I²SDS

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2008-4

May 24, 2008

Computational Issues in Semiparametric Bayesian Replacement Models

Jason R. Merrick
Statistical Sciences and Operations Research
Virginia Commonwealth University, Richmond, VA

Refik Soyer
Department of Decision Sciences
The George Washington University, Washington, DC

2

Computational Issues in Semiparametric Bayesian Replacement Models

Jason R. W. Merrick, Refik Soyer
Virginia Commonwealth University, The George Washington University

CONTENTS

In this paper we discuss certain issues that arise in the Bayesian analysis of replacement models where the cumulative intensity function is modeled using a nonparametric approach. The problem we consider is motivated by maintenance of railroad tracks which experience wear as a result of traffic. We consider a semi-parametric model to describe the failure characteristics of a rail section by specifying a nonparametric form for cumulative intensity function and by taking into account effect of covariates by a parametric form. Use of a gamma process prior for the cumulative intensity function complicates the Bayesian analysis when the updating is based on failure count data. We develop a Bayesian analysis of the model using Markov chain Monte Carlo (MCMC) methods and determine replacement strategies. Adoption of MCMC methods involves a data augmentation algorithm. We show the implementation of our approach using actual data.

Keywords: *Gamma process prior*, *Optimal replacement*, *Data augmentation* and *Proportional intensities model*

1-58488-344-8/04/$0.00+$1.50

2.1 Introduction

Planned replacement strategies are commonly used for systems such as railroad tracks that experience aging or wear. This is done to prevent in-service failures that may be very costly relative to the cost associated with a planned replacement/repair. Railroad tracks experience wear as a function of traffic usage, which is measured in millions of gross tons (MGT). A failure of a railroad track takes the form of a crack in a rail section. Though this does not affect the use of the rail immediately, it can possibly lead to a fracture which is potentially hazardous. The replacement of rail tracks is a major expense for railroad companies. Thus, it is important for railroad companies to develop decision models to determine effective replacement strategies.

Most of the replacement strategies literature assumes that the failure characteristics of systems are known and does not address statistical issues; see for example Cho and Parlar (1991) for a general review. More recently, statistical issues in the development of optimal replacement strategies have been considered by Mazzuchi and Soyer (1995, 1996) and Dayanik and Gurler (2002) using Bayesian approaches. These authors considered parametric Bayesian approaches that do not allow a flexible modeling strategy in determining optimal strategies. Furthermore, their approaches do not allow incorporation of covariate effects on failure intensity. Nonparametric replacement strategies have been considered from a sampling theory perspective in Frees and Ruppert (1985) where adaptive age replacement policies are developed. Such non-parametric approaches have not been considered from a Bayesian point of view.

In this paper, we present a Bayesian decision theoretic approach to the optimal replacement problem by focusing on systems such as railroad tracks that are subject to wear. In so doing, we present a semi-parametric model to describe the failure characteristics of rail tracks by specifying a nonparametric form for modeling wear and by taking into account effect of covariates by a parametric function. We develop a Bayesian analysis of the model based on failure/replacement data using Markov chain Monte Carlo methods (MCMC) and determine replacement strategies using our model. Adoption of MCMC methods for determining optimal strategies requires development of a data augmentation algorithm, in the sense of Tanner and Wong (1987), to evaluate posterior predictive distributions.

Synopsis of our paper is as follows. In Section 2,a modulated Poisson process model is presented for describing the failure behavior of as railroad tracks, that are subject to minimal repair. The modulated Poisson process model was first proposed in Cox (1972b) to consider covariate effects in counting processes. We refer to the model as proportional intensities model (PIM) as it is a counting process alternative to the proportional hazards model (PHM) of Cox (1972a). We introduce a semi-parametric PIM by using a gamma process prior for the baseline cumulative intensity and specifying parametric priors for covariate effects. In Section 3 Bayesian analysis of the semiparametric PIMs is considered where the number of rail track failures are described by a non-homogeneous Poisson process (NHPP). The MCMC based procedures that are used for parametric type models are adopted for inference for the semiparametric PIM. The analysis of rail track data is straightforward if the failure counts are observed in identical traffic usage intervals for each rail section. However, data augmentation steps must be introduced to handle the overlapping, but not identical, intervals that occur in the railroad data analyzed and to perform prediction for development of replacement strategies. We discuss the basics of block replacement with minimal repair protocol that applies to repairable systems such as railroad tracks in Section 4 using cost-based utility (loss) functions. In Section 5, Bayesian replacement strategies are developed for rail tracks the semiparametric model and an illustration of the approach is presented using actual rail track failure data.

2.2 Proportional Intensities Model for Rail Section Failures

As pointed out in Section 1, railroad tracks experience wear as a function of traffic usage and the wear causes a failure of a railroad track in the form of a crack in a rail section. Such a crack can possibly lead to a fracture if it is not repaired. When a crack is found on the rail, a small piece of rail section around the crack is cut out and replaced with a new rail piece. Since this does not significantly change the performance of the rail section which can be miles in length, the rail sections are assumed to be minimally repaired.

As the railroad tracks are assumed to be minimally repaired upon failure, point processes, and specifically non-homogeneous Poisson processes (NHPP), are used to model their failure behavior. Most of the models applied to repairable systems do not consider the effect of covariates on the intensity function of the NHPP. Under the minimal repair (MR) protocol of Barlow and Hunter (1960), the number of failures of the i-th rail section is described by a nonhomogeneous Poisson process (NHPP) with cumulative intensity (or mean value) function $\Lambda_i(t|\Theta)$. In what follows we will present a generalization of the NHPP to incorporate covariate effects in the cumulative intensity function.

Let $N_i(t)$ denote the number of failures for the i-th rail section in an interval of length t MGT and let Z_i denote the p-dimensional vector of available covariates that describe the characteristics of the i-th rail section. In the data on rail failures used for the analysis, the available covariates are constant with respect to traffic usage. $N_i(t)$ is described by a NHPP with intensity function

$$\lambda_i(t) = \frac{d}{dt}E[N_i(t)]. \tag{2.1}$$

To reflect the fact that the intensity function is affected by covariates, $\lambda_i(t)$ can be modulated by a function of Z_i. Such a modulation was introduced in Cox (1972b) by considering

$$\lambda_i(t;Z_i) = \lambda_0(t)e^{\beta^T Z_i} \tag{2.2}$$

where $\lambda_0(t)$ is the baseline intensity function and β is a vector of p parameters. The Poisson process model defined by the intensity (2.2) was referred to as the modulated Poisson process by Cox (1972b). The model can be thought as a counting process alternative to the proportional hazards model (PHM) of Cox (1972a) where a similar form was used for the failure rate of a non-repairable system. Thus, we will refer to the model as the *proportional intensities model* (PIM).

Under the PIM the cumulative intensity function of the rail track failure process is given by $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$ which can be written as

$$\Lambda_i(t; Z_i) = \Lambda_0(t) e^{\beta^T Z_i}, \tag{2.3}$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the baseline cumulative intensity function, that is, $E[N_i(t)] = \Lambda_i(t)$. We note that the baseline cumulative intensity $\Lambda_0(t)$ may have a parametric or a nonparametric form. In the former case, $\Lambda_0(t)$ will depend on some vector of parameters, say, θ. Thus, we will write the above as $\Lambda_0(t) = \Lambda_0(t;\theta)$ where $\Lambda_0(t;\theta) = \int_0^t \lambda_0(s;\theta)\ ds$. In the nonparametric case $\Lambda_0(t)$ will be modeled by a stochastic process. In both cases, the distribution of $N_i(t)$ given Z_i and $\Theta = (\Lambda_0(t), \beta)$ is specified using $\Lambda_i(t; Z_i, \Theta)$, explicitly

$$P(N_i(t) = n | \Lambda_0(t), \beta, Z_i) = \frac{\Lambda_0(t)^n e^{n\beta^T Z_i}}{n!} \exp\{-\Lambda_0(t) e^{\beta^T Z_i}\}. \tag{2.4}$$

Thus, $N_i(t)$ given Z_i and Θ is a NHPP and conditional on Z_i and Θ, all the properties of NHPPs will hold for the PIM. For example, for the i-th rail section, probability of number of failures in any MGT interval $[s,t)$, is obtained as

$$P(N_i(t) - N_i(s) = n | \Lambda_0(t), \beta, Z_i) = \frac{[\Lambda_0(t) - \Lambda_0(s)]^n e^{n\beta^T Z_i}}{n!} \exp\{-[\Lambda_0(t) - \Lambda_0(s)] e^{\beta^T Z_i}\}.$$

In the optimal replacement problem setup of Section 4, evaluation of the expected cost requires $E[N_i(t_B)|\Theta]$ where t_B is the replacement interval.

2.2.1 Modeling the Baseline Intensity Function

In modeling the baseline intensity function of the PIM, one strategy is to specify a parametric form $\lambda_0(t;\theta)$. For example, one can specify a power law model for $\lambda_0(t;\theta)$ which is widely used in reliability modeling of repairable systems. The power law model is given by $\lambda_0(t;\theta) = \alpha\gamma t^{\gamma-1}$ implying that

$$\Lambda_0(t;\theta) = \alpha t^{\gamma}, \tag{2.5}$$

where $\theta = (\alpha, \gamma)$ and $\alpha > 0$, $\gamma > 0$. In the power law model, values of $\gamma > 1$ imply that the system, in our case the rail track, deteriorates by usage, that is, by MGT. This is typically what is expected in rail tracks that are subject to wear. Under the parametric modeling strategy, the Bayesian formulation of the optimal replacement problem is completed by specifying the prior distribution $\pi(\Theta|D_0)$ of the unknown parameters $\Theta = (\theta, \beta)$, that is, $\pi(\alpha, \gamma, \beta|D_0)$ for the power law model.

Railroad tracks show great deal of variation in their physical characteristics and in terms of the environments under which they operate. A fully parametric model is not flexible enough to account for such variation. An alternative modeling strategy is to consider a nonparametric form for the baseline intensity $\lambda_0(t)$ or equivalently for the cumulative the baseline intensity $\Lambda_0(t)$ of the PIM. In the Bayesian framework this can be achieved by specifying a prior distribution on the baseline cumulative intensity function $\Lambda_0(t)$. In order to provide flexibility in modeling, it is important that such a prior allows a wide variety of different forms for $\Lambda_0(t)$. Since, the baseline cumulative intensity function is proportional to the expected number of failures up to traffic usage t in the PIM, there is no restriction on the size of any instantaneous jumps of the $\Lambda_0(t)$. Thus a gamma process is a suitable prior for $\Lambda_0(t)$ in the PIM.

To construct a gamma process prior, we consider a partition of $[0, \infty)$ into k intervals can be defined as $[t_0, t_1)$, $[t_1, t_2), \ldots, [t_{k-1}, t_k = \infty)$, where $\Lambda_0(t_0) = 0$ and $r_l = \Lambda_0(t_l) - \Lambda_0(t_{l-1})$, implying that

$$\Lambda_0(t_j) = \sum_{l=1}^{j} r_l. \tag{2.6}$$

for $j = 1, \ldots, k$. Doksum (1974) considered such a construction and showed that a probability distribution can be specified on the space of positive increasing functions, $\{\Lambda_0(t)\}$, by specifying the k-dimensional distribution of $r_1, \ldots, r_k$, for each possible partition $[t_0, t_1)$, $[t_1, t_2), \ldots,$ $[t_{k-1}, \infty)$. In this construction the distributional assumptions must hold for any partition of $[0, \infty)$ and must be consistent between partitions. The process obtained is non-decreasing and the increments are independent. If the increments have gamma distributions, the resulting process is called a gamma process, see Singpurwalla (1997). Let c be a positive real number, $\Lambda_0^*(t)$ be a best guess for baseline cumulative intensity function and assume that the distribution of the r_j's is given by

$$r_j G(c\Lambda_0^*(t_j) - c\Lambda_0^*(t_{j-1}), c), \tag{2.7}$$

where $XG(a, b)$ denotes that X has a gamma distribution with shape parameter a and scale parameter b. It follows from this construction that $\Lambda_0(t)$ is a gamma process with $\Lambda_0^*(t)$ being a best guess and c is a measure of certainty about the best guess given the prior history D_0,

$$(\Lambda_0(t)|D_0) G(c\Lambda_0^*(t), c), \tag{2.8}$$

for all values of t. The above implies that $E[\Lambda_0(t)|D_0] = \Lambda_0^*(t)$ and $V[\Lambda_0(t)|D_0] = \Lambda_0^*(t)/c$.

Treatment of $\Lambda_0(t)$ as a stochastic process in the above enables us to develop a Bayesian version of the replacement models considered by Ozekici (1995). Note that using the nonparametric approach we have specified the prior only for $\Lambda_0(t)$. We can complete the Bayesian formulation of the optimal replacement problem by specifying a parametric form for the prior distribution $\pi(\beta|D_0)$ of β as independent of $\Lambda_0(t)$. Our Bayesian modeling strategy consists of a nonparametric treatment of the baseline cumulative intensity and a parametric specification of the effect of covariates in the $\Lambda_i(t;Z_i,\Theta) = \Lambda_0(t)e^{\beta^T Z_i}$. This approach is usually referred to as a *semiparametric Bayesian approach* and thus, we will refer to the corresponding PIM as the *semiparametric PIM*.

2.3 Bayesian Inference for the Semiparametric PIM

In this section, we will present Bayesian inference for the semiparametric PIMs. In so doing, we first discuss the Bayesian analysis of the parametric PIM using an adoption of MCMC methods of Dellaportas and Smith (1993) presented for the PHM. Analysis of the semiparametric model is nontrivial when the failure counts are observed in the overlapping, but not identical traffic usage intervals as is typically the case with actual data coming from different rail sections. This requires development of a new MCMC algorithm for the Bayesian analysis.

Under the parametric Bayesian approach, the baseline cumulative intensity $\Lambda_0(t)$ is assumed to be a differentiable function $\Lambda_0(t;\theta)$ where θ is a vector of unknown parameters. Thus the baseline intensity function $\lambda_0(t)$ is given by $\lambda_0(t;\theta) = \frac{d}{dt}\Lambda_0(t;\theta)$. If $N_i(t)$ for each rail section $i = 1,\ldots,n$ is observed at traffic usages $t = t_{i,1},\ldots,t_{i,r_i}$ then the data for the i-th rail section is given by $D_i = \{N_i(t) = n_i(t), j = 1,\ldots,r_i, Z_i\}$. Using the independent increments property of the NHPP, the likelihood function of θ and β given D_i is written as

$$L_i(\theta,\beta;D_i) = \prod_{j=1}^{r_i} \frac{\left(\left\{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\right\} e^{\beta^T Z_i}\right)^{n_i(t_{i,j})-n_i(t_{i,j-1})}}{\left(n_i(t_{i,j}) - n_i(t_{i,j-1})\right)!} \times \exp\{-\left\{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\right\} e^{\beta^T Z_i}\},$$

where $\Lambda_0(t_{i,0};\theta) = 0$.

Given m rail sections, conditional on the cumulative intensities, that is, $\Lambda_i(t)$'s $i = 1,\ldots,m$, the $N_i(t)$'s are assumed to be independent. Thus, given the failure counts for each $N_i(t)$ at traffic usage $t = t_{i,1},\ldots,t_{i,r_i}$ for $i = 1,\ldots,m$, the likelihood function of θ and β given $D = (D_i; i = 1,...,m)$ is given by

$$L(\theta,\beta;D) = \prod_{i=1}^{m} L_i(\theta,\beta;D_i). \tag{2.9}$$

The joint posterior distribution of θ and β given D can not be obtained analytically for any given form of the prior $\pi(\theta,\beta)$, but a Gibbs sampler can be used to draw samples from the joint posterior $\pi(\theta,\beta|D)$. For any reasonable choice of the forms of $\Lambda_0(t;\theta)$, $\pi(\theta,\beta)$, the full conditonal distributions are logconcave densities and therefore the adaptive rejection sampling algorithm of Gilks and Wild (1992) can be used to draw samples from these distributions at each iteration of the Gibbs sampler. Typically independent priors are assumed for θ and β and a reasonable form for $\pi(\beta)$ is the multivariate normal density.

The use of a gamma process prior for the cumulative intensity function of a NHPP was considered by Kuo and Ghosh (2001). The model considered by the authors excluded the covariate information and the inference was introduced only for the case of failure time data. If the data is only available as failure counts at different points in traffic usage, as in the case of the data for the railroad tracks, then the semiparametric Bayesian inference in the PIM is not straightforward. In the railroad track data, the rail sections are observed over different intervals, some of which overlap. In this case the implementation of the Gibbs sampler requires a data augmentation step. Such a step is also required for the case of a single rail section for predictive estimation which is needed in development of replacement strategies. In the sequel, inference for the semiparametric PIM will be discussed for the single and multiple item cases and then a general algorithm will be presented.

2.3.1 Analysis for a Single Rail Section

Suppose that for the i-th rail section, the process $N_i(t)$ is observed in the traffic usage intervals $[t_1, t_2)$ and $[t_2, t_3)$ so that the data is given by

$$D_i = \{N_i(t_2) - N_i(t_1) = n_1, N_i(t_3) - N_i(t_2) = n_2, Z_i\}$$

as shown in Figure 2.1. The cumulative intensity for the i-th rail section is given by

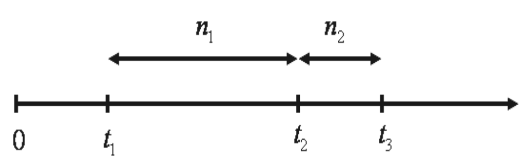

FIGURE 2.1

The data observed for a single rail section.

$$\Lambda_i(t) = \Lambda_0(t) e^{\beta^T Z_i},$$

where $\Lambda_0(t)$ is defined by (2.8). Using the independent increments property of the NHPP, the likelihood function of $\Lambda_0(t)$ and β given D_i is written as

$$L_i(\Lambda_0(t), \beta | D) = \prod_{j=1}^{2} \frac{\left(\Lambda_0(t_{j+1}) e^{\beta^T Z_i} - \Lambda_0(t_j) e^{\beta^T Z_i}\right)^{n_j}}{n_j!} \tag{2.10}$$

$$\times \exp\{-(\Lambda_0(t_{j+1}) e^{\beta^T Z_i} - \Lambda_0(t_j) e^{\beta^T Z_i})\}. \tag{2.11}$$

As in the parametric case, a Gibbs sampler can be used for developing posterior inference. Assuming that the prior on β, $\pi(\beta)$ is independent of the gamma process prior on $\Lambda_0(t)$, the full conditional posterior of $\Lambda_0(t)$ given β is obtained, by using the independent increments property of the gamma process, as

$$\begin{aligned}\pi(\Lambda_0|\beta,D_i) \propto\ & [\Lambda_0(t_2)-\Lambda_0(t_1)]^{c\left[\Lambda_0^*(t_2)-\Lambda_0^*(t_1)\right]+n_1-1} \\ & \times \exp\left\{-[\Lambda_0(t_2)-\Lambda_0(t_1)]\,(c+e^{\beta^T Z_i})\right\} \\ & \times [\Lambda_0(t_3)-\Lambda_0(t_2)]^{c\left[\Lambda_0^*(t_3)-\Lambda_0^*(t_2)\right]+n_2-1} \\ & \times \exp\left\{-[\Lambda_0(t_3)-\Lambda_0(t_2)]\,(c+e^{\beta^T Z_i})\right\}.\end{aligned}$$

Thus the posterior distribution of $\Lambda_0(t)$ conditional on β can be written as

$$(\Lambda_0(t)|\beta,D_i) \quad G(c\Lambda_0^*(t),c),\ \text{for } t<t_1 \tag{2.12}$$

$$(\Lambda_0(t_2)-\Lambda_0(t_1)|\beta,D_i) \quad G(c\{\Lambda_0^*(t_2)-\Lambda_0^*(t_1)\}+n_1,c+e^{\beta^T Z_i}), \tag{2.13}$$

$$(\Lambda_0(t_3)-\Lambda_0(t_2)|\beta,D_i) \quad G(c\{\Lambda_0^*(t_3)-\Lambda_0^*(t_2)\}+n_2,c+e^{\beta^T Z_i}), \tag{2.14}$$

$$(\Lambda_0(t)-\Lambda_0(t_3)|\beta,D_i) \quad G(c\{\Lambda_0^*(t)-\Lambda_0^*(t_3)\},c),\ \text{for } t>t_3. \tag{2.15}$$

It follows from (2.13) and (2.14) that

$$\Lambda_0(t_2)=[\Lambda_0(t_2)-\Lambda_0(t_1)]+\Lambda_0(t_1) \tag{2.16}$$

is a sum of two independent gamma random variables with different scale parameters. Thus, the distribution of $(\Lambda_0(t_2)|\beta,D_i)$ can be simulated as the sum of two gamma random variables. If the process was also observed during the interval $[0,t_1)$ with, say, $N_i(t_1)=n_0$ then the distribution of $\Lambda_0(t_1)$ would be updated as

$$(\Lambda_0(t_1)|\beta,D_i)G(c\Lambda_0^*(t_1)+n_0,c+e^{\beta^T Z_i}) \tag{2.17}$$

and the distribution of $\Lambda_0(t_2)$ would be the sum of two independent gamma random variables with the same scale implying that

$$(\Lambda_0(t_2)|\beta,D_i)G(c\Lambda_0^*(t_2)+n_0+n_1,c+e^{\beta^T Z_i}).$$

It can be seen that conditional on β, the effect of the covariates is on the scale parameter of the gamma distribution. As previously dicussed, sampling from the full conditional of β can be done via the adaptive rejection sampling method of Gilks and Wild (1992) as the joint density is log-concave.

The Gibbs sampler is used to sample from the joint posterior distribution of

$$(\Lambda_0(t_2)-\Lambda_0(t_1),\Lambda_0(t_3)-\Lambda_0(t_2),\beta|D_i), \tag{2.18}$$

using (2.13) and (2.14) whose sum is a sample point for $[\Lambda_0(t_3)-\Lambda_0(t_1)]$, which in turn yields a sample point from the full conditional distribution of β using adaptive rejection sampling in an iterative manner.

The conditional posterior distribution of $\Lambda_0(t)$ given β is known for $t < t_1$ and $t > t_3$ and for the instants of traffic usage t_2 and t_3. However, as the number of failures of the rail section up to traffic usage t is not known for $t \in [t_1, t_2)$ and $t \in [t_2, t_3)$, the posterior distribution is not immediately available. This causes problems when making predictive statements, such as in the optimal replacement problem discussed in Section 5.

2.3.1.1 The Prediction Problem for a Single Rail Section

Suppose that the posterior distribution of $\Lambda_0(t^*)$ is required, where t^* is in the interval (t_1, t_2) and thus the number of failures between t_1 and t^* is unknown as shown in Figure 2.2. One way to update the distribution of $\Lambda_0(t^*)$ is through a data aug-

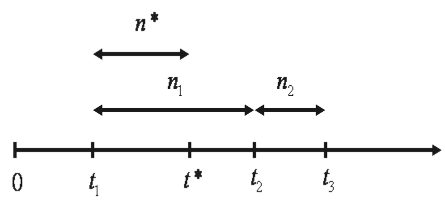

FIGURE 2.2
The prediction problem for a single rail section.

mentation step within the Gibbs sampler. If $N_i(t^*) - N_i(t_1) = n^*$ is known then the distribution of $\Lambda_0(t^*)$ can be updated as the sum of two independent gamma random variables as

$$\Lambda_0(t^*) = [\Lambda_0(t^*) - \Lambda_0(t_1)] + \Lambda_0(t_1), \tag{2.19}$$

where $(\Lambda_0(t_1)|\beta, D_i)$ is given by (2.12) and

$$(\Lambda_0(t^*) - \Lambda_0(t_1)|\beta, D_i, n^*) G(c\{\Lambda_0^*(t^*) - \Lambda_0^*(t_1)\} + n^*, c + e^{\beta^T Z_i}). \tag{2.20}$$

Similarly, the updating for the other increments of the gamma process can be obtained as

$$(\Lambda_0(t_2) - \Lambda_0(t^*)|\beta, D_i) G(c\{\Lambda_0^*(t_2) - \Lambda_0^*(t^*)\} + (n_1 - n^*), c + e^{\beta^T Z_i}), \tag{2.21}$$

and $\Lambda_0(t_3) - \Lambda_0(t_2)$ is still given by (2.14). The above results follow from the independent increments property of the gamma process.

The implementation of the Gibbs sampler requires draws from $(\beta|\ \Lambda_0(t), D_i, n^*)$ and the adaptive rejection sampling algorithm can be used to draw samples from this distribution. The final component of the Gibbs sampler is the full conditional for $(N_i(t^*) - N_i(t_1)|\Lambda_0(t), \beta, D_i)$. By using independent increments property of the NHPP and adopting a well known result in NHPP's given by Ross (1989, p. 242), it can be shown that

$$(N_i(t^*) - N_i(t_1) = n^* | \Lambda_0(t), \beta, D_i) \sim Bin\left[n_1, \frac{\Lambda_0(t^*) - \Lambda_0(t_1)}{\Lambda_0(t_2) - \Lambda_0(t_1)}\right], \quad (2.22)$$

which is a Binomial distribution where the terms involving β are implicit in the generated values of $\Lambda_0(\bullet)$.

2.3.2 Analysis for Two Rail Sections

Analysis for Two Rail Sections Not Requiring Data Augmentation

Suppose that data from multiple, say $m = 2$, rail sections are observed. Let $\{N_1(t)\}$ and $\{N_2(t)\}$ denote the corresponding NHPP's with the same baseline cumulative intensity function, $\Lambda_0(t)$. As in the previous section, given the $\Lambda_i(t)$'s, $i = 1, 2$, $N_1(t)$ and $N_2(t)$ are assumed to be independent. For illustrative purposes, consider the case where a single interval is observed for each rail section i with n_i failures in $[t_{i,1}, t_{i,2})$, for $i = 1, 2$. Then the likelihood function of $\Lambda_0(t)$ and β given $D = \{n_{1,1}, [t_{1,1}, t_{1,2}),$ $n_{2,1}, [t_{2,1}, t_{2,2}), Z_1, Z_2\}$ is obtained as

$$L(\Lambda_0(t), \beta; D) = \prod_{i=1}^{2} \frac{\left[(\Lambda_0(t_{i,2}) - \Lambda_0(t_{i,1}))\, e^{\beta^T Z_i}\right]^{n_{i,1}}}{n_{i,1}!} \exp\left\{-(\Lambda_0(t_{i,2}) - \Lambda_0(t_{i,1}))\, e^{\beta^T Z_i}\right\}. \quad (2.23)$$

The likelihood function for the case of multiple traffic usage intervals for each process can be easily obtained by using the independent increments property of each NHPP.

The posterior inference in the above case follows along the lines of the last section if the observed intervals for the two rail sections are not overlapping as shown in Figure 2.3. In this particular case, using the independent increments property of the

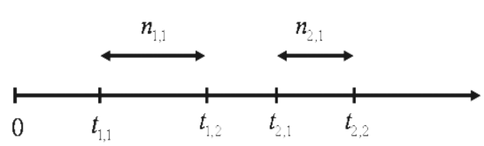

FIGURE 2.3
The case of non-overlapping intervals for two rail sections.

gamma process prior, the full conditional posterior of $\Lambda_0(t)$ given β is obtained as

$$\begin{aligned}\pi(\Lambda_0(t)|\beta,D) \propto\ & (\Lambda_0(t_{1,2})-\Lambda_0(t_{1,1}))^{c\left(\Lambda_0^*(t_{1,2})-\Lambda_0^*(t_{1,1})\right)+n_{1,1}-1} \\ & \times \exp\left\{-(\Lambda_0(t_{1,2})-\Lambda_0(t_{1,1}))\,(c+e^{\beta^T Z_1})\right\} \\ & \times (\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1}))^{c\left(\Lambda_0^*(t_{2,2})-\Lambda_0^*(t_{2,1})\right)+n_{2,1}-1} \\ & \times \exp\left\{-(\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1}))\,(c+e^{\beta^T Z_2})\right\}\end{aligned} \tag{2.24}$$

implying that

$$\begin{aligned}(\Lambda_0(t)|\beta,D) &\quad G(c\Lambda_0^*(t),c),\ \text{for } t<t_{1,1} \\ (\Lambda_0(t_{1,2})-\Lambda_0(t_{1,1})|\beta,D) &\quad G(c\{\Lambda_0^*(t_{1,2})-\Lambda_0^*(t_{1,1})\}+n_{1,1},c+e^{\beta^T Z_1}), \\ (\Lambda_0(t_{2,1})-\Lambda_0(t_{1,2})|\beta,D) &\quad G(c\{\Lambda_0^*(t_{2,1})-\Lambda_0^*(t_{1,2})\},c), \\ (\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1})|\beta,D) &\quad G(c\{\Lambda_0^*(t_{2,2})-\Lambda_0^*(t_{2,1})\}+n_{2,1},c+e^{\beta^T Z_2}), \\ (\Lambda_0(t)-\Lambda_0(t_{2,2})|\beta,D) &\quad G(c\{\Lambda_0^*(t)-\Lambda_0^*(t_{2,2})\},c),\ \text{for } t>t_{2,2}.\end{aligned} \tag{2.25}$$

Updating for other portions of $\Lambda_0(t)$ follows along the same lines as presented in the last section. Similarly, sampling from the full conditional of β given $\Lambda_0(t)$ and D is achieved via the use of adaptive rejection sampling.

Note that if both rail sections are observed for the same traffic usage interval, as in Figure 2.4, then updating is again straightforward.

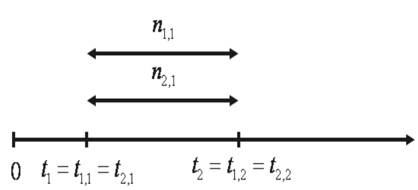

FIGURE 2.4
The case of identical intervals for two rail sections.

Analysis for Two Rail Sections Requiring Data Augmentation

In the railroad track data, there are cases where two rail sections are observed for different but overlapping traffic usage intervals, as shown in Figure 2.5. Updating $\Lambda_0(t)$ given β then requires the use of a data augmentation step in the Gibbs sampler as discussed before.

As the intervals overlap, $N_1(t_{12})-N_1(t_{11})$ and $N_2(t_{22})-N_2(t_{21})$ are no longer independent a priori. Thus $(\Lambda_0(t_{12})-\Lambda_0(t_{11}))$ and $(\Lambda_0(t_{22})-\Lambda_0(t_{21}))$ cannot be updated separately. However, if the counts over the non-overlapping intervals $[t_{11},t_{21})$

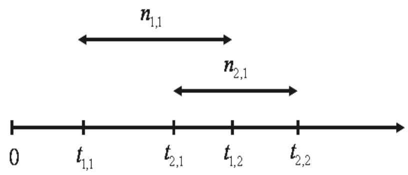

FIGURE 2.5
The case of overlapping intervals for two rail sections.

and $[t_{21}, t_{12})$ from $N_1(\bullet)$ and $[t_{21}, t_{12})$ and $[t_{12}, t_{22})$ from $N_2(\bullet)$ were available then updating could be performed on each interval separately. This is possible due to the independent increments properties of the Poisson and gamma processes. Assume that $N_1(t_{2,1}) - N_1(t_{1,1}) = n_1^*$ and $N_2(t_{2,2}) - N_2(t_{1,2}) = n_2^*$ as shown in figure 2.6. Then it follows from the above that

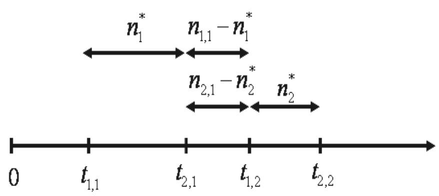

FIGURE 2.6
The failure counts required for data augmentation.

$$(\Lambda_0(t_{21}) - \Lambda_0(t_{11}) \ |\beta, n_1^*, D) \sim G(c[\Lambda_0^*(t_{21}) - \Lambda_0^*(t_{11})] + n_1^*, c + e^{\beta^T Z_1}), \tag{2.26}$$

$$(\Lambda_0(t_{12}) - \Lambda_0(t_{21})|\beta, n_1^*, n_2^*, D)|\beta, n_1^*, n_2^*, D) \sim G(c[\Lambda_0^*(t_{12}) - \Lambda_0^*(t_{21})] + (n_{1,1} - n_1^*) + (n_{2,1} - n_2^*), c + \sum_{i=1}^{2} e^{\beta^T Z_i} \ ]) \tag{2.27}$$

and

$$(\Lambda_0(t_{22}) - \Lambda_0(t_{12}) \ |\beta, n_2^*, D) \sim G(c[\Lambda_0^*(t_{22}) - \Lambda_0^*(t_{12})] + n_2^*, c + e^{\beta^T Z_2}). \tag{2.28}$$

In implementing the Gibbs sampler, data augmentation is needed on the number of failures of the railroad tracks in the non-overlapping periods. Again, using the properties of the Poisson process, it can be shown that

$$(N_1(t_{2,1}) - N_1(t_{1,1}) \ |n_{1,1}, \tfrac{\Lambda_0(t_{2,1}) - \Lambda_0(t_{1,1})}{\Lambda_0(t_{1,2}) - \Lambda_0(t_{1,1})}) \sim Bin(n_{1,1}, \tfrac{\Lambda_0(t_{2,1}) - \Lambda_0(t_{1,1})}{\Lambda_0(t_{1,2}) - \Lambda_0(t_{1,1})}) \tag{2.29}$$

and

$$\begin{aligned}(N_2(t_{2,2}) - N_2(t_{1,2}) \; |n_{2,1}, \tfrac{\Lambda_0(t_{2,2})-\Lambda_0(t_{1,2})}{\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1})}) \\ \sim Bin(n_{2,1}, \tfrac{\Lambda_0(t_{2,2})-\Lambda_0(t_{1,2})}{\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1})}),\end{aligned} \tag{2.30}$$

where (2.29) and (2.30) are independent binomial random variables.

2.3.3 A General Data Augmentation Algorithm

The data augmentation is not overly complex for the case of two rail sections with only one overlapping interval. If the number of overlapping intervals increases, deciding which intervals upon which to data augment is more complicated and, therefore, requires a systematic approach. One alternative is to break the possible traffic usages in to a partition defined by the endpoints of all intervals, such as in Figure 2.7 for the case of three NHPP's. For any observed interval that has now been broken up

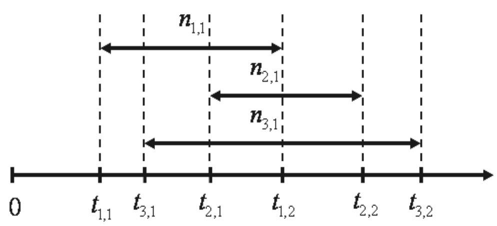

FIGURE 2.7

The case of three overlapping intervals.

in to sub-intervals, data augmentation is used on all but the one of these sub-intervals; the number of failures in the remaining interval is known given the total number of failures in the whole interval and the number of failures in the other sub-intervals. The distribution of the augmented failure counts will be a multinomial distribution. We now describe a general data augmentation algorithm that follows this approach.

We are given data $D = (D_i; i = 1,\dots,m)$ from m rail sections where $D_i = \{N_i(t) = n_i(t), j = 1,\dots,r_i, Z_i\}$. Our data consists of r_i inspection runs for the i-th rail section where the inspections are performed at $t_{i,0},\dots,t_{i,r_i}$ MGTs and $n_i(t_{i,1}) - n_i(t_{i,0}),\dots,$ $n_i(t_{i,r_i}) - n_i(t_{i,r_i-1})$ failures are discovered. In order to generalize the algorithm that we have presented in the previous sections, we need to first determine the intervals that will be used for the data augmentation steps within the Gibbs sampler.

Let $t_1^*,\dots,t_q^*$ denote the ordered list of q unique values amongst the interval endpoints $t_{i,j}$ for $j = 1,\dots,r_i$ and $i = 1,\dots,m$. The ordered values t_k^* for $k = 1,\dots,q$ will be used for the data augmentation. Also, let $N_{i,k}^*$ denote the unknown number of failures in the interval $[t_k^*, t_{k+1}^*)$ for rail section i and $B_k^* = \{i : \exists j \mid t_k^* \le t_{i,j} < t_{k+1}^*\}$ for $k = 1,\dots,q-1$, denote the set of all rail indices i that have a failure count that spans the interval $[t_k^*, t_{k+1}^*)$. Furthermore, let $S_{i,j}^* = \{k : t_{i,j} \le t_k^* < t_{i,j+1}\}$ be the set of all interval endpoints for all rails that fall within the j-th observed interval for the i-th rail and define $m^*_{i,j} = |S^*_{i,j}|$ be the number of interval endpoints in this set. We will also define the ordered list of members of $S^*_{i,j}$ by $\{l^1_{i,j}, \ldots, l^{m^*_{i,j}}_{i,j}\}$.

Given the above setup, at each iteration of the Gibbs sampler, the full posterior conditional distribution of $\Lambda_0(t)$ given β and D can be obtained by data augmenting on $N^*_{i,k}$. Similar to the development in the previous sections, given $N^*_{i,k}$, β and D, we can update $\Lambda_0(t^*_{k+1}) - \Lambda_0(t^*_k)$ by using the independent increments property of the gamma process. More specifically, given $N^* = (N^*_{i,k};\ i = 1, \ldots, m, k = 1, \ldots, q)$, we can easily show that

$$(\Lambda_0(t^*_{k+1}) - \Lambda_0(t^*_k)\ |N^*, \beta, D) G(c[\Lambda^*_0(t^*_{k+1}) - \Lambda^*_0(t^*_k)] + \sum_{i\varepsilon B^*_k} N^*_{i,k}, c + \sum_{i\varepsilon B^*_k} e^{\beta^T z_i}) \tag{2.31}$$

for $k = 1, \ldots, q$. Note that for $t < t^*_1$ we still have $(\Lambda_0(t)|\beta, D) G(c\Lambda^*_0(t), c)$. In order to obtain the distribution of $N^*_{i,k}$'s we define the vector $\underline{N^*_{i,j}} = (N^*_{i,k}; k\varepsilon S^*_{i,j})$ containing $N^*_{i,k}$'s that lie in the interval $[t_{i,j}, t_{i,j+1})$. Given $\Delta = \{\Lambda_0(t^*_{k+1}) - \Lambda_0(t^*_k);\, k = 1, \ldots, q\}$, using the properties of NHPPs we can obtain the full conditional of $\underline{N^*_{i,j}}$'s as

$$(\underline{N^*_{i,j}}|\Delta, \beta, D)\ Mult(n_i(t_{i,j+1}) - n_i(t_{i,j}), p^*_{i,1}, \ldots, p^*_{i,m^*_{i,j}}) \tag{2.32}$$

which is a multinomial of order $m^*_{i,j} - 1$, where

$$p^*_{i,l} = \frac{\Lambda_0(l^{l+1}_{i,j}) - \Lambda_0(l^l_{i,j})}{\Lambda_0(l^{m^*_{i,j}}_{i,j}) - \Lambda_0(l^1_{i,j})}. \tag{2.33}$$

Note that $\underline{N^*_{i,j}}$'s are drawn as independent multinomials at each iteration of the Gibbs sampler.

2.4 Block Replacement of Railroad Tracks with Minimal Repair

Under the block replacement protocol, all units are replaced at time points t_B, $2t_B$, …, irrespective of their ages and an in-service replacement or repair is made whenever failures occur; see Cox (1962). However, under the block replacement with minimal repair protocol of Barlow and Hunter (1960)items are minimally repaired upon failure but replaced at times t_B, $2t_B$, …, irrespective of their ages. The replacement problem involves optimal choice of the interval t_B typically by minimizing a cost function.

To introduce some notation let c_P to denote the cost of a planned replacement and c_R to denote the cost of a minimal repair such that $c_R < c_P$. As pointed out by Mazzuchi and Soyer (1996), the cost per unit time for the i-th rail section is given by

$$C(t_B, N_i(t_B)) = \frac{c_P + c_R N_i(t_B)}{t_B}, \tag{2.34}$$

where $N_i(t)$ represents the number of in-service failures for the i-th section that occur in an interval of length t, Assuming that m rail sections will be replaced at time t_B, the total cost per unit time is given by

$$C(t_B, N(t_B)) = \sum_{i=1}^{m} C(t_B, N_i(t_B)), \tag{2.35}$$

where $N(t_B) = (N_1(t_B), \ldots, N_m(t_B))$ .The optimal block replacement strategy t_B^* is determined by minimizing $E[C(t_B, N(t_B))]$ when a model is specified for the $N_i(t_B)$'s. The expectation $E[C(t_B, N(t_B))]$ is taken with respect to the unknown quantity $N(t_B)$ in $C(t_B, N(t_B))$. It is important to note that the counting process $N_i(t)$ is based on an unknown parameter vector Θ and it is more appropriate to write down $E[C(t_B, N(t_B))]$ as

$$E[C(t_B, N(t_B))|\Theta] = \frac{mc_P + c_R \sum_{i=1}^{m} E[N_i(t_B)|\Theta]}{t_B} \ . \tag{2.36}$$

Thus, a Bayesian optimal block replacement interval is determined by minimizing

$$E[C(t_B)] = E_\Theta\{E_{N(t_B)}[C(t_B, N(t_B))|\Theta]\}$$

with respect to t_B. The above requires evaluation of

$$E[C(t_B)] = \int E_{N(t_B)}[C(t_B, N(t_B))|\Theta] \ \pi(\Theta|D) \ d\Theta, \tag{2.37}$$

where D denotes the information available when the decision is made and $\pi(\Theta|D)$ is the probability distribution that represents the analyst's uncertainty about Θ when D is available and is referred to as the posterior distribution of Θ.

$N_i(t_B)$ is described by a semi-parametric PIM with cumulative intensity function $\Lambda_i(t_B)$. Determination of the optimal Bayesian block replacement interval requires the evaluation of $E[C(t_B)]$ given by (2.37), which involves the conditional cumulative intensity function $\Lambda_i(t_B|\Theta)$. Using the gamma process prior for the baseline cumulative intensity function, $\Lambda_0(t)$, $E[C(t_B)]$ can be evaluated by using posterior samples from $\pi(\Lambda_0(t), \beta|D)$. The posterior samples are obtained using the Gibbs sampler with a data augmentation step as discussed in the previous section and the expected cost is evaluated as

$$E[C(t_B)] \approx \frac{1}{S} \sum_{l=1}^{S} \frac{mc_P + c_R \sum_{i=1}^{m} [\Lambda_0(t_B)]_l \exp(\beta_l^T Z_i)}{t_B}. \tag{2.38}$$

The above can be minimized to obtain the optimal replacement interval t_B^*.

2.5 Application to Failure Data on Rail Sections

We have data on 132 sections of rail with observations varying over the life of each section, ranging from 3 MGT to 800 MGT. Grinding has been performed on the rail

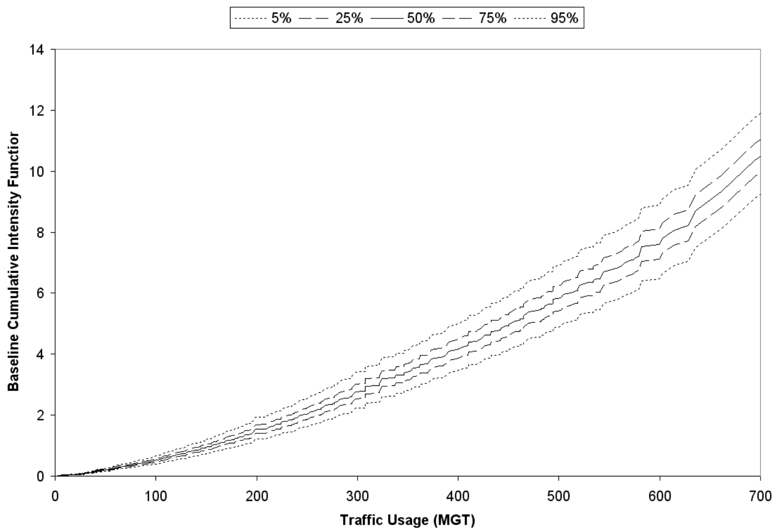

FIGURE 2.8
The posterior distribution of the baseline cumulative intensity function under the semiparametric model.

sections, but at different rates, varying from none to 1 mm per year. This a maintenance operation which is used for preventing derailments caused by rail fractures. This was the only covariate used in the analysis.

We considered the semiparametric model and applied the general data augmentation algorithm. In so doing, a priori, we assumed that the baseline cumulative intensity function took the power law form $\Lambda_0^*(t) = \alpha t^\gamma$, with $\alpha = 0.0005$ and $\gamma = 1.5$ are specified. This corresponds to an expected total of 11.3 failures over an 800 MGT lifetime with a moderately increasing failure intensity. However, to represent our uncertainty about this prior assumption, we set $c = 25$. A diffused normal prior was used for the covariate coefficient.In implementation of the algorithm we found 254 different interval endpoints and created a large data augmentation structure to analyze the data.

Based on the posterior analysis, as expected, grinding had a negative effect on failure intensity. Thus, higher levels of grinding and rail weight will result in less frequent replacements of rail sections.

In Figure 2.8 we present the posterior distribution of the cumulative intensity function $\Lambda_0(t)$ for the semiparametric model. We can see the jumps in the cumulative intensity as a result of using the gamma process prior.

To demonstrate our optimal replacement decision method, we assume that the cost of a planned replacement is 10 times the cost of a minimal repair, thus $c_R = 1$ and $c_P = 10$. We assume that we have two rail sections which must be replaced as a block. One of the sections is ground at 0.75 mm per year, while the other is ground

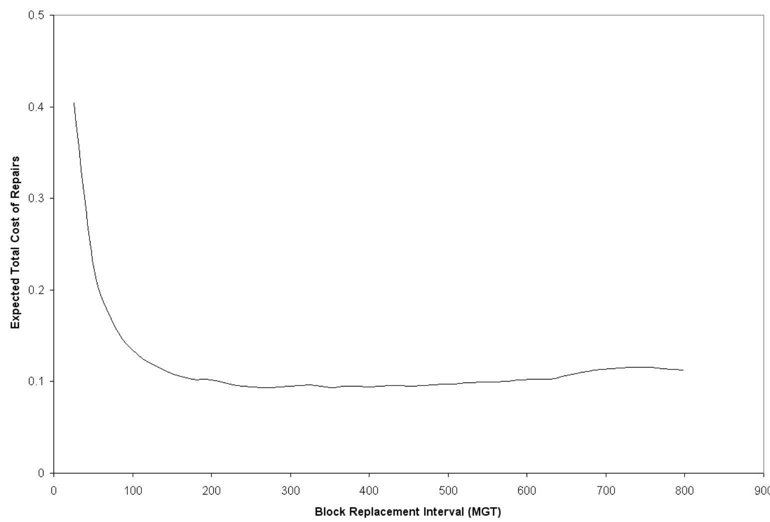

FIGURE 2.9
The expected total cost of repairs under the semiparametric model.

at 1 mm per year.

Figure 2.9 shows the expected total cost of curve for the semiparametric model. The optimal replacement interval under the semiparametric failure model (the minimum of the curve in Figure 2.9) is found to be 400, whereas under a comparable parametric model it is found to be 600 MGT. This is a reflection of the difference in the wear characteristics under the two failure models.

Overall, the semiparametric failure model is not restricted in its representation of the failure process. While the prior assumption takes the power law form, the posterior distribution of the baseline cumulative intensity function does not have to. Thus, the optimal replacement decision is driven by the actual characteristics of the failure process, not the parametric assumptions. While the analysis is more complex with the semiparametric model, our data augmentation algorithm simplifies this to iterative sampling from known distributions, thus allowing a more represenative model to be used in the optimal replacement decision.

References

Barlow, R. E and Hunter, L. (1960). Optimum preventive maintenance policies, *Operations Research*, , pp. 90-100.

Cho, D. I. and Parlar, M. (1991). A survey of maintenance models for multi-unit systems. *European Journal of Operational Research*, 51, pp. 1-23.

Cox, D. R. (1962). *Renewal Theory*. Methuen, London.

Cox, D. R. (1972a). Regression models and life tables. *J. R. Statist. Soc. B*, 34, pp. 187-220.

Cox, D. R. (1972b). The statistical analysis of dependencies in point processes. In *Stochastic Point Processes*. Ed. P. A. W. Lewis, pp. 55-66, New York Wiley.

Dayanik, S. and Gurler, U. (2002). An adaptive Bayesian replacement policy with minimal repair, *Operations Research*, 50, pp. 552-558.

Dellaportas, P., and Smith, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Appl. Statist.*, 42, pp. 443-459.

Doksum, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probability*, 2, pp. 183-201.

Frees, E. W. and Ruppert, D. (1985). Sequential non-parametric age replacement policies. *Ann. Statist.*, 13, pp. 650-652.

Gilks, W. R. and Wild, P. (1993). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, 41, pp. 337-348.

Kuo, L. and Ghosh, S. K. (2001). Bayesian Nonparametric Inference for Nonhomogeneous Poisson Processes, *Technical Report, North Carolina State University. Institute of Statistics Mimeo Series* No. 2530.

Mazzuchi, T. A. and Soyer R. (1995). A Bayesian perspective on some replacement strategies. *Reliability Engineering and System Safety*, 51, pp. 295-303.

Mazzuchi, T. A. and Soyer R. (1996). Adaptive Bayesian Replacement Strategies. In *Bayesian Statistics 5*. Ed. J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. Smith, pp. 667-674, Oxford University Press.

Ozekici S. (1995). Optimal maintenance policies in random environments. *European Journal of Operational Research*, 82, pp. 283-294.

Ross, S. M. (1989). *Introduction to Probability Models*. Academic Press, NY.

Singpurwalla, N. D. (1997). Gamma processes and their generalizations: an overview. In *Engineering Probabilistic Design and Maintenance for Flood Protection*, Ed. R. Cooke et al, pp. 67-75, Kluwer, New York.

Tanner, T. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *J. Am. Statist. Ass.*, 82, pp. 528-549.