The Korean Communications
in Statistics Vol. 4, No. 1, 1997
pp. 259-269

A Bayesian Approach for Record Value Statistics Model Using Nonhomogeneous Poisson Process

Kiheon Choi[1)], Hee-chual Kim[2)]

Abstract

Bayesian inference for a record value statistics(RVS) model of nonhomogeneous Poisson process is considered. We deal with Bayesian inference for double exponential, Gamma, Rayleigh, Gumble RVS models using Gibbs sampling and Metropolis algorithm and also explore Bayesian computation and model selection.

1. Introduction

The nonhomogeneous Poisson process(*NHPP*) is widely used in modeling the number of failures $M(t)$ in $(0, t]$for a repairable system and for software reliability(cf. Ascher & Feingold, 1984; Musa, Iannino & Okumoto, 1987; Crowder, Kimber, Smith & Sweening, 1991). The intensity function(rate of occurrence of failure, ROCOF) $\lambda(t) = dE[M(t)]/dt$ is often assumed to be a monotonic function of t. For examples, the homogeneous Poisson, the Musa-Okumoto(1984), the Weibull (Power law) and the Cox-Lewis(1966) processes assume that the ROCOF is a constant function, a fraction function, a power function and a log-linear function of time, respectively. To relax the monotonic ROCOF assumption, we consider in this paper a RVS model and develop Bayesian inference and model selection methodologies for a RVS model.

We model the failure time by two different classes. One is generalized order statistics(GOS), and the other is record value statistics(RVS). In the GOS model, we assume there is an unknown number of faults N at the beginning of software testing. We model the observed epochs of failures to be the first n order statistics taken from N $i.i.d$ observations with density f supported in R^+. The GOS model is limited to testing whether no new fault is introduced at each repair.

To incorporate the situation that new faults may be added during repairs, we use the RVS model where we assume the epochs of failures to be the record breaking statistics of $i.i.d.$

1) Assistant Professor, Dept. of Statistics, Duksung Women's University, Seoul, Korea
2) Lecturer, Dept. of Statistics, Dongguk University, Seoul, Korea

observations taken from f.

In addition to modeling the failure times, we can also consider its dual $M(t)$ i.e., the number of failures discovered in the interval $(0, t]$. Let $m(t) = EM(t)$, a nondecreasing function of t. For the RVS model, we can relate $M(t)$ to *NHPP*.

We assume $m(t)$ is differentiable. We use $\lambda(t) = dm(t)/dt$ to denote the intensity function or ROCOF for the *NHPP*.

We suppose $m(t)$ is indexed by the unknown parameters β, where β can be a vector of unknown parameters. For simplicity, we write $m(t) = m(t|\beta)$. Given the time truncated model testing until time t, the ordered epochs of the observed n failure times are denoted by $x_1, x_2, \cdots, x_n$. Therefore, the data set D_t consists of $\{n, x_1, x_2, \cdots, x_n; t\}$, Given the failure truncated model observed until n failures, the data set D_{x_n} consists of $\{x_1, x_2, \cdots, x_n\}$.

The likelihood for the time truncated model is

$$L_{NHPP}(\beta | D_t) = [\Pi_{i=1}^{n} \lambda(x_i)] \exp(-m(t)). \tag{1.1}$$

It is developed in many textbooks, for example, Cox and Lewis(1966), Crowder et al(1991). For the failure truncated model, a similar expression (1.1) can be applied with t replaced by x_n. There are two types of *NHPP* processes according to the limiting behavior of $m(t)$:

(1) $NHPP(1)$, if $\lim_{t\to\infty} m(t) < \infty$;

(2) $NHPP(2)$, if $m(t) \to \infty$ as $t \to \infty$.

In this paper, we only consider $NHPP(2)$ model. Section 2 develops the likelihood of $NHPP(2)$ processes and clarifies the relationships among the RVS and $NHPP(2)$. Section 3 develops the Gibbs algorithms for evaluating the posterior distributions. Section 4 discuss methods for the Bayesian inference and model selection. A numerical example is given and some concluding remarks are given in Section 5.

2. Record Value Statistics and *NHPP*(2)

In the RVS model, we assume the epochs of failures are the record values of *i.i.d.* random variables distributed according to a density $f(S|\beta)$. Let $S_1, S_2, \cdots$ denote the *i.i.d.* random variables distributed according to $f(S|\beta)$. We define the sequence of record values, $\{X_n\}_{n\geq 1}$ by

$$R_1 = 1,$$

$$R_{k+1} = \min\{i : i > R_k,\ S_i > S_{R_k}\} \quad \text{where, } k = 2, 3, \ldots.$$

$$X_n = S_{R_n} \quad \text{where, } n \geq 1.$$

We model the observed epochs failures $x_1 < x_2 < \cdots$ as the record values $X_1, X_2, \cdots$. Glick(1978) showed that this sequence of record values can be infinite. Therefore, we can model failure times when the number of failures is unbounded as $t \to \infty$.

To relate the RVS model to the *NHPP*, We need the following theorem that follows from Dwass(1964). Suppose F is continues with full support on R^+. Then the record values constructed above with values in $(0, t]$ are the points of a *NHPP* in $(0, t]$ with the mean measures

$$m(t) = -\ln(1 - F(t)).$$

Consequently, the ROCOF of the *NHPP* is $\lambda(t) = m'(t) = f(t)/(1 - F(t))$, which is also a hazard function of F. It is straightforward to verify that $f(t)/(1-F(t))$ is β, $\alpha/(\beta+t)$, $\beta\alpha t^{\alpha-1}$, $\exp(\alpha+\beta t)$, and $\beta_2^2 t/(\beta_2 t+1)$, for the exponential, Pareto, Weibull, truncated extreme value, and Erlang distributions, respectively. Therefore, the point processes associated with the RVS model constructed from the exponential, Pareto, Weibull, truncated extreme value, and Erlang densities are exactly the homogeneous Poisson process, the Musa-Okumoto, the Weibull(Duane), Cox & Lewis, and Erlang processes, respectively where the latter four are the nonhomogeneous Poisson process.

Now let us consider the likelihood function. Plugging in $\lambda(t) = m'(t) = f(t)/(1-F(t))$ in (1.1), the likelihood *NHPP* is

$$L_{NHPP}(\beta | D_t) = [\Pi_{i=1}^{n} f(x_i|\beta)/(1 - F(x_i|\beta))](1 - F(t|\beta)) \qquad (2.1)$$

For the failure truncated model, we obtain the likelihood $L_{NHPP}(\beta | D_{x_n})$, similar to (2.1) with t replaced by x_n.

3. Gibbs sampling for RVS models

In this section, we develop the conditional distributions used in the Gibbs sampling algorithm. Gibbs sampling is Markov Chain Monte Carlo(MCMC) technique. The stationary distribution of this Markov chain is the posterior distribution we desire. we can also replicate this chain with independent starting points to obtain multiple samples from the posterior distribution. Please refer to Gelfand & Smith(1990), Kuo & Yang(1993) for detailed discussions of Gibbs sampling. We used the likelihood for the time truncated situation in the *NHPP*. For the failure truncated case, it suffices to change t to x_n. The Metropolis algorithm is often used in the Gibbs algorithm. In the prior specifications, we use the notation $X \perp Y$ to mean X and Y are independent, $\Gamma(a, b)$ to denote the gamma distribution with mean a/b. The vector $\underset{\sim}{\beta}$ denotes either β or (α, β).

3.1 Double exponential RVS model

The likelihood function is

$$L_{NHPP(2)}(\beta|D_t) = 1/2\, \beta^n \exp(-\beta t).$$

Prior: $\beta \sim \Gamma(a, b)$.

$$\begin{aligned} p(\beta \mid D_t) &\sim L(\beta|D_t) \cdot p(\beta) \\ &\sim 1/2\, \beta^n \exp(-\beta t) \cdot \frac{b^a}{\Gamma(a)} \beta^{a-1} \cdot e^{-b\beta} \\ &\sim \beta^{a-1+n} \cdot e^{-\beta(b+t)} \cdot \text{constant} . \end{aligned}$$

Posterior distribution: $\beta|D_t \sim \Gamma(n+a, t+b)$.

3.2 Gamma-2 RVS model

The likelihood function is

$$L_{NHPP(2)}(\theta|D_t) = \left(\Pi_{i=1}^{n} \frac{\beta^2 x_i}{\beta x_i + 1} \right) e^{-\beta t}(1+\beta t).$$

Prior: $\beta \sim \pi_1$, where π_1 is a prior on $\beta > 0$.

Posterior distribution:

$$\beta|D_t \propto \beta^{2n} \cdot e^{-\beta t}(1+\beta t) \cdot \Pi_{i=1}^{n} \frac{x_i}{(1+\beta x_i)} \times \pi_1(\beta) ,$$

where β can be generated by the Mertopolis algorithm .

3.3 Rayleigh RVS model

The likelihood function is

$$L_{NHPP(2)}(\beta|D_t) = (\Pi_{i=1}^{n} 2\beta x_i) \exp[-\beta t^2].$$

Prior: $\beta \sim \Gamma(a, b)$.

Posterior distribution: $\beta|D_t \sim \Gamma(n+a, t^2+b)$.

3.4 Gompertz RVS model

The likelihood function is

$$L_{NHPP(2)}(\beta|D_t) = \{ \Pi_{i=1}^{n} [\exp(\beta + \alpha x_i]\} \exp\left(\frac{-e^{\beta+\alpha t} + e^{\beta}}{\alpha}\right).$$

Prior: $\alpha \sim \pi_2(\beta)$, $\beta \sim \pi_3(\beta)$, where π_2 and π_3 are priors for $-\infty < \alpha < \infty$ and $\beta > 0$, respectively, and $\alpha \perp \beta$.

Gibbs algorithm :

$$\beta|\alpha, D_t \propto \exp(n\beta - e^{\beta+\alpha t} + e^{\beta})\, \pi_2(\beta) ;$$

$$\alpha|\beta, D_t \propto \exp(\alpha \sum x_i - e^{\beta+\alpha t} - \alpha)\, \pi_3(\alpha) ,$$

where $\alpha(\beta)$ can be generated by the Metropolis algorithm.

3.5 Gumble RVS model

The likelihood function is

$$L_{NHPP(2)}(\beta|D_t)=\{\Pi_{i=1}^{n}\ \beta e^{-\beta x_i}\}\exp(-\ e^{t\beta}).$$

Prior: $\beta \sim \pi_4(\beta)$, where π_4 is prior $\beta > 0$.

Posterior distribution: $\beta|D_t \propto \beta^n \exp[\beta \sum x_i + e^{t\beta}] \times \pi_4(\beta)$,

where β can be generated by the Metropolis algorithm.

4. Bayesian Inference and Model Selection for *NHPP*(2)

We are also interested in predicting the mean time between failures, estimating the conditional predictive ordinates and the future reliability function. For inferences on the parameters, such as N and β, we can use the empirical measure of the Gibbs samplers. It would be easier to consider the failure truncated situation, i.e., testing until the n^{th} failure. All the posterior distributions described in Section 4 need to be modified by replacing t with x_n. Moreover, instead of the full data D_{x_n}, we can consider the sequential setup, where Bayesian inference is applied for each $D_{x_n}=(x_1, x_2, \cdots, x_i)$, $i=1, \cdots, n$.

Inference for the future survival function evaluated at x distance away from x_i can be obtained by(cf. Cinlar 1975, p.97)

$$\begin{aligned} E(S(x)|D_{x_i}) &= E[E(P(X_{i+1} > x_i+x)|\beta, D_{x_i})|D_{x_i}] \\ &= E[\exp(-m(x_i+x)+m(x_i))|D_{x_i}] . \end{aligned} \tag{4.1}$$

The Prequential Conditional Predictive Ordinate(PCPO) for the future epoch x_{i+1} is defined to be $c_{i+1} = p(x_{i+1}|D_{x_i})$, the conditional density of X_{i+1} evaluated at the future observed time x_{i+1} given $(x_1, \cdots, x_i)$. The PCPO is particularly relevant for model selection because it assesses the predictability of X_{i+1} given the past data. Given the sequence $\{X_i\}_{i\ge 1}$, we can write $c_{i+1} = p(x_{i+1}|x_i)$. The PCPO can be computed by

$$\begin{aligned} p(x_{i+1}|x_i) &= \int p(x_{i+1}|\beta, D_{x_i})p(\beta|D_{x_i})d\beta \\ &= \int \lambda(x_{i+1})\exp(-m(x_{i+1})+m(x_i))p(\beta|D_{x_i})d\beta . \end{aligned} \tag{4.2}$$

Both (4.1) and (4.2) can be evaluated using the Gibbs samplers or Rao-Blackwell versions. We will illustrate the Monte Carlo integration methods using (5.2). We apply Gelman & Rubin (1992) approach in the Monte Carlo method.

Let $\alpha^{(r,s)}$ and $\beta^{(r,s)}$ denote the Gibbs sampler for α and β, respectively generated in the

r^{th} replication and the s^{th} iteration of the MCMC algorithm. We can derive an estimator of (4.2) using direct substitution from Gibbs samplers with S a sufficiently large even number.

We consider Rayleigh RVS model, then

$$\hat{p}(x_{i+1} | x_i) = \frac{2}{RS} \sum_{r=1}^{R} \sum_{s=\frac{S}{2}+1}^{S} 2\beta^{(r,s)} x_{i+1} \exp (\beta^{(r,s)} x_{i+1}^2 - \beta^{(r,s)} x_i^2). \qquad (4.3)$$

We have presented several $NHPP(2)$ models. Which model or models is or are appropriate? We address this issue using the Bayesian PCPO. The PCPO for x_{i+1} is defined by $c_{i+1} = p(x_{i+1} | x_1, \cdots, x_i)$ for $i \ge 2$. Usually, we select diffuse priors to let the data do the talking. We can plot c_i over i, $i = 2, 3, \ldots, n$, for each model. The bigger the c are on the average, the better the model. Moreover, the plot may reveal outliers, if any, where the c values are unusually small. The Bayesian predictive likelihood criterion is to select the model that maximizes the predictive density $C(l) = \Pi_{i=1}^{n} c_i(l) = p(x_1, x_2, \cdots, x_n)$, where l is the index on the models. This computation is straightforward from (4.2).

The PCPO defines the conditional predictive ordinates(CPO) of x_{i+1} by conditioning on the set of past failure epochs $x_1, x_2, \cdots, x_n$.

5. Numerical Examples

The data set of 26 interfailure times listed in Table 1 were based on the trouble report for one of the large modules of the Naval Tactical Data System. Geol & Okumoto(1979), Kuo & Yang(1995) also use it for illustration. We found 2000 iterations to be enough even for the most diffuse prior being considered. All the following numerical results are obtained with 2000 iterations and 70 replications in the Gibbs sampler. We monitor the convergence of the Gibbs sampler using the Gelman & Rubin(1992) method that uses the analysis of variance technique to determine whether further iterations are needed.

The priors on β are $\beta \sim \Gamma(1, 0.0001)$ and $\pi(\beta) = 1, \beta > 0$. We find $c(i)$ are outliers for $i = 22, 24$. Therefore, they are deleted from the calculations in $C(l)$. We use C' to denote the modified product. Table 2 lists densities and mean functions. Table 3 lists Posterior mean, using FORTRAN IMSL. Figure 1 expresses predictive reliability function. Model selection is given Table 4, using PCPO. Figure 2 expresses PCPO. Overall, double exponential and Rayleigh RVS models are preferred.

Error no.	Interfailure time	Error no.	Interfailure time
1	9	14	9
2	12	15	4
3	11	16	1
4	4	17	3
5	7	18	3
6	2	19	6
7	5	20	1
8	8	21	11
9	5	22	33
10	7	23	7
11	1	24	91
12	6	25	2
13	1	26	1

<Table 1> Software Failure Data

	Density Function	RVS/NHPP(2) (mean function)
Dou EXP	$1/2\beta \exp(-\lvert t \rvert \beta)$	$\beta t - \log(1/2)$
Gamma-2	$\beta^2 t \exp^{-\beta t}$	$t\beta - \log(1+\beta t)$
Rayleigh	$2\beta t \exp(-\beta t^2)$	βt^2
Gumble	$\beta e^{\beta t} \cdot \exp(-e^{\beta t})$	$e^{t\beta}$
Gompertz	$\exp\left\{(\beta+\alpha t) - \frac{1}{\alpha}\exp(e^{\beta+\alpha t} - e^{\beta})\right\}$	$\frac{e^{\beta+\alpha t} - e^{\beta}}{\alpha}$

<Table 2 > Density and Mean function

1. Double exponential RVS model

S	R	$\hat{\beta}$
500	50	0.107165
	70	0.106772
2000	50	0.106609
	70	0.107789

2. Gamma-2 RVS model

S	R	$\hat{\beta}$
500	50	0.009285
	70	0.000993
2000	50	0.001039
	70	0.000971

3. Rayleigh RVS model

S	R	$\hat{\beta}$
500	50	0.000431
	70	0.000436
2000	50	0.000433
	70	0.000429

4. Gumble RVS model

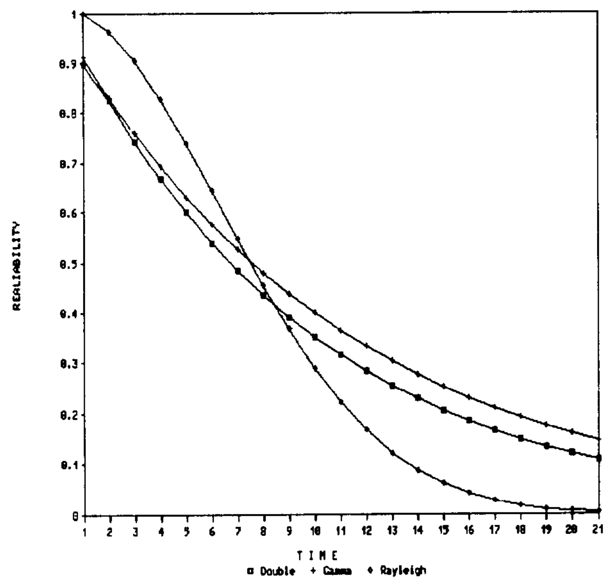
S	R	$\hat{\beta}$
500	50	0.006144
	70	0.006143
2000	50	0.006144
	70	0.006143

<Table 3> Posterior Mean $\hat{\beta}$

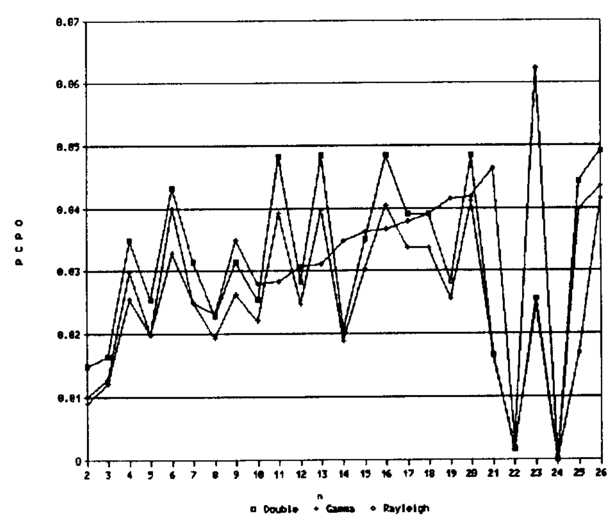
($NHPP(2)$, R^{th} replication and S^{th} iteration)

	$RVS / NHPP(2)$		
	Double exponential	Gamma-2	Rayleigh
$\ln C(l)$	-67.5	-71.2	-69.3

<Table 4> Model Selection



<Figure 1> *NHPP*(2) $\hat{p}(x|D_x)$



<Figure 2> *NHPP*(2) PCPO

References

[1] Aitkin, M. (1991), Posterior Bayes Factors, *Journal of the Royal Statistical Society, B* 111-142.

[2] Ascher, H. and Feing, H (1984), *Repairable Systems Reliability*, New York: Marcel Dekker.

[3] Casella, G. and George, E. I. (1992), Explaining the Gibbs Sampler, *The American Statistician*, 46, 167-174.

[4] Cinlar, E. (1975), *Introduction to Stochastic Process*, New Jersey: Prentice-Hall.

[5] Cox, D. R. and Lewis, P. A. (1966), *Statistical Analysis of Series of Events,* London: Methuen.

[6] Crowder, M. J., Kimber, A. C., Smith, R. L and Sweening, T. J. (1991), *Statistical Analysis of Reliability*, London: Chapman and Hall.

[7] Dawid, A. P. (1984), Statistical Theory: The Prequential Approach, *Journal of the Royal Statistical Society, Ser. A*, 147, 278-292.

[8] Gelfand, A. E., Dey, D. K., and Chang, H. (1992), Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods (with discussion), *in Bayesian Statistics* 4, eds. J. M. Bermardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Cambridge, MA; Oxford University Press, pp. 147-169.

[9] Gelfand, A. E. and Smith, A. F. M. (1990), Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, 85, 398-409.

[10] Gelman, A. E. and Rubin D. (1992), Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 7, 457-472.

[11] Glick, N. (1978), Breaking Records and Breaking Boards, *American Mathematical Monthly*, 85, 2-26.

[12] Goel, A. L. and Okumoto, K. (1979), Time Dependent Error Detection Rate Model for Software Reliability and Other Performance Measures, *IEEE Transactions on Reliability*, 28, 206-211.

[13] Gaver, D. P. and Acar, M. (1979), Analytical Hazard Representations for Use in Reliability, Mortality, and Simulation Studies, *Communications in Statistics, Simulation and Computation*, B8, 91-111.

[14] Hammersley, J. M. and Handscomb, D. C. (1964), *Monte Carlo Methods*, London: Chapman and Hall.

[15] Hastings, W, K., (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, 87, 97-109

[16] Jelinski, Z. and Moranda, P. B. (1972), Software Reliability Research, *in Statistical Computer Performance Evaluation*, ed. W. Freiberger, New York: Academic Press, pp. 465-497.

[17] Kuo, L. and Yang, T. (1993), Bayesian computation of Software Reliability, *Technical Report No.* 83-18, University of Connecticut, Statistics Dept.

[18] Kuo, L. and Yang, T. Y. (1995), Bayesian Computation of Software Reliability, *Journal of Computational and Graphical Statistics,* 65-82.

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

[20] Musa, J. D. and Okumoto, K. (1984), *A Logarithmic Poisson Execution Time Model for Software Reliability Measurement*, in Proceedings Seventh International Conference on Software Engineering Orlando, pp. 230-238.

[21] Musa, J. D., Iannino, A. and Okumoto, K. (1987), *Software Reliability: Measurement, Prediction, Application*, New York: McGraw Hill.

[22] Meinhold, R. J. and Singpurwalla, N. D. (1983), Bayesian Analysis of a Commonly Used Model for Describing Software Failures, *The Statistician*, 32, 168-173.