Bayesian Predictive Analyses for Logarithmic Non-Homogeneous Poisson Process in Software Reliability

Nickson Cheruiyot, Luke Akong'o Orawo, Ali Salim Islam

Department of Mathematics, Egerton University, Egerton, Kenya
Email: mutaicher21@gmail.com, orawo2000@yahoo.com, asislam54@yahoo.com

Abstract

This paper discusses the Bayesian approach to estimation and prediction of the reliability of software systems during the testing process. A Non-Homogeneous Poisson Process (NHPP) arising from the Musa-Okumoto (1984) software reliability model is proposed for the software failures. The Musa-Okumoto NHPP reliability model consists of two components—the execution time component and the calendar time component, and is a popular model in software reliability analysis. The predictive analyses of software reliability model are of great importance for modifying, debugging and determining when to terminate software development testing process. However, Bayesian and Classical predictive analyses on the Musa-Okumoto (1984) NHPP model is missing on the literature. This paper addresses four software reliability issues in single-sample prediction associated closely with development testing program. Bayesian approach based on non-informative prior was adopted to develop explicit solutions to these problems. Examples based on both real and simulated data are presented to illustrate the developed theoretical prediction results.

Subject Areas

Mathematical Statistics

Keywords

Non-Informative Priors, Non-Homogeneous Poisson Process, Bayesian Approach, Intensity Function, Software Reliability Model

1. Introduction

Software has become a driver for everything in the 21$^{\text{st}}$ century from elementary

education to genetic engineering. Thus due to high dependency, the size and complexity of computer systems have grown and these pose a great problem in their reliability as failures are prone to happen during their operations. To avoid the failures and faults, reliability of software needs to be studied during development of software so as to come up with reliable software. Reliability of software is of a lot of concern to the developers.

Software reliability is defined as the probability of failure free software operations for a specified period of time in a specified environment [1]. With the increasing need of software with zero defects, predicting reliability of software systems is gaining more and more importance [2]. Software reliability is achieved through testing during the software development stage [3]. Software Reliability modeling is done to estimate the form of the curve of the failure rate by statistically estimating the parameters associated with the selected model. In most cases, the reliability development of a complex system often take place by testing a system until it fails, then making repairs and design changes and testing it again. This process continues until a desired level of reliability is achieved [4]. The purpose of this measure is to estimate the extra execution time during test required to meet a specified reliability objective and to identify the expected reliability of the software when the product is released. During reliability modeling, the software systems are tested in an environment that resembles to the operational environment [5].

Over the past decades many software reliability models that can be used for predictive analyses have been proposed by different authors [6] [7]. The Musa-Okumoto reliability model, also known as logarithmic was developed by Musa and Okumoto in 1984; which they confirmed to be more accurate than the exponential model. The Musa-Okumoto software reliability model is one of non-homogeneous Poisson process software models with the intensity function given by;

$$\lambda(t) = \frac{\alpha\beta}{1 + \beta t} \tag{1}$$

The model is based on the assumptions that failures are observed during execution time caused by remaining faults in the software; whenever a failure is observed, an instantaneous effort is made to find what caused the failure and the faults are removed prior to future tests and whenever a repair is done it reduces the number of future faults not like other models. The model must remain stable during the entire testing period for any particular testing environment and a reasonably accurate prediction of reliability must be provided by the model. These are the two main aspects of a good reliability model [8]. The Musa-Okumoto (1984) model has been used in various testing environment and in many instances, it provides good estimation and prediction of software reliability. Compared to other models when used in testing industrial data set, Musa-Okumoto model is the best performer in terms of fitting and predictive capability to the data [5].

There has been a lot of application of Musa-Okumoto software reliability growth model as it one of the best predictive models, it belongs to the selected models in the AIAA recommended practice standard on software reliability [9], [10]. Musa-Okumoto model have been also used in software cost estimation models with high accuracy [11] [12] [13]. A critical review and categorization of software reliability have been done by many researchers [14] [15]. Predictive analyses on this model is missing in literature and this paper presents predictive analyses on Musa-Okumoto software reliability model using Bayesian approach. This paper presents Bayesian single-sample predictive inference for Musa-Okumoto software reliability model using Bayesian approach.

2. Bayesian Methodology

Bayesian method owes its name to the fundamental role of Bayes' theorem. In Bayesian reasoning, uncertainty is attributed not only to data but also to the parameters. Therefore, all parameters are modelled by distributions. Before any data are obtained, the knowledge about the parameters of a problem are expressed in the prior distribution of the parameters. Given actual data, the prior distribution and the data are combined into the posterior distribution of the parameters. The posterior distribution summarizes our knowledge about the parameters after observing the data.

In this paper we assume that a reliability growth testing is performed on a computer software system and the number of failures in the time interval $(0,t]$, denoted by $N(t)$ is observed. We also assume that $\{N(t), t > 0\}$ follows the NHPP with intensity given in Equation (1). Let $0 < t_1 < t_2 < \cdots$ be the successive failure times. When testing stops after a pre-determined n number of failures is observed, the failure data is said to be failure-truncated. We denote the n failures time by $Y_{obs}^f = [t_i]_{i=1}^n$ where $0 < t_1 < t_2 < \cdots < t_n$, a time-truncated data is when testing is observed for fixed time *t*. We denote the corresponding observed data by $Y_{obs}^t = \{n, t_1, \cdots, t_n; t\}$, where $0 < t_1 < \cdots < t_n \leq t$.

2.1. Issues in Prediction

In this paper we present four issues 1) 2) 3) and 4) as listed below in single-sample prediction which are associated closely with development testing program of a software. Here, we consider one software and assume that its cumulative time between failure times obey Musa-Okumoto software reliability growth model with observed data as either Y_{obs}^f or Y_{obs}^t. Based on Y_{obs}^f or Y_{obs}^t, we are interested in the following problems:

1) What is the probability that at most *k* software failure will occur in the future time period $(T, \tau]$ with $\tau > T$?

2) Given that the pre-determined target value λ_{tv} for the failure rate of the software undergoing development testing is not achieved at time *T*, what is the probability that the target value λ_{tv} will be achieved at time $\tau, \tau > T$?

3) Suppose that the target value λ_{tv} for the software failure rate is not

achieved at time T, how long will it take so that the software failure rate will be attained at λ_{tv}?

4) What is the upper prediction limit (UPL) of $\lambda_{\tau} = \alpha\beta/(1+\beta\tau)$ with level γ. τ being a pre-determined value greater than T?

2.2. Prior, Posterior and Predictive Distributions

Let Y_{obs} represent Y_{obs}^f or Y_{obs}^t. The joint density of Y_{obs} is therefore:

$$f\left(Y_{obs}/\alpha,\beta\right) = \left(\alpha\beta\right)^n \prod_{i=1}^{n}\left(1+\beta t_i\right)^{-1} e^{\left(-\alpha\ln(1+\beta T)\right)}, \ \alpha > 0, \beta > 0 \quad (2)$$

Case 1: β, the shape parameter is known, we adopt the following non-informative prior distribution for α:

$$\pi(\alpha) \propto \frac{1}{\alpha}, \ \alpha > 0 \quad (3)$$

The posterior distribution of α is thus given by;

$$h\left(\alpha/Y_{obs}\right) = \left[\Gamma(n)\right]^{-1}\alpha^{n-1}\left(\ln(1+\beta T)\right)^n e^{\left(-\alpha\ln(1+\beta T)\right)} \quad (4)$$

Let $\tilde{y}$ be the random variable being predicted. The predictive density of $\tilde{y}$ is;

$$f\left(\tilde{y}/Y_{obs}\right) = \int_0^{\infty} f\left(\tilde{y}/Y_{obs}\right)h\left(\alpha/Y_{obs}\right)\mathrm{d}\alpha \quad (5)$$

Hence, the Bayesian UPL of $\tilde{y}$ with level γ, denoted as $y_U^{(\beta)}$, must satisfy

$$\gamma = \int_{-\infty}^{y_U^{(\beta)}} f\left(y/Y_{obs}\right)\mathrm{d}y \quad (6)$$

Case 2: The shape parameter β is unknown; we consider the following joint prior distribution of α and β where both parameters are assumed to be independent.

$$\pi(\alpha,\beta) \propto \frac{1}{\alpha\beta}, \alpha, \beta > 0 \quad (7)$$

Thus the corresponding joint posterior distribution for α and β is given as;

$$h\left(\alpha,\beta/Y_{obs}\right) = \left[k\Gamma(n)\right]^{-1}\alpha^{n-1}\beta^{n-1}\prod_{i=1}^{n}\left(1+\beta t_i\right)^{-1}e^{\left(-\alpha\ln(1+\beta T)\right)} \quad (8)$$

Equation (8) is similar to Equation (4), let $\tilde{y}$ be the random variable predicted. The predictive density of $\tilde{y}$ is;

$$f\left(\tilde{y}/Y_{obs}\right) = \int_0^{\infty}\int_0^{\infty} f\left(\tilde{y}/Y_{obs},\alpha,\beta\right)h\left(\alpha,\beta/Y_{obs}\right)\mathrm{d}\alpha\mathrm{d}\beta \quad (9)$$

and the Bayesian UPL denoted by y_U of $\tilde{y}$ with level γ similar to Equation (6) is;

$$\gamma = \int_{-\infty}^{y_U} f\left(\tilde{y}/Y_{obs}\right)\mathrm{d}y \quad (10)$$

3. Main Results for Prediction Using Non-Informative Priors

In this section we address the four issues stated in Section 2.1 using the Bayesian approach. The main results are presented as propositions and their proof given in the **Appendix**. Below, we use $\chi^2(n;\gamma)$ to represent γ the percentage point of the chi-square distribution with n degrees of freedom such that $\Pr\left\{\chi^2(n) \le \chi^2(n;\gamma)\right\} = \gamma$, and define Poisson $(h/\theta) = \theta^h \mathrm{e}^{-\theta}/h!$ and gamma $(x/n,\lambda) = \lambda^n x^{n-1} \mathrm{e}^{-\lambda n}/\Gamma(n)$. The prior is assumed to be Equation (3) and Equation (7) in all subsequent propositions.

Preposition 1 (issue 1)

The probability that at most k failures will occur in the time interval $(T,\tau]$ with $\tau > T$ is

$$
\gamma_k = \begin{cases}
\dfrac{\left[\ln(1+\beta T)\right]^n}{\left[\ln\left(\dfrac{1+\beta\tau}{1+\beta T}\right)\right]} \displaystyle\sum_{j=n}^{n+k} \binom{j-n}{n-1} \dfrac{\left[\ln\left(\dfrac{1+\beta\tau}{1+\beta T}\right)\right]^j}{\left[\ln(1+\beta\tau)\right]^j} & \text{if } \beta \text{ is known} \\[4em]
\displaystyle\sum_{j=n}^{n+k} \dfrac{\Gamma(j)}{d(j-n)!\,\Gamma(n)} \int_0^\infty \dfrac{\beta^{n-1}\prod_{i=1}^n (1+\beta t_i)^{-1}}{\left[\ln(1+\beta\tau)\right]^j}\left[\ln\left(\dfrac{1+\beta\tau}{1+\beta T}\right)\right]^{j-n} \mathrm{d}\beta & \text{if } \beta \text{ is unknown}
\end{cases}
\tag{11}
$$

Preposition 2 (issue 2)

The probability that the target value λ_{tv} will be achieved at time τ $(\tau > T)$ is

$$
\gamma_k = \begin{cases}
1 - \displaystyle\sum_{h=0}^{n-1} \dfrac{\left[\left(\dfrac{1+\beta\tau}{\beta}\right)\lambda_\tau \ln(1+\beta T)\right]^h}{h!}\, \mathrm{e}^{-\lambda_\tau\left(\frac{1+\beta\tau}{\beta}\right)\ln(1+\beta T)} & \text{if } \beta \text{ is known} \\[4em]
1 - \dfrac{1}{k}\displaystyle\sum_{h=0}^{n-1}\int_0^\infty \dfrac{\left[\left(\dfrac{1+\beta\tau}{\beta}\right)\lambda_{tv}\ln(1+\beta T)\right]^h}{h!}\,\dfrac{\beta^{n-1}\prod_{i=1}^n(1+\beta t_i)^{-1}}{\left[\ln(1+\beta T)\right]^n}\,\mathrm{e}^{-\lambda_{tv}\left(\frac{1+\beta\tau}{\beta}\right)\ln(1+\beta T)}\,\mathrm{d}\beta & \text{if } \beta \text{ is unknown}
\end{cases}
\tag{12}
$$

Preposition 3 (issue 3)

For a given level γ, the time τ^* required to attain λ_{tv} is

$$
\tau^* = \begin{cases}
\left[\dfrac{\chi^2(2n;\gamma)}{2\lambda_{tv}\ln(1+\beta T)} - \dfrac{1}{\beta}\right] - T & \text{if } \beta \text{ is known} \\[2em]
\tau - T & \text{if } \beta \text{ is unknown}
\end{cases}
\tag{13}
$$

Remark 1: For the second part of Equation (13), τ is the solution to the equation

$$
\gamma = 1 - \frac{1}{k}\sum_{h=0}^{n-1}\int_0^\infty \frac{\left[\lambda_\tau \dfrac{1+\beta\tau}{\beta}\ln(1+\beta T)\right]^h}{h!}\,\mathrm{e}^{-\lambda_\tau \frac{1+\beta\tau}{\beta}\ln(1+\beta T)}\cdot\frac{\beta^{n-1}\prod_{i=1}^n(1+\beta t_i)^{-1}}{\left[\ln(1+\beta T)\right]^n}\,\mathrm{d}\beta .
\tag{14}
$$

Preposition 4 (issue 4)

The Bayesian UPL of $\lambda_\tau = \dfrac{\alpha\beta}{1+\beta\tau}$ with level γ is

$$\lambda_U^{(\beta)}(\tau) = \begin{cases} \dfrac{\beta\chi^2(2n;\gamma)}{2(1+\beta\tau)\ln(1+\beta T)} & \text{if } \beta \text{ is known} \\ \lambda_{tv} & \text{if } \beta \text{ is unknown} \end{cases} \tag{15}$$

Remark 2: For the second part of Equation (15), λ_{tv} is the solution to

$$\gamma = 1 - \frac{1}{k}\sum_{h=0}^{n-1}\int_0^\infty \frac{\left[\lambda_{tv}\dfrac{1+\beta\tau}{\beta}\ln(1+\beta T)\right]^h}{h!} e^{-\lambda_{tv}\frac{1+\beta\tau}{\beta}\ln(1+\beta T)} \cdot \frac{\beta^{n-1}\prod_{i=1}^{n}(1+\beta t_i)^{-1}}{\left[\ln(1+\beta T)\right]^n}\,\mathrm{d}\beta \tag{16}$$

4. Real Example

We have used the time between failures data described in [16] to illustrate the developed methodologies for the single-sample Bayesian predictive analysis. We conducted the goodness of fit test presented in [17] and found that the data obey the Musa-Okumoto process. On the basis of this data set the maximum likelihood estimates for the parameters α and β of the Musa-Okumoto growth model were obtained as $\hat{\beta} = 0.008282448$ and $\hat{\alpha} = 15.285550499$, respectively.

1) Suppose we are interested in the probability γ_k that at most *k* failures will occur in a future time period $(T,\tau] = (180,250]$. a) For the case β known, we take its maximum likelihood estimate as its true value, *i.e.* $\beta = 0.008282448$. Using the first formula in Equation (11), we have $\gamma_0 = 0.00204337$, $\gamma_1 = 0.01347748$, $\gamma_2 = 0.04653484$, $\gamma_3 = 0.11230530$, $\gamma_4 = 0.21351423$, $\gamma_5 = 0.34188371$, $\gamma_6 = 0.48155675$, $\gamma_7 = 0.61554018$, $\gamma_8 = 0.73112395$, $\gamma_9 = 0.82215131$, $\gamma_{10} = 0.88836847$, $\gamma_{11} = 0.93328146$, $\gamma_{12} = 0.96190403$, $\gamma_{13} = 0.97915241$, $\gamma_{14} = 0.98903392$, $\gamma_{15} = 0.99444044$. b) When β is unknown, from the second formula of Equation (33), we obtain $\gamma_0 = 0.002423218$, $\gamma_1 = 0.015348190$, $\gamma_2 = 0.052195351$, $\gamma_3 = 0.122789747$, $\gamma_4 = 0.229662151$, $\gamma_5 = 0.362653362$, $\gamma_5 = 0.362653362$, $\gamma_7 = 0.639743372$, $\gamma_8 = 0.750133795$, $\gamma_9 = 0.836483617$, $\gamma_{10} = 0.897655409$, $\gamma_{11} = 0.941055333$, $\gamma_{12} = 0.963420189$, $\gamma_{13} = 0.981411270$, $\gamma_{13} = 0.981411270$, $\gamma_{15} = 0.991582449$ **Figure 1** shows the graph of desired probabilities when β is known and when it is unknown.

From the graph it can be seen that there is high probability that at most 15 failures will occur during that time interval when β is unknown as compared to when it is known. 2) Suppose the target value is given by $\lambda_{tv} = 0.03$ chosen arbitrarily. At the time $T = 182.21$, the MLE of the achieved failure rate for this software is $\hat{\lambda}(182.21) = \hat{\alpha}\hat{\beta}/(1+182.21\hat{\beta}) = 0.05045615$, which is greater than λ_{tv} thus it cannot be achieved at time $T = 182.21$ and development testing will continue. Suppose we want to find the probability that the target value λ_{tv} will be achieved at the time $\tau = 277.83\,\mathrm{h}$. a) When β is known (say, $\beta = 0.008282448$), from the first formula in Equation (12), we obtain
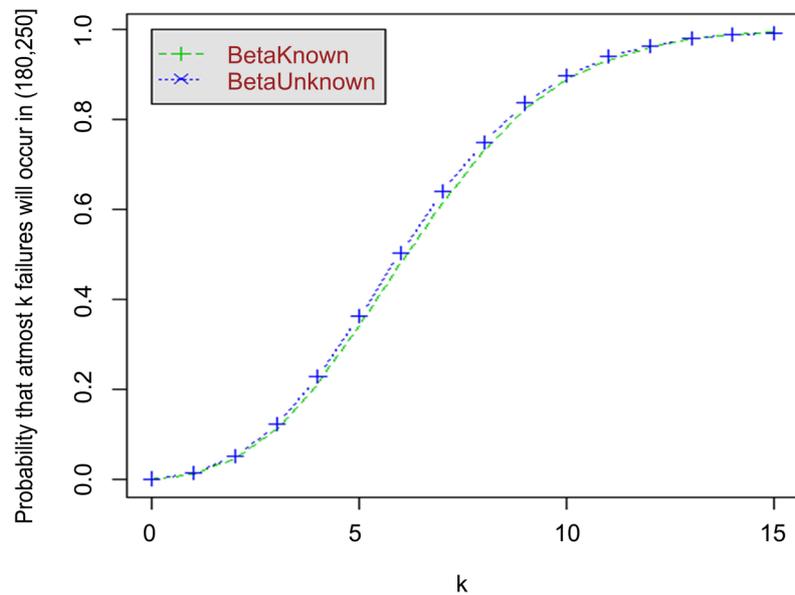
Figure 1. The graph of the probabilities γ_k that at most *k* failures will occur in the time interval (180, 250] for the cases of β known and unknown.

$\gamma = 1.687506\text{e} - 06$, which is very small and hence the target value will not be achieved. b) when β is unknown, from the second formula in Equation (12) we have $\gamma = 0.193896$ computed by the Monte Carlo Method of integration based on a sample of size $L = 1000$. This shows that, when β is unknown there is a possibility of achieving the target value at time $\tau = 277.83\,\text{h}$.

3) Since the target value λ_{tv} was not achieved at $T = 182.21$, we want to know how long it will take for the target value to be achieved. a) when β is known (say, $\beta = 0.008282448$), let $\gamma = 0.90$, from the first formula in Equation (13) we obtain $\tau^* = 538.7523\,\text{h}$. This means that, it will take another 538.7523 hours in order to achieve the desired failure rate. b) when β is unknown, from second formula in Equation (13) and Remark1, we obtain $\tau^* = 414\,\text{h}$. Thus, it takes another 414 hours in order to achieve the desired failure rate when β is unknown this shows a high reduction in time as compared to when β is known. 4) Given $\tau = 900\,\text{h}$, from first formula in Equation (15) the Bayesian UPL of $\lambda_\tau = \dfrac{\alpha\beta}{1 + \beta\tau}$ with level $\gamma = 0.90$ is given by

$$\lambda_U^{(\beta)}(\tau) = 0.02473799.$$

5. Conclusions

In software development, predictive analysis is very important as it helps the software developer to make a trade-off decision at the right time. In this paper, explicit solution to predictive issues that may arise during development process were derived using Bayesian approach. These solutions are helpful to software developers in many instances such as resource allocation, when to terminate the testing process, modification needed in the software before termination.

The study used Bayesian approach with non-informative priors to derived explicit solutions for predictive issues that may arise during software development process. In all the cases when the shape parameter was known, solutions to posterior and predictive distributions had closed forms while when it is unknown, solutions had no closed forms and the study used Markov Chain Monte Carlo (MCMC). Bayesian approach was used as it is advantageous over classical approach. Bayesian approach is available for small sample sizes and allows the input of prior information about reliability growth process and provides full posterior and predictive distributions [6].

However, it will be interesting to look at two-sample prediction for Musa-Okumoto (1984) model considering procedures that [3] used. These procedures presented in this paper can also be extended to other NHPP models such as Cox-Lewis process and the delayed S-shaped process. This is left open for future research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Nuria, T.R. (2011) Stochastic Comparisons and Bayesian Inference in Software Reliability. Ph.D. Thesis, Universidad Carlos III de Madrid, Madrid.

[2] Sonia, D. and Renu, D. (2014) A Study of Various Reliability Growth Models. *International Journal of Advanced Research in Computer Science and Software Engineering*, **4**, 1213-1219.

[3] Daniel, R.J. and Hoang, P. (2001). On the Maximum Likelihood Estimates for the Goel-Okumoto Software Reliability Model. *The American Statistician*, **55**, 219-222.
https://doi.org/10.1198/000313001317098211

[4] Muralidharan, K., Rupal, S. and Deepak, H.D. (2008) Future Reliability Estimation Based on Predictive Distribution in Power Law Process. *Quality Technology & Quantitative Management*, **5**, 193-201.
https://doi.org/10.1080/16843703.2008.11673396

[5] Ullah, N., Morisio, M. and Vetro, A. (2013) A Comparative Analysis of Software Reliability Growth Models Using Defects Data of Closed and Open Source Software. In: 35*th Annual IEEE Software Engineering Workshop*, Heraclion, 12-13 October 2012, 187-192.

[6] Yu, J.-W., Tian, G.-L. and Tang, M.-L. (2007) Predictive Analyses for Non-Homogeneous Poisson Processes with Power Law Using Bayesian Approach. *Computational Statistics & Data Analysis*, **51**, 4254-4268.
https://doi.org/10.1016/j.csda.2006.05.010

[7] Akuno, A.O., Orawo, L.A. and Islam, A.S. (2014) One-Sample Bayesian Predictive Analyses for an Exponential Non-Homogeneous Poisson Process in Software Reliability. *Open Journal of Statistics*, **4**, 402-411. https://doi.org/10.4236/ojs.2014.45039

[8] Kapur, P.K., Pham, H., Gupta, A. and Jha, P.C. (2011) Software Reliability Assessment with OR Applications. Springer Series in Reliability Engineering, Springer-Verlag, London, 58. https://doi.org/10.1007/978-0-85729-204-9

[9] Lyu, M. (1996) Handbook of Software Reliability Engineering. McGraw-Hill, New York.

[10] Malaiya, Y.K. and Denton, J. (1997) What Do the Software Reliability Growth Model Parameters Represent? Technical Report CS-97-115, Department of Computer Science, Colorado State University, Fort Collins.

[11] Xia, W., Capretz, L.F. and Ho, D. (2008) A Neuro-Fuzzy Model for Function Point Calibration. *WSEAS Transactions on Information Science and Applications*, **5**, 22-30.

[12] Nassif, A.B., Capretz, L.F. and Ho, D. (2010) Software Estimation in the Early Stages of the Software Life Cycle. In: *International Conference on Emerging Trends in Computer Science, Communication and Information Technology*, Nanded, 9-11 January, 5-13.

[13] Nassif, A.B., Ho, D. and Capretz, L.F., (2013) Towards an Early Software Estimation Using Log-Linear Regression and a Multilayer Perceptron Model. *Journal of Systems and Software*, **86**, 144-160. https://doi.org/10.1016/j.jss.2012.07.050

[14] Yadav, A. and Khan, R.A. (2009) Critical Review on Software Reliability Models. *International Journal of Recent Trends in Engineering*, **2**, 114-116.

[15] Sheakh, T.H., Quadri, S.M.K. and Singh, V. (2012) Critical Review of Software Reliability Model. *International Journal of Emerging Technology and Advance Engineering*, **2**, 496-499.

[16] Xie, M., Goh, T.N. and Ranjan, P. (2002) Some Effective Control Chart Procedures for Reliability Monitoring. *Reliability Engineering & System Safety*, **77**, 143-150. https://doi.org/10.1016/S0951-8320(02)00041-8

[17] Zhao, J. and Wang, J. (2005) A New Goodness-of-Fit Test Based on the Laplace Statistic for a Large Class of NHPP Models. *Communications in Statistics—Simulation and Computation*, **34**, 725-736. https://doi.org/10.1081/SAC-200068389

Appendix: Proof of Preposition 1 - 4

We first state the following identity without proof: That is

$$\int_{D(m;a,b)} dF(t_1)\cdots dF(t_m) = \left[F(a)-F(b)\right]^m \Big/ m! \tag{A.1}$$

where m is any positive integer, a and b are two real numbers such that $a < b$, $F(t)$ is an increasing and differentiable function and

$$D(m;a,b) \triangleq \left\{(t_1,\cdots,t_m): a < t_1 < \cdots < t_m < b\right\}.$$

Proof of Proposition 1

The probability that at most k failures will occur in the interval $(T,\tau]$ is $\gamma_k = \Pr\left\{N(\tau) \le n+k/Y_{obs}\right\}$. When β is known, we have

$$\gamma_k = \int_0^\infty \Pr\left\{N(\tau) \le n+k/Y_{obs},\alpha\right\} \cdot h\left(\alpha/Y_{obs}\right) d\alpha. \tag{A.2}$$

where $h\left(\alpha/Y_{obs}\right)$ is given by equation (4) and

$$\Pr\left[N(\tau) \le n+k/Y_{obs},\alpha\right] = \sum_{j=n}^{n+k} f\left(Y_{obs},N(\tau)=j/\alpha\right) \Big/ f\left(Y_{obs}/\alpha\right) \tag{A.3}$$

From Equation (2), we have $f\left(Y_{obs}/\alpha\right) = \beta^n \alpha^n \prod_{i=1}^n \left(1+\beta t_i\right)^{-1} e^{\left[-\alpha\ln(1+\beta T)\right]}$, and

$$\begin{aligned}
&f\left(Y_{obs},N(\tau)=j/\alpha\right) \\
&= \int_{D(j-n;T,\tau)} f\left(Y_{obs},x_{n+1},\cdots,x_j,N(\tau)=j\right) \prod_{\ell=n+1}^j dx_\ell \\
&= \int_{D(j-n,T,\tau)} \alpha^j \beta^j e^{\left[-\alpha\ln(1+\beta T)\right]} \prod_{i=1}^j \left(1+\beta t_i\right)^{-1} \prod_{\ell=n+1}^j dt_\ell \\
&= \alpha^j \beta^j \prod_{i=1}^n \left(1+\beta t_i\right)^{-1} e^{\left[-\alpha\ln(1+\beta T)\right]} \int_{D(j-n,T,\tau)} \prod_{\ell=n+1}^j \left(1+\beta t_\ell\right)^{-1} \prod_{\ell=n+1}^j dt_\ell
\end{aligned}$$

Thus Equation (A.3) becomes

$$\Pr\left[N(\tau) \le n+k/Y_{obs}\right] = \frac{\sum_{j=n}^{n+k} \alpha^{j-n} e^{\left[-\alpha\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]} \left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^{j-n}}{(j-n)!} \tag{A.4}$$

And hence Equation (A.2) becomes

$$\begin{aligned}
\gamma_k &= \int_0^\infty \sum_{j=n}^{n+k} \alpha^{j-n} e^{\left[-\alpha\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]} \frac{\left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^{j-n} \alpha^{n-1} \left[\ln(1+\beta T)\right]^n e^{\left[-\alpha\ln(1+\beta T)\right]}}{(j-n)!\Gamma(n)} d\alpha \\
&= \sum_{j=n}^{n+k} \frac{\left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^{j-n} \left[\ln(1+\beta T)\right]^n}{(j-n)!\Gamma(n)} \frac{\Gamma(j)}{\left[\ln(1+\beta\tau)\right]^j} \\
&\quad \times \int_0^\infty \frac{\left[\ln(1+\beta\tau)\right]^j}{\Gamma(j)} \alpha^{j-n} e^{\left[-\alpha\ln(1+\beta\tau)\right]} d\alpha
\end{aligned} \tag{A.5}$$

The integral part of Equation (A.5) integrates to 1 since it is a gamma distri-

bution with parameters j and $\ln(1+\beta\tau)$ and hence Equation (A.5) reduces to

$$\gamma_k = \frac{\left[\ln(1+\beta T)\right]^n}{\left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^n} \sum_{j=n}^{n+k} \binom{j-n}{n-1} \frac{\left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^j}{\left[\ln(1+\beta\tau)\right]^j}. \tag{A.6}$$

This is the first formula of Equation (11).

When β is unknown, noting that $\Pr\{N(\tau) \le n+k/Y_{obs}, \alpha, \beta\}$ and $h(\alpha, \beta/Y_{obs})$ are given by Equation (A.4) and Equation (8) respectively, we obtain

$$\gamma_k = \int_0^\infty \int_0^\infty \Pr\left[N(\tau) \le n+k/Y_{obs}, \alpha, \beta\right] \cdot h(\alpha, \beta/Y_{obs}) \, \mathrm{d}\alpha \, \mathrm{d}\beta$$

$$= \sum_{j=n}^{n+k} \frac{1}{(j-n)!\Gamma(n)} \int_0^\infty \int_0^\infty \alpha^{j-1} \beta^{n-1} \prod_{i=1}^n (1+\beta t_i)^{-1} \left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^{j-n} \mathrm{e}^{\left[-\alpha \ln(1+\beta\tau)\right]} \mathrm{d}\alpha \, \mathrm{d}\beta \tag{A.7}$$

$$= \sum_{j=n}^{n+k} \frac{\Gamma(j)}{d(j-n)!\Gamma(n)} \int_0^\infty \frac{\beta^{n-1} \prod_{i=1}^n (1+\beta t_i)^{-1}}{\left[\ln(1+\beta\tau)\right]^j} \left[\ln\left(\frac{1+\beta\tau}{1+\beta T}\right)\right]^{j-n} \mathrm{d}\beta$$

Since the summation of k is from n to $n+k$ and k's are not the same, we substitute letter k with d in Equation (A.7) where $d=k$ as used in equation (8). Equation (A.7) implies the second formula in Equation (11). □

Proof of preposition 2

Let $f(\lambda_\tau/Y_{obs})$ denote the posterior of $\lambda_\tau = \alpha\beta/(1+\beta\tau)$. Hence, the probability that the target value λ_{tv} will be achieved at time τ is given by

$$\gamma = \Pr\{\lambda_\tau \le \lambda_{tv}/Y_{obs}\} = \int_0^{\lambda_{tv}} f(\lambda_\tau/Y_{obs}) \, \mathrm{d}\lambda_\tau \tag{A.8}$$

when β is known, making transformation $\lambda_\tau = \alpha\beta/(1+\beta\tau)$, we have $\alpha = \lambda_\tau \frac{1+\beta\tau}{\beta}$ and $\frac{\mathrm{d}\alpha}{\mathrm{d}\lambda_\tau} = \frac{1+\beta\tau}{\beta}$. Consequently, the posterior density of λ_τ is

$$f(\lambda_\tau/Y_{obs}) = h(\alpha/Y_{obs}) \left|\frac{\mathrm{d}\alpha}{\mathrm{d}\lambda_\tau}\right|$$

$$f(\lambda_\tau/Y_{obs}) = \frac{1}{\Gamma(n)} \left[\frac{\lambda_\tau(1+\beta\tau)}{\beta}\right]^{n-1} \left[\ln(1+\beta T)\right]^n \mathrm{e}^{-\lambda_\tau\left(\frac{1+\beta\tau}{\beta}\right)\ln(1+\beta T)} \cdot \frac{1+\beta\tau}{\beta}$$

$$= \frac{1}{\Gamma(n)} \left[\left(\frac{1+\beta\tau}{\beta}\right)\ln(1+\beta T)\right]^n \lambda_\tau^{n-1} \mathrm{e}^{-\lambda_\tau\left[\left(\frac{1+\beta\tau}{\beta}\right)\ln(1+\beta T)\right]} \tag{A.9}$$

From Equation (A.9), it can easily be noted that λ_τ has gamma distribution with parameters n and $\frac{1+\beta\tau}{\beta} \ln(1+\beta T)$. Noting that gamma and Poisson distributions have a relationship defined as

$$\frac{b^a}{\Gamma(a)} \int_0^\lambda x^{a-1} \mathrm{e}^{-bx} \mathrm{d}x = 1 - \sum_{h=0}^{a-1} \frac{(b\lambda)^h}{h!} \mathrm{e}^{-b\lambda}. \tag{A.10}$$

By substituting Equation (A.9) and Equation (A.10) into Equation (A.8), we obtain the first formula of Equation (12).

When β is unknown, making transformation $\lambda_\tau = \alpha\beta/(1+\beta\tau)$ and $\beta = \beta$, we obtain $\alpha = \lambda_\tau \dfrac{1+\beta\tau}{\beta}$ and $\beta = \beta$. Note that the Jacobian is $\dfrac{\partial(\alpha,\beta)}{\partial(\lambda_\tau,\beta)} = \dfrac{1+\beta\tau}{\beta}$. From Equation (8), the joint posterior density of (λ_τ,β) is

$$f(\lambda_\tau,\beta/Y_{obs}) = h(\alpha,\beta/Y_{obs}) \left| \frac{\mathrm{d}(\alpha,\beta)}{\mathrm{d}(\lambda_\tau,\beta)} \right|.$$

$$
\begin{aligned}
f(\lambda_\tau,\beta/Y_{obs}) &= \left[\frac{\lambda_\tau(1+\beta\tau)}{\beta}\right]^{n-1} \frac{\beta^{n-1}\prod_{i=1}^{n}(1+\beta t_i)^{-1}}{k\Gamma(n)} e^{-\lambda_\tau\frac{1+\beta\tau}{\beta}\ln(1+\beta T)} \cdot \frac{1+\beta\tau}{\beta} \\
&= \frac{\beta^{n-1}\prod_{i=1}^{n}(1+\beta t_i)^{-1}}{k\Gamma(n)\left[\ln(1+\beta T)\right]^n} \left[\frac{1+\beta\tau}{\beta}\ln(1+\beta T)\right]^n \lambda_\tau^{n-1} e^{-\lambda_\tau\frac{1+\beta\tau}{\beta}\ln(1+\beta T)}
\end{aligned}
\tag{A.11}
$$

By substituting Equation (A.10) and Equation (A.11) into Equation (A.8), we obtain the second formula of Equation (12). □

Proof of preposition 3

For given level γ, the time required to attain the target value λ_{tv} is $\tau^* = \tau - T$, where τ satisfies Equation (44). When β is known, from Equation (46), it can easily be seen that

$$2\left[\frac{1+\beta\tau}{\beta}\ln(1+\beta T)\right]\lambda_\tau$$

follows a chi-square distribution with $2n$ degrees of freedom. Thus we have

$$2\left[\frac{1+\beta\tau}{\beta}\ln(1+\beta T)\right]\lambda_{tv} = \chi^2(2n,\gamma). \tag{A.12}$$

and Equation (13) follows immediately.

The time required to attain the target λ_{tv} with level γ when β is unknown is $\tau^* = \tau - T$ where τ is the solution to

$$\gamma = 1 - \frac{1}{k}\sum_{h=0}^{n-1}\int_0^\infty \frac{\left[\lambda_\tau\frac{1+\beta\tau}{\beta}\ln(1+\beta T)\right]^h}{h!} e^{-\lambda_\tau\frac{1+\beta\tau}{\beta}\ln(1+\beta T)} \cdot \frac{\beta^{n-1}\prod_{i=1}^{n}(1+\beta t_i)^{-1}}{\left[\ln(1+\beta T)\right]^n}\mathrm{d}\beta. \tag{A.13} \; □$$

Proof of preposition 4

For a pre-determined $\tau(\tau > T)$, the Bayesian upper prediction limit for $\lambda_\tau = \dfrac{\alpha\beta}{1+\beta\tau}$ with level γ is $\lambda_U^{(\beta)}(\tau)$ satisfying $\gamma = \Pr\left\{\lambda_\tau \le \lambda_U^{(\beta)}(\tau)/Y_{obs}\right\}$. From Equation (A.8) and Equation (A.12), we have $\lambda_U^{(\beta)}(\tau) = \dfrac{\beta\chi^2(2n;\gamma)}{2(1+\beta\tau)\ln(1+\beta T)}$, thus follows Equation (15). The second part follows similarly. □