Semiparametric Bayesian Decision Models for Optimal Replacement

Jason R. Merrick[1] and Refik Soyer[2]
[1]Department of Statistical Sciences and Operations Research
Virginia Commonwealth University
[2]Department of Decision Sciences
The George Washington University

April 30, 2007

Abstract

We present a Bayesian decision theoretic approach for developing replacament strategies. In so doing, we consider a semi-parametric model to describe the failure characteristics of systems 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.

1 Introduction

Many systems experience aging or wear as a function of time or usage. In industry it is common practice to use planned replacement strategies for these systems to prevent in-service failures that may be very costly relative to the cost associated with a planned replacement/repair. For example, 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 operations research literature on preventive maintenance assumes that the failure characteristics (models) of the systems are known and does not address statistical issues. Thus, the replacement strategies are dependent on such failure characteristics and they are not adaptive to learning from observed failure/replacement data; see for example Cho & Parlar (1991) for a general review of such models. More recently, statistical issues in the development of optimal replacement strategies have been considered by Mazzuchi & Soyer (1995, 1996) and Dayanik & Gurler (2002) using Bayesian approaches. These authors considered parametric Bayesian approaches that does not allow a flexible modeling strategy and their approaches do not allow use of covariate information in determining optimal strategies. For example, in addition to wear, the failure of a rail track is affected by factors such as rail weight, rail curvature and speed as well as the implementation of preventive measures such as the grinding of the surface of the rail and the lubrication of the track to reduce friction. Thus, these factors as well as the wear behavior should be considered by any model describing the failure behavior of rail tracks and should be taken into account developing replacement strategies. This is essential for systems such as rail tracks that are subject to block replacement. Non-parametric replacement strategies have been considered from a sampling theory perspective in Frees & Ruppert (1985) and Aras & Whitaker (1991) where adaptive age replacement policies are developed without using covariate information. 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 & Wong (1987), to evaluate posterior predictive distributions. Our approach enables us to obtain adaptive replacement strategies via updating our uncertainities and policies as we learn from the failure/replacement process.

Synopsis of our paper is as follows. In Section 2 we discuss the basics of block replacement and motivate the block replacement with minimal repair protocol that applies to repairable systems such as railroad tracks. We present the Bayesian decision theoretic set up for optimal replacement problem and its components using cost-based utility (loss) functions. In Section 3, a modulated Poisson process model is presented for describing the failure behavior of systems, such as rail 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). Analogous to PHM, in the PIM, the intensity function of the process is modeled as a product of a baseline intensity, which is a function of traffic usage in the railroad application. First a parametric approach is considered for the PIM. Due to the lack of controlled testing facilities for railroad tracks and the wide variation in the physical characteristics and operating environments of tracks in use, there is little evidence to support a choice of a fully parametric PIM. Thus, the parametric assumption is relaxed and a semi-parametric model is proposed for the PIM, using a gamma process prior for the baseline cumulative intensity. In Section 4 Bayesian analyses of the parametric and semiparametric PIMs are considered. The fully parametric approach adopts the MCMC methods of Dellaportas & Smith (1993) to perform parametric inference on the PIM where the number of rail track failures are described by a nonhomogeneous Poisson process (NHPP). The MCMC based procedures 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. In Section 5, Bayesian replacement strategies are developed for rail tracks by accounting for covariate information using the parametric and semiparametric models and an illustration of the approach is presented using actual rail track failure data.

2 Block Replacement Problem

Two of the most commonly used replacement strategies for systems/items subject to aging/wear are the age and block replacement; see Cox (1962) for an earlier introduction. Under the age replacement protocol, a planned replacement is made at age t_A, if the item survives until then, or an in-service replacement is made whenever the item fails. 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. For non-repairable systems/components failed units are always replaced by a new one and thus they are assumed to operate under the *good as new* (GN) replacement protocol. Analysis of this protocol involves modeling via the renewal function, but this is not applicable for systems like railroad tracks. Another block replacement protocol is the block replacement with minimal repair that was originally introduced by Barlow & Hunter (1960). Under this protocol 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.

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. Thus, in what follows we will consider the block replacement with minimal repair protocol and present the Bayesian decision theoretic formulation.

2.1 Bayesian Formulation of the Block Replacement Problem

The Bayesian approach, in addition to providing coherent inference, also provides a coherent framework for decision making. In Bayesian paradigm optimal strategies are chosen by maximizing expected utility. As in any decision problem, the Bayesian approach to the optimal replacement problem requires specification of three components:

(i) a utility (loss) function that reflects the consequences of selecting a specific replacement interval t_B;

(ii) a probability model describing the failure behavior of the system (or the component);

(iii) a prior distribution reflecting the analyst's a priori beliefs about all unknown components of the model.

As mentioned in the above, when a crack is found on a section of the rail track, the rail track is minimally repaired in the sense of Barlow & Hunter (1960). This implies that the rail tracks can be repaired in such a way that their failure characteristics (failure rates, reliability, etc.) are as they were just prior to the observation of the crack. To introduce some notation let c_P to denote the cost of a planned replacement and c_F to denote the cost of a minimal repair such that $c_F > c_P$. As pointed out by Mazzuchi & Soyer (1996), the cost per unit time for the i-th rail section is given by

$$C(t_B, N_i(t_B)) = \frac{mc_P + c_F N_i(t_B)}{t_B}, \tag{1}$$

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

$$C(t_B, N(t_B)) = \sum_{i=1}^{m} C(t_B, N_i(t_B)), \tag{2}$$

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

$$E[C(t_B, N(t_B))|\Theta] = \frac{mc_P + c_F \sum_{i=1}^{m} E[N_i(t_B)|\Theta]}{t_B} \ . \tag{3}$$

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

$$E[C(t_B)] = \int E_{N(t_B)}[C(t_B, N(t_B))|\Theta] \ \pi(\Theta|D) \ d\Theta, \tag{4}$$

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. An attractive feature of the Bayesian optimal replacement strategies is that they are adaptive. As one learns about the processes $N_i(t)$, $i = 1, ..., m$ based on D, the optimal interval t_B will be revised accordingly. The ability to update strategies as well as uncertainities over time is a natural outcome of the Bayesian approach. If the decision is made based on prior information D_0 then $\pi(\Theta|D_0)$ is referred to as the prior distribution of Θ. Thus, the components of the Bayesian decision framework consist of the cost-based utility function $C(t_B, N(t_B))$, the probability model of $N_i(t_B|\Theta)$ that describes the failure characteristics of the rail sections and the prior distribution $\pi(\Theta|D_0)$ representing uncertainty about parameters Θ of the probability model.

Often in reliability studies, upon failure items/systems can be restored to some satisfactory level of performance without replacing the whole unit. Such systems are referred to as repairable systems, see Crowder et al (1991). 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 & Hunter (1960), the number of failures of the i-th rail section in the replacement interval, $N_i(t_B)$, 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.

3 Proportional Intensities Model for Rail Section Failures

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 grinding problem analysis, the available covariates are constant with respect to traffic usage. $N_i(t)$ is described by a

NHPP with intensity function

$$\lambda_i(t) = \frac{d}{dt} E[N_i(t)]. \tag{5}$$

To reflect thefact 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

$$\lambda_i(t; Z_i) = \lambda_0(t) e^{\beta^T Z_i}, \tag{6}$$

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 (6) 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. In the modulated Poisson process model the ratio of the intensity functions of two rail sections at traffic usage t is given by

$$\frac{\lambda_i(t; Z_i)}{\lambda_j(t; Z_j)} = e^{\beta^T (Z_i - Z_j)}, \tag{7}$$

which does not depend on t. 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

$$\Lambda_i(t; Z_i) = \Lambda_0(t) e^{\beta^T Z_i}, \tag{8}$$

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

$$P(N_i(t) = n | \Lambda_0(t), \beta, Z_i) = \frac{\Lambda_0(t)^n e^{n\beta^T Z_i}}{n!} \exp\{-\Lambda_0(t) e^{\beta^T Z_i}\}. \tag{9}$$

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

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

In the optimal replacement problem setup, evaluation of the expected cost $E[C(t_B, N(t_B))|\Theta]$ requires $E[N_i(t_B)|\Theta]$ where $\Theta = (\Lambda_0(t), \beta)$. This can be obtained as

$$E[C(t_B, N(t_B))|\Theta] = \frac{mc_P + c_F \sum_{i=1}^{m} \Lambda_0(t_B)e^{\beta^T Z_i}}{t_B}. \tag{10}$$

3.1 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

$$\Lambda_0(t; \theta) = \alpha t^{\gamma}, \tag{11}$$

where $\theta = (\alpha, \gamma)$ and $\alpha > 0$, $\gamma > 0$. In the power law model (11), 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 systems such as rail tracks that are subject to wear. The power law model implies that the distribution of the time to the first failure call arrival is a Weibull distribution and thus sometimes the NHPP with the power law intensity function is referred to as *Weibull process*. We will refer to the PIM with a parametric form of $\lambda_0(t; \theta)$ as the parametric PIM. 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.

As previously discussed, 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 $\Lambda_0(t)$. Furthermore, $\Lambda_0(t)$ is a function taking values in $[0, \infty)$. 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

$$\Lambda_0(t_j) = \sum_{l=1}^{j} r_l. \tag{12}$$

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

$$r_j \backsim G(c\Lambda_0^*(t_j) - c\Lambda_0^*(t_{j-1}), c), \tag{13}$$

where $X \backsim G(a, b)$ 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,

$$(\Lambda_0(t)|D_0) \backsim G(c\Lambda_0^*(t), c), \tag{14}$$

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.*

4 Bayesian Analysis of the PIMs

In this section, we will present Bayesian inference for the semiparametric PIMs. In so doing, we first present the Bayesian analysis of the parametric PIM using an adoption of MCMC methods of Dellaportas & Smith (1993) presented for 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.

4.1 Bayesian Inference for the Parametric PIM

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

$$L_i(\theta,\beta;D_i) = \prod_{j=1}^{r_i} \frac{\left(\{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\}\, e^{\beta^T Z_i}\right)^{n_i(t_{i,j})-n_i(t_{i,j-1})}}{(n_i(t_{i,j}) - n_i(t_{i,j-1}))!} \times \exp\{-\{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\}\, e^{\beta^T Z_i}\}, \quad (15)$$

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

$$L(\theta,\beta;D) = \prod_{i=1}^{m} L_i(\theta,\beta;D_i). \quad (16)$$

The joint posterior distribution of θ and β given D is

$$\pi(\theta,\beta|D) \propto L(\theta,\beta;D)\pi(\theta,\beta),$$

which 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 $\pi(\theta,\beta|D)$. Implementation of the Gibbs sampler for the parametric PIM requires draws from the full conditional posterior distributions $\pi(\theta|\beta,D)$ and $\pi(\beta|\theta,D)$. If we assume θ and β are independent apriori, that is, $\pi(\theta,\beta) = \pi(\theta)\ \pi(\beta)$, then the full conditionals are given by

$$\pi(\theta|\beta,D) \propto \{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\}^{\sum_{i=1}^{m}\sum_{j=1}^{r_i} n_i(t_{i,j})-n_i(t_{i,j-1})} \times \exp\{-\sum_{i=1}^{n}\sum_{j=1}^{r_i}\{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\}\, e^{\beta^T Z_i}\}\pi(\theta) \quad (17)$$

and

$$\pi(\beta|\theta,D) \propto e^{\sum_{i=1}^{m}\sum_{j=1}^{r_i}(n_i(t_{i,j})-n_i(t_{i,j-1}))\beta^T Z_i} \times \exp\{-\sum_{i=1}^{m}\sum_{j=1}^{r_i}\{\Lambda_0(t_{i,j};\theta) - \Lambda_0(t_{i,j-1};\theta)\}\, e^{\beta^T Z_i}\}\pi(\beta). \quad (18)$$

For any choice of the forms of $\Lambda_0(t;\theta)$, $\pi(\theta)$ and $\pi(\beta)$, (17) and (18) are logconcave densities and therefore the adaptive rejection sampling algorithm of Gilks & Wild (1992) can be used to draw samples from (17) and (18) at each iteration of the Gibbs sampler. However, some specific forms may yield known distributions for components of θ in (17). For example, if the power law form is specified for $\Lambda_0(t;\alpha,\gamma) = \alpha t^\gamma$, with $\theta = (\alpha,\gamma)$, then a gamma prior density for α shape a and scale b, yields the posterior full conditional distribution as a gamma distribution with shape $a + \sum_{i=1}^m \sum_{j=1}^{r_i} n_i(t_{i,j}) - n_i(t_{i,j-1})$ and scale $b+\sum_{i=1}^m \sum_{j=1}^{r_i} \left\{ t_{i,j}^\gamma - t_{i,j-1}^\gamma \right\} e^{\beta^T Z_i}$. But the full conditional of γ

$$\pi(\gamma|\alpha,\beta,D) \propto \prod_{i=1}^{m}\prod_{j=1}^{r_i} \left\{ t_{i,j}^\gamma - t_{i,j-1}^\gamma \right\}^{n_i(t_{i,j})-n_i(t_{i,j-1})} \times \exp\{-\sum_{i=1}^{m}\sum_{j=1}^{r_i} \alpha \left\{ t_{i,j}^\gamma - t_{i,j-1}^\gamma \right\} e^{\beta^T Z_i}\}\pi(\gamma). \tag{19}$$

does not simplify to a known distribution for any form of $\pi(\gamma)$ and thus still requires use of the adaptive rejection sampling algorithm to obtain samples. For any choice of the prior $\pi(\beta)$, sampling from (18) requires adaptive rejection sampling algorithm. A reasonable form for $\pi(\beta)$ is the multivariate normal density.

Once a posterior sample has been obtained from $\pi(\theta,\beta|D)$, all the marginal distributions of θ and β as well as their moments can be approximated from this posterior sample. Let $(\theta_l, \beta_{l,1}, \ldots, \beta_{l,p})$, for $l = 1, \ldots, S.$, be a sample from the posterior distribution generated using the method outlined above. An estimate of the posterior mean of β_j is $\hat{\beta}_j = \frac{1}{S}\sum_{l=1}^S \beta_{l,j}$ and the posterior cumulative distribution function of β_j can be approximated by

$$P(\beta_j \le b) \approx \frac{1}{S}\sum_{l=1}^{S} I(\beta_{l,j} \le b). \tag{20}$$

Similar estimates can be obtained for the parameters of the baseline cumulative intensity function. The expected posterior predictive probability of observing n failures of the i-th rail section up to traffic usage t can be approximated by

$$P(N_i(t) = n|D) \approx \frac{1}{S}\sum_{l=1}^{S} \frac{\left(\Lambda_0(t;\theta_l)e^{\beta_l^T Z_i}\right)^n}{n!} \exp\{-\Lambda_0(t;\theta_l)e^{\beta_l^T Z_i}\}. \tag{21}$$

4.2 Bayesian Inference for the Semiparametric PIM

The use of a gamma process prior for the cumulative intensity function of a NHPP was considered by Kuo & 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 semi-parametric 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 a general algorithm will be presented.

4.2.1 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 1. The cumulative intensity for the i-th rail section is

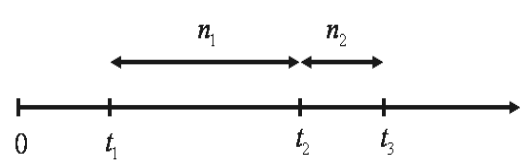

Figure 1: The data observed for a single rail section.

given by

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

where $\Lambda_0(t)$ is defined by (14). Using the independent increments property of the NHPP, the likelihood function of $\Lambda_0(t)$ and β given D_i is written as

$$\begin{aligned} L_i(\Lambda_0(t), \beta|D) &= \prod_{j=1}^{2} \frac{\left(\Lambda_0(t_{j+1})e^{\beta^T Z_i} - \Lambda_0(t_j)e^{\beta^T Z_i}\right)^{n_j}}{n_j!} && (22) \\ &\quad \times \exp\{-(\Lambda_0(t_{j+1})e^{\beta^T Z_i} - \Lambda_0(t_j)e^{\beta^T Z_i})\}. && (23) \end{aligned}$$

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

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

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

$$\begin{aligned} (\Lambda_0(t)|\beta, D_i) &\backsim G(c\Lambda_0^*(t), c), \text{ for } t < t_1 && (24) \\ (\Lambda_0(t_2) - \Lambda_0(t_1)|\beta, D_i) &\backsim G(c\{\Lambda_0^*(t_2) - \Lambda_0^*(t_1)\} + n_1, c + e^{\beta^T Z_i}), && (25) \\ (\Lambda_0(t_3) - \Lambda_0(t_2)|\beta, D_i) &\backsim G(c\{\Lambda_0^*(t_3) - \Lambda_0^*(t_2)\} + n_2, c + e^{\beta^T Z_i}), && (26) \\ (\Lambda_0(t) - \Lambda_0(t_3)|\beta, D_i) &\backsim G(c\{\Lambda_0^*(t) - \Lambda_0^*(t_3)\}, c), \text{ for } t > t_3. && (27) \end{aligned}$$

It follows from (25) and (26) that

$$\Lambda_0(t_2) = [\Lambda_0(t_2) - \Lambda_0(t_1)] + \Lambda_0(t_1) \tag{28}$$

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

$$(\Lambda_0(t_1)|\beta, D_i) \backsim G(c\Lambda_0^*(t_1) + n_0, c + e^{\beta^T Z_i}) \tag{29}$$

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

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

It can be seen that conditional on β, the effect of the covariates is on the scale of the gamma distributed posterior of the cumulative intensity function. However, unconditional on β this is not the case. Following the prior treatment of the parametric PIM, the prior distributions of the parameters β_j are assumed to be independent normals

$$\beta_j \sim Normal(\mu_j, \sigma_j^2). \tag{30}$$

Implementation of the Gibbs sampler requires the full conditional distribution of each β_j

$$(\beta_j|\beta^{(-j)}, \Lambda_0(t_2) - \Lambda_0(t_1), \Lambda_0(t_3) - \Lambda_0(t_2), D), \tag{31}$$

which is proportional to

$$e^{(n_1+n_2)Z_{i,j}\beta_j} \exp\{-([\Lambda_0(t_3) - \Lambda_0(t_1)]e^{\beta^{(-j)T} Z_i^{(-j)}})e^{\beta_j Z_{i,j}}\} \exp\{-\frac{1}{2\sigma_j^2}(\beta_j - \mu_j)^2\}, \tag{32}$$

for the specific example with likelihood (22), where $\beta^{(-j)} = (\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_p)$ and $Z_i^{(-j)} = (Z_{i,1}, \ldots, Z_{i,j-1}, Z_{i,j+1}, \ldots, Z_{i,p})$. Again samples from (32) can be obtained using the adaptive rejection sampling as each density is log-concave.

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

$$(\Lambda_0(t_2) - \Lambda_0(t_1), \Lambda_0(t_3) - \Lambda_0(t_2), \beta|D_i), \tag{33}$$

using (25) and (26) whose sum is a sample point for $[\Lambda_0(t_3) - \Lambda_0(t_1)]$, which in turn yields a sample point from (32) 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 which will be discussed in Section 5.

4.2.2 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. One way to update the distribution of $\Lambda_0(t^*)$ is through

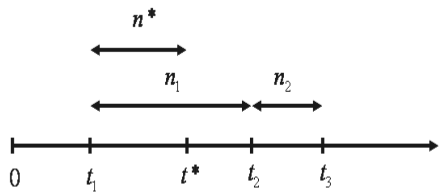

Figure 2: The prediction problem for a single rail section.

a data augmentation 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

$$\Lambda_0(t^*) = [\Lambda_0(t^*) - \Lambda_0(t_1)] + \Lambda_0(t_1), \tag{34}$$

where $(\Lambda_0(t_1)|\beta, D_i)$ is given by (24) and

$$(\Lambda_0(t^*) - \Lambda_0(t_1)|\beta, D_i, n^*) \backsim G(c\{\Lambda_0^*(t^*) - \Lambda_0^*(t_1)\} + n^*, c + e^{\beta^T Z_i}). \tag{35}$$

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

$$(\Lambda_0(t_2) - \Lambda_0(t^*)|\beta, D_i) \backsim G(c\{\Lambda_0^*(t_2) - \Lambda_0^*(t^*)\} + (n_1 - n^*), c + e^{\beta^T Z_i}), \tag{36}$$

and $\Lambda_0(t_3) - \Lambda_0(t_2)$ is still given by (26). The above results follow from the independent increments property of the gamma process.

The implementation of the Gibbs sampler requires specification of ($\beta|$ $\Lambda_0(t)$, D_i, n^*) which is a distribution similar to (32) in the specific example 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

$$(N_i(t^*) - N_i(t_1) = n^*|\Lambda_0(t), \beta, D_i) \sim Bin\left[n_1, \frac{\Lambda_0(t^*) - \Lambda_0(t_1)}{\Lambda_0(t_2) - \Lambda_0(t_1)}\right], \tag{37}$$

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

4.2.3 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

$$\begin{aligned} L(\Lambda_0(t), \beta; D) &= \prod_{i=1}^{2} \frac{\left[(\Lambda_0(t_{i,2}) - \Lambda_0(t_{i,1}))\, e^{\beta^T Z_i}\right]^{n_{i,1}}}{n_{i,1}!} \\ &\quad \exp\left\{-\,(\Lambda_0(t_{i,2}) - \Lambda_0(t_{i,1}))\, e^{\beta^T Z_i}\right\}. \end{aligned} \tag{38}$$

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 section 4.2.1 if the observed intervals for the two rail sections are not overlapping as shown in Figure 3. In this particular case, using the independent increments

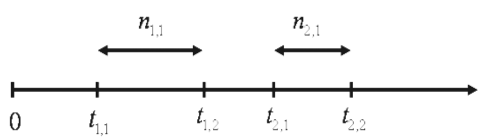

Figure 3: The case of non-overlapping intervals for two rail sections.

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

$$\begin{aligned} \pi(\Lambda_0(t)|\beta, D) &\propto (\Lambda_0(t_{1,2}) - \Lambda_0(t_{1,1}))^{c\left(\Lambda_0^*(t_{1,2}) - \Lambda_0^*(t_{1,1})\right) + n_{1,1} - 1} \\ &\quad \times \exp\left\{-\,(\Lambda_0(t_{1,2}) - \Lambda_0(t_{1,1}))\,(c + e^{\beta^T Z_1})\right\} \\ &\quad \times (\Lambda_0(t_{2,2}) - \Lambda_0(t_{2,1}))^{c\left(\Lambda_0^*(t_{2,2}) - \Lambda_0^*(t_{2,1})\right) + n_{2,1} - 1} \\ &\quad \times \exp\left\{-\,(\Lambda_0(t_{2,2}) - \Lambda_0(t_{2,1}))\,(c + e^{\beta^T Z_2})\right\} \end{aligned} \tag{39}$$

implying that

$$
\begin{aligned}
(\Lambda_0(t)|\beta, D) &\backsim G(c\Lambda_0^*(t), c),\ \text{for } t < t_{1,1} \\
(\Lambda_0(t_{1,2}) - \Lambda_0(t_{1,1})|\beta, D) &\backsim G(c\{\Lambda_0^*(t_{1,2}) - \Lambda_0^*(t_{1,1})\} + n_{1,1}, c + e^{\beta^T Z_1}), \\
(\Lambda_0(t_{2,1}) - \Lambda_0(t_{1,2})|\beta, D) &\backsim G(c\{\Lambda_0^*(t_{2,1}) - \Lambda_0^*(t_{1,2})\}, c), \qquad (40) \\
(\Lambda_0(t_{2,2}) - \Lambda_0(t_{2,1})|\beta, D) &\backsim G(c\{\Lambda_0^*(t_{2,2}) - \Lambda_0^*(t_{2,1})\} + n_{2,1}, c + e^{\beta^T Z_2}), \\
(\Lambda_0(t) - \Lambda_0(t_{2,2})|\beta, D) &\backsim G(c\{\Lambda_0^*(t) - \Lambda_0^*(t_{2,2})\}, c),\ \text{for } t > t_{2,2}.
\end{aligned}
$$

Updating for other portions of $\Lambda_0(t)$ follows along the same lines as presented in section 4.2.1. Similarly, sampling from the full conditional of β given $\Lambda_0(t)$ and D is achieved via the use of adaptive rejection sampling.

If both rail sections are observed for the same traffic usage interval, as in Figure 4, then updating is again straightforward. In this case, it can be shown that the full conditional of $\Lambda_0(t)$ can be obtained as

$$
\begin{aligned}
(\Lambda_0(t)|\beta, D) &\backsim G(c\Lambda_0^*(t), c), \text{for } t < t_1 \\
(\Lambda_0(t_2) - \Lambda_0(t_1)|\beta, D) &\backsim G(c\{\Lambda_0^*(t_2) - \Lambda_0^*(t_1)\} + \sum_{i=1}^{2} n_{i,1}, c + \sum_{i=1}^{2} e^{\beta^T Z_i}), \\
(\Lambda_0(t) - \Lambda_0(t_2)|\beta, D) &\backsim G(c\{\Lambda_0^*(t) - \Lambda_0^*(t_2)\}, c),\ \text{for } t > t_2. \qquad (41)
\end{aligned}
$$

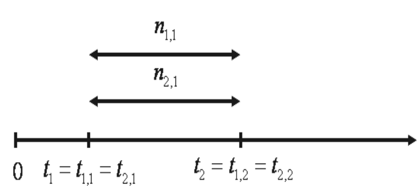

Figure 4: The case of identical intervals for two rail sections.

4.2.4 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 5. Updating $\Lambda_0(t)$ given β then requires the use of a data augmentation step in the Gibbs sampler as discussed in section 4.2.1. As the intervals overlap,

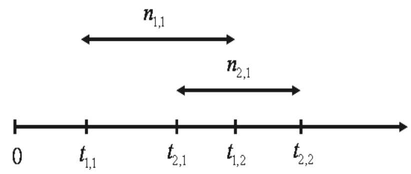

Figure 5: The case of overlapping intervals for two rail sections.

$N_1(t_{12}) - N_1(t_{11})$ 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 $[t_{11}, t_{21})$ 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 6. Then it follows from the above that

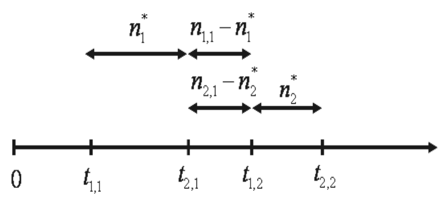

Figure 6: The failure counts required for data augmentation.

$$\begin{aligned}(\Lambda_0(t_{21}) - \Lambda_0(t_{11}) \quad & |\beta, n_1^*, D) \\ & \sim G(c[\Lambda_0^*(t_{21}) - \Lambda_0^*(t_{11})] + n_1^*, c + e^{\beta^T Z_1}),\end{aligned} \tag{42}$$

$$\begin{aligned}&(\Lambda_0(t_{12}) - \Lambda_0(t_{21})|\beta, n_1^*, n_2^*, D)|\beta, n_1^*, n_2^*, D) \\ &\sim G(c[\Lambda_0^*(t_{12}) - \Lambda_0^*(t_{21})] + (n_{1,1} - n_1^*) + (n_{2,1} - n_2^*), c + \sum_{i=1}^{2} e^{\beta^T Z_i} \ ])\end{aligned} \tag{43}$$

and

$$\begin{aligned}(\Lambda_0(t_{22}) - \Lambda_0(t_{12}) \quad & |\beta, n_2^*, D) \\ & \sim G(c[\Lambda_0^*(t_{22}) - \Lambda_0^*(t_{12})] + n_2^*, c + e^{\beta^T Z_2}).\end{aligned} \tag{44}$$

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

$$
\begin{aligned}
(N_1(t_{2,1}) - N_1(t_{1,1}) \quad & |n_{1,1}, \tfrac{\Lambda_0(t_{2,1})-\Lambda_0(t_{1,1})}{\Lambda_0(t_{1,2})-\Lambda_0(t_{1,1})}) \\
& \sim Bin(n_{1,1}, \tfrac{\Lambda_0(t_{2,1})-\Lambda_0(t_{1,1})}{\Lambda_0(t_{1,2})-\Lambda_0(t_{1,1})})
\end{aligned} \tag{45}
$$

and

$$
\begin{aligned}
(N_2(t_{2,2}) - N_2(t_{1,2}) \quad & |n_{2,1}, \tfrac{\Lambda_0(t_{2,2})-\Lambda_0(t_{1,2})}{\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1})}) \\
& \sim Bin(n_{2,1}, \tfrac{\Lambda_0(t_{2,2})-\Lambda_0(t_{1,2})}{\Lambda_0(t_{2,2})-\Lambda_0(t_{2,1})}),
\end{aligned} \tag{46}
$$

where (45) and (46) are independent binomial random variables.

4.2.5 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 7 for the case of three NHPP's. For any

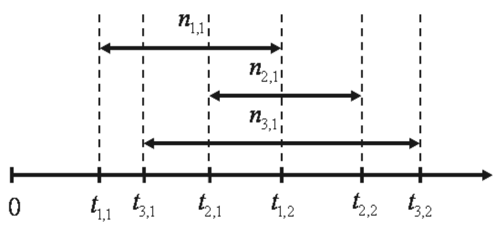

Figure 7: The case of three overlapping intervals.

observed interval that has now been broken up 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 $D = (D_i; i = 1, \ldots, m)$ from m rail sections where $D_i = \{N_i(t) = n_i(t), j = 1, \ldots, r_i, Z_i\}$. Our data consists of r_i inspection runs for the i-th rail section where the inspections are performed at $t_{i,0}, \ldots, t_{i,r_i}$

MGTs and $n_i(t_{i,1}) - n_i(t_{i,0}), \ldots, n_i(t_{i,r_i}) - n_i(t_{i,r_i-1})$ 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 $t_1^*, \ldots, t_q^*$ denote the ordered list of q unique values amongst the interval endpoints $t_{i,j}$ for $j = 1, \ldots, r_i$ and $i = 1, \ldots, m$. The ordered values t_k^* for $k = 1, \ldots, q$ 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 $k = 1, \ldots, q-1$, 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 $\{l_{i,j}^1, \ldots, l_{i,j}^{m_{i,j}^*}\}$.

In the example given in Figures 5 and 6 of 4.2.2, we have $r_1 = r_2 = 1$, with $q = 4$ interval endpoints as $t_1^* = t_{11}, t_2^* = t_{21}, t_3^* = t_{12}$ and $t_4^* = t_{22}$. According to our notation we have the index sets $B_1^* = \{1\}$, $B_2^* = \{1, 2\}$ and $B_3^* = \{2\}$. For rail section 1, the number of unknown failures in $[t_1^*, t_2^*)$ is $N_{1,1}^*$ which is given by n_1^* in Figure 6. Similarly, $N_{1,2}^*$, number of unknown failures in $[t_2^*, t_3^*)$ is given by $(n_{11} - n_1^*)$ and $N_{1,3}^* = n_2^*$ in the figure. For the second rail section we have $N_{2,1}^* = n_1^*$, $N_{2,2}^* = (n_{21} - n_2^*)$ and $N_{2,3}^* = n_2^*$ in Figure 6. Since each rail section is observed for a single interval in the example, the set of endpoints are given by $S_{1,1}^* = \{t_1^*, t_2^*\}$ and $S_{2,1}^* = \{t_2^*, t_3^*\}$ with $m_{1,1}^* = 2$ and $m_{2,1}^* = 2$ implying that $\{l_{1,1}^1 = t_1^*, l_{1,1}^2 = t_2^*\}$ and $\{l_{2,1}^1 = t_2^*, l_{2,1}^2 = t_3^*\}$.

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

$$(\Lambda_0(t_{k+1}^*) - \Lambda_0(t_k^*)\ |N^*, \beta, D) \backsim G(c[\Lambda_0^*(t_{k+1}^*) - \Lambda_0^*(t_k^*)] + \sum_{i \epsilon B_k^*} N_{i,k}^*, c + \sum_{i \epsilon B_k^*} e^{\beta^T z_i}) \tag{47}$$

for $k = 1, \ldots, q$. Note that for $t < t_1^*$ we still have $(\Lambda_0(t)|\beta, D) \backsim G(c\Lambda_0^*(t), c)$. In order to obtain the distribution of $N_{i,k}^*$'s we define the vector $\underline{N_{i,j}^*} = (N_{i,k}^*; k \epsilon S_{i,j}^*)$ 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

$$(\underline{N^*_{i,j}}|\Delta, \beta, D) \backsim\ Mult(n_i(t_{i,j+1}) - n_i(t_{i,j}), p^*_{i,1}, \ldots, p^*_{i,m^*_{i,j}}) \tag{48}$$

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

$$p^*_{i,l} = \frac{\Lambda_0(l^{l+1}_{i,j}) - \Lambda_0(l^l_{i,j})}{\Lambda_0(l^{m^*_{i,j}}_{i,j}) - \Lambda_0(l^1_{i,j})}. \tag{49}$$

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

5 Block Replacement of Railroad Tracks with Minimal Repair

Under the minimal repair assumption for each rail section, in the cost equation

$$C(t_B, N_i(t_B)) = \frac{mc_P + c_F N_i(t_B)}{t_B},$$

$N_i(t_B)$ is described by a 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 (4), which involves the conditional cumulative intensity function $\Lambda_i(t_B|\Theta)$. Under the parametric set-up of section 3, the conditional cumulative intensity function for the i-th rail section is given by $\Lambda_i(t_B|\theta, \beta, Z_i) = \Lambda_0(t_B; \theta)e^{\beta^T Z_i}$. As discussed in Section 4, once the data is observed and the Bayesian updating is completed using the Gibbs sampler, we can obtain the optimal replacement interval using the posterior samples from the joint distribution of (θ, β). If a posterior sample, denoted (θ_l, β_l) for $l = 1, \ldots, S$, is available from $\pi(\theta, \beta|D)$, then $E[C(t_B)]$ can be approximated as

$$E[C(t_B)|Z_1, \ldots, Z_n] \approx \frac{1}{S}\sum_{l=1}^{S} \frac{mc_P + c_F \sum_{i=1}^{m} \Lambda_0(t_B; \theta_l) \exp(\beta_l^T Z_i)}{t_B} \tag{50}$$

and the optimal t^*_B is obtained by minimizing (50) with respect to t_B. If an adaptive strategy is used and if each $N_i(t)$ is observed at instants of traffic usage $t = t_{i,1}, \ldots, t_{i,r_i}$, $t_{i,r_i} < t_B$, $i = 1, \ldots, m$, during the replacement cycle then the distribution $\pi(\theta, \beta|D)$ can be updated using the likelihood function (16) and new replacement interval t_B^* can be determined by minimizing (50).

The semi-parametric PIM and its Bayesian approach presented in Section 4.2 can also be used in a similar manner for developing optimal block replacement policies. Using the gamma process prior for the baseline cumulative intensity function, $\Lambda_0(t)$, under the semi-parametric PIM $E[C(t_B)]$ 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 Section 4.2 and the expected cost is evaluated as

$$E[C(t_B)] \approx \frac{1}{S} \sum_{l=1}^{S} \frac{mc_P + c_F \sum_{i=1}^{m} [\Lambda_0(t_B)]_l \exp(\beta_l^T Z_i)}{t_B}. \tag{51}$$

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

5.1 Application to Failure Data on Rail Sections

We have data supplied by the Association of American railroads 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 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.

We first performed the parametric analysis of Section 4.1 with the power law form where $\theta = (\alpha, \gamma)$. The prior on α was the conditionally conjugate gamma distribution, with a mean of 0.0005 and high variance, and the prior on γ was a truncated normal distribution, with a mean of 1.5 and a high variance. Then we considered the semiparametric model and applied the general data augmentation algorithm of Section 4.2.5. 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 as equal to the prior means used in the parametric model. 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 $c = 25$. In implementation of the algorithm of Section 4.2.5, we found 254 different interval endpoints and created a large data augmentation structure to analyze the data. For both models we ran the Gibbs sampler, collecting 1000 samples after a warm-up of 200 samples. For each model, we assumed a normal prior on β with a mean of 0 and a standard deviation of 20, a diffuse prior.

In Figures 8 and 9, we present the posterior distribution of the cumulative intensity function $\Lambda_0(t)$ for the semiparametric and parametric models. We can see from the figures that the forms of the two functions are different. While the overall expected number of failures over the lifetime of a rail are similar, the speed of wear (rate of change of the slope) is different. The restrictions of the parametric form do not seem to allow the specific wear pattern shown in the data. This is also reflected in the optimal replacement intervals found under the two approaches. This difference in the intensity functions leads to a difference in the parameters reflecting the effect of the covariate, namely the level of maintenance grinding performed. Under the semiparametric model, the posterior distribution of the parameter β has a mean of 3.00 with a standard deviation of 0.23, which corresponds to a 0.5 mm per year level of grinding providing a 78% reduction in failures compared to no maintenance grinding. This can be contrasted to the parametric model which, due to its restriction on the form of the intensity function, predicts that the same 0.5 mm per year of grinding will provide a 58% reduction in the number of failures. This result seems to imply that we can get a better estimate of the effect of covariates under the semiparametric model. However, before we can jump to this conclusion, we must ask whether the semiparametric model actually fits the data better.

In order to infer which model describes the data better one can use Bayes factors. However, computation of the Bayes factors is difficult since the marginal likelihoods under the two competing models can not be directly approximated from the Gibbs sampler. An alternative is to use a model selection criterion such as the Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002). For a generic parameter vector Θ, DIC is defined as

$$DIC = \overline{\mathrm{D}} + p_{\mathrm{D}},$$

where D$= -2log\mathcal{L}(\Theta)$, is two times the negative loglikelihood, $\overline{\mathrm{D}} = E_{\Theta|data}[\mathrm{D}]$ and $p_{\mathrm{D}} = \overline{\mathrm{D}} - \mathrm{D}(\widehat{\Theta})$, where $\widehat{\Theta}$ is the posterior mean. The DIC has the general "fit+complexity" form used by many model selection criteria. In the above $\overline{\mathrm{D}}$ represents the "goodness of the fit" of the model where p_{D} represents a complexity penalty as reflected by the effective number of parameters of the model. In the table below we present the (DIC) for the parametric and semiparametric models. Since the lower values of DIC is preferable, the results show that there is strong evidence in favor of the semiparametric model.

Table 1: DIC Comparison of the Models

Model	$\overline{\mathrm{D}}$	p_{D}	DIC
Parametric	1447.17	5.84	1453.01
Semiparametric	1327.21	37.59	1364.80

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 $c_F = 1$ 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 at 1 mm per year.

Figures 10 and 11 show the expected total cost of repairs curves for the semiparametric and parametric failure models respectively. The optimal replacement interval under the semiparametric failure model (the minimum of the curve in Figure 10) is found to be 400, whereas under the parametric failure model it is found to be 600 MGT. This is a reflection of the difference in the wear characteristics under the two failure models. As the rate of wear appears to accelerate in the semiparametric model (Figure 8), so the rail should be replaced earlier. Whereas under the parametric model, the rate of wear increases slowly (Figure 9), implying that the replacement can be delayed.

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.

6 Conclusions

In this paper, we presented a Bayesian decision theoretic approach to the optimal replacement problem using semiparametric models. The type of

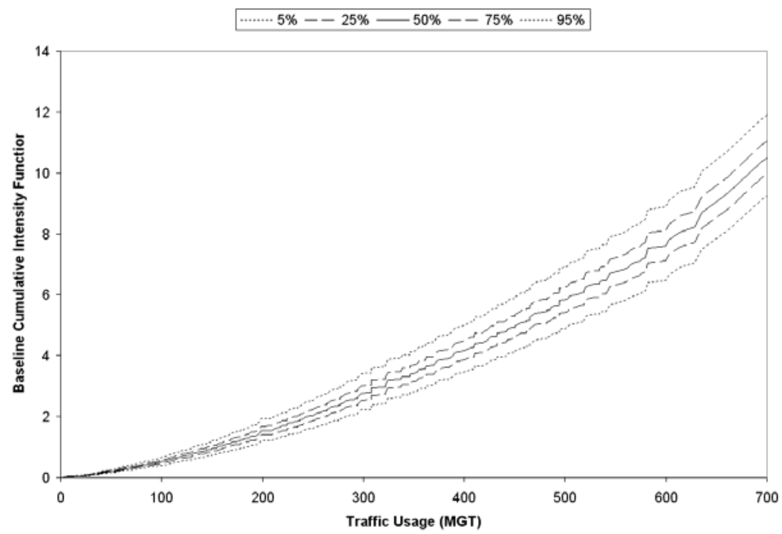

Figure 8: The posterior distribution of the baseline cumulative intensity function under the semiparametric model.

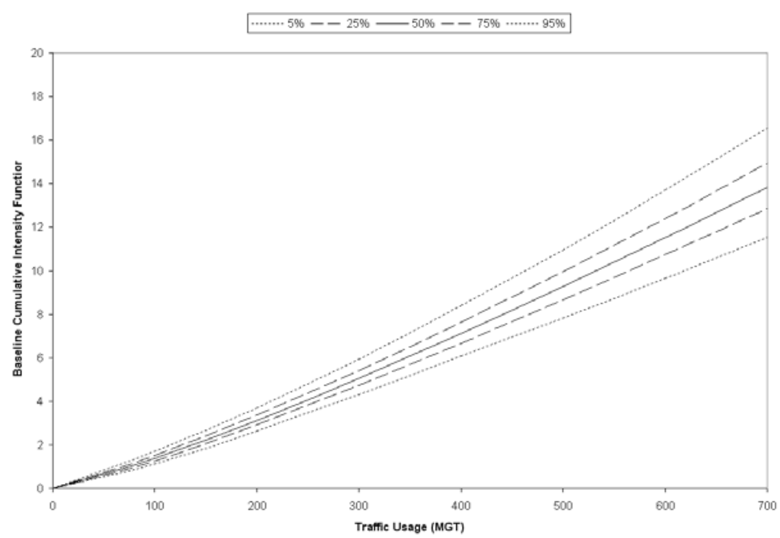

Figure 9: The posterior distribution of the baseline cumulative intensity function under the parametric model.

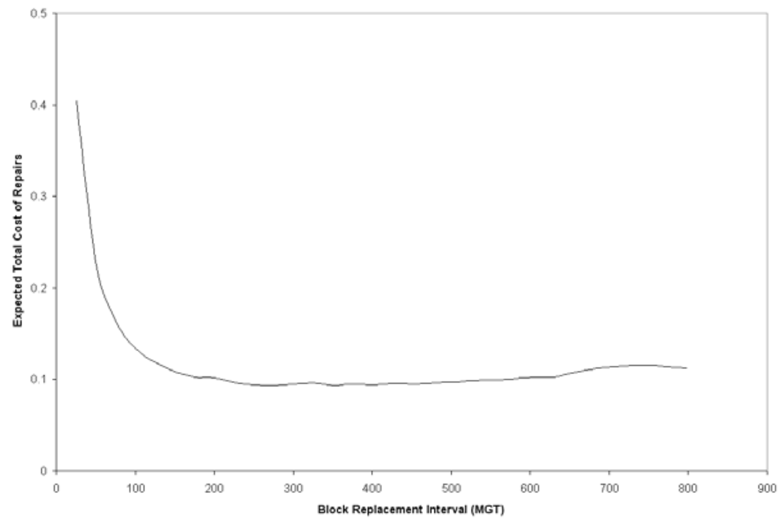

Figure 10: The expected total cost of repairs under the semiparametric model.

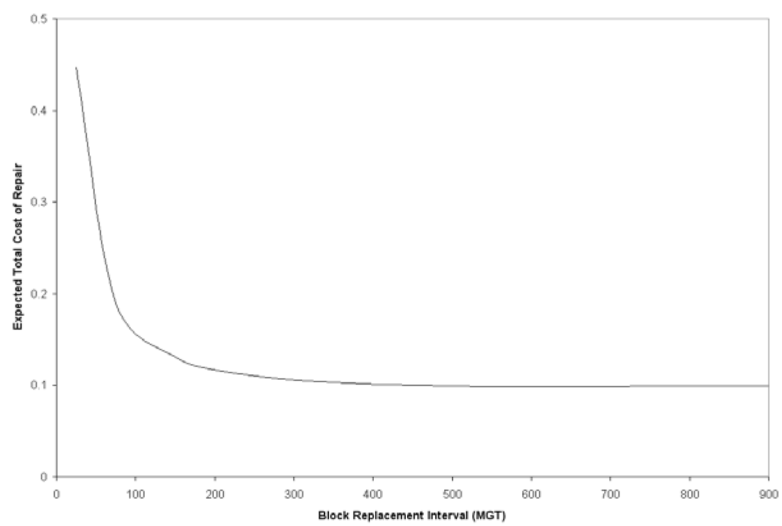

Figure 11: The expected total cost of repairs under the parametric model.

data that arise in analysis of these models required new MCMC methods. We developed a data augmentation algorithm within the Gibbs sampler which allowed us to do the posterior analysis and to obtain optimal replacement intervals. Although our semiparametric Bayesian approach was motivated by analysis of failure data on rail sections, our methods may have use in other applications. For example, one potential area of application is software reliability where the need for nonparametric Bayesian models was recently considered by Wilson & Samaniego (2004).

References

[1] Aras, G. and Whitaker, R. L. (1991). Sequential non-parametric estimation of an optimal age replacement policy. Tech. Rep. 158, Univ. of California, Santa Barbara.

[2] Campodonico, S. and Singpurwalla, N. D. (1995). Inference and predictions from Poisson point processes incorporating expert knowledge. J. of the Am. Statist. Assoc., Vol. 90, pp. 220-226.

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

[4] Cox, D. R. (1962). Renewal Theory. Methuen, London.

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

[6] 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.

[7] Crowder, M. J., Kimber A. C., Smith, R. L. and Sweeting, T. J. (1991). In Statistical Analysis of Reliability Data, Chapman & Hall, London.

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

[9] Dellaportas, P., & Smith, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. Appl. Statist., 42, No. 3, pp. 443-459.

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

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

[12] Gelfand, A. E. & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. J. Am. Statist. Ass., 85, pp. 398-409.

[13] Gelfand, A. E. (1997). Approaches for Bayesian semi-parametric regression. Technical report at the University of Connecticut. TR9711.

[14] Gelfand, A. E. (1996). Model determination using sampling-based methods. In Markov Chain Monte Carlo in Practice. Ed. W. R. Gilks, S. Richardson and D. J. Spiegelhalter, pp. 145-161, Chapman & Hall, London.

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

[16] Hjort, N. L. (1990). Non-parametric Bayes estimates based on Beta processes in models for life history data. Ann. Statist., Vol. 18, No. 3, pp. 1259-1294.

[17] Kalbfleisch, J. D. (1978). Non-parametric Bayes analysis of survival time data. J. R. Statist. Soc. B, Vol. 40, No. 2, pp. 214-221.

[18] Kalbfleisch, J. D. and Prentice, R. L. (1982). The Statistical Analysis of Failure Time Data. John Wiley, New York.

[19] 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.

[20] Law, A. M. and Kelton, W. D. (1991). Simulation modeling and analysis. 2nd edition, McGraw Hill, New York.

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

[22] 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.

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

[24] 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.

[25] Spiegelhalter, D. J., Best, N. G, Carlin, B. P. and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society, Series B, Vol. 64, pp. 1-34.

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

[27] Wilson, S. P. and Samaniego, F. Z. (2004). Nonparametric Analysis of Bug Counting Models in Software Reliability. Technical Report.