A NEW SOFTWARE RELIABILITY GROWTH MODEL

R.R.L.Kantam[1], V.Ramakrishna[2], M.S.Ravikumar[3]

Abstract: The proposed model is adopted as a mean value function of a Non-Homogeneous Poisson Process (NHPP) to define a Software Reliability Growth Model (SRGM). The model is approximated for software failure data to measure the software reliability. The suitability of the model is also discussed.

Keywords: Parallel-Series Configurations, HLD, NHPP, SRGM.

1. INTRODUCTION

Computer software is aimed to instruct the computer's hardware in performing the necessary computations. An important goal is to know how to initially estimate and subsequently measure quality, reliability of developed software. For a software product also, the methodologies of reliability can be applied. Since softwares are intellectual creation, generally it is difficult to measure the software quality. The only way for a software quality is to identify systematically the opportunities of improvement at every stage of software development. The associated preferability of a software product may be known as software reliability. The imperfectness in software can be known by the existing discrepancy between what it can do versus what the computing environment wants it to do. These discrepancies are known as software faults. Even if we know that the software contains faults, we do not know generally their exact identity. The available two approaches for indicating software faults are program processing and program testing. Possible imperfectness of these approaches induces the need of a metric that reflects the degree of program correctness. One such metric in software engineering practice is software reliability.

A commonly used approach for measuring software reliability is via an analytical model whose parameters are generally estimated from available data on software failures. Reliability and other relevant measures are computed from the fitted model. In a way, software reliability is a probability measure and is defined as the probability that software faults do not cause a failure during a specified exposure period in a specified environment. It is useful to give the user, confidence about software correctness. A number of analytical models proposed to address the problem of software reliability measurement can be classified into four main groups according to the nature of failure process studied. One among them is failure count models. In this group the number of experienced software failures within a given time period is a random process- time dependent, stochastic process. Since faults are removed from the system as and when failures are experienced, it is expected that the observed failures per unit time (known as failure intensity) will decrease. The basic idea behind most of the failure count models is that of a Poisson distribution whose parameter takes different forms for different models. Letting N(t) be the cumulative number of failures observed by time t, N(t) can be modeled as a Poisson process with a time dependent failure rate.

ISBN - 978-93-81583-55-5

$$P[N(t)=y]=\frac{[m(t)]^{y}e^{-m(t)}}{y!},\ y=0,1,2,......\quad .\ (1.1)$$

Here m(t) is called mean value function and its derivative is called the failure intensity function μ(t). If the limiting value of N (t) is a finite constant that limit is called eventual expected number of failures of the system. If F(t) is the cdf of a positive valued continuous random variable, ‘a’ is eventual number of failures, generally ‘a.F(t)’ is taken as the mean value function of failure count Poisson process. In the above Poisson mass function all such models are termed as non-homogeneous Poisson processes (NHPP) measuring software reliability generally known as ‘software reliability growth model’(SRGM). In literature a number of models exist under this category as can be known from [3], [4] and [5]. In this paper we adopt a generalized HLD to define mean value function of NHPP in order to study software reliability growth. The rest of the paper is organized as follows. Estimation of parameters of the corresponding NHPP from failure count data is presented in Section 2. Goodness of fit of model for data sets is discussed in Section 3.

2. SOFTWARE RELIABILITY GROWTH MODEL (SRGM) WITH TYPE-II GHLD

Consider the non-homogeneous Poisson process given by refer to (1.1) with mean value function given by

$$m(t)=a\left[1-\left(\frac{2e^{-bt}}{1+e^{-bt}}\right)\right]^{\theta},$$

where ‘a’ is limiting value of m (t) known as expected number of software failures eventually, b = 1/σ. If $t_1<t_2<t_3<\ldots\ldots\ldots.t_k$ are time instants at which cumulative number of experienced failures are recorded as $n_1,n_2,\ldots\ldots\ldots.n_k,$ then the likelihood function to estimate the parameters of the SRGM is given by

$$L=\prod_{i=1}^{n}\frac{e^{-[m(t_i)-m(t_{i-1})]}\left[m(t_i)-m(t_{i-1})\right]^{n_i-n_{i-1}}}{(n_i-n_{i-1})!},$$

$$\log L=\sum_{i=1}^{k}-\left[m(t_i)-m(t_{i-1})\right]+\sum_{i=1}^{k}(n_i-n_{i-1})\log\left[m(t_i)-m(t_{i-1})\right]$$
$$-\sum_{i=1}^{k}\log(n_i-n_{i-1})!\quad .$$

The likelihood equations to estimate the parameters a and b after simplification give the estimates as

ISBN - 978-93-81583-55-5

$$\hat{a} = \frac{\sum_{i=1}^{k} (n_i - n_{i-1})}{\sum_{i=1}^{k} \left[\left(\frac{2e^{-bt_i}}{1+e^{-bt_i}} \right)^{\theta} - \left(\frac{2e^{-bt_{i-1}}}{1+e^{-bt_{i-1}}} \right)^{\theta} \right]}, \tag{2.1}$$

and $\hat{b}$ as the solution of the equation

$$\frac{\partial \log L}{\partial b} = -a\theta \sum_{i=1}^{k} \left[\left(\frac{1-e^{-bt_i}}{1+e^{-bt_i}} \right)^{\theta-1} \frac{2t_i e^{-bt_i}}{(1+e^{-bt_i})^2} - \left(\frac{1-e^{-bt_{i-1}}}{1+e^{-bt_{i-1}}} \right)^{\theta-1} \frac{2t_{i-1} e^{-bt_{i-1}}}{(1+e^{-bt_{i-1}})^2} \right]$$

$$+\theta \sum_{i=1}^{k} \frac{(n_i - n_{i-1}) \left[\left(\frac{2e^{-bt_i}}{1+e^{-bt_{i-1}}} \right)^{\theta-1} \frac{2t_i e^{-bt_{i-1}}}{(1+e^{-bt_{i-1}})^2} - \left(\frac{1-e^{-bt_i}}{1+e^{-bt_i}} \right)^{\theta-1} \frac{2t_{i-1} e^{-bt_i}}{(1+e^{-bt_i})^2} \right]}{\left[\left(\frac{2e^{-bt_i}}{1+e^{-bt_{i-1}}} \right)^{\theta} - \left(\frac{2e^{-bt_i}}{1+e^{-bt_i}} \right)^{\theta} \right]} = 0. \tag{2.2}$$

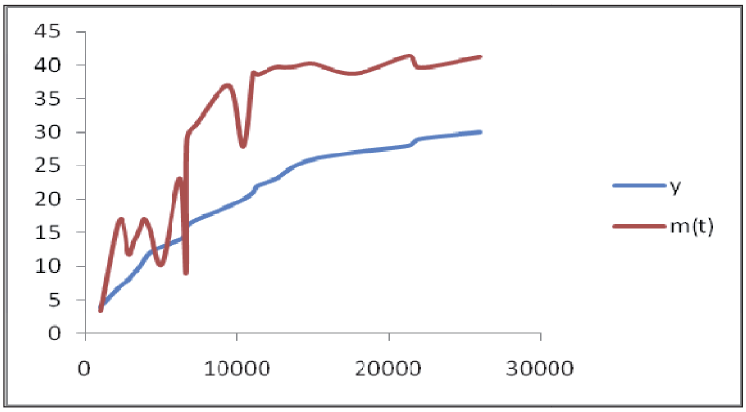
As a matter of illustration we have applied (2.1) and (2.2) to three Data Sets borrowed from [1]. In fact these are browsed from www.dacs.dtic.mul/databases/sled/swrel.hts. These data sets are Software Reliability Data for Project 2, Project 5 and Project 17.The results are shown in the Table 2.1.

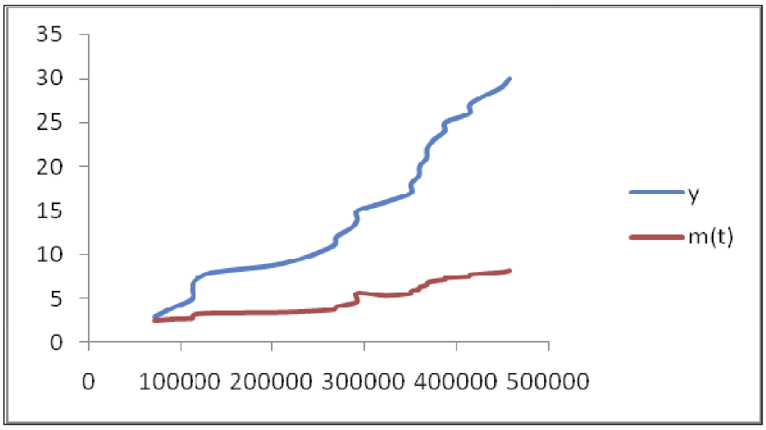
Table 2.1: Estimated values of a and b in the GHLD-II based SRGM

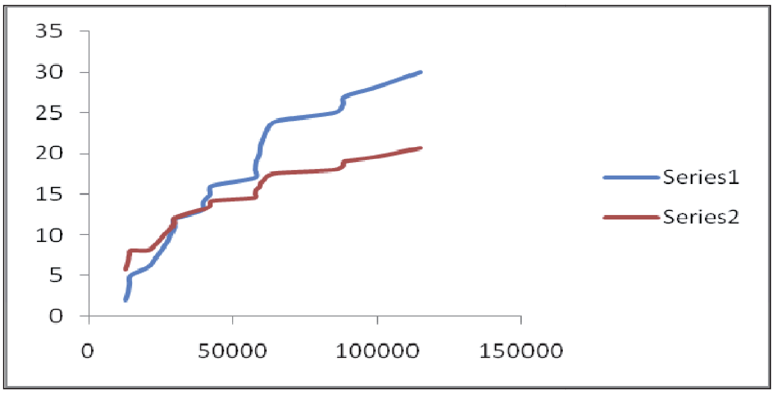
Project Name	**GHLD – II**			
	θ = 2		**θ = 3**	
	a	**b**	**a**	**b**
Project 2	41.2820	0.000281	41.2820	0.00028
Project 5	8.3662	0.000282	8.3663	0.000281
Project 17	26.3026	0.000281	26.3026	0.000281

Using these estimates of a, and b, the observed and expected cumulative number of failures are graphically shown for GHLD-II in the following Figures of 2.1.

Figure 2.1
Project 2, Project 5 and Project 17 for GHLD-II with
$\theta = 2$

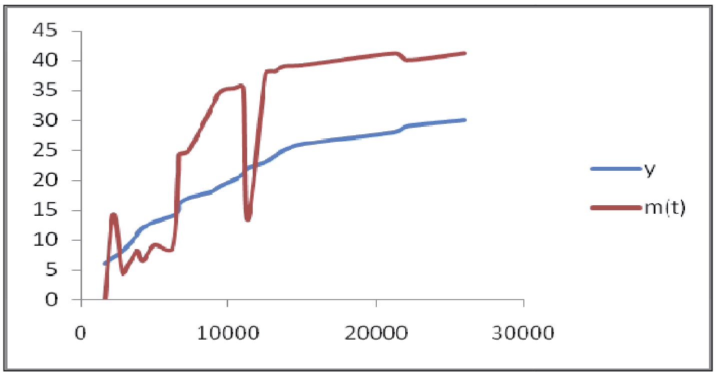


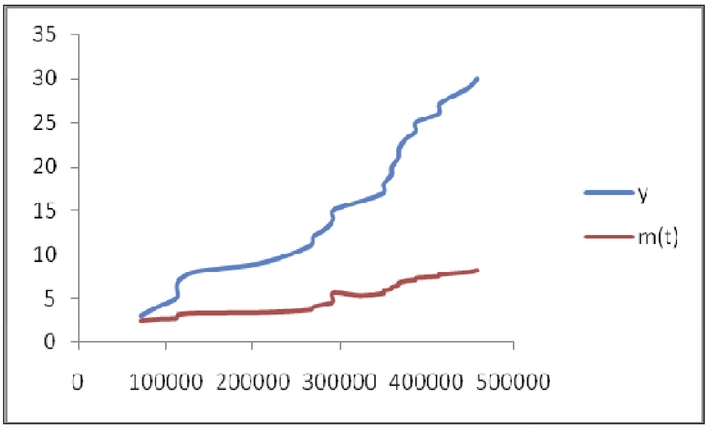


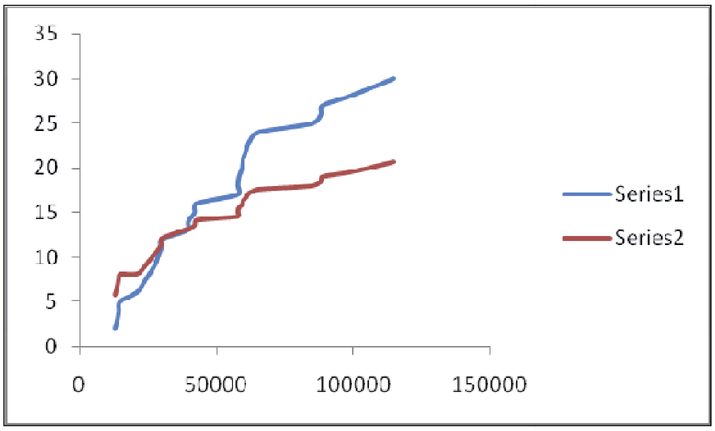


ISBN - 978-93-81583-55-5

Figure 2.1 (Continued)
Project 2, Project 5 and Project 17 for GHLD-II with
$\theta = 3$







Though there is some overlapping between observed and expected values of m(t) for project 2 data set, in the rest of the data sets for the rest of the models, m(t) generally over estimates the observed number of failures.

Sometimes the conditional probability of failure free operation of a software given it is tested for a specified period of time is called software reliability and is calculated from the following formula,

$$R(x/t) = e^{-[m(t+x)-m(t)]},$$

where m(t) is the mean value function and t is total time duration for which the software is already executed and x indicates the lapse of time till the occurrence of next failure. For any dataset, this can be calculated provided the estimates of the parameters of the mean value function are available.

3. GOODNESS OF FIT

The proposed SRGM can be compared for their adoptability to a live data with a goodness of fit criterion. Generally the goodness of fit is measured with the well-known Kolmogrov-Smirnov (K-S) test statistic defined as follows. The same can be used as a criterion of preferability also between the models.If $X_1, X_2,$ ------, X_n is a random sample of size n from a distribution function F(x) and $F_n^*(x)$ is the corresponding empirical distribution function, then $D_n = \text{Sup}_x \left| F_n^*(x) - F(x) \right|$ .D_n is called the K-S statistic and it is the maximum vertical distance between $F_n^*(x)$ and F(x). Goodness of fit is achieved when $D_n \leq$ the critical value as per the K-S table. We have calculated $F_n^*(x)$ and F(x) for the Type-II GHLD applying to three live data sets given in [1], and obtained the following values of the D_n statistic. In this process we have divided the data sets into 5 parts approximately at 20%, 40%, 60%, 80% and 100% of the cumulative test times. In each portion, the cumulative number of failures observed is also noted down from the data sets. The D_n statistic is evaluated by estimating the model parameters with the 5 parts of the data set separately. The values of D_n are given in the Table 3.1. The critical value for K-S statistic (n=5, α = 0.05) is 0.5632.

Table 3.1
Calculated Kolmogrov – SmirnovD$_n$ Statistic Values

Project Name	**GHLD - II**	
	θ = 2	**θ = 3**
Project 2	0.469	0.334
Project 5	0.79	0.79
Project 17	0.79	0.79

ISBN - 978-93-81583-55-5

Form this table we see that on the basis of D_n statistic, Type—II GHLD is suitable model for the data sets of Projects 5, and 17.

4. REFERENCES

1. Grottke M., and Trivedi K.S., "Truncated Non-Homogeneous Poisson Process Models- Properties and Performance",Opsearch, Vol.42 (4), pp: 310-321, 2005.
2. Kantam R.R.L., Satya Prasad R., and Ramakrishna V., "Software Reliability- Generalized Half Logistic GrowthModel", International Journal of Systems and Technology, Vol.2 (2), pp: 217-222, 2009.
3. Musa J.D., "Software Reliability Engineering",TataMcGraw-Hill, New Delhi,2004.
4. Pham H., "A Generalized Logistic Software ReliabilityGrowth Model", Opsearch, Vol.42 (4),pp: 322-331, 2005.
5. Satya Prasad R., "Half Logistic Software Reliability Growth Model", Ph.D Thesis, Acharya Nagarjuna University, India, 2007.
6. Satya Prasad R., and Kantam R.R.L., "Half Logistic Software Reliability Growth Model",International Journal of Advanced Research in Computer Science, Vol.1 (2), pp: 31-36, 2010.

****

----

[1]*Author 1: Department of Statistics, Acharya Nagarjuna University,*
Nagarjunanagar-522510, India.
kantam.rrl@gmail.com
[2]*Author 2: Department of Computer Science and Engineering,*
KoneruLakshmayyaUniversity, Vaddeswaram-522502, India.
vramakrishna2006@gmail.com
[3]*Author 3: Department of Statistics, Acharya Nagarjuna University,*
Nagarjunanagar-522510, India.
msrk.raama@gmail.com