International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 1, ISSN No. 2455-2143, Pages 116-120
Published Online November-December 2016 in IJEAST (http://www.ijeast.com)

ESTIMATION OF PARAMETERS OF DAGUM DISTRIBUTION SOFTWARE RELIABILITY GROWTH MODEL

K. Harishchandra
Department of Statistics
Bangalore University, Bengalur,
Karnataka, India-560 056

Balakrishna
Department of Statistics
Vijaya Evening College, Bengalur,
Karnataka, Indian- 560 004

Abstract— Non Homogeneous Poisson process models with expected number of faults detected in given testing time are proposed in the literature. These models show either constant, monotonic increasing or monotonic decreasing failure occurrence rate per fault. In this article we propose a software reliability model in which the distribution of time between two failures assumed to be Dagum distribution. The model can capture increasing/decreasing nature of failure rate. The parameters of the model are estimated using maximum likelihood method. A simulation study and real data are used to verify the model.

Keywords: *Software reliability, fault detection process, failure occurrence rate per fault, Maximum likelihood estimation.*

I. INTRODUCTION

A computer is an electronic device used for quick and accurate function of any phenomena. Today it has been widely used for control of many complex systems. The efficiency of a computer depends on its two main ingredients namely hardware and the software. The quality of hardware and software of a system can be described by many parameters such as complexity, portability, maintainability, availability, reliability, etc. The Reliability of the computer both in terms of hardware and software is of very much essential for development of computer field. Software Reliability received great attention to deal with statistical problems. Software reliability is defined as *the probability of failure –free operation of a computer program in a specified environment for a specified period of time* (Musa and Okumoto, [5]). Over the last three decades many software reliability growth models have appeared in the literature. Many of these models describe the failure behavior in terms of failure occurrence rate per fault which is assumed to be constant, increasing and decreasing. In this article we propose a software reliability growth model in which the time between two failures assumed to be Dagum distribution. The motivation to choose this distribution is the inadequacy of existing models to describe the nature of failure process.

II. FINITE FAILURE NHPP MODELS

Non Homogeneous Poisson Process group of models provides an analytical framework for describing the software failure process during testing. The main issue of NHPP models is to estimate $m(t)$, the expected number of faults experienced up to a certain point of time, which is also called mean value function. These models differ in mean value function $m(t)$. The NHPP models can be further classified into two categories, namely finite failure and infinite failure categories. Finite failure NHPP models assumes that the expected number of faults experienced detected given infinite amount of testing time will be finite, where as in Infinite failure models assume that the expected number of faults detected given infinite amount of testing time will be infinite. The parameters of the finite failure NHPP model are described as under.
If $N(t)$ be the cumulative number of faults detected by the time t , $F(t) = P[T \leq t]$ is the distribution function and a denote the expected number of faults that would be detected in a given infinite testing time, then the mean value function is given by

$$\boldsymbol{m(t) = E[N(t)] = aF(t)} \qquad (1)$$

The failure intensity function is given by

$$\boldsymbol{\lambda(t) = aF'(t)} \qquad (2)$$

The reliability function is given by

$$R(t) = P[T \geq t] = 1 - P[T \leq t] = 1 - F(t)$$

The hazard rate or failure occurrence rate per fault of the software is given by

$$\boldsymbol{h(t) = \frac{F'(t)}{1 - F(t)}} \qquad (3)$$

From (2) and (3) the failure intensity function can also expressed in terms of hazard rate as

$$\lambda(t) = aF'(t) = [a - m(t)]\frac{F'(t)}{1 - F(t)} = [a - m(t)]h(t)$$

here $h(t)$ is the rate at which individual faults manifest themselves during testing. The quantity $[a - m(t)]$ is the expected number of faults remaining in the software at the time t. It is monotonically non deceasing function of the time. Therefore the nature if the intensity $\lambda(t)$ is governed by the nature of the future occurrence rate per fault $h(t)$

III. MODELS WITH SEMI INFINITE INTERVAL DISTRIBUTION

In the NHPP model the time between two failures takes the value between zero and infinity and assumes a probability distribution. The probability distribution of these kinds is called semi infinite distribution. In this section we discuss existing models in which the time between failures follow a semi infinite distribution and we propose a model based on Dagum distribution.

In the Goel-Okumoto (GO) model (Goel and Okumoto [3]) the time between failure assumes exponential distribution consequently failure occurrence rate per fault of the software $h(t)$ is assumed constant b. In generalized GO model proposed by (Goel [2]) the time between failures assumes Weibull distribution in which the nature of the failure occurrence rate per fault is determined by the parameter γ. If $\gamma < 1$ then the failure occurrence rate per fault is increasing and if $\gamma > 1$ then the failure occurrence rate per fault is decreasing.

Harishchandra and Manjunath [4] developed a SRGM in which behavior of failure occurrence rate per fault i.e., hazard function is described by generalized inverse exponential distribution. The hazard rate function $h(t)$ of the generalized inverse exponential SRGM is given as

$$h(t) = \frac{\alpha b}{t^2}\left(\frac{e^{-b/t}}{1 - e^{-b/t}}\right) \qquad (5)$$

The corresponding mean value function $m(t)$ and failure intensity function $\lambda(t)$ are

$$m(t) = a\left[1 - \left(1 - e^{-bt}\right)^{\alpha}\right] \qquad (6)$$

$$\lambda(t) = \frac{a\alpha b}{t^2}e^{-bt}\left(1 - e^{-bt}\right)^{\alpha-1} \qquad (7)$$

The software reliability is the conditional probability that the i^{th} software failure does not occur between $(t, t + x], (x \geq 0)$ on the condition that the $(i-1)th$ software failure has occurred at testing time t, is given by

$$R(x|t) = exp\left\{a\left[\left(1 - e^{-b/(t+x)}\right)^{\alpha} - \left(1 - e^{-bt}\right)^{\alpha}\right]\right\} \qquad (8)$$

The expressions of $m(t)$, $\lambda(t)$ and $h(t)$ of the above are (presented in **Error! Reference source not found.** table 3.1

Table 3.1

Coverage Function	$m(t)$	$\lambda(t)$	$h(t)$
Goel Okumoto	$a(1 - e^{-bt})$	abe^{-bt}	b
Weibull	$a(1 - e^{-bt^{\gamma}})$	$ab\gamma t^{\gamma-1}e^{-bt^{\gamma}}$	$ab\gamma t^{\gamma-1}$
S-shaped	$a[1 - (1 + bt)e^{-bt}]$	ab^2te^{-bt}	$\frac{b^2t}{1+bt}$
GIED	$a[1 - (1 - e^{-bt})^{\alpha}]$	$\frac{a\alpha b}{t^2}e^{-bt}(1 - e^{-bt})^{\alpha-1}$	$\frac{\alpha b}{t^2}\left(\frac{e^{-b/t}}{1 - e^{-b/t}}\right)$

IV. DAGUM DISTRIBUTION MODEL

In this section we propose another finite failure non homogeneous Poisson process model based on the Dagum distribution in which failure occurrence rate per fault of the software $h(t)$ is depends on time. The model proposed here assumes Dagum distribution. If t is the time between two failures, then the probability density function of the Dagum distribution with parameters is given in (9)

$$f(t) = \frac{\theta p}{t}\left(\frac{\left(\frac{t}{b}\right)^{\theta p}}{\left[1 + \left(\frac{t}{b}\right)^{\theta}\right]^{p+1}}\right), t \geq 0,$$

$$here\ \theta > 0, p > 0\ called\ shape\ parameters\ and$$
$$b > 0\ scale\ parameter \qquad (9)$$

The corresponding distribution function, mean value function, failure intensity function and the hazard function of the software reliability growth model with Dagum distribution are given by (10)-(14).

$$F(t) = \left(1 + \left(\frac{t}{b}\right)^{-\theta}\right)^{-p}$$

$$m(t) = aF(t) = a\left(1 + \left(\frac{t}{b}\right)^{-\theta}\right)^{-p} \qquad (11)$$

$$\lambda(t) = aF'(t) = \frac{ap\theta}{b}\left(\frac{t}{b}\right)^{-\theta-1}\left(1 + \left(\frac{t}{b}\right)^{-\theta}\right)^{-p-1} \qquad (12)$$

$$h(t) = \frac{F'(t)}{1 - F(t)} = \frac{\frac{p\theta}{b}\left(\frac{t}{b}\right)^{-\theta-1}\left(1 + \left(\frac{t}{b}\right)^{-\theta}\right)^{-p-1}}{1 - \left(1 + \left(\frac{t}{b}\right)^{-\theta}\right)^{-p}} \qquad (13)$$

$$R(x|t) = exp\left\{a\left[\left(1 + \left(\frac{t+x}{b}\right)^{-\theta}\right)^{-p} - \left(1 + \left(\frac{t}{b}\right)^{-\theta}\right)^{-p}\right]\right\} \qquad (14)$$

V. PARAMETER ESTIMATION FOR INTERVAL DOMAIN DATA

The interval domain data are available in terms of cumulative number of faults y_i up to a specified time t_i or numbers of detected faults $x_i = y_i - y_{i-1}$ in a given interval (t_{i-1}, t_i). For interval domain data the likelihood function and log likelihood function are given in the following equations

$$L = \prod_{i=1}^{n} \frac{e^{-[m(t_i)-m(t_{i-1})]}[m(t_i) - m(t_{i-1})]^{}}{[y_i - y_{i-1}]!} \tag{15}$$

The log likelihood function is given by

$$lnL = \sum_{i=1}^{n} [y_i - y_{i-1}] ln[m(t_i) - m(t_{i-1})] \tag{16}$$

Substituting the mean value function given in (11) in the log likelihood function given in (16) we get the log likelihood function as

$$lnL = \sum_{i=1}^{n} (y_i - y_{i-1}) ln\left[a\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p} - a\left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p}\right] - a\Big(\tag{17}$$

This can also be presented as

$$lnL = -m(s_n) + \sum_{i=1}^{n} ln\lambda(s_i)$$

$$= -a\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)^{-p} + nlna + nlnp + nln\theta - n\theta lnb$$

$$-(p+1)\sum_{i=1}^{n} ln\left(1+\left(\frac{s_i}{b}\right)^{-\theta}\right) \tag{18}$$

Taking partial derivative of the above expression (18) with respect to a, b, θ and p and equating them to zero we get the system of equations, which are

$$a = \frac{y_n}{\left(1+\left(\frac{t_n}{b}\right)^{-\theta}\right)^{-p}},$$

$$pat_n\left(1+\left(\frac{t_n}{b}\right)^{-\theta}\right)^{-p-1}\left(\frac{t_n}{b}\right)^{-\theta-1}$$

$$= \sum_{i=1}^{n} \frac{(y_i - y_{i-1})p\theta\left[\left(\left(t_{i-1}\left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p-1}\left(\frac{t_{i-1}}{b}\right)^{-\theta-1}\right)\right) - \left(t_i\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p-1}\left(\frac{t_i}{b}\right.\right.\right.}{\left(\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p} - \left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p}\right)}$$

$$pa\left(\frac{t_n}{b}\right)^{-\theta}\left(1+\left(\frac{t_n}{b}\right)^{-\theta}\right)^{-p-1} ln\left(\frac{t_n}{b}\right)^{-1} \tag{19}$$

$$= \sum_{i=1}^{n} \frac{(y_i - y_{i-1})p\theta\left[\left(p\left(\frac{t_i}{b}\right)^{-\theta}\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p-1} ln\left(\frac{t_i}{b}\right)^{-1}\right) - \left(p\left(\frac{t_{i-1}}{b}\right)^{-\theta}\left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p-1} ln\left(\frac{t_{i-1}}{b}\right)\right.\right.}{\left(\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p} - \left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p}\right)}$$

and

$$a\left(1+\left(\frac{t_n}{b}\right)^{-\theta}\right)^{-p} ln\left(1+\left(\frac{t_n}{b}\right)^{-\theta}\right)$$

$$= \sum_{i=1}^{n} \frac{(y_i - y_{i-1})p\theta\left[\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p} ln\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right) - \left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p} ln\left(1+\left(\frac{t_{i-1}}{b}\right)^{-}\right.\right.}{\left(\left(1+\left(\frac{t_i}{b}\right)^{-\theta}\right)^{-p} - \left(1+\left(\frac{t_{i-1}}{b}\right)^{-\theta}\right)^{-p}\right)}$$

We can use numerical method to solve the above equations (19) to obtain MLEs a, b, θ and p

4.1.1. Parameter estimation for Time domain data

The time domain data are available in terms of time between $(i-1)^{th}$ and i^{th} failures t_i or in terms of time of occurrence of i^{th} failure $s_i = \sum_{j=0}^{i} t_j$, $t_0 = 0$. For time domain data the likelihood function and log likelihood function are given by

$$L = \prod_{i=1}^{n} \lambda(s_i) e^{-\int_{s_{i-1}}^{s_i} \lambda(x)} \tag{20}$$

$$lnL = -m(s_n) + \sum_{i=1}^{n} ln[\lambda(s_i)] \tag{21}$$

Substituting the mean value function and the failure intensity function given in (11) and (12) in the log likelihood function given in (21) we get the log likelihood function as

$$lnL = -m(s_n) + \sum_{i=1}^{n} ln\lambda(s_i)$$

$$= -a\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)^{-p} + nlna + nlnp + nln\theta - n\theta lnb - (\theta+1)\sum_{i=}^{i}$$

$$-(p+1)\sum_{i=1}^{n} ln\left(1+\left(\frac{s_i}{b}\right)^{-\theta}\right) \tag{22}$$

Taking partial derivative of the above expression (22) with respect to a, b, θ and p and equating them to zero we get the system of equations, which are given in (23)

$$a = \frac{n}{\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)^{-p}},$$

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 2, Issue 1, ISSN No. 2455-2143, Pages 116-120
Published Online November-December 2016 in IJEAST (http://www.ijeast.com)

$$pas_n\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)^{-p-1}\left(\frac{s_n}{b}\right)^{-\theta-1}=\frac{n\theta}{b}+(p+1)\theta b^{\theta+1}\sum_{i=1}^{n}\frac{(s_i)^{-\theta-1}}{\left(1+\left(\frac{s_i}{b}\right)^{-\theta}\right)}$$

$$pa\left(\frac{s_n}{b}\right)^{-\theta}\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)^{-p-1}ln\left(\frac{s_n}{b}\right)^{-1}=(p+1)\sum_{i=1}^{n}\frac{\left(\frac{s_i}{b}\right)^{-\theta}ln\left(\frac{s_i}{b}\right)^{-1}}{\left(1+\left(\frac{s_i}{b}\right)^{-\theta}\right)}+nlnb-\frac{n}{\theta}+\sum_{i=1} s_i$$

and

$$a\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)^{-p}ln\left(1+\left(\frac{s_n}{b}\right)^{-\theta}\right)=\frac{n}{p}+\sum_{i=1}^{n}ln\left(1+\left(\frac{s_i}{b}\right)^{-\theta}\right)$$

…..(23)

We can use numerical method to solve the above equations to obtain MLEs a, b, θ *and* p

4.2. Numerical Analysis

We now present a numerical example for finite failure Software Reliability Growth models based on the actual testing-data. Here we considered two types of data one is interval domain data and the other one is time domain data

The interval domain data was reported by Musa [6] based on failure data from a real-time command and control system, which represents the failures observed during system testing for 25 hours of CPU time. The delivered number of object instructions for this system was 21700 and was developed by Bell Laboratories. The estimates of the parameters along with of the above model using the this interval domain data are given in the following table 4.1

Table 4.1

Model	Estimates of Parameters				AIC
G-O	$a = 142.31$	$b = 00.124$			273.283
Weibull	$a = 164.82$	$b = 00.182$	$\gamma = 0.701$		283.818
GIED	$a = 150.00$	$b = 01.300$	$\alpha = 0.491$		259.274
Dagum	$a = 149.99$	$b = 01.202$	$b = 0.717$	$b = 2.589$	270.010

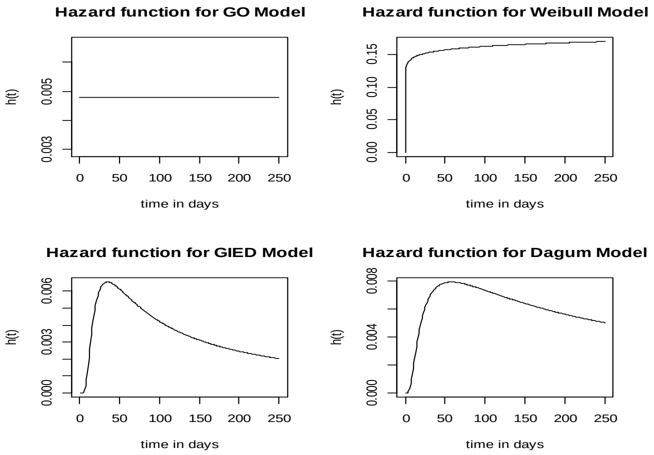
The time domain data was extracted from information about failures in the development of software for real-time multi computer complex of the US Naval Fleet Computer Center of the US Naval Tactical Data System (NTDS) (Goel and Okumoto [3]). The data consists of 26 software failure occurrence time $\{s_i(days); i = 1,2,\dots,26\}$ out of 250 days. Here we obtain the estimation results and AIC value for existing and our proposed models The output are summarized in the table 4.2

Table 4.2

Model	Estimates of Parameters				AIC
G-O	$a = 35.00$	$b = 00.005$			170.012
Weibull	$a = 39.97$	$b = 00.003$	$\gamma = 1.048$		172.262
GIED	$a = 38.16$	$b = 56.015$	$\alpha = 0.566$		170.377
Dagum	$a = 34.97$	$b = 03.137$	$\theta = 0.815$	$p = 16.391$	171.256

Using the above estimates of the parameters, the graph of hazard function of the existing and our proposed model based on log logistic distribution are plotted against the time in days. The graphs are shown below in figure 4.1

Figure 4.1



VI. CONCLUSIONS

In the NHPP it is assumed that the failure rate assumed to follow exponential distribution In this chapter, we have proposed finite failure SRGM based on various distribution because of the fact that the existing finite failure NHPP models were inadequate to describe failure pattern. We use the mean value function $m(t)$ and failure intensity function $\lambda(t)$ to determine the likelihood function for both interval domain data and time domain data. The estimates of the parameters are obtained by the method of maximum likelihood.

The models are verified numerically using two real time data sets, namely interval domain data and time domain data.

From the table 4.1 the AIC value for Weibull model is less than the other models. Therefore for interval domain data Weibull model fit better compare to other model.

From the table 4.2 the AIC values for Dagum distribution are less than the other models. Therefore for time domain data Dagum model fits better compare to the other model.

We also plotted the graph of the hazard function for time domain data for the models discussed in this chapter. Among them GO model shows constant hazard function. For Weibull model it is increasing then at certain point of time and onwards it is almost constant. The Raleigh model shows increasing trend. Whereas the GIED, Dagum and Log logistic model shows increasing/decreasing behavior of the failure occurrence rate per fault.

VII. REFERENCES

[1]. Abouammoh A. M, and A. M. Alshingiti.,(2009) '*Reliability estimation of generalized inverted exponential distribution',* J. Stat. Comput. Simulation. 79(11):1301-1315.

[2]. Goel, A. L.,(1985): "*Software Reliability Models: Assumptions, Limitations and Applicability",* IEEE Trans. Software Engineering; SE-11, pp. 1411-1423

[3]. Goel, A. L., and Okumoto. K,. (1979):'*Time-dependent error detection rate model for software reliability and other performance, Measure',* IEEE Trans. Reliability R-28(3): 206-211

[4]. Harishchandra K and Manjunath K.M., (2012): '*Modeling and statistical Inference on Generalized Inverse Exponential Software Reliability Growth Model*', Far East Journal of Theoretical Statistics 39(1) 67-77

[5]. Lyu. M. R., (1996): '*Handbook of Software Reliability Engineering*'. McGraw Hill,

[6]. Musa J. D and Okumoto K., (1982): *'Software reliability models: Concepts classification, comparisons, and practice'*, in: Skwirzynski J.K (ed) Electronic Systems Effectiveness and Life cycle Costing. Springer, New York, 395-423

[7]. Musa J. D, Iannino A, and Okumoto K., (1987): '*Software Reliability: Measurement, Prediction, Application'*, McGraw-Hill, Inc., USA

[8]. Yamada. S, Ohba. M and Osoki. S., (1983)'*S shaped reliability growth modeling for software error detection',* IEEE Trans. Reliability R-32(5):475-484

[9]. Zhang X, Pham H., (2000) *'Comparisons of non homogeneous Poisson process software reliability models and its applications',* International Journal of Systems 31(9):1115- 1123