BAYESIAN STATISTICS 7, pp. 163–179
J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid,
D. Heckerman, A. F. M. Smith and M. West (Eds.)

Bayesian Nonparametric Inference for Mixed Poisson Processes

E. GUTIÉRREZ-PEÑA
Universidad Nacional Autónoma de México, Mexico
eduardo@sigma.iimas.unam.mx

L.E. NIETO-BARAJAS
Instituto Tecnológico Autónomo de México, Mexico
lnieto@itam.mx

SUMMARY

This paper presents a Bayesian nonparametric approach to the analysis of two different types of non-homogeneous mixed Poisson processes. The unknown mean function is modelled *a priori* as a process with independent increments and the corresponding posteriors are derived. Posterior inferences are carried out via a Gibbs sampling scheme.

Keywords: BAYESIAN NONPARAMETRICS, EXTENDED GAMMA PROCESS, LÉVY PROCESS, LOG-BETA PROCESS, NEGATIVE-BINOMIAL PROCESS.

1. INTRODUCTION

In the theory of stochastic processes there are two that are fundamental. One is the Weiner model of Brownian motion; the other is the Poisson process. Both are important instances of exponential families of stochastic processes (Küchler and Sorensen, 1997). The Poisson process has several interesting properties and characterisations. For one thing, it is the only point process with stationary and independent increments. In addition, Poisson processes are ordinary renewal processes with exponentially distributed inter-occurrences times; they are the only ordinary renewal processes that are also stationary. For a thorough account of the structure and properties of general Poisson processes, the reader is referred to Kingman (1993).

Despite its nice properties, the Poisson process is often too simple to be useful in applications, and so several generalisations of it have arisen. Non-homogeneous Poisson processes (NHPP), for instance, are widely used in reliability (see Kuo & Yang, 1996; Kuo & Ghosh, 1997; and the references therein). On the other hand, the mixed Poisson process (MPP) has played a prominent role in areas such as actuarial science (Grandell, 1991, 1997; and Albrecht, 1982) and flood frequency analysis in hydrology (*e.g.*, Lang, 1999). Non-homogeneous mixed Poisson processes are a natural generalisation thereof (Grandell, 1997).

A further generalisation, introduced by Cox (1955) and usually referred to as doubly stochastic Poisson processes, has also been extensively used in risk theory (Grandell, 1976, 1991). Such processes (in two dimensional spaces) are widely used as models for point patterns which are thought to reflect underlying environmental heterogeneity (Wolpert & Ickstadt, 1998a; Brix & Diggle, 2001).

Lo (1982) discusses Bayesian nonparametric inference for Poisson point processes based on i.i.d. multiple samples. Kuo & Yang (1996) and Kuo & Ghosh (1997) address Bayesian parametric and nonparametric inference, respectively, for non-homogeneous Poisson processes. In the latter paper, either the intensity or the cumulative intensity of the NHPP is assigned one of a number of alternative independent increment processes.

This paper is concerned with Bayesian nonparametric inference for non-homogeneous mixed Poisson processes (NHMPP) of two different types, each defined by a specific mixing distribution or process (see Sections 2.2 and 2.3). For both types of NHMPP, we assign a general Lévy process prior to simple transformations of the mean function and derive the corresponding posteriors. Such priors exhibit a sort of partial conjugacy which makes the analysis relatively simple. Posterior analysis is carried out via a Gibbs sampling scheme. Each type of NHMPP gives rise to one of two different types of negative binomial processes when the mixing distribution (process) is gamma. These are perhaps the most important instances of NHMPP and we will focus on them throughout the paper.

In the next section, we briefly review some basic facts concerning Poisson, mixed Poisson, Cox and Lévy Processes. Section 3 deals with Bayesian nonparametric inference for the class of NHPP (cf. Kuo & Ghosh, 1997). The results of the latter section are then used in Section 4 in order to analyse the NHMPP of type 1 and type 2 via the Gibbs sampler. Section 5 describes in some detail how the required posterior simulations can be carried out, and the methods are illustrated using a data set of 31 failure times previously analysed in the literature. Finally, Section 6 contains some concluding remarks.

2. PRELIMINARIES

In this section we review basic aspects of Poisson, mixed Poisson and Cox processes, and introduce some notation. A more detailed account can be found in Grandell (1997). We also briefly review the theory of Lévy processes. See Gikhman & Skorokhod (1965) and Sato (1999) for details.

2.1. *Poisson Processes*

A point process $\tilde{\mathcal{N}}(\cdot)$ is called a Poisson process with intensity λ, denoted $\tilde{\mathcal{N}}(t) \sim \text{Po}(\lambda t)$, if

(i) $\tilde{\mathcal{N}}(\cdot)$ has independent increments; and

(ii) $[\tilde{\mathcal{N}}(t) - \tilde{\mathcal{N}}(s)] \sim \text{Po}\{\lambda(t-s)\}$ for all $0 < s < t$.

A Poisson process is homogeneous, *i.e.*, $[\tilde{\mathcal{N}}(t+s) - \tilde{\mathcal{N}}(s)]$ has the same distribution for all $s \geq 0$. Moreover, Poisson processes are the only point processes with stationary and independent increments.

In the sequel, $\mathcal{A} = \{A(t) : t \geq 0\}$ will denote the set of all functions $A(\cdot)$ such that (i) $A(0) = 0$; (ii) $A(t) < \infty$ for all $0 \leq t < \infty$; and (iii) $A(\cdot)$ is non-decreasing and right continuous. Let $\Lambda(\cdot) \in \mathcal{A}$. A point process $\tilde{\mathcal{N}}(\cdot)$ is called a non-homogeneous Poisson process with mean function (intensity measure, cumulative intensity) $\Lambda(\cdot)$ if

(i) $\tilde{\mathcal{N}}(\cdot)$ has independent increments; and

(ii) $[\tilde{\mathcal{N}}(t) - \tilde{\mathcal{N}}(s)] \sim \text{Po}\{\Lambda(t) - \Lambda(s)\}$ for all $0 < s < t$.

In this case the increments are no longer stationary unless $\Lambda(t)$ is linear in t.

2.2. *Mixed Poisson Processes of Type 1*

Let θ be a positive random variable and let $\mathcal{N}(\cdot)$ be a Poisson process with intensity θ. Then the point process $\mathcal{N}(\cdot)$ is called a (homogeneous) mixed Poisson process. This process has stationary increments, but they will not be independent unless the distribution of θ is concentrated at one point.

Now, for a fixed function $\Lambda(\cdot) \in \mathcal{A}$, let θ be a positive random variable and let $\mathcal{N}(\cdot)$ be a NHPP with mean function $\mathcal{M}_\theta(\cdot) = \theta\Lambda(\cdot)$. Then the point process $\mathcal{N}(\cdot)$ is termed a non-homogeneous mixed Poisson process of type 1 (NHMPP1). Such processes have the following intuitive interpretation: first, a realisation of a positive random variable θ is generated; conditional on θ, $\mathcal{N}(\cdot)$ is a Poisson process with mean function $\mathcal{M}_\theta(\cdot)$. As in the case of the NHPP, the increments of this process are not stationary unless $\Lambda(t)$ is linear in t. In addition, the increments of this process are not independent unless the distribution of θ is concentrated at one point.

Suppose, in particular, that $\theta \sim \text{Ga}(a, b)$. Then $\mathcal{N}(\cdot) \sim \text{Nb}_1\{a, p(\cdot)\}$ with $p(t) = \Lambda(t)/\{b + \Lambda(t)\}$ for all $t \geq 0$, *i.e.*, $\mathcal{N}(\cdot)$ is a negative binomial process of type 1. Let $\mu = \text{E}(\theta) = a/b$. Then $\text{E}\{\mathcal{N}(\cdot)\} = \mu\Lambda(\cdot)$ and $\text{Var}\{\mathcal{N}(\cdot)\} = \mu\Lambda(\cdot)\{1 + \Lambda(\cdot)/b\}$. Thus, $\mathcal{N}(\cdot)$ is overdispersed relative to the NHPP in the sense that $\text{E}\{\mathcal{N}(t)\} \leq \text{Var}\{\mathcal{N}(t)\}$ for all $t \geq 0$. For fixed μ, the parameter b controls both the degree of overdispersion and the strength of the dependence between the increments of $\mathcal{N}(\cdot)$.

2.3. *Mixed Poisson Processes of Type 2*

Let $\mathcal{N}(\cdot)$ be a NHPP with mean function $\Lambda(\cdot)$. When $\Lambda(\cdot)$ is assumed to be a random measure, we call $\mathcal{N}(\cdot)$ a non-homogeneous mixed Poisson process of type 2 (NHMPP2). Elsewhere in the literature, such processes are known as doubly stochastic Poisson Processes or Cox processes. We assume that the realisations of $\Lambda(\cdot)$ belong to the set $\mathcal{A}$ defined in Section 2.1 in order for $\mathcal{N}(\cdot)$ to be well defined. Similarly to the previous case, a NHMPP2 has the following interpretation: first, a realisation of the stochastic process $\Lambda(\cdot)$ is generated; conditional on $\Lambda(\cdot)$, $\mathcal{N}(\cdot)$ is a Poisson process with mean function $\mathcal{M}_\Lambda(\cdot) = \Lambda(\cdot)$. If $\Lambda(\cdot)$ is assumed to have independent increments, then $\mathcal{N}(\cdot)$ will also have independent (but not stationary) increments.

Suppose, for example, that $\Lambda(\cdot) \sim \text{GaP}\{a(\cdot), b\}$, *i.e.*, $\Lambda(\cdot)$ is a gamma process (*e.g.*, Ferguson, 1973; van der Weide, 1997). Then $\mathcal{N}(\cdot) \sim \text{Nb}_2\{a(\cdot), 1/(b + 1)\}$. In other words, $\mathcal{N}(\cdot)$ is a negative binomial process of type 2. In this case, $\text{E}\{\mathcal{N}(\cdot)\} = a(\cdot)/b$ and $\text{Var}\{\mathcal{N}(\cdot)\} = \{a(\cdot)/b\}\{1 + (1/b)\}$, and so $\mathcal{N}(\cdot)$ is also overdispersed relative to the NHPP. As in the case of the NHMPP1, b controls the degree of overdispersion.

As pointed out by Grandell (1997, Ch. 5), Cox processes, and not mixed Poisson processes, are actually the process version of the mixed Poisson distribution. Homogeneous mixed Poisson processes can be seen as Cox processes with underlying random measure $\Lambda(dt) = \theta dt$ where θ is a positive random variable. This is why we refer to Cox processes as mixed Poisson processes (of type two, so as to distinguish them from the usual MPP described in the previous section).

2.4. *Lévy Processes*

A continuous time process $L(\cdot)$ with independent stationary increments is called a *Lévy process* if its sample paths are right continuous with limits from the left and $L(0) = 0$. If the stationarity requirement is dropped, then $L(\cdot)$ is usually called an additive process (*e.g.*, Sato, 1999). Recently, however, especially in the Bayesian literature, the term 'Lévy process' has been used to refer to the more general additive process. Throughout this paper we will use the latter terminology.

According to Gikhman & Skorokhod (1965), every Lévy process $L(\cdot)$ can be written as the sum of a fixed jump component and a 'continuous' component, *i.e.*, if $t_1, t_2, \ldots$ correspond to fixed points of discontinuity having independent non-negative jumps $L\{t_1\}, L\{t_2\}, \ldots$ (also independent of the rest of the process), then

$$L(t) = L_c(t) + \sum_j L\{t_j\} I(t_j \le t),$$

where $L_c(\cdot)$ is a non-decreasing process with no fixed points of discontinuity and therefore has a Lévy representation for the Laplace transform

$$\mathrm{E}\left\{e^{-\psi L_c(t)}\right\} = \exp\left\{-\int_0^\infty \left(1 - e^{-\psi\nu}\right) dN_t(\nu)\right\}, \tag{1}$$

with $N_t(\nu)$ a Lévy measure satisfying:

1. For every Borel set B, $N_t(B)$ is continuous and nondecreasing as a function of t;
2. For every real $t > 0$, $N_t(\cdot)$ is a measure on the Borel sets of $(0, \infty)$;
3. $\int_0^1 \nu dN_t(\nu) < \infty$; and
4. $\int_1^\infty dN_t(\nu) < \infty$.

3. NON-HOMOGENEOUS POISSON PROCESS

In this section we present a detailed Bayesian analysis of NHPP, which will be used in Section 4 in order to study NHMPP.

3.1. *Likelihood and Prior*

Following Kuo & Ghosh (1997), let us consider a time-truncated model where the process is observed up to a fixed time τ. We denote the ordered epochs of the n observed jumps by $0 = x_0 < x_1 < x_2 < \cdots < x_n \le \tau$. If we allow ties in the observed jump times, then the probability of observing no jumps in the interval $(0, x_1)$, d_1 jumps at x_1, no jumps in (x_1, x_2), and so on up to no jumps in (x_n, τ), is given by

$$\mathrm{Lik}(\Lambda \mid D, \tau) = \left\{\prod_{i=1}^n [\,\Lambda(x_i) - \Lambda(x_i^-)\,]^{d_i}\, e^{-[\,\Lambda(x_i) - \Lambda(x_i^-)\,]}\, e^{-[\,\Lambda(x_i^-) - \Lambda(x_{i-1})\,]}\right\} e^{-[\,\Lambda(\tau) - \Lambda(x_n)\,]}.$$

Here, *D* denotes the data and d_i is the number of multiple jumps observed at time x_i. The total number of observed jumps in the interval $(0, \tau]$ is then $d = \sum_{i=1}^n d_i$.

We will assign a Lévy process prior to $\Lambda(\cdot)$. As pointed out in Section 2.4, the prior can then be characterised by

$$M = \{t_1, t_2, \ldots\}, \quad \{f_{t_1}, f_{t_2}, \ldots\},$$

the set of fixed points of discontinuity together with the corresponding density functions for the jumps, and $N_t(\cdot)$, the Lévy measure for the part of the process without fixed points of discontinuity. We assume the Lévy measure to be of the form

$$dN_t(\nu) = d\nu \int_0^t K(\nu, u) du. \tag{2}$$

Typically, $K(\cdot, \cdot)$ is parametrised in terms of a non-negative measure $\alpha(\cdot)$ and a non-negative, piece-wise continuous function $\beta(\cdot)$, *i.e.*,

$$\Lambda(\cdot) \sim \text{LévyProcess}\{\alpha(\cdot), \beta(\cdot)\}.$$

For instance, if

$$K(\nu,u)du = \nu^{-1}\exp\{-\nu\beta(u)\}d\alpha(u) \tag{3}$$

then $\Lambda(\cdot)$ is an extended gamma process (see Dykstra & Laud, 1981).

On the other hand, if

$$K(\nu,u)du = \{1-\exp(-\nu)\}^{-1}\exp\{-\nu\beta(u)\}d\alpha(u) \tag{4}$$

then $\Lambda(\cdot)$ is a log-beta process (see Walker & Muliere, 1997).

3.2. *Posterior Distributions*

Experience with updating Lévy processes for modeling cumulative hazard functions (see, for example, Walker & Muliere, 1997) and updating Lévy-driven processes for modeling hazard rate functions (see Nieto-Barajas, 2001) gives rise to the following posterior distributions.

Theorem 1. *Let $\Lambda(\cdot)$ be a Lévy process and let $x_1,\ldots,x_n$ be the ordered epochs of the observed jump times from a NHPP $\mathcal{N}(\cdot)$. Let $x_1^*,\ldots,x_m^*$ be the time epochs not equal to any of the prior fixed jumps. Then the posterior distribution of $\Lambda(\cdot)$ given the data is again a Lévy process with the following characteristics (we will use an (*) to denote an updated parameter/function):*

$M^* = M\cup\{x_1^*,\ldots x_m^*\}$, *with*

$$f_j^*(\nu\,|\,D) \propto \begin{cases} \nu^{d_i}e^{-\nu}f_j(\nu) & \text{if } t_j = x_i, \\ e^{-\nu}f_j(\nu) & \text{if } t_j \neq x_i, \end{cases}$$

$$f_{x_i^*}(\nu\,|\,D) \propto \nu^{d_i}e^{-\nu}K(\nu,x_i),$$

for $i=1,\ldots,m$ and

$$K^*(\nu,u) = e^{-\nu}K(\nu,u).$$

Proof. Posterior distribution of the jumps at the prior fixed points of discontinuity are obtained by looking at the likelihood and using standard Bayesian updating. We will concentrate in obtaining the posterior distribution of the continuous part of the process for a single time epoch x_1. The idea is to characterise the posterior distribution by its Laplace transform. If we define

$$\phi(\psi,\Lambda,x_1,\epsilon) = \mathrm{E}\left[\exp\left\{-\int_0^\tau \psi(s)d\Lambda(s)\right\}\,\middle|\,\mathcal{N}(x_1-\epsilon,x_1]=1\right],$$

then our aim is to find

$$\lim_{\epsilon\to 0}\phi(\psi,\Lambda,x_1,\epsilon).$$

Thus, $\phi(\psi,\Lambda,x_1,\epsilon)$ can be expressed as

$$\frac{\mathrm{E}\left[\int_{x_1-\epsilon}^{x_1} d\Lambda(s)\exp\left\{-\int_0^\tau\{\psi(s)+1\}d\Lambda(s)\right\}\right]}{\mathrm{E}\left[\int_{x_1-\epsilon}^{x_1} d\Lambda(s)\exp\left\{-\int_0^\tau d\Lambda(s)\right\}\right]}.$$

Splitting up the integral and using independence properties this becomes

$$H(x_1-\epsilon,x_1)\frac{\mathrm{E}\left[\exp\left\{-\int_{(0,x_1-\epsilon]\cup(x_1,\infty)}\{\psi(s)+1\}d\Lambda(s)\right\}\right]}{\mathrm{E}\left[\exp\left\{-\int_{(0,x_1-\epsilon]\cup(x_1,\infty)}d\Lambda(s)\right\}\right]},$$

where

$$H(x_1-\epsilon, x_1) = \frac{\mathrm{E}\left[\int_{x_1-\epsilon}^{x_1} d\Lambda(s)\exp\left\{-\int_0^\tau\{\psi(s)+1\}d\Lambda(s)\right\}\right]}{\mathrm{E}\left[\int_{x_1-\epsilon}^{x_1} d\Lambda(s)\exp\left\{-\int_0^\tau d\Lambda(s)\right\}\right]}$$

Taking limits, we can express $\lim_{\epsilon\to 0}\phi(\psi,\Lambda,x_1,\epsilon)$ as

$$H(x_1)\exp\left[-\int_0^\tau\int_0^\infty\left\{1-e^{-\psi(s)\nu}\right\}e^{-\nu}K(\nu,s)d\nu ds\right],$$

with

$$H(x_1) = \frac{\int_0^\infty e^{-\psi(x_1)\nu}\nu e^{-\nu}K(\nu,x_1)d\nu}{\int_0^\infty \nu e^{-\nu}K(\nu,x_1)d\nu},$$

which can be written as

$$\mathrm{E}\left\{e^{-\psi(x_1)\Lambda^*\{x_1\}}\right\}\mathrm{E}\left[\exp\left\{-\int_0^\tau\psi(s)d\Lambda^*(s)\right\}\right],$$

with $\Lambda^*\{x_1\}$ and $\Lambda_c^*(t)$ as stated in the theorem. The general case for $x_1,\ldots,x_n$ time epochs can be obtained by a straightforward extension of this procedure. This completes the proof. □

In particular, if $\Lambda(\cdot)\sim\mathrm{EGaP}_c\{\alpha(\cdot),\beta(\cdot)\}$, *i.e.*, $\Lambda(\cdot)$ is an extended gamma process without prior fixed points of discontinuity, then the posterior distribution of $\Lambda(\cdot)$ is again an extended gamma process with parameters $\alpha^*(\cdot)=\alpha(\cdot)$ and $\beta^*(\cdot)=\beta(\cdot)+1$, and with fixed points of discontinuity at $M^*=\{x_1,\ldots,x_n\}$ with posterior distributions for the jumps $f_{x_i}^*(\nu)=\mathrm{Ga}(\nu\mid d_i,\beta(x_i)+1)$, $i=1,\ldots,n$.

Moreover, if $\Lambda(\cdot)\sim\mathrm{logBeP}_c\{\alpha(\cdot),\beta(\cdot)\}$, that is, if $\Lambda(\cdot)$ is a log-beta process without prior fixed points of discontinuity, then the posterior distribution of $\Lambda(\cdot)$ is again a Lévy process with a log-beta measure for the continuous part, with parameters $\alpha^*(\cdot)=\alpha(\cdot)$ and $\beta^*(\cdot)=\beta(\cdot)+1$, and with fixed points of discontinuity at $M^*=\{x_1,\ldots,x_n\}$ with posterior distributions for the jumps $f_{x_i}^*(\nu)\propto\nu^{d_i}e^{-\nu}K(\nu,x_i)$, $i=1,\ldots,n$, where $K(\nu,u)$ is given by (4).

In both cases, $\alpha(\cdot)$ is a non-negative measure and $\beta(\cdot)$ is a non-negative, piece-wise continuous function.

It must be pointed out that Kuo & Ghosh (1997) derive the posterior distributions for several particular priors, both on the cumulative intensity $\Lambda(t)$ and on the intensity function $\lambda(t)=d\Lambda(t)/dt$. Specifically, these authors assume gamma and beta (Hjort, 1990) process priors for $\Lambda(\cdot)$ and an extended gamma process prior on $\lambda(\cdot)$.

4. NON-HOMOGENEOUS MIXED POISSON PROCESSES

4.1. *Mixed Poisson Processes of Type 1*

Let $\mathcal{N}(\cdot)$ be a NHMPP1 such that $[\mathcal{N}(\cdot)\mid\theta]\sim\mathrm{Po}\{\theta\Lambda(\cdot)\}$ and $\theta\sim f(\theta)$ for some distribution f. This implies that $\mathcal{N}(\cdot)$ has mean function $\mathcal{M}(\cdot)=\mu_\theta\,\Lambda(\cdot)$, where $\mu_\theta=\mathrm{E}_f[\theta]$. We can make Bayesian inferences about $\mathcal{M}(\cdot)$ by assigning a nonparametric prior to $\Lambda(\cdot)$. Because $\Lambda(\cdot)$ is a non-decreasing function, a reasonable choice is to use a Lévy process prior.

Posterior distributions. The corresponding posterior distributions are given by the next result.

Theorem 2. *Let $\Lambda(\cdot)$ be a Lévy process and let $x_1, \ldots, x_n$ be the ordered epochs of the observed jump times from a NHMPP1 $\mathcal{N}(\cdot)$. Let $x_1^*, \ldots, x_m^*$ be time epochs not equal to any of the prior fixed jumps. Then the posterior conditional distribution of $\Lambda(\cdot)$ given θ is again a Lévy process with the following characteristics (we will use an (*) to denote an updated parameter/function):*

$M^* = M \cup \{x_1^*, \ldots x_m^*\}$ *with*

$$f_j^*(\nu \mid D, \theta) \propto \begin{cases} \nu^{d_i} e^{-\theta\nu} f_j(\nu) & \text{if } t_j = x_i, \\ e^{-\theta\nu} f_j(\nu) & \text{if } t_j \neq x_i, \end{cases}$$

$$f_{x_i^*}(\nu \mid D, \theta) \propto \nu^{d_i} e^{-\theta\nu} K(\nu, x_i),$$

for $i = 1, \ldots, m$ and

$$K^*(\nu, u) = e^{-\theta\nu} K(\nu, u).$$

Furthermore, the posterior conditional distribution of θ given $\Lambda(\cdot)$ is given by

$$f^*(\theta \mid D, \Lambda) \propto \theta^{\sum_{i=1}^n d_i} e^{-\theta\Lambda(\tau)} f(\theta).$$

Proof. The posterior conditional distribution of $\Lambda(\cdot)$ comes from conditioning on θ and extending the proof given to Theorem 1. The posterior conditional distribution of θ is obtained by using standard Bayesian updating. □

In particular, if $\Lambda(\cdot) \sim \text{EGaP}_c\{\alpha(\cdot), \beta(\cdot)\}$, *i.e.*, if $\Lambda(\cdot)$ is an extended gamma process without prior fixed points of discontinuity, and if $\theta \sim \text{Ga}(a, b)$, then the conditional posterior distribution of $\Lambda(\cdot)$ given θ is again an extended gamma process with parameters $\alpha^*(\cdot) = \alpha(\cdot)$ and $\beta^*(\cdot) = \beta(\cdot) + \theta$. It has fixed points of discontinuity at $M^* = \{x_1, \ldots, x_n\}$ with posterior distributions for the jumps $f_{x_i}^*(\nu) = \text{Ga}(\nu \mid d_i, \beta(x_i) + \theta)$, $i = 1, \ldots, n$. Also, the posterior conditional of θ given $\Lambda(\cdot)$ is $\text{Ga}(a + d, b + \Lambda(\tau))$.

If, on the other hand, $\Lambda(\cdot) \sim \text{logBeP}_c\{\alpha(\cdot), \beta(\cdot)\}$, then the posterior conditional of $\Lambda(\cdot)$ given θ is again a Lévy process with a log-beta measure for the continuous part and a D-distribution (Walker, 1995) for the jumps. Also, provided that $\theta \sim \text{Ga}(a, b)$, the posterior conditional of θ given $\Lambda(\cdot)$ is $\text{Ga}(a + d, b + \Lambda(\tau))$.

Prior elicitation. An important issue when carrying out a Bayesian analysis is how to determine the hyper-parameters of the prior process. Recall that our interest lies on the mean function $\mathcal{M}(\cdot) = \mu_\theta \Lambda(\cdot)$. A simple idea is to 'centre' the process $\Lambda(\cdot)$ in such a way so that the first two moments of $\mathcal{M}(\cdot)$ equal two arbitrary functions $\mathcal{M}_0(\cdot)$ and $\Psi_0(\cdot)$, respectively (see, for example, Walker and Damien, 1998). By way of illustration, let us consider the case when $\Lambda(\cdot)$ is an extended gamma process.

Lemma 1. *Let $\Lambda(\cdot)$ be an extended gamma process without prior fixed points of discontinuity, i.e., $\Lambda(\cdot)$ has a representation of the form (1) and Lévy measure given by (2) and (3). Let $\mathcal{M}_0(\cdot)$ and $\Psi_0(\cdot)$ be non-negative and differentiable functions on $(0, \infty)$. Then, there exist functions $\alpha(\cdot)$ and $\beta(\cdot)$ such that*

$$\mathcal{M}_0(t) = \text{E}\{\mathcal{M}(t)\} \quad \textit{and} \quad \Psi_0(t) = \text{Var}\{\mathcal{M}(t)\},$$

for all $t \geq 0$.

Proof. We require

$$\mathcal{M}_0(t) = \mathrm{E}\{\mathcal{M}(t)\} = \mu_\theta \int_0^t \frac{d\alpha(u)}{\beta(u)}$$

and

$$\Psi_0(t) = \mathrm{Var}\{\mathcal{M}(t)\} = \mu_\theta^2 \int_0^t \frac{d\alpha(u)}{\beta^2(u)}.$$

Differentiating both equations with respect to t and solving the simultaneous equations we get

$$\beta(t) = \mu_\theta \left\{ \frac{d\mathcal{M}_0(t)}{d\Psi_0(t)} \right\}$$

and

$$d\alpha(t) = \frac{\{d\mathcal{M}_0(t)\}^2}{d\Psi_0(t)}.$$

This completes the proof. □

Following Walker and Damien (1998), we can centre the extended gamma process at a Bayesian parametric model. For example, if we use the Weibull model $\mathcal{M}(t) = \gamma t^\delta$ with 'prior' distribution $\gamma \sim \mathrm{Ga}(p, q)$, and if $\theta \sim \mathrm{Ga}(a, b)$, from Lemma 1 we obtain $\beta(t) = (aq/2b)t^{-\delta}$ and $d\alpha(t) = \delta(p/2)t^{-1}dt$.

4.2. *Mixed Poisson Processes of Type 2*

Suppose that $\mathcal{N}(\cdot)$ is a NHMPP2 such that

$$[\,\mathcal{N}(\cdot) \mid \Lambda(\cdot)\,] \sim \mathrm{Po}\{\Lambda(\cdot)\}$$

and

$$\Lambda(\cdot) \sim \text{Lévy Process}\{a(\cdot), b(\cdot)\},$$

where $b(\cdot)$ is a fixed known function. For the sake of simplicity, we consider only Lévy processes without prior fixed points of discontinuity as distributions for $\Lambda(\cdot)$.

As in the previous case, we want to make inference about the mean function $\mathcal{M}(\cdot)$, which depends on $a(\cdot)$. In a Bayesian approach we can assign a Lévy process prior to $a(\cdot)$.

Posterior distributions. The following proposition provides the posterior distribution for the NHMPP2 model.

Proposition 1. *Let $a(\cdot)$ be a Lévy process such that $a(ds) \sim f\{a(ds)\}$, and let $x_1, \ldots, x_n$ be the ordered epochs of the observed jump times from a NHMPP2 $\mathcal{N}(\cdot)$. Then the posterior conditional distribution of $a(ds)$ given the data and $\Lambda(ds)$ can be obtained as (we will use an (*) to denote an updated parameter/function):*

$$f^*\{a(ds) \mid D, \Lambda(ds)\} \propto f\{\Lambda(ds) \mid a(ds), b(s)\} f\{a(ds)\}.$$

Furthermore, the posterior conditional distribution of $\Lambda(\cdot)$ given the data and $a(\cdot)$ is again a Lévy process with the following characteristics:

$M^* = \{x_1, \ldots x_n\}$ *with*

$$f^*_{x_i}(\nu \mid D, a) \propto \nu^{d_i} e^{-\nu} K(\nu, x_i),$$

for $i = 1, \ldots, n$ and

$$K^*(\nu, u) = e^{-\nu} K(\nu, u).$$

Proof. Straightforward, given the independence of the increments. □

If, in particular, we take

$$\Lambda(\cdot) \sim \text{EGaP}\{a(\cdot), b(\cdot)\} \tag{5}$$

and an extended gamma process prior for $a(\cdot)$, *i.e.*,

$$a(\cdot) \sim \text{EGaP}\{\alpha(\cdot), \beta(\cdot)\}, \tag{6}$$

then the mean function of $\mathcal{N}(\cdot)$ becomes

$$\mathcal{M}(t) = \int_0^t \frac{da(u)}{b(u)}.$$

It follows from Proposition 1 that the conditional posterior distributions are given by

$$[\,\Lambda(ds)\,|\,D, a\,] \sim \text{Ga}\left\{a(ds) + \sum_i d_i I(x_i \in ds), b(s) + 1\right\}$$

and

$$[\,a(ds)\,|\,D, \Lambda\,] \propto \frac{\{b(s)\Lambda(ds)\}^{a(ds)}}{\Gamma\{a(ds)\}}\, a(ds)^{\alpha(ds)-1} e^{-\beta(s)a(ds)} I\{a(ds) > 0\}.$$

Prior elicitation. Consider the special case where $\Lambda(\cdot)$ and $a(\cdot)$ are both extended gamma processes. We can centre the prior process $a(\cdot)$ in such a way that the prior expected value and the prior variance of the mean function $\mathcal{M}(\cdot)$ be equal to two given functions $\mathcal{M}_0(\cdot)$ and $\Psi_0(\cdot)$, respectively.

Lemma 2. *Let $a(\cdot)$ be an extended gamma process without prior fixed points of discontinuity, as specified in Lemma 1. Let $\mathcal{M}_0(\cdot)$ and $\Psi_0(\cdot)$ be non-negative and differentiable functions defined on $(0, \infty)$. Then, there exist functions $\alpha(\cdot)$ and $\beta(\cdot)$ such that*

$$\mathcal{M}_0(t) = \text{E}\{\mathcal{M}(t)\} \;\; \textit{and} \;\; \Psi_0(t) = \text{Var}\{\mathcal{M}(t)\},$$

for all $t \geq 0$.

Proof. We require

$$\mathcal{M}_0(t) = \text{E}\{\mathcal{M}(t)\} = \int_0^t \frac{d\alpha(u)}{b(u)\beta(u)}$$

and

$$\Psi_0(t) = \text{Var}\{\mathcal{M}(t)\} = \int_0^t \frac{d\alpha(u)}{b^2(u)\,\beta^2(u)}.$$

Differentiating both equations with respect to t and solving the simultaneous equations we get

$$\beta(t) = \frac{d\mathcal{M}_0(t)}{b(t)\, d\Psi_0(t)}$$

and

$$d\alpha(t) = \frac{\{d\mathcal{M}_0(t)\}^2}{d\Psi_0(t)}.$$

This completes the proof. □

Again, if we use a Weibull model $\mathcal{M}(t) = \gamma t^\delta$ with 'prior' distribution $\gamma \sim \text{Ga}(p, q)$ to centre our prior process, then, using Lemma 2, we obtain $\beta(t) = [q/\{2b(t)\}]\, t^{-\delta}$ and $d\alpha(t) = (\delta p/2)\, t^{-1} dt$.

5. POSTERIOR SIMULATIONS

In this section we describe in some detail how the required posterior simulations can be carried out. The methods described in the last section are illustrated using data previously analysed by Kuo & Ghosh (1997) and other authors. The data set consists of the following 31 failure epochs in days: 9, 21, 32, 36, 43, 45, 50, 58, 63, 70, 71, 77, 78, 87, 91, 92, 95, 98, 104, 105, 116, 149, 156, 247, 249, 250, 337, 384, 396, 405 and 540, and relates to trouble reports for one module of the US Naval Tactical Data System.

5.1. *Example 1: NHMPP1*

First we briefly describe a way of simulating a Lévy process as defined in Section 2.4, since this will be required in order to sample from the posterior distribution of $\mathcal{M}(\cdot)$ under the NHMPP1 model. The method is based on a representation of a Lévy process derived by Ferguson & Klass (1972) and further discussed in Walker & Damien (2000). A related algorithm appears in Wolpert & Ickstadt (1998b).

Consider a Lévy process $L(\cdot)$, as defined in Section 2.4, and recall that the process is assumed to be observed up to a fixed time τ. Now let

$$M(x) = -N_\tau[x,\infty) = -\int_x^\infty dN_\tau(\nu),$$

where $N_t(\nu)$ denotes the Lévy measure of the process $L_c(t)$.

Define positive random variables $J_1 \geq J_2 \geq \cdots$ by

$$\text{pr}(J_1 \leq x_1) = \exp\{M(x_1)\}$$

and

$$\text{pr}(J_i \leq x_i \mid J_{i-1} \leq x_{i-1}) = \exp\{M(x_i) - M(x_{i-1})\}$$

for $x_i < x_{i-1}$.

We can obtain the J_i via $\omega_i = -M(J_i)$, where $\omega_1, \omega_2, \omega_3, \ldots$ are the jump times of a standard Poisson process (*i.e.*, with intensity 1). In other words, $\omega_1, \omega_2-\omega_1, \omega_3-\omega_2, \ldots \overset{iid}{\sim} \text{Ga}(1,1)$. The Ferguson-Klass representation of the process $L_c(t)$ is then given by

$$L_c(t) = \sum_i J_i\, I\{U_i \leq n_t(J_i)\},$$

where $U_1, U_2, \ldots \overset{iid}{\sim} \text{Un}(0,1)$, and $n_t(\nu) = \frac{dN_t}{dN_\tau}(\nu)$, for $t \in [0,\tau]$.

Posterior simulation of $\mathcal{M}(\cdot)$ can be carried out by means of a Gibbs sampling scheme (*e.g.*, Smith & Roberts, 1993) using the full conditionals given in Theorem 2. In order to sample from the full conditional of $\Lambda(\cdot)$, we use the algorithm describe above based on the Ferguson-Klass representation of a Lévy process.

We now illustrate this using the data set described above. We take $\theta \sim \text{Ga}(5,1)$, so that $\mathcal{N}(\cdot) \sim \text{Nb}_1(5, p(\cdot))$ with $p(t) = \Lambda(t)/\{1+\Lambda(t)\}$. Also, we give $\Lambda(\cdot)$ an extended gamma process prior EGaP$\{\alpha(\cdot), \beta(\cdot)\}$. Hence the corresponding full conditionals are those described in Section 4.1. We centred the prior on a Weibull model, as discussed in Section 4.1, and took $p = q = \delta = 1$. We ran the Gibbs sampler for 10000 iterations with a burn-in period of 1000 iterations. Convergence was assessed informally based on plots of ergodic averages for θ and $\Lambda(t)$ for a few selected values of t. Figure 1 shows the data together with the posterior mean (continuous line), the median (dotted line), and a 95% credible band (dashed lines) for the mean function $\mathcal{M}(\cdot)$. The straight line appearing on the left of the plot represents the function at which we centred the prior of $\mathcal{M}(\cdot)$, namely $\mathcal{M}_0(t) = t$.

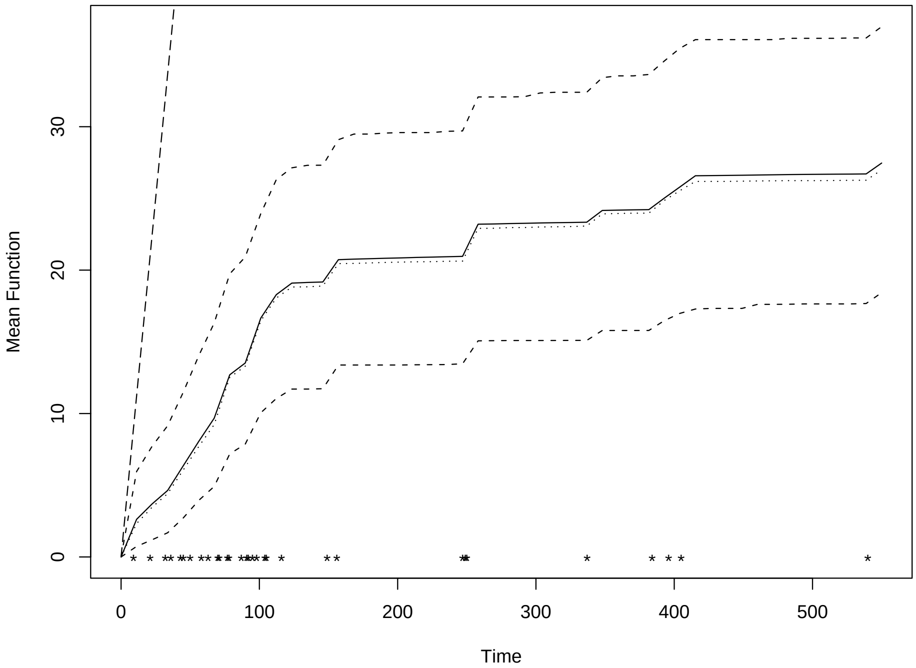

Figure 1. *Posterior mean, median, and 95% band for $\mathcal{M}(\cdot)$ (NHMPP1; extended gamma prior)*

5.2 *Example 2: NHMPP2*

In this example we will illustrate the analysis of a NHMPP2 using the same data as in Example 1. For the mixing distribution we used a gamma process with parameters $a(\cdot)$ and b, as in (5) with $b(t) \equiv b$ for all t, and an extended gamma process prior for $a(\cdot)$ as in (6). Thus, $\mathcal{N}(\cdot)$ is a negative binomial process of type 2. We centred $a(\cdot)$ at a Weibull model as described in Section 4.2, and took $\delta = 1$, $p = 0.5$, $q = 0.5$ and $b = 100$. One way of approximating the posterior process $a^*(\cdot)$ is to take a fine partition, say $t_0, t_1, \ldots, t_R$ for R sufficiently large, and implement a Gibbs sampler.

Let $\Lambda_r = \Lambda(t_{r-1}, t_r]$, $a_r = a(t_{r-1}, t_r]$, $\alpha_r = \alpha(t_{r-1}, t_r]$ and $\beta_r = \beta(t_r)$, for $r = 1, \ldots, R$. Then the posterior conditional distributions of Λ_r and a_r are,

$$[\,\Lambda_r \,|\, \boldsymbol{D}, a_r\,] \sim \text{Ga}\left\{a_r + \sum_i d_i\, I(t_{r-1} < x_i \le t_r), b+1\right\}$$

and

$$[\,a_r \,|\, \boldsymbol{D}, \Lambda_r\,] \propto \frac{(b\Lambda_r)^{a_r}}{\Gamma(a_r)}\; a_r^{\alpha_r - 1} e^{-\beta_r a_r} I(a_r > 0),$$

for $r = 1, \ldots, R$.

We can simulate from the posterior conditional distribution of a_r introducing a Metropolis-Hastings step (see, for example, Tierney, 1994) in the following way: at iteration $(k+1)$, for each $r = 1, \ldots, R$, generate $a_r^\star \sim \text{Ga}(\alpha_r, \beta_r)$; then take $a_r^{(k+1)} = a_r^\star$ with probability $\pi(a_r^\star, a_r^{(k)})$ and $a_r^{(k+1)} = a_r^{(k)}$ with probability $\{1 - \pi(a_r^\star, a_r^{(k)})\}$, where

$$\pi(a_r^\star, a_r^{(k)}) = \frac{\Gamma(a_r^{(k)})}{\Gamma(a_r^\star)}\;(b\Lambda_r)^{(a_r^\star - a_r^{(k)})}.$$

We implemented this algorithm with $t_0 = 0$, $t_r = t_{r-1} + 0.54$ and $R = 1000$. We ran the Markov chain for 10000 iterations with a burn-in period of 1000. Convergence was also assessed visually by plotting ergodic averages of some of the a_r and Λ_r. Figure 2 shows the data together with the posterior mean (continuous line), the median (dotted line), and a 95% credible band (dashed lines) for the mean function $\mathcal{M}(\cdot)$. As in the previous example, the straight line appearing on the left of the graph represents the function at which we centred the prior of $\mathcal{M}(\cdot)$, namely $\mathcal{M}_0(t) = t$.

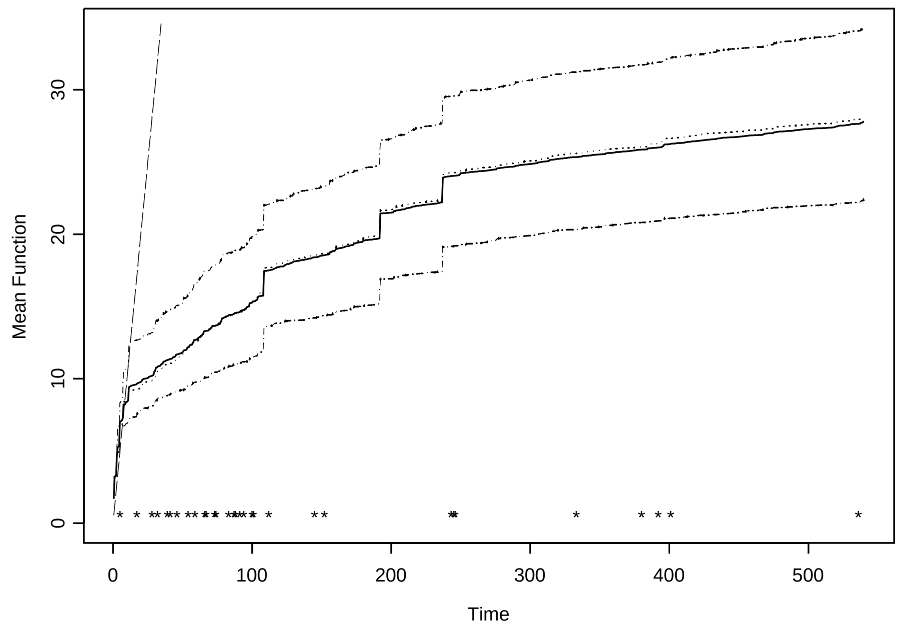

Figure 2. *Posterior mean, median and 95% band for $\mathcal{M}(\cdot)$ (NHMPP2; extended gamma prior)*

6. CONCLUDING REMARKS

In this paper we have presented a relatively simple procedure to make Bayesian inferences for two types of non-homogeneous mixed Poisson processes. No parametric form for the mean functions was assumed; instead, the mean functions were given Lévy process priors 'centred' at suitable parametric models. Due to the nature of the parameter space, this approach can be regarded as nonparametric. For each of the two types of mixed Poisson processes, we showed how one can centre the prior in order to reflect prior knowledge about the mean function.

Posterior inference for the NHMPP1 was achieved by implementing a Gibbs sampling scheme, using an efficient algorithm (Ferguson and Klass, 1972) which allowed us to simulate whole paths from the required Lévy processes (Section 5.1). However, for the NHMPP2, a partitioning simulation scheme proved to be a simpler way to carry out the required posterior simulations (Section 5.2). In both cases, the algorithms were easy to code, although it must be pointed out that in general the algorithm described in Section 5.1 is computationally more expensive. The methods were illustrated with the analysis of two different types of negative binomial processes.

One possible extension to the models discussed in this paper would be to use a Lévy-driven process prior (Nieto-Barajas, 2001; Nieto-Barajas & Walker, 2001) for the intensity function of a NHMPP. Such priors would be especially useful in cases where the mean function (cumulative intensity) is constrained to be continuous, and the required posterior simulations could be carried out without much additional effort with respect to the methods discussed here.

ACKNOWLEDGEMENTS

The first and second authors were supported by grants 32256-E and I39357-E, respectively, from CONACyT, Mexico. The first author wishes also to acknowledge partial support from the Sistema Nacional de Investigadores, Mexico.

REFERENCES

Albrecht, P. (1982). On some statistical methods connected with the mixed Poisson process. *Scand. Actuarial J.* 1–14.

Brix, A. and Diggle, P.J. (2001). Spatiotemporal prediction for log-Gaussian Cox Processes. *J. Roy. Statist. Soc. B* **63**, 823–841.

Cox, D.R. (1955). Some statistical methods connected with series of events (with discussion). *J. Roy. Statist. Soc. B* **17**, 129–164.

Dykstra, R.L. and Laud, P.W. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–367.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

Ferguson, T.S. and Klass, M.J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43**, 1634–1643.

Gikhman, I.I. and Skorokhod, A.V. (1965). *An Introduction to the Theory of Random Processes*. New York: Dover.

Grandell, J. (1976). *Doubly Stochastic Processes*. Lecture Notes in Mathematics 529. New York: Springer.

Grandell, J. (1991). *Aspects of Risk Theory*. New York: Springer.

Grandell, J. (1997). *Mixed Poisson Processes*. London: Chapman and Hall.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18**, 1259–1294.

Kingman, J.F.C. (1993). *Poisson Processes*. Oxford: Oxford University Press

Küchler, U. and Sorensen, M. (1997). *Exponential Families of Stochastic Processes*. New York: Springer.

Kuo. L. and Ghosh, S.K. (1997). Bayesian nonparametric inference for non-homogeneous Poisson processes. *Tech. Rep.*, University of Connecticut, Storrs, USA.

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

Lang, M. (1999). Theoretical discussion and Monte-Carlo simulations for the negative binomial process paradox. *Stoch. Environ. Res. Risk Assess.* **13**, 183–200.

Lo, A.Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Wahrscheinlichkeitstheorie und verwandte Gebiete* **59**, 55-66.

Nieto-Barajas, L.E. (2001). *Bayesian Nonparametric Survival Analysis via Markov Processes*. Ph.D. Thesis, University of Bath, UK.

Nieto-Barajas, L.E. and Walker, S.G. (2001). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Tech. Rep.*, University of Bath, UK.

Sato, (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge Studies in Advanced Mathematics 68. Cambridge: University Press.

Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B* **55**, 3–23.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701–1762. (with discussion).

van der Weide, H. (1997). Gamma processes. In *Engineering Probabilistic design and Maintenance for Flood Protection* (R. Cooke *et al.* , Eds.) Dordrecht: Kluwer, 77–83.

Walker, S.G. (1995). Generating random variates from D-distributions via substitution sampling. *Statist. Computing* **5**, 311-315.

Walker, S.G. and Damien, P. (1998). A full Bayesian nonparametric analysis involving a neutral to the right process. *Scandinavian J. Statist.* **25**, 669–680.

Walker, S.G. and Damien, P. (2000). Representation of Lévy processes without Gaussian components. *Biometrika* **87**, 477–483.

Walker, S.G. and Muliere, P. (1997). Beta-Stacy processes and a generalisation of the Pólya-urn scheme. *Ann. Statist.* **25**, 1762–1780.

Wolpert, R.L. and Ickstadt, K. (1998a). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

Wolpert, R.L. and Ickstadt, K. (1998b). Simulation of Lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics*. (D. Dey, P. Müller and D. Sinha, eds.) New York: Springer, 227–242.

DISCUSSION

LUKE TIERNEY *(University of Minnesota, USA)*

I congratulate the authors on a very interesting paper. I have just a few technical comments and a few general remarks on the models used.

The NHMPP1 prior distribution is a semi-parametric model with a one-dimensional parametric component. One advantage of using a Lévy process for the non-parametric part is that one may be able to integrate out the non-parametric part to obtain a computable expression for the marginal density of the parametric component. This is the approach used by Kalbfleisch (1978) for the proportional hazards model. For an extended Gamma process prior distribution on Λ and assuming no ties in the data this yields

$$f(\theta|D) \propto \theta^n \left[\prod_{i=1}^{n} \frac{\alpha'(x_i)}{\theta + \beta(x_i)}\right] \exp\left\{-\int_0^{\tau} \log\left(1 + \frac{\theta}{\beta(u)}\right) d\alpha(u)\right\} f(\theta)$$

for the marginal posterior density of θ. The one-dimensional integral in this representation can easily be computed numerically. For the software reliability example the marginal posterior density of $\log\theta$ is shown as the solid curve in Figure 3. With a little experimentation one can easily find a scaled t distribution with 10 degrees of freedom that can be used as an envelope for rejection sampling from this marginal distribution; such an envelope is shown as the dashed curve in Figure 3. Thus a strategy for exact sampling of the joint posterior distribution is available: sample θ from its marginal distribution by rejection sampling, and then sample Λ from the extended Gamma process conditional posterior distribution using the Ferguson-Klass method.

In this particular example the envelope was determined empirically; it may be useful to explore methods for automatic determination of the envelope. Similar strategies may be useful in other non-parametric settings with a low dimensional parametric component.

A second technical point concerns the use of approximations in the sampling methods. Some form of approximation seems to be unavoidable when dealing with infinite dimensional parameters. The paper uses different approaches for the two settings considered: In the NHMPP1 model the Ferguson-Klass steps for the non-parametric component are truncated in some form (the paper does not appear to say explicitly how this is done); in the NHMPP2 model the infinite dimensional posterior distribution is replaced by a finite dimensional approximation. A concern with the NHMPP1 approach is that the use of an approximate conditional distribution within a Gibbs sampler may lead to a Markov chain that does not converge or converges to a distribution that is not close to the target distribution. Perhaps some theoretical results can be obtained to show that there is no need for this concern, but at this point I am not aware of such results. The NHMPP2 approach, while in some sense less sophisticated, is a bit safer in that the sampler is assured to have the approximate posterior distribution as its stationary distribution.

At the modeling level, one point on which I was not clear is how the additional hierarchy level in the two models can best be used to answer questions relevant to a particular application.

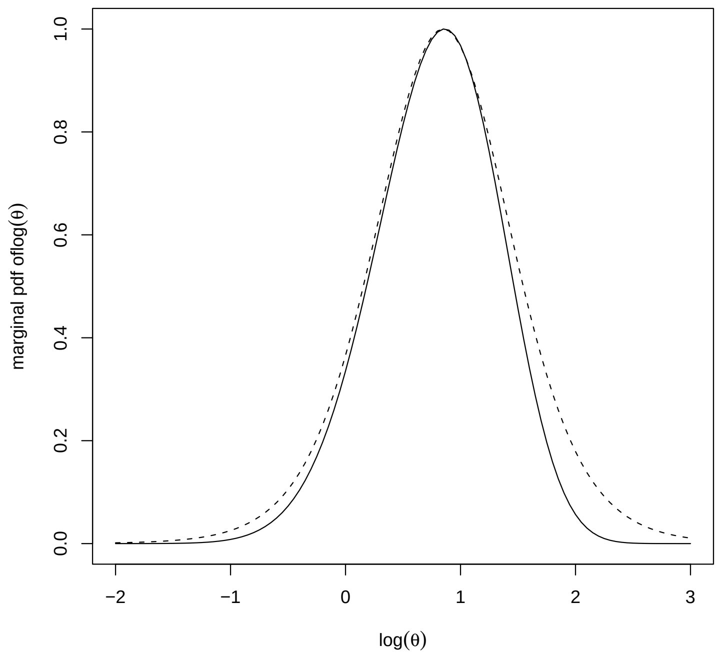

Figure 3. *Marginal Posterior Density of* $\log\theta$ *with Scaled* t *Envelope*

For the software reliability example it seems that there would be two primary types of questions. The first concerns burn-in: does the data we have so far on a particular software system suggest that we have found enough bugs to warrant release of the software? The second type of question concerns generalization: what does our experience with this project tell us about other similar projects? For the first kind of question it would seem that we would be most interested in inferences about Λ, the rate process for the particular software system under study, rather than the mean rate process $\mathcal{M}$. In this context the added hierarchy provides a richer family of prior distributions for Λ than a simple Lévy process but at the cost of considerable added complexity. It also may not by itself provide enough structure for extrapolation beyond the period of observation, which would be quite important in this context.

For answering questions of generalization the NHMPP models can be viewed as hierarchical models representing variability of individual software projects from an overall mean rate $\mathcal{M}$ at the first stage, and uncertainty about the overall mean rate $\mathcal{M}$ at the second stage. To allow for learning about the magnitude of the first stage variability it would be very useful to extend the analysis of the paper to allow for the use of data on multiple software projects, i.e. multiple realizations $\mathcal{N}$, and to allow for a prior distribution on the magnitude of the dispersion of the first stage. For the NHMPP1 model this might mean taking $f(\theta \mid \alpha)$ to have a $\text{Ga}(\alpha, \beta)$ distribution and adding a prior distribution for α; for the NHMPP2 model it might involve adding a prior distribution for at least a parametric component of the first stage Lévy process dispersion $b(\cdot)$.

As a final comment, while Lévy processes are quite popular in the literature as formal prior distributions on functional parameters I am somewhat uneasy about their use in cases where the functions are directly interpretable and would typically be expected to have some smoothness properties. In practical Bayesian analyses it is rarely possible to completely elicit the actual prior distribution that should be used. Instead, a formal prior distribution is used that is matched to certain key features that have been elicited and that is hoped to be reasonably close in other features. In well-behaved finite dimensional problems the resulting formal posterior distributions can be expected to be close to those that would have been obtained with a more complete elicitation. As the elicitation sections of the paper show, Lévy processes can be chosen to match many features that one might like a formal prior distribution to have. However, they also have some peculiar properties, such as assigning probability one to pure jump functions with jumps that are everywhere dense. In applications such as the software reliability example my prior probability that the rate function is of this form would be zero. Given this wide separation between the real prior distribution and the formal Lévy process prior distribution, it is not clear whether the resulting formal posterior distribution will be close to my actual posterior distribution in the features I am most concerned about. The inconsistency results of Diaconis and Freedman (1986) suggest there is reason for concern. Formal prior distributions based on splines or Gaussian processes that produce smooth realizations seem less likely to create such obvious discrepancies, though such distributions may have their own undesirable features.

REPLY TO THE DISCUSSION

We would like to thank Prof. Tierney for his comments, all of which raise interesting issues.

Prof. Tierney's strategy for exact sampling will undoubtedly be useful in many nonparametric (or semiparametric) settings with a low dimensional parametric component where the required integrals can be computed in closed form. For instance, for the log-beta process prior without prior fixed points of discontinuity, the posterior marginal distribution of θ would be given by

$$f(\theta \mid D) \propto \theta^n \left[\prod_{i=1}^{n} \alpha'(x_i)\psi'\{\theta + \beta(x_i)\} \, \exp\left\{ -\int_0^{\tau} \left(\psi\{\theta + \beta(u)\} - \psi\{\beta(u)\}\right) d\alpha(u) \right\} \right],$$

where $\psi(x) = d \log \Gamma(x)/dx$ denotes the digamma function. This turns out to be very similar to the one obtained by Prof. Tierney for the extended gamma case, and could therefore be also sampled from using a rejection method.

Concerning the second technical point, we would like to emphasise that in the Ferguson and Klass's representation of a Lévy process as an infinite sum of random jumps at random locations (see Section 5.1), the sequence of jumps $J_1, J_2, \ldots$ is decreasing. Therefore, when truncating the sum in the representation of $L_c(t)$, the dropped terms always correspond to the smallest jumps. This is in contrast to alternative algorithms such as that of Wolpert and Ickstadt (1998b), for which truncation is not as straightforward.

As for the actual truncation method, one possibility is to monitor, at each iteration of the Gibbs sampler, the relative error

$$\mathrm{RE}(r) = \left| \frac{S_r - S_{r-1}}{S_r} \right|$$

until $\mathrm{RE}(r) \leq \epsilon$, where $S_r = \sum_{i=1}^{r} J_i$ is the cumulative sum of the jump sizes and ϵ is a small positive constant. Thus, at each iteration the truncation point would be given by the minimum value of r satisfying the last condition.

Alternatively, one could truncate the sum uniformly, fixing the number of terms at R, say, across all iterations of the Gibbs sampler. The value of R can be determined on the basis of a few preliminary runs. Actually, this was the procedure we followed in the example of Section 5.1, where we used $R = 500$.

As far as modelling is concerned, we regard both NHMPP1 and NHMPP2 as providing alternatives to the usual non-homogeneous Poisson process at the "likelihood level", and not necessarily as providing a richer family of prior distributions for $\Lambda(\cdot)$. Acknowledgedly our notation is slightly confusing, but what we are assuming is that the observed data are samples from the marginal distribution of $\mathcal{N}(\cdot)$ with respect to the corresponding mixture representation. We then put a prior on $\mathcal{M}(\cdot) = \mathrm{E}\{\mathcal{N}(\cdot)\}$ via

(i) a Lévy prior for $\Lambda(\cdot)$ in the NHMPP1 case, where $\mathcal{M}(t) = \mu_\theta \Lambda(t)$;

(ii) a Lévy prior for $a(\cdot)$ in the NHMPP2 case, where, for example, $\mathcal{M}(t) = a(t)/b$ if we let $\Lambda(\cdot)$ have a gamma process GaP$\{a(\cdot), b\}$.

In this latter case we do not see the distribution on $\Lambda(\cdot)$ as a "prior", but rather as a mixing distribution which gives rise to a model that is overdispersed relative to the usual NHPP. Thus, in the example of Section 5.2, where we used a gamma process for $\Lambda(\cdot)$, we ended up with a negative binomial process (of type 2) for $\mathcal{N}(\cdot)$, a process parameterised by $a(\cdot)$.

In both cases the hierarchical representation is a convenient device allowing us to carry out the analysis via a Gibbs sampler. In view of the discussion above, we feel that the function of interest should be the mean $\mathcal{M}(\cdot)$ of the marginal process $\mathcal{N}(\cdot)$, and not $\Lambda(\cdot)$.

The hierarchical model discussed by Prof. Tierney, describing the variability of individual software projects, looks very interesting and certainly deserves further study.

We agree with Prof. Tierney that care must be taken when choosing nonparametric priors, and that consistency is a genuine concern in such cases. However, recent and ongoing work by several authors (Barron *et al.*, 1999; Kim and Lee, 2001; Walker and Hjort, 2001; Walker, 2002ab) shows that, under certain conditions on the prior processes, many of the priors commonly used in applications yield consistent posteriors.

Concerning the last point of the discussion about the continuity of the prior process, apart from the priors suggested by Prof. Tierney, as mentioned in the article one can also consider kernel mixture (Lévy-driven) priors of the form

$$\Lambda(t) = \int h(t, \nu) dL(\nu)$$

where $L(\nu)$ is the driving Lévy process and $h(\cdot, \cdot)$ is a suitable kernel function. Such priors have been successfully used by Nieto-Barajas (2001) and Nieto-Barajas and Walker (2001) in the context of survival analysis.

ADDITIONAL REFERENCES IN THE DISCUSSION

Barron, A., Schervish, M.J. and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27**, 536–561.

Diaconis, P. and Freedman, D. (1986). On inconsistent Bayes estimates of location. *Ann. Statist.* **14**, 68–87.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. B* **40**, 31–38.

Kim, Y. and Lee, J. (2001). On posterior consistency of survival models. *Ann. Statist.* **29**, 666–686.

Walker, S.G. and Hjort, N.L. (2001). On Bayesian consistency. *J. Roy. Statist. Soc. B* **63**, 811–821.

Walker, S.G. (2002a). On sufficient conditions for Bayesian consistency. *Tech. Rep.*, University of Bath, UK.

Walker, S.G. (2002b). A new approach to Bayesian consistency. *Tech. Rep.*, University of Bath, UK.