Statistica Sinica **14**(2004), 1127-1146

BAYESIAN NONPARAMETRIC SURVIVAL ANALYSIS VIA LÉVY DRIVEN MARKOV PROCESSES

Luis E. Nieto-Barajas and Stephen G. Walker

ITAM, Mexico and University of Kent, UK

Abstract: In this paper we present and investigate a new class of nonparametric priors for modelling a cumulative distribution function. We take $F(t) = 1 - \exp\{-Z(t)\}$, where $Z(t) = \int_0^t x(s)\,\mathrm{d}s$ is continuous and $x(\cdot)$ is a Markov process. This is in contrast to the widely used class of neutral to the right priors (Doksum (1974)) for which $Z(\cdot)$ is discrete and has independent increments. The Markov process allows the modelling of trends in $Z(\cdot)$, not possible with independent increments. We derive posterior distributions and present a full Bayesian analysis.

Key words and phrases: Bayes nonparametrics, consistency, Lévy process, gamma process, Markov process, stationary process, Lévy driven Markov process.

1. Introduction

In this paper, we present a new Bayesian nonparametric approach to modeling an unknown survival function. Previous work in this area has focused on independent increment processes (Lévy processes), used to construct neutral to the right processes (Doksum (1974)). The Dirichlet process (Ferguson (1973)) is a particular case. Ferguson and Phadia (1979) detailed the posterior distributions for censored and uncensored survival data and, in particular, worked with the simple homogeneous and gamma processes, as well as the Dirichlet process. Recent work on neutral to the right priors has been done by Hjort (1990), Walker and Muliere (1997) and Walker and Damien (1998). In a more general context, Kim (1999) used independent increment processes as a prior distribution for the cumulative intensity function of multiplicative counting processes.

As is well known, the main drawback of models based on independent increment processes is that, with probability one, the survival function is discrete. In order to overcome this problem, Dykstra and Laud (1981) considered an independent increment process to model the *hazard rate* function. The corresponding cumulative hazard function, and hence the survival function, are continuous. A disadvantage of this is that the model only allows monotone hazard rate functions. Arjas and Gasbarra (1994) proposed a Markov jump process having a martingale structure. However, this model is somewhat complicated, lacks interpretability and relies on a partition.

In this article we use a piecewise continuous Markov process to model the hazard rate function and, by integration, obtain a continuous process to model the cumulative hazard and survival functions. Hence we retain the desirable continuous models without the limitations of a monotone hazard rate. It should be pointed out that our model is perhaps the simplest extension of the neutral to the right prior which eliminates the drawbacks associated with it. Indeed, the updates from prior to posterior for our model closely resemble those for the neutral to the right prior, once a number of strategic latent variables have been introduced.

Lo and Weng (1989), Brix (1999) and James (2002) consider a mixture model for the hazard rate function which is similar to our own. We will comment more on their prior models in Section 2. Wolpert and Ickstadt (1998), on the other hand, use a kernel mixture to model the cumulative intensity function in a Cox process and, in particular, they take an extended gamma process as the mixing measure.

The outline of the article is as follows. In Section 2 we present relevant background material and introduce the new model. In Section 3 we describe posterior distributions based on observed censored and uncensored observations. Section 4 considers prior elicitation and Section 5 deals with the simulation methods for inference purposes. In Section 6 we present two numerical examples and Section 7 deals with consistency issues. The paper concludes in Section 8 with a discussion.

Before this, we introduce some notation. Let $\mathrm{Ga}(\alpha,\beta)$ denotes a gamma density with mean α/β; $\mathrm{Po}(c)$ a Poisson density with mean c; $\mathrm{U}(a,b)$ a uniform density on $[a,b]$; $T_1,\ldots,T_n$ are independent failure times, possibly with the inclusion of random right censored times.

2. Background and New Model

Let T be a continuous random variable defined on $(0,\infty)$ with (conditional) cumulative distribution function given by $F(t|Z)=1-\exp\{-Z(t)\}$ and density function $f(t|Z)=x(t)\exp\{-Z(t)\}$, where $Z(\cdot)$ is the cumulative hazard function given by $Z(t)=\int_0^t x(s)\,\mathrm{d}s$, with $x(t)$ the hazard rate function.

Let $L(t)$ be an independent increments (Lévy) process defined on $[0,\infty)$ without Gaussian components, so $L(t)$ is a pure jump process (Ferguson and Klass (1972)). If L has Lévy measure $\mathrm{d}N_\tau(\nu)$, a non-negative measure satisfying $\int_0^\infty \min\{1,\nu\}\,\mathrm{d}N_\tau(\nu)<\infty$, then

$$-\log \mathrm{E}[\exp\{-\theta L(\tau)\}]=\int_{\nu>0}\left(1-e^{-\theta\nu}\right)\,\mathrm{d}N_\tau(\nu).$$

Also, for $\mathrm{d}L(\tau) = L(\tau + \mathrm{d}\tau) - L(\tau)$,

$$-\log \mathrm{E}[\exp\{-\theta(\tau)\mathrm{d}L(\tau)\}] = \int_{\nu>0} [1 - \exp\{-\theta(\tau)\nu\}]\{\mathrm{d}N_{\tau+\mathrm{d}\tau}(\nu) - \mathrm{d}N_\tau(\nu)\}.$$

So, if $\mathrm{d}N_\tau(\nu) = \mathrm{d}\nu \int_0^\tau \gamma(\nu, s)\,\mathrm{d}s$, which will be the case for us, then

$$-\log \mathrm{E}[\exp\{-\theta(\tau)\mathrm{d}L(\tau)\}] = \mathrm{d}\tau \int_{\nu>0} [1 - \exp\{-\theta(\tau)\nu\}]\gamma(\nu, \tau)\,\mathrm{d}\nu. \tag{1}$$

For the neutral to the right (NTR) model (Doksum (1974)),

$$Z_{NTR}(t) = \int_0^t \mathrm{d}L(\nu) = L(t).$$

For example, $L(\cdot)$ can be a simple homogeneous process, gamma process, log-beta process and so on. Dykstra and Laud (1981) considered

$$Z_{DL}(t) = \int_0^t (t - \nu)\,\mathrm{d}L(\nu)$$

and, in particular, they took $L(\cdot)$ to be an extended gamma process. Here $Z_{DL}(\cdot)$ has continuous sample paths but it is easy to see that in this case the hazard rate function is given by $x_{DL}(t) = \int_0^t \mathrm{d}L(\nu)$, and so is monotone.

Our proposal for the cumulative hazard function is continuous but we have removed the monotone condition of the hazard rate function. We consider

$$Z(t) = \int_0^t \frac{1}{a} \left\{1 - e^{-a(t-\nu)}\right\} \mathrm{d}L(\nu) \tag{2}$$

for some $a > 0$. In this case, the hazard rate function is given by

$$x(t) = \int_0^t \exp\{-a(t - \nu)\}\ \mathrm{d}L(\nu). \tag{3}$$

Due to the fact that for $a > 0$ and $t > \nu$, the factors $\{\exp^{-a(t-\nu)}\}$ are decreasing functions of t for each ν, the sample paths of (3) are non-monotone functions. Moreover, the sample paths of the hazard rate process $x(t)$ are piece-wise continuous functions. Therefore, the cumulative hazard process $Z(t)$, given by (2), is continuous with probability one. As will be shown, models based on (3) are tractable. We can force L to have a jump at $t = 0$ and in this case $x(0) = L\{0\}$. The interpretation of a will be given later in Section 4.

Some other authors have considered kernel mixture models of the type

$$x(t) = \int k(t, u)\,\mathrm{d}L(u). \tag{4}$$

Lo and Weng (1989) take $L(u)$ to be a weighted gamma process and k any suitable kernel function. They provide a representation or derivation of the posterior distribution but, due to the complicated form of their posterior, they concentrate on finding posterior means and present a way to approximate them using a Monte Carlo method.

We take L to be fundamentally different in that L will have a finite number of jumps in any bounded interval. This leads to certain advantages which will become clear later. Also, we provide a full posterior analysis via Markov chain Monte Carlo methods.

Brix (1999) takes k to be an arbitrary kernel and $L(u)$ to be a generalized gamma process. The homogeneous version of the L process we consider here (see Theorem 1 in Section 2) can be seen as a generalized gamma process. Brix also describes how to simulate from this type of process.

James (2002, Section 4) also considers models of the type (4), taking $L(u)$ to be a general size-based random measure. This general measure contains the generalized gamma processes and the weighted gamma processes as particular cases. James (2002) presents a characterization of the posterior distribution and shows a procedure to approximate posterior quantities for all kernels k and all measures L via a MCMC method based on a "Chinese restaurant" type algorithm.

The approach we follow to derive posterior distributions relies on latent variables which facilitates an easy-to-implement MCMC method for simulation. Similarly, James (2002) allows the use of missing observations, which can be seen as latent variables, in his derivations of the posterior. However, the procedure we use to obtain the posterior is considerably different.

Our kernel is novel and has roots in a discrete time model (Nieto-Barajas and Walker (2002)), the continuous time model being constructed by allowing the time intervals to collapse to zero. More details about this are included in the discussion.

The process (3) is a shot-noise process with exponentially decaying shocks, that is

$$x(t) = \sum_i \exp\{-a(t-\theta_i)\} J_i \, I(\theta_i \le t), \tag{5}$$

where the $\{J_i\}$ are the random shocks and the $\{\theta_i\}$ are the random times when they occur. For different choices of $L(\nu)$ we obtain different processes $x(t)$. Here we state a useful result detailing conditions under which $x(t)$ is a stationary gamma process. This will then enable us to build up a model based on this type of process.

Theorem 1. *If $L(\tau)$ is the Lévy process with Lévy measure $dN_\tau(\nu) = \delta\, \tau\, d\nu \exp(-\nu)$ and $x(t) = \int_0^t \exp(-a\nu)\, dL(t-\nu)$, then marginally $x(t) \sim Ga(\alpha, 1)$, where $\alpha = \delta/a$.*

An alternative representation of $x(t)$, and a proof of Theorem 1 based on compound Poisson processes, can be found in Ross ((1996), Chap.8). In addition, Barndorff-Nielsen and Shephard (1999) obtained a more general result to Theorem 1 involving Generalised-Inverse-Gaussian families. While stationary Lévy processes play a large role in many models, we will be employing non-stationary processes, which have not been so widely used. Where they have been used, full Bayesian posterior analyses on real data sets have been rare. We have to note that for the Lévy measure of Theorem 1, and for the more general Lévy measure that will be introduced in Section 3, representation (5) of the process $x(t)$ has a finite number of jumps. In Section 3 we introduce novel auxiliary variables which permit a full and quite straightforward Bayesian analysis via MCMC methods.

3. Posterior Distributions

In this section we show how to obtain posterior distributions. From now on we write

$$Z(t) = \int_0^\infty k(t,\nu)\,\mathrm{d}L(\nu), \tag{6}$$

$$k(t,\nu) = \frac{1}{a}\left[1 - \exp\left\{-a(t-\nu)^+\right\}\right]. \tag{7}$$

Let $\mathcal{B}$ be the space of cumulative hazard functions, that is, the set of all non-decreasing, right continuous functions such that if $Z \in \mathcal{B}$ then $Z(0) = 0$ and $Z(t) \to \infty$ as $t \to \infty$. Then we consider the probability P' defined on $([0,\infty) \times \mathcal{B}, \mathcal{A} \times \sigma(\mathcal{B}))$, where $\mathcal{A}$ is Borel's σ-algebra on $[0,\infty)$ and $\sigma(\mathcal{B})$ is the σ-algebra generated by the Borel sets on $\mathcal{B}$ with the Skorokov metric;

$$\mathrm{P}'(T > t, Z \in B) = \mathrm{E}_Z\left[\exp\{-Z(t)\}\, I(Z \in B)\right],$$

for $B \in \sigma(\mathcal{B})$. We are concerned with probability measures on $\{\mathcal{B}, \sigma(\mathcal{B})\}$. If P_0 is one of those measures, with integral operator E_0 defined by $\mathrm{E}_0[\psi(Z)] = \int \psi(Z)\mathrm{P}_0(\mathrm{d}Z)$, it is specified when all finite-dimensional $\mathrm{P}_0\{Z[t_{j-1},t_j) \in D_j,\ j = 1,\ldots,k\}$, for $D_j \in \mathcal{A}$, are known. However, this is equivalent to knowing the Laplace transforms $\mathrm{E}_0[\exp\{-\sum_{j=1}^k \theta_j Z[t_{j-1},t_j)\}]$. In other words, knowledge of $\mathrm{E}_0[\exp\{-\int_0^\infty \theta(s)\mathrm{d}Z(s)\}]$ for all $\theta(s)$, and in particular for $\theta(s) = \sum_{j=1}^k \theta_j I\{s \in [t_{j-1},t_j)\}$, is sufficient to specify P_0 on $\{\mathcal{B}, \sigma(\mathcal{B})\}$ completely. See also Hjort (1990).

In our case

$$\mathrm{E}\left[\exp\left\{-\int \theta(s)\,\mathrm{d}Z(s)\right\}\right] = \mathrm{E}\left[\exp\left\{-\int K_\theta(\nu)\,\mathrm{d}L(\nu)\right\}\right],$$

where $K_\theta(\nu) = \int_{s>\nu} \theta(s)\exp\{-a(s-\nu)\}\,\mathrm{d}s$. Therefore, the aim is to find and understand, for generic $\kappa(\cdot)$,

$$\mathrm{E}\left[\exp\left\{-\int \kappa(\nu)\,\mathrm{d}L(\nu)\right\}\middle| A\right]$$

for two types of observations: $A = \{T > t\}$ and $A = \{T = t\}$. From there we build up a picture of the posterior distributions via changes in the Lévy process.

Before we do this, we can adjust the Lévy measure to make it more general:

$$\mathrm{d}N_\tau(\nu) = \mathrm{d}\nu \int_0^\tau \exp\{-\nu\beta(u)\}\,\mathrm{d}\alpha(u), \tag{8}$$

where $\alpha(\cdot)$ is a continuous measure and $\beta(\cdot)$ is a non-negative piece-wise continuous function. Note that $x(t)$ is the stationary gamma process of Theorem 1 when $\alpha(u) = \delta u$ and $\beta(u) = 1$.

Remark. In Theorems 2 and 3, without lost of generality, we disregard the fixed jump of the Lévy process at zero and consider it again in Theorem 4.

Theorem 2. *The posterior Lévy measure given an observation $T > t$ is given by*

$$dN'_\tau(\nu) = d\nu \int_0^\tau \exp\{-\nu\beta'(u)\}d\alpha(u),$$

where $\beta'(u) = \beta(u) + k(t,u)$.

Proof. Let

$$\varphi(\kappa,t) = \mathrm{E}\left[\exp\left\{-\int \kappa(\nu)\,\mathrm{d}L(\nu)\right\}\middle| T > t\right].$$

Using (6) and noting that $\mathrm{P}(T > t) = \mathrm{E}\{\mathrm{P}(T > t|L)\}$, $\varphi(\kappa,t)$ can be expressed as

$$\frac{\mathrm{E}\left(\exp\left[-\int\{\kappa(\nu) + k(t,\nu)\}\,\mathrm{d}L(\nu)\right]\right)}{\mathrm{E}\left[\exp\left\{-\int k(t,\nu)\,\mathrm{d}L(\nu)\right\}\right]}.$$

Since

$$\mathrm{E}\left[\exp\left\{-\psi(\nu)\,\mathrm{d}L(\nu)\right\}\right] = \exp\left\{-\frac{\psi(\nu)\mathrm{d}\alpha(\nu)}{\beta(\nu)\left\{\psi(\nu) + \beta(\nu)\right\}}\right\},$$

see (1), and using independence properties, $\varphi(\kappa,t)$ becomes

$$\begin{aligned} &\exp\left\{-\int\left[\frac{\kappa(\nu) + k(t,\nu)}{\beta(\nu)\{\kappa(\nu) + k(t,\nu) + \beta(\nu)\}} - \frac{k(t,\nu)}{\beta(\nu)\{k(t,\nu) + \beta(\nu)\}}\right]\,\mathrm{d}\alpha(\nu)\right\} \\ &= \exp\left\{-\int \frac{\kappa(\nu)}{\{k(t,\nu) + \beta(\nu)\}\{\kappa(\nu) + k(t,\nu) + \beta(\nu)\}}\,\mathrm{d}\alpha(\nu)\right\}. \end{aligned}$$

Finally, we have $\varphi(\kappa,t) = \mathrm{E}\{\exp(-\int \kappa(\nu)\,\mathrm{d}L'(\nu))\}$, where $L'(\cdot)$ is a Lévy process with Lévy measure as stated in the Theorem.

Consequently, given n observations $T_1 > t_1, \ldots, T_n > t_n$, the posterior Lévy measure becomes $\mathrm{d}N'_\tau(\nu) = \mathrm{d}\nu \int_0^\tau \exp\{-\nu\beta'(u)\}\,\mathrm{d}\alpha(u)$, where now $\beta'(u) = \beta(u) + \sum_{1\le i\le n} k(t_i,u)$.

It is more difficult to find the posterior distribution given an observation $T = t$. In order to do this we need to introduce a latent parameter or observation.

We assume we have witnessed, along with $T = t$, the observation $S = s$, where the random mass function for S is given by

$$f(s|t, L) = \mathrm{P}(S = s|t, L) \propto e^{-a(t-s)} L\{s\}\, I(0 \le s \le t), \tag{9}$$

coming from the joint density

$$f(t, s|L) = e^{-a(t-s)} L\{s\} \exp\left\{-\int k(t, \nu)\, \mathrm{d}L(\nu)\right\} I(0 \le s \le t), \tag{10}$$

where $L\{s\} = L(s) - L(s-)$. Of course, this ensures that

$$f(t|L) = x(t) \exp\left\{-\int_0^t x(s)\, \mathrm{d}s\right\}.$$

This idea was also useful for the extended gamma process model for $x(\cdot)$ (see Laud, Damien and Walker (1999)). Experience with updating Lévy processes for modelling cumulative hazard functions (see, for example, Walker and Muliere (1997)) leads to the following result.

Theorem 3. *The posterior Lévy measure given an observation $T = t$ and $S = s$ is given by*

$$dN'_\tau(\nu) = d\nu \int_0^\tau \exp\{-\nu\beta'(u)\}\, d\alpha(u),$$

where $\beta'(u) = \beta(u) + k(t, u)$ and the updated Lévy process has a fixed point of discontinuity at s with the distribution of $L'\{s\}$ being $Ga(2, \beta(s) + k(t, s))$.

Proof. Let

$$\varphi(\kappa, t, s, \epsilon) = \mathrm{E}\left[\exp\left\{-\int \kappa(\nu)\, \mathrm{d}L(\nu)\right\}\middle| T = t, S \in [s, s+\epsilon]\right],$$

which can be obtained by

$$\frac{\mathrm{E}\left[\exp\left\{-\int \kappa(\nu)\, \mathrm{d}L(\nu)\right\} \int_s^{s+\epsilon} f(t, \omega)\, \mathrm{d}\omega\right]}{\mathrm{E}\left\{\int_s^{s+\epsilon} f(t, \omega)\, \mathrm{d}\omega\right\}}.$$

Using (6) and (10), $\varphi(\kappa, t, s, \epsilon)$ becomes

$$\frac{\mathrm{E}\left(\int_s^{s+\epsilon} h(t, \omega)\, \mathrm{d}L(\omega) \exp\left[-\int_0^\infty \{\kappa(\nu) + k(t, \nu)\}\, \mathrm{d}L(\nu)\right]\right)}{\mathrm{E}\left[\int_s^{s+\epsilon} h(t, \omega)\, \mathrm{d}L(\omega) \exp\left\{-\int_0^\infty k(t, \nu)\, \mathrm{d}L(\nu)\right\}\right]},$$

where $h(t, \omega) = \exp\{-a(t - \omega)\}$. Splitting up the integral,

$$\varphi(\kappa, t, s, \epsilon) = \frac{\mathrm{E}\left(\exp\left[-\int_{[0,s)\cup(s+\epsilon,\infty)} \{\kappa(\nu) + k(t, \nu)\}\, \mathrm{d}L(\nu)\right]\right)}{\mathrm{E}\left[\exp\left\{-\int_{[0,s)\cup(s+\epsilon,\infty)} k(t, \nu)\, \mathrm{d}L(\nu)\right\}\right]} D(s, s+\epsilon),$$

where

$$D(s,s+\epsilon)=\frac{\mathrm{E}\left(\int_s^{s+\epsilon}h(t,\omega)\,\mathrm{d}L(\omega)\exp\left[-\int_s^{s+\epsilon}\left\{\kappa(\nu)+k(t,\nu)\right\}\,\mathrm{d}L(\nu)\right]\right)}{\mathrm{E}\left[\int_s^{s+\epsilon}h(t,\omega)\,\mathrm{d}L(\omega)\exp\left\{-\int_s^{s+\epsilon}k(t,\nu)\,\mathrm{d}L(\nu)\right\}\right]}.$$

Since

$$\mathrm{E}\left[\,\mathrm{d}L(\nu)\exp\left\{-\psi(\nu)\,\mathrm{d}L(\nu)\right\}\right]=\frac{\mathrm{d}\alpha(\nu)}{\left\{\psi(\nu)+\beta(\nu)\right\}^2}\exp\left[-\frac{\psi(\nu)\mathrm{d}\alpha(\nu)}{\beta(\nu)\left\{\psi(\nu)+\beta(\nu)\right\}}\right],$$

and using independence properties, $D(s,s+\epsilon)$ can be written as

$$A(s,s+\epsilon)\exp\left[-\int_s^{s+\epsilon}\frac{\kappa(\nu)}{\left\{k(t,\nu)+\beta(\nu)\right\}\left\{\kappa(\nu)+k(t,\nu)+\beta(\nu)\right\}}\mathrm{d}\alpha(\nu)\right],$$

where the second part comes using the proof of Theorem 2 and

$$A(s,s+\epsilon)=\int_s^{s+\epsilon}\frac{h(t,\omega)\,\mathrm{d}\alpha(\omega)}{\left\{\kappa(\omega)+k(t,\omega)+\beta(\omega)\right\}^2}\bigg/\int_s^{s+\epsilon}\frac{h(t,\omega)\,\mathrm{d}\alpha(\omega)}{\left\{k(t,\omega)+\beta(\omega)\right\}^2}\,.$$

Now,

$$\lim_{\epsilon\to 0}A(s,s+\epsilon)=\left\{\frac{k(t,s)+\beta(s)}{\kappa(s)+k(t,s)+\beta(s)}\right\}^2,$$

then, based again on the proof of Theorem 2, we have

$$\begin{aligned}\lim_{\epsilon\to 0}\varphi(\kappa,t,s,\epsilon)=&\left\{\frac{k(t,s)+\beta(s)}{k(t,s)+\beta(s)+\kappa(s)}\right\}^2\\ &\times\exp\left[-\int_0^{\infty}\frac{\kappa(\nu)}{\left\{k(t,\nu)+\beta(\nu)\right\}\left\{\kappa(\nu)+k(t,\nu)+\beta(\nu)\right\}}\,\mathrm{d}\alpha(\nu)\right]\\ =&\,\mathrm{E}\left\{\exp\left(-\kappa(s)L'\{s\}\right)\right\}\ \mathrm{E}\left[\exp\left\{-\int\kappa(\nu)\,\mathrm{d}L'(\nu)\right\}\right]\end{aligned}$$

with $L'\{s\}$ as stated in the theorem. This completes the proof.

We are now in a position to write down the full posterior distributions, allowing the process to have prior fixed jumps.

Theorem 4. *Let $L(\cdot)$ be a Lévy process such that $L(t)=L_c(t)+\sum_j L\{\tau_j\}I(\tau_j\leq t)$ with Lévy measure for the "continuous" part $dN_\tau(\nu)=d\nu\int_0^\tau\exp\{-\nu\beta(u)\}d\alpha(u)$, and let $M=\{\tau_1,\tau_2,\ldots\}$ be the set of prior fixed points of discontinuity of $L(\cdot)$ and the density function of $L\{\tau_j\}$ be f_j.*

(i) *Given an observation $T>t$, the posterior Lévy measure is given by*

$$dN'_\tau(\nu)=d\nu\int_0^\tau\exp\{-\nu\beta'(u)\}\,d\alpha(u),$$

where $\beta'(u) = \beta(u) + k(t, u)$. *Furthermore,* $M' = M$ *and*

$$f_j'(x) \propto \begin{cases} e^{-k(t,\tau_j)x} f_j(x) & \text{if } t > \tau_j \\ f_j(x) & \text{otherwise.} \end{cases}$$

(ii) *Given an observation* $T = t$ *and* $S = s$, *the posterior Lévy measure is given by*

$$dN_\tau'(\nu) = d\nu \int_0^\tau \exp\{-\nu\beta'(u)\}\, d\alpha(u),$$

where $\beta'(u) = \beta(u) + k(t, u)$.

(a) *If* $s \notin M$ *then* $M' = M \cup \{s\}$ *and* f_s' *is* $Ga(2, \beta(s) + k(t, s))$. *Furthermore,*

$$f_j'(x) \propto \begin{cases} e^{-k(t,\tau_j)x} f_j(x) & \text{if } t > \tau_j \\ f_j(x) & \text{otherwise.} \end{cases}$$

(b) *If* $s \in M$ *then* $M' = M$ *and* $f_s'(x) \propto x e^{-k(t,s)x} f_s(x)$. *Furthermore, for* $\tau_j \neq s$,

$$f_j'(x) \propto \begin{cases} e^{-k(t,\tau_j)x} f_j(x) & \text{if } t > \tau_j \\ f_j(x) & \text{otherwise.} \end{cases}$$

Proof. The proof follows from Theorems 2 and 3.

It is now possible to build the full posterior based on a sample of size n. Recall the s are latent and are required whenever an uncensored observation is witnessed. The density for s given t and $L(\cdot)$ has been given in (9). In practice, a Gibbs sampler (see, for example, Smith and Roberts (1993)) would be needed to sample from the joint posterior distribution of $L(\cdot)$ and s.

4. Prior Specifications

In Section 3 we introduced a more general Lévy measure given by (8) Let $L_c(\cdot)$ represent the Lévy process induced by (8) without any fixed points of discontinuity. Then

$$\mathrm{E}\left[\exp\left\{-\int_0^t \psi(\nu)\,\mathrm{d}L_c(\nu)\right\}\right] = \exp\left[-\int_0^t \frac{\psi(\nu)\,\mathrm{d}\alpha(\nu)}{\beta(\nu)\{\psi(\nu) + \beta(\nu)\}}\right].$$

This Lévy process is not necessarily homogeneous. Consequently, the hazard rate function $x(t)$ will not be stationary. For this more general process we require that $x(t)$, as in (3), exists. Noticing that $L_c(t) < \infty$ for all finite t ensures that the process $x(t)$ exists in a pathwise sense with probability one. Additionally, we also need that $S(t) = \exp\{-Z(t)\} \to 0$ as $t \to \infty$, to ensure that $Z(t)$, defined in (6), is a cumulative hazard function with probability one. This occurs provided

$$\int_0^t \frac{k(t,\nu)\,\mathrm{d}\alpha(\nu)}{\beta(\nu)\,\{k(t,\nu) + \beta(\nu)\}} \to \infty$$

as $t \to \infty$.

An important issue is how to specify the prior process, that is, how to choose α and β. One way of doing this is to match the first two moments of $x(t)$ with functions $\mu(t) = \mathrm{E}\{x(t)\}$ and $\sigma(t) = \mathrm{Var}\{x(t)\}$. This will be based on the two equations

$$\mu(t) = \mathrm{E}(L\{0\}) \exp(-at) + \int_0^t \exp\{-a(t-\nu)\}\, \mathrm{d}\alpha(\nu)/\beta^2(\nu), \tag{11}$$

$$\sigma(t) = \mathrm{Var}(L\{0\}) \exp(-2at) + 2\int_0^t \exp\{-2a(t-\nu)\}\, \mathrm{d}\alpha(\nu)/\beta^3(\nu). \tag{12}$$

We consider a prior process with a fixed jump at zero in order to allow non-zero hazard rates at time $t = 0$. The following lemma gives the conditions for achieving this.

Lemma 5. *Let $L(t) = L\{0\} + L_c(t)$ be a Lévy process which includes a fixed jump at zero, and let $\mu(t)$ and $\sigma(t)$ be nonnegative and differentiable functions on $[0, \infty)$. Let $a > 0$ be a constant. If $\mu(t)$, $\sigma(t)$ and a satisfy*

$$a \geq \max\left\{-\frac{d}{dt}\log\mu(t), -\frac{1}{2}\frac{d}{dt}\log\sigma(t)\right\} \tag{13}$$

for all $t \geq 0$, then the parameters $\alpha(\cdot)$ and $\beta(\cdot)$ which ensure that $\mu(t) = E\{x(t)\}$ and $\sigma(t) = Var\{x(t)\}$ for all $t \geq 0$, are given by

$$\beta(t) = \frac{a\mu(t) + d\mu(t)/dt}{a\sigma(t) + 0.5 d\sigma(t)/dt} \quad \text{and} \quad d\alpha(t) = \{a\mu(t)dt + d\mu(t)\}\,\beta^2(t).$$

To satisfy the initial values we also need $\mu(0) = E(L\{0\})$ and $\sigma(0) = Var(L\{0\})$.

Proof. Differentiating (11) and (12) with respect to t, we obtain

$$\mathrm{d}\mu(t) = -a\mu(t)\mathrm{d}t + \mathrm{d}\alpha(t)/\beta^2(t) \quad \text{and} \quad \mathrm{d}\sigma(t) = -2a\sigma(t)\,\mathrm{d}t + 2\mathrm{d}\alpha(t)/\beta^3(t).$$

Then, solving these simultaneous differential equations we obtain the required expressions. The conditions on $\mu(t)$ and $\sigma(t)$ arise when constraining $\beta(t) \geq 0$ and $\alpha(t) \geq 0$ for $t \geq 0$. This completes the proof.

If $\mu(t) = \exp(-at)$ then $-\mathrm{d}/\mathrm{d}t \log\mu(t) = a$. So we can interpret the quantity $-\mathrm{d}/\mathrm{d}t \log\mu(t)$ as the rate of decay of the function $\mu(t)$. Therefore, the prior process can be centred on any non-negative function $\mu(t)$ whose local rate of decay (if any) is slower than the rate of decay of the negative exponential function $\exp(-at)$. In the same way, a similar interpretation can be derived for $\sigma(t)$.

One simple choice for the parameters of the prior density of the fixed jump at $t = 0$ would be $L\{0\} \sim \mathrm{Ga}(\alpha_0, \beta_0)$, with α_0 and β_0 such that $\mu(0) = \alpha_0/\beta_0$ and

$\sigma(0) = \alpha_0/\beta_0^2$. One way of proposing the functions $\mu(\cdot)$ and $\sigma(\cdot)$ is to obtain them from the mean and variance of the hazard rate of a Bayesian parametric model (see Walker, Damien, Laud and Smith (1999)). This allows us to express prior opinion about location and uncertainty using a recognisable model. For example, we might want to centre $x(t)$ on a constant hazard rate, i.e., $S(t) = \exp(-\eta t)$ with prior distribution $\eta \sim \mathrm{Ga}(p,q)$. Solving the two differential equations from Lemma 5, we obtain $\beta(t) = q$ and $\mathrm{d}\alpha(t) = apq\,\mathrm{d}t$, with prior distribution for $L\{0\} \sim \mathrm{Ga}(p,q)$. An important remark here is that without the fixed jump at zero we would not be able to centre the prior process on a constant hazard rate.

Moreover, if we want to centre $x(t)$ on a monotone hazard rate, we can use the Weibull parametric model $S(t) = \exp(-\eta t^b)$ with the same gamma prior for η as in the constant hazard model. Keeping b fixed, we get $\beta(t) = (q/b)t^{-(b-1)}$ and $\mathrm{d}\alpha(t) = (pq/b)t^{-b}(b-1+at)\,\mathrm{d}t$, with no fixed prior jump at $t = 0$ if $b > 1$. If $b = 1$ we get the same specifications of the constant hazard model. The nonparametric prior is not available if $b < 1$.

Besides satisfying (13), one way of determining the appropriate value of the parameter a is setting the degree of correlation of the (stationary) process between t and $t+1$, that is, $\mathrm{Corr}\,\{x(t), x(t+1)\} = e^{-a}$. For smaller values of a the correlation between t and $t+1$ is larger. We can view a as a smoothing parameter.

The conditions imposed by Lemma 5 on the functions $\mu(t)$ and $\sigma(t)$ are not restrictive. For example, if $\mu(t)$ is a rough fluctuating function then we need a low correlated prior process to model the fluctuations, that is to say, we need a large value of a. This is precisely what (13) requires because $\mu(t)$ is expected to have a large rate of decay. In the same way, if $\mu(t)$ is a fairly smooth function (like a bathtub shape for example) then we need a small value of a to produce a highly correlated process and model trends. It is worth mentioning that, for non-decreasing functions $\mu(t)$ and $\sigma(t)$, the parameter a can take any positive value.

5. Posterior Simulation

In order to undertake a full Bayesian analysis, we need to simulate the Lévy process $L(\cdot)$. One can notice that the Lévy measure for our $x(t)$ process is finite, whereas previously employed measures used for neutral to the right processes, such as the beta-Stacy process (Walker and Muliere (1997)), the beta process (Hjort (1990)) and the extended gamma process (Dykstra and Laud (1981)), are infinite. The importance of having a finite Lévy measure is that the induced Lévy process can be simulated exactly, whilst a Lévy process with infinite Lévy measure cannot. One way of implementing the simulation is to use the result of Ferguson and Klass (1972), which is now briefly outlined.

Let $\mathrm{d}N_\tau(\nu)$ be the Lévy measure as in (8). Let Υ be the largest value of t for which we are interested in simulating the process. Also let

$$M(x) = -N_\Upsilon[x,\infty) = -\int_x^\infty \mathrm{d}N_\Upsilon(\nu) = -\int_0^\Upsilon \frac{\exp\{-x\beta(u)\}}{\beta(u)}\,\mathrm{d}\alpha(u).$$

Define the positive random variables $J_1 \geq J_2 \geq \cdots$ by $\mathrm{P}(J_1 \leq x_1) = \exp\{M(x_1)\}$,

$$\mathrm{P}(J_i \leq x_i | J_{i-1} = x_{i-1}) = \exp\{M(x_i) - M(x_{i-1})\}, \qquad x_i < x_{i-1}.$$

Here the J_i's are an ordering of those appearing in (5). If $-M(0) < \infty$ (which is true in our case), then the distribution of $J_i|[J_{i-1} = x_{i-1}]$ has mass $\exp\{M(0) - M(x_{i-1})\}$ at zero and is otherwise continuous. This means that we will have a finite number of non-zero J_i's in $[0,\Upsilon]$.

We can obtain the J_i via $\vartheta_i = -M(J_i)$ and $J_i = 0$ if $\vartheta_i > -M(0)$, where $\vartheta_1, \vartheta_2, \ldots$ are the jump times of a standard Poisson process at unit rate, that is $\vartheta_1, \vartheta_2 - \vartheta_1, \ldots$ are i.i.d. $\mathrm{Ga}(1,1)$. Then the Ferguson and Klass (1972) representation of the process is given by $L(t) = \sum_i J_i I\{U_i \leq n_t(J_i)\}$, where, $U_1, U_2, \ldots$ are i.i.d. $\mathrm{Un}(0,1)$ and

$$n_t(\nu) = \frac{\mathrm{d}N_t}{\mathrm{d}N_\Upsilon}(\nu) = \frac{\int_0^t \exp\{-\nu\beta(u)\}\,\mathrm{d}\alpha(u)}{\int_0^\Upsilon \exp\{-\nu\beta(u)\}\,\mathrm{d}\alpha(u)},$$

for $t \in [0,\Upsilon]$. This form of simulating the Lévy process is not difficult to implement.

To carry out posterior inference we need to implement a Gibbs sampler in the following way. Let $t' = (t_1, \ldots, t_n)$ be a random sample of size n such that, without loss of generality, t_i is uncensored for $i = 1, \ldots, n_u$ and right censored for $i = n_u + 1, \ldots, n$. For each t_i, $i = 1, \ldots, n_u$ let s_i be an auxiliary variable. The simulation from the joint posterior distribution of $L(\cdot)$ and s can be achieved through simulating from the full conditional distributions of $L(\cdot)$ given s written as $\Pi(L|s,t)$, and of s given $L(\cdot)$, which is sampling from $f(s|L,t)$ given in (9). The algorithm is as follows.

Let $M^{(0)} = \{\tau_1^{(0)}, \tau_2^{(0)}, \ldots\}$ be the set of prior fixed points of discontinuity and let us consider the distribution for the corresponding jumps to be $L\{\tau_l^{(0)}\} \sim \mathrm{Ga}(\alpha_{0_l}, \beta_{0_l})$. Initiate the algorithm by generating $s_i \sim \mathrm{Un}(0, t_i)$, $i = 1, \ldots, n_u$. and for $h = 1, \ldots, H$.

1. Generate $L^{(h)} \sim \Pi(L|s^{(h-1)}, t)$, with the following specifications:
   - the Lévy measure is given by

$$\mathrm{d}N'_\tau(\nu) = \mathrm{d}\nu \int_0^\tau \exp\{-\nu\beta'(u)\}\,\mathrm{d}\alpha(u),$$

   where $\beta'(u) = \beta(u) + \sum_{i=1}^n k(t_i, u)$;

- the set of fixed jumps $M^{(h)} = \{\tau_1^{(h)}, \ldots, \tau_m^{(h)}\}$ is formed by all different $\{s_i^{(h-1)}, \tau_l^{(0)}\}$, together with $r_j^{(h)}$, $j = 1, \ldots, m$, the number of $\{s_i^{(h-1)}, \tau_l^{(0)}\}$ equal to $\tau_j^{(h)}$;
- the distribution for the fixed jumps $L\{\tau_j^{(h)}\}$ is given by

$$f_j^{(h)}(x) \propto \begin{cases} \mathrm{Ga}\left(x|\alpha_{0_l} + r_j^{(h)} - 1, \beta_{0_l} + \sum_{i=1}^n k(t_i, \tau_j^{(h)})\right), \\ \qquad\qquad\qquad\qquad \text{if } \tau_j^{(h)} = \tau_l^{(0)}, l = 1, 2, \ldots, \\ \\ \mathrm{Ga}\left(x|2 + r_j^{(h)} - 1, \beta(s_i^{(h-1)}) + \sum_{i=1}^n k(t_i, \tau_j^{(h)})\right), \\ \qquad\qquad\qquad\qquad \text{if } \tau_j^{(h)} = s_i^{(h-1)}, i = 1, \ldots, n_u \end{cases}$$

  for $j = 1, \ldots, m$.

2. Generate $s_i^{(h)} \sim f(s_i|L^{(h)}, t)$, $i = 1, \ldots, n_u$, where

$$f(s_i|L^{(h)}, t) \propto e^{as_i} L^{(h)}\{s_i\} I(0 \le s_i \le t_i).$$

6. Numerical Examples

Example 1. The first example involves the well-known Kaplan and Meier (1958) data set. This data set has been used by many authors to compare their Bayes estimates with the Product-limit estimate (see, for example, Ferguson and Phadia (1979); Walker and Muliere (1997)). The data consists of observations measured in months: 0.8, 1.0*, 2.7*, 3.1, 5.4, 7.0*, 9.2, 12.1*, where $*$ denotes a right censored observation.

To centre the prior process we chose the constant hazard model, that is $\beta(t) = q$, $\mathrm{d}\alpha(t) = apq\,\mathrm{d}t$ and $L\{0\} \sim \mathrm{Ga}(p, q)$. We took $p = 0.05$ and $q = 0.1$ in order to place the prior away from the data. The only remaining parameter to be determined is a. We took $a = 0.001$ to introduce a high correlation in the prior process (see Section 4). The Gibbs sampler was run for 10,000 iterations with a burn-in of 1,000 taking the last 9,000 simulations to estimate the curves.

Prior and posterior hazard rate estimates are presented in the top graph of Figure 1. Here we can see that the prior hazard rate estimate is constant at 0.5 for all t and the posterior hazard rate estimate is also constant at 0.1 for all t. According to Walker and Damien (1998) the Kaplan-Meier data were generated from an exponential model with hazard rate 0.1. This means that our posterior estimate reproduces the true hazard. The Kaplan-Meier, the prior and posterior survival curve estimates, together with 95% predictive bands, are presented in the bottom graph of Figure 1. As can be seen from Figure 1, the posterior estimate follows the Kaplan-Meier path in a smooth way, as expected.

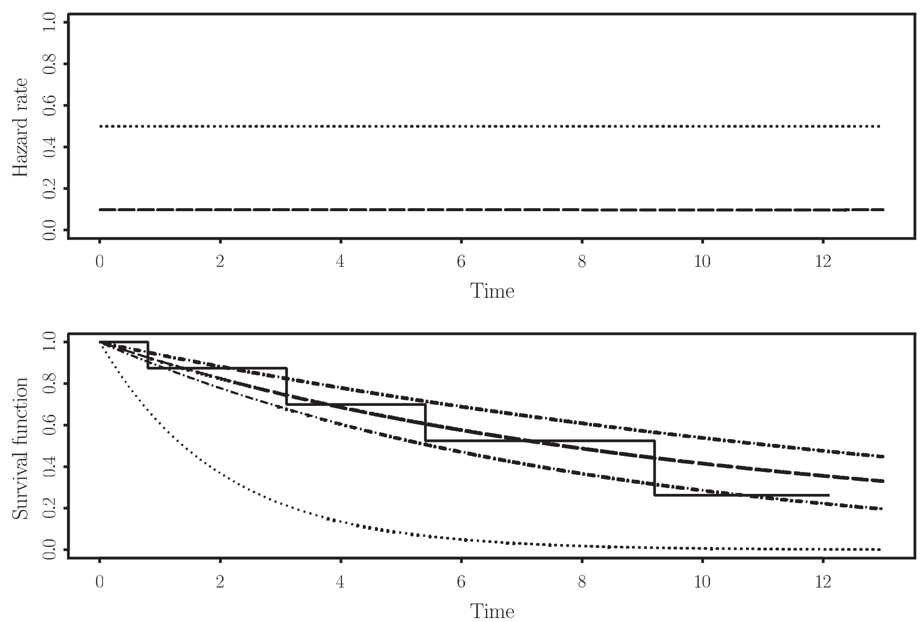

Figure 1. Hazard rate estimates (top) and survival curve estimates (bottom) for the K-M data. (——) Kaplan-Meier estimate, (····) prior estimate (----) posterior estimate and (-·-·) 95% bands.

Example 2. The second example involves a data set taken from Moreau, O'Quigley and Mesbah (1985). This data consists of survival times of gastric cancer patients and is divided into two treatment groups. Here we only consider the combined chemotherapy/radiation group with data:

17, 42, 44, 48, 60, 72, 94, 95, 103, 108, 122, 144, 167, 170, 183, 185, 193, 195, 197, 208, 234, 235, 254, 307, 315, 401, 445, 464, 484, 528, 542, 567, 577, 580, 795, 855, 882*, 892*, 1031*, 1033*, 1306*, 1335*, 1366, 1452*, 1472*,

where, as before, $*$ denotes a right censored observation.

In this example we employ a prior $\Pi(a)$ for a. The conditional likelihood function of a is not log-concave so we rely on a Metropolis-Hastings algorithm (Tierney (1994)) for sampling this conditional posterior distribution.

We centred the prior process on a constant hazard model and took $p = 0.0005$ and $q = 0.1$ to locate the prior away from the data. For illustrative purposes we took $\Pi(a) = \mathrm{Ga}(a|1, 2)$ as the prior distribution for a, i.e., a prior mean value for a of 0.5. The Gibbs-Metropolis algorithm was run for 10,000 iterations with a burn-in period of 1,000.

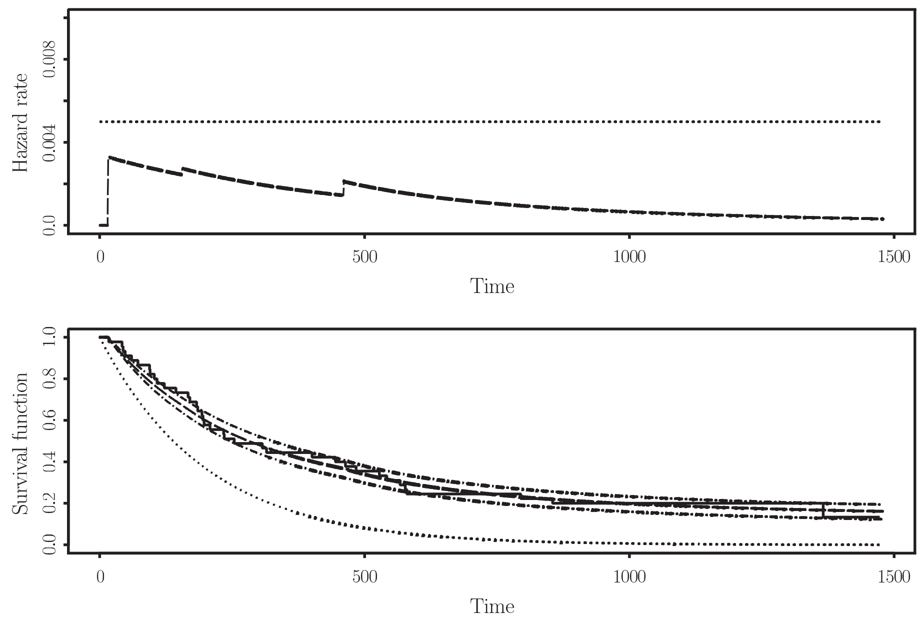

Figure 2. Hazard rate estimates (top) and survival curve estimates (bottom) for the gastric data. (——) Kaplan-Meier estimate, (····) prior estimate (----) posterior estimate and (-·-·) 95% bands.

Prior and posterior hazard rate estimates are presented in the top graph of Figure 2. In this graph we can see that the prior estimate is 0.005 for all t, whereas the posterior estimate is step-wise decreasing with exponential decaying shocks. Sudden changes in the hazard rate allow the survival curve to model changes in the rate of decay. The Kaplan-Meier, the prior and posterior survival curve estimates, together with 95% predictive bands, are plotted in the bottom graph of Figure 2.

From Figure 2 (bottom graph) we can see that the posterior estimate of the survival curve is quite smooth and is in agreement with the Kaplan-Meier estimate. The predictive bands are very tight, which means that the posterior variance is small. The posterior estimate for the parameter a is 0.0019.

7. Consistency

Substantial theory has recently been developed on Hellinger consistency (Barron, Shervish and Wasserman (1999), Ghosal, Ghosh and Ramamoorthi (1999)). In survival studies interest focuses on the survival function and hence our concern is with weak consistency, in the sense that the posterior distributions accumulate in all weak neighbourhoods of the true density function f_0, with corresponding distribution function F_0. Schwartz (1965) and Barron (1988) (see

Theorem 1 in Ghosal et al. (1999)) proved that, provided the prior puts positive mass in Kullback-Leibler neighborhoods of the true distribution function F_0, then the posterior concentrates for large samples on weak neighborhoods of F_0.

Let d_K denote the Kullback-Leibler divergence, that is

$$d_K(f,g) = \int \log\{f(t)/g(t)\}\, f(t)\,\mathrm{d}t.$$

Theorem 6. *Let Π be the probability measure governing $f(t) = x(t)\exp\{-Z(t)\}$, that is, the joint measure for L and a. Assuming that $f_0(t)$ satisfies*
(i) (a) $\int t\, f_0(t)\,\mathrm{d}t < \infty$ *and* (b) $\int \mathrm{E}\{Z(t)\}\, f_0(t)\,\mathrm{d}t < \infty$,
(ii) $\int f_0(t)\,\log\{f_0(t)\}\, dt < \infty$, *and*
(iii) $x_0(t) > 0$ *except possibly* $x_0(0) = 0$,
and that the corresponding hazard rate $x_0(t)$ of f_0 satisfies that $x_0(0) < \infty$ and there exists a non-decreasing function $G(\nu)$ defined on $[0,\infty)$ such that

$$x_0(t) = \int_0^t e^{-a_0(t-\nu)}\,\mathrm{d}G(\nu),$$

then the nonparametric prior Π is weak consistent.

Proof. We need to prove that $\Pi\{f : d_K(f_0, f) < \varepsilon\} > 0$ for all $\varepsilon > 0$. We start by showing that Π puts positive mass on $\{f : \int f_0 \log(f_0/f) < \infty\}$. To do this we need to consider $\int \log\{f(t)\}\, f_0(t)\,\mathrm{d}t$ and show that this is finite (in fact $> -\infty$) with positive probability. This combined with (ii) is sufficient for this. Now

$$\int \log\{f(t)\}\, f_0(t)\,\mathrm{d}t = \int \log\{x(t)\}\, f_0(t)\,\mathrm{d}t - \int Z(t)\, f_0(t)\,\mathrm{d}t$$

and the second term on the right is finite with probability one by assumption (i)(b). Observe that

$$\int \log\{x(t)\}\, f_0(t)\,\mathrm{d}t = \int \log\{L(t)\}\, f_0(t)\,\mathrm{d}t + \int \log\left\{\frac{\int_0^t e^{-a(t-s)}\mathrm{d}L(s)}{L(t)}\right\} f_0(t)\,\mathrm{d}t.$$

The first term here is $> -\infty$ with positive probability, that is,

$$\Pi\left\{f : \int \log\{L(t)\}\, f_0(t)\,\mathrm{d}t > -\infty\right\} > 0,$$

and the second term is greater than

$$-a \int \int_0^t \frac{(t-s)\,\mathrm{d}L(s)}{L(t)}\, f_0(t)\,\mathrm{d}t > -a \int t\, f_0(t)\,\mathrm{d}t$$

which is finite with probability one by assumption (i)(a).

Therefore, for any $\varepsilon > 0$, there exists an M such that

$$\Pi\left\{f : \int_M^\infty \log(f_0/f)\, f_0 < \varepsilon/2\right\} > 0.$$

To finish we need to show that there is positive probability for the event

$$\left\{f : \int_0^M \log(f_0/f)\, f_0 < \varepsilon/2\right\}$$

for all $\varepsilon > 0$ and all M. Now

$$\begin{aligned}&\int_0^M \log\{f_0(t)/f(t)\}\, f_0(t)\,\mathrm{d}t\\ &= \int_0^M \log\{x_0(t)/x(t)\}\, f_0(t)\,\mathrm{d}t + \int_0^M \int_0^t \{x(s) - x_0(s)\}\,\mathrm{d}s f_0(t)\,\mathrm{d}t.\end{aligned}$$

Remembering that x_0 is a mixture of exponentials, and since $x(t) = \int_0^t e^{-a(t-s)}\,\mathrm{d}L(s)$ and Π_L has full support on step functions with finite jumps in $[0, M]$ and $\Pi(a)$ has full support on $[0,\infty)$, it follows that Π puts positive probability on both

$$\left\{f : \sup_{t\in[\delta,M]} |\log\{x_0(t)\} - \log\{x(t)\}| < \varepsilon/4\right\},$$

$$\left\{f : \sup_{t\in[0,M]} |x_0(t) - x(t)| < \varepsilon/4\right\},$$

for any $\varepsilon > 0$ and $\delta \geq 0$. The $\delta \geq 0$ here is to cover the possibility that $x_0(0) = 0$. If $x_0(0) > 0$ then we can put $\delta = 0$. If $x_0(0) = 0$ then we can find a small enough δ so that Π puts positive mass on $\int_0^\delta \log(x_0/x)\, f_0$ being arbitrarily small. With (iii) we do not worry about this phenomenon for any other t. This then demonstrates that Π puts positive mass on all Kullback-Leibler neighbourhoods of f_0.

8. Discussion

The idea of using Lévy driven Markov processes with an exponential kernel for modeling hazard rate functions arose from an attempt to find the continuous time version of the discrete time gamma process used by Nieto-Barajas and Walker (2002). Let $x_n(t)$ be a piecewise constant process defined as

$$x_n(t) = \lambda_{n,0}\, I(t = 0) + \lambda_{n,k}\, I\left\{(k-1)/n < t \leq k/n\right\}, \quad k = 1, 2, \ldots,$$

where $\lambda_{n,0} \sim \mathrm{Ga}(\alpha, 1)$ and $\lambda_{n,k}$ is obtained from $\lambda_{n,k-1}$ via $\lambda_{n,k}|u_{n,k} \sim \mathrm{Ga}(\alpha + u_{n,k}, 1 + c_n)$ and $u_{n,k}|\lambda_{n,k-1} \sim \mathrm{Po}(c_n\, \lambda_{n,k-1})$. Then for all n, $x_n(t)$ is strictly stationary with marginals $\mathrm{Ga}(\alpha, 1)$.

It can be seen that $\lambda_{n,k}$ can be expressed as

$$\lambda_{n,k} = \gamma_{n,k}/(1 + c_n) + \mathrm{Ga}\{\mathrm{Po}(c_n\lambda_{n,k-1}), 1\}/(1 + c_n),$$

where $\gamma_{n,k}$ are i.i.d. $\mathrm{Ga}(\alpha, 1)$. If $\lambda_{n,k-1}$ is $\mathrm{Ga}(\alpha, 1)$ then so is $\lambda_{n,k}$. Consequently, we can switch $\gamma_{n,k}$ and $\lambda_{n,k-1}$ leaving $\lambda_{n,k}$ also marginally $\mathrm{Ga}(\alpha, 1)$. We need to keep $\mathrm{E}(\lambda_k|\lambda_{n,k-1}) = (\alpha + c_n\lambda_{n,k-1})/(1 + c_n)$, so the new stationary model becomes

$$\lambda_{nk} = c_n\lambda_{n,k-1}/(1 + c_n) + c_n\xi_{n,k}/(1 + c_n),$$

where $\xi_{n,k} = \mathrm{Ga}\{\mathrm{Po}(\gamma_{n,k}/c_n), 1\}$, which is in the form of a stochastic difference equation of the kind considered by Wolfe (1982). Then, it can be proven that the piece-wise constant process $x_n(t)$ defined with the new $\lambda_{n,k}$ converges in distribution to the Markov gamma process $x(t)$ defined in (3), as $n \to \infty$.

In Section 3 we presented the posterior conditional distributions needed to achieve a full Bayesian analysis via Gibbs sampling. However, it is not straightforward to assume that the algorithm in fact converges to the true posterior distribution. Athreya, Doss and Sethuraman (1996) presented some theorems which guarantee convergence of the MCMC simulation method in general state spaces. In our case, the MCMC simulation process involves simulating from a stochastic process; nevertheless, it satisfies the ρ-irreducibility and aperiodicity conditions required to ensure convergence to the posterior distribution $\Pi(L|\mathrm{data})$.

Acknowledgements

The research of the first author was supported by grant I39357-E from the National Council for Science and Technology of Mexico (CONACYT), and an Overseas Research Studentship. The work of the second author was supported by an EPSRC Advanced Research Fellowship. The authors are grateful for the comments of referees.

References

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data using the Gibbs sampler. *Statist. Sinica* **4**, 505-524.

Athreya, K. B., Doss, H. and Sethuraman, J. (1996). On the convergence of the Markov chain simulation method. *Ann. Statist.* **24**, 69-100.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Orstein-Uhlenbeck based models and some of their uses in financial economics. *J. Roy. Statist. Soc. Ser. B* **63**, 167-241.

Barron, A. R. (1988). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical Report 7. Department of Statistics, University of Illinois.

Barron, A., Shervish, M. J. and Wasserman, L. (1999). Consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27**, 536-561.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Probab.* **31**, 929-953.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-202.

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

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

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

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

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7**, 163-186.

Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27**, 143-158.

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

James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. Available at http://arXiv.org/abs/math/0205093.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27**, 562-588.

Laud, P. W., Damien, P. and Walker, S. G. (1999). Bayesian analysis of survival models with bathtub hazard rates. Technical Report 31. Division of Biostatistics Medical College of Wisconsin.

Lo, A. Y. and Weng, C-S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41**, 227-245.

Moreau, T., O'Quigley, J. and Mesbah, M. (1985). A global goodness of fit statistics for the proportional hazards model. *Appl. Statist.* **34**, 212-218.

Nieto-Barajas, L. E. and Walker, S. G. (2002). Markov beta and gamma processes for modeling hazard rates. *Scand. J. Statist.* **29**, 413-424.

Ross, S. M. (1996). *Stochastic Processes.* Second edition. Wiley, New York.

Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10-26.

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. Ser. B* **55**, 3-23.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22**, 1701-1722.

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

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

Walker, S. G., Damien P., Laud, P. W. and Smith, A. F. M. (1999). Bayesian nonparametrics inference for random distributions and related functions. *J. Roy. Statist. Soc. Ser. B* **61**, 485-527.

Wolfe, S. J. (1982). On a stochastic analogue of the stochastic difference equation $X_n = \rho X_{n-1} + B_n$. *Stochastic Process. Appl.* **12**, 301-312.

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

Departamento de Estadística, ITAM, Río Hondo 1, Tizapán San Angel, 01000 México, D.F., México.

E-mail: lnieto@itam.mx

Institute of Mathematics, Statistics and Actuarial Science, University of Kent, CT2 7NZ Canterbury, Kent, UK.

E-mail: S.G.Walker@Kent.ac.uk

(Received October 2002; accepted January 2004)