

## A SENSITIVITY ANALYSIS FOR BAYESIAN NONPARAMETRIC DENSITY ESTIMATORS

Luis E. Nieto-Barajas<sup>1</sup> and Igor Prünster<sup>2</sup>

<sup>1</sup>ITAM, Mexico, <sup>2</sup>University of Turin, Collegio Carlo Alberto and ICER, Italy

*Abstract:* Bayesian nonparametric methods have recently gained popularity in the context of density estimation. In particular, the density estimator arising from the mixture of Dirichlet process is now commonly exploited in practice. In this paper we perform a sensitivity analysis for a wide class of Bayesian nonparametric density estimators, including the mixture of Dirichlet process and the recently proposed mixture of normalized inverse Gaussian process. Whereas previous studies focused only on the tuning of prior parameters, our approach consists of perturbing the prior itself by means of a suitable function. In order to carry out the sensitivity analysis we derive representations for posterior quantities and develop an algorithm for drawing samples from mixtures with a perturbed nonparametric component. Our results bring out some clear evidence for Bayesian nonparametric density estimators, and we provide an heuristic explanation for the neutralization of the perturbation in the posterior distribution.

*Key words and phrases:* Bayesian nonparametric inference, density estimation, increasing additive process, latent variables, Lévy process, mixture model, sensitivity.

### 1. Introduction

The problem of density estimation is an inferential problem that falls naturally in the realm of nonparametric statistics. Many estimators proposed in the literature are of the form

$$f(x) = \int k(x, y)P(dy), \quad (1.1)$$

where  $k$  is a parametric kernel and  $P$  is a mixing distribution. The popular kernel density estimator is of the form (1.1) with  $P$  the empirical distribution function. The Bayesian counterpart is achieved by replacing the empirical distribution function with a nonparametric prior and, hence, (1.1) is a random density function. The Bayesian density estimator, with respect to the integrated squared loss function, is then given by the predictive distribution arising from (1.1),

$$\hat{f}_n(x) = E \left[ \int k(x, y)P(dy) \mid X_1, \dots, X_n \right]. \quad (1.2)$$

A typical choice for  $P$  is the Dirichlet process, giving rise to the popular Dirichlet process mixture (DPM). An early use of the DPM can be found in Berry and Christensen (1979), whereas the first systematic treatment of density estimation is due to Lo (1984), who also provided an expression for (1.2). However, such an expression, involving a sum over partitions, turns out to be inapplicable to concrete examples of datasets. Ferguson (1983) noted that the DPM can be formulated in hierarchical form as

$$\begin{aligned} X_i | Y_i &\stackrel{\text{ind}}{\sim} k(\cdot, Y_i) \\ Y_i | P &\stackrel{\text{i.i.d.}}{\sim} P \\ P &\sim \mathcal{D}_\alpha, \end{aligned} \tag{1.3}$$

where  $\mathcal{D}_\alpha$  stands for the Dirichlet process prior with parameter measure  $\alpha(\cdot) = aP_0(\cdot)$ , with  $P_0(\cdot) = E[P(\cdot)]$  being the prior guess at the shape. This idea turned out to be very fruitful for practical purposes. Indeed, Escobar (1988) (see also Escobar (1994) and Escobar and West (1995)) developed an MCMC algorithm for simulating from the posterior distribution of a DPM, which allowed the DPM and semiparametric variations of it to be exploited in many different applied contexts such as longitudinal data models (Müller and Rosner (1997), Müller, Quintana and Rosner (2004)), clustering and product partition models (Petrone and Raftery (1997) and Quintana and Iglesias (2003)), survival analysis (Doss (1994)), analysis of randomized block experiments (Bush and MacEachern (1996)), multivariate ordinal data (Kottas, Müller and Quintana (2005)), and differential gene expression (Do, Müller and Tang (2005)). See Müller and Quintana (2004) for a review.

In the recent literature, various large classes of discrete random probability measures that generalize the Dirichlet process have been proposed. Among them we mention species sampling models (Pitman (1996)), stick breaking priors (Ishwaran and James (2001)), normalized random measures with independent increments (Regazzini, Lijoi and Prünster (2003)), Poisson–Kingman models (Pitman (2003)) and spatial neutral to the right models (James (2006)). Indeed, any specific prior contained in the classes listed above is a potential candidate for replacing the Dirichlet process in (1.3). For instance, Lijoi, Mena and Prünster (2005, 2007) used specific normalized random measures with independent increments (NRMI), and highlighted their different clustering behavior with respect to the DPM model. Recently interest has also focussed on generalizations of (1.3) by the introduction of general dependence structures on time or covariates. This important research line was initiated in the seminal papers of MacEachern (1999, 2000), whose dependent

nonparametric priors constitute the theoretical framework for many models now exploited in practice. Among other contributions, Griffin (2007) proposed a class of time-varying NRMI in view of financial applications, Dunson and Park (2008) and Dunson, Xue and Carin (2008) introduced dependent stick-breaking priors for epidemiologic and bioassay problems, and De Iorio, Müller, Rosner and MacEachern (2004) designed an ANOVA model based on the dependent Dirichlet process. It has to be remarked that in all the papers mentioned above, some sensitivity analysis with respect to the choice of the parameters of the mixing measure  $P$ , chosen to be the Dirichlet process or a suitable tractable alternative, was performed. However, at least to the authors knowledge, no sensitivity analysis involving a perturbation of  $P$  itself has been carried out.

In this paper we consider mixture models like (1.3) where  $P$  is a perturbed NRMI. To be more specific, consider a Dirichlet process; it is well known (Ferguson (1973)) that it can be constructed as a suitable normalization of a gamma process  $\Gamma$ . Now, take some positive deterministic function  $\bar{k}$  and define a new random probability measure as

$$P(dy) = \frac{\bar{k}(y)\Gamma(dy)}{\int \bar{k}(s)\Gamma(ds)} = \frac{\sum_{i \geq 1} \bar{k}(y) J_i \delta_{Y_i}(dy)}{\sum_{i \geq 1} \bar{k}(y) J_i}, \quad (1.4)$$

where  $\delta_a$  is a point mass at  $a$ , the  $Y_i$ 's are the jump locations and the  $J_i$ 's are the jump heights. Note that  $P$  in (1.4) is obviously not a Dirichlet process anymore, unless  $\bar{k}$  is constant. The role of the function  $\bar{k}$  in (1.4) consists of amplifying or squeezing the heights of the jumps of the gamma process. Hence,  $\bar{k}$  can be used to perturb the original Dirichlet process: by selecting a  $\bar{k}$  with a relatively large (low) value on a certain interval  $B$ , the expected probability assigned to  $B$  increases (decreases). In this paper we try to answer the question of whether the Bayesian density estimator (1.2) is robust with respect to such a perturbation. Our answer is positive and we also provide a description of how the posterior distribution of (1.4) is able to neutralize the effect of  $\bar{k}$ . Indeed, we face the problem in a greater generality by allowing  $\Gamma$  in (1.4) to be any increasing additive process (IAP): mixtures with such a driving measure have been termed normalized priors driven by IAPs in Nieto-Barajas, Prünster and Walker (2004), where the distribution of their mean functionals is studied.

In order to carry out our sensitivity analysis, two main ingredients are needed: the posterior distribution of a perturbed NRMI and a simulation algorithm for drawing samples from a mixture like (1.3), with mixing measure a perturbed NRMI. These are given in Section 2, together with a precise definition

of perturbed NRMI. In Section 3 we study in great detail a simulated data example, and analyze the posterior behaviour of the perturbed NRMI highlighting its ability to neutralize the influence of  $\bar{k}$ . The section is then completed by the analysis of two classical datasets, namely the galaxy and acidity data. Section 4 contains some concluding remarks.

## 2. Tools for the Analysis of a General Perturbed Mixture Model

### 2.1. The perturbed mixture model

Let us start this section by describing the mixture model we are going to consider in some detail. Let  $(X_i)_{i \geq 1}$  be a sequence of observable random variables, whereas  $(Y_i)_{i \geq 1}$  is a sequence of latent random variables. We assume a mixture model for the observations, namely

$$\begin{aligned} X_i | Y_i &\stackrel{\text{ind}}{\sim} k(\cdot, Y_i) \\ Y_i | P &\stackrel{\text{i.i.d.}}{\sim} P \\ P &\sim \mathcal{P}, \end{aligned} \tag{2.1}$$

where  $\mathcal{P}$  is the distribution of a perturbed NRMI given by

$$P(dy) = \frac{\bar{k}(y)A(dy)}{\int_{\mathbb{R}} \bar{k}(s)A(ds)}, \tag{2.2}$$

with  $\bar{k}$  being some non-negative function, and  $A$  an increasing additive process (IAP) that is an increasing process with independent but not necessarily stationary increments. See Sato (1999) for an exhaustive account of IAPs. Note that if  $\bar{k}$  is a constant, then the perturbed NRMI reduces to an NRMI as defined in Regazzini, Lijoi and Prünster (2003). In the following, the process  $A_k(y) := \int_{(-\infty, y]} \bar{k}(s)A(ds)$  is termed a *weighted IAP*. The random probability in (2.2) is uniquely characterized by  $\bar{k}$  and the Poisson intensity measure  $\nu$  corresponding to  $A$ . This can be seen from the Laplace transform of  $A_k(y)$  which, for any  $\lambda \geq 0$ , is given by

$$E \left[ e^{-\lambda \int_{(-\infty, y]} \bar{k}(s)A(ds)} \right] = \exp \left\{ - \int_{(-\infty, y] \times \mathbb{R}^+} \left( 1 - e^{-\lambda \bar{k}(s)v} \right) \nu(ds, dv) \right\}. \tag{2.3}$$

In the sequel, it is useful to write the Poisson intensity as

$$\nu(dy, dv) = \rho(dv|y)\alpha(dy), \tag{2.4}$$

where  $\alpha(\cdot)$  is a measure on  $\mathbb{R}$  and  $\rho$  a measurable kernel such that  $\rho(\cdot|y)$  is a measure on  $\mathbb{R}^+$  for every  $y \in \mathbb{R}$ . Now, the normalization in (2.2) leads to a well-defined random probability measure if the denominator in (2.2) is (almost surely)

finite and positive. This can be achieved by requiring  $\bar{k}$ ,  $\alpha$  and  $\rho$  to satisfy  $\int_{\mathbb{R} \times \mathbb{R}^+} [1 - \exp\{-\lambda v \bar{k}(y)\}] \rho(dv|y) \alpha(dy) < \infty$  for every  $\lambda > 0$  (finiteness), and  $\rho(\mathbb{R}^+|y) = \infty$  for every  $y \in \mathbb{R}$  (positiveness). See Nieto-Barajas, Prünster and Walker (2004) for details.

As recalled in the Introduction, the Dirichlet process with parameter  $\alpha$  arises when  $\bar{k}(y) = c$  and  $A$  is a gamma process or, equivalently,  $\rho(dv|y) = v^{-1} e^{-v} dv$ . The normalized inverse Gaussian (N-IG) process (Lijoi, Mena and Prünster (2005)) with parameter measure  $\alpha$  is obtained by setting  $\bar{k}(y) = c$  and  $\rho(dv|y) = \delta(\sqrt{2\pi})^{-1} v^{-3/2} e^{-\gamma^2 v/2} dv$ , which corresponds to  $A$  being an inverse Gaussian process. On the other hand, if  $\bar{k}$  is not a constant, the resulting  $P$  in (2.2) is not Dirichlet or N-IG anymore, but is a normalized perturbed gamma or inverse Gaussian measure. It is worth noting that the un-normalized process  $\bar{k}(y)\Gamma(dy)$  is also known as an extended gamma process, which has been introduced in Dykstra and Laud (1981) in the context of survival analysis. In Section 3 we compare the density estimates arising from these two priors, and from their perturbed variations.

## 2.2. The posterior distribution of the mixing measure

Given the complexity of a mixture model like (2.1), inference is necessarily simulation-based. Nonetheless the derivation of some analytical quantities is required to set up an algorithm for drawing samples from (2.1). A common strategy is to integrate out  $P$  in (2.1), and then resort to its predictive distributions within a Gibbs sampler to obtain posterior samples. Such an approach falls within the class of marginal methods and can be applied to DPM and N-IG mixtures. However, for general mixing distributions like (2.2), even when  $A$  is a gamma or inverse Gaussian process, the predictive distributions are intractable. Hence we have to derive, according to the terminology of Papaspiliopoulos and Roberts (2008), a conditional method, which means that we have to set up an algorithm without integrating out (2.2) in (2.1).

Consequently the basic analytical building block of the algorithm, to be set forth in the next paragraph, is represented by a posterior characterization of (2.2), given the latent variables  $Y := (Y_1, \dots, Y_n)$ . Indeed, it is enough to characterize the un-normalized perturbed IAP  $A_k(dy)$ , since the normalization can be carried out within the algorithm. Due to the (almost sure) discreteness of the driving process  $A_k(dy)$ , ties usually appear in the  $Y_i$ 's. Hence, define  $Y^* := (Y_1^*, \dots, Y_r^*)$  as the set of distinct values within  $Y$ , and denote by  $n_j$  the frequency of  $Y_j^*$ , for  $j = 1, \dots, r \leq n$ . Obviously,  $\sum_{j=1}^r n_j = n$ . We are now in a position to provide the posterior representation that characterizes the posterior distribution  $A_k(dy)$  as a mixture with respect to a suitable latent variable.

**Proposition 1.** *Let  $Y$  be a set of latent variables sampled from (2.2). Then the posterior distribution of  $A_k(ds) = \bar{k}(s)A(ds)$ , given  $Y$ , is a mixture with respect to the distribution of a latent variable  $U$ . More explicitly, we have the following.*

- (a)  $[A_k(ds)|U, Y]$  coincides in distribution with a weighted IAP with fixed points of discontinuity

$$\bar{k}(s)A^*(ds) + \sum_{j=1}^r \bar{k}(s)J_j^*\delta_{Y_j^*}(ds). \quad (2.5)$$

- (a.1)  $A^*(ds)$  is an IAP without fixed points of discontinuity, characterized by the intensity

$$\nu^*(ds, dv) = e^{-u\bar{k}(s)v}\nu(ds, dv). \quad (2.6)$$

- (a.2) The  $Y_j^*$ ,  $j = 1, \dots, r$ , are fixed points of discontinuity with jump heights  $J_j^*$  that are absolutely continuous with density

$$f_j(v) \propto v^{n_j} e^{-U\bar{k}(Y_j^*)v} \rho(dv|Y_j^*). \quad (2.7)$$

- (a.3) The jumps  $J_j^*$ ,  $j = 1, \dots, r$  are mutually independent and independent of  $A^*$ .

- (b)  $[U|Y]$  is absolutely continuous with density

$$f_{U|Y}(u) \propto u^{n-1} \exp\{-\psi_{\bar{k}}(u)\} \prod_{j=1}^r \tau_{n_j}(u|Y_j^*), \quad (2.8)$$

where  $\psi_{\bar{k}}(u) = \int_{\mathbb{R} \times \mathbb{R}^+} (1 - e^{-u\bar{k}(s)v})\nu(ds, dv)$  and  $\tau_{n_j}(u|y_j^*) = \int_0^\infty v^{n_j} e^{-u\bar{k}(y_j^*)v} \rho(dv|y_j^*)$ .

The proof is postponed to the Appendix. We remark that an indirect proof of this result can also be derived from the posterior characterization of an NRMI due to James, Lijoi and Prünster (2009), via some suitable modifications and changes of variables.

### 2.3. Posterior simulation from the perturbed mixture model

Given the posterior characterization of the perturbed IAP  $A_k(ds)$ , we are in a position to set up a Gibbs sampling scheme for simulating from the posterior of a general mixture model (2.1). This sampling strategy represents a generalization of the algorithm set forth in Nieto-Barajas and Prünster (2008) for a simple NRMI.

Recall that we do not observe the latent  $Y_i$ 's (nor the  $Y_j^*$ 's), we only observe  $X := (X_1, \dots, X_n)$ . Therefore, in order to make posterior inference, we need to implement a Gibbs sampling scheme with conditional distributions

$$[A_k|X, Y] \text{ and } [Y|X, A_k]. \quad (2.9)$$

First consider  $[A_k|X, Y]$ : due to conditional independence properties, this conditional distribution does not depend on  $X$ . By Proposition 1, we can sample from this distribution in two steps by sampling first  $[U|Y]$  and then  $[A_k|U, Y]$ . The distribution  $[U|Y]$  is univariate and absolutely continuous with density given in (2.8). Therefore samples can be obtained by implementing a Metropolis–Hastings step within the Gibbs sampler.  $[A_k|U, Y]$  is an updated weighted IAP of the form (2.5). For simulating the random process  $A^*$  in (2.5), we resort to the representation of Ferguson and Klass (1972) combined with an inverse Lévy measure algorithm. See Walker and Damien (2000) for a discussion of the Ferguson and Klass algorithm. To this end, write

$$A^*(s) = \sum_i J_i \mathbb{I}_{(\tau_i \leq s)}, \quad (2.10)$$

where  $\mathbb{I}_A$  denotes the indicator of set  $A$ . The positive random variables  $J_i$  are obtained via  $\theta_i = -M(J_i)$ , with  $M(v) = -\nu(\mathbb{R}, [v, \infty))$ ,  $J_i = 0$  if  $\theta_i > -M(0)$ , and  $\theta_1, \theta_2, \dots$  are jump times of a standard Poisson process at unit rate, i.e.,  $\theta_1, \theta_2 - \theta_1, \dots \stackrel{\text{i.i.d.}}{\sim} \text{Ga}(1, 1)$ , where  $\text{Ga}(a, b)$  stands for the gamma distribution with mean  $a/b$ . The random locations  $\tau_i$  are obtained via  $\omega_i = n_{\tau_i}(J_i)$ , where  $\omega_1, \omega_2, \dots$  are i.i.d. from a uniform distribution on  $(0, 1)$ , independent of the  $J_i$ 's and

$$n_\tau(v) = \frac{\nu((-\infty, \tau], dv)}{\nu(\mathbb{R}, dv)}.$$

Hence, the posterior weighted IAP  $A_k(y)|U, Y$  can be expressed as

$$A_k(y) = \sum_i \bar{k}(z_i) G_i \mathbb{I}_{(z_i \leq y)},$$

where, for any  $i \geq 1$ ,  $G_i \in \{J_1, \dots\} \cup \{J_1^*, \dots, J_k^*\}$  and  $z_i \in \{\tau_1, \dots\} \cup \{y_1^*, \dots, y_k^*\}$ , with the  $\tau_j$ 's and  $J_j$ 's denoting the locations and jumps in (2.10), the  $y_j^*$ 's and  $J_j^*$ 's being the fixed points of discontinuity and the corresponding jumps sampled via (2.7), respectively.

Now, concerning  $[Y|X, A_k]$ , since an IAP is a pure jump process, its support is given by the locations of the jumps of  $A_k$ , that is the  $z_i$ 's, and therefore

$$f_{Y_i|X_i, A_k}(v) \propto \sum_j k(x_i, v) G_j \delta_{z_j}(v). \quad (2.11)$$

Note that sampling from (2.11) will produce ties in the  $Y_i$ 's since the support of the conditional density (2.11) for each  $Y_i$  is the same; however, the probability assigned to each point mass is different.

Once we have a sample from the posterior distribution of the driving measure  $A_k(\cdot)$ , a realization of the posterior random density  $f(x)$ , given in (2.1), can be expressed as a discrete mixture of the form

$$f(x|A_k) = \sum_i k(x, z_i) \frac{\bar{k}(z_i)G_i}{\sum_j \bar{k}(z_j)G_j}. \quad (2.12)$$

When sampling from the posterior distribution of nonparametric mixture models, a phenomenon of “sticky cluster locations” often appears causing a slow-down in the convergence of the algorithm. Such an issue was first noted in Bush and MacEachern (1996), where an important device for overcoming this problem was proposed; it consists in a re-sampling step, sometimes termed *acceleration step*, to be added to the algorithms. Specifically, one has to re-sample the locations of the distinct latent  $Y_j^*$ 's from the full conditional distribution  $f_{Y_j^*}(v|x, \text{cluster configuration})$ , whose density is proportional to

$$\prod_{i:y_i=y_j^*} k(x_i, v)P_0(dv). \quad (2.13)$$

Sampling from this distribution can be done by implementing a Metropolis-Hastings step, if  $P_0(\cdot) = E[P(\cdot)]$  is known explicitly. This is the case for the Dirichlet and the N-IG processes, where  $P_0(\cdot) = \alpha(\cdot)/a$ . In the perturbed case  $P_0$  has more complicated structure, but if  $\bar{k}$  is a step-function it can still be computed explicitly. Details for its determination are given in the Appendix.

Therefore, the algorithm for simulating from the posterior distributions in (2.9) can be summarized as follows. Given the starting points  $y_1^{(0)}, \dots, y_n^{(0)}$ , and using the notation in terms of distinct  $y_j^*$ 's with frequency  $n_j$ , for  $j = 1, \dots, r$ , one has to iterate the following steps.

- (i) Sample  $u$  from the conditional density (2.8).
- (ii) Sample the process  $A^*(\cdot)$  using the Ferguson and Klass algorithm with Lévy intensity (2.6).
- (iii) For each  $y_j^*$ , with  $j = 1, \dots, r$ , sample a jump height  $J_j^*$  according to its density (2.7).
- (iv) Compute a realization of the posterior random density  $f(x)$  by (2.12).
- (v) For  $i = 1, \dots, n$ , sample  $y_i$  from its discrete conditional density (2.11).
- (vi) Group the  $y_i$ 's obtained in (v) into the distinct  $y_j^*$ 's with frequency  $n_j$ . Re-sample the values of the fixed locations  $y_j^*$  using (2.13).

### 3. Sensitivity Analysis

In this section we perform the sensitivity analysis and show that the posterior density estimates are not significantly affected by the perturbation arising through  $\bar{k}$ . The setup is as follows. The kernel in (2.1) is chosen to be

$$k(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - y)^2\right\}, \quad (3.1)$$

which represents the typical choice in the context of mixture models. In the simulated data example,  $\sigma$  in (3.1) is fixed, whereas in data examples a prior will be placed on  $\sigma$  leading to a semiparametric mixture model. Hence, the nonparametric prior (2.2) controls the means of the normal components in the mixture. We focus attention on the following two choices: (1)  $A$  is a gamma process with parameter  $\alpha(\cdot)$ ; (2)  $A$  is an inverse Gaussian process with parameter  $\alpha(\cdot)$ . These choices produce the Dirichlet and N-IG priors if  $\bar{k}$  is constant; otherwise perturbed versions of them arise. Even if their perturbed versions are not Dirichlet or N-IG processes, for terminological simplicity, we refer to them as perturbed Dirichlet and N-IG processes. Another option for  $A$  is represented by the stable subordinator; however, by the properties of stable distribution, in such a case  $\bar{k}$  would be absorbed by  $A$  and the whole model would reduce to the mixture of normalized stable process studied in Ishwaran and James (2001).

As for the choice of the parameter measure of the processes we used  $\alpha(dy) = a P_0(dy) = a \text{Ga}(y|p, q)dy$ , where  $\text{Ga}(\cdot|p, q)$  stands for gamma density with mean  $p/q$  and  $a$  some positive constant. This choice was motivated by the fact that we consider positive data, and it serves also to highlight the fact that our algorithm poses no additional computational difficulties if the kernel  $k$  and  $P_0$  do not form a conjugate pair. Hence, we were left with the selection of the perturbing function  $\bar{k}$ . We considered functions  $\bar{k}$  of the type

$$\bar{k}(y) = \begin{cases} \kappa & \text{if } y \in B \subset \mathbb{R}^+, \\ 1 & \text{otherwise.} \end{cases}$$

If  $\kappa > 1$ , then  $\bar{k}$  amplifies the jumps of  $A$  in the region  $B$ , thus increasing, with respect to the unperturbed model, the prior expected probability that the means of the normal components lie in  $B$ . If  $\kappa < 1$ , then the prior probability of selecting means in  $B$  decreases. Posterior inference was obtained via the Gibbs sampler outlined in Section 2.3. For each case we took 20,000 iterations with a burn-in of 2,000 and kept the last 18,000 for computing the estimates.

Before moving on to the analysis of the various datasets, it is important to remark that the following results do not depend on the choices listed above. Indeed, extensive simulation studies, not reported here, indicate that the same

conclusions are reached when changing the kernel, the process  $A$ , its parameter measure  $\alpha$ , and the perturbing function  $\bar{k}$ . In the analysis of the following simulated data example, we try to explain why this is the case.

### 3.1. A simulated data example

The first example consists of simulated data obtained from a mixture of two gamma distributions of the form

$$f(x) = 0.7 \text{Ga}(x|25, 5) + 0.3 \text{Ga}(x|64, 8). \quad (3.2)$$

Both components in the mixture  $f(x)$  have variance 1, but the weight in the mixture is different. This hypothetical situation occurs when there is a dominating group linked to a smaller group with slightly larger values; this situation would result in observing mainly values from the first group and few values from the other group.

Here we consider the case that 30 observations were available from the previous mixture, and density estimation was required. We set  $\sigma = 1$  for the kernel (3.1),  $\alpha(dy) = \text{Ga}(y|3, 0.5)dy$  for the gamma process, and  $\alpha(dx) = 0.1 \text{Ga}(x|3, 0.5)dx$  for the inverse Gaussian process. This implies that the mode of the prior distribution on the number of components is four for both the Dirichlet and the N-IG process. See Lijoi, Mena and Prünster (2005) for details on this strategy of tuning the distributions on the number of components in order to make the priors comparable. As a perturbing function we selected  $\bar{k}(y) = 10\mathbb{I}_{[0,3]}(y) + \mathbb{I}_{(3,\infty)}(y)$ , which significantly amplifies the jumps in the interval  $[0, 3]$ . Note that the means of the two components of (3.2) are 5 and 8, respectively.

Let us note that the prior guess at the shape of the distribution of the means in (2.1), which is given by the expected value of (2.2), corresponds to  $P_0(dy) = E[P(dy)] = \text{Ga}(y|3, 0.5)dy$  for the Dirichlet and N-IG processes (that is, with  $\kappa = 1$ ), whereas it is given by  $P_0(dy) = E[P(dy)] = 2.07\text{Ga}(y|3, 0.5)\mathbb{I}_{[0,3]}(dy) + 0.75\text{Ga}(y|3, 0.5)\mathbb{I}_{(3,\infty)}(dy)$  for the perturbed Dirichlet process, and by  $P_0(dy) = E[P(dy)] = 2.34\text{Ga}(y|3, 0.5)\mathbb{I}_{[0,3]}(dy) + 0.68\text{Ga}(y|3, 0.5)\mathbb{I}_{(3,\infty)}(dy)$  for the perturbed N-IG process. The two expression for  $P_0$  are determined using (A.3) and (A.4), respectively. The prior density estimates of the mixture (2.1) are then given by  $\hat{f}(x) = E[f(x)] = \int_{\mathbb{R}^+} (\sqrt{2\pi})^{-1} e^{-(x-y)^2/2} P_0(dy)$ . Figure 3.1 depicts the prior distribution on the means and the prior density estimate corresponding to the Dirichlet process and its perturbed version. The great influence of the perturbation on the prior structure is apparent: it induces the mixing measure to select means in the interval  $[0, 3]$  and the overall prior density estimate, though smooth, is evidently biased towards allocating mass in the region including  $[0, 3]$ . The effect of the perturbation on the N-IG process mixture is similar.

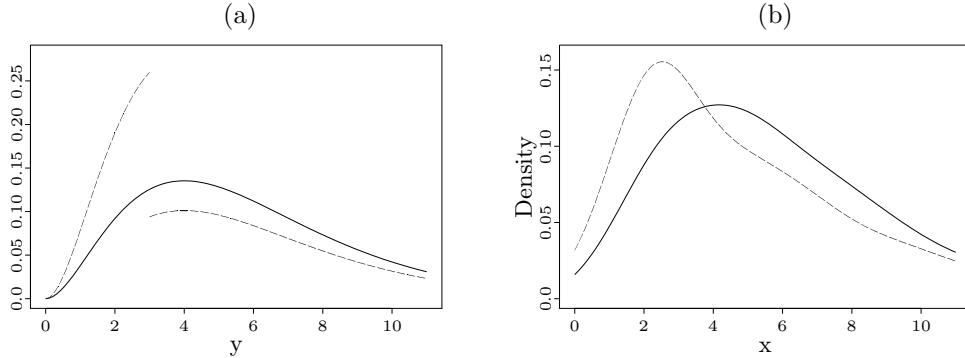


Figure 3.1. (a) Prior density on the means of the normal components: Dirichlet (solid line) and perturbed Dirichlet (dashed line) processes; (b) Prior density estimate of the mixture: Dirichlet (solid line) and perturbed Dirichlet (dashed line) processes.

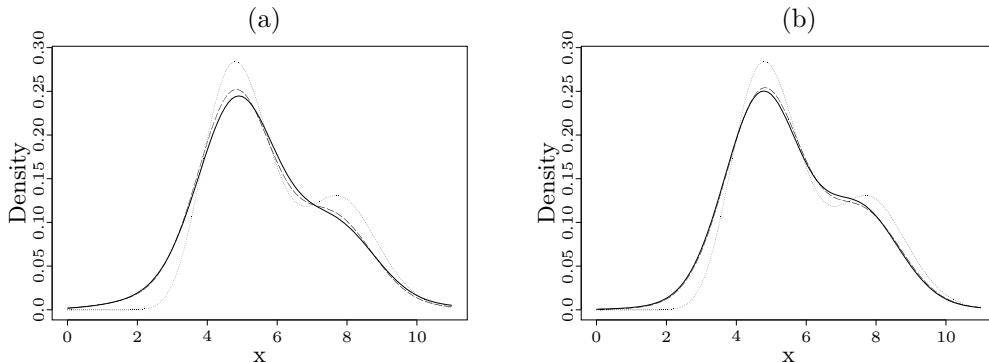


Figure 3.2. True density (dotted line) together with: (a) Posterior density estimates for Dirichlet (solid line) and perturbed Dirichlet (dashed line) mixtures; (b) Posterior density estimates for the N-IG (solid line) and perturbed N-IG (dashed line) mixtures.

Such a heavy perturbation leads one to expect that the posterior density estimates would be influenced by it. However, this appears not to be the case. Figure 3.2 displays the posterior density estimates for both the Dirichlet and N-IG mixtures, together with the density estimates arising from their perturbations.

Now we try to provide an explanation of why the posterior density estimate is not much affected by the perturbation. We do this by looking at the specific example at issue, but the behaviour is actually quite general. To understand the mechanism that neutralizes the perturbation we have to reason at the level of the latent variables that control the location of the means of the normal components. The key quantity is the latent variable  $U$ , whose density is given in (2.8).

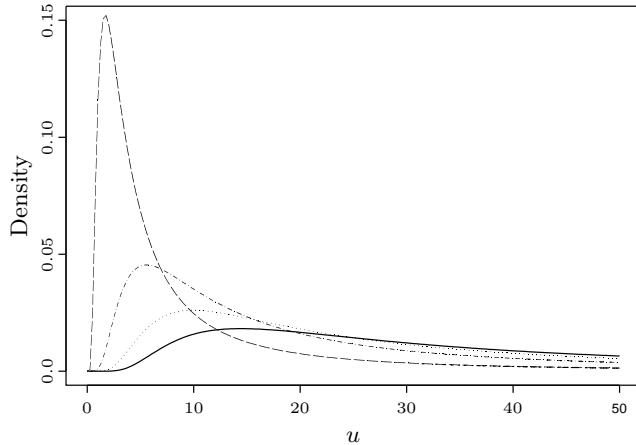


Figure 3.3. Distribution of  $U|Y$  for  $Y$  containing 0 (solid line), 10 (dotted line), 20 (dashed-dotted line) and 30 (dashed line) latent variables in  $[0, 3]$ .

First consider the case where  $A$  is a gamma process: it is easy to see that the distribution of  $U$  given the latent variables  $Y = (Y_1, \dots, Y_{30})$  is

$$f_{U|Y}(u) \propto u^{30-1} (1 + 10u)^{-[\alpha([0,3]) + s]} (1 + u)^{-[\alpha([0,3]^c) + 30 - s]}, \quad (3.3)$$

where  $s$  stands for the number of latent variables which fall into the interval  $[0, 3]$ . Moreover, since in our case  $\alpha(dy) = \text{Ga}(y|3, 0.5)dy$ , we have  $\alpha([0, 3]) \cong 0.19$ . Now focus attention on the set  $[0, 3]$  where the perturbation takes place. In the Dirichlet case this set has a prior probability  $P_0([0, 3]) \cong 0.19$ , whereas in the perturbed case the probability of selecting means in the set is doubled, since  $P_0([0, 3]) \cong 0.4$ . The question is now what is the posterior expected probability of the set  $[0, 3]$ , which clearly depends on the configuration of the latent variables. For illustrative purposes, we consider four cases, namely that 30, 20, 10 and none of the latent variables fall in  $[0, 3]$ . First note that the configuration of the latent variables heavily influences the shape of the density of  $U|Y$  in (3.3). Figure 3.3 displays (3.3) corresponding to the four choices of the latent variable configuration: the more the latent variables fall in  $[0, 3]$ , the more likely it is that (3.3) generates a small number. As an average representative of the distribution in (3.3) we take the median. For the four considered cases, we have medians equal to 5.1 when 30 latent variables belong to  $[0, 3]$ , 16.9 for 20, 29.7 for 10 and 42.6 for none.

Given this, we now compute the posterior probability of selecting a mean in  $[0, 3]$  conditional on the latent variable  $U|Y$  and on  $Y$ , which corresponds to

Table 1.  $E[P([0, 3])|U, Y]$  corresponding to different configurations of  $Y$  for the Dirichlet process and the perturbed Dirichlet process with different values for  $U$ .

# of latents in $[0, 3]$	Dirichlet	Perturbed Dirichlet						Median of $U Y$
		$u = 1$	$u = 5$	$u = 10$	$u = 20$	$u = 30$	$u = 50$	
30	0.974	0.984	0.971	0.970	0.970	0.969	0.968	0.971
20	0.651	0.767	0.684	0.663	0.655	0.650	0.648	0.657
10	0.328	0.465	0.362	0.343	0.335	0.328	0.325	0.329
0	0.006	0.011	0.007	0.006	0.005	0.004	0.003	0.004

calculating

$$E \left[ P([0, 3]) \mid U, Y \right] = E \left[ \frac{\int_{[0, 3]} \bar{k}(y) A^*(dy) + \sum_{Y_i^* \in [0, 3]} k(Y_i^*) J_i^*}{\int_{\mathbb{R}^+} \bar{k}(y) A^*(dy) + \sum_{Y_i^* \in \mathbb{R}^+} k(Y_i^*) J_i^*} \right], \quad (3.4)$$

where  $A^*$  is an IAP with intensity  $\nu^*(dy, dv) = e^{-\{1+u\bar{k}(y)\}v} v^{-1} dv \alpha(dy)$  and with fixed points of discontinuity at the distinct latent variables  $\{Y_1^*, \dots, Y_r^*\}$ , with jump heights  $J_i^*$  having the  $Ga(n_i, 1 + u\bar{k}(Y_j^*))$  distribution. As shown in James, Lijoi and Prünster (2009) for the unperturbed Dirichlet process, (3.4) is independent of  $U$  and coincides with the usual a posteriori estimate of the Dirichlet process, namely  $(\alpha([0, 3]) + s)(\alpha(\mathbb{R}^+) + n)^{-1}$ , where  $s$  stands, as before, for the number of latent variables taking value in  $[0, 3]$ . The first column of Table 1 reports the posterior estimates  $E[P([0, 3])|Y]$  for the Dirichlet process corresponding to different configurations of  $Y$ ; the following columns show the posterior expected values in (3.4) for the perturbed Dirichlet process for various values of  $U$  and different configurations of  $Y$ , and are obtained by exploiting the algorithm described in Section 2.3.

In Table 1, first note that small values of  $U$  tend to overestimate, whereas large values of  $U$  tend to underestimate, the posterior probability with respect to the unperturbed Dirichlet case. Second, there is an intermediate value of  $U$  (which heavily depends on the configuration of the latents) so that one matches the a posteriori probability of the unperturbed case. Indeed, the median values of the distributions (3.3) plugged into (3.4) lead to a discrepancy between perturbed and unperturbed Dirichlet of less than 1%. Now, since the median represents an average value of (3.3), the posterior probability of  $[0, 3]$  not conditioned on  $U$  and, consequently, also the density estimate in Figure 3.2 exhibits no real difference between the perturbed and the unperturbed Dirichlet mixtures. It is important to remark that if the distribution of  $U|Y$  did not produce sufficiently “good” values, where “good” depends on the configuration of  $Y$ , the effect of  $\bar{k}$  would still be evident in the posterior estimate.

At this point, it is natural to look at the values of  $U|Y$  generated inside the Gibbs sampler. Given that the posterior density estimates are very similar, we can deduce from the median value of  $U|Y$  the average configuration of the latent variables. The median value of  $U|Y$  is 27: this suggests that approximately ten latent variables fall within  $[0, 3]$ . Also, the histogram is quite similar to the density of  $U|Y$  with ten latents in  $[0, 3]$ . One may argue that around ten latents in  $[0, 3]$  is larger than what one would expect given the true density is (3.2); however, (3.2) is an unequal mixture with weight 0.7 on the first component, thus favoring small rather than large latent variables. Moreover, and more importantly, we are fitting a mixture of gamma distributions that are not symmetric and have mode smaller than the mean. For a normal mixture, to reproduce the behaviour of a gamma distribution near the origin, a shift of one of the normal components toward 0 is necessary.

Similar considerations can be made for the N–IG mixture with the proviso that, in this case, the distribution of  $U|Y$  depends not only on the location of the latents but also on the number of distinct latents  $r$  and, moreover, on the frequency of the classes  $n_1, \dots, n_r$ . A description of this more elaborate clustering mechanism is provided in Lijoi, Mena and Prünster (2005). However, the structure of  $U|Y$  is once again able to neutralize the effect of  $\bar{k}$ .

### 3.2. Galaxy data

For the second example we consider the well-known galaxy data, first analyzed in Roeder (1990). The data consists of the relative velocities, in thousands of kilometers per second, of 82 galaxies from six well-separated conic sections of space. These data have been analyzed by several authors, including Escobar and West (1995) and Lijoi, Mena and Prünster (2005). To analyze this dataset, we extend our model (2.1) to a semiparametric mixture by placing an independent hyper-prior on  $\sigma$ , namely  $\sigma \sim \text{Ga}(1, 2)$ . The Gibbs sampler is then extended to include the conditional posterior distribution for  $\sigma$ ,

$$\pi(\sigma|x, u, y, A_k) \propto \pi(\sigma) \prod_{i=1}^n k(x_i|y_i, \sigma),$$

and a Metropolis-Hastings step is added for simulating from it.

Here we compare the posterior density estimates corresponding to the N–IG and perturbed N–IG mixtures. To this end, set the parameter measure to be  $\alpha(dy) = 0.1 \text{Ga}(y|1, 0.01)dy$  and the perturbing function to  $\bar{k}(y) = 20 \mathbb{I}_{[11, 16]}(y) + \mathbb{I}_{[0, 11] \cup (16, \infty)}(y)$ . By (A.4), this implies that  $P_0(dy) = 4.13 \text{Ga}(y|1, 0.001)\mathbb{I}_{[11, 16]}(dy) + 0.86 \text{Ga}(y|1, 0.001)\mathbb{I}_{[0, 11] \cup (16, \infty)}(dy)$ . The prior and posterior density estimates arising from the N–IG and perturbed N–IG mixture are shown in Figure 3.4. It has to be noted that the prior estimate for the N–IG mixture is

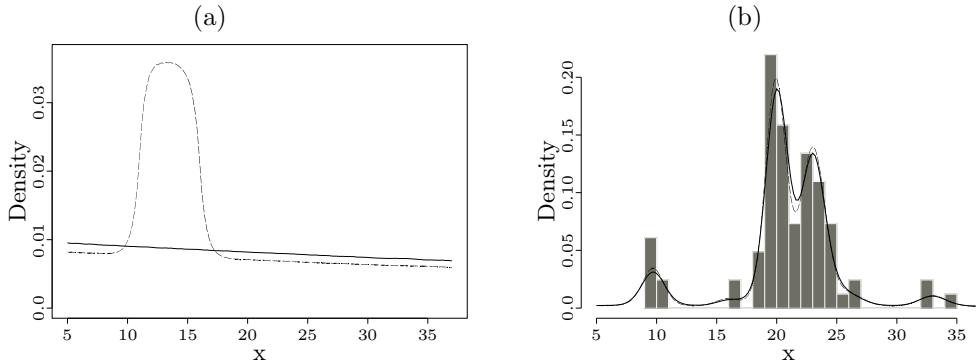


Figure 3.4. Galaxy data: (a) Prior density estimates for the N-IG (solid line) and perturbed N-IG (dashed line) mixtures; (b) Posterior density estimates for the N-IG (solid line) and perturbed N-IG (dashed line) mixtures together with histogram of the data.

essentially flat, whereas its perturbed version has a high peak around (11, 16]. In contrast the posterior density estimates are close and the posterior estimates for  $\sigma$  are similar, being 0.88 for the N-IG mixture and 0.76 for the perturbed N-IG mixture. When considering Dirichlet and perturbed Dirichlet mixtures, conclusions do not change all that much.

### 3.3. Acidity data

The third data we analyze concern the environmental problem of acidification. The data consist of measurements of an acid neutralizing capacity (ANC) index in a sample of 155 lakes in North-Central Wisconsin, USA; a low value of ANC can lead to a loss of biological resources. These data were studied by several authors, among others Richardson and Green (1997), and they were considered on a log-scale there, as here.

As with the galaxy data, we consider a semiparametric variation of the mixture model (2.1) and assign an independent gamma hyper-prior to  $\sigma$ ,  $\sigma \sim \text{Ga}(1, 10)$ . We compute the posterior density estimates corresponding to the Dirichlet and perturbed Dirichlet mixtures. To this end, set the parameter measure to be  $\alpha(dy) = \text{Ga}(y|5, 1)dy$  and the perturbing function to  $\bar{k}(y) = 0.1\mathbb{I}_{[4,5]}(y) + \mathbb{I}_{[0,4] \cup (5,\infty)}(y)$ . This choice, in contrast to the previous ones, has the effect of squeezing, by a factor of 10, the jumps of the process in the interval (4, 5]. From (A.3), we have that  $P_0(dy) = 0.32\text{Ga}(y|5, 1)\mathbb{I}_{[4,5]}(dy) + 1.16\text{Ga}(y|5, 1)\mathbb{I}_{[0,4] \cup (5,\infty)}(dy)$ . Figure 3.5. displays the prior and posterior density estimates arising from the Dirichlet and perturbed Dirichlet mixture. The perturbation creates a bimodal prior density estimate with minimum in the interval (4, 5]. However, the posterior density estimates do no exhibit any appreciable

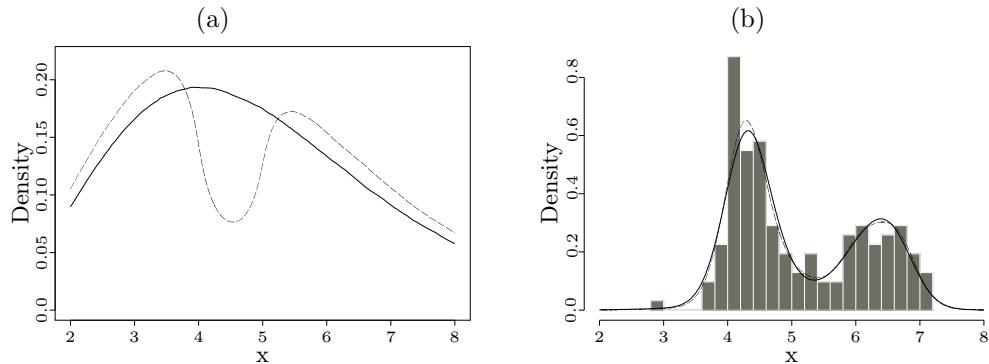


Figure 3.5. Acidity data: (a) Prior density estimates for the Dirichlet (solid line) and perturbed Dirichlet (dashed line) mixtures; (b) Posterior density estimates for the for the Dirichlet (solid line) and perturbed Dirichlet (dashed line) mixtures together with histogram of the data.

difference. The posterior estimates for  $\sigma$  are basically the same, 0.14 for the Dirichlet mixture and 0.12 for the perturbed Dirichlet mixture. Although not reported here, a similar behaviour arises when considering N-IG and perturbed N-IG mixtures.

#### 4. Conclusions

In this paper we have performed a new type of sensitivity analysis for Bayesian density estimators that consists of perturbing the nonparametric component in the mixture model by means of a suitable function. Such a perturbation heavily affects the prior structure of the model by increasing (or decreasing) the mass in certain regions. However, the model is able to absorb and neutralize such a perturbation in the posterior distribution by a quite interesting mechanism whose heuristics we have tried to describe. Being robust with respect to functional perturbations seems to represent a clear point in favor for mixtures of NRMI and, in particular, of their important special cases represented by the mixture of Dirichlet process and the mixture of normalized inverse Gaussian process. This clearly speaks also in favor of the use of dependent variations of these priors in a regression setup, where the fundamental ingredient is always a model like (1.3) with the Dirichlet replaced by a dependent random measure.

#### Acknowledgement

The authors are grateful to an associate editor and a referee for valuable comments and suggestions which lead to a clear improvement of the paper. Special thanks are due to Antonio Lijoi and Peter Müller for several helpful conversations.

L.E. Nieto-Barajas is supported by grant J48072-F from the National Council for Science and Technology of Mexico (CONACYT). I. Prünster acknowledges support from the Italian Ministry of University and Research (MiUR), grant 2006/133449.

## Appendix

**Proof of Proposition 1.** The strategy of the proof consists of deriving the posterior Laplace functional of  $\bar{k}(y)A(dy)$ , and it exploits the technique set forth in Prünster (2002). Let  $C_1, \dots, C_r$  be such that  $C_j \subset \mathbb{R}$  and  $C_i \cap C_j = \emptyset$  for all  $i \neq j$ . Then

$$\psi(\lambda, Y^*, C) = E \left\{ e^{-\int \lambda(s)\bar{k}(s)A(ds)} \mid (Y_1^*, \dots, Y_r^*) \in \times_{j=1}^r C_j^{n_j} \right\},$$

which takes the form

$$\frac{E \left[ e^{-\int \lambda(s)\bar{k}(s)A(ds)} \prod_{j=1}^r \left\{ \int_{C_j} \bar{k}(s)A(ds) \right\}^{n_j} \left\{ \int \bar{k}(s)A(ds) \right\}^{-n} \right]}{E \left[ \prod_{j=1}^r \left\{ \int_{C_j} \bar{k}(s)A(ds) \right\}^{n_j} \left\{ \int \bar{k}(s)A(ds) \right\}^{-n} \right]}.$$

We work on the numerator and denominator separately. Using the gamma identity  $\lambda^{-n} = \int_0^\infty u^{n-1} e^{-u\lambda} du / \Gamma(n)$ , the numerator can be expressed as

$$E \left[ \int_0^\infty \frac{u^{n-1}}{\Gamma(n)} e^{-\int \{\lambda(s)+u\}\bar{k}(s)A(ds)} \prod_{j=1}^r \left\{ \int_{C_j} \bar{k}(s)A(ds) \right\}^{n_j} du \right].$$

Now, noting that  $x^n = e^{ux}(-1)^n(d^n/dx^n)e^{-ux}$ , the numerator is

$$\int_0^\infty \frac{u^{n-1}}{\Gamma(n)} E \left( e^{-\int_{C_{r+1}} \{\lambda(s)+u\}\bar{k}(s)A(ds)} \prod_{j=1}^r \left\{ (-1)^{n_j} \frac{d^{n_j}}{du^{n_j}} e^{-\int_{C_j} \{\lambda(s)+u\}\bar{k}(s)A(ds)} \right\} \right) du,$$

where  $C_{r+1} = \mathbb{R} - \bigcup_{j=1}^r C_j$ . Considering that the regions  $C_1, \dots, C_{r+1}$  are disjoint we can apply the independence properties of the IAP and, using (2.3), the numerator is now expressed as

$$\begin{aligned} & \int_0^\infty \frac{u^{n-1}}{\Gamma(n)} \exp \left\{ - \int_{C_{r+1} \times (0, \infty)} \left( 1 - e^{-\{\lambda(s)+u\}\bar{k}(s)v} \right) \nu(ds, dv) \right\} \\ & \times \prod_{j=1}^r \left[ (-1)^{n_j} \frac{d^{n_j}}{du^{n_j}} \exp \left\{ - \int_{C_j \times (0, \infty)} \left( 1 - e^{-\{\lambda(s)+u\}\bar{k}(s)v} \right) \nu(ds, dv) \right\} \right] du. \end{aligned}$$

The denominator of  $\psi$  can be obtained from the numerator by setting  $\lambda(s) = 0$  for all  $s$ .

Now assuming that the  $C_j$ 's are of the form  $C_j = [y_j^* - \epsilon, y_j^* + \epsilon]$ , and recalling that  $\nu(ds, dv) = \rho(dv|s)\alpha(ds)$ , then taking the limit when  $\epsilon \rightarrow 0$ , we obtain

$$\begin{aligned} \lim_{\epsilon \rightarrow \infty} \psi(\lambda, Y^*, C) &= \lim_{\epsilon \rightarrow \infty} \mathbb{E} \left\{ e^{-\int \lambda(s) \bar{k}(s) A(ds)} \mid (Y_1^*, \dots, Y_r^*) \in \times_{j=1}^r [y_j^* - \epsilon, y_j^* + \epsilon] \right\} \\ &= \frac{\int_0^\infty u^{n-1} \mathbb{E} \left[ e^{-\int \{\lambda(s) + u\} \bar{k}(s) A(ds)} \right] \prod_{j=1}^r \left[ \int_0^\infty v^{n_j} e^{-\{\lambda(y_j^*) + u\} \bar{k}(y_j^*) v} \rho(dv|y_j^*) \right] du}{\int_0^\infty u^{n-1} \mathbb{E} \left\{ e^{-\int u \bar{k}(s) A(ds)} \right\} \prod_{j=1}^r \left\{ \int_0^\infty v^{n_j} e^{-u \bar{k}(y_j^*) v} \rho(dv|y_j^*) \right\} du}. \end{aligned}$$

Finally, multiplying and dividing by  $\mathbb{E} \left\{ e^{-\int u \bar{k}(s) A(ds)} \right\}$  inside the integral in the numerator, the previous expression becomes

$$\begin{aligned} &\int_0^\infty \exp \left\{ - \int_{\mathbb{R} \times (0, \infty)} \left( 1 - e^{-\lambda(s) \bar{k}(s)v} \right) e^{-u \bar{k}(s)v} \rho(dv|s) \alpha(ds) \right\} \\ &\quad \times \prod_{j=1}^r \int_0^\infty e^{-\lambda(y_j^*) \bar{k}(y_j^*) v} \frac{v^{n_j} e^{-u \bar{k}(y_j^*) v} \rho(dv|y_j^*)}{\tau_{n_j}(u|y_j^*)} \\ &\quad \times \frac{u^{n-1} \exp \left\{ - \int_{\mathbb{R} \times (0, \infty)} \left( 1 - e^{-u \bar{k}(s)v} \right) \rho(dv|s) \alpha(ds) \right\} \prod_{j=1}^r \tau_{n_j}(u|y_j^*) du}{\int_0^\infty u^{n-1} \exp \left\{ - \int_{\mathbb{R} \times (0, \infty)} \left( 1 - e^{-u \bar{k}(s)v} \right) \rho(dv|s) \alpha(ds) \right\} \prod_{j=1}^r \tau_{n_j}(u|y_j^*) du}, \end{aligned}$$

where  $\tau_{n_j}(u|y_j^*) = \int_0^\infty v^{n_j} e^{-u \bar{k}(y_j^*) v} \rho(dv|y_j^*)$ . This completes the proof.

**Details for the determination of  $\mathbf{P}_0(\cdot) = \mathbb{E}[\mathbf{P}(\cdot)]$ .** Using the gamma identity, the independence properties of  $A$ , and some suitable rearrangement, one has

$$\begin{aligned} \mathbb{E}[\mathbf{P}(B)] &= \mathbb{E} \left\{ \frac{\int_B \bar{k}(y) A(dy)}{\int_{\mathbb{R}} \bar{k}(s) A(ds)} \right\} = \int_0^\infty \mathbb{E} \left\{ \int_B \bar{k}(y) A(dy) e^{-u \int_{\mathbb{R}} \bar{k}(s) A(ds)} \right\} du \\ &= \int_0^\infty \mathbb{E} \left\{ \int_B \bar{k}(y) A(dy) e^{-u \int_B \bar{k}(s) A(ds)} \right\} \mathbb{E} \left\{ e^{-u \int_{B^c} \bar{k}(s) A(ds)} \right\} du \\ &= \int_0^\infty e^{-\psi_{\bar{k}}(u)} \int_B \tau_1(u|y) \bar{k}(y) \alpha(dy) du, \end{aligned} \tag{A.1}$$

where  $\tau_1$  and  $\psi_{\bar{k}}$  are defined as in point b) of Proposition 1. Let  $(B_1, \dots, B_m)$  be a partition of  $\mathbb{R}$  and  $\bar{k}(y) = \sum_{j=1}^m k_j \mathbb{I}_{B_j}(y)$  with  $k_j > 0$  for  $j = 1, \dots, m$ . Assume the IAP  $A$  is such that its Poisson intensity measure (2.4) can be written as  $\rho(dv)\alpha(dy)$ , which means that jump-sizes and jump-locations are independent. Set  $\psi_{k_j}(u) = \int_{\mathbb{R}^+} (1 - e^{-uk_j v}) \rho(dv)$  and  $\tau_{1,k_j}(u) = \int_0^\infty v e^{-uk_j v} \rho(dv)$ , then note that (A.1) becomes

$$\mathbb{E}[\mathbf{P}(B)] = \sum_{j=1}^m k_j \alpha(B \cap B_j) \int_0^\infty e^{-\sum_{i=1}^m \psi_{k_i}(u) \alpha(B_i)} \tau_{1,k_j}(u) du. \tag{A.2}$$

If  $A$  is a gamma process, (A.2) reduces to

$$E[P(B)] = \sum_{j=1}^m k_j \alpha(B \cap B_j) \int_0^\infty (1+uk_j)^{-\alpha(B_j)-1} \prod_{i=1, i \neq j}^m (1+uk_i)^{-\alpha(B_i)} du, \quad (\text{A.3})$$

whereas if  $A$  is an inverse Gaussian process, we have

$$E[P(B)] = \sum_{j=1}^m k_j \alpha(B \cap B_j) \int_0^\infty (1 + 2uk_j)^{-\frac{1}{2}} e^{-\sum_{i=1}^m (1+2uk_i)^{\frac{1}{2}} \alpha(B_i)} du. \quad (\text{A.4})$$

By specifying the forms for  $\bar{k}$  and  $\alpha$ , the previous two expressions allow an explicit evaluation of  $P_0$  via numerical integration.

## References

- Berry, D. A. and Christensen, R. (1979). Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet processes. *Ann. Statist.* **7**, 558-568.
- Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275-285.
- De Iorio, M., Müller, P., Rosner, G. and MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.* **99**, 205-215.
- Do, K.-A., Müller, P. and Tang, F. (2005). A Bayesian mixture model for differential gene expression. *J. Roy. Statist. Soc. Ser. C* **54**, 627-644.
- Doss, H. (1994). Bayesian nonparametric estimation for incomplete data via successive substitution sampling. *Ann. Statist.* **22**, 1763-1786.
- Dunson, D. B. and Park, J-H. (2008). Kernel stick breaking processes. *Biometrika* **95**, 307-323.
- Dunson, D. B., Xue, Y., and Carin, L. (2008). The matrix stick breaking process: Flexible Bayes meta analysis. *J. Amer. Statist. Assoc.* **103**, 317-327.
- Dykstra, R. L. and Laud, P. W. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356-367.
- Escobar, M. D. (1988). Estimating the means of several normal populations by nonparametric estimation of the distribution of the Means. Ph.D. thesis, Dept. of Statistics, Yale University.
- Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *J. Amer. Statist. Assoc.* **89**, 268-277.
- Escobar, M. D. and 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.
- Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics*, 287-302, Academic Press, New York.
- Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Statist.* **43**, 1634-1643.
- Griffin, J. E. (2007). The Ornstein—Uhlenbeck Dirichlet process and other time-varying processes for Bayesian nonparametric inference. *CRISM Working Papers* 07-03.

- Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161-173.
- James, L. F. (2006). Poisson calculus for spatial neutral to the right processes. *Ann. Statist.* **34**, 416-440.
- James, L. F., Lijoi, A. and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.* **36**, 76-97.
- Kottas, A., Müller, P. and Quintana, F. (2005). Nonparametric Bayesian modeling for multivariate ordinal data. *J. Comput. Graph. Statist.* **14**, 610-625.
- Lijoi, A., Mena, R. and Prünster, I. (2005). Hierarchical mixture modelling with normalized inverse Gaussian priors. *J. Amer. Statist. Assoc.* **100**, 1278-1291.
- Lijoi, A., Mena, R. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B* **69**, 715-740.
- Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimation. *Ann. Statist.* **12**, 351-357.
- MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statist. Assoc., Alexandria, VA.
- MacEachern, S. N. (2000). Dependent Dirichlet processes. Tech. Rep., Ohio State University.
- Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis, *Statist. Sci.* **19**, 95-110.
- Müller, P., Quintana, F. and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *J. R. Stat. Soc. B* **66**, 735-749.
- Müller, P. and Rosner, G. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *J. Amer. Statist. Assoc.* **92**, 1279-1292.
- Nieto-Barajas, L. E. and Prünster, I. (2008). Bayesian nonparametric estimation based on normalized measures driven mixtures: a general sampling scheme. Tech. Report.
- Nieto-Barajas, L. E., Prünster, I. and Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Statist.* **32**, 2343-2360.
- Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169-186.
- Petrone, S. and Raftery, A. E. (1997). A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability. *Statist. Probab. Lett.* **36**, 69-83.
- Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell* (Edited by Ferguson, T. S., et al.). Lecture Notes, Monograph Series, **30**, 245-267. IMS, Hayward.
- Pitman, J. (2003). Poisson-Kingman partitions. In *Science and Statistics: A Festschrift for Terry Speed* (Edited by Goldstein, D. R.). Lecture Notes, Monograph Series, **40**, 1-35. IMS, Hayward.
- Prünster, I. (2002). Random probability measures derived from increasing additive processes and their application to Bayesian statistics. Ph.D. thesis, University of Pavia.
- Quintana, F. A. and Iglesias, P. L. (2003). Bayesian clustering and product partition models. *J. Roy. Statist. Soc. Ser. B* **65**, 557-574.
- Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist.* **31**, 560-585.
- Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 731-792.

- Roeder, K. (1990). Density estimation with confidence sets exemplified by superclusters and voids in the galaxies. *J. Amer. Statist. Assoc.* **85**, 617-624.
- Sato, K. (1999). *Lévy processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.
- Walker, S. G. and Damien, P. (2000). Representations of Lévy processes without Gaussian components. *Biometrika* **87**, 477-483.

Departamento de Estadistica, ITAM, Rio Hondo 1, San Angel, 01000 Mexico, D.F.

E-mail: lnieto@itam.mx

Dipartimento di Statistica e Matematica Applicata, Università degli Studi di Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy.

E-mail: igor@econ.unito.it

(Received April 2007; accepted December 2007)