CHARACTERIZATION OF POISSON TYPE LENGTH BIASED EXPONENTIAL CLASS SOFTWARE RELIABILITY GROWTH MODEL AND PARAMETER ESTIMATION

Rajesh Singh[1], Kailash R. Kale[2] , Pritee Singh[3]

1. R. T. M. Nagpur University, Nagpur-440033,
rsinghamt@hotmail.com
2. G. N. A. ACS College, Barshitakli, Dist-Akola.
kailashkale10@gmail.com
3. Institute of Science, Nagpur;
priteesingh25@gmail.com

Abstract

The authors of this study set out to build a software reliability growth model (SRGM). Software reliability is a crucial attribute that has to be quantified and evaluated. In most cases, software errors happen at unpredictable times. In this article, the failure intensity of the single parameter length-biased exponential class SRGM has been characterized taking into account the Poisson process of the incidence of software faults. The parameters of the proposed SRGM under investigation are the scale parameter (θ_1) and the total number of failures (θ_0). It is considered that the experimenter may have previous knowledge of the parameters from past or earlier experiences in the form of gamma priors. The posterior probability may be obtained by combining the prior probability with the likelihood of the data, and Bayes estimators can then be suggested.

Keywords: Binomial process, gamma prior, maximum likelihood estimator (MLE), Rayleigh class, software reliability growth model (SRGM), incomplete gamma function, confluent hyper-geometric function.

I. Introduction

Beginning in the early 1970s of the previous century, research on software reliability has advanced until the present. Various kinds of software have dominated many fields of the humanities, sciences, and technology as well as daily life for all people. The film industry, education sector, E-commerce sector, medical and healthcare sectors, space agencies, banking sector and various government agencies, all employ different types of software for the convenience and development the fields.

The software is the end product of several intricate code sequences developed by the humans according to the needs of above sectors within stipulated time period. Due to huge magnitude of complicated code sequences, there is a greater likelihood of failures or ineffective performance. These software failures may result from a variety of issues, including memory faults, language-specific issues, calling third-party libraries, standard library issues, etc. Such flaws may have operational repercussions that cause system failure and unanticipated dangerous outcomes.

As a result, the aforementioned areas require software that operates reliably. Hence, the evaluation and quantification of the software's performance are therefore crucial. The reliability of the software is one of the performance indicators. To put it another way, it becomes crucial to create reliable software that serves the needs of users or systems.

Software reliability is now thought to be a crucial factor in determining customer satisfaction, along with software functionality and performance. Software reliability growth models (SRGM) outline the broad link between software failure occurrences and the key process influences (such as fault introduction, fault removal, operational profile, etc.). The statistical relation between data on defects and the known characteristics of probabilistic behavior is known as the SRGM. The basic goal of software reliability modeling is to represent a relationship in which, when defects are found and removed, there is a reduction in the number of failures per time interval or an increase in the time interval between failures. The SRGM is often characterized by the mean failure function or failure intensity function. The pattern of occurrence of software failure is its type, and the mathematical functional form of failure intensity is its class. The software reliability growth models are categorized according to the system described by [8].

The length-biased distributions have been presented by [2] and formalized by [9]. These distributions are sometimes referred to as size-biased probability distributions. Reliability theory may also use these distributions (see [4], [5], and [6]). Modeling software reliability may be done using length-biased distributions. In this study, using the Poisson pattern of occurrence of software failure and the length-biased exponential form of failure intensity, the Poisson type length-biased exponential class model is introduced as per the classification system provided by [8]. As this SRGM is being described, it is assumed that the failure occurring at time t has a Poisson occurrence (i.e., Type) and that the mean failure function's functional form is characterized by a length-biased exponential distribution (i.e., Class). The software failures in this model are presumed to be independent of one another and dependent on the duration of the time interval that comprises the same software failure. For the estimation part of parameters the gamma priors taking into account. The Bayes estimators of the parameters are obtained in this study by the methods of [7], [12], [10], and [11], and they are compared with MLEs in subsequent parts.

II. Model Section

Suppose time to failure follows length biased exponential distribution denoted by $f(t)$ with scale parameter θ_1 and software failures occur in Poisson pattern then

$$f(t) = \begin{cases} t\theta_1^2 e^{-\theta_1 t} \, ; \; t > 0, \theta_1 > 0, E[t] \neq 0 \\ 0 \qquad ; \qquad otherwise \end{cases} \tag{1}$$

Also let the total number of faults remaining in the software at time $t = 0$ is a Poisson random variate with mean θ_0 then the failure intensity $\lambda(t) = \theta_0 f(t)$ (cf. [7]) can be obtained as

$$\lambda(t) = \theta_0 t \theta_1^2 e^{-\theta_1 t}\,; \qquad t > 0,\, \theta_1 > 0,\, \theta_0 > 0 \qquad \text{and } E[t] \neq 0 \tag{2}$$

The mean failures function at time t comes out to be

$$\mu(t) = \theta_0 \left[1 - (1 + \theta_1 t) e^{-\theta_1 t}\right] \tag{3}$$

The details about number of failures experienced by time t, performance of failure intensity $\lambda(t)$ and $\mu(t)$ have been discussed in [13].

III. Maximum Likelihood Estimation

The most important and extensively used technique of point estimation is maximum likelihood estimation when underlying distribution of data is known. The maximum likelihood estimation is considered for failure times. The base of maximum likelihood estimation is likelihood function

which can be obtained by assuming that m_e failures are experienced at times $t_i, i = 1,2, \dots, m_e$ up to execution time is t_e ($\geq t_{m_e}$). Also using the failure intensity at each $t_i, i = 1,2, \dots, m_e$ obtained in (2) and mean failure function at time t_e obtained by replacing $t = t_e$, the likelihood function of θ_0 and θ_1 can be obtained as $L(\theta_0, \theta_1) = \{\prod_{i=1}^{m_e} \lambda(t_i)\} exp\ (-\mu(t_e))$ (cf. Musa et al. (1987)).
The $L(\theta_0, \theta_1)$ can take following form for this model

$$L(\theta_0, \theta_1) = \theta_0^{m_e}\theta_1^{2m_e}\left[\prod_{i=1}^{m_e} t_i\right] e^{-T\theta_1}e^{-\theta_0\left[1-(1+\theta_1 t_e)e^{-\theta_1 t_e}\right]} \tag{4}$$

where
$\sum_{i=1}^{m_e} t_i = T$
The Maximum Likelihood Estimators for the parameters θ_0 and θ_1 are

$$\hat{\theta}_{m0} = m_e\left(1-(1+\hat{\theta}_{m1}t_e)e^{-\hat{\theta}_{m1}t_e}\right)^{-1} \tag{5}$$

and

$$\hat{\theta}_{m1} = \left[\hat{\theta}_{m0}^{-1}t_e^{-2}(2m_e - T\theta_{m1})e^{\hat{\theta}_{m1}t_e}\right]^{1/2} \tag{6}$$

respectively. The values of $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ can be obtained after simultaneous solution of equations (5) and (6).

IV. Bayesian parameter estimation

The bayesian technique is used to put the subjective and objective data sources together into the analysis. In this technique the parameters are considered as a random variables having known probability pattern. This known probability pattern is termed as prior in Bayesian technique. Whole the analysis is based on this prior and using Bayes theorem combines this prior and likelihood of data. In present case, it is considered that the experimenter have prior information about both the parameters θ_0 and $\theta_1$in the form of gamma probability function. Then the following prior distributions $g(\theta_0)$ and $g(\theta_1)$ can be considered for parameters θ_0 and θ_1 respectively.

$$g(\theta_0) \propto \begin{cases} \theta_0^{\nu-1}e^{-\eta\theta_0} & ,\theta_0 \epsilon\ [0,\infty) \\ 0 & ,otherwise \end{cases} \tag{7}$$

and

$$g(\theta_1) \propto \begin{cases} \theta_1^{\alpha-1}e^{-\beta\theta_1} & ,\theta_1 \epsilon\ [0,\infty) \\ 0 & ,otherwise \end{cases} \tag{8}$$

Now, Consider the total execution time is t_e and during this time m_e failures are experienced at times t_i, $i = 1,2, \dots, m_e$ then, the joint posterior of θ_0 and $\theta_1$given $\underline{t}$($= t_i$, $i = 1,2, \dots, m_e$) is

$$\pi(\theta_0, \theta_1|\underline{t}) \propto \theta_0^{m_e+\nu-1}\theta_1^{2m_e+\alpha-1}e^{-(T+\beta)\theta_1}e^{-(\eta+1)\theta_0}\ e^{\left[\theta_0(1+\theta_1 t_e)e^{-\theta_1 t_e}\right]}\ \theta_0 > m_e\ , \theta_1 > 0 \tag{9}$$

In this section, the point estimates (posterior mean) of both the parameters θ_0 and θ_1 under study are obtained by Bayesian technique considering the squared error loss as

$$\hat{\theta}_{B0} = D^{-1}\sum_{j=0}^{\infty}\frac{\Gamma(m_e+\nu+j+1,(\eta+1)m_e)}{j!(\eta+1)^{(j+1)}}\ \Psi(2m_e+\alpha, 2m_e+\alpha+j+1, T^*t_e^{-1}) \tag{10}$$

and

$$\hat{\theta}_{B1} = \frac{(2m_e+\alpha)}{Dt_e}\sum_{j=0}^{\infty}\frac{\Gamma(m_e+\nu+j,(\eta+1)m_e)}{j!(\eta+1)^{j}}\ \Psi(2m_e+\alpha+1, 2m_e+\alpha+j+2, T^*t_e^{-1}) \tag{11}$$

where $\Psi(\alpha, \beta; x)$ is Confluent Hypergeometric Function (cf. [1] and [3]), normalizing constant is
$D = \sum_{j=0}^{\infty}\frac{\Gamma(m_e+\nu+j,(\eta+1)m_e)}{j!(\eta+1)^{j}}\ \Psi(2m_e+\alpha, 2m_e+\alpha+j+1, T^*t_e^{-1})$
and
$T^* = T+\beta+jt_e$.

V. Discussion

I. Subsection One

The proposed Bayes estimators i.e. $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ are compared with corresponding maximum likelihood estimators i.e. $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ respectively on the basis of risk efficiencies $RE_j = R'_j R_j^{-1}$ where $R_j = E\left[\hat{\theta}_{Bj} - \theta_j\right]^2$ and $R'_j = E\left[\hat{\theta}_{mj} - \theta_j\right]^2$; $j = 0,1$. Here, the performance of proposed Bayes estimators $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ over MLEs $\hat{\theta}_{m0}$ and $\hat{\theta}_{m1}$ have been compared on the basis of risks efficiencies using Monte Carlo simulation technique. The risks efficiencies are obtained by generating sample of size, say m_e failures upto total execution time t_e and it was repeated 10^3 times from the length biased exponential distribution. Then, using Monte Carlo simulation technique risks efficiencies has been evaluated and is presented in the graphs Figure 1 to 9.

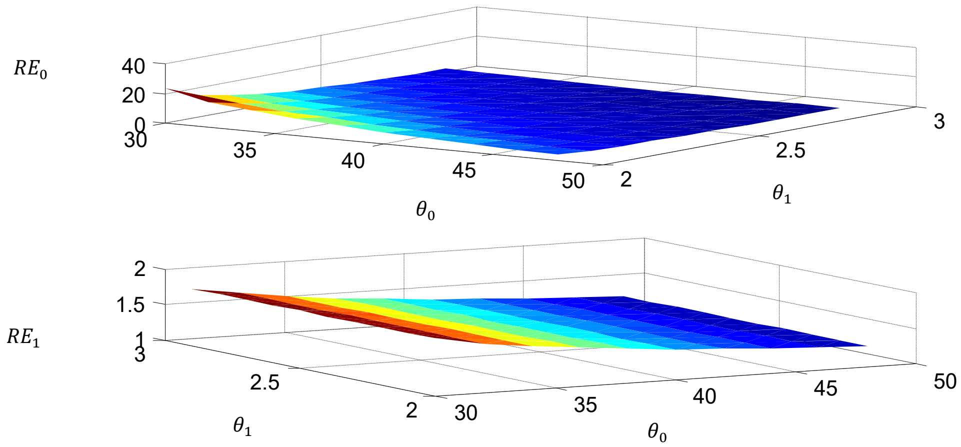

Figure 1: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ for t_e = 100; ϑ = 1, η = 1; α = 1, β = 1,*

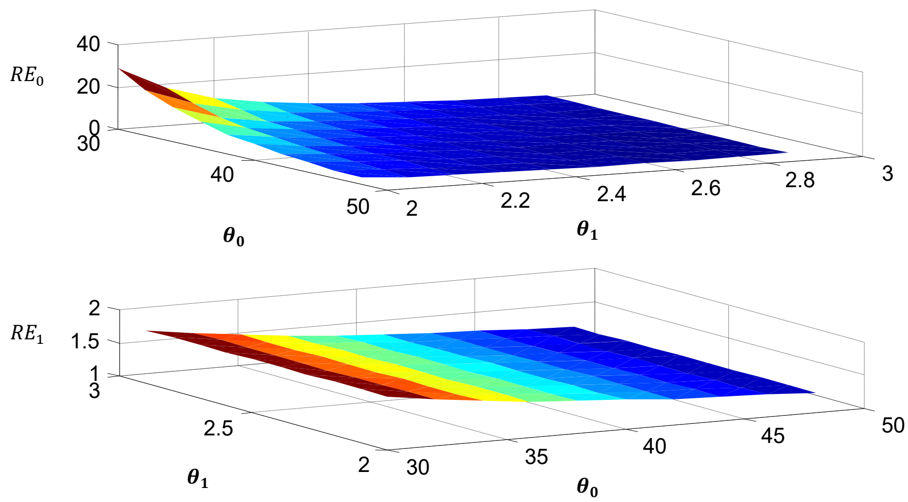

Figure 2: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 125; ϑ = 1, η = 1; α = 1, β = 1*

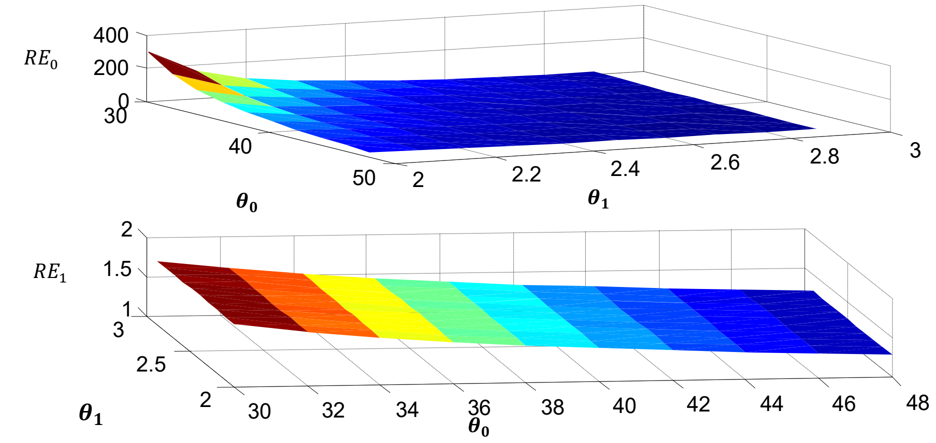

Figure 3: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 150; ϑ = 1, η = 1; α = 1, β = 1,*

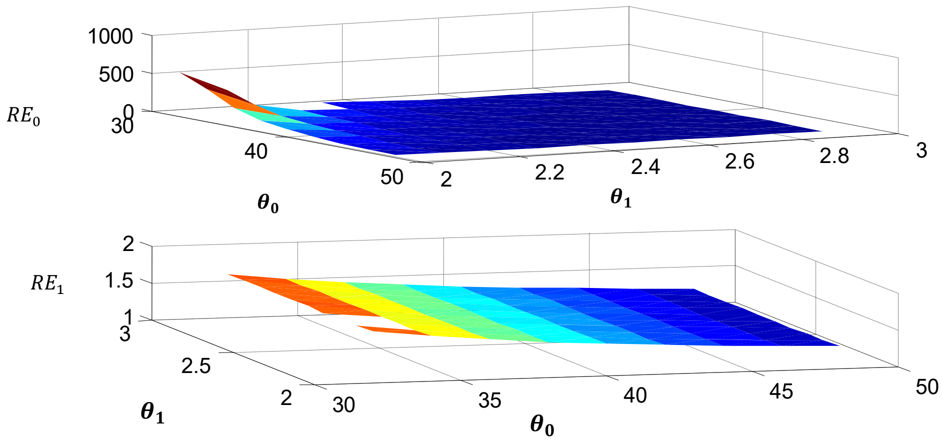

Figure 4: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 200; ϑ = 1, η = 1; α = 1, β = 1*

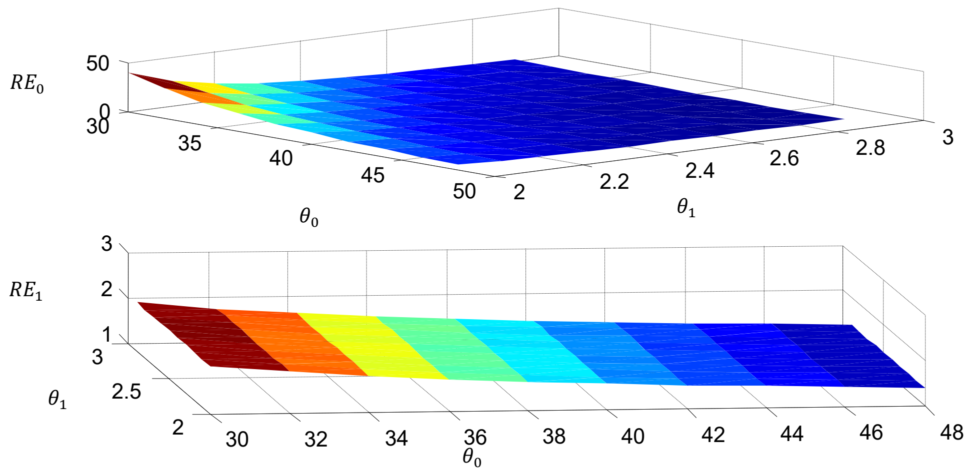

Figure 5: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 100; ϑ = 10, η = 1; α = 10, β = 1,*

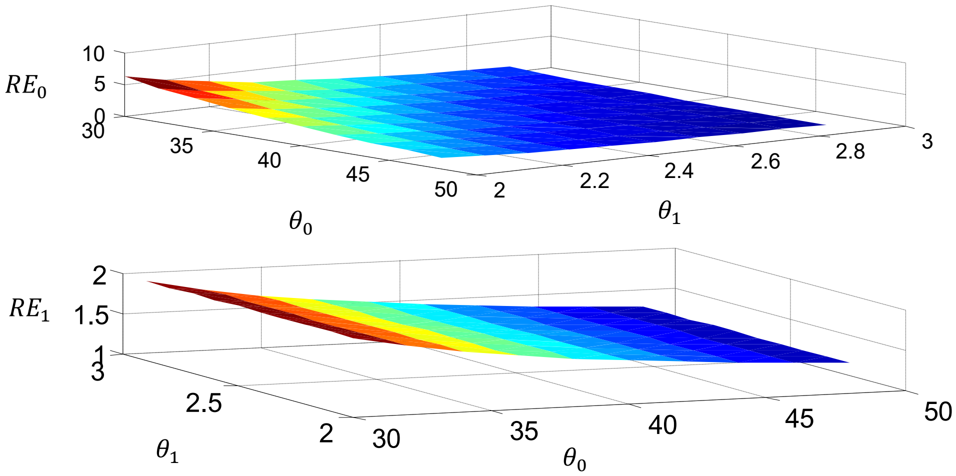

Figure 6: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 100; ϑ = 10, η = 5; α = 10, β = 5,*

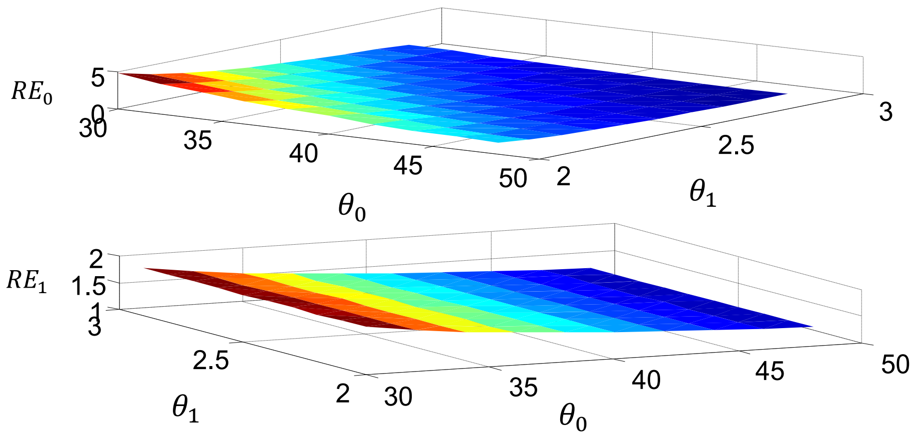

Figure 7: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 100; ϑ = 10, η = 10; α = 10, β = 10,*

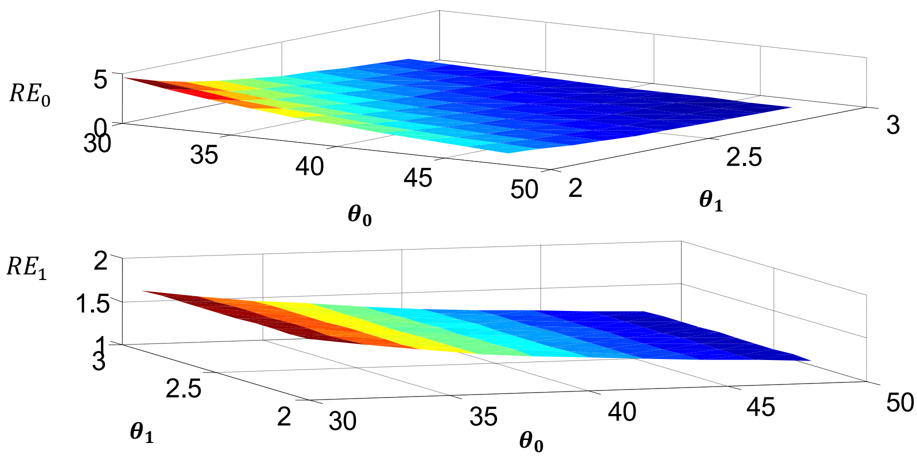

Figure 8: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ t_e = 100; ϑ = 1, η = 10; α = 1, β = 10,*

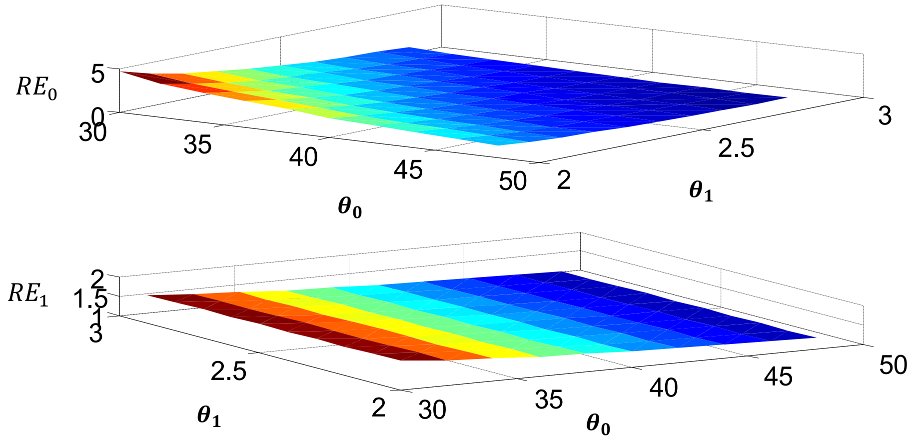

Figure 9: *Risk Efficiencies $\hat{\theta}_{B0}$ and $\hat{\theta}_B$ t_e = 100; ϑ = 3, η = 10; α = 3, β = 10,*

II. Subsection Two

Based on the above graphical representation from Figure 1 to Figure 9 of performance of proposed Bayes estimators against their corresponding MLE, it can be seen that the risk efficiencies RE_0 of $\hat{\theta}_{B0}$ decrease as θ_0 and θ_1 increase. It can also be seen that for large values of θ_1 and θ_0 the proposed Bayes estimator of θ_0 is not better than MLE otherwise proposed Bayes estimator $\hat{\theta}_{B0}$ is better than MLE. Moreover, when the value of t_e is small, the values of RE_0 first increase, attain a maxima and then decrease as the value of t_e increase. Similarly, it can be observed that the risk efficiencies of $\hat{\theta}_{B1}$ i.e. RE_1 decrease as the value of θ_1 and $\theta_0$increase but the values of risk efficiencies RE_1 are almost constant for the increase in values of θ_0. Further, The values of RE_1 are uniform over the variation in value of t_e. It is important to note that the proposed Bayes estimator $\hat{\theta}_{B1}$ is always better than MLE. Due to increase in values of shape and scale parameter of both the priors the values of RE_0 decrease for constant values of scale parameters.

On the basis of better performance of risk efficiencies of $\hat{\theta}_{B0}$ and $\hat{\theta}_{B1}$ over $\hat{\theta}_{m0}$ and $\hat{\theta}_{m0}$ following conclusions are drawn.

VI. Conclusions

After having experience or prior knowledge about the software failure process to researchers. These proposed Bayes estimators can perform better than their corresponding MLEs for the proper choices of prior parameters. The proposed Bayes estimator of θ_0 can be preferred over MLE if it is felt that total number of failures may not be very large and failure rate may be small. The proposed Bayes estimator of θ_1 can be preferred over MLE. Under this prior belief these estimators can be preferred for calendar time modeling.

References

[1] Abramowitz, M. and Stegun, I. A. "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", New York: Dover publications, 1965.

[2] Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies," *Ann. Eugenics*, 6:13-25.

[3] Gradshteyn, I. S. and Ryzhik, I. M. Table of Integrals, Series, and Products, Alan Jeffrey (editor) 5th Ed., New York: Academic Press, 1994.

[4] Gupta, R. C. and Keating, J. P. (1986). Relations for Reliability Measures under Length Biased Sampling, *Scand Journal of Statistics*, 13: 49-56.

[5] Gupta, R. C. and Tripathi, R. C. (1990). Effect of length-biased sampling on the modeling error, *Communication in statistics –Theory and Methods*, 19(4):1483-1491.

[6] Khatree, R. (1989). Characterization of Inverse-Gaussian and Gamma distributions through their length-biased distributions, *IEEE Transactions on Reliability*, 38(5):610-611.

[7] Musa, J. D. Iannino, A. and Okumoto, K. Software Reliability: Measurement, Prediction, Application, New York:McGraw-Hill, 1987.

[8] Musa, J. D. and Okumoto, K. (1984). A logarithmic Poisson execution time model for software reliability measurement, *Proceedings of Seventh International conference on software engineering Orlando*, 230-238.

[9] Rao, C. R. (1965). On discrete distributions arising out of methods of ascertainment." In Classical and Contagious Discrete Distributions, Eds. G.P. Patil, *Pergamon Press and Statistical Publishing Society. Calcutta*, 320-332.

[10] Singh, R. and Andure, N. W. (2008). Bayes estimators for the parameters of the Poisson type exponential distribution", *IAPQR transactions*, 33(2):121-128.

[11] Singh, R. and Singh, P. (2012). Characterization of some Software Reliability Growth Models and Parameter Estimation, *MRP-UGC, New Delhi*, 2012 (Submitted to UGC New Delhi).

[12] Singh, R. Vidhale, A. A. and Carpenter, M. (2009). Bayes estimators of parameters of Poisson Type Exponential Class Software Model considering generalized Poisson and Gamma priors, *Journal of Model Assist. Statis. Appl.*, 4(2)83-89.

[13] Singh, R. Singh, P. and Kale, K. R. (2016). Bayes estimators for the parameters of Poisson Type Length Biased Exponential Class Model using Non-Informative Priors, *Journal of Reliability and Statistical Studies*, 9(1):21-28.

[14] Singh, R. Singh, P. and Kale, K. R. (2022). Estimation of Parameters of PTRC SRGM using Non-informative Priors. *International Journal of Advanced Research in Science, Communication and Technology*, 2(1)172-178.