdown.

- S(t): Phase of the service process at time t
- M(t): Phase of the Markovian Arrival Process at time t
- λ: Rate of arrival and is defined as $\lambda = \pi D_1 e$ where π is the invariant probability vector of

the generator matrix $D = D_0 + D_1$

- γ: Rate of service, where $\gamma = [\alpha(-T)^{-1}e]^{-1}$

It is obvious and can proved that$\{(N(t), Y(t), S(t), M(t))$:t 0$\}$ is a continuous time Markov chain (CTMC) whose state space is given below:

$$\mathbf{Y} = l^* \cup l(i)$$

where

$$l(i) = \{(i, 0, l): i \geq 1, 1 \leq l \leq n\} \cup \{(i, 1, l): i \geq 1, 1 \leq l \leq n\} \cup \{(i, 2, k, l): i \geq 1, 1 \leq k \leq m, 1 \leq l \leq n\}$$

$$\cup \{(i, 3, l): i \geq 1, 1 \leq l \leq n\}$$

The generator matrix of the Markov chain is as follows:

$$Q = \begin{bmatrix} B_{00} & B_{01} & 0 & 0 & 0 & 0 & 0 & B_{10} & A_1 & A_0 \\ 0 & 0 & 0 & 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ 0 & 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 & \cdots & \cdots \\ \cdots & \ddots & \ddots & \ddots & \cdots & & & & & \end{bmatrix}$$

where,

$B_{00} = [D_0 - \eta I_n \quad \eta I_n \quad 00 \quad D_0 \quad \alpha \otimes D_1 p_1 p_2 T^0 \otimes I_n \quad p_1 q_2 T^0 \otimes I_n \quad T \oplus (D_0 - \sigma I_n)] B_{01} = [D_1 \quad 0 \quad 0 \quad 00 \quad 0 \quad 0 \quad 0 q_1 p_2 T^0 \otimes I_n \quad q_1 q_2 T^0 \otimes I_n \quad I_m \otimes D_1 \quad e_m \otimes \sigma I_n]$

$B_{10} = [0 \quad 0 \quad 00 \quad 0 \quad \mu\alpha \otimes I_n 0 \quad 0 \quad 00 \quad 0 \quad 0] A_1 =$
$\begin{bmatrix} D_0 - \eta I_n & \eta I_n & 0 & 00 & D_0 - \mu I_n & \alpha \otimes D_1 & 0 p_1 p_2 T^0 \otimes I_n & p_1 q_2 T^0 \otimes I_n & T \oplus (D_0 - \sigma I_n) & 00 \\ \delta I_n & 0 & D_0 - \delta I_n \end{bmatrix}$

$A_2 = \begin{bmatrix} 0 & 0 & 0 & 00 & 0 & \mu\alpha \otimes I_n & 00 & 0 & 0 & 00 \\ 0 & 0 & 0 \end{bmatrix} A_0 =$
$\begin{bmatrix} D_1 & 0 & 0 & 00 & 0 & 0 & 0 q_1 p_2 T^0 \otimes I_n & q_1 q_2 T^0 \otimes I_n & I_m \otimes D_1 & e_m \otimes \sigma I_n 0 \\ 0 & 0 & D_1 \end{bmatrix}$

IV Analysis of the System

The analysis of the our model has been discussed in this section under steady state.

4.1 Stability Condition

Let us define $A = A_0 + A_1 + A_2$. Then, clearly A is an infinitesimal generator matrix and as a result, we can find an invariant probability vector Ψ of A which obeys

$$\Psi A = 0; \Psi e = 1$$

where the vector Ψ is given by $\Psi = (\psi_0, \psi_1, \psi_2, \psi_3)$.

The vector Ψ, partitioned as $\Psi = (\psi_0, \psi_1, \psi_2, \psi_3)$ is computed by solving the following equations:

$$\psi_0[D - \eta I_n] + \psi_2[p_2 T^0 \otimes I_n] = 0$$

$$\psi_1[\eta I_n] + \psi_1[D_0 - \mu I_n] + \psi_2[q_2 T^0 \otimes I_n] + \psi_3[\delta I_n] = 0$$

$$\psi_1[\alpha \otimes D_1 + \mu\alpha \otimes I_n] + \psi_2[T \oplus (D - \sigma I_n)] = 0$$

$$\psi_2[e_m \otimes \sigma I_n] + \psi_3[D - \delta I_n]$$

subject to

$$\psi_0 + \psi_1 + \psi_2 + \psi_3 = 1.$$

The necessary and sufficient condition required by the system to attain stability is $\Psi A_0 e < \Psi A_2 e$ i.e.,

$$\psi_0[D_1 e_n] + \psi_2[q_1 T^0 \otimes e_n + e_m \otimes D_1 e_n + e_m \otimes \sigma e_n] + \psi_3 D_1 e_n < \psi_1 \mu e_n.$$

4.2 The Transition Probability Vector

Let the transition probability vector of the infinitesimal generator Q be indicated as **x**.

This probability vector can be partitioned as: $x = (x_0, x_1, x_2, \ldots)$, where x_0 is of dimension $2n + mn$ and x_i is of dimension $3n + mn$, for $i \geq 1$.

Since **x** is a transition probability vector of Q, the following two conditions will be satisfied by it:

$$xQ = 0 and xe = 1$$

Once the condition for the system to be stable is achieved, the invariant probability vector **x** can be computed using

$$x_{i+1} = x_1 \mathrm{R}^i, i \geq 0$$

and the remaining vectors namely, $x_0 and x_1$ can be evaluated by solving the equations given below:

$$x_0 \mathrm{B}_{00} + x_1 \mathrm{B}_{10} = 0$$

$$x_0 \mathrm{B}_{01} + x_1 [\mathrm{A}_1 + \mathrm{RA}_2] = 0$$

based on the normalizing condition

$$x_0 e_{2n+mn} + x_1 [\mathrm{I} - \mathrm{R}]^{-1} e_{3n+mn} = 1$$

The rate matrix R can be evaluated by making use of Logarithmic Reduction Algorithm proposed by Latouche et al.(1999).

Logarithmic Reduction Algorithm:

$$Step0: H \leftarrow (-A_1)^{-1} A_0, L \leftarrow (-A_1)^{-1} A_2, G = L and T = H.$$

$$Step1:$$
$$U = HL + LH$$
$$M = H^2$$
$$H \leftarrow (I - U)^{-1} M$$
$$M \leftarrow L^2$$
$$L \leftarrow (I - U)^{-1} M$$
$$G \leftarrow G + TL$$
$$T \leftarrow TH$$
$$continue Step1 until \parallel e - Ge \parallel_\infty < \varepsilon.$$

$Step2: R = -A_0 (A_1 + A_0 G)^{-1}$

V Analysis of the busy period

Under this section, we perform the analysis of busy period of our system.

The duration of time between the customer's arrival to the void system and the instant at which the system size reaches zero for the first time is termed as the busy period. Hence, the first passage time from level 1 to level 0 is the analogue of the busy period.

The first return time to level zero with the condition that there should be atleast one visit to a state in any other levels is termed as the busy cycle. Let us first bring in the idea of fundamental period for the purpose of analysing the busy period. As far as the QBD process is concerned, it is nothing but the first passage time from the level i to the level $i-1$, where $i \geq 2$. The discussion has to be done separately for the boundary states (i.e.) for the cases $i = 0,1$. It can be easily seen that there are $3n + mn$ states for each level i where $i \geq 1$. Therefore, while arranging the states in the lexicographic order, j^{th} state of level i may be indicated as (i, j).

NOTATIONS:

- $G_{jj'}(\mathrm{k}, \mathrm{x})$ - The conditional probability that the QBD process enters the level $i-1$ by making precisely k transitions to the left and also by entering the state (i, j') given that it started in the state (i, j) at time $t = 0$.
- $\bar{G}_{jj'}(\mathrm{z}, \mathrm{s}) = \sum_{k=1}^{\infty} \mathrm{z}^{\mathrm{k}} \int_0^{\infty} e^{-\mathrm{sx}} dG_{jj'}(\mathrm{k}, \mathrm{x}) \colon |\mathrm{z}| \leq 1, \mathrm{Re}(\mathrm{s}) \geq 0$
- $\bar{G}(\mathrm{z}, \mathrm{s}) - The matrix (\bar{G}_{jj'}(\mathrm{z}, \mathrm{s}))$
- $G = (G_{jj'}) = \bar{G}(1,0)$- The matrix which concerns about the first passage times without including the boundary states.
- $G_{jj'}^{(1,0)}(\mathrm{k}, \mathrm{x})$ - The conditional probability that the QBD process enters the level 0 by making precisely k transitions to the left given that it started in the level 1 at time $t = 0$.
- $G_{jj'}^{(0,0)}(\mathrm{k}, \mathrm{x})$ - The first return time to the level 0.
- $\mathbb{E}_{1j}$ - The expected first passage time from the level i to the level $i-1$, given that at time $t = 0$, the process is in the state (i, j).
- $\vec{\mathbb{E}}_1$ - The column vector with $\mathbb{E}_{1j}$ as its entries.
- $\mathbb{E}_{2j}$ - The expected number of customers who received service during the first passage

time from the level i to the level $i-1$, given that the first passage time begins in the state (i,j).

- $\vec{\mathbb{E}}_2$ - The column vector with $\mathbb{E}_{2j}$ as its entries.
- $\vec{\mathbb{E}}_1^{(1,0)}$ - The vector which gives the expected first passage times from the level 1 to the level 0.
- $\vec{\mathbb{E}}_2^{(1,0)}$ - The vector which gives the expected number of service completions in the first passage time from the level 1 to the level 0.
- $\vec{\mathbb{E}}_1^{(0,0)}$ - The expected first return time to the level 0.
- $\vec{\mathbb{E}}_2^{(0,0)}$ - The expected number of service completions in the course of first return time to the level 0.

It can be easily seen that the matrix $\bar{G}(\mathrm{z},\mathrm{s})$ satisfies the following equation:

$$\bar{G}(\mathrm{z},\mathrm{s}) = \mathrm{z}[\mathrm{sI}-\mathrm{A}_1]^{-1}\mathrm{A}_2 + [\mathrm{sI}-\mathrm{A}_1]^{-1}\mathrm{A}_0\bar{G}^2(\mathrm{z},\mathrm{s})$$

Once the rate matrix R is evaluated, we can easily find the matrix G by making use of the result

$$G = -[\mathrm{A}_1 + \mathrm{RA}_2]^{-1}\mathrm{A}_2$$

The matrix G may also be evaluated by employing logarithmic reduction algorithm.

As far as the boundary states are concerned, namely 0 and 1, we have the following equations which are satisfied by $\bar{G}^{(1,0)}(\mathrm{z},\mathrm{s})$ and $\bar{G}^{(0,0)}(\mathrm{z},\mathrm{s})$ respectively.

$$\bar{G}^{(1,0)}(\mathrm{z},\mathrm{s}) = \mathrm{z}[\mathrm{sI}-\mathrm{A}_1]^{-1}\mathrm{B}_{10} + [\mathrm{sI}-\mathrm{A}_1]^{-1}\mathrm{A}_0\bar{G}(\mathrm{z},\mathrm{s})\bar{G}^{(1,0)}(\mathrm{z},\mathrm{s})$$

$$\bar{G}^{(0,0)}(\mathrm{z},\mathrm{s}) = [\mathrm{sI}-\mathrm{B}_{00}]^{-1}\mathrm{B}_{01}\bar{G}^{(1,0)}(\mathrm{z},\mathrm{s})$$

. Since, the three matrices namely, G, $\bar{G}^{(1,0)}(1,0)$ and $\bar{G}^{(0,0)}(1,0)$ are stochastic, we may easily evaluate the following moments:

$$\vec{\mathbb{E}}_1 = -\frac{\partial}{\partial \mathrm{s}}\bar{G}(\mathrm{z},\mathrm{s})|_{\mathrm{s}=0;\mathrm{z}=1} = -[\mathrm{A}_1 + \mathrm{A}_0(G+\mathrm{I})]^{-1}e$$

$$\vec{\mathbb{E}}_2 = \frac{\partial}{\partial \mathrm{z}}\bar{G}(\mathrm{z},\mathrm{s})|_{\mathrm{s}=0;\mathrm{z}=1,} = -[\mathrm{A}_1 + \mathrm{A}_0(G+\mathrm{I})]^{-1}\mathrm{A}_2 e$$

$$\vec{\mathbb{E}}_1^{\ (1,0)} = -\frac{\partial}{\partial \mathrm{s}}\bar{G}^{(1,0)}(\mathrm{z},\mathrm{s})|_{\mathrm{s}=0;\mathrm{z}=1} = -[\mathrm{A}_1 + \mathrm{A}_0 G]^{-1}[\mathrm{A}_0\vec{\mathbb{E}}_1 + e]$$

$$\vec{\mathbb{E}}_2^{\ (1,0)} = \frac{\partial}{\partial \mathrm{z}}\bar{G}^{(1,0)}(\mathrm{z},\mathrm{s})|_{\mathrm{s}=0;\mathrm{z}=1} = -[\mathrm{A}_1 + \mathrm{A}_0 G]^{-1}[\mathrm{B}_{10}e + \mathrm{A}_0\vec{\mathbb{E}}_2]$$

$$\vec{\mathbb{E}}_1^{\ (0,0)} = -\frac{\partial}{\partial \mathrm{s}}\bar{G}^{(0,0)}(\mathrm{z},\mathrm{s})|_{\mathrm{s}=0;\mathrm{z}=1} = -\mathrm{B}_{00}^{-1}[e + \mathrm{B}_{01}\vec{\mathbb{E}}_1^{\ (1,0)}]$$

$$\vec{\mathbb{E}}_2^{\ (0,0)} = \frac{\partial}{\partial \mathrm{z}}\bar{G}^{(0,0)}(\mathrm{z},\mathrm{s})|_{\mathrm{s}=0;\mathrm{z}=1} = -\mathrm{B}_{00}{}^{-1}\mathrm{B}_{01}\vec{\mathbb{E}}_2^{\ (1,0)}$$

VI Performance Measures

In order to examine the behaviour of our model in the steady state, a few performance measures for our model are enumerated in this section.

- Probability of orbit being empty:

  $P_{empty} = \sum_{j=0}^{1}\sum_{l=1}^{n} x_{0jl} + \sum_{k=1}^{m}\sum_{l=1}^{n} x_{02kl}$

- Probability of server to be idle:

  $P_{idle} = \sum_{i=0}^{\infty}\sum_{l=1}^{n} x_{i1l}$

- Probability of server to be in vacation:

  $P_{vacation} = \sum_{i=0}^{\infty}\sum_{l=1}^{n} x_{i0l}$

- Probability of server to be busy:

  $P_{busy} = \sum_{i=0}^{\infty}\sum_{k=1}^{m}\sum_{l=1}^{n} x_{i2kl}$

- Probability of server to be in breakdown:

$$P_{breakdown} = \sum_{i=1}^{\infty} \sum_{l=1}^{n} x_{i3l}$$

- Probability of a new arrival getting into service directly:

$$P_s = \frac{1}{\lambda}\{\sum_{i=0}^{\infty} x_{i1} D_1 e\}$$

- Probability of a new arrival getting to receive service directly with a minimum of one customer waiting in the orbit:

$$P_{sw} = \frac{1}{\lambda}\{\sum_{i=1}^{\infty} x_{i1} D_1 e\}$$

- The total retrial rate at which the orbital customers appeal for service:

$$\mu^* = \mu\{\sum_{i=1}^{\infty} \sum_{l=1}^{n} x_{i0l} + \sum_{i=1}^{\infty} \sum_{l=1}^{n} x_{i1l} + \sum_{i=1}^{\infty} \sum_{k=1}^{m} \sum_{l=1}^{n} x_{i2kl} + \sum_{i=1}^{\infty} \sum_{l=1}^{n} x_{i3l}\}$$

- The effective retrial rate :

$$\mu_s = \mu\{\sum_{i=1}^{\infty} \sum_{l=1}^{n} x_{i1l}\}$$

- Expected orbit size

$$E_{orbit} = \sum_{i=1}^{\infty} i x_i e_{3n+mn}$$

- Average system size:

$$E_{system} = E_{orbit} + P_{busy}$$

- Probability of a successful retrial:

$$P_{sr} = \frac{\mu}{\mu+\lambda}\{\sum_{i=1}^{\infty} \sum_{l=1}^{n} x_{i1l}\}$$

- Mean number of successful retrial:

$$E_{srt} = \frac{\mu}{\mu+\lambda}\{\sum_{i=1}^{\infty} \sum_{l=1}^{n} i x_{i1l}\}$$

VII Numerical Results

In this section, we will analyse the behaviour of our model numerically as well as graphically. The following five different MAP representations, all of which have the same mean, say 1, are taken into consideration for the arrival process.

Erlang of order 2:

$$D_0 = [-2 \quad 20 \quad -2]; D_1 = [0 \quad 02 \quad 0]$$

Exponential:

$$D_0 = [-1]; D_1 = [1]$$

Hyperexponential:

$$D_0 = [-1.90 \quad 00 \quad -0.19]; D_1 = [1.710 \quad 0.1900.171 \quad 0.019]$$

Since, all these three arrival process are renewal, their correlation is zero.

Consider the following three phase type distribution for the service times.

Erlang of order 2:

$$\alpha = (1,0); T = [-12 \quad 120 \quad -12]$$

Exponential:

$$\alpha = (1); T = [-6]$$

Hyperexponential:

$$\alpha = (0.8,0.2); T = [-16.8 \quad 00 \quad -1.68]$$

Illustrative Example 1:

In the following tables, we examine the impact of the retrial rate μ against the expected orbit size. Fix $\lambda = 1; \gamma = 6; \mathfrak{p}_1 = 0.6; \eta = 2; \mathfrak{p}_2 = 0.4; \sigma = 1; \mathfrak{q}_1 = 0.4; \delta = 2; \mathfrak{q}_2 = 0.6$.

Table 1: *Expected orbit size - Exponential Service*

μ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
11	10.0011675	7.48381067	28.7824173
13	7.70793519	5.79663583	21.8380220
15	6.53032933	4.92315841	18.2956968
17	5.81352572	4.38926486	16.1464306
19	5.33131245	4.02923081	14.7031629
21	4.98472155	3.77006250	13.6669608

Table 2: *Expected orbit size - Erlang Service*

μ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
11	12.7430548	9.45396003	37.5253785
13	9.29498782	6.94117250	26.9239288
15	7.66359572	5.73985615	21.9503036
17	6.71251431	5.03600320	19.0617824
19	6.08961722	4.57372333	17.1738503
21	5.65003242	4.24691754	15.8431278

Table 3: *Expected orbit size - Hyperexponential Service*

μ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
11	4.86515668	3.72736456	12.8474069
13	4.18208704	3.21276505	10.8735917
15	3.76653401	2.89837776	9.67760920
17	3.48707744	2.68644585	8.87528324
19	3.18123908	2.53393179	8.29971350
21	3.13499689	2.41893523	7.86667776

From the Table 1-3, we have the following observations.

• While maximizing the retrial rate, the expected orbit size minimizes for different arrangements of service and arrival times.

• While comparing to Erlang and Exponential arrival times, the expected orbit size decreases more rapidly in the case of hyperexponential arrival time. Similarly, the expected orbit size decreases slowly in the case of Erlang arrival time.

Illustrative Example 2:

In the following tables, we examine the impact of the service rate γ on the expected orbit size. Fix $\lambda = 1; \mu = 8; \mathfrak{p}_1 = 0.60; \eta = 2; \mathfrak{p}_2 = 0.40; \sigma = 1; \mathfrak{q}_1 = 0.40; \delta = 2; \mathfrak{q}_2 = 0.60$.

Table 4: *Expected orbit size - Exponential Service*

γ	ARRIVAL		
	EXPONENTIAL	**ERLANG**	**HYPEREXPONENTIAL**
7	8.62406523	6.38645821	24.9106128
8	5.34880881	3.98252157	14.9745343
9	3.99796043	2.98468674	10.9073389
10	3.26376783	2.44125073	8.70686572
11	2.80380376	2.10058231	7.33376198
12	2.48925463	1.86760598	6.39840730
13	2.26091389	1.69851986	5.72205463

Table 5: *Expected orbit size - Erlang Service*

γ	ARRIVAL		
	EXPONENTIAL	**ERLANG**	**HYPEREXPONENTIAL**
7	10.0886405	7.41091127	29.7073972
8	5.78766750	4.28690000	16.4457714
9	4.20009203	3.12293216	11.6004235
10	3.37742675	2.51789282	9.10389213
11	2.87562366	2.14837061	7.58839020
12	2.53827076	1.89982294	6.57424954
13	2.29625267	1.72148762	5.85002618

Table 6: *Expected orbit size - Hyperexponential Service*

γ	ARRIVAL		
	EXPONENTIAL	**ERLANG**	**HYPEREXPONENTIAL**
7	4.97313477	3.77351943	13.3210642
8	3.82708589	2.90365725	10.0520703
9	3.17297465	2.40685867	8.19104494
10	2.75103836	2.08653058	6.99299693
11	2.45682845	1.86336640	6.15918869
12	2.24027528	1.69928391	5.54661758
13	2.07439882	1.57374500	5.07829007

From the Table 4-6, we have the following observations.

• While maximizing the service rate, the expected orbit size minimizes for various possible arrangements of arrival and service times.

• While comparing to Erlang and Exponential arrival times, expected orbit size decreases more rapidly forhyperexponential arrival time. Similarly, the expected orbit size reduces gradually for Erlang arrival time.

Illustrative Example 3:

In the following tables, we examine the impact of the vacation rate η on expected orbit size. Fix $\lambda = 1; \gamma = 6; \mathfrak{p}_1 = 0.60; \mu = 8; \mathfrak{p}_2 = 0.40; \sigma = 1; \mathfrak{q}_1 = 0.40; \delta = 2; \mathfrak{q}_2 = 0.60$.

Table 7: *Expected orbit size - Exponential Service*

η	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
4	3.37760655	2.48794146	9.31560352
6	2.38366783	1.76179679	6.27202950
8	2.04477644	1.51618610	5.22962692
10	1.87519953	1.39392122	4.70752570
12	1.77370727	1.32100067	4.39510928
14	1.70624181	1.27264837	4.18755323
16	1.65818186	1.23826766	4.03979561

Table 8: *Expected orbit size - Erlang Service*

η	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
4	3.76004704	2.74496640	10.6614094
6	2.59592138	1.90098311	7.03743508
8	2.20977550	1.62305018	5.82862515
10	2.01854731	1.48608798	5.22879429
12	1.90472438	1.40484025	4.87154080
14	1.82932015	1.35114669	4.63485557
16	1.77572971	1.31305560	4.46666909

Table 9: *Expected orbit size - Hyperexponential Service*

η	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
4	2.20929153	1.67412072	5.46231400
6	1.65814092	1.26336768	3.87581037
8	1.45463836	1.11330793	3.29231358
10	1.34958578	1.03632469	2.99270300
12	1.28565431	0.989662976	2.81122828
14	1.24271322	0.958408469	2.68981153
16	1.21190604	0.936030889	2.60298246

From the Table 7-9, we have the following observations.

• While maximizing the vacation rate, the expected orbit size minimizes for various possible arrangements of arrival and service times.

• While comparing to Erlang and Exponential arrival times, the expected orbit size decreases more rapidly for hyperexponential arrival time. Similarly, the expected orbit size reduces gradually for Erlang arrival time.

Illustrative Example 4:

In the following tables, we examine the impact of the repair rate δ against the expected orbit size. Fix $\lambda = 1; \gamma = 6; \mathfrak{p}_1 = 0.60; \eta = 2; \mathfrak{p}_2 = 0.40; \sigma = 1; \mathfrak{q}_1 = 0.40; \mu = 8; \mathfrak{q}_2 = 0.60$.

Table 10: *Expected orbit size - Exponential Service*

δ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
4	7.59759409	5.51219114	22.7455157
6	5.87750594	4.25241592	17.5011470
8	5.24414894	3.78930406	15.5590737
10	4.91583865	3.54965983	14.5484770
12	4.71517696	3.40339073	13.9291648
14	4.57989976	3.30488730	13.5108519
16	4.48255049	3.23406112	13.2093878

Table 11: *Expected orbit size - Erlang Service*

δ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
4	9.01354432	6.47340005	27.6296447
6	6.68529225	4.79331402	20.3620356
8	5.87341675	4.20725586	17.8186074
10	5.46167788	3.91034115	16.5247530
12	5.21302166	3.73120654	15.7416246
14	5.04665662	3.61145470	15.2167986
16	4.92756297	3.52578803	14.8406217

Table 12: *Expected orbit size - Hyperexponential Service*

δ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
4	4.31052599	3.21966516	11.8178685
6	3.70054138	2.75276167	10.1062609
8	3.44417977	2.55760891	9.37940438
10	3.30337635	2.45078331	8.97785988
12	3.21444535	2.38346547	8.72330076
14	3.15320451	2.33718412	8.54755003
16	3.10847255	2.30342084	8.41893285

From the Table 10-12, we have the following observations.

- While maximizing the repair rate, the expected orbit size minimizes for various possible arrangements of arrival and service times.
- While comparing to Erlang and Exponential arrival times, the expected orbit size decreases more rapidly in the case hyperexponential arrival time. Similarly, the expected orbit size decreases gradually for Erlang arrival time.

Illustrative Example 5:

In the following tables, we examine the impact of the breakdown rate σ against expected orbit size. Fix $\lambda = 1; \gamma = 6; \mathfrak{p}_1 = 0.60; \eta = 2; \mathfrak{p}_2 = 0.40; \delta = 2; \mathfrak{q}_1 = 0.40; \mu = 8; \mathfrak{q}_2 = 0.60$.

Table 13: *Expected orbit size - Exponential Service*

σ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
0.2	4.08040357	2.96155493	11.8191902
0.3	4.73131254	3.44689911	13.7479083
0.4	5.54494873	4.05191775	16.1683587
0.5	6.59105242	4.82715640	19.2937651
0.6	7.98585710	5.85628764	23.4817132
0.7	9.93858363	7.28856770	29.3811856
0.8	12.8676733	9.41875439	38.3052798

Table 14: *Expected orbit size - Erlang Service*

σ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
0.2	4.08450150	2.94457362	11.9867048
0.3	4.80393168	3.47566946	14.1600116
0.4	5.73618355	4.16175105	16.9879714
0.5	6.99170138	5.08201017	20.8145996
0.6	8.77333791	6.38066758	26.2765620
0.7	11.4985662	8.35096411	34.6983724
0.8	16.1853911	11.6946701	49.3649830

Table 15: *Expected orbit size - Hyperexponential Service*

σ	ARRIVAL		
	EXPONENTIAL	ERLANG	HYPEREXPONENTIAL
0.2	4.02198460	3.02084111	10.8591758
0.3	4.32444739	3.25474173	11.6877493
0.4	4.65258139	3.50785258	12.5914617
0.5	5.01076886	3.78348382	13.5827069
0.6	5.40425830	4.08558108	14.6764439
0.7	5.83942470	4.41891647	15.8909556
0.8	6.32412046	4.78934454	17.2488808

From the Table 13-15, we have the following observations.

- While maximizing the breakdown rate, the mean orbit size also maximizes for all possible arrangements of arrival and service times.
- While comparing to exponential and hyperexponential service times, the mean orbit size increases more rapidly for Erlang service time. Similarly, the mean orbit size increases slowly for hyperexponential service time.

Illustrative Example 6:

In the following figures, we observe the impact of both the retrial rate μ and vacation rate η against expected system size. Fix $\lambda = 1; \mathfrak{p}_1 = 0.60; \gamma = 6; \mathfrak{p}_2 = 0.40; \sigma = 1; \mathfrak{q}_1 = 0.40; \delta = 2; \mathfrak{q}_2 = 0.60$.

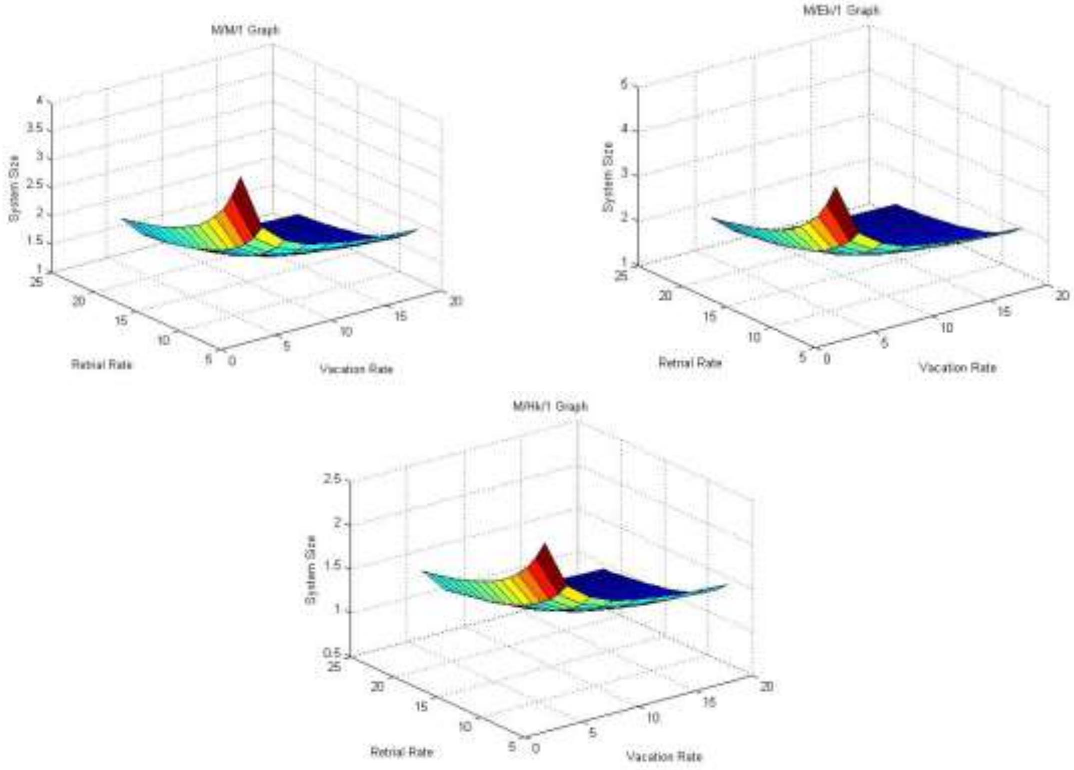


Figure 1: *Exponential arrival*

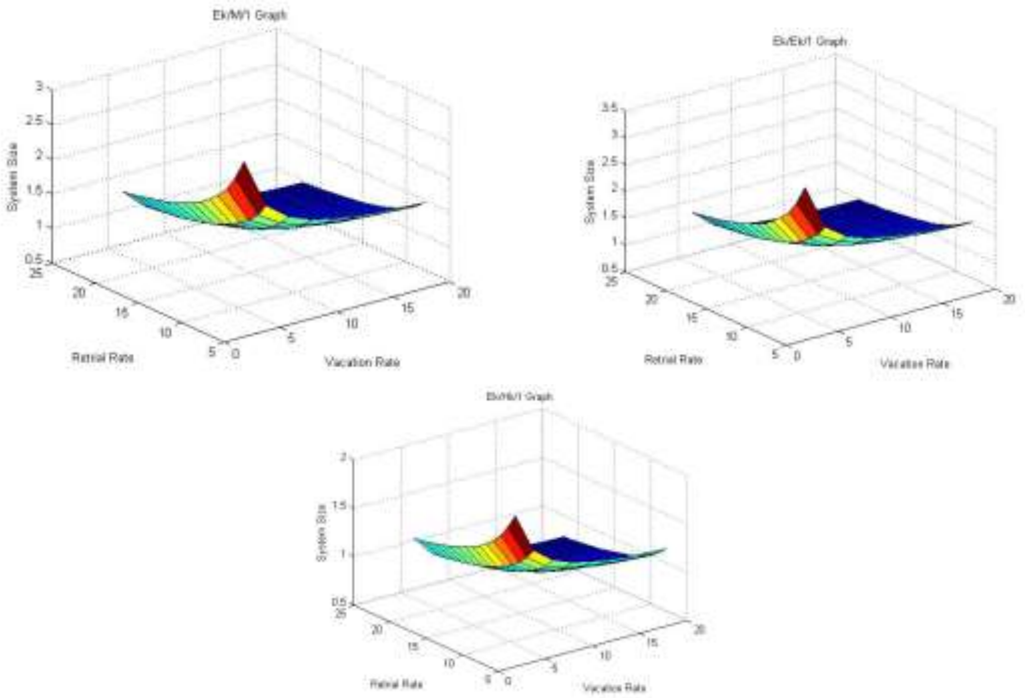


Figure 2: *Erlang arrival*

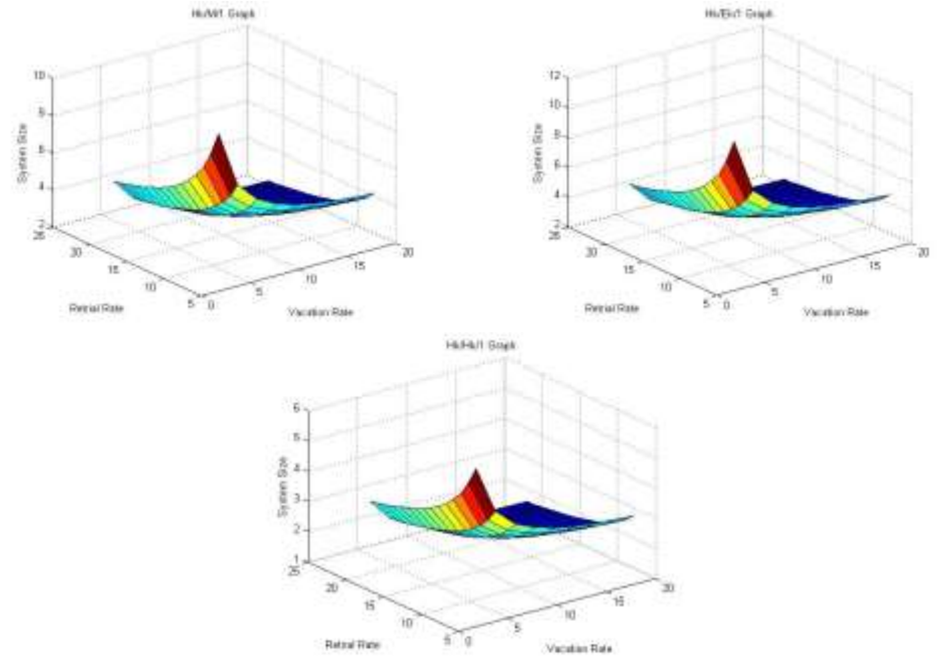


Figure 3: *Hyperexponential arrival*

A quick observation of Figure 1-3 discloses the fact that expected system size decreases while maximizing both the retrial rate and vacation rate for all possible arrangements of arrival and service times.

Illustrative Example 7:

In the following figures, we analyse the influence of both the repair rate δ and the service rate γ against the probability of server being idle.

We fix $\lambda = 1; \mathfrak{p}_1 = 0.60; \eta = 2; \mathfrak{p}_2 = 0.40; \mu = 8; \mathfrak{q}_1 = 0.40; \sigma = 1\mathfrak{q}_2 = 0.60$.

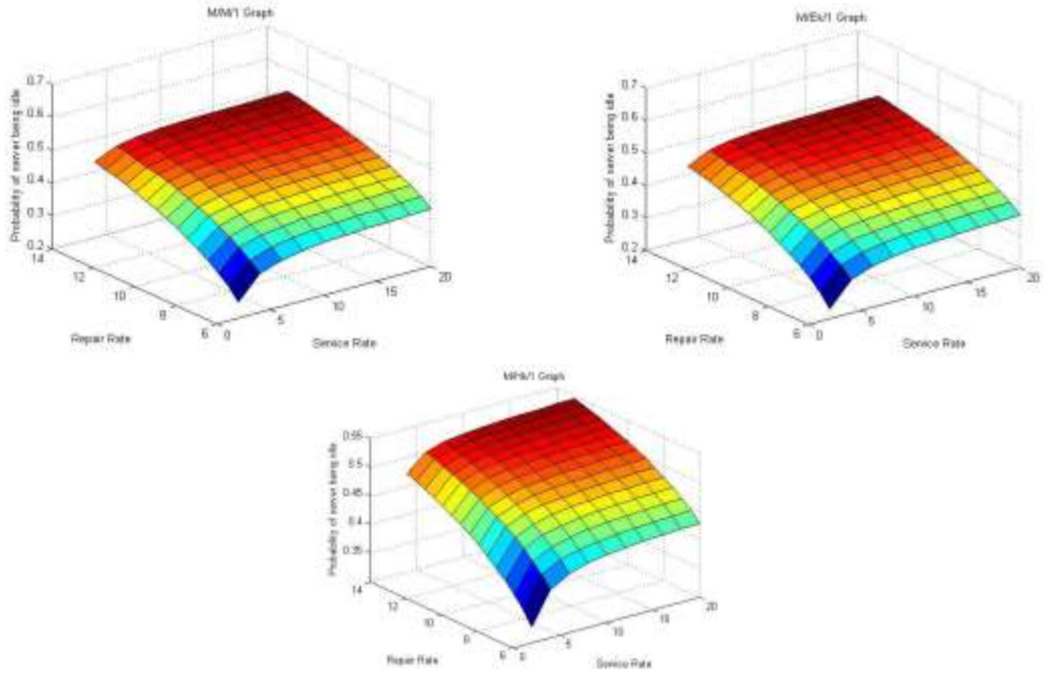


Figure 4: *Exponential arrival*

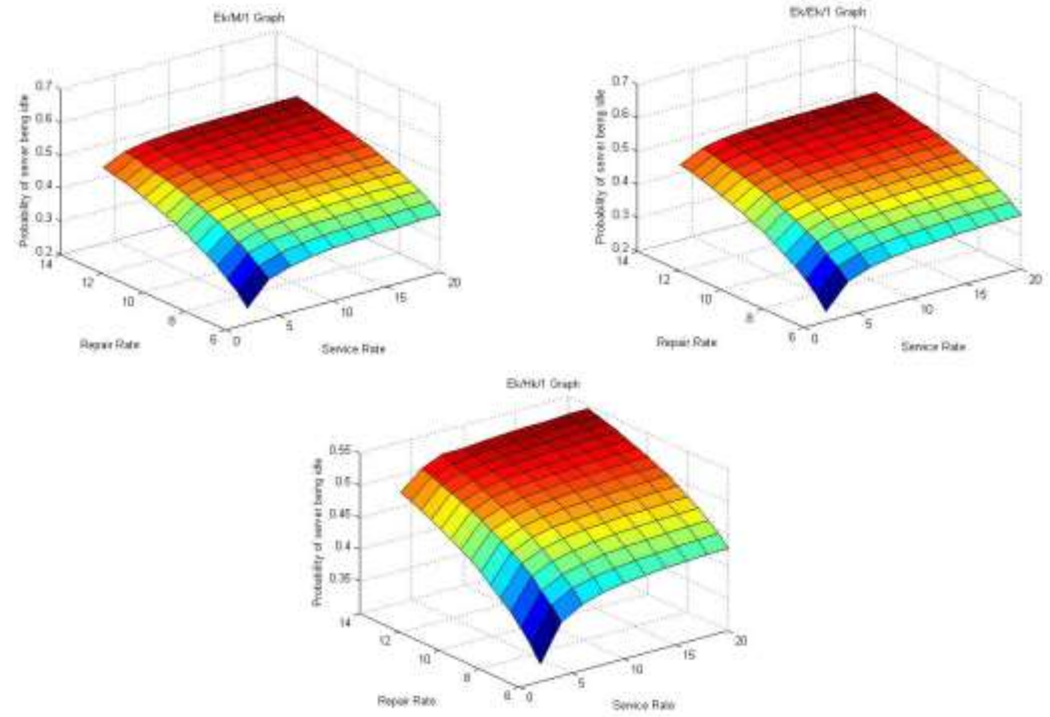


Figure 5: *Erlang arrival*

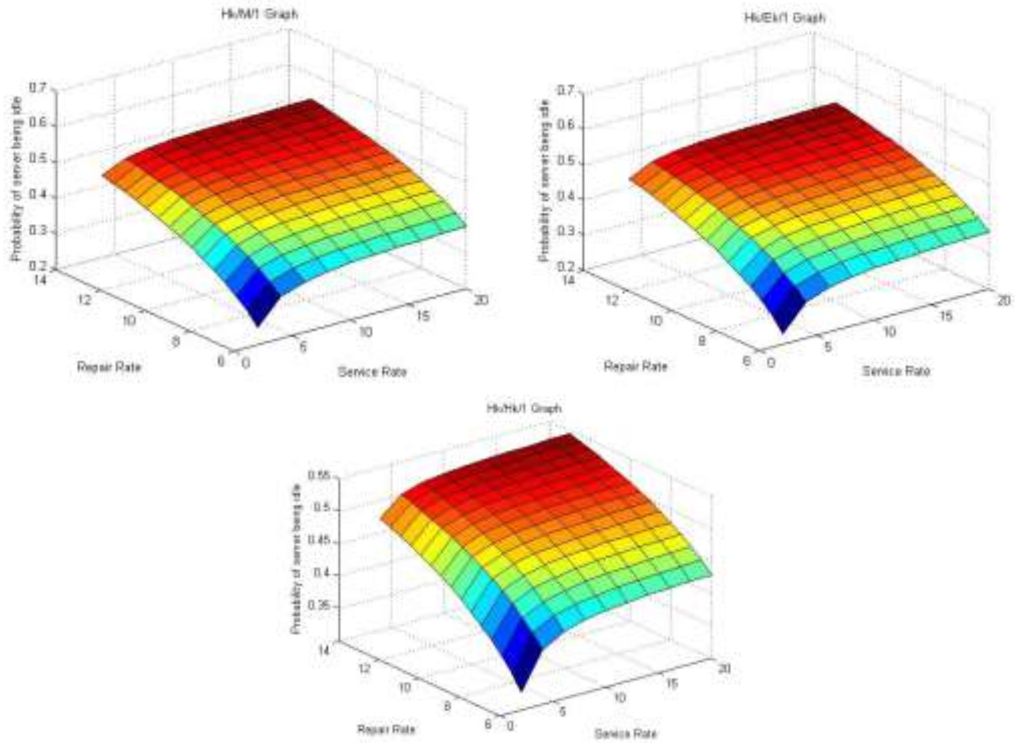


Figure 6: *Hyperexponential arrival*

A quick view of Figure 4-6 reveals the fact that the probability of server being idle increases while increasing both the repair rate and the service rate for all possible arrangements of arrival and service times.

VIII Conclusion

In our work, we have discussed about the retrial queuing system in which arrival follows MAP and service time follows PH-distribution together with Bernoulli schedule vacation, Bernoulli feedback, breakdown and repair. The effect of retrial rate, repair rate, vacation rate, breakdown rate and service rate on the average orbit size has been analysed through numerical values. Also, the influence of both the repair rate and the service rate against the probability of server being idle and the effect of retrial rate and the vacation rate on the average system size have been clearly visualized with the support of graphical representations. We have also analysed the busy period for our model. Our work can be extended to queueing models in which arrival follows BMAP.

References

[1] Aissani, A. and Artalejo, J.R. (1998). On the single server retrial queue subject to breakdowns. *Queueing Systems*, 30:309–321.

[2] Artalejo, J.R., Comez-Corral, A. and Neuts, M.F. (2000). Analysis of multiserver queues with constant retrial rate. *European Journal of Operational Research*, 135:569–581.

[3] Badamchi Zadeh, A. and Shankar, G.H. (2008). A two phase queue system with Bernoulli feedback and Bernoulli schedule server vacation. *Information and Management Science*, 19-2:329–338.

[4] Chakravarthy, S.R. (2009). Analysis of multi-server queue with Markovian arrivals and synchronous phase type vacations. *Asia-Pacific Journal of Operational Research*, 26-1:85–113.

[5] Chakravarthy, S.R. (2010). Markovian arrival processes. *Wiley Encyclopaedia of Operations Research and Management Science*.

[6] Cohen, J.W. (1957). Basic problems of telephone traffic theory and the influence of repeated calls. *Philips Telecommunication Review*, 18-2:49-100.

[7] Dimitry Efrosin and Anastasia Winkler (2011). Queueing system with a constant retrial rate, non-reliable server and threshold-based recovery. *European Journal of Operational Research*, 210:594-605.

[8] Fayolle, G. (1986). A simple telephone exchange with delayed feedbacks. *International Seminar on Teletraffic Analysis and computer Performance Evaluation*, 245–253.

[9] Fu-Min Chang, Tzu-Hsin Liu, Jau-Chuan Ke (2018). On an unreliable server retrial queue with customer feedback and impatience. *Applied Mathematical Modelling*, 55:171–182.

[10] Gautam Choudhury and Madhuchanda Paul (2005). A two phase queueing system with Bernoulli feedback. *Information and Management Science*, 16-1:35-52.

[11] Gautam Choudhury and Mitali Deka (2018). A batch arrival unreliable server delaying repair queue with two phases of service and Bernoulli vacation under multiple vacation policy.. *Quality Technology and Quantitative Management*,15-2:157–186.

[12] Ioannis Dimitriou and Christos Langaris (2010). A repairable queueing model with two-phase service, start-up times and retrial customers. *Computers and Operations Research*, 37:1181–1190.

[13] Jeffrey J. Hunter (1983). Filtering of Markov Renewal Queues, I: Feedback Queues. *Advances in Applied Probability*, 15-2:349-375.

[14] Jin-ting Wang and Peng Zhang (2009). A single-server discrete-time retrial G-queue with server breakdowns and repairs. *Acta Mathematicae Applicatae Sinica, English Series*, 25-4:675–684.

[15] Keilson, J. and Servi, L.D. (1986). Oscillating random walks doles for GI/G/1 vacation system with Bernoulli schedules. *Journal of Applied Probability*, 23-3:790–802.

[16] Krishna Kumar, B., Vijayalakshmi, G., Krishnamoorthy, A. and Sadiq Basha, S. (2010). A single server feedback retrial queue with collisions. *Computers and Operations Research*, 37:1247–1255.

[17] Kulkarni, V.G. and Choi, B.D. (1990). Retrial queues with server subject to breakdowns and repairs. *Queueing Systems*, 7:191–208.

[18] Latouche, G. and Ramaswami, V. (1999). Introduction to Matrix Analytic Methods in Stochastic Modelling. *American Statistical Association, SIAM*, Alexandria, Philadelphia.

[19] Lucantoni, D, Meier-Hellstern, K.S. and Neuts, M.F. (1990). A single- server queue with server vacations and a class of nonrenewal arrival processes. *Advanced Applied Probability*, 22:676–705.

[20] Nawel Gharbi and Claude Dutheillet (2011). An algorithmic approach for analysis of finite-source retrial systems with unreliable server. *Computers and Mathematics with Applications*, 62:2535–2546.

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

[22] Neuts, M.F. (1979). A versatile markovian point process. *Journal of Applied Probability*, 16:764–779.

[23] Peishu Chen, Yongwu Zhou, Changawen Li (2015). Discrete-time retrial queue with Bernoulli vacation, preemptive resume and feedback customers. *Journal of Industrial Engineering and Management*, 8-4:1236–1250.

[24] Quan-Lin Li, Yu Ying, Yiqiang Q.Zhao (2006). A BMAP/G/1 retrial queue with a server subject to breakdowns and repairs. *Annals of Operations Research*, 141-1:233–270.

[25] Ralph L. Disney Donald C. McNickle Burton Simon (1980). The M/G/1 queue with instantaneous bernoulli feedback. *Naval Research Logistics: A Journal dedicated to Advances in Operations and Logistics Research*, 27-4:635–644.

[26] Robert D. Foley and Ralph L. Disney (1983). Queues with Delayed Feedback. *Advances in Applied Probability*, 15-1:162–182.

[27] Servi, L.D. (1986) Average delay approximation of M/G/1 cyclic service queues with Bernoulli schedules. *IEEE Journal on Selected Ares in Communications*, 4-6:813–822.

[28] Takács, L. (1963). A single-server queue with feedback. *The Bell system Technical Journal*, 42-2:505–519.

ISSN 1932-2321