International Journal of Reliability and Applications
Vol. 2, No. 3, pp. 199-207, 2001

An Adaptive Failure Rate Change-Point Model for Software Reliability

Kwang Mo Jeong *
Department of Statistics
Pusan National University, 609-735, Pusan, Korea

Abstract. The failure rate functions between successive failures are of concatenated form. We allow the parameters of failure rate function change after a certain failure and its fixing. We confine out attention to a model wherein the interfailure times are described by its failure rate function. We suggest an adative failure rate function with a change-point under the assumption that interfailure times are record value statistics from a Weibull distribution. The proposed model will be applied through a practical example of software failure data.

Key Words : *Failure rate, Software reliability, Change-Point, Poisson process.*

1. INTRODUCTION

In early models of software reliability the failure rate of a program is assumed to be a constant multiple of the unknown number of faults remaining. Usually there is assumed an unknown number of faults N at the beginning of software testing and also no new fault is introduced at the time of debugging. This means that each fault affects the same amount of contribution to the failure rate of the program, so that the debugging process results in the failure rate improvement in a series of equally sized steps. The model of Jelinski and Moranda(1972) is the most commonly cited in earlier software reliability context. Musa(1975), and Goel and Okumoto(1978) are the researches on the same assumption as mentioned before. But the assumption is very restricted and so it is challenged to have an alternative assumption. For example, earlier fault fixes may result in a greater effect than later ones and hence make a more contribution to the overall failure rate. Littlewood(1981) proposed a model under this kind of assumption by assuming a gamma prior on each failure rate.

Models of software reliability based on the interfailure times have been classified into several types, for example, by Sinpurwalla and Wilson(1994). Kuo and

* *E-mail address : kmjung@pusan.ac.kr*

Yang(1996) modeled failure times in two different classes; one is the general order statistics model(GOS) and the other is the record value statistics(RVS) model. We note that the observed epochs of failures are assumed to be the first n order statistics taken from N independent and identically distributed(iid) observations with probability density function(pdf) f. The Jelinski-Moranda model is the case when f is an exponential pdf. Knafl and Morgan(1996) discussed 2-parameter Poisson process models for ungrouped software failure data. They treated the maximum likelihod estimation(MLE) problem for several types of GOS models including exponential, Weibull, and Pareto distributions.

We proceed our discussion on RVS models in relation to the nonhomogeneous Poisson process(NHPP), which has been used extensively in software reliability models. NHPP is routinely used to model failures for repairable systems. We may regard a software testing and debugging processes as repairable system because at the time of software failure it can be returned to the state of functioning by fixing the bugs causing the failure. As discussed by Kuo and Yang(1996) RVS models are just the models describing the epochs of failures in the NHPP. RVS models seem to be more reasonable than GOS models in the sense that new faults may be added during software debugging in RVS model. The epochs of failures are assumed ss the record-breaking statistics of unobserved iid outcomes from f.

If we let $m(t)$ be the mean number of failures observed in $(0, t]$ for the NHPP, the failure rate function is defined by $\lambda(t) = dm(t)/dt$. The relation between $\lambda(t)$ and pdf $f(t)$ will be deferred until Section 2. The failure rate functions between successive failures are of concatenated form and hence we term it as the concatenated failure rate function for the software reliability model. It is greatly intuitive to consider that frequent failures should result in a judgement of poor reliability and rare failures in a judgement of high reliability. Al-Mutairi et al.(1998) suggested an adaptive concatenated failure rate function. The term adaptive means that failure history is represented in terms of failure intensity until the previous failure. Al-Mutairi et al.(1998) assumed an adaptive form of Musa-Okumoto failure rate function in a Bayesian framework.

Additionally we allow the parameters of failure rate function change after a certain failure and its fixing. This is the so called change-point model. In this paper we confine out attention to a model wherein the interfailure times are described by its failure rate function. We suggest an adaptive failure rate function with a change-point under the assumption that interfailure times are RVS from a Weibull distribution. This motivation seems to be reasonable in the sense that the failure rate varies according to the number of failures and the elapsed time until the previous failure. The proposed model will be compared with the results of Kuo and Yang(1996) through a practical example of software failure data.

2. A CHANGE-POINT FAILURE RATE MODEL

First we briefly review some of the NHPP which are associated with the existing software reliability models. For the GOS model the random number of failures $M(t)$ in $(0,t]$ is NHPP with $m(t)=\theta F(t)$ when N has a Poisson distribution with mean θ, where $F(t)$ is a cumulative distribution function corresponding to the *pdf* $f(t)$. When $F(t)=1-\exp(-t/\beta)$ then $M(t)$ is a NHPP with the mean function $m(t)=\theta(1-\exp(-t/\beta)$. This is called the Goel-Okumoto process (Goel and Okumoto,1979) with the corresponding Jelinski-Moranda model(Langberg and Singpurwalla, 1985).

The RVS model with pdf f corresponds to a NHPP with mean function $m(t)=-\log[1-F(t)]$. According to the failure rate function $\lambda(t)$ several well-known point processes follow: Duane(1964) process with $\lambda(t)=\gamma t^{\gamma-1}/\beta$, and the Cox and Lewis(1966) process with $\lambda(t)=\exp(\alpha+\beta t)$, and also the Musa-Okumoto(1984) process with $\lambda(t)=\alpha/(\beta+t)$. We note that the set of observed failure epochs of the Duane process is the counting process of the RVS from a Weibull distribution. We also note that the failure epochs of a NHPP with failure intensity $\lambda(t)$ obey the RVS model with the pdf

$$f(t)=\lambda(t)\exp\left(-\int_0^t \lambda(u)du\right). \tag{2.1}$$

Let T_i be the interfailure time between the $(i-1)st$ failure and the ith failure of a software. In this section we formulate the interfailure times in terms of failure rate function $\lambda_{i+1}(t)$, $i=1,2,\cdots$. Let S_i be the successive times to the ith failure, so that $T_i=S_i-S_{i-1}$. We assume hereafter the failure rate function of the form

$$\lambda_{i+1}(t|\gamma,\beta)=\frac{\gamma}{\beta}t^{\gamma-1}, \tag{2.2}$$

that is, the interfailure time T_{i+1} follows from a Weibull distribution. From (2.1) the pdf of interfailure time T_{i+1} is given by

$$f_{i+1}(t)=\frac{\gamma}{\beta}t^{\gamma-1}\exp(-\frac{t^\gamma}{\beta}).$$

The interfailure times are assumed to be independent but they are in general not identically distributed because the failure intensity varies according to the occurrence of each failure.

Given observed data $t_1,t_2,\cdots,t_n$, with $s_i=\sum_{j\le i}t_j$ denoting the failure time duration until the ith failure, we introduce an adaptive failure rate function of the form

$$\lambda_{i+1}(t|s_i,\gamma,\beta)=\frac{\gamma}{\beta}w_i t^{\gamma-1}, \tag{2.3}$$

where $w_i=\frac{i}{s_i}$ denotes a failure intensity during the first i failures. It would be more reasonable to expect that the interfailure times change dramatically after a specified failure and hence its fixing. We call this point of observation number a change-point. Define an adaptive failure rate function with a change-point k as

$$\lambda_{i+1}(t|s_i,\gamma,\beta)=\begin{cases}\frac{\gamma_1}{\beta_1}w_i t^{\gamma_1-1}, & i<k\\ \frac{\gamma_2}{\beta_2}w_i t^{\gamma_2-1}, & i\ge k.\end{cases} \tag{2.4}$$

In the change-point failure rate function we allow the shape and scale parameters change after a certain observation number k. This type of model seems to be more flexible in the sense of representing the contribution of a particular bug and its fixing. When the shape parameters are assumed to be equal, i.e., $\gamma_1 = \gamma_2 = 1$ the model (2.4) is reduced to a constant failure rate model of exponential distribution.

3. MAXIMUM LIKELIHOOD ESTIMATION

From the observed data $t_1, t_2, \cdots, t_n$ we consider the likelihood function and the estimation of parameters. Under the assumed failure rate function of (2.4) the pdf of T_{i+1} is given by

$$f_{i+1}(t) = \begin{cases} \frac{\gamma_1}{\beta_1} w_i t^{\gamma_1 - 1} \exp(-w_i \frac{t^{\gamma_1}}{\beta_1}) \ , & i < k \\ \frac{\gamma_2}{\beta_2} w_i t^{\gamma_2 - 1} \exp(-w_i \frac{t^{\gamma_2}}{\beta_2}) \ , & i \geq k. \end{cases} \tag{3.1}$$

Then for any fixed k the log-likelihood function can be written in the form

$$\begin{aligned} l_n(\boldsymbol{\gamma}, \boldsymbol{\beta} | k, \underline{t}) &\propto (k-1)\log(\frac{\gamma_1}{\beta_1}) + (\gamma_1 - 1)\sum_{i<k} \log(t_{i+1}) \\ &- \frac{1}{\beta_1}\sum_{i<k} w_i t_{i+1}^{\gamma_1} + (n-k)\log(\frac{\gamma_2}{\beta_2}) \\ &+ (\gamma_2 - 1)\sum_{i \geq k} \log(t_{i+1}) - \frac{1}{\beta_2}\sum_{i \geq k} w_i t_{i+1}^{\gamma_2} \end{aligned} \tag{3.2}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$, and $\underline{t} = (t_1, \cdots, t_n)$.

Differentiating $l_n(\boldsymbol{\gamma}, \boldsymbol{\beta} | k, \underline{t})$ with respect to β_1 and equating to zero we find that

$$\beta_1(\gamma_1, k) = \frac{1}{k-1}\sum_{i<k} w_i t_{i+1}^{\gamma_1} \ . \tag{3.3}$$

Similarly we obtain

$$\beta_2(\gamma_2, k) = \frac{1}{n-k}\sum_{i \geq k} w_i t_{i+1}^{\gamma_2} \ . \tag{3.4}$$

For simplicity of notations we denote $\beta_i(\gamma_i, k)$ as β_i. From the pseudo-likelihood function $\tilde{l}_n(\boldsymbol{\gamma}, k)$, which is obtained by substituting β_i into (3.2), we find the MLE of $(\boldsymbol{\gamma}, k)$ as the maximizer of $\tilde{l}_n(\boldsymbol{\gamma}, k)$, that is,

$$(\hat{\boldsymbol{\gamma}}, \hat{k}) = \arg\max_{k_1 \leq k \leq k_2} \sup_{\boldsymbol{\gamma} > 0} \tilde{l}_n(\boldsymbol{\gamma}, k) \ ,$$

where k_1 and k_2 are the serial orders of observations taken at both sides of data set so that k is assumed to lie between k_1 and k_2.

In practical calculations we first differentiate $l_n(\gamma, \beta | k, \underline{t})$ given in (3.2) with respect to γ_1 to obtain the following likelihood equation

$$\frac{k-1}{\gamma_1} + \sum_{i<k} \log(t_{i+1}) = \frac{1}{\beta_1} \sum_{i<k} w(i) t_{i+1}^{\gamma_1} \log(t_{i+1}) . \quad (3.5)$$

We next solve two equations (3.3) and (3.5) simultaneously to find the MLEs of β_1 and γ_1 for fixed k. The Newton-Raphson method is a typical one to find the iterative solutions. In a similar way we can obtain the MLEs of γ_2 and β_2 by simultaneously solving the equation (3.4) with the following likelihood equation given by

$$\frac{n-k}{\gamma_2} + \sum_{i \geq k} \log(t_{i+1}) = \frac{1}{\beta_2} \sum_{i \geq k} w(i) t_{i+1}^{\gamma_2} \log(t_{i+1}) . \quad (3.6)$$

After the β_i and γ_i are obtained as functions of k we find the pseudo-likelihood function $\tilde{l}_n(k)$, which depends only on k, by substituting β_i and γ_i into (3.2). The MLE of k is the point of observation order maximizing $\tilde{l}_n(k)$ between k_1 and k_2.

To infer for the future mean time between failures we suggest the one-step-ahead prediction given the previous epoch of failure. Given the epoch s_i of *ith* failure we can calculate the conditional mean of T_{i+1} before change-point k, i.e., $i < k$, as

$$E(T_{i+1} | s_i, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \left(w_i \frac{1}{\beta_1} \right)^{-\frac{1}{\gamma_1}} \Gamma(1 + \frac{1}{\gamma_1}) , \quad (3.7)$$

where $\Gamma(a)$ is a gamma function defined by

$$\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt .$$

On the other hand when $i \geq k$ the conditional mean $E(T_{i+1} | s_i, \boldsymbol{\gamma}, \boldsymbol{\beta})$ is the same form as (3.7) except that γ_1 and β_1 are replaced by γ_2 and β_2, respectively. From the equation (3.7) and so on the one-step-ahead predictions for the mean times between failures are obtained by substituting the MLEs of parameters γ_i, β_i and k.

We can also derive the conditional variance of T_{i+1} given s_i, when $i < k$, as follows

$$V(T_{i+1} | s_i, \boldsymbol{\gamma}, \boldsymbol{\beta}) = \left(w_i \frac{1}{\beta_1} \right)^{-\frac{2}{\gamma_1}} \left[\Gamma(1 + \frac{2}{\gamma_1}) - \Gamma^2(1 + \frac{1}{\gamma_1}) \right] , \quad i < k \quad (3.8)$$

In a similar way $V(T_{i+1} | s_i, \boldsymbol{\gamma}, \boldsymbol{\beta})$, when $i \geq k$, is the same form as (3.8) except that γ_1 and β_1 are replaced by γ_2 and β_2, respectively.

Nextly the mean residual life $E(T_{i+1} - \tau_0 | T_{i+1} > \tau_0)$ for any fixed time period τ_0 can be written in the form

$$E(T_{i+1} - \tau_0 | T_{i+1} > \tau_0) = \frac{1}{1 - F_{i+1}(\tau_0)} \int_{\tau_0}^\infty t f_{i+1}(t) dt.$$

It's not difficult to express the right hand side of the above equation in terms of complementary incomplete gamma function but we omit it here.

4. AN EXAMPLE

We explain the proposed adaptive change-point software reliability model through a practical data set. The data set of 31 failure times, as shown in the 2nd column of Table 4.1, was based on the trouble report for one of the larger modules of the Naval Tactical Data System(NTDS). The first 26 failures were found during the production phase, and the remaining 5 errors were observed during the test phase. This data set was previously illustrated by Goel and Okumoto(1979), and Mazzuchi and Soyer(1988). In particular Kuo and Yang(1996) suggested a RVS model from Weibull distribution in Bayesian framework with some priors on Weibull parameters.

Under the change-point failure rate function we obtain the estimated change-point $\hat{k} = 21$. The corresponding estimated parameters before $\hat{k}$ is $\hat{\gamma}_1 = 1.63$, and $\hat{\beta}_1 = 2.22$. On the other hand the corresponding MLEs after change-point are $\hat{\gamma}_2 = 0.82$ and $\hat{\beta}_2 = 2.08$. We see the failure rate function changed to be monotone decreasing after the change-point against the monotone increasing state before it. Therefore we expect a growth of reliability for the NTDS data set after the first $21st$ failures.

This advantage comes from the change-point Weibull process model for the software reliability. The 3rd column of the table represents the one-step-ahead mean times between failures predicted by the proposed model. For comparison the corresponding results of Kuo and Yang(1996) are given in the 4th column of the table. The proposed model is modestly good compared to the Bayesian famework of Kuo and Yang(1996). The predicted one-step-ahead mean times are plotted against the observed serial order in Figure 4.1. The observed failure times fluctuate dramatically after 22nd, and hence both the proposed model and the Bayesian approach are not good in prediction at the rear of data set as was already mentioned by Kuo and Yang(1996).

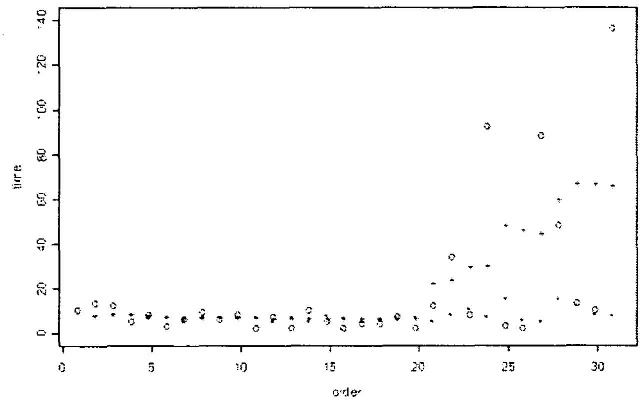

Figure 4.1. Plot of One-Step-Ahead Mean Times Against the Observed Orders for the Data of NTDS ; the circle(o) denotes the observed failure times, and the plus(+) denotes the ones by the proposed method, and the star(*) corresponds to Bayesian model

Table 4.1. Predicted One-step-Ahead Interfailure Times

serial no	Observed time	proposed model	Bayesian model
1	9.00	NA	NA
2	12.00	5.68	4.84
3	11.00	6.24	5.39
4	4.00	6.30	5.21
5	7.00	5.68	3.64
6	2.00	5.52	4.41
7	5.00	5.08	2.92
8	8.00	4.93	3.92
9	5.00	4.97	4.63
10	7.00	4.87	3.92
11	1.00	4.87	4.41
12	6.00	4.63	2.41
13	1.00	4.61	4.18
14	9.00	4.43	2.41
15	4.00	4.52	4.84
16	1.00	4.46	3.64
17	3.00	4.31	2.41
18	3.00	4.24	3.31
19	6.00	4.17	3.31
20	1.00	4.18	4.18
21	11.00	20.27	2.41
22	33.00	21.56	5.21
23	7.00	27.60	8.03
24	91.00	27.65	4.41
25	2.00	45.85	12.36
26	1.00	44.06	2.92
27	87.00	42.22	2.41
28	47.00	57.93	12.27
29	12.00	64.94	9.29
30	9.00	64.60	5.39
31	135.00	63.71	4.84

5. SUMMARY AND FUTURE RESEARCH

We briefly reviewed several types of software reliability models and mainly confined our attention to the modeling in terms of failure rate function. Among which we assume the NHPP from the Weibull distribution. The failure rate function is adapted by the empirical intensity until the previous failure. Additionally we suggested change-point model by allowing the parameters to change after a certain failure observation. This type of change-point model seems to be reasonable because we can expect an improvement on software reliability after a certain bug, which may affect great contribution to a running of software, is fixed. This is an advantage of change-point model against a simple adaptive model not considering a change-point.

We applied the proposed model to a data set of 31 failure times from NTDS which was previously analyzed by Kuo and Yang(1996) in Bayesian framework under the Weibull process assumption. The performance of the proposed model against that of Kuo and Yang(1996) is moderately good except few observations in the rear of data set. Finding alternative models which fit well for the anomalous observations in the data set is remained as a future research work.

ACKNOWLEDGEMENT

This paper was supported by the Ministry of Information and Communication, Korea.

REFERENCES

Al-Mutairi, D. and Chen, Y. and Singpurwalla, N.D.(1998), An Adaptive Concatenated Failure Rate Model for Software Reliability, *Journal of the American Statistical Association*, **93**, 1150-1163.

Duane, J.T.(1964), Learning Curve Approach To Reliability Monitoring, *IEEE Transactions on Aerospace*, **AS-2**, 563-566.

Goel, A.K. and Okumoto, K.(1978), *Bayesian Software Prediction Models: An Imperfect Debugging Model for Reliability and Other Quantitative Measures of Software Systems*, RADC-TR-78-155, Rome Air Development Center.

Jelinski, Z. and Moranda, P.B.(1972), Software Reliability Research, in *Statistical Computer Performance Evaluation*, Academic Press, 465-484.

Joe, H.(1989), Statistical Inference for General Order Statistics and Nonhomogeneous Poisson Process Software Reliability Models, *IEEE Transaction On Software Engineering*, **15**, 1485-1490

Knafl, G.J. and Morgan, J.(1996), Solving ML Equations for 2-Parameter Poisson Process Models for Ungrouped Software-Failure Data, *IEEE Transactions on Reliability*, **45(1)**.

Kuo, L. and Yang, T.Y.(1996), Bayesian Computation for Nonhomogeneous Poisson Processes in Software Reliability, *Journal of the American Statistical Association*, **91**, 763-773.

Langberg, N. and Singpurwalla, N.D.(1985), A Unification of Some Software Reliability Models, *SIAM Journal on Scientific and Statistical Computing*, **6**, 781-790.

Littlewood, B.(1981), Stochastic Reliability-Growth: A Model for Fault-Removal in Computer-Programs and Hardware-Designs, *IEEE Transactions on Reliability*, **30**, 313-320.

Mazzuchi, T.A. and Soyer, R.(1988), A Bayes-Empirical Bayes Model for Software Reliability, *IEEE Transactions on Reliability*, **37**, 248-254.

Musa, J.D.(1975), A Theory of Software Reliability and its Application, *IEEE Transaction on Software Engineering*, **1**, 312-327.

Musa, J.D. and Okumoto, K.(1984), A Logarithmic Poisson Execution Time Model for Software Reliability Measurement, in *Proceedings of the Seventh International Conference On Software Engineering*, 230-238.

Raftery, A.E.(1987), Inference and Prediction for a General Order Statistic Model With Unknown Population Size, *Journal of the American Statistical Association*, **82**, 1163-1168.

Singpurwalla, N.D. and Wilson, S.P.(1994), Software Reliability Modeling, *International Statistical Review*, **62**, 289-317.

Singpurwalla, N.D. and Wilson, S.P.(1999), *Statistical Methods in Software Engineering*, Springer.