A nonparametric Bayesian approach to inference for non-homogeneous Poisson processes

Athanasios Kottas [1]

Department of Applied Mathematics and Statistics, Baskin School of Engineering,

University of California, 1156 High Street, Santa Cruz, CA 95064, USA

(REVISED VERSION – August 23, 2006)

Abstract

We propose modeling for Poisson processes over time, exploiting the connection of the Poisson process intensity with a density function. Nonparametric mixture models for this density induce flexible prior models for the intensity function. We work with Beta densities for the mixture kernel and a Dirichlet process prior for the mixing distribution. We also discuss modeling for monotone intensity functions through scale uniform mixtures. Simulation-based model fitting enables posterior inference for any feature of the Poisson process that might be of interest. A data example illustrates the methodology.

Keywords: Beta mixtures; Dirichlet process; Markov chain Monte Carlo; Point processes; Scale uniform mixtures

---

[1] Tel.: 831-459-5536; Fax: 831-459-4829; *E-mail address*: thanos@ams.ucsc.edu

This work was supported in part by NSF grant DMS-0505085.

1. Introduction

Poisson processes play a fundamental role in the theory and applications of point processes (see, e.g., Kingman, 1993; Møller and Waagepetersen, 2004). From a modeling perspective, of interest for a non-homogeneous Poisson process (NHPP) over time is its intensity function λ, a non-negative and locally integrable function, that is, $\int_B \lambda(u)du < \infty$, for all bounded B. The mean measure (or cumulative intensity function) of the process is given by $\Lambda(t) = \int_0^t \lambda(u)du$, $t \in R^+$. Formally, a point process over time, $\mathcal{Y} = \{Y(t) : t \geq 0\}$, is a NHPP if $\mathcal{Y}$ has independent increments and, for any $t > s \geq 0$, $Y(t) - Y(s)$ follows a Poisson distribution with mean $\Lambda(t) - \Lambda(s)$.

We propose a Bayesian nonparametric modeling approach for NHPPs. The method is based on the direct connection of the intensity function with an associated density function. To model the density function, we employ flexible nonparametric mixtures of Beta densities. The resulting nonparametric prior for the intensity function enables model-based, data-driven inference for non-standard intensity shapes and allows quantification of the associated uncertainty. Moreover, we develop prior models that incorporate monotonicity restrictions for the intensity function, using scale mixtures of uniform densities.

The plan of the paper is as follows. Section 2 presents the methodology for intensity functions with general shapes, including details for posterior inference and prior specification. A data illustration is provided in Section 3. Section 4 discusses alternative model formulations for monotone intensity functions, and Section 5 concludes with a discussion.

2. The modeling approach

2.1. The probability model

We consider a NHPP observed over the time interval $(0, T]$ with events that occur at times

$0 < t_1 \leq t_2 \leq ... \leq t_n \leq T$. The likelihood for its intensity function λ is given by

$$\exp\{-\int_0^T \lambda(u)du\} \prod_{i=1}^{n} \lambda(t_i). \tag{1}$$

Let $\gamma = \int_0^T \lambda(u)du$. To cast the problem in a density estimation framework, the key observation is that $f(t) = \lambda(t)/\gamma$, $t \in (0, T]$, is a density function on $(0, T]$. Hence, since (f, γ) provides an equivalent representation for λ, a nonparametric prior model for f, with a parametric prior for γ, will induce a semiparametric prior for λ. A flexible specification for f arises through a Dirichlet process (DP) mixture, $f(t; G) = \int k(t; \boldsymbol{\theta})\mathrm{d}G(\boldsymbol{\theta})$. Here, $k(t; \boldsymbol{\theta})$ is a parametric kernel (with parameter $\boldsymbol{\theta} \in \Theta \subseteq R^d$, $d \geq 1$) supported by $(0, T]$, and the random mixing distribution G is assigned a DP prior (Ferguson, 1973; Antoniak, 1974) with precision parameter α and base distribution G_0, denoted by $G \sim \mathrm{DP}(\alpha G_0)$. In this context, the choice of the DP allows us to draw from the existing theory for this prior, and to utilize well-established techniques for simulation-based model fitting. Note that we are creating a prior model for random intensity functions induced by the prior model for the associated random density functions. In fact, since γ only scales λ, it is f that determines the shape of the intensity function λ, and thus a flexible model for f will capture non-standard shapes in λ.

To allow modeling of general density, and thus also intensity, shapes on $(0, T]$, we employ a Beta distribution for the kernel of the mixture $f(t; G)$. (In Section 4 we develop different mixture model formulations for applications where it is of interest to model monotone intensity functions.) Mixtures of Beta densities yield a wide range of distributional shapes, in fact, they can be used to approximate arbitrarily well any density defined on a bounded interval (Diaconis and Ylvisaker, 1985). We parameterize the rescaled Beta distribution, with support on $(0, T)$, in terms of its mean $\mu \in (0, T)$ and a scale parameter

$\tau > 0$. Specifically, letting $\boldsymbol{\theta} = (\mu, \tau)$,

$$k(t; \mu, \tau) = \frac{1}{\text{Be}\{\mu\tau T^{-1}, \tau(1-\mu T^{-1})\}T^{\tau-1}} t^{\mu\tau T^{-1}-1}(T-t)^{\tau(1-\mu T^{-1})-1}, \; t \in (0, T), \quad (2)$$

where $\text{Be}(a, b)$ denotes the Beta function, $\int_0^1 u^{a-1}(1-u)^{b-1}du$, $a > 0$, $b > 0$.

Hence the mixture model for the random density f is given by

$$f(t; G) = \int k(t; \mu, \tau)\text{d}G(\mu, \tau), \quad G \sim \text{DP}(\alpha G_0). \quad (3)$$

We assume random α with a gamma(a_α, b_α) prior distribution $p(\alpha)$ such that $\text{E}(\alpha) = a_\alpha/b_\alpha$. To specify the base distribution G_0, we assume independent components, $G_0(\mu, \tau) = G_{01}(\mu)G_{02}(\tau)$, and note that the variance under (2) is $\mu(T-\mu)/(\tau+1)$. Hence μ determines the location of a mixture component and, for specified μ, τ controls its dispersion. The default choice of a uniform distribution on $(0, T)$ for $G_{01}(\mu)$ is appealing and, in fact, proves to be sufficiently flexible in applications. For $G_{02}(\tau)$ we take an inverse gamma distribution with fixed shape parameter a_τ and random scale parameter β, which is assigned an exponential prior distribution $p(\beta)$ with mean $1/d$.

Denote by $\boldsymbol{\theta}_i = (\mu_i, \tau_i)$ the latent mixing parameter associated with t_i. The discrete countable nature of the DP is a key feature as it enables data-driven clustering in the $\boldsymbol{\theta}_i$. The discreteness for the DP is immediate from its constructive definition (Sethuraman, 1994), according to which, a realization G, given α and β, is (almost surely) of the form

$$G = \sum_{j=1}^{\infty} \omega_j \delta_{(\tilde{\mu}_j, \tilde{\tau}_j)}, \quad (4)$$

where δ_y denotes a point mass at y, $\omega_1 = z_1$, $\omega_j = z_j \prod_{s=1}^{j-1}(1-z_s)$, $j = 2,3,...$, with $z_s \mid \alpha$ independent from a Beta$(1, \alpha)$ distribution, and, independently, $(\tilde{\mu}_j, \tilde{\tau}_j) \mid \beta$ independent from G_0.

Using (1), the full Bayesian model for γ and f, equivalently, for γ and G, becomes

$$\exp(-\gamma)\gamma^n \left\{ \prod_{i=1}^{n} \int k(t_i; \mu_i, \tau_i)\text{d}G(\mu_i, \tau_i) \right\} p(\gamma)p(G \mid \alpha, \beta)p(\alpha)p(\beta) \quad (5)$$

with the prior structure $p(G \mid \alpha,\beta)p(\alpha)p(\beta)$ for G, and its hyperparameters, discussed above, and a prior $p(\gamma)$, with support $(0,\infty)$, for γ.

Because it seems difficult to specify parameters for a prior distribution for γ, we use the reference prior (e.g., Bernardo, 2005). To obtain it, we work with the marginal likelihood, $\mathcal{L}^*(\gamma)$, for γ, which arises from (5) by integrating out all other parameters over their (proper) priors. Specifically, $\log \mathcal{L}^*(\gamma) \propto -\gamma + n\log\gamma$, and, hence, the Fisher's information based on this marginalized likelihood yields $p(\gamma) \propto \gamma^{-1}$ as the reference prior.

2.2. Posterior inference

Based on the model structure in (5), and under the $p(\gamma) \propto \gamma^{-1}$ prior for γ, it is straightforward to verify that the joint posterior, $p(\gamma, G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$, is proper. Here, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ and data $= (t_1, ..., t_n)$. In fact, the marginal posterior $p(\gamma|\text{data}) = \text{gamma}(n, 1)$, and $p(\gamma, G, \boldsymbol{\theta}, \alpha, \beta|\text{data}) = p(\gamma|\text{data})\ p(G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$. Hence, to explore the full posterior distribution $p(\gamma, G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$, it suffices to implement a Markov chain Monte Carlo method to obtain draws from $p(G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$, the posterior for the DP mixture part of model (5). To this end, we follow the approach proposed in Gelfand and Kottas (2002).

Using results from Antoniak (1974), $p(G, \boldsymbol{\theta}, \alpha, \beta|\text{data}) = p(G \mid \boldsymbol{\theta}, \alpha, \beta)p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$, where the distribution for $G \mid \boldsymbol{\theta}, \alpha, \beta$ is a DP, with updated precision parameter $\alpha + n$ and base distribution $G_0^*(\mu, \tau|\boldsymbol{\theta}, \alpha, \beta) = (\alpha+n)^{-1}\{\alpha G_0(\mu, \tau|\beta) + \sum_{i=1}^n \delta_{(\mu_i,\tau_i)}(\mu, \tau)\}$, and $p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$ is the posterior that results after marginalizing G over its DP prior. Several posterior simulation methods have been suggested for this marginal posterior; see, e.g., the review in Müller and Quintana (2004). We have used algorithm 5 from Neal (2000) to obtain posterior samples from $p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$. Next, posterior draws $\boldsymbol{\theta}_b = \{(\mu_{ib}, \tau_{ib}) : i = 1, ..., n\}$, α_b, β_b from $p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$ can be used to draw G_b from $p(G \mid \boldsymbol{\theta}_b, \alpha_b, \beta_b)$ using a truncation approximation to (4). Specifically, we take $G_b =$

$\sum_{j=1}^{J} w_{jb}\delta_{(\mu'_{jb},\tau'_{jb})}$, where $w_{1b} = z_{1b}$, $w_{jb} = z_{jb}\prod_{s=1}^{j-1}(1 - z_{sb})$, $j = 2, ..., J - 1$, $w_{Jb} = 1 - \sum_{j=1}^{J-1} w_{jb} = \prod_{s=1}^{J-1}(1 - z_{sb})$, with z_{sb} independent Beta$(1, \alpha_b + n)$, and, independently, (μ'_{jb}, τ'_{jb}) independent $G_0^*(\mu, \tau \mid \boldsymbol{\theta}_b, \alpha_b, \beta_b)$. The approximation can be made arbitrarily accurate. For instance, because $\mathrm{E}(\sum_{j=1}^{J-1} w_{jb} \mid \alpha_b) = 1 - \{(\alpha_b + n)/(\alpha_b + n + 1)\}^{J-1}$, we can choose J that makes, say, $\{(n + \max_b \alpha_b)/(n + 1 + \max_b \alpha_b)\}^{J-1}$ arbitrarily small.

Now, $f_{b0} = \int k(t_0; \mu, \tau)\mathrm{d}G_b(\mu, \tau) = \sum_{j=1}^{J} w_{jb}k(t_0; \mu'_{jb}, \tau'_{jb})$ is a realization from the posterior of $f(t_0; G)$, for any time point t_0 in $(0, T)$. Hence, if γ_b is a draw from $p(\gamma|\text{data})$, $\gamma_b f_{b0}$ is a posterior draw for $\lambda(t_0; \gamma, G) = \gamma f(t_0; G)$, the intensity function functional at t_0. Analogously, $F_{b0} = \sum_{j=1}^{J} w_{jb}K(t_0; \mu'_{jb}, \tau'_{jb})$, where K is the distribution function for the density k in (2), is a posterior realization for $F(t_0; G) = \int_0^{t_0} f(u; G)du = \int K(t_0; \mu, \tau)\mathrm{d}G(\mu, \tau)$, and $\gamma_b F_{b0}$ is a draw from the posterior of the cumulative intensity function functional at t_0, $\Lambda(t_0; \gamma, G) = \int_0^{t_0} \lambda(u; \gamma, G)du = \gamma F(t_0; G)$. Hence full posterior inference for the intensity and the cumulative intensity functions at any point in the time interval $(0, T)$ is available. For instance, posterior point estimates and associated uncertainty bands for λ and Λ can be obtained using point and interval estimates from $p\{\lambda(t_0; \gamma, G) \mid \text{data}\}$ and $p\{\Lambda(t_0; \gamma, G) \mid \text{data}\}$ over a grid of time points t_0.

2.3. Prior specification

Regarding α, recall that this parameter of the DP prior controls the number n^* (where $n^* \leq n$) of distinct components in the DP mixture (3) (Antoniak, 1974; Escobar and West, 1995). For instance, for moderately large n, it can be shown that $\mathrm{E}(n^*) \approx (a_\alpha/b_\alpha)\log\{1 + (nb_\alpha/a_\alpha)\}$.

To specify the mean, $1/d$, of the exponential prior for β, we consider a single component of mixture (3), for which the variance is $\mu(T - \mu)/(\tau + 1)$. Setting $a_\tau = 2$, which yields an inverse gamma distribution $G_{02}(\tau)$ with infinite variance, and using marginal prior means for μ and τ, based on G_0, a rough estimate for the variance above is $0.25T^2/(1+d^{-1})$. Let r be a prior guess at the range of the t_i; $r = T$ is a natural default choice. Then we specify d through $0.25T^2/(1 + d^{-1}) = (r/6)^2$. This approach is fairly automatic and, in fact, yields a noninformative specification as it is based on the special case of the mixture with a single component, whereas, in applications, more components are needed to capture the intensity function shape.

3. Data example

To illustrate the methodology, we consider a standard data set from the literature, the coal-mining disasters data, as compiled by Jarrett (1979). The data are the times of 191 explosions in mines, leading to coal-mining disasters involving 10 or more men killed, over a total time period of 40,550 days, from 15 March 1851 to 22 March 1962.

We employ the Beta DP mixture model (5) to obtain inference for the intensity of coal-mining disasters. We have conducted prior sensitivity analysis, considering several combinations of gamma(a_α, b_α) priors for α and exponential priors for β (with mean $1/d$), which revealed robustness for posterior results. To illustrate, Figure 1 shows pointwise posterior means and 95% central posterior intervals for the intensity function under three prior choices. Specifically, with $a_\alpha = 2$ in all cases, the posterior estimates correspond to priors with $b_\alpha = 1.4$, $d = 0.125$; $b_\alpha = 0.52$, $d = 0.125$; and $b_\alpha = 1.4$, $d = 0.333$, based on $r = T$, $\mathrm{E}(n^*) \approx 7$; $r = T$, $\mathrm{E}(n^*) \approx 15$; and $r = 1.5T$, $\mathrm{E}(n^*) \approx 7$, respectively. Posterior point estimation under model (5) and kernel estimation based on certain bandwidth selection methods, as discussed in, e.g., Diggle and Marron (1988), reveal comparable shapes for the intensity function. However, the Bayesian model yields more general inference than point estimation, including the interval estimates reported in Figure 1.

4. Modeling for monotone intensity functions

The focus so far has been on modeling general shapes for NHPP intensities. To this end, we have demonstrated the utility of a mixture formulation for the intensity function, using DP mixtures of Beta densities. However, in certain applications, such as software reliability (see, e.g., Kuo and Yang, 1996), it might be of interest to place monotonicity restrictions on the shape of the intensity function.

Here, we show how such restrictions can be incorporated in the prior model for $f(\cdot)$, and, thus, for $\lambda(\cdot)$, retaining the DP mixture framework of Section 2, albeit with different choices for the mixture kernel. Without loss of generality, we take $T = 1$ and, thus, present models for non-increasing and non-decreasing densities on the unit interval.

The key result is a representation for non-increasing densities on R^+. Specifically, for any non-increasing density $h(\cdot)$ on R^+ there exists a distribution function G, with support on R^+, such that $h(t) \equiv h(t;G) = \int \theta^{-1} 1_{(0,\theta)}(t) \mathrm{d}G(\theta)$, i.e., $h(\cdot)$ can be expressed as a scale mixture of uniform densities. The result involves a general mixing distribution G and thus, for Bayesian modeling, invites the use of a nonparametric prior for G; see, e.g., Brunner and Lo (1989), Brunner (1995), and Kottas and Gelfand (2001) for applications of this representation, which utilize DP priors.

To construct non-increasing densities on the unit interval, we use the representation discussed above, restricting the support of the mixing distribution on $(0,1)$. Therefore, the model for the intensity function becomes $\lambda(\cdot) \equiv \lambda(t;\gamma,G) = \gamma f_*(t;G)$, where $\gamma = \int_0^1 \lambda(u) du$, and $f_*(t;G) = \int_0^1 \theta^{-1} 1_{(0,\theta)}(t) \mathrm{d}G(\theta)$, for $t \in (0,1]$, with G a distribution function on $(0,1)$. A DP prior, $\mathrm{DP}(\alpha G_0)$, is assigned to G, where now G_0 is a distribution on $(0,1)$, say, a Beta distribution. It is straightforward to verify that $f_*(t;G)$ is a non-increasing density on $(0,1]$. Hence, the resulting mixture formulation for $\lambda(t;\gamma,G)$, with the DP prior for G and a prior for γ, defines a prior model for non-increasing intensities.

Turning to a model for non-decreasing intensities, consider the mixture density given by $\int_0^1 \theta^{-1} 1_{(-\theta,0)}(x) \mathrm{d}G(\theta)$, $x \in (-1, 0]$, where, again, G is a distribution function on $(0, 1)$. This is a non-decreasing density on $(-1, 0]$, and, thus, $f^*(t; G) = \int_0^1 \theta^{-1} 1_{(-\theta,0)}(t-1) \mathrm{d}G(\theta)$, for $t \in (0, 1]$, yields non-decreasing densities on $(0, 1]$. A DP mixture model for $f^*(t; G)$ arises by placing a $\mathrm{DP}(\alpha G_0)$ prior on G, with a Beta distribution for G_0. Finally, the prior model for non-decreasing intensities emerges through $\lambda(t; \gamma, G) = \gamma f^*(t; G)$, and is completed with a prior for γ. Under both modeling scenarios presented above, posterior inference proceeds as discussed in Section 2.2 for the more general Beta DP mixture model.

5. Discussion

We have proposed a modeling approach for NHPP intensity functions employing Beta DP mixtures for the associated density functions. The method yields flexible data-driven inference for the intensity function as well as for any functional of the Poisson process that might be of interest. We have discussed how such inferences can be obtained, illustrating with a data set from the literature. Finally, we have developed alternative model formulations, based on DP mixtures of uniform densities, for monotone intensity functions.

Although this paper is focused on inference for NHPPs, the nonparametric mixtures developed in Section 2.1 and Section 4 can also be utilized for density estimation on bounded intervals; see Kottas (2006) for details as well as for an additional data illustration involving a different temporal point pattern than the one discussed in Section 3.

Regarding the existing literature on intensity function estimation, classical approaches build on kernel density estimation ideas; see, e.g., Diggle (1985) and Diggle and Marron (1988). Hence, they rely on specification of a smoothing parameter (the bandwidth), corrections for boundary effects, and asymptotic arguments to provide tolerance bands for the intensity function point estimates. The proposed Bayesian mixture approach, which yields full and exact model-based inference, might be a useful alternative.

Bayesian nonparametric work has focused mainly on the mean measure Λ, including priors based on gamma, Beta, and Lévy processes; see Lo (1992), Kuo and Ghosh (1997), Gutiérrez-Peña and Nieto-Barajas (2003) and further references therein. Potential drawbacks in working with Λ might include the lack of smoothness in the resulting posterior estimates, induced by properties of the stochastic processes used as priors, and the fact that inference for λ is typically not readily available. Regarding prior models for the intensity function, the existing work includes the method suggested by Lo and Weng (1989), which was recently extended in Ishwaran and James (2004). Under this approach, $\lambda(t; H) = \int m(t; v)H(dv)$, $t \in (0, T]$, where m is a specified non-negative kernel (typically, not a density) with parameters v, and the mixing measure H is assigned a weighted gamma process prior. A similar formulation arises under the approach of Wolpert and Ickstadt (1998) applied to one-dimensional NHPPs. The proposed DP mixture modeling approach might be a useful addition to the existing methods as it builds on a familiar Bayesian density estimation framework, facilitating prior choice and posterior simulation.

An extension of the methodology developed in this paper to modeling for spatial NHPP intensities has been recently proposed by Kottas and Sansó (2006). A practically important extension of the model (for NHPPs either over time or over space) would be to semiparametric regression settings for data that include individual-specific covariates, i.e., for point patterns that can be assumed to arise from a marked NHPP. We will report on this work in a future manuscript.

References

Antoniak, C.E. (1974), Mixtures of Dirichlet processes with applications to nonparametric problems, *Ann. Statist.* **2**, 1152-1174.

Bernardo, J.M. (2005), Reference analysis, in: Handbook of Statistics 25, D.K. Dey and C.R. Rao, eds. (Amsterdam: Elsevier) pp. 17-90.

Brunner, L.J. (1995), Bayesian linear regression with error terms that have symmetric unimodal densities, *J. Nonparametr. Stat.* **4**, 335-348.

Brunner, L.J., Lo, A.Y. (1989), Bayes methods for a symmetric unimodal density and its mode, *Ann. Statist.* **17**, 1550-1566.

Diaconis, P., Ylvisaker, D. (1985), Quantifying prior opinion, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds. *Bayesian statistics 2* (North-Holland, Amsterdam).

Diggle, P. (1985), A kernel method for smoothing point process data, *Appl. Statist.* **34**, 138-147.

Diggle, P., Marron, J.S. (1988), Equivalence of smoothing parameter selectors in density and intensity estimation, *J. Amer. Statist. Assoc.* **83**, 793-800.

Escobar, M., West, M. (1995), Bayesian density estimation and inference using mixtures, *J. Amer. Statist. Assoc.* **90**, 577-588.

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

Gelfand, A.E., Kottas, A. (2002), A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models, *J. Comput. Graph. Statist.* **11**, 289-305.

Gutiérrez-Peña, E., Nieto-Barajas, L.E. (2003), Bayesian nonparametric inference for mixed Poisson processes, in: J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds. *Bayesian Statistics 7* (University Press, Oxford) pp. 163-179.

Ishwaran, H., James, L.F. (2004), Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data, *J. Amer. Statist. Assoc.* **99**, 175-190.

Jarrett, R.G. (1979), A note on the intervals between coal-mining disasters, *Biometrika* **66**, 191-193.

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

Kottas, A. (2006), Dirichlet process mixtures of Beta distributions, with applications to density and intensity estimation, *Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods*, 23rd International Conference on Machine Learning, Pittsburgh, PA. [Available at `http://www.ams.ucsc.edu/reports/trview.php`]

Kottas, A., Gelfand, A.E. (2001), Bayesian semiparametric median regression modeling, *J. Amer. Statist. Assoc.* **96**, 1458-1468.

Kottas, A., Sansó, B. (2006), Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis, *J. Statist. Plann. Inference* (to appear).

Kuo, L., Ghosh, S.K. (1997), Bayesian nonparametric inference for nonhomogeneous Poisson processes, Technical Report, Department of Statistics, University of Connecticut.

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

Lo, A.Y. (1992), Bayesian inference for Poisson process models with censored data, *J. Nonparametr. Stat.* **2**, 71-80.

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

Moller, J., Waagepetersen, R.P. (2004), *Statistical Inference and Simulation for Spatial Point Processes* (Chapman & Hall, Boca Raton).

Müller, P., Quintana, F.A. (2004), Nonparametric Bayesian data analysis, *Statist. Science* **19**, 95-110.

Neal, R.M. (2000), Markov chain sampling methods for Dirichlet process mixture models, *J. Comput. Graph. Statist.* **9**, 249-265.

Sethuraman, J. (1994), A constructive definition of Dirichlet priors, *Statist. Sinica* **4**, 639-650.

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

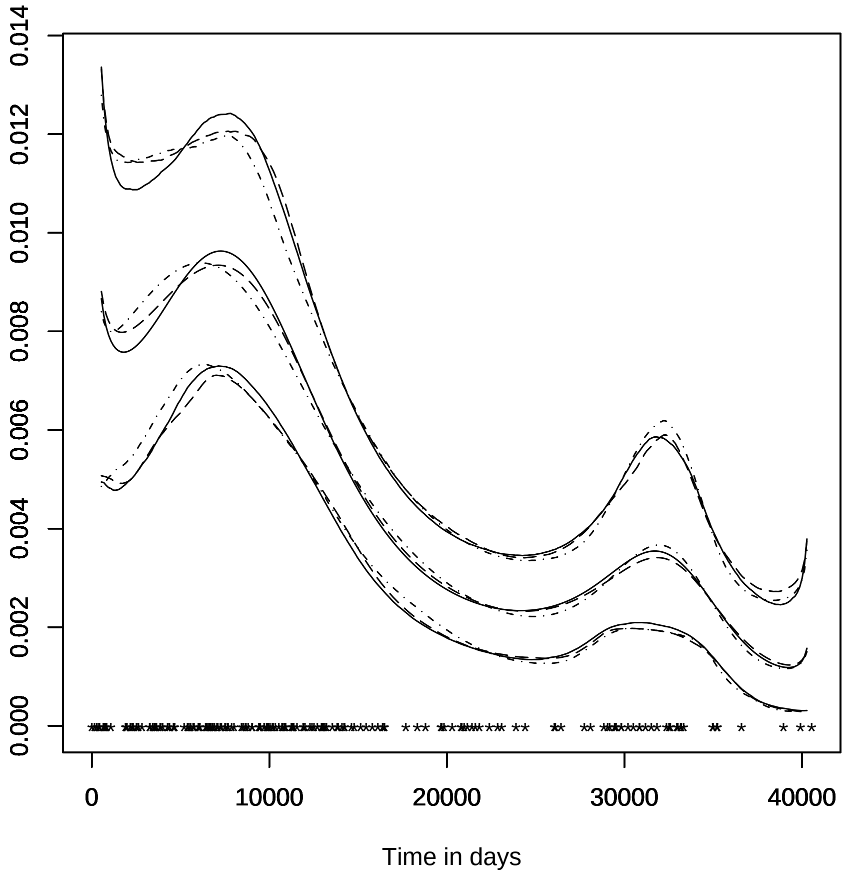

Figure 1: Posterior point estimates and 95% pointwise interval estimates for the intensity function under three prior settings. The dashed lines correspond to the $b_\alpha = 1.4$, $d = 0.125$ prior choice; the solid lines to $b_\alpha = 0.52$, $d = 0.125$; and the dashed-dotted lines to $b_\alpha = 1.4$, $d = 0.333$. The observed times of the 191 explosions in mines are plotted on the horizontal axis.