Dirichlet Process Mixtures of Beta Distributions, with Applications to Density and Intensity Estimation

Athanasios Kottas THANOS@AMS.UCSC.EDU
Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064 USA

Abstract

We propose a class of Bayesian nonparametric mixture models with a Beta distribution providing the mixture kernel and a Dirichlet process prior assigned to the mixing distribution. Motivating applications include density estimation on bounded domains, and inference for non-homogeneous Poisson processes over time. We present the mixture model formulation, discuss prior specification, and develop a computational approach to posterior inference. The model is illustrated with two data sets.

1. Introduction

In looking beyond standard parametric families one is naturally led to mixture models. In particular, finite mixture distributions provide flexible modeling and are now feasible to implement due to advances in simulation-based model fitting. Though it may appear paradoxical, rather than handling the large number of parameters resulting from finite mixture models with a large number of mixands, it may be easier to work with an infinite dimensional specification by assuming a random mixing distribution which is not restricted to a specified parametric family. The Dirichlet process (DP) (Ferguson, 1973) is the most widely used in this context, following Antoniak (1974), Lo (1984) and Ferguson (1983). Under the DP mixture setting, the generic form of the random mixture density is given by $f(\cdot; G) = \int k(\cdot; \boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta})$, where $k(\cdot; \boldsymbol{\theta})$ is a parametric kernel (with parameter vector $\boldsymbol{\theta}$), and the random mixing distribution G is assigned a DP prior. The DP prior can be specified in terms of two hyperparameters, a precision parameter $\alpha > 0$, and a base distribution G_0, and, will thus be denoted by $G \sim \mathrm{DP}(\alpha, G_0)$.

Appearing in *Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods*, *23rd* ICML, Pittsburgh, PA, 2006.

Within the field of Bayesian nonparametrics, the class of DP mixture models has witnessed several applications (see, e.g., the reviews by MacEachern & Müller, 2000, and Müller & Quintana, 2004). This is, arguably, due to the availability of posterior simulation techniques (mainly Markov chain Monte Carlo, MCMC, algorithms) for DP mixtures, following the work of Escobar (1994) and Escobar & West (1995). Moreover, more recent work, has been focusing on extensions of the DP, including development of priors with more flexible structure in their hyperparameters (e.g., Ishwaran & James, 2001; Lijoi, Mena & Prünster, 2005), as well as prior models for dependent distributions (e.g., MacEachern, 2000; De Iorio et al., 2004; Gelfand, Kottas & MacEachern, 2005; Griffin & Steel, 2006; Teh et al., 2006).

Most of the work with DP mixtures is based on normal kernels, either univariate or multivariate. However, there are several applications where different choices for the kernel are more natural/appropriate. Relevant examples include models for unimodal densities on the real line (e.g., Brunner & Lo, 1989; Brunner, 1995; Kottas & Gelfand, 2001; Kottas & Krnjajić, 2005) and models for survival and reliability analysis (e.g., Merrick, Soyer & Mazzuchi, 2003; Kottas, 2006; Hanson, 2006).

In this paper we study the utility of DP mixtures of Beta distributions. This class of DP mixture models emerges as a natural candidate for density estimation problems when the support of the distribution is restricted on a bounded interval. Moreover, we consider applications of the mixture model to inference for non-homogeneous Poisson processes over time. The method is based on the direct connection of the Poisson process intensity function with an associated density function. To model the density function, we employ the Beta DP mixture model. The resulting nonparametric prior for the intensity function enables model-based, data-driven inference for non-standard intensity shapes.

The plan of the paper is as follows. Section 2 presents the formulation for the Beta DP mixture model, including discussion of prior specification and posterior predictive inference. In Section 3 the mixture model is utilized to develop a modeling approach for Poisson processes. Full posterior inference for the intensity function and the mean measure of the Poisson process is discussed. The methodology is illustrated with two data sets in Section 4. Finally, Section 5 concludes with a summary and discussion of possible extensions.

2. The probability model for density estimation

Consider observations y_i, $i = 1, ..., n$, assumed to arise from a distribution, with density f, which is supported by a bounded interval, $(0, T)$, on the positive real line. Of interest is inference for f and for other features of the distribution.

The well-studied DP mixture of normals model is not optimal in this context, since inference for the density would be subject to edge effects for data sets with measurements close to the endpoints of $(0, T)$. Mixtures of Beta densities emerge as a natural alternative. With appropriate mixing, this model yields a wide range of distributional shapes, in fact, it can be used to approximate arbitrarily well densities defined on bounded intervals (Diaconis & 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 $\boldsymbol{\theta} = (\mu, \tau)$, the kernel density of the DP mixture can be written as

$$k(y; \mu, \tau) = \frac{y^{\mu\tau T^{-1}-1}(T-y)^{\tau(1-\mu T^{-1})-1}}{\mathrm{Be}\{\mu\tau T^{-1}, \tau(1-\mu T^{-1})\}T^{\tau-1}}, \qquad (1)$$

where $y \in (0, T)$, and $\mathrm{Be}(a, b)$ 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

$$f(y; G) = \int k(y; \mu, \tau)\mathrm{d}G(\mu, \tau),\ G \sim \mathrm{DP}(\alpha, G_0). \quad (2)$$

We assume random α with a gamma(a_α, b_α) prior distribution $p(\alpha)$ such that $\mathrm{E}(\alpha) = a_\alpha / b_\alpha$. 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 (1) 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 $1/d$. In Section 4 we discuss prior specification for α and β as well as sensitivity of posterior estimates to these prior choices.

The probability model for the observables y_i, $i = 1, ..., n$, can be expressed in hierarchical form by introducing a latent mixing parameter vector $\boldsymbol{\theta}_i = (\mu_i, \tau_i)$ associated with each y_i. Then, given G, the $\boldsymbol{\theta}_i$ are i.i.d. from G, and, at the first stage of the model, given the $\boldsymbol{\theta}_i$, the y_i are conditionally independent with densities $k(\cdot; \mu_i, \tau_i)$ as in (1). The discrete countable nature of the DP (Ferguson, 1973; Sethuraman, 1994) 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

$$G = \sum_{j=1}^{\infty} \omega_j \delta_{(\tilde{\mu}_j, \tilde{\tau}_j)}, \qquad (3)$$

where δ_x denotes a point mass at x, $\omega_1 = z_1$, $\omega_j = z_j \prod_{s=1}^{j-1}(1 - z_s)$, j = 2,3,..., with $z_s \mid \alpha$ independent from a Beta$(1, \alpha)$ distribution (where this notation refers to the standard parameterization for the Beta distribution), and, independently, $(\tilde{\mu}_j, \tilde{\tau}_j) \mid \beta$ independent from G_0.

To obtain posterior predictive inference under model (2), it suffices to work with the marginalized version of the hierarchical model that arises by integrating G over its DP prior. Several posterior simulation techniques have been suggested for the resulting marginal posterior, $p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$, and data $= (y_1, ..., y_n)$ (see, e.g., the references in Müller & Quintana, 2004). We have used algorithm 5 from Neal (2000), which, though the least efficient among the MCMC algorithms discussed by Neal (2000), is the easiest to implement, and provided an acceptable exploration of the posterior for the data examples discussed in Section 4 as well as for artificial data in a simulation study (not reported here). However, it would be of interest to experiment with other posterior simulation methods as is briefly discussed in Section 5.

The posterior draws $\boldsymbol{\theta}_b = \{(\mu_{ib}, \tau_{ib}) : i = 1, ..., n\}$, α_b, β_b, $b = 1, ..., B$, from $p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$ can be used to estimate the posterior predictive density $p(y_0|\text{data})$ at any point $y_0 \in (0, T)$,

$$p(y_0|\text{data}) = \int k(y_0; \mu_0, \tau_0) p(\mu_0, \tau_0|\boldsymbol{\theta}, \alpha, \beta) \, p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data}). \qquad (4)$$

Here, $p(\mu_0, \tau_0|\boldsymbol{\theta}, \alpha, \beta)$ is a mixed distribution with point masses (equal to $(\alpha + n)^{-1}$) at the $\boldsymbol{\theta}_i$, $i =$

$1, ..., n$, and continuous mass (equal to $(\alpha+n)^{-1}\alpha$) on $G_0(\mu_0, \tau_0|\beta)$. The posterior predictive density provides the standard Bayesian density estimate. Note, however, that $p(y_0|\text{data})$ is only the posterior expectation for $f(y_0; G)$, the density functional at y_0. More general inference is discussed in Section 3 in the context of modeling for Poisson processes.

To our knowledge, general Beta DP mixtures of the form in (2) have not appeared in the literature. Regarding the use of Beta mixtures in Bayesian modeling, Robert & Rousseau (2003) developed a goodness of fit method using discrete Beta mixtures with unknown number of components. Petrone (1999a; 1999b) and Petrone & Wasserman (2002) focus on density estimation based on random Bernstein densities, $\sum_{j=1}^{k} w_{j,k}\text{be}(x; j, k-j+1)$, where k is assumed random and $\text{be}(x; a, b)$ denotes the Beta density with parameters a and b (again, under the standard parameterization for the Beta distribution). Thus Bernstein densities form a specific class of discrete Beta mixtures such that, for a given number of components, the component parameters are specified and only the mixing weights are random. In particular, the Bernstein-Dirichlet prior model induces a prior on the weights through $w_{j,k} = F(j/k) - F\{(j-1)/k\}$, where F is a distribution function, with support on $[0, 1]$, modeled with a DP prior. We also refer to Mallick & Gelfand (1994) and Gelfand & Mallick (1995) for related work, which employs mixtures of Beta distribution functions to model random monotonic functions.

3. Applications to inference for Poisson process intensities

Point processes are stochastic process models for events that occur separated in time or space. Applications of point process modeling can be found in several scientific fields, including traffic engineering, software reliability, neurophysiology, weather modeling, and forestry. Poisson processes play a fundamental role in the theory and applications of point processes. Focusing on point processes over time, let $N(t)$ denote the number of event occurrences in the time interval $(0, t]$. Then, formally, the point process $\mathcal{N} = \{N(t) : t \geq 0\}$, is a non-homogeneous Poisson process if for any $t > s \geq 0$, $N(t) - N(s)$ follows a Poisson distribution with mean $\Lambda(t) - \Lambda(s)$, and $\mathcal{N}$ has independent increments, that is, for any $0 \leq t_1 < t_2 \leq t_3 < t_4$, the random variables $N(t_2) - N(t_1)$ and $N(t_4) - N(t_3)$ are independent. Here, Λ is the mean measure (or cumulative intensity function) of the Poisson process. For any $t \in R^+$, it is defined as $\Lambda(t) = \int_0^t \lambda(u)du$, where λ is the Poisson process intensity function, which is a non-negative and locally integrable function, that is, $\int_B \lambda(u)du < \infty$, for all bounded $B \subset R^+$. See, e.g., Kingman (1993), Guttorp (1995), and Moller & Waagepetersen (2004) for background on the theory and applications of Poisson processes.

Based on its definition, from a modeling perspective, of interest for a non-homogeneous Poisson process is its intensity function. We consider such a process 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 the Poisson process intensity function λ is given by

$$\exp\{-\int_0^T \lambda(u)du\} \prod_{i=1}^{n} \lambda(t_i). \qquad (5)$$

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. We use the Beta DP mixture model (2) for f. 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.

Using (5), the full Bayesian model for γ and f, equivalently, for γ and G, becomes

$$\exp(-\gamma)\gamma^n \left\{\prod_{i=1}^{n} \int k(t_i; \mu_i, \tau_i)\mathrm{d}G(\mu_i, \tau_i)\right\} p(\gamma)p(G \mid \alpha, \beta)p(\alpha)p(\beta) \qquad (6)$$

with the prior structure $p(G \mid \alpha, \beta)p(\alpha)p(\beta)$ for G and its hyperparameters discussed in Section 2, and a prior $p(\gamma)$, with support $(0, \infty)$, for γ.

Because, in general, it seems difficult to specify parameters for a prior distribution for γ, we use Jeffreys prior based on the marginal likelihood, $\mathcal{L}^*(\gamma)$, for γ, which arises from (6) 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 the prior $p(\gamma) \propto \gamma^{-1}$. In fact, this is the reference prior for γ with the DP mixture prior playing the role of the conditional prior for the *nuisance* parameters (all parameters other than γ). The reference prior approach (see, e.g., Bernardo, 2005) is too technical to discuss here, but its basic idea is to choose the prior which, in a certain asymptotic sense, maximizes the information in the posterior that is provided by the data.

Based on the model structure in (6), 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. In fact, the marginal posterior $p(\gamma|\text{data}) = \text{gamma}(n, 1)$, and $p(\gamma, G, \boldsymbol{\theta}, \alpha, \beta|\text{data}) = p(\gamma|\text{data})\, p(G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$, where $\text{data} = (t_1, ..., t_n)$. Hence, to explore the full posterior distribution $p(\gamma, G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$, it suffices to implement an MCMC method to obtain draws from $p(G, \boldsymbol{\theta}, \alpha, \beta|\text{data})$, the posterior for the DP mixture part of model (6). We use the approach proposed in Gelfand & Kottas (2002), and Kottas (2006), which yields full inference for functionals of the random mixture density $f(t; G)$.

Using results from Antoniak (1974),

$$p(G, \boldsymbol{\theta}, \alpha, \beta|\text{data}) = p(G \mid \boldsymbol{\theta}, \alpha, \beta) p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data}).$$

Here the distribution for $G \mid \boldsymbol{\theta}, \alpha, \beta$ is a DP with updated precision parameter $\alpha + n$ and base distribution $G_0^*(\mu, \tau|\boldsymbol{\theta}, \alpha, \beta) = (\alpha + n)^{-1}\{\alpha G_0(\mu, \tau|\beta) + \sum_{i=1}^n \delta_{(\mu_i, \tau_i)}(\mu, \tau)\}$. Posterior draws $\boldsymbol{\theta}_b$, α_b, and β_b from $p(\boldsymbol{\theta}, \alpha, \beta \mid \text{data})$ (obtained as in Section 2) can be used to draw G_b from $p(G \mid \boldsymbol{\theta}_b, \alpha_b, \beta_b)$ using a truncation approximation to (3). 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 $\text{Beta}(1, \alpha_b + n)$, 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 $\text{E}(\sum_{j=1}^{J-1} w_{jb} \mid \alpha_b) = 1 - \{(\alpha_b + n)/(\alpha_b + n + 1)\}^{J-1}$, we can choose J that makes, say, $\{(n + \max_b \alpha_b)/(n + 1 + \max_b \alpha_b)\}^{J-1}$ arbitrarily small. Now,

$$f_{b0} = \int k(t_0; \mu, \tau) \text{d}G_b(\mu, \tau) = \sum_{j=1}^J w_{jb} k(t_0; \mu'_{jb}, \tau'_{jb})$$

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 (1), is a posterior realization for $F(t_0; G) = \int_0^{t_0} f(u; G) du = \int K(t_0; \mu, \tau) \text{d}G(\mu, \tau)$, 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.

Regarding the Bayesian nonparametric literature on modeling for Poisson processes, most of the work has focused on the cumulative intensity function Λ, including priors based on gamma processes, weighted gamma processes, Beta processes, and Lévy processes; see Lo (1992), Kuo & Ghosh (1997), Gutiérrez-Peña & 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 & Weng (1989), which was recently extended in Ishwaran & 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 & Ickstadt (1998) applied to one-dimensional Poisson processes. Wolpert & Ickstadt (1998) developed a hierarchical model to account for spatial variation in the intensity function of a spatial Poisson process. The approaches by Wolpert & Ickstadt (1998) and Ishwaran & James (2004) are useful for general multiplicative point processes. However, for the important special case of non-homogeneous Poisson processes, the proposed approach based on DP mixture models might be a useful addition to the existing methods as it builds on a familiar Bayesian density estimation framework, facilitating the choice of kernel and prior for the mixing distribution.

4. Data illustrations

Here, we illustrate the methodology with two data sets. Regarding the priors for the DP hyperparameters, recall that α controls the number n^* of distinct components in the DP mixture (2) (Antoniak, 1974; Escobar & West, 1995). For instance, for moderately large n, $\text{E}(n^*) \approx (a_\alpha/b_\alpha) \log\{1 + (n b_\alpha/a_\alpha)\}$. To specify the mean, $1/d$, of the exponential prior for β, we consider a single component of mixture (2), 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 y_i (Section 2) or the t_i (Section 3). 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 density or intensity function shape. For both data sets discussed below, we used $r = T$ (which can be viewed as a default choice) leading to an exponential prior for β with mean 8. We have also experimented with less informative priors for β, based on $T < r \leq 1.5T$, revealing little sensitivity of the resulting posterior estimates.

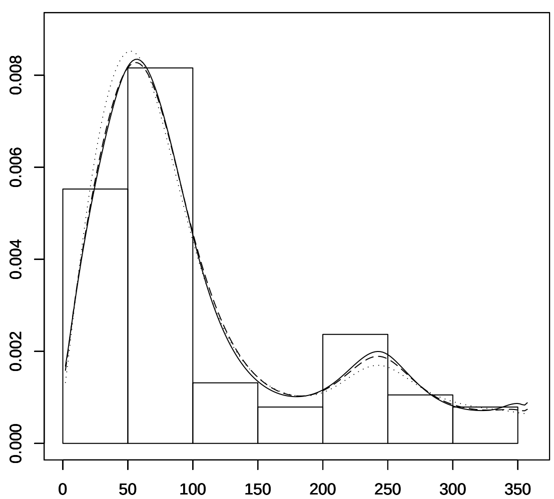

Figure 1. Turtle data. Posterior predictive densities under three prior choices (see Section 4 for details) overlaid on histogram of the data.

The first data set is a standard example from the field of directional statistics. It consists of directions (in degrees, clockwise from north) in which each of 76 female turtles moved after laying their eggs on a beach. A histogram of the data is provided in Figure 1. Most of the turtles show a preference for swimming approximately in the 60° direction, while a substantial minority prefer the opposite direction. See, e.g., Ferreira, Juárez & Steel (2006) for recent work on model-based approaches to the analysis of directional data as well as for references to earlier analyses of the turtle data. Here, we work only with the data on the direction angle to study the performance of the Beta DP mixture model in density estimation.

Figure 1 includes estimates for the posterior predictive density (4) under three priors for α, a gamma$(2, 5.25)$ prior (dotted line), a gamma$(2, 2.19)$ prior (dashed line), and a gamma$(2, 0.75)$ prior (solid line). These choices correspond to $\mathrm{E}(n^*) \approx 2$, 4, and 9, respectively.

Regarding inference for n^*, the posterior 25% percentile, posterior median, and posterior 75% percentile are given by 3, 4, and 5, under the gamma$(2, 5.25)$ prior for α; by 4, 6, and 7, under the gamma$(2, 2.19)$ prior; and by 6, 8, and 10, under the gamma$(2, 0.75)$ prior. We note that, although the posterior summaries for n^* are affected to some extent by the prior for α, the posterior density estimates depict relatively little sensitivity to sensible prior choices for α.

We next turn to the analysis of a temporal point pattern, to illustrate the modeling approach for Poisson processes over time. The data set consists of 31 failure times, in days, based on the trouble report for one of the larger modules of the Naval Tactical Data System, and has been previously analyzed, among others, by Kuo & Yang (1996), Kuo & Ghosh (1997) and Gutiérrez-Peña & Nieto-Barajas (2003), using both parametric and nonparametric Bayesian methods.

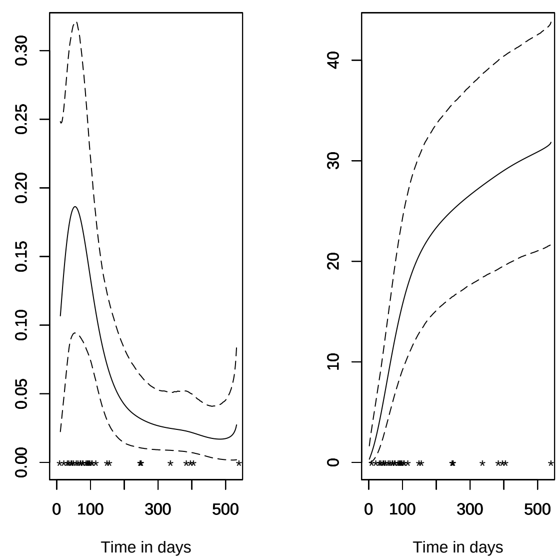

Figure 2. Software reliability data. Posterior point estimates (solid lines) and interval estimates (dashed lines) for the intensity function (left panel) and the cumulative intensity function (right panel). In both panels, the observed failure times are plotted on the horizontal axis.

We fit model (6) using a gamma$(2, 1.64)$ prior for α, which implies $\mathrm{E}(n^*) \approx 4$. We also obtained posterior results under different priors for α, yielding $\mathrm{E}(n^*)$ between 2 and 7. Encouragingly, even with the smaller sample size which is available here, posterior inference was again robust to these prior choices.

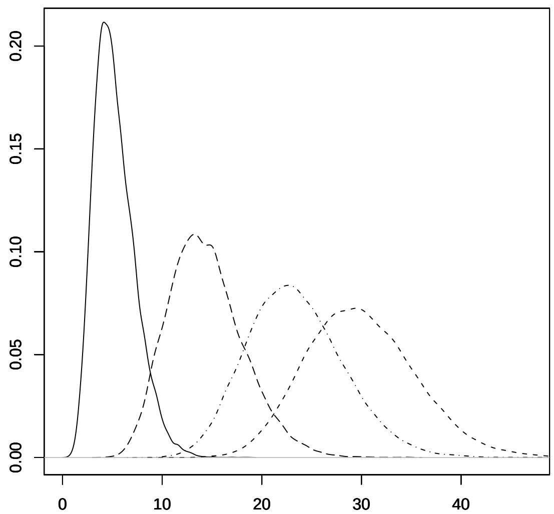

Figure 3. Software reliability data. Posteriors for the cumulative intensity at time points, t_0 = 40, 90, 200, and 450, plotted by the solid, dashed, dashed-dotted, and small dashed lines, respectively.

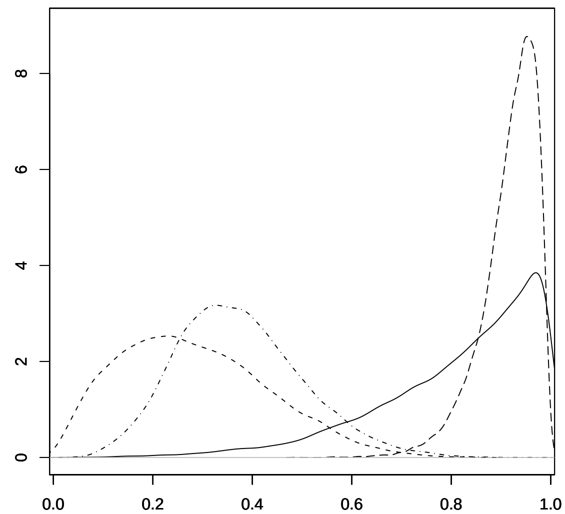

Figure 4. Software reliability data. Posterior for the probability of at least one failure in the time intervals $(0, 20]$, $(90, 110]$, $(335, 355]$, and $(490, 510]$, plotted by the solid, dashed, dashed-dotted, and small dashed lines, respectively.

Figure 2 shows point estimates (based on posterior means), and interval estimates (based on 95% pointwise central posterior intervals) for the intensity function and the cumulative intensity function. Results for the cumulative intensity function are similar to the ones reported in Kuo & Ghosh (1997) and Gutiérrez-Peña & Nieto-Barajas (2003), although the Beta DP mixture model yields smoother estimates, which is, arguably, a desirable property for a prior probability model in this context. Moreover, it readily provides inference for the intensity function depicting more emphatically the pattern of failures over time.

To indicate the range of inferences the method provides, Figure 3 plots the entire posterior for the cumulative intensity functional, $\Lambda(t_0; \gamma, G)$, at four time points, t_0 = 40, 90, 200, and 450, and Figure 4 plots the posterior for $1 - \exp\{\Lambda(s) - \Lambda(t)\}$, that is, for the probability of at least one failure in the time interval $(s, t]$; included are the posteriors corresponding to four intervals of length equal to 20 days.

5. Summary

We have proposed a modeling approach for density functions on bounded intervals, and for Poisson process intensity functions. The method is based on DP mixtures of Beta distributions. We have discussed how posterior inference can be obtained for both applications considered. Finally, we have illustrated the model with two data examples.

An extension of the methodology discussed in this paper to modeling for spatial non-homogeneous Poisson process intensities has been recently proposed by Kottas and Sansó (2006), where particular emphasis was placed on applications to extreme value analysis problems. The DP mixture model used in Kottas and Sansó (2006) is based on a bivariate Beta density. When covariate information is available, a practically important extension of the model (for Poisson processes either over time or over space) would be to semiparametric regression settings. In particular, current work studies new modeling formulations for data that include individual-specific covariates, that is, for point patterns that can be assumed to arise from a marked Poisson process.

In addition to extensions to new modeling scenarios, of practical importance would be a study of the performance of different posterior simulation methods for the proposed Beta DP mixture model. Such a study might include, for instance, alternative MCMC algorithms for non-conjugate DP mixtures (e.g., MacEachern & Müller, 1998; Neal, 2000) as well as alternative approaches to MCMC methods, such as variational inference methods (Blei & Jordan, 2006).

Acknowledgments

The author wishes to thank an anonymous reviewer for useful comments. This work was supported in part by NSF grant DMS-0505085.

References

Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to nonparametric problems. *Annals of Statistics*, 2, 1152–1174.

Bernardo, J.M. (2005). Reference analysis. In D.K. Dey and C.R. Rao (Eds.), *Handbook of Statistics 25*. Amsterdam: Elsevier, pp. 17–90.

Blei, D.M., & Jordan, M.I. (2006). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 1, 121–144.

Brunner, L.J., & Lo, A.Y. (1989). Bayes methods for a symmetric unimodal density and its mode. *Annals of Statistics*, 17, 1550–1566.

Brunner, L.J. (1995). Bayesian linear regression with error terms that have symmetric unimodal densities. *Journal of Nonparametric Statistics*, 4, 335–348.

De Iorio, M., Müller, P., Rosner, G.L., & MacEachern, S.N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, 99, 205-215.

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*. Amsterdam: North-Holland.

Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89, 268–277.

Escobar, M., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577–588.

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

Ferguson, T.S. (1983). Bayesian density estimation by mixtures of normal distributions. In M.H. Rizvi, J.S. Rustagi and D. Siegmund (Eds.), *Recent advances in statistics*. New York: Academic Press.

Ferreira, J.T.A.S., Juárez, M.A., & Steel, M.F.J. (2006). *Directional Log-spline Distributions* (Technical Report). Department of Statistics, University of Warwick, UK.

Gelfand, A.E., & Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 11, 289–305.

Gelfand, A.E., & Mallick, B.K. (1995). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics*, 51, 843–852.

Gelfand, A.E., Kottas, A., & MacEachern, S.N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, 100, 1021–1035.

Griffin, J.E. & Steel, M.F.J. (2006). Order-based dependent Dirichlet processes. *Journal of the American statistical Association*, 101, 179-194.

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*. Oxford: University Press.

Guttorp, P. (1995). *Stochastic Modelling of Scientific Data*. Boca Raton: Chapman & Hall.

Hanson, T.E. (2006). Modeling censored lifetime data using a mixture of gammas baseline. *Bayesian Analysis*, 1, 575–594.

Ishwaran, H., & James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96, 161–173.

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. *Journal of the American Statistical Association*, 99, 175–190.

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

Kottas, A. (2006). Nonparametric Bayesian survival analysis using mixtures of Weibull distributions. *Journal of Statistical Planning and Inference*, 136, 578–596.

Kottas, A., & Gelfand, A.E. (2001). Bayesian semi-parametric median regression modeling. *Journal of the American Statistical Association*, 96, 1458–1468.

Kottas, A., & Krnjajić, M. (2005). *Bayesian Nonparametric Modeling in Quantile Regression* (Technical Report AMS 2005-06). Department of Applied Mathematics and Statistics, University of California, Santa Cruz.

Kottas, A., & Sansó, B. (2006). Bayesian Mixture Modeling for Spatial Poisson Process Intensities, with Applications to Extreme Value Analysis. To appear in *Journal of Statistical Planning and Inference*, Special Issue on "Bayesian Inference in Stochastic Processes".

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. *Journal of the American Statistical Association*, 91, 763–773.

Lijoi, A., Mena, R.H., & Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association*, 100, 1278–1291.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics*, 12, 351–357.

Lo, A.Y. (1992). Bayesian inference for Poisson process models with censored data. *Journal of Nonparametric Statistics*, 2, 71–80.

Lo, A.Y., & Weng, C.-S. (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Annals of the Institute of Statistical Mathematics*, 41, 227–245.

MacEachern, S.N. (2000). *Dependent Dirichlet Processes.* (Technical Report). Department of Statistics, The Ohio State University.

MacEachern, S.N., & Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 7, 223–238.

MacEachern, S.N., & Müller, P. (2000). Efficient MCMC schemes for robust model extensions using encompassing Dirichlet process mixture models. In F. Ruggeri and D. Rios-Insua (Eds.), *Robust Bayesian Analysis*. New York: Springer.

Mallick, B.K., & Gelfand, A.E. (1994). Generalized linear models with unknown link functions. *Biometrika*, 81, 237–245.

Merrick, J.R.W., Soyer, R., & Mazzuchi, T.A. (2003). A Bayesian semiparametric analysis of the reliability and maintenance of machine tools. *Technometrics*, 45, 58–69.

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

Müller, P., & Quintana, F.A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, 19, 95–110.

Neal, R.M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9, 249–265.

Petrone, S. (1999a). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, 26, 373–393.

Petrone, S. (1999b). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics*, 27, 105–126.

Petrone, S., & Wasserman, L. (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society, Series B*, 64, 79–100.

Robert, C.P., & Rousseau, J. (2003). *A mixture approach to Bayesian goodness of fit* (Technical Report). CEREMADE, Université Paris Dauphine.

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

Teh, Y.W., Jordan, M.I., Beal, M.J. & Blei, D.M. (2006). Hierarchical Dirichlet processes. To appear in the *Journal of the American Statistical Association.*

Wolpert, R.L., & Ickstadt, K. (1998). Poisson/Gamma random field models for spatial statistics. *Biometrika*, 85, 251–67.