Bayesian Computation for the Superposition of Nonhomogeneous Poisson Processes

Tae Young Yang and Lynn Kuo

Myongji University and University of Connecticut

Abstract

Bayesian inference for the superposition of nonhomogeneous Poisson processes is studied. A Markov chain Monte Carlo method with data augmentation is developed to compute the features of the posterior distribution. For each observed failure epoch, a latent variable is introduced that indicates which component of the superposition model gives rise to the failure. This data augmentation approach facilitates specification of the transitional kernel in the Markov chain. Moreover, new Bayesian tests are developed for the full superposition model against simpler submodels. Model determination by a predictive likelihood approach is studied. A numerical example based on a real data set is given.

Key words and phrases: Additive intensity function, Data augmentation, Gibbs sampling, Metropolis algorithm, Model selection, Predictive reliability function.

AMS 1991 subject classifications: Primary 62F15, secondary 62M20.

Abbreviated Title: Superposed Poisson Processes

1. Introduction

The nonhomogeneous Poisson process is widely used in modeling the number of failures in the time interval $(0, t]$ for a repairable system and for software reliability. The intensity function of the nonhomogeneous Poisson process is often assumed to be a monotonic function of t. Such is the case, *inter alia* for the homogeneous Poisson, the Musa-Okumoto (1984), the Weibull (Power Law), and the Cox-Lewis (1966) processes, whose intensity function is respectively constant, a fractional function, a power function, and a log-linear function of time. To relax the monotonic intensity function assumption, we consider in this paper a model that superimposes several independent nonhomogeneous Poisson processes with simple intensity functions.

Let $M_j(t)$ denote the nonhomogeneous Poisson process that counts the number of failures from the j^{th} component in $(0, t]$, with intensity function $\lambda_j(t|\boldsymbol{\theta}_j) = dE\{M_j(t)\}/dt$, where the functional form of $\lambda_j(t|\boldsymbol{\theta}_j)$ is assumed to be specified but with unknown parameter $\boldsymbol{\theta}_j$ that may be a vector. Moreover, the numbers of failures from the j^{th} component $M_j(t), j = 1, \dots, J$, are assumed to be independent. Then a process $M(t) = \sum_{j=1}^{J} M_j(t)$, which counts the number of failures in the interval $(0, t]$ for the superposition model, is also a nonhomogeneous Poisson process with intensity function $\lambda(t|\boldsymbol{\theta}) = dE\{M(t)\}/dt = \lambda_1(t|\boldsymbol{\theta}_1) + \cdots + \lambda_J(t|\boldsymbol{\theta}_J)$, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J\}$ (cf. Çinlar, pp. 87-88). In our notation, we often omit $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_j$ for simplicity.

The family of intensity functions for the superposition model is richer than that of its components. It includes bathtub-shaped functions, polynomials with peaks and valleys, and simple monotonic functions. For example, the intensity function for a superposition model with two nonhomogeneous Poisson processes with Weibull intensities is $\lambda(t) = \alpha_1\beta_1 t^{\alpha_1 - 1} + \alpha_2\beta_2 t^{\alpha_2 - 1}$. If $\alpha_1 < 1$ and $\alpha_2 > 1$, then the intensity function has a bathtub shape. When $\alpha_1 = \alpha_2$, the intensity function of the superposition model reduces to that of a simple Weibull intensity, and thus β_1 and β_2 are not separately identifiable. However, Bayesian analysis with continuous prior distribution assigns zero probability to the case $\alpha_1 = \alpha_2$. Therefore, the identifiability issue in β_1 and β_2 is not a serious problem in common Bayesian analysis. On the other hand, if one assumes a superposition model and the data actually come from a nonhomogeneous Poisson process with a simple intensity function, then the posterior distributions of α_1 and α_2 would heavily overlap and the separation in β_1 and β_2 is only inherent in the prior distribution. To avoid this problem, we assume that the number of components J is known. A similar identifiability issue has been addressed in Berger and Sun (1993). They study inference for poly-Weibull densities. In fact, we could also reproduce poly-Weibull densities in our framework. Let $\lambda(t) = \sum_{j=1}^{J} \alpha_j \beta_j t^{\alpha_j - 1}$ for the intensity of a superimposed model of J Weibull nonhomogeneous Poisson processes. Applying the unified approach of Kuo and Yang (1996), the points of failures for the superposition model are the record value statistics derived from an i.i.d. unobserved sample distributed with $f(t) = \lambda(t) \exp\{-\int_0^t \lambda(u) du\}$. Then $f(t) = (\sum_{j=1}^{J} \alpha_j \beta_j t^{\alpha_j - 1}) \exp(-\sum_{j=1}^{J} \beta_j t^{\alpha_j})$. This is the poly-Weibull density considered by Berger and Sun except we consider different likelihoods. See Yang and Kuo (1995) for more discussion of the different likelihoods.

The major contribution of this paper is to show that Bayesian inference for the superposition model is accessible now by means of a Markov chain Monte Carlo technique with data augmentation. The technique consists of iteratively generating latent variables using multinomial distributions and applying the technique in Kuo and Yang (1996) to update the parameters given the latent variables.

In the full model, we assume none of the λ_j's can be zero. Therefore, continuous priors on the unknown parameters suffice. In order to incorporate nested submodels with positive prior probability, we consider a mixture of discrete and continuous (discrete/continuous) distributions for λ_j. That is, λ_j equals zero with prior probability p_{0j} and $\lambda_j > 0$ is distributed according to a continuous density $\pi_j(\lambda_j)$ on $(0, \infty)$ with probability $1 - p_{0j}$. When λ_j is indexed by $\boldsymbol{\theta}_j$ (vector), it is often sufficient to assign a discrete/continuous prior to just one of the parameters in $\boldsymbol{\theta}_j$. Consequently, we develop a Gibbs sampler for the Bayesian test of the full model against a nested submodel.

In addition to the Bayesian test, we also discuss the predictive likelihood method for model selection among the superposition model and all its submodels. We split the data into two parts "future" and "past". We select the model that maximizes the predictive likelihood of the future given the past.

Gibbs algorithms are given in Section 2 for evaluating the posterior distributions for the full model. In Section 3, the Gibbs algorithm is extended to develop a Bayesian test for choosing between a full model and a reduced model. Issues for model determination are discussed in Section 4, and a numerical example is presented in Section 5.

2. Gibbs Sampling for the Full Model

In this section, we develop the conditional distributions used in the Gibbs sampler for the full model. Please refer to Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), Tierney (1994), Casella and George (1992), Chen and Singpurwalla (1994), Kuo and Yang (1995, 1996), and Kuo *et al.* (1997) for detailed discussions of Gibbs sampling.

Suppose that the software is to be tested until time t. The ordered epochs of the observed n failure times are denoted by $x_1, \ldots, x_n$. Therefore the data set D_t consists of $\{n, x_1, x_2, \ldots, x_n; t\}$. The likelihood function of the nonhomogeneous Poisson process given the data for the time truncated model is

$$L(\boldsymbol{\theta}|D_t) = \left\{ \prod_{i=1}^{n} \lambda(x_i) \right\} \exp\{-m(t)\}, \tag{2.1}$$

where $m(t) = \int_0^t \lambda(u)du$, and $0 < x_1 < \cdots < x_n < t$. This likelihood is developed in many textbooks, for example, Cox and Lewis (1966), Basawa and Prakasa Rao (1980), and Crowder *et al.* (1991). Given the failure truncated model (observed until the n^{th} failure), the data set D_{x_n} consists of $\{x_1, x_2, \ldots, x_n\}$. Then the likelihood is equal to (2.1) with t replaced by x_n.

We assume the prior has independent components for the $\boldsymbol{\theta}_j$s. The posterior density of $\boldsymbol{\theta}$ is

$$f(\boldsymbol{\theta}|D_t) \propto \left[\prod_{i=1}^{n} \{\lambda_1(x_i) + \cdots + \lambda_J(x_i)\} \right] \exp\left[- \{m_1(t) + \cdots + m_J(t)\} \right] \times \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j), \tag{2.2}$$

where $\pi(\boldsymbol{\theta}_j)$ denotes the prior marginal density of $\boldsymbol{\theta}_j$.

The expression $\prod_{i=1}^{n} \{\lambda_1(x_i) + \cdots + \lambda_J(x_i)\}$ in (2.2) usually prevents us from specifying a convenient form for the conditional density of $\boldsymbol{\theta}$ given D_t. We, therefore, introduce latent variables, denoted by $\mathbf{I}$, so we can obtain the posterior distribution of $\boldsymbol{\theta}$ given D_t by iteratively sampling $\mathbf{I}$ from $f(\mathbf{I}|\boldsymbol{\theta}, D_t)$ and sampling $\boldsymbol{\theta}$ from $f(\boldsymbol{\theta}|\mathbf{I}, D_t)$. We use the notation f to denote a generic name for the densities.

For each i, $i = 1, \ldots, n$, let us consider the latent variable $\mathbf{I}_i = (I_{i1}, \ldots, I_{iJ})^T$, where $I_{ij} = 1$ if the i^{th} failure is caused by the j^{th} type and $I_{ij} = 0$ otherwise. Note $\sum_{j=1}^{J} I_{ij} = 1$ for $i = 1, \ldots, n$. Let the conditional density of $\mathbf{I}_i$ given $\boldsymbol{\theta}$ and D_t be a multinomial distribution MN with parameters 1 and the cell probabilities $(p_{i1}, \ldots, p_{iJ})$, where $p_{ij} = \lambda_j(x_i)/\left\{\lambda_1(x_i) + \cdots + \lambda_J(x_i)\right\}$. Let $\mathbf{I} = (\mathbf{I}_1, \ldots, \mathbf{I}_n)$, an $J \times n$ matrix with the i^{th} column indicating which component causes the i^{th} failure. Then we simulate $\mathbf{I}_1, \ldots, \mathbf{I}_n$ independently, where $\mathbf{I}_i \sim MN\{1, (p_{i1}, \ldots, p_{iJ})\}$ for $i = 1, \ldots, n$. The posterior density of $\boldsymbol{\theta}$ given $\mathbf{I}$ and D_t is

$$\begin{aligned} f(\boldsymbol{\theta}|\mathbf{I}, D_t) &\propto L(\boldsymbol{\theta}|D_t) \times \prod_{i=1}^{n} p(\mathbf{I}_i|\boldsymbol{\theta}, D_t) \times \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j) \\ &\propto \prod_{j=1}^{J} \prod_{i:I_{ij}=1} \lambda_j(x_i) \times \prod_{j=1}^{J} \exp\{-m_j(t)\} \times \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j). \end{aligned} \tag{2.3}$$

This conditional density is a joint density composed of independent posterior densities of the $\boldsymbol{\theta}_j$'s given $\mathbf{I}$ and D_t. We can therefore update the $\boldsymbol{\theta}_j$'s independently for $j = 1, \ldots, J$ as in Kuo and Yang (1996), thus providing a practical way of computing the Bayes estimates for superposition models.

We can also use the above algorithm to obtain the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$. This can be done by choosing $\pi(\boldsymbol{\theta}_j) = 1$, for all j, with $\boldsymbol{\theta}_j$ in the suitable domain, and using the posterior mode to estimate $\boldsymbol{\theta}$. Because this flat prior may be improper, the condition $\int L(\boldsymbol{\theta}|D_t) d\boldsymbol{\theta} < \infty$ is required to make this method valid.

Next we provide more detailed algorithms for two examples.

2.1 Superposition of Musa-Okumoto and Weibull Processes

Let the intensities for the two component processes be $\lambda_1(t) = \beta_1/(\alpha_1 + t)$ and $\lambda_2(t) = \alpha_2\beta_2 t^{\alpha_2 - 1}$, where $\alpha_i > 0$ and $\beta_i > 0$ for $i = 1, 2$. Hjorth (1980) points out that this intensity function $\lambda_1(t) + \lambda_2(t)$ has three possible types of shapes: increasing, decreasing, and bathtub. We use $\boldsymbol{\theta}_1$ to denote (α_1, β_1) and $\boldsymbol{\theta}_2$ to denote (α_2, β_2), and $\boldsymbol{\theta}$ to denote $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$.

The likelihood function of the superposition model follows from (2.1),

$$L(\boldsymbol{\theta}|D_t) = \left\{ \prod_{i=1}^{n} \left(\frac{\beta_1}{x_i + \alpha_1} + \alpha_2\beta_2 x_i^{\alpha_2 - 1} \right) \right\} \exp\left\{ -\beta_1 \ln\left(1 + \frac{t}{\alpha_1}\right) - \beta_2 t^{\alpha_2} \right\}. \tag{2.4}$$

We consider the prior that has four independent components: $\alpha_1 \sim \pi(\alpha_1)$; $\beta_1 \sim \Gamma(a_1, b_1)$; $\alpha_2 \sim \pi(\alpha_2)$; $\beta_2 \sim \Gamma(a_2, b_2)$, where $\pi(\alpha_1)$ and $\pi(\alpha_2)$ are arbitrary priors on $\alpha_1 > 0$ and $\alpha_2 > 0$, respectively. We use $\Gamma(a, b)$ to denote the gamma density $f(\beta) = b^a \beta^{a-1} \exp(-b\beta)/\Gamma(a)$. Although the independent marginals and the gamma distribution for the β's are chosen for convenience, this prior is quite flexible. The prior on $\boldsymbol{\theta}_2 = (\alpha_2, \beta_2)$ includes the class of customary priors for the Weibull family as discussed in Sinha (1986, p. 160). In the following we list the conditional densities used in the Gibbs algorithm.

- Construct $\mathbf{I}$ given $\boldsymbol{\theta}$ and D_t by generating independent I_{i1} from the Bernoulli distribution with parameter p_{i1}, for $i = 1, \ldots, n$, where $p_{i1} = \{\beta_1/(x_i + \alpha_1)\}/\{\beta_1/(x_i + \alpha_1) + \alpha_2 \beta_2 x_i^{\alpha_2 - 1}\}$. Let $I_{i2} = 1 - I_{i1}$;
- $\alpha_1 | \beta_1, \alpha_2, \beta_2, \mathbf{I}, D_t \propto \frac{1}{\prod_{i:I_{i1}=1}(x_i+\alpha_1)} \exp\{-\beta_1 \ln(1 + t/\alpha_1)\} \times \pi(\alpha_1)$;
- $\beta_1 | \alpha_1, \alpha_2, \beta_2, \mathbf{I}, D_t \sim \Gamma\{a_1 + \sum_{i=1}^n I_{i1}, b_1 + \ln(1 + t/\alpha_1)\}$;
- $\alpha_2 | \alpha_1, \beta_1, \beta_2, \mathbf{I}, D_t \propto \alpha_2^{\sum_{i=1}^n I_{i2}} (\prod_{i:I_{i2}=1} x_i)^{\alpha_2} \exp(-\beta_2 t^{\alpha_2}) \times \pi(\alpha_2)$;
- $\beta_2 | \alpha_1, \beta_1, \alpha_2, \mathbf{I}, D_t \sim \Gamma(a_2 + \sum_{i=1}^n I_{i2}, b_2 + t^{\alpha_2})$.

The variates α_1 and α_2 can be generated by the Metropolis algorithm with the above expressions in the conditional densities for α_1 and α_2 as the target densities. Kuo and Yang (1995) provide a brief discussion of the Metropolis algorithm (cf. Metropolis *et al.*, 1953).

2.2 Polynomial Intensity Function

Let us consider a nonhomogeneous Poisson process with intensity function $\lambda(t) = \beta_0 + \beta_1 t + \cdots + \beta_{J-1} t^{J-1}$. It can be considered as a superposition of a homogeneous Poisson and $J - 1$ Weibull processes with known shape parameters $2, \ldots, J$. We assume $\beta_j > 0$ for $j = 0, \ldots, J - 1$. For better notation, we are changing the index space on j from 1 to J to 0 to $J - 1$. We can apply the Gibbs sampler developed earlier in this section with $p_{ij} = \beta_j x_i^j / (\sum_{j=0}^{J-1} \beta_j x_i^j)$, for $j = 0, \ldots, J - 1$. Then the conditional posterior density of $\boldsymbol{\theta}$ in (2.3) is

$$f(\boldsymbol{\theta}|\mathbf{I}, D_t) \propto \prod_{j=0}^{J-1} \left\{ \beta_j^{\sum_{i=1}^n I_{ij}} \exp\left(-\frac{t^{j+1}}{j+1} \beta_j \right) \pi(\beta_j) \right\}.$$

Therefore, we can update β_j, $j = 0, \ldots, J-1$, independently given $\mathbf{I}$ and D_t. If $\pi(\beta_j)$ is a gamma density or a flat prior density for β_j, we can easily generate β_j using an updated gamma density.

REMARKS. Bayesian inference for any functionals can be developed easily from the posterior distribution. For example, the predictive reliability function evaluated at $x_n + x$ is

$$\begin{aligned} E(S(x)|D_{x_n}) &= E[E\{P(X_{n+1} - x_n > x)|\boldsymbol{\theta}, D_{x_n}\}|D_{x_n}] \\ &= E[\exp\{-m(x_n + x|\boldsymbol{\theta}) + m(x_n|\boldsymbol{\theta})\}|D_{x_n}] \\ &= \int \exp\{-m(x_n + x|\boldsymbol{\theta}) + m(x_n|\boldsymbol{\theta})\} f(\boldsymbol{\theta}|D_{x_n}) d\boldsymbol{\theta}. \end{aligned} \tag{2.5}$$

Let $\boldsymbol{\theta}^{(r,\ell)}$ denote the Gibbs variate for $\boldsymbol{\theta}$ generated in the r^{th} replication and the ℓ^{th} iteration of the MCMC algorithm based on the whole data D_{x_n}. Then, we can obtain a Monte Carlo estimate of the predictive reliability function in (2.5) for the superimposed Musa-Okumoto and Weibull process with L being a sufficiently large even number:

$$\begin{aligned} \hat{S}(x|D_{x_n}) &= \frac{2}{RL} \sum_{r=1}^{R} \sum_{\ell=\frac{L}{2}+1}^{L} \exp\left\{-m(x_n + x|\boldsymbol{\theta}^{(r,\ell)}) + m(x_n|\boldsymbol{\theta}^{(r,\ell)})\right\} \\ &= \frac{2}{RL} \sum_{r=1}^{R} \sum_{\ell=\frac{L}{2}+1}^{L} \exp\left[-\beta_1^{(r,\ell)} \ln\left(1 + \frac{x_n + x}{\alpha_1^{(r,\ell)}}\right) - \beta_2^{(r,\ell)} (x_n + x)^{\alpha_2^{(r,\ell)}} \right. \\ &\qquad \left. + \beta_1^{(r,\ell)} \ln\left(1 + \frac{x_n}{\alpha_1^{(r,\ell)}}\right) + \beta_2^{(r,\ell)} x_n^{\alpha_2^{(r,\ell)}} \right]. \end{aligned}$$

3. Gibbs Sampling for Nested Models

The priors $\pi(\beta_j)$ given in Subsections 2.1 and 2.2 are assumed to be continuous. They have the drawback that they assign zero probability *a priori* to any reduced model. To incorporate the possibility that a reduced model can be a reasonable contender in model fitting, we need to assign reduced models *a priori* positive probabilities. Therefore, we consider discrete/continuous priors for some parameters in the model.

To present an illustration, we consider the superposition model of Musa-Okumoto and Weibull with intensities $\beta_1/(t + \alpha_1)$ and $\alpha_2 \beta_2 t^{\alpha_2 - 1}$, respectively. Suppose we are interested in comparing the full model and the reduced model with the Musa-Okumoto model only. We can consider the prior for β_2 to be a mixture distribution, where β_2 is equal to 0 with probability p_0 and is distributed according to a continuous density $\pi(\beta_2)$ with probability $1-p_0$. We assume independent continuous priors for the other parameters. So this model assigns *a priori* the reduced model ($\beta_2 = 0$) with probability p_0. Therefore, we would prefer the reduced model ($\beta_2 = 0$) to the full model ($\beta_2 > 0$) provided $p^* = \Pr(\beta_2 = 0|D_t) > 1/2$, where p^* can be approximated from the Gibbs sample.

To describe the Gibbs sampler, let us first consider the conditional density of β_2 given $\beta_1, \alpha_1, \alpha_2$ and D_t. It is a mixture of a point mass at 0 and a continuous density $\pi_c(\beta_2|\alpha_1, \beta_1, \alpha_2, D_t)$ with mass p^{**} and $1-p^{**}$, respectively. The continuous part of the posterior density of β_2, $\pi_c(\beta_2|\alpha_1, \beta_1, \alpha_2, D_t)$, can be shown to be proportional to $L(\boldsymbol{\theta}|D_t)\pi(\beta_2)$, where $L(\boldsymbol{\theta}|D_t)$ is given in (2.4).

Moreover, let

$$m_1(\alpha_1, \beta_1, \alpha_2, D_t) = \int_0^\infty L(\boldsymbol{\theta}|D_t)\pi(\beta_2)d\beta_2,$$

and

$$L(\boldsymbol{\theta}_1|D_t) = L(\boldsymbol{\theta}_1, \alpha_2, \beta_2 = 0|D_t) = \left(\prod_{i=1}^{n} \frac{\beta_1}{x_i + \alpha_1}\right) \exp\left\{-\beta_1 \ln\left(1 + \frac{t}{\alpha_1}\right)\right\},$$

where $\boldsymbol{\theta}_1 = (\alpha_1, \beta_1)$. The Bayes factor (BF) in favor of $H_0 : \beta_2 = 0$ (the Musa-Okumoto model only), against the alternative $H_1 : \beta_2 > 0$ (the full model), given the other parameters α_1, α_2, β_1, and D_t is given by

$$\mathrm{BF} = \frac{L(\boldsymbol{\theta}_1|D_t)}{m_1(\alpha_1, \beta_1, \alpha_2, D_t)}.$$

Then, the posterior probability p^{**} for $\beta_2 = 0$ is given by $p^{**} = p_0/\{p_0 + (1-p_0)/BF\}$. The probability p^{**} is a function of $\alpha_1, \beta_1, \alpha_2$, and D_t. We omit its arguments for simplicity of notation. Most mathematical softwares provide routines to compute numerically one-dimensional integrals such as needed to evaluate m_1. In the absence of such software, we can approximate this integral using numerical integration techniques, for example, the Simpson's rule or the trapezoidal rule.

The MCMC algorithm which consists of simulating each parameter from the full conditional density can be described as follows:

1. Generate β_2 given $\alpha_2, \boldsymbol{\theta}_1$, and D_t:

Simulate $\beta_2 = 0$ or $\beta_2 > 0$ with probabilities p^{**} and $1 - p^{**}$, respectively. If $\beta_2 > 0$, then generate β_2 from the conditional density $\pi_c(\beta_2|\alpha_1, \beta_1, \alpha_2, D_t)$ which is proportional to $L(\boldsymbol{\theta}|D_t)\pi(\beta_2)$. This can be done by a Metropolis algorithm or by following the procedure in Subsection 2.1 with data augmentation.

2. Generate α_2 given $\beta_2, \alpha_1, \beta_1$, and D_t:

   (a) If $\beta_2 = 0$, generate α_2 conditional on α_1, β_1 and D_t from the conditional density $\pi(\alpha_2|\alpha_1, \beta_1, D_t)$ (that has β_2 integrated out). The conditional density $\pi(\alpha_2|\alpha_1, \beta_1, D_t)$ is proportional to $p_0\pi(\alpha_2)L(\boldsymbol{\theta}_1|D_t) + (1 - p_0)\pi(\alpha_2)m_1(\alpha_1, \beta_1, \alpha_2, D_t)$. So α_2 can be sampled by the Metropolis algorithm.

   (b) If $\beta_2 > 0$, simulate α_2 from the conditional density $\pi_c(\alpha_2|\beta_2, \alpha_1, \beta_1, D_t)$ that is proportional to $L(\boldsymbol{\theta}|D_t)\pi(\alpha_2)$. This can be done by a Metropolis algorithm, or by following the procedure in Subsection 2.1 with data augmentation.

3. Simulate α_1 and β_1 given $\boldsymbol{\theta}_2$ and D_t as follows :

   If $\beta_2 > 0$, then generate $\boldsymbol{\theta}_1$ from the full model as in Subsection 2.1; if $\beta_2 = 0$, then generate $\boldsymbol{\theta}_1$ from the reduced model $L(\boldsymbol{\theta}_1|D_t)$ as in Kuo and Yang (1996).

4. Iterate and replicate steps 1–3 to obtain R multiple chains of length L. Then estimate p^* by using the relative frequency of $\beta_2 = 0$ in the second half of the Gibbs sample.

Similarly, we can develop a Bayesian procedure for testing the nested Weibull model versus the full superposition model. We assume that β_1 is assigned a discrete/continuous prior with a continuous density $\pi(\beta_1)$ of prior probability $1 - q_0$ and with a discrete point mass at zero with prior probability q_0. Then we can derive a new test similarly as before, by interchanging β_1 and β_2 and interchanging α_1 and α_2 in steps 1–3 with the appropriate changes on prior and the Bayes factor: $\mathrm{BF} = L(\boldsymbol{\theta}_2|D_t)/\{m_2(\alpha_1, \alpha_2, \beta_2, D_t)\}$, where

$$L(\boldsymbol{\theta}_2|D_t) = \left(\prod_{i=1}^{n} \alpha_2\beta_2 x_i^{\alpha_2-1}\right) \exp\left(-\beta_2 t^{\alpha_2}\right) \text{ and } m_2(\alpha_1, \alpha_2, \beta_2, D_t) = \int_0^\infty L(\boldsymbol{\theta}|D_t)\pi(\beta_1)d\beta_1.$$

Then we use q^* to denote the posterior probability assigned to the Weibull submodel (β_1=0).

If we consider discrete/continuous priors for both β_1 and β_2 simultaneously, this will allow us to select among the Musa-Okumoto, the Weibull, and the superposition models.

However, the independence assumption for the parameters in the prior implies that we choose an intensity to be identically equal to zero *a priori* with probability $p_0 q_0$. It is not a very attractive prior unless we believe the intensity can be identically zero. Dependent priors can be constructed to eliminate this drawback.

4. Model Determination

Although the formulation in Section 3 allows us to select models between a full model and its nested submodels, it is quite complicated even for $J = 2$. Therefore, we propose another model selection method based on the predictive likelihood evaluation. We consider $2^J - 1$ submodels with the following intensity functions:

$$\lambda_1(t), \ldots, \lambda_J(t),\ \lambda_1(t) + \lambda_2(t), \ldots, \lambda_{J-1}(t) + \lambda_J(t), \ldots, \lambda_1(t) + \cdots + \lambda_J(t),\ \text{for } J \geq 2,$$

where none of the λ_j's can be zero.

The predictive likelihood modifies the posterior Bayes factor (Aitkin, 1991). Kuo and Yang (1996) discussed the advantage of using the predictive likelihood for model selection. We divide our data into two parts for a fixed i: the first part, or the "past", consists of the first i failure times; the second part, or the "future", consists of the remaining data. It may be a reasonable decision to define the "future" as the last half portion of the whole data and the "past" as the first half portion of the data. We use $\overline{D}_{x_i}$ to denote the second part, that is, $(x_{i+1}, \ldots, x_n)$. Then we define the predictive likelihood of the "future" $\overline{D}_{x_i}$ given the "past" D_{x_i} for a model denoted by M:

$$\begin{aligned} p_M(\overline{D}_{x_i}|D_{x_i}) &= \int p(\overline{D}_{x_i}|\boldsymbol{\theta}_M, D_{x_i})\pi_M(\boldsymbol{\theta}_M|D_{x_i})d\boldsymbol{\theta}_M \\ &= \int \left\{ \prod_{\ell=i+1}^{n} \lambda(x_\ell|\boldsymbol{\theta}_M) \right\} e^{m(x_i|\boldsymbol{\theta}_M) - m(x_n|\boldsymbol{\theta}_M)} \pi_M(\boldsymbol{\theta}_M|D_{x_i})d\boldsymbol{\theta}_M. \end{aligned} \tag{4.1}$$

The last equality follows from the Markovian property of the inter-arrival times for the nonhomogeneous Poisson process. Here $\boldsymbol{\theta}_M$ is the set of all unknown parameters in the model M. The posterior density of $\boldsymbol{\theta}_M$ given D_{x_i} is π_M. The modified posterior Bayes factor criterion prefers model M_1 to M_2 if the ratio $p_{M_2}(\overline{D}_{x_i}|D_{x_i})/p_{M_1}(\overline{D}_{x_i}|D_{x_i}) < 1$. This ratio can easily be approximated by applying the Monte Carlo integration technique to $p_{M_1}(\overline{D}_{x_i}|D_{x_i})$

and $p_{M_2}(\overline{D}_{x_i}|D_{x_i})$ separately; that is, for $j = 1, 2$, the predictive likelihood $p_{M_j}(\overline{D}_{x_i}|D_{x_i})$ is estimated by averaging $p_{M_j}(\overline{D}_{x_i}|\boldsymbol{\theta}_{M_j}, D_{x_i})$ over the $\boldsymbol{\theta}_{M_j}$s drawn in the second half of the iteration and all replications of the Gibbs sampler based on the data D_{x_i}. For model selection among the $2^J - 1$ models, the modified posterior Bayes factor criterion would choose the best model with the largest $\hat{p}_M(\overline{D}_{x_i}|D_{x_i})$.

We are also interested in the model adequacy issue. For each i, $i = 1, \ldots, n$, the conditional predictive density for the future failure epoch X_{i+1} given $D_{x_i} = (x_1, \ldots, x_i)$ can be computed by

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

Therefore, we can construct a $1 - \alpha$ level prediction interval for X_{i+1} from the Gibbs sample of X_{i+1} using (4.2). For example, the 95% prediction interval for X_{i+1} can be constructed from the 2.5% and 97.5% quantiles of the sample $\{x_{i+1}^{(r,\ell)}\}_{r=1,\ell=L/2+1}^{R\ \ \ L}$ generated from (4.2). We construct intervals for each X_i based on past data $D_{x_{i-1}}$ for $i = 2, \ldots, n$. We judge the model adequate if 95% of the intervals contain the actual observed data x_i.

Now we elaborate on how to generate X_{i+1} from the predictive density $p(X_{i+1}|D_{x_i})$ in (4.2). Observe the expression $\lambda(X_{i+1}) \exp\{-m(X_{i+1}) + m(x_i)\}$ is a density for a truncated variable with value greater than x_i. Therefore, we can simulate X_{i+1} from the truncated density with parameters $\boldsymbol{\theta}^{(r,\ell)}$ that are Gibbs variates from the posterior density $p(\boldsymbol{\theta}|D_{x_i})$. Let G denote the untruncated distribution, where $G(x) = 1 - \exp\{-m(x)\}$ that is a function of $\boldsymbol{\theta}^{(r,\ell)}$. Then x_{i+1} can be generated by setting $x_{i+1}^{(r,\ell)} = G^{-1}[G(x_i) + U^{(r,\ell)}\{1 - G(x_i)\}]$, with $U^{(r,\ell)}$ being the independent uniform variates generated in the ℓ^{th} iteration and r^{th} replication of the Markov chain. When this inverse G^{-1} is complicated to solve explicitly, we can obtain x_{i+1} by solving for it as the root of the equation $G(x) = c$, where $c = G(x_i) + U^{(r,\ell)}\{1 - G(x_i)\}$, a known constant. The root can be solved easily using standard techniques or software. This method is more efficient and therefore to be preferred to the rejection method for generating a truncated random variate. In Section 2, we describe the Gibbs sampler for the whole data D_{x_n}. We can easily change the program to handle D_{x_i}. Then we repeat this for each $i = 1, \ldots, n - 1$.

5. Numerical Examples

To illustrate the MCMC algorithm discussed so far, we consider a data set given in Hossain and Dahiya (1993). It consists of 34 failures epochs in days (9, 21, 32, 36, 43, 45, 50, 58, 63, 70, 71, 77, 78, 87, 91, 92, 95, 98, 104, 105, 116, 149, 156, 247, 249, 250, 337, 384, 396, 405, 540, 798, 814, 849). This data set is based on the trouble report for one of the larger modules of the U.S. Naval Tactical Data System; the first 31 failure epochs of this data set are exactly that used in Kuo and Yang (1996); the first 26 failure epochs of this data set have also been used extensively in the literature. To avoid underflow problems in the numerical computation, we have converted the scale of the data set to 1/10 of its original.

We consider three nonhomogeneous Poisson processes with the following intensity functions: the Musa-Okumoto with $\lambda_1(t) = \beta_1/(t+\alpha_1)$, the Weibull with $\lambda_2(t) = \alpha_2\beta_2 t^{\alpha_2-1}$, and the superposition model with $\lambda(t) = \lambda_1(t) + \lambda_2(t)$, respectively. We monitor the convergence of the Gibbs sampler using the Gelman and Rubin (1992) method that uses the analysis of variance technique to determine whether further iterations are needed. We found 20,000 iterations to be enough for the priors being considered. All the following numerical results are obtained with 20,000 iterations and 50 replications in the Gibbs sampler.

Table 1 summarizes the priors being considered for the full superposition model and for the two nested submodels. We choose $p_0 = 1/2$ for the prior probability assigned to the point mass $\beta_2 = 0$ in the first nested model. Similarly, for the second nested model, we choose $q_0 = 1/2$ for the prior mass assigned to the point mass at $\beta_1 = 0$. With this much prior weight on the reduced model, the algorithm will visit the reduced model a sufficient number of times to obtain an accurate estimate of the posterior probabilities. For the continuous part of the mixture density and other parameters with continuous densities, we always choose priors with moderate variance, so the posterior distribution is relatively stable. Table 1 illustrates the posterior probabilities using the MCMC algorithm. The posterior probability $p^* = 0.147$ in Table 1 indicates that the full superposition model is preferred to the Musa-Okumoto model when we compare these two models. Moreover, the posterior probability $q^* = 0.093$ is solid evidence that the superposition model is much better than the Weibull model.

For model selection based on the predictive likelihood, we also consider the Musa-

Okumoto, the Weibull, and the superposition model. We use the same proper priors as in Table 1. Table 2 lists the logs of the predictive likelihoods for the last half of the data (18^{th} to 34^{th}) with respect to the first half of the data (1^{st} to 17^{th}). The table shows that the superposition model is preferred to both the Musa-Okumoto and the Weibull model. This is consistent with the result from the Bayesian tests for the nested submodels in Table 1. Moreover, this table shows that the Musa-Okumoto model fits the data better than the Weibull model, a result that perhaps cannot be directly obtained from the Bayesian tests developed in Section 3.

Acknowledgments

Yang's research was supported by the Korea Science and Engineering Foundation under contract number 971-0105-027-2.

References

Aitkin, M. (1991). Posterior Bayes factors. *J. Roy. Statist. Soc. Ser. B,* 111-142.

Basawa, I.V., and Prakasa Rao, B.L.S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

Berger, J. O., and Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *J. Amer. Statist. Assoc.,* 88, 1412-1418.

Casella, G., and George, E.I. (1992). Explaining the Gibbs sampler. *Amer. Statist.*, 46, 167-174.

Chen, Y.P., and Singpurwalla, N. D. (1994). A non-Gaussian Kalman filter model for tracking software reliability. *Statistica Sinica,* 4, 535-548.

Çinlar, E. (1975). *Introduction to Stochastic Processes.* Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Cox, D.R., and Lewis, P.A. (1966). *Statistical Analysis of Series of Events.* Methuen, London, United Kingdom.

Crowder, M.J., Kimber, A.C., Smith, R.L., and Sweeting T.J. (1991). *Statistical Analysis of Reliability.* Chapman and Hall, London, United Kingdom.

Gelfand, A.E., and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.,* 85, 398-409.

Gelman, A.E., and Rubin D. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.,* 7, 457-472.

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. and Machine Intelligence,* 6, 721-741.

Hjorth, U. (1980). A reliability distribution with increasing, decreasing, constant, and bathtub-shaped failure rates. *Technometrics*, 22, 99-107.

Hossain, S., and Dahiya, R., (1993). Estimating the parameters of a non-homogeneous Poisson-process model for software reliability. *IEEE Trans. Reliab.,* 42, 604-611.

Kuo, L., Lee, J. C., Choi, K., and Yang, T.Y. (1997). Bayes inference for S-shaped software reliability growth models. *IEEE Trans. Reliab.,* 46, 76-80.

Kuo, L., and Yang, T.Y. (1995). Bayesian computation of software reliability. *J. Comput. Graphical Statist.,* 4, 65-82.

Kuo, L., and Yang, T.Y. (1996). Bayesian computation for nonhomogeneous Poisson processes in software reliability. *J. Amer. Statist. Assoc.,* 91, 763-773.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chemical Physics,* 21, 1087-1092.

Musa, J.D., and Okumoto, K. (1984). A logarithmic Poisson execution time model for software reliability measurement. *Proceedings Seventh International Conference on Software Engineering*, Orlando, 230-238.

Sinha, S.K. (1986). *Reliability and Life Testing.* Wiley, New York, USA.

Tanner, M., and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.,* 82, 528-550.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.,* 22, 1701-1758.

Yang, T., and Kuo, L. (1995). Bayesian computation for the superposition of nonhomogeneous Poisson processes. Technical Report No. 9530, Statistics Department, University

of Connecticut.

Table 1: Prior Distributions and Posterior Inference for Full and Nested Models

	Full Superposition	Nested Model 1 with Musa-Okumoto	Nested Model 2 with Weibull
Priors	$\alpha_1 \sim \Gamma(1, 10^{-2})$	$\alpha_1 \sim \Gamma(1, 10^{-2})$	$\alpha_1 \sim \Gamma(1, 10^{-2})$
	$\beta_1 \sim \Gamma(1, 10^{-2})$	$\beta_1 \sim \Gamma(1, 10^{-2})$	$\beta_1 \begin{cases} = 0 : q_0 = 0.5 \\ \sim \Gamma(1, 10^{-2}) : \ 1 - q_0 \end{cases}$
	$\alpha_2 \sim U(0, 1)$	$\alpha_2 \sim U(0, 1)$	$\alpha_2 \sim U(0, 1)$
	$\beta_2 \sim \Gamma(1, 10^{-2})$	$\beta_2 \begin{cases} = 0 : p_0 = 0.5 \\ \sim \Gamma(1, 10^{-2}) : 1 - p_0 \end{cases}$	$\beta_2 \sim \Gamma(1, 10^{-2})$
Posterior for the Reduced Model		$p^* = 0.147$	$q^* = 0.093$

Table 2: Log-Scaled Predictive Likelihood

	Superposition	Musa-Okumoto	Weibull
Prior	$\alpha_1 \sim \Gamma(1, 10^{-2})$	$\alpha_1 \sim \Gamma(1, 10^{-2})$	
	$\beta_1 \sim \Gamma(1, 10^{-2})$	$\beta_1 \sim \Gamma(1, 10^{-2})$	
	$\alpha_2 \sim U(0, 1)$		$\alpha_2 \sim U(0, 1)$
	$\beta_2 \sim \Gamma(1, 10^{-2})$		$\beta_2 \sim \Gamma(1, 10^{-2})$
Log-predictive likelihood	-45.53	-46.67	-47.73

Tae Young Yang
Department of Mathematics
Myongji University
Yongin, Kyonggi 449-728
Republic of Korea
e-mail: tyang@wh.myongji.ac.kr

Lynn Kuo
Department of Statistics U-120
University of Connecticut
Storrs CT 06269-3120
U.S.A.
e-mail: lynn@stat.uconn.edu