Hierarchical Mixture Modeling With Normalized Inverse-Gaussian Priors

Antonio LIJOI, Ramsés H. MENA, and Igor PRÜNSTER

In recent years the Dirichlet process prior has experienced a great success in the context of Bayesian mixture modeling. The idea of overcoming discreteness of its realizations by exploiting it in hierarchical models, combined with the development of suitable sampling techniques, represent one of the reasons of its popularity. In this article we propose the normalized inverse-Gaussian (N–IG) process as an alternative to the Dirichlet process to be used in Bayesian hierarchical models. The N–IG prior is constructed via its finite-dimensional distributions. This prior, although sharing the discreteness property of the Dirichlet prior, is characterized by a more elaborate and sensible clustering which makes use of all the information contained in the data. Whereas in the Dirichlet case the mass assigned to each observation depends solely on the number of times that it occurred, for the N–IG prior the weight of a single observation depends heavily on the whole number of ties in the sample. Moreover, expressions corresponding to relevant statistical quantities, such as a priori moments and the predictive distributions, are as tractable as those arising from the Dirichlet process. This implies that well-established sampling schemes can be easily extended to cover hierarchical models based on the N–IG process. The mixture of N–IG process and the mixture of Dirichlet process are compared using two examples involving mixtures of normals.

KEY WORDS: Bayesian nonparametrics; Density estimation; Dirichlet process; Inverse-Gaussian distribution; Mixture models; Predictive distribution; Semiparametric inference.

1. INTRODUCTION

Since the appearance of the article by Ferguson (1973), illustrating the statistical implications of the Dirichlet process, the Bayesian nonparametric literature has grown rapidly. Research has focused on both the analysis of various properties of the Dirichlet process and the proposal of alternative priors. In particular, Blackwell (1973) showed that the Dirichlet prior selects discrete distributions almost surely, which implies positive probability of ties in a sample drawn from it, as highlighted by Antoniak (1974). It is clear that such a property is undesirable when one is interested in modeling continuous data.

Various extensions of the Dirichlet process have been proposed to overcome this drawback. For instance, Mauldin, Sudderth, and Williams (1992) and Lavine (1992) introduced Pólya-tree priors. Although these can be constructed in such a way as to put probability one on continuous distributions, inference based on them depends heavily on the tree of partitions used for their construction. Paddock, Ruggeri, Lavine, and West (2003) tried to get rid of this unpleasant feature by making the partition itself random. The most fruitful approach to exploiting the Dirichlet process in inferential procedures is represented by the mixture of Dirichlet process (MDP), introduced by Lo (1984) and further developed by Ferguson (1983). This prior is concentrated on the space of densities, thus overcoming the discreteness problem, and it has been the focus of much attention during the 1990s as a consequence of the introduction of simulation-based inference, first developed by Escobar (1988). Relevant contributions include those by, among others, Escobar (1994), Escobar and West (1995), MacEachern (1994), MacEachern and Müller (1998), and Green and Richardson (2001). The model has been comprehensively reviewed by Dey, Müller, and Sinha (1998). The reason of the success of the MDP, as pointed out by Green and Richardson (2001), is that it exploits the discreteness of the Dirichlet process rather than combating it, thus providing a flexible model for clustering of items in a hierarchical setting.

This article proposes an alternative prior, the *normalized inverse-Gaussian* (N–IG) prior, to be used in the context of mixture modeling. Our proposed approach consists of the hierarchical model (or a suitable semiparametric variation of it) given by

$$\begin{aligned} (Y_i|X_i) &\overset{\text{ind}}{\sim} \mathscr{L}(Y_i|X_i), \qquad i = 1, \ldots, n, \\ (X_i|\tilde{P}) &\overset{\text{iid}}{\sim} \tilde{P}, \\ \tilde{P} &\sim \text{N–IG}. \end{aligned} \tag{1}$$

This represents a valid alternative to the corresponding model based on the Dirichlet process, because the N–IG prior, although sharing the discreteness property, turns out to have some advantages, whereas it preserves almost the same tractability as the Dirichlet process, it is characterized by a more elaborate clustering property that makes use of all of the information contained in the data. It is well known that the mass assigned to each observation by the Dirichlet process depends solely on the number of times that it occurs; in contrast, for the N–IG process, the weight given to a single observation depends heavily on the whole number of ties in the sample. For making inference with nonparametric hierarchical mixture models, the prior distribution of the number of distinct observations within the sample, which takes on the interpretation of distribution of the number of components in the mixture, is of great interest. Such a distribution, obtained by Antoniak (1974) in the Dirichlet case, admits a simple closed-form expression for the N–IG process. Using two examples involving mixtures of normals, we compare the behavior of the N–IG process mixture and the MDP.

Antonio Lijoi is Assistant Professor, Department of Economics and Quantitative Methods, University of Pavia, Pavia, Italy, and Research Associate at the Institute of Applied Mathematics and Computer Science (IMATI), National Research Council (CNR), Milan, Italy (E-mail: *lijoi@unipv.it*). Ramsés H. Mena is Associate Researcher, Research Institute for Applied Mathematics and Systems (IIMAS), National University of Mexico (UNAM), Mexico City, Mexico (E-mail: *ramses@sigma.iimas.unam.mx*). Igor Prünster is Assistant Professor, Department of Economics and Quantitative Methods, University of Pavia, Pavia, Italy, and Fellow of the International Center of Economic Research (ICER), Turin, Italy (E-mail: *igor.pruenster@unipv.it*) A. Lijoi and I. Prünster were supported in part by the Italian Ministry of University and Research (MIUR) research project "Bayesian Nonparametric Methods and Their Applications." The authors are grateful to two anonymous referees for valuable comments that have lead to a substantial improvement in the presentation. Special thanks also to Eugenio Regazzini for some helpful suggestions.

Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Theory and Methods
DOI 10.1198/016214505000000132

Like the Dirichlet process, the N–IG process is a special case of various classes of random probability measures. Indeed, it is a particular case of normalized random measure with independent increments (RMI). This class was introduced by Kingman (1975) and, subsequently studied by Perman, Pitman, and Yor (1992) and Pitman (2003). In a Bayesian context, it was first considered by Regazzini, Lijoi, and Prünster (2003), who generalized it also to nonhomogeneous measures (see also Prünster 2002). Further extensions and results have been given by James (2002). The N–IG process is also, when its parameter measure is nonatomic, a special case of species sampling model, a family of random probability measures introduced by Pitman (1996) and studied, in a Bayesian framework, by James (2002) and Ishwaran and James (2003). The N–IG and Dirichlet processes play special roles within these classes (and, indeed, within all random probability measures); to our knowledge, they are the only priors for which completely explicit expressions of their finite-dimensional distributions are available. Moreover, the N–IG process leads to explicit and tractable expressions for quantities of statistical relevance. What distinguishes the Dirichlet process from normalized RMIs and species-sampling models (and thus also from the N–IG process) is its conjugacy, as shown by Prünster (2002). However, this is no longer a problem, given the availability of suitable sampling schemes. It is worth noting that a posterior characterization of the N–IG process, in terms of a latent variable, can be deduced immediately from work of James (2002). Based on the families of discrete random probability measures mentioned earlier, recently general classes of mixture models have been proposed. Ishwaran and James (2001) proposed replacing the Dirichlet process by stick-breaking priors and study mixtures based on the two-parameter Poisson–Dirichlet process, due to Pitman (1995), in great detail. Ishwaran and James (2003) extended the analysis to cover species-sampling mixture models. Another proposal for generalizing MDP is represented by the normalized random measures driven by increasing additive processes due to Nieto-Barajas, Prünster, and Walker (2004) (see also Prünster 2002).

The outline of the article is as follows. We start with the construction of an N–IG process. We do this by first deriving the N–IG distribution on $(0, 1)$ and its multivariate analog on the $(n-1)$-dimensional simplex in Section 2, where we compute some of its moments as well. In Section 3 we show that there exists a random probability measure with these finite-dimensional distributions, which we term the N–IG process. Such a construction mimics the procedure adopted by Ferguson (1973) in defining the Dirichlet process. The role played by the gamma distribution in the Dirichlet case is attributed to the inverse-Gaussian distribution in our case. Assuming exchangeability of observations, we obtain the predictive distributions corresponding to an N–IG process. We further develop the analysis of this prior by illustrating the mechanism through which the weights are allocated to the observations. This represents a key feature for its potential success in applications. Moreover, we obtain a closed-form expression for the prior probability distribution, $p(\cdot|n)$, of the number of distinct observations in a sample of size n. In Section 4 we consider the mixture of N–IG process and show that well-known simulation techniques, such as those set forth by Escobar and West (1995) and Ishwaran and James (2001), can be extended in a straightforward way to deal with a mixture of N–IG process. For illustrative purposes, we consider simulated datasets, where observations are drawn from uniform mixtures of normals, and compare the performance of the mixture of N–IG process and the MDP in terms of both posterior probabilities on the correct number of components and log-Bayes factors. Finally, the extensively studied "galaxy" dataset is analyzed. Another fruitful use of discrete nonparametric priors in a continuous setting, is represented by the possibility of considering their filtered-variate version, as proposed by Dickey and Jiang (1998). This naturally leads to considering prior and posterior distributions of means of random probability measures. In Section 5 we provide a description of the posterior distribution of means of N–IG processes.

2. THE NORMALIZED INVERSE–GAUSSIAN DISTRIBUTION

The beta distribution and its multidimensional analog, the Dirichlet distribution, are commonly used as priors for the binomial and multinomial models. It is well known that, given n independent gamma random variables with $V_i \sim \mathrm{Ga}(\bar{\alpha}_i, 1)$, the Dirichlet distribution is defined as the distribution of the vector $(W_1, \ldots, W_n)$, where $W_i = V_i / \sum_{j=1}^n V_j$ for $i = 1, \ldots, n$. If $\bar{\alpha}_i > 0$ for every i, then the $(n-1)$-dimensional distribution of $(W_1, \ldots, W_{n-1})$ is absolutely continuous with respect to the product measure λ^{n-1}, where λ is the Lebesgue measure on the real line, and its density function on the simplex $\Delta_{n-1} = \{(w_1, \ldots, w_{n-1}) : w_i \geq 0, i = 1, \ldots, n-1, \sum_{i=1}^{n-1} w_i \leq 1\}$ is given by

$$f(w_1, \ldots, w_{n-1}) = \frac{\Gamma(\sum_{i=1}^n \bar{\alpha}_i)}{\prod_{i=1}^n \Gamma(\bar{\alpha}_i)} \left(\prod_{i=1}^{n-1} w_i^{\bar{\alpha}_i - 1} \right) \left(1 - \sum_{i=1}^{n-1} w_i \right)^{\bar{\alpha}_n - 1}. \quad (2)$$

Clearly, if $n = 2$, then (2) reduces to the beta density with parameter $(\bar{\alpha}_1, \bar{\alpha}_2)$ (see, e.g., Bilodeau and Brenner 1999).

In this section we derive a distribution on the unit interval and its multivariate analog on the simplex by substituting the gamma distribution with the inverse-Gaussian distribution in the foregoing construction. A random variable V has inverse-Gaussian distribution with shape parameter $\alpha \geq 0$ and scale parameter $\gamma > 0$, denoted by $V \sim \mathrm{IG}(\alpha, \gamma)$, if it has density with respect to the Lebesgue measure given by

$$f(v) = \frac{\alpha}{\sqrt{2\pi}} v^{-3/2} \exp\left\{ -\frac{1}{2} \left(\frac{\alpha^2}{v} + \gamma^2 v \right) + \gamma \alpha \right\}, \quad v \geq 0, \quad (3)$$

for $\alpha > 0$, whereas $V = 0$ (almost surely) if $\alpha = 0$. Seshadri (1993) has provided an exhaustive account of the inverse-Gaussian distribution. In what follows we assume, without loss of generality, that $\gamma = 1$.

Let $V_1, \ldots, V_n$ be independent random variables with $V_i \sim \mathrm{IG}(\alpha_i, 1)$, where $\alpha_i \geq 0$ for all i and $\alpha_i > 0$ for some i, $i = 1, \ldots, n$. We define the N–IG distribution with parameter $(\alpha_1, \ldots, \alpha_n)$, denoted by $\mathrm{N–IG}(\alpha_1, \ldots, \alpha_n)$, as the distribution of the random vector $(W_1, \ldots, W_n)$, where $W_i = V_i / \sum_{j=1}^n V_j$ for $i = 1, \ldots, n$. Clearly, if $\alpha_j = 0$ for some j, then $W_j = 0$ almost surely.

The following proposition provides the density function over Δ_{n-1} of an N–IG-distributed random vector.

Proposition 1. Suppose that the random vector $(W_1, \ldots, W_m)$ is N–IG$(\alpha_1, \ldots, \alpha_m)$ distributed. If $\alpha_i > 0$ for every $i = 1, \ldots, m$, then the distribution of $(W_1, \ldots, W_{m-1})$ is absolutely continuous with respect to λ^{m-1}, and its density function on Δ_{m-1} coincides with

$$\begin{aligned} f(w_1, \ldots, w_{m-1}) &= \frac{e^{\sum_{i=1}^m \alpha_i} \prod_{i=1}^m \alpha_i}{2^{m/2-1}\pi^{m/2}} \\ &\quad \times K_{-m/2}\left(\sqrt{\mathcal{A}_m(w_1, \ldots, w_{m-1})}\right) \\ &\quad \times \left({w_1}^{3/2} \cdots {w_{m-1}}^{3/2} \left(1 - \sum_{j=1}^{m-1} w_j\right)^{3/2} \right. \\ &\quad \left. \times \left[\mathcal{A}_m(w_1, \ldots, w_{m-1})\right]^{m/4} \right)^{-1}, \end{aligned} \tag{4}$$

where $\mathcal{A}_m(w_1, \ldots, w_{m-1}) = \sum_{i=1}^{m-1} \frac{\alpha_i^2}{w_i} + \alpha_m^2/(1 - \sum_{j=1}^{m-1} w_j)$ and K denotes the modified Bessel function of the third type. If $m = 2$, then (4) reduces to a univariate density on $(0, 1)$ of the form

$$f(w) = \frac{e^{\alpha_1+\alpha_2}\alpha_1\alpha_2}{\pi} \frac{K_{-1}(\sqrt{\mathcal{A}_2(w)})}{w^{3/2}(1-w)^{3/2}(\mathcal{A}_2(w))^{1/2}}. \tag{5}$$

One of the most important features of the Dirichlet distribution is the additive property inherited from the gamma distribution through which the Dirichlet distribution is defined. The same happens for the N–IG distribution in the sense that, if $(W_1, \ldots, W_n)$ is N–IG$(\alpha_1, \ldots, \alpha_n)$ distributed and $m_1 < m_2 < \cdots < m_k = n$ are integers, then the vector $(\sum_{i=1}^{m_1} W_i, \sum_{i=m_1+1}^{m_2} W_i, \ldots, \sum_{i=m_{k-1}+1}^{n} W_i)$ has distribution N–IG$(\sum_{i=1}^{m_1} \alpha_i, \sum_{i=m_1+1}^{m_2} \alpha_i, \ldots, \sum_{i=m_{k-1}+1}^{n} \alpha_i)$. This can be easily verified on the basis of the additive property of the inverse-Gaussian distribution.

A common desire is to associate to a probability distribution some descriptive indexes that summarize its characteristics. The following proposition provides some of these for the N–IG distribution. Set $a = \sum_{j=1}^n \alpha_j$ and $p_i = \alpha_i/a$ for every $i = 1, \ldots, n$.

Proposition 2. Suppose that the random vector $(W_1, \ldots, W_n)$ is N–IG$(\alpha_1, \ldots, \alpha_n)$ distributed; then

$$\begin{aligned} \mathrm{E}[W_i] &= p_i, \\ \mathrm{var}[W_i] &= p_i(1 - p_i)a^2 e^a \Gamma(-2, a), \end{aligned}$$

and

$$\mathrm{cov}(W_i, W_j) = -p_i p_j a^2 e^a \Gamma(-2, a) \quad \text{for } i \neq j,$$

where $\Gamma(\cdot, \cdot)$ denotes the incomplete gamma function.

It is worth noting that the moments corresponding to an N–IG distribution are quite similar to those of the Dirichlet distribution. Indeed, the structure is the same and they differ just by a multiplicative constant. Figures 1 and 2 depict various features of Dirichlet and N–IG distributions. For comparative purposes, the parameters have been chosen in such a way that their means and variances coincide. Recall that if a random vector is Dirichlet distributed, then $\mathrm{E}[W_i] = \bar{\alpha}_i/\bar{a} := \bar{p}_i$ and $\mathrm{var}[W_i] = \bar{p}_i(1 - \bar{p}_i)/(1 + \bar{a})$, for any $i = 1, \ldots, n$, having set $\bar{a} := \sum_{i=1}^n \bar{\alpha}_i$. Thus the means coincide if the $p_i = \bar{p}_i$, for any i, whereas given a, the variance match is achieved by solving for $\bar{a}$ the equation $a^2 e^a \Gamma(-2, a) = (\bar{a} + 1)^{-1}$. The plots seems to suggest a slightly lower concentration of the mass of the N–IG distribution around the barycenter of the simplex. Such a feature can be interpreted as the N–IG being less informative than the Dirichlet prior. This will become apparent in the sequel.

3. THE NORMALIZED INVERSE–GAUSSIAN PROCESS

3.1 Definition and Existence of the Normalized Inverse-Gaussian Process

In this section we define and show existence of the N–IG process. Toward this end, we suppose that the observations take values on some complete and separable metric space $\mathbb{X}$, endowed with its Borel σ-field $\mathscr{X}$. The N–IG process, $\tilde{P}$, is a random probability measure on $(\mathbb{X}, \mathscr{X})$ that can be defined via its family of finite-dimensional distributions, as was done by Ferguson (1973) for the Dirichlet process.

Let $\mathscr{P} = \{Q_{A_1,\ldots,A_n} : A_1, \ldots, A_n \in \mathscr{X}, n \geq 1\}$ be a family of probability distributions, and let α be a finite measure on $(\mathbb{X}, \mathscr{X})$ with $\alpha(\mathbb{X}) = a > 0$. If $\{A_1, \ldots, A_n\}$ is a measurable partition of $\mathbb{X}$, then set

$$Q_{A_1,\ldots,A_n}(C) = \int_{C \cap \Delta_{n-1}} f(v_1, \ldots, v_{n-1})\, dv_1 \cdots dv_{n-1}, \tag{6}$$

$\forall\, C \in \mathbb{R}_n$, where f is defined as in (4) with $\alpha_i = \alpha(A_i) > 0$, for all i. If $\{A_1, \ldots, A_n\}$ is not a partition of $\mathbb{X}$, then consider the partition $\{B_1, \ldots, B_m\}$ that it generates and set

$$Q_{A_1,\ldots,A_n}(C) = Q_{B_1,\ldots,B_m}\left(\left\{ (x_1, \ldots, x_m) \in [0, 1]^m : \left(\sum_{(1)} x_i, \ldots, \sum_{(n)} x_i \right) \in C \right\} \right),$$

where $\sum_{(j)}$ means that the sum extends over all indices $i \in \{1, \ldots, m\}$ for which $B_i \subset A_j$. We can show that $\mathscr{P}$ satisfies the following conditions:

(C1) For any $n \geq 1$ and any finite permutation π of $(1, \ldots, n)$,

$$Q_{A_1,\ldots,A_n}(C) = Q_{A_{\pi(1)},\ldots,A_{\pi(n)}}(\pi C), \qquad \forall\, C \in \mathscr{B}(\mathbb{R}^n),$$

where $\pi C = \{(x_{\pi(1)}, \ldots, x_{\pi(n)}) : (x_1, \ldots, x_n) \in C\}$.

(C2) $Q_{\mathbb{X}} = \delta_1$, where δ_x represents the point mass at x.

(C3) For any family of sets $\{A_1, \ldots, A_n\}$ in $\mathscr{X}$, let $\{D_1, \ldots, D_h\}$ be a measurable partition of $\mathbb{X}$ such that it is finer than the partition generated by $\{A_1, \ldots, A_n\}$. Then, for any $C \in \mathscr{B}(\mathbb{R}^n)$,

$$Q_{A_1,\ldots,A_n}(C) = Q_{D_1,\ldots,D_h}(C'),$$

where

$$C' = \left\{ (x_1, \ldots, x_h) \in [0, 1]^h : \left(\sum_{(1)} x_i, \ldots, \sum_{(n)} x_i \right) \in C \right\}.$$

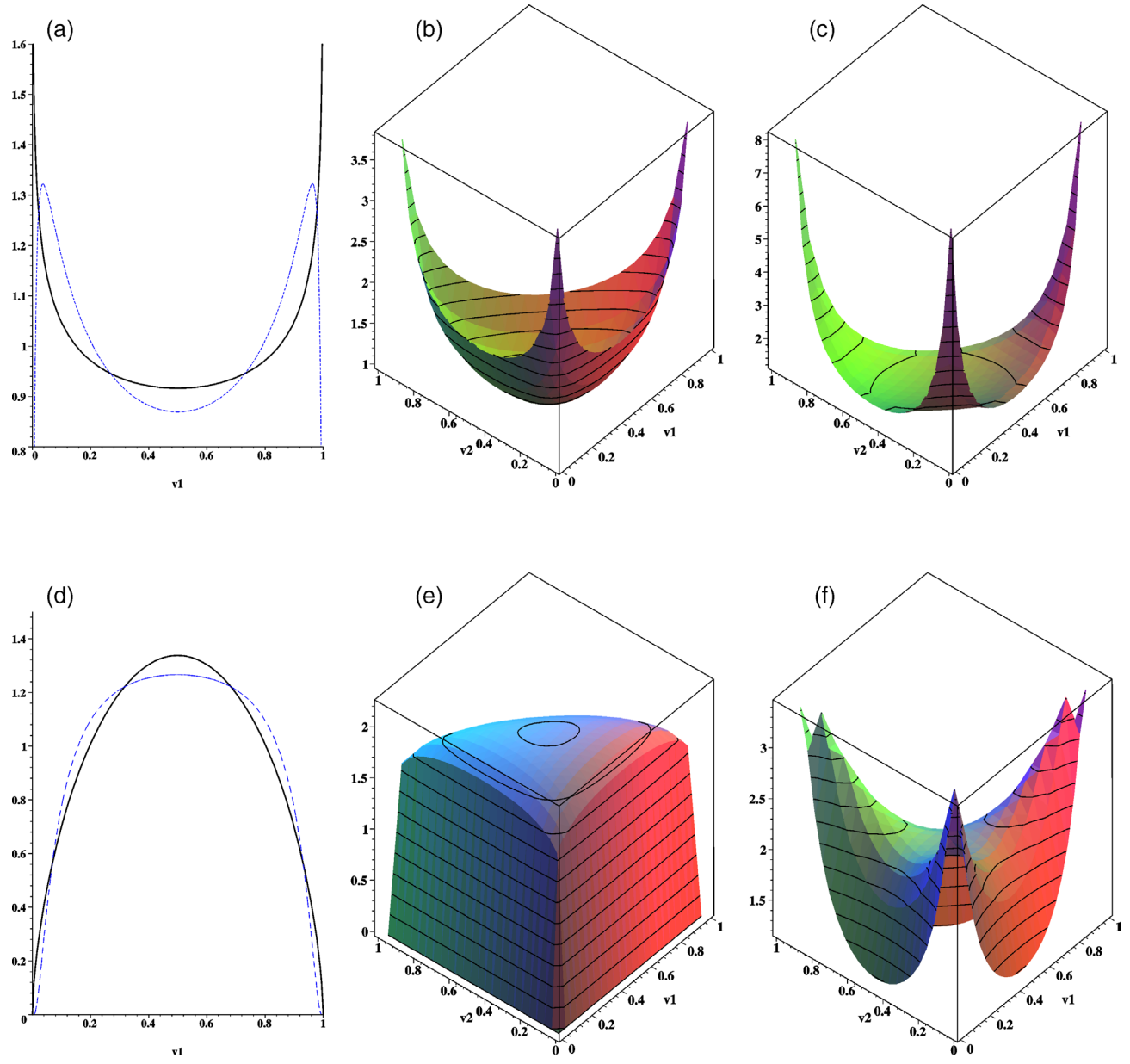

Figure 1. N–IG and Dirichlet Densities. (a) N–IG (·····) and Dirichlet (——) densities on (0, 1) with $p_1 = \bar{p}_1 = 1/2$, $a = .5$, and $\bar{a} = 1.737$. (b) Dirichlet density on Δ_2 with $\bar{p}_1 = \bar{p}_2 = 1/3$ and $\bar{a} = 1.737$. (c) N–IG density on Δ_2 with $p_1 = p_2 = 1/3$ and $a = .5$. (d) N–IG (·····) and Dirichlet (——) densities on (0, 1) with $p_1 = \bar{p}_1 = 1/2$, $a = 1.8$, and $\bar{a} = 3.270$. (e) Dirichlet density on Δ_2 with $\bar{p}_1 = \bar{p}_2 = 1/3$ and $\bar{a} = 3.270$. (f) N–IG density on Δ_2 with $p_1 = p_2 = 1/3$ and $a = 1.8$.

(C4) For any sequence $(A_n)_{n\geq 1}$ of measurable subsets of $\mathbb{X}$ such that $A_n \downarrow \emptyset$,

$$Q_{A_n} \Rightarrow \delta_0,$$

where the symbol $\Rightarrow$ denotes, as usual, weak convergence of a sequence of probability measures.

Hence, according to proposition 3.9.2 of Regazzini (2001), there exists a unique random probability measure admitting $\mathscr{P}$ as its family of finite-dimensional distributions. We term such a random probability measure $\tilde{P}$ an *N–IG process*.

3.2 Relationship to Other Classes of Discrete Random Probability Measures

In this section we discuss the relationship of the N–IG process to other classes of discrete random probability measures. First, recall that Ferguson (1973) also proposed an alternative construction of the Dirichlet process as a normalized gamma process. The same can be done in this case by replacing the gamma process with an inverse-Gaussian process, that is, an increasing Lévy process, $\xi := \{\xi_t : t \geq 0\}$, which is uniquely characterized by its Lévy measure, $\nu(\mathrm{d}v) = (2\pi v^3)^{-1/2} \mathrm{e}^{-v/2}\,\mathrm{d}v$. If α is a finite and nonnull measure on $\mathbb{R}$, then the time change $t = A(x) = \alpha((-\infty, x])$ yields a reparameterized process, ξ_A, which still has independent, but generally not stationary, increments. Because ξ_A is finite and positive almost surely, we are in a position to consider

$$\tilde{F}(x) = \frac{\xi_{A(x)}}{\xi_a} \quad \text{for every } x \in \mathbb{R} \tag{7}$$

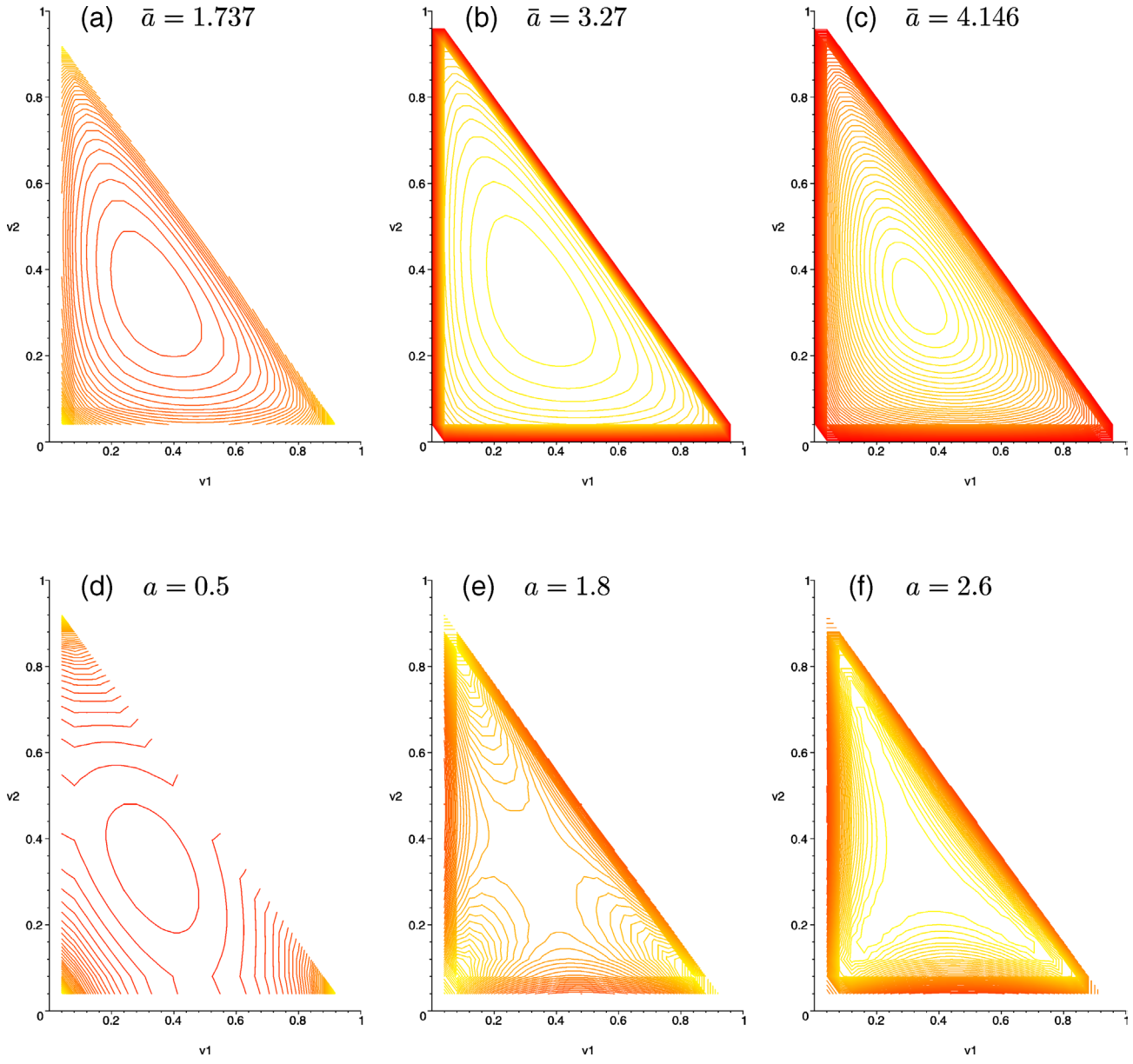

Figure 2. Contour Plots for the Dirichlet, (a)–(c), and N–IG, (d)–(f), Densities on Δ_2 With $p_1 = p_2 = \tilde{p}_1 = \tilde{p}_2 = 1/3$ for Different Choices of a and $\bar{a}$.

as a random probability distribution function on $\mathbb{R}$, where $a := \alpha(\mathbb{R})$. The corresponding random probability $\tilde{P}$ is an N–IG process on $\mathbb{R}$. As shown by Regazzini et al. (2003), such a construction holds for any increasing additive process, provided that the total mass of the Lévy measure is unbounded, giving rise to the class of normalized RMI. James (2002) extended normalized RMIs to Polish spaces. Thus the N–IG process represents clearly a particular case of normalized RMI. We remark that the idea of defining a general class of discrete random probability measures via normalization is due to Kingman (1975), and previous stimulating work was done by Perman et al. (1992) and Pitman (2003).

The N–IG process, provided that α is nonatomic, turns out to also be a species-sampling model. This class of random probability measures, due to Pitman (1996), is defined as

$$\tilde{P}(\cdot) = \sum_{i \geq 1} P_i \delta_{X_i}(\cdot) + \left(1 - \sum_{i \geq 1} P_i\right) H(\cdot), \qquad (8)$$

where $0 < P_i < 1$ are random weights such that $\sum_{i>1} P_i \leq 1$, independent of the locations X_i, which are iid with some nonatomic distribution H. A species-sampling model is completely characterized by H, which is the prior guess at the shape of $\tilde{P}$, and the so-called "exchangeable partition probability function" (EPPF) (see Pitman 1996). Although the definition of species-sampling model provides an intuitive and quite general framework, it leaves a difficult open problem, namely the concrete assignment of the random weights P_i. It is clear that such an issue is crucial for applicability of these models. Up to now, all tractable species sampling models have been defined by exploiting the so-called "stick-breaking" procedure, already adopted by Halmos (1944) and Freedman (1963). (See Ishwaran and James 2001 and references therein for examples of what they call "stick-breaking priors.") More recently, Pitman (2003) considered Poisson–Kingman models, a subclass of normalized RMI, which are defined via normaliza-

tion of homogeneous random measures with independent increments under the additional assumption of α being nonatomic. He showed that Poisson–Kingman models are a subclass of species-sampling models and characterized them by deriving an expression of the EPPF. From these considerations, it follows that an N–IG process, provided that α is nonatomic, is also a species-sampling model. Note that in the N–IG case, a closed-form expression of the EPPF can be derived; based on work of Pitman (2003) or James (2002) combined with the arguments used in the proof of Proposition 3, one obtains (A.1) in the Appendix. In this article we do not pursue the approach based on Poisson process partition calculus and refer the reader to the article by James (2002) for extensive calculus relative to EPPFs. We just remark that the N–IG process represents the first example (at least to our knowledge) of a tractable species-sampling model, which cannot be derived via a stick-breaking procedure.

Before proceeding, it seems worthwhile to point out that the peculiarities of the N–IG and Dirichlet processes compared with other members of these classes (and, indeed, within all random probability measures) is represented by the fact that their finite-dimensional distributions are known explicitly.

3.3 Some Properties of the Normalized Inverse-Gaussian Process

In this section we study some properties of the N–IG process. In particular, we discuss discreteness, derive some of its moments and the predictive distribution, and consider the distribution of the number of distinct observations with a sample drawn from an N–IG process.

A first interesting issue is the almost-sure discreteness of an N–IG process. On the basis of the construction provided in Section 3.1, it follows that the distribution of the N–IG process admits a version that selects discrete probability measures almost surely. By a result of James (2003) that makes use of the construction in (7), we have that all versions of $\tilde{P}$ are discrete. Hence we can affirm that the N–IG process is discrete almost surely.

The moments of an N–IG process with parameter α follow immediately from Proposition 2. Let $B, B_1, B_2 \in \mathscr{X}$ and set $C := B_1 \cap B_2$ and $P_0(\cdot) = \alpha(\cdot)/a$. Then

$$\mathrm{E}[\tilde{P}(B)] = P_0(B),$$

$$\mathrm{var}[\tilde{P}(B)] = P_0(B)(1 - P_0(B))a^2 \mathrm{e}^a\, \Gamma(-2, a),$$

and

$$\mathrm{cov}\big(\tilde{P}(B_1), \tilde{P}(B_2)\big) = [P_0(C) - P_0(B_1)P_0(B_2)]a^2 \mathrm{e}^a \Gamma(-2, a).$$

Note that if P_0 is diffuse, then the mean of $\tilde{P}$ follows also from the fact that the N–IG process is a species-sampling model. These quantities are usually exploited for incorporating real qualitative prior knowledge into the model. For instance, Walker and Damien (1998) suggested controlling, the prior guess P_0, as well as also the variance of the random probability measure at issue. Walker, Damien, Laud, and Smith (1999) provided a detailed discussion on the prior specification in nonparametric problems.

An important goal in inferential procedures is predicting future values of a random quantity based on its past outcomes. Suppose that a sequence $(X_n)_{n\geq 1}$ of exchangeable observations is defined in such a way that, given $\tilde{P}$, the X_i's are iid with distribution $\tilde{P}$. Moreover, let $X_1^*, \ldots, X_k^*$ denote the k distinct observations within the sample, $\mathbf{X}^{(n)} = (X_1, \ldots, X_n)$, with $n_j > 0$ terms being equal to X_j^*, for $j = 1, \ldots, k$, and such that $\sum_j n_j = n$. The next proposition provides the predictive distributions corresponding to an N–IG process.

Proposition 3. If $\tilde{P}$ is an N–IG process with diffuse parameter α, then the predictive distributions are of the form

$$\Pr\big(X_{n+1} \in B \big| \mathbf{X}^{(n)}\big) = w_0^{(n)} P_0(B) + w_1^{(n)} \sum_{j=1}^{k} (n_j - 1/2)\delta_{X_j^*}(B)$$

for every $B \in \mathscr{X}$, with

$$w_0^{(n)} = \frac{\sum_{r=0}^{n} \binom{n}{r} (-a^2)^{-r+1} \Gamma(k + 1 + 2r - 2n; a)}{2n \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \Gamma(k + 2 + 2r - 2n; a)}$$

and

$$w_1^{(n)} = \frac{\sum_{r=0}^{n} \binom{n}{r} (-a^2)^{-r+1} \Gamma(k + 2r - 2n; a)}{n \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \Gamma(k + 2 + 2r - 2n; a)}.$$

Note that the proof of Proposition 3 exploits a general result of Pitman (2003), later derived independently by means of a different technique by Prünster (2002) (see also James 2002).

Thus the predictive distribution is a linear combination of the prior guess P_0 and a weighted empirical distribution with explicit expression for the weights. Moreover, a generalized Pólya urn scheme follows immediately, which is given in (11). To highlight its distinctive features, it is useful to compare it with the prediction rule of the Dirichlet process. In the latter case, one has that X_{n+1} is new with probability $a/(a + n)$ and that it coincides with X_j^* with probability $n_j/(a + n)$ for $j = 1, \ldots, k$. Thus the probability allocated to previous observations is $n/(a + n)$ and does not depend on the number k of distinct observations within the sample. Moreover, the weight assigned to each X_j^* depends only on the number of observations equal to X_j^*, a characterizing property of the Dirichlet process that at the same time represents one if its drawbacks (see Ferguson 1974). In contrast, for the N–IG process, the mechanism is quite interesting and exploits all available information. Given $\mathbf{X}^{(n)}$, the $(n + 1)$st observation is new with probability $w_0^{(n)}$ and coincides with one of the previous with probability $(n - k/2)w_1^{(n)}$. In this case, the balancing between new and old observations takes the number of distinct values k into account; it is easy to verify numerically that $w_0^{(n)}$ is decreasing as k decreases and, thus one removes more mass from the prior if fewer distinct observations (i.e., more ties) are recorded. Moreover, allocation of the probability to each X_j^* is more elaborate than for the Dirichlet case; instead of increasing the weight proportionally to the number of ties, the probability assigned to X_j^* is reinforced more than proportionally each time a new tie has been recorded. This can be explained by the argument that the more often X_j^* is observed, the stronger is the statistical evidence in its favor and, thus it is sensible to reallocate mass toward it.

A small numerical example may clarify this point. When comparing the Dirichlet and N–IG processes, it is sensible to fix their parameter such that they have the same mean and variance, as was done in Section 2. Recall that this is tantamount

Table 1. Posterior Probability Allocation of the Dirichlet and N–IG Processes

$n = 10$	*Dirichlet process*	*N–IG process*
New observation	.1480	.2182
$X_1^* = .3$, $n_1 = 4$	.3408	.3420
$X_2^* = .1$, $n_2 = 3$	.2556	.2443
$X_3^* = .6$, $n_3 = 2$	.1704	.1466
$X_4^* = .5$, $n_4 = 1$	.0852	.0489

to requiring them to have the same prior guess P_0 and imposing $\operatorname{var}(\mathscr{D}_{\tilde{\alpha}}(B)) = \operatorname{var}(\tilde{P}(B))$ for every $B \in \mathscr{X}$, which is equivalent to $(\bar{a}+1)^{-1} = a^2 e^a \Gamma(-2, a)$, where $\bar{a}$ and a are the total masses of the parameter measures of the Dirichlet and N–IG processes.

To highlight the reinforcement mechanism, let $\mathbb{X} = [0, 1]$, P_0 equal to the uniform distribution with $a = .5$ and $\bar{a} = 1.737$. Suppose that one has observed $\mathbf{X}^{(10)}$ with four observations equal to .3, three observations equal to .1, two observations equal to .6, and one observation equal to .5. Table 1 gives the probabilities assigned by the predictive distribution of the Dirichlet and N–IG processes.

From this table, the reinforcement becomes apparent. In the Dirichlet case, having four ties in .3 means assigning to it four times the probability given to an observation that appeared once. For the N–IG process, things are very different; the mass assigned to the prior (i.e., to observing a new value) is higher, indicating the greater flexibility of the N–IG prior. Moreover, the probability assigned to .6 is more than two times the mass assigned to .5, the probability assigned to .1 is greater than $3/2$ times the mass given to .6, and so on. Having a sample with many ties means having significant statistical evidence, with reference to the associated clustering structure, and the N–IG process makes use of this information.

A further remarkable issue concerns determination of the prior distribution for the number k of distinct observations X_i^*'s, among the n being observed. Antoniak (1974) remarked that such a distribution is induced by the distribution assigned to the random probability measure and gave an expression for the Dirichlet case. The corresponding formula for the N–IG process is given in the following proposition.

Proposition 4. The distribution of the number of distinct observations k in a sample of size n is given by

$$p(k|n) = \binom{2n-k-1}{n-1} \frac{e^a(-a^2)^{n-1}}{2^{2n-k-1}\Gamma(k)} \times \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \times \Gamma(k+2+2r-2n;\, a) \tag{9}$$

for $k = 1, \ldots, n$.

Note that the expression for $p(k|n)$ in the Dirichlet case is

$$c_n(k) a^k \frac{\Gamma(a)}{\Gamma(a+n)}, \tag{10}$$

where $c_n(k)$ is the absolute value of a Stirling number of the first kind (for details, see Green and Richardson 2001). Unlike in (10), the evaluation of which can be achieved by resorting to recursive formulas for Stirling numbers, evaluation of the expression in (9) is straightforward. Table 2 compares the priors $p(k|n)$ corresponding to the Dirichlet and N–IG processes, in the same setup of Table 1.

In this toy example, note that the distribution $p(\cdot|n)$ induced by the N–IG prior is flatter than that corresponding to the Dirichlet process. This fact plays a role in the context of mixture modeling to be examined in the next section.

4. THE MIXTURE OF NORMALIZED INVERSE–GAUSSIAN PROCESS

In light of the previous results, it is clear that the reinforcement mechanism makes the N–IG prior an interesting alternative to the Dirichlet process. Nowadays the Dirichlet process is seldom used directly for assessing inferences; indeed, it is commonly exploited as the crucial building block in a hierarchical mixture model of type (1). Here we aim to study a mixture of N–IG process model as set in (1) and suitable semiparametric variations of it. The mixture of N–IG process represents a particular case of normalized random measures driven by increasing additive processes (see Nieto-Barajas et al. 2004) and also of species-sampling mixture models (see Ishwaran and James 2003). With respect to both classes, the mixture of N–IG process stands out for its tractability.

To exploit a mixture of N–IG process for inferential purposes, it is essential to derive an appropriate sampling scheme. In such a framework, knowledge of the predictive distributions, determined in Proposition 3, is crucial. Indeed, most of the simulation algorithms developed in Bayesian nonparametrics rely on variations of the Blackwell–MacQueen Pólya urn scheme and on the development of appropriate Gibbs sampling procedures. (See Escobar 1994; MacEachern 1994; Escobar and West 1995 for the Dirichlet case and Pitman 1996; Ishwaran and James 2001 for extensions to stick-breaking priors.) In the case of an N–IG process, it follows that the joint distribution of $\mathbf{X}^{(n)}$ can be characterized by the following generalized Pólya urn scheme. Let $Z_1, \ldots, Z_n$ be an iid sample from P_0. Then a sample $\mathbf{X}^{(n)}$ from an N–IG process can be generated as follows. Set $X_1 = Z_1$, and for $i = 2, \ldots, n$,

$$\left(X_i \middle| \mathbf{X}^{(i-1)}\right) = \begin{cases} Z_i & \text{with probability } w_0^{(i-1)} \\ X_{i,j}^* & \text{with probability} \\ & (n_j - 1/2) w_1^{(i-1)}, j = 1, \ldots, k_i, \end{cases} \tag{11}$$

Table 2. Prior Probabilities for k Corresponding to the Dirichlet and N–IG Processes

$n = 10$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
Dirichlet $\bar{a} = 1.737$	.0274	.1350	.2670	.2870	.1850	.0756	.0196	.0031	.00029	.00001
N–IG $a = .5$	.0523	.1286	.1813	.1936	.1716	.1296	.0824	.042	.0155	.0031

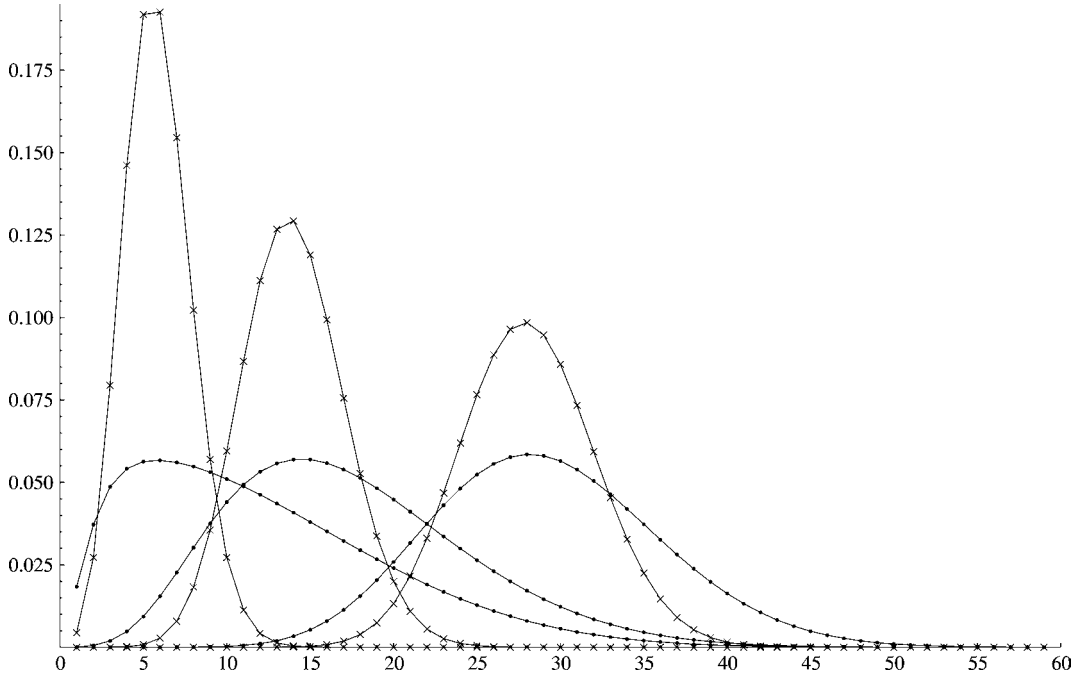

Figure 3. Prior Probabilities for k (n = 100) Corresponding to the N–IG Process and the Dirichlet Process for Different Values of a (ā). These values were chosen to match the modes at 6, 14, and 28 [⋆⋆ Dirichlet (ā = 1.2, 4.2, 12.6) (left ā = 1.2 to right ā = 12.6); ●● N–IG (a = .1, 1, 5) (left a = .1 to right a = 5)].

where k_i represents the number of distinct observations, denoted by $X^*_{i,1}, \ldots, X^*_{i,k_i}$, in $\mathbf{X}^{(i-1)}$, and $w_0^{(i-1)}$ and $w_1^{(i-1)}$ are as given in Proposition 3.

Moreover, the result in Proposition 4 can be interpreted as providing the prior distribution of the number k of components in the mixture, for a given sample size n. As in the Dirichlet case, a smaller total mass a yields a $p(\cdot|n)$ more concentrated on smaller values of k. This can be explained by the fact that a smaller a gives rise to a smaller $w_0^{(n)}$; that is, it generates new data with lower probability. However, the $p(\cdot|n)$ induced by the N–IG prior is apparently less informative than that corresponding to the Dirichlet process prior and thus is more robust with respect to changes in a. A qualitative illustration is given in Figure 3, where the distribution of k given $n = 100$ observations is depicted for the N–IG and Dirichlet processes. We connected the probability points by straight lines only for visual simplification. Note that the mode of the distribution of k corresponding to the N–IG never has probability larger than .07.

Recently, new algorithms have been proposed for dealing with mixtures. Extending the work of Brunner, Chan, James, and Lo (2001) on an iid sequential importance sampling procedure for fitting MDPs, Ishwaran and James (2003) proposed a generalized weighted Chinese restaurant algorithm that covers species-sampling mixture models. They formally derived the posterior distribution of a species-sampling mixture. To draw approximate samples from this distribution, they devised a computational scheme that requires knowledge of the conditional distribution of the species-sampling model, $\tilde{P}$, given the missing values $X_1, \ldots, X_n$. When feasible, such an algorithm has the merit of reducing the Monte Carlo error (see Ishawaran and James 2003 for details and further references). In the case of an N–IG process, sampling from its posterior law is not straightforward, because it is characterized in terms of a latent variable (see James 2002). Further investigations are needed to implement these interesting sampling procedures efficiently in the N–IG case and, more generally, for normalized RMI. Problems of the same type arise if one is willing to implement the scheme proposed by Nieto-Barajas et al. (2004).

To carry out a more detailed comparison between Dirichlet and N–IG mixtures, we next consider two illustrative examples.

4.1 Simulated Data

Here we consider simulated data sampled from uniform mixtures of normal distributions with different number of components and compare the behavior of the MDP and mixture of N–IG process for different choices of priors on the number of components. We then evaluate the performance, in terms of posterior probabilities, on the correct number of components and in terms of log-Bayes factors for testing $\{k = k^*\}$ against $\{k \neq k^*\}$, where k^* represents the correct number of components.

We first analyze a dataset $Y^{(100)}$, simulated from a mixture of three normal distributions with means -4, 0, and 8; variance 1; and corresponding weights .33, .34, and .33. Let $\mathrm{N}(\cdot|m, v)$ denote the normal distribution with mean m and variance $v > 0$. We consider the following hierarchical model to represent such data:

$$\begin{aligned}(Y_i|X_i) &\overset{\text{ind}}{\sim} \mathrm{N}(Y_i|X_i, 1), \qquad i = 1, \ldots, 100,\\ (X_i|\tilde{P}) &\overset{\text{iid}}{\sim} \tilde{P},\\ \tilde{P} &\sim \mathscr{P},\end{aligned}$$

where $\mathscr{P}$ refers to either an N–IG process or a Dirichlet process. Both are centered at the same prior guess, $\mathrm{N}(\cdot|\bar{Y}, t^2)$, where t is the data range, that is, $t = \max Y_i - \min Y_i$. To fix the total masses a and $\bar{a}$, we no longer consider the issue of

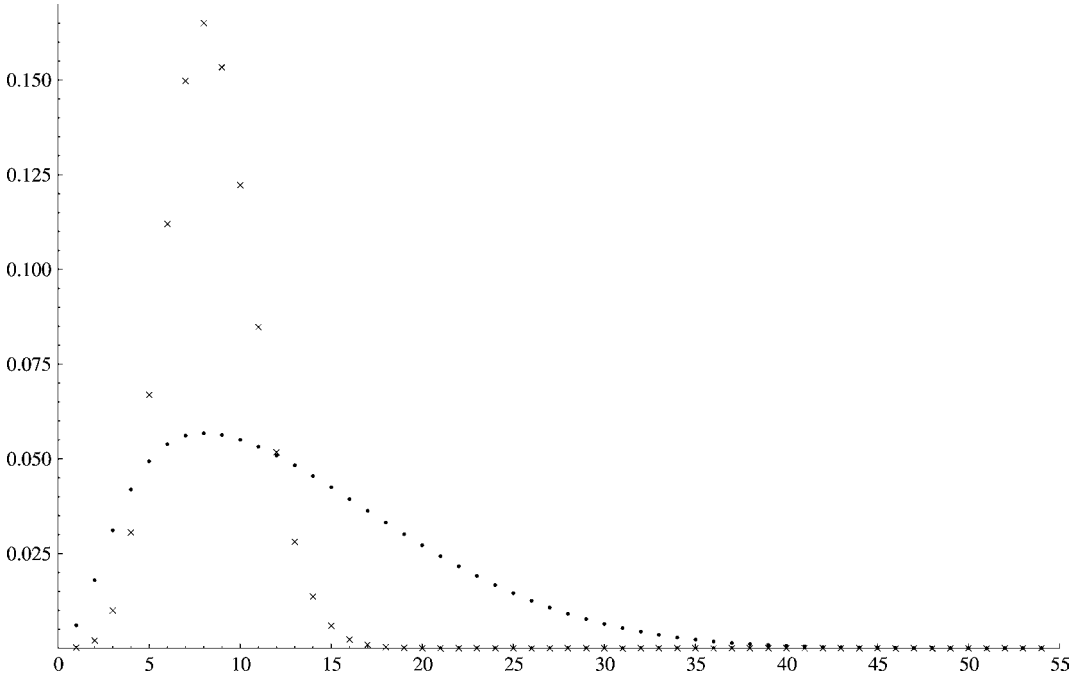

Figure 4. Prior Probabilities for k (n = 100) Corresponding to the Dirichlet and N–IG Processes [× ×Dirichlet ($\bar{a} = 2$); • •N–IG ($a = .2365$)].

matching variances. For comparative purposes in this mixture-modeling framework, it seems more plausible to start with similar priors for the number of components k. Hence a and $\bar{a}$ are such that the mode of $p(\cdot|n)$ is the same in both the Dirichlet and N–IG cases. We choose to set the mode equal to $k = 8$ in both cases, yielding $\bar{a} = 2$ and $a = .2365$. The corresponding distributions are plotted in Figure 4.

In this setup we draw a sample from the posterior distribution $\mathscr{L}(\mathbf{X}^{(100)}|Y^{(100)})$ by iteratively drawing samples from the distributions of $(X_i|\mathbf{X}_{-i}^{(100)}, Y^{(100)})$, for $i = 1, \ldots, 100$, given by

$$P\left(X_i \in \cdot \middle| \mathbf{X}_{-i}^{(100)}, Y^{(100)}\right) = q_{i,0}^* \mathrm{N}\left(X_i \in \cdot \middle| \frac{t^2 Y_i + \bar{Y}}{1+t^2}, \frac{t^2}{t^2+1}\right) + \sum_{j=1}^{k_i} q_{i,j}^* \delta_{X_j^*}(\cdot), \quad (12)$$

where $\mathbf{X}_{-i}^{(100)}$ is the vector $\mathbf{X}^{(100)}$ with the ith coordinate deleted and k_i refers to the number of X_j^*'s in $\mathbf{X}_{-i}^{(100)}$. The mixing proportions are given by

$$q_{i,0}^* \propto w_{i,0}^{(99)} \mathrm{N}(\cdot|\bar{Y}, t^2 + 1)$$

and

$$q_{i,j}^* \propto \left\{(n_j - 1/2) w_{i,1}^{(99)}\right\} \mathrm{N}(\cdot|X_j^*, 1),$$

subject to the constraint $\sum_{j=0}^{k_i} q_{i,j}^* = 1$. The values for $w_{i,0}^{(99)}$ and $w_{i,1}^{(99)}$ are given as in Proposition 3 when applied for determining the "predictive" distribution of X_i given $\mathbf{X}_{-i}^{(100)}$; they are computed with a precision of 14 decimal places. Note that in general, $w_{i,0}^{(99)}$ and $w_{i,1}^{(99)}$ depend only on a, n, and k_i and not on the different partition ξ of $n = n_{\xi_1} + \cdots + n_{\xi_k}$. This implies that a table containing the values for $w_{i,0}^{(99)}$ and $w_{i,1}^{(99)}$, for a given a, n, and $k = 1, \ldots, 99$ can be generated in advance for use in the Pólya urn Gibbs sampler. Further details are available on request from the authors.

To obtain our estimates, we resorted to the Pólya urn Gibbs sampler, such as the one set forth by Escobar and West (1995). As pointed out by Ishwaran and James (2001), such a generalized Pólya urn characterization also can be considered for the N–IG case, the only ingredients being the weights of the predictive distribution determined in Proposition 3. All of the following results were obtained by implementing the MCMC algorithm using 10,000 iterations, after 2,000 burn-in sweeps. Figure 5 depicts the posterior density estimates and the true density that generated the data.

The predictive density corresponding to the mixture of N–IG process clearly represents a more accurate estimate. Table 3 provides the prior and posterior probabilities of the number of components. The prior probabilities are computed according to Proposition 4, whereas the posterior probabilities follow from the MCMC output. One may note that the posterior mass assigned to $k = 3$ by the N–IG process is significantly higher than that corresponding to the MDP.

As far as the log-Bayes factors are concerned, one obtains $BF_{\mathscr{D}_\alpha} = 5.46$ and $BF_{\mathrm{N–IG}_\alpha} = 4.97$, thus favoring the MDP. The reason of this outcome seems to be due to the fact that the prior probability on the correct number of components is much lower for the MDP.

It is interesting to look also at the case in which the mode of $p(k|n)$ corresponds to the correct number of components $k^* = 3$. In such a case the prior probability on $k^* = 3$ is much higher for the MDP, being .285 against .0567 of the mixture of N–IG process. This results in a posterior probability on $k^* = 3$ of .9168 for the MDP and .86007 for the mixture of N–IG process, whereas the log-Bayes factors favor the N–IG mixture, being $BF_{\mathscr{D}_\alpha} = 3.32$ and $BF_{\mathrm{N–IG}_\alpha} = 4.63$. This again seems to be due to the fact that the prior probability of one process on the correct number of components is much lower than the one corresponding to the other.

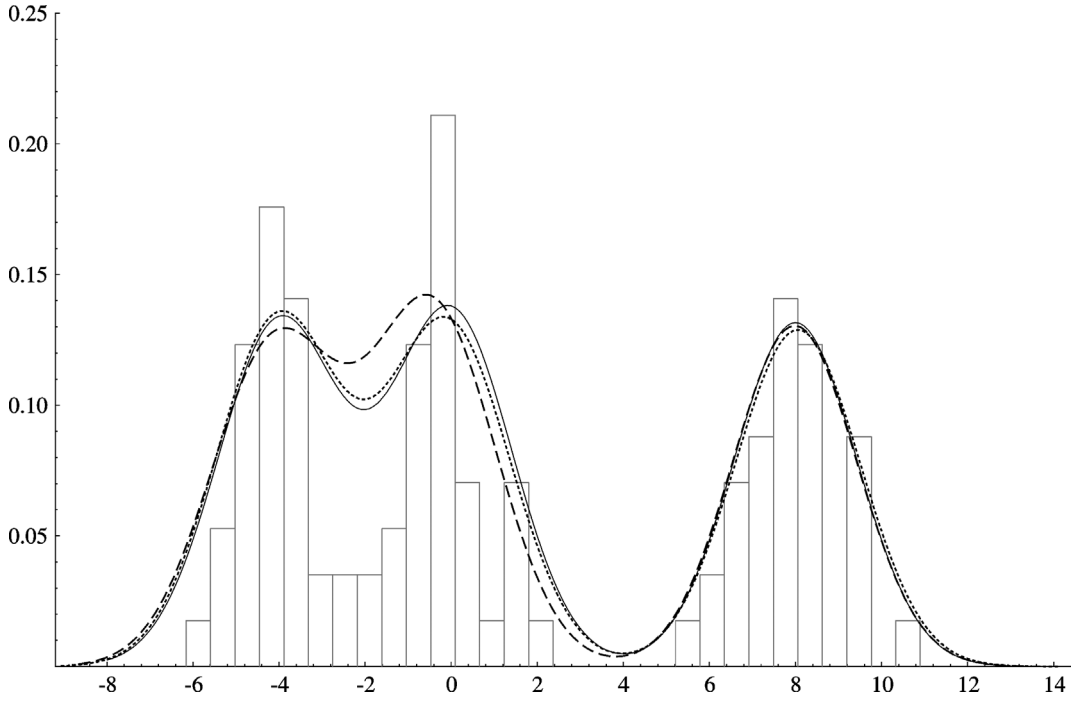

Figure 5. Posterior Density Estimates for the Mixture of Three Normals With Means −4, 0, 8, Variance 1, and Mixing Proportions .33, .34, and .33 [– – · Dirichlet ($\bar{a} = 2$); —— model; ····· N–IG ($a = .2365$)].

The next dataset, $Y^{(120)}$, that we deal with in this framework is drawn from a uniform mixture of six normal distributions with means $-10, -6, -2, 2, 6$, and 10 and variance .7. Both the Dirichlet and N–IG processes are again centered at the same prior guess, $N(\cdot|\bar{Y}, t^2)$, where t represents for the data range. Set $a = .01$ and $\bar{a} = .5$ such that $p(k|n)$ have mode at $k = 3$ for both the MDP and mixture of N–IG process. Such an example is meant to shed some light on the ability of the mixture of N–IG process to detect clusters when starting from a prior with mode at a small number of clusters and data coming from a mixture with more components. In this case the mixture of N–IG process performs better than the MDP both in terms of posterior probabilities, as shown in Table 4, and also in terms of log-Bayes factors, because $BF_{\mathscr{D}_\alpha} = 3.13$, whereas $BF_{\mathrm{N-IG}_\alpha} = 3.78$.

If one considers data drawn from mixtures with more components, then the mixture of N–IG process does systematically better in terms of posterior probabilities, but the phenomenon of "relatively low probability on the correct number of components" of the MDP reappears, leading it to be favored in terms of Bayes factors. For instance, we have taken 120 data from an uniform mixture of eight normals with means $-14, -10, -6, -2, 2, 6, 10$, and 14 and variance .7 and set $a = .01$ and $\bar{a} = .5$, so the $p(k|120)$ have mode at $k = 3$. This yields a priori probabilities on $k^* = 8$ of .00568 and .0489 and a posteriori probabilities of .0937 and .4338 for the MDP and mixture of N–IG process. In contrast, the log-Bayes factors are $BF_{\mathscr{D}_\alpha} = 2.90$ and $BF_{\mathrm{N-IG}_\alpha} = 2.73$. One can alternatively match the a priori probability on the correct number of components such that $p(8|120) = .04765$ for both priors. This happens if we set $\bar{a} = .86$, which results in the MDP having mode at $k = 4$, closer to the correct number $k^* = 8$, whereas the mixture of N–IG process remains unchanged. This results in a posterior probability on $k^* = 8$ for the MDP of .1134 and, having removed the influence of the low probability $p(8|120)$, in a log-Bayes factor of .94.

4.2 Galaxy Data

In this section we reanalyze the galaxy dataset popularized by Roeder (1990). These data, consisting of the relative velocities (in km/sec) of 82 galaxies, have been widely studied within the framework of Bayesian statistics (see, e.g., Roeder and Wasserman 1997; Green and Richardson 2001; Petrone and Veronese 2002). For comparison purposes, we use the following semiparametric setting, also analyzed by Escobar and West (1995):

$$\begin{aligned}(Y_i|m_i, V_i) &\overset{\text{ind}}{\sim} N(Y_i|m_i, V_i), \qquad i = 1, \ldots, 82,\\ (m_i, V_i|\tilde{P}) &\overset{\text{iid}}{\sim} \tilde{P},\\ \tilde{P} &\sim \mathscr{P},\end{aligned}$$

Table 3. Prior and Posterior Probabilities for k Corresponding to the Dirichlet and N–IG Processes Such That p(k|100) Has Mode at k = 8

$n = 100$		$k \le 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k \ge 9$
Dirichlet	Prior	.00225	.00997	.03057	.06692	.11198	.14974	.16501	.46356
$\bar{a} = 2$	Posterior		.7038	.2509	.0406	.0046	.0001		
N–IG	Prior	.02404	.0311	.0419	.04937	.05386	.05611	.05677	.68685
$a = .2365$	Posterior		.8223	.1612	.0160	.0005			

Table 4. Prior and Posterior Probabilities for k Corresponding to the Dirichlet and N–IG Processes Such That p(k|120) Has Mode at 3

$n = 120$		$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k \geq 9$
Dirichlet	Prior	.08099	.21706	.27433	.21954	.12583	.05532	.01949	.005678	.00176
$\bar{a} = .5$	Posterior				.0148	.3455	.5722	.064	.0033	.0002
N–IG	Prior	.04389	.05096	.05169	.05142	.05082	.04997	.0489	.04765	.6047
$a = .01$	Posterior					.0085	.6981	.2395	.0479	.006

where, as before, $\mathscr{P}$ represents either the N–IG process or the Dirichlet process. Again, the prior guess P_0 is the same for both nonparametric priors and is characterized by

$$P_0(A) = \int_A \mathrm{N}(x|\mu, \tau v^{-1})\mathrm{Ga}(v|s/2, S/2)\,\mathrm{d}x\,\mathrm{d}v,$$

where $A \in \mathscr{B}(\mathbb{R} \times \mathbb{R}^+)$ and $\mathrm{Ga}(\cdot|c, d)$ is the density corresponding to a gamma distribution with mean c/d. Similar to Escobar and West (1995), we assume a further hierarchy for μ and τ, namely $\mu \sim \mathrm{N}(\cdot|g, h)$ and $\tau^{-1} \sim \mathrm{Ga}(\cdot|w, W)$. We choose the hyperparameters for this illustration to fit those used by Escobar and West (1995), namely $g = 0$, $h = .001$, $w = 1$, and $W = 100$. For comparative purposes, we again choose a to achieve the mode match; if the mode of $p(\cdot|82)$ is in $k = 5$, as it is for the Dirichlet process with $\bar{a} = 1$, then $a = .0829$. From Table 5, it is apparent that the N–IG process provides a noninformative prior for k compared with the distribution induced by the Dirichlet process.

The details of the Pólya urn Gibbs sampler are as provided by Escobar and West (1995), with the only difference being the replacement of the weights with those corresponding to the N–IG process given in Proposition 3. Figure 6 shows the posterior density estimates based on 10,000 iterations considered after a burn–in period of 2,000 sweeps. Table 5 displays the prior and the posterior probabilities of the number of components in the mixture.

Some comments on the results in Table 5 are in order. Escobar and West (1995) extensively discussed the issue of learning about the total mass $\bar{a}$. This is motivated by the sensitivity of posterior inferences on the choice of the parameter $\bar{a}$. Hence they randomized $\bar{a}$ and specified a prior distribution for it. It seems worth noting that the posterior distribution of the number of components for the N–IG process, with a fixed a, essentially coincides with the corresponding distribution arising from an MDP with random $\bar{a}$, given in table 6 of Escobar and West (1995). Some explanations of this phenomenon might be that the prior $p(k|n)$ is essentially noninformative in a broad neighborhood of its mode, thus mitigating the influence of a on posterior inferences, and the aggressiveness in reducing/detecting clusters of the N–IG mixture, shown in the previous example, seems also to be a plausible reason of such a gain in "parsimony." However, for a mixture of N–IG process, it may be of interest to have a random a. To achieve this, a Metropolis step must be added in the algorithm. In principle, this is straightforward. The only drawback would be an increase in the computational burden, because computing the weights for the N–IG process is, after all, not as quick as computing those corresponding to the Dirichlet process.

5. THE MEAN OF A NORMALIZED INVERSE–GAUSSIAN PROCESS

An alternative use of discrete nonparametric priors for inference with continuous data is represented by histogram smoothing. Such a problem can be handled by exploiting the so-called "filtered-variate" random probability measures as defined by Dickey and Jiang (1998). These quantities essentially coincide with suitable means of random probability measures. Here we focus attention on means of N–IG processes. After the recent breakthrough achieved by Cifarelli and Regazzini (1990), this has become a very active area of research in Bayesian nonparametrics (see, e.g., Diaconis and Kemperman 1996; Regazzini, Guglielmi, and Di Nunno 2002 for the Dirichlet case and Epifani, Lijoi, and Prünster 2003; Hjort 2003; James 2002 for results beyond the Dirichlet case). In particular, Regazzini et al. (2003) dealt with means of normalized priors, providing conditions for their existence, their exact prior, and approximate posterior distribution. Here we specialize their general results to the N–IG process and derive the exact posterior density.

Letting $\mathbb{X} = \mathbb{R}$ and $\mathscr{X} = \mathscr{B}(\mathbb{R})$, we study the distribution of the mean $\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)$. The first issue to face is its finiteness. A necessary and sufficient condition for this to hold can be derived from proposition 1 of Regazzini et al. (2003) and is given by $\int_{\mathbb{R}} \sqrt{2\lambda x + 1}\alpha(\mathrm{d}x) < +\infty$ for every $\lambda > 0$.

As far as the problem of the distribution of a mean is concerned, proposition 2 of Regazzini et al. (2003) and some calculations lead to an expression of the prior distribution function of the mean as

$$\begin{aligned}\mathbb{F}(\sigma) = \frac{1}{2} - \frac{1}{\pi}\mathrm{e}^a \int_0^{+\infty} \frac{1}{t} \exp\Bigg\{ &- \int_{\mathbb{R}} \sqrt[4]{1 + 4t^2(x-\sigma)^2} \\ &\times \cos\left[\frac{1}{2}\arctan(2t(x-\sigma))\right]\alpha(\mathrm{d}x)\Bigg\} \\ &\times \sin\Bigg\{\int_{\mathbb{R}} \sqrt[4]{1 + 4t^2(x-\sigma)^2} \\ &\times \sin\left[\frac{1}{2}\arctan(2t(x-\sigma))\right]\alpha(\mathrm{d}x)\Bigg\}\,\mathrm{d}t,\end{aligned} \tag{13}$$

Table 5. Posterior Probabilities for k Corresponding to the Dirichlet and N–IG Processes in the Galaxy Data Example

$n = 82$		$k \leq 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$	$k \geq 11$
Dirichlet	Prior	.2140	.2060	.214	.169	.106	.0548	.0237	.0088	.0038
$\bar{a} = 1$	Posterior		.0465	.125	.2524	.2484	.2011	.08	.019	.0276
N–IG	Prior	.1309	.0617	.0627	.0621	.0607	.0586	.0562	.0534	.4537
$a = .0829$	Posterior	.003	.0754	.168	.2301	.2338	.1225	.0941	.0352	.0379

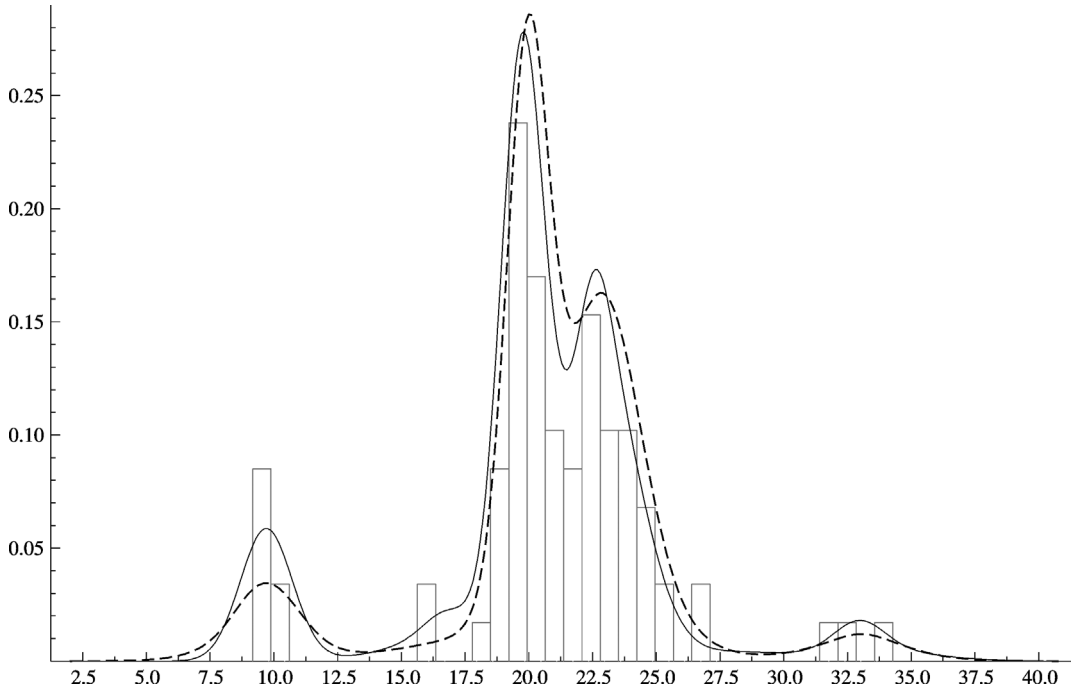

Figure 6. Posterior Density Estimates for the Galaxy Dataset [—— N–IG (a = .0829); – – - Dirichlet (a = 2)].

for any $\sigma \in \mathbb{R}$. Before providing the posterior density of a mean of an N–IG process, it is useful to introduce the Liouville–Weyl fractional integral, defined as

$$I^n_{c^+} h(\sigma) = \int_c^\sigma \frac{(\sigma - u)^{n-1}}{(n-1)!} h(u)\, \mathrm{d}u$$

for $n \geq 1$, and set equal to the identity operator for $n = 0$ for any $c < \sigma$. Moreover, let $\mathrm{Re}\, z$ and $\mathrm{Im}\, z$ denote the real and imaginary part of $z \in \mathbb{C}$.

Proposition 5. If $\tilde{P}$ is an N–IG process with diffuse parameter α and $-\infty \leq c = \inf \mathrm{supp}(\alpha)$, then the posterior density of the mean is of the form

$$\rho_{x^{(n)}}(\sigma) = \frac{1}{\pi} I^{n-1}_{c^+} \mathrm{Im}\, \psi(\sigma) \quad \text{if } n \text{ is even} \tag{14}$$

and

$$\rho_{x^{(n)}}(\sigma) = \frac{-1}{\pi} I^{n-1}_{c^+} \mathrm{Re}\, \psi(\sigma) \quad \text{if } n \text{ is odd}, \tag{15}$$

with

$$\psi(\sigma) = \frac{(n-1)!\, 2^{n-1} \int_0^\infty t^{n-1} \mathrm{e}^{-\int_{\mathbb{R}} \sqrt{1 - it2(x-\sigma)}\, \alpha(\mathrm{d}x)}}{a^{2n-(2+k)} \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \Gamma(k + 2 + 2r - 2n; a)} \times \prod_{j=1}^{k} \left(1 - it2(x_j^* - \sigma)\right)^{-n_j + 1/2} \mathrm{d}t.$$

6. CONCLUDING REMARKS

In this article we have studied some of the statistical implications of the use of the N–IG process as an alternative to the Dirichlet process. Both priors are almost surely discrete, allow for explicit derivation of their finite-dimensional distribution, and lead to tractable expressions of relevant quantities. Their natural use is in the context of mixture modeling, where they present remarkably different behaviors. Indeed, it has been shown that the prior distribution on the number of components, induced by the N–IG process, is wider than that induced by the Dirichlet process. This seems to mitigate the influence of the choice of the total mass a of the parameter measure. Moreover, the mixture of N–IG process seems to be more aggressive in reducing/detecting clusters on the basis of ties present in the data. Such conclusions are, however, preliminary and still need to be validated by more extensive applications, which will constitute the object of our future research.

APPENDIX: PROOFS

A.1 Proof of Proposition 1

Having (3) at hand, the computation of (4) is as follows. One operates the transformation $W_i = V_i (\sum_{i=1}^n V_i)^{-1}$, for $i = 1, \ldots, n-1$, and $W_n = \sum_{j=1}^n V_j$. Some algebra and formula 3.471.12 of Gradshteyn and Rhyzik (2000) leads to the desired result.

A.2 Proof of Proposition 2

In proving this result, we do not use the distribution of an N–IG random variable directly. The key point is exploiting the independence of the variables $(V_1, \ldots, V_n)$ used for defining $(W_1, \ldots, W_n)$. Set $V = \sum_{j=1}^n V_j$ and $V_{-i} = \sum_{j \neq i} V_j$, and recall that the moment-generating function of an $\mathrm{IG}(\alpha, \gamma)$ distributed random variable is given by $E[\mathrm{e}^{-\lambda X}] = \mathrm{e}^{-a(\sqrt{2\lambda + \gamma^2} - \gamma)}$. As far as the mean is concerned, for any $i = 1, \ldots, n$, we have

$$\begin{aligned}
\mathrm{E}[W_i] &= \mathrm{E}[V_i V^{-1}] \\
&= \int_0^{+\infty} \mathrm{E}[\mathrm{e}^{-uV_{-i}}] \mathrm{E}[\mathrm{e}^{-uV_i} V_i]\, \mathrm{d}u \\
&= \int_0^{+\infty} \mathrm{E}[\mathrm{e}^{-uV_{-i}}] \mathrm{E}\left[-\frac{\mathrm{d}}{\mathrm{d}u} \mathrm{e}^{-uV_i} \right] \mathrm{d}u \\
&= -\int_0^{+\infty} \mathrm{e}^{-(a-\alpha_i)(\sqrt{2u+1}-1)} \frac{\mathrm{d}}{\mathrm{d}u} \mathrm{e}^{-\alpha_i(\sqrt{2u+1}-1)}\, \mathrm{d}u
\end{aligned}$$

$$= \alpha_i \int_0^{+\infty} (2u+1)^{-1/2} \mathrm{e}^{-a(\sqrt{2u+1}-1)}\,\mathrm{d}u$$
$$= \frac{\alpha_i}{a} \int_0^{+\infty} -\frac{\mathrm{d}}{\mathrm{d}u} \mathrm{e}^{-a(\sqrt{2u+1}-1)}\,\mathrm{d}u = \frac{\alpha_i}{a} = p_i,$$

having applied Fubini's theorem and theorem 16.8 of Billingsley (1995). The variance and covariance can be deduced by similar arguments combined with integration by parts and some tedious algebra.

A.3 Proof of Proposition 3

From Pitman (2003) (see also James 2002; Prünster 2002), we can deduce that the predictive distribution associated with an N–IG process with diffuse parameter α is of the form

$$P\bigl(X_{n+1} \in \cdot \bigl| \mathbf{X}^{(n)}\bigr) = w^{(n)} \frac{\alpha(\cdot)}{a} + \frac{1}{n} \sum_{j=1}^{k} w_j^{(n)} \delta_{X_j^*}(\cdot),$$

with weights equal to

$$w^{(n)} = \frac{a \int_{\mathbb{R}^+} u^n \mathrm{e}^{-a(\sqrt{2u+1}-1)} \mu_{n_1}(u) \cdots \mu_{n_k}(u) \mu_1(u)\,\mathrm{d}u}{n \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-a(\sqrt{2u+1}-1)} \mu_{n_1}(u) \cdots \mu_{n_k}(u)\,\mathrm{d}u}$$

and

$$w_j^{(n)} = \frac{\int_{\mathbb{R}^+} u^n \mathrm{e}^{-a(\sqrt{2u+1}-1)} \mu_{n_1}(u) \cdots \mu_{n_j+1}(u) \cdots \mu_{n_k}(u)\,\mathrm{d}u}{\int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-a(\sqrt{2u+1}-1)} \mu_{n_1}(u) \cdots \mu_{n_k}(u)\,\mathrm{d}u},$$

where $\mu_n(u) := \int_{\mathbb{R}^+} v^n \mathrm{e}^{-uv} \nu(\mathrm{d}v)$ for any positive u and $n = 1, 2, \ldots$, and $\nu(\mathrm{d}v) = (2\pi v^3)^{-1/2} \mathrm{e}^{-v/2}\,\mathrm{d}v$. We can easily verify that $\mu_n(u) = 2^{n-1}\Gamma(n - \frac{1}{2})(\sqrt{\pi}[2u+1]^{n-1/2})^{-1}$ and, moreover, that

$$\int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-a(\sqrt{2u+1}-1)} \mu_{n_1}(u) \cdots \mu_{n_k}(u)\,\mathrm{d}u$$
$$= \frac{2^{n-k}\mathrm{e}^a}{(\sqrt{\pi})^k} \left\{ \prod_{j=1}^{k} \Gamma(n_j - 1/2) \right\} \int_0^{+\infty} \frac{u^{n-1} \mathrm{e}^{-a\sqrt{2u+1}}}{[2u+1]^{n-k/2}}\,\mathrm{d}u.$$

By the change of variable $\sqrt{2u+1} = y$, the latter expression is equal to

$$\frac{2^{n-k}\mathrm{e}^a}{(\sqrt{\pi})^k} \left\{ \prod_{j=1}^{k} \Gamma(n_j - 1/2) \right\} \frac{1}{2^{n-1}} \int_1^{+\infty} \frac{(y^2-1)^{n-1} \mathrm{e}^{-ay}}{y^{2n-k-1}}\,\mathrm{d}y$$
$$= \frac{\mathrm{e}^a}{2^{k-1}(\sqrt{\pi})^k} \left\{ \prod_{j=1}^{k} \Gamma(n_j - 1/2) \right\} \times \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^{n-1-r} \int_1^{+\infty} \frac{\mathrm{e}^{-ay}}{y^{2n-k-2r-1}}\,\mathrm{d}y$$
$$= \frac{\mathrm{e}^a}{2^{k-1}(\sqrt{\pi})^k} \left\{ \prod_{j=1}^{k} \Gamma(n_j - 1/2) \right\} \times \sum_{r=0}^{n-1} \binom{n-1}{r} (-1)^{n-1-r} a^{2n-2-k-2r} \Gamma(k+2r-2n+2; a).$$

Now the result follows by rearranging the terms appropriately combined with some algebra.

A.4 Proof of Proposition 4

According to the proof of Proposition 3, the joint distribution of the distinct observations $(X_1^*, \ldots, X_k^*)$ and the random partition coincides with

$$\alpha(\mathrm{d}x_1) \cdots \alpha(\mathrm{d}x_k) \frac{\mathrm{e}^a(-a^2)^{n-1}}{a^k 2^{k-1} \pi^{k/2} \Gamma(n)} \left\{ \prod_{j=1}^{k} \Gamma\left(n_j - \frac{1}{2}\right) \right\} \times \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \Gamma(k+2+2r-2n; a).$$

The conditional distribution of the vector $(k, n_1, \ldots, n_k)$, given n, can be obtained by marginalizing with respect to $(X_1^*, \ldots, X_k^*)$, thus yielding

$$\frac{\mathrm{e}^a(-a^2)^{n-1}}{2^{k-1}\pi^{k/2}\Gamma(n)} \left\{ \prod_{j=1}^{k} \Gamma\left(n_j - \frac{1}{2}\right) \right\} \times \sum_{r=0}^{n-1} \binom{n-1}{r} (-a^2)^{-r} \Gamma(k+2+2r-2n; a). \quad (A.1)$$

To determine $p(k|n)$, one needs to compute

$$\sum_{(*)} \prod_{j=1}^{k} \Gamma\left(n_j - \frac{1}{2}\right),$$

where $(*)$ means that the sum extends over all partitions of the set of integers $\{1, \ldots, n\}$ into k komponents,

$$\sum_{(*)} \prod_{j=1}^{k} \Gamma\left(n_j - \frac{1}{2}\right) = \pi^{k/2} \sum_{(*)} \prod_{j=1}^{k} \left(\frac{1}{2}\right)_{n_j - 1} = \binom{2n-k-1}{n-1} \frac{\Gamma(n)}{\Gamma(k)} 2^{2k-2n},$$

where, for any positive b, $(b)_n = \Gamma(b+n)/\Gamma(b)$, and the result follows.

A.5 Proof of Proposition 5

Let $\Pi_m = \{A_{m,i} : i = 1, \ldots, k_m\}$ be a sequence of partitions constructed in a similar fashion as was done in section 4 of Regazzini et al. (2003). We accordingly discretize α by setting $\alpha_m = \sum_{i=1}^{k_m} \alpha_{m,i}\,\delta_{s_{m,i}}$, where $\alpha_{m,i} = \alpha(A_{m,i})$ and $s_{m,i}$ is a point in $A_{m,i}$. Hence, whenever the jth element in the sample X_j belongs to A_{m,i_j} it is as if we observed s_{m,i_j}. Then, if we apply proposition 3 of Regazzini et al. (2003), some algebra leads to express the posterior density function for the discretized mean as (14) and (15) with

$$\psi_m(\sigma) = -C_m\bigl(x^{(n)}\bigr) \frac{\mathrm{e}^a 2^{n-k}}{\sqrt{\pi}^k} \left(\prod_{j=1}^{k} \alpha_{m,i_j} \Gamma(n_j - 1/2) \right) \times \int_0^{+\infty} \Bigl(t^{n-1} \mathrm{e}^{-\sum_{j=1}^{k_m} \sqrt{1-it2(s_j-\sigma)}\alpha_{m,j}} \Bigr) \times \left(\prod_{r=1}^{k} \Bigl[\bigl(1 - it2\bigl(s_{i_r} - \sigma\bigr)\bigr)^{n_r - 1/2} + g_m\bigl(\alpha_{m,i_r}, s_{m,i_r}, t\bigr) \Bigr] \right)^{-1} \mathrm{d}t, \quad (A.2)$$

where $C_m(x^{(n)})^{-1}$ is the marginal distribution of the discretized sample and $g_m(\alpha_{m,i_r}, s_{m,i_r}, t) = O(\alpha_{m,i_r}^2)$ as $m \to +\infty$, for any t. On the basis of proposition 4 of Regazzini et al. (2003), one could use the

previous expression as an approximation to the actual distribution, because it converges almost surely, in the weak sense, along the tree of partitions Π_m. But in the particular N–IG case, we are able to compute $C_m(x^{(n)})^{-1}$ by mimiking the technique used in Proposition 1 and exploiting the diffuseness of α. Thus we have

$$C_m\big(x^{(n)}\big)^{-1} = \frac{\mathrm{e}^a 2^{n-k}}{(n-1)!\sqrt{\pi}^k}\left(\prod_{j=1}^{k}\alpha_{m,i_j}\Gamma(n_j-1/2)\right) \times \int_{(0,+\infty)} \frac{u^{n-1}\mathrm{e}^{-a\sqrt{2u+1}}}{\prod_{r=1}^{k}[[1+2u]^{n_r-1/2}+h_m(\alpha_{m,i_r},u)]}\,\mathrm{d}u,$$

where $h_m(\alpha_{m,i_r},u) = O(\alpha_{m,i_r}^2)$ as $m\to+\infty$, for any u. Insert the previous expression in (A.2) and simplify. Then apply theorem 35.7 of Billingsley (1995) and dominated convergence to obtain that the limiting posterior density is given by (14) and (15) with

$$\psi(\sigma) = -\frac{(n-1)!\int_0^{\infty} t^{n-1}\mathrm{e}^{-\int_{\mathbb{R}}\sqrt{1-it2(x-\sigma)}\alpha(\mathrm{d}x)}}{\int_0^{+\infty}u^{n-1}\mathrm{e}^{-a\sqrt{2u+1}}(\sqrt{2u+1})^{-n+k/2}\,\mathrm{d}u} \times \prod_{j=1}^{k}\big(1-it2(x_j^*-\sigma)\big)^{-n_j+1/2}\,\mathrm{d}t.$$

Now arguments similar to those in the proof of Proposition 3 lead to the desired result.

[Received February 2004. Revised December 2004.]

REFERENCES

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152–1174.

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley.

Bilodeau, M., and Brenner, D. (1999), *Theory of Multivariate Statistics*, New York: Springer-Verlag.

Blackwell, D. (1973), "Discreteness of Ferguson Selections," *The Annals of Statistics*, 1, 356–358.

Brunner, L. J., Chan, A. T., James, L. F., and Lo, A. Y. (2001), "Weighted Chinese Restaurant Processes and Bayesian Mixture Models," unpublished manuscript.

Cifarelli, D. M., and Regazzini, E. (1990), "Distribution Functions of Means of a Dirichlet Process," *The Annals of Statistics*, 18, 429–442.

Dey, D., Müller, P., and Sinha, D. (eds.) (1998), *Practical Nonparametric and Semiparametric Bayesian Statistics*, Lecture Notes in Statistics 133, New York: Springer-Verlag.

Diaconis, P., and Kemperman, J. (1996), "Some New Tools for Dirichlet Priors," in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, New York: Oxford University Press, pp. 97–106.

Dickey, J. M., and Jiang, T. J. (1998), "Filtered-Variate Prior Distributions for Histogram Smoothing," *Journal of the American Statistical Association*, 93, 651–662.

Epifani, I., Lijoi, A., and Prünster, I. (2003), "Exponential Functionals and Means of Neutral-to-the-Right Priors," *Biometrika*, 90, 791–808.

Escobar, M. D. (1988), "Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means," unpublished doctoral dissertation, Yale University, Dept. of Statistics.

——— (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268–277.

Escobar, M. D., and 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," *The Annals of Statistics*, 1, 209–230.

——— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

——— (1983), "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics*, eds. M. H. Rizvi, J. S. Rustagi, and D. Siegmund, New York: Academic Press, pp. 287–302.

Freedman, D. A. (1963), "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case," *The Annals of Mathematical Statistics*, 34, 1386–1403.

Gradshteyn, I. S., and Ryzhik, J. M. (2000), *Table of Integrals, Series, and Products* (6th ed.), New York: Academic Press.

Green, P. J., and Richardson, S. (2001), "Modelling Heterogeneity With and Without the Dirichlet Process," *Scandinavian Journal of Statistics*, 28, 355–375.

Halmos, P. (1944), "Random Alms," *The Annals of Mathematical Statistics*, 15, 182–189.

Hjort, N. L. (2003), "Topics in Non-Parametric Bayesian Statistics," in *Highly Structured Stochastic Systems*, eds. P. J. Green et al., Oxford, U.K.: Oxford University Press, pp. 455–478.

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of American Statistical Association*, 96, 161–173.

——— (2003), "Generalized Weighted Chinese Restaurant Processes for Species Sampling Mixture Models," *Statistica Sinica*, 13, 1211–1235.

James, L. F. (2002), "Poisson Process Partition Calculus With Applications to Exchangeable Models and Bayesian Nonparametrics," *Mathematics ArXiv*, math.PR/0205093.

——— (2003), "A Simple Proof of the Almost Sure Discreteness of a Class of Random Measures," *Statistics & Probability Letters*, 65, 363–368.

Kingman, J. F. C. (1975), "Random Discrete Distributions," *Journal of the Royal Statistical Society*, Ser. B, 37, 1–22.

Lavine, M. (1992), "Some Aspects of Pólya Tree Distributions for Statistical Modelling," *The Annals of Statistics*, 20, 1222–1235.

Lo, A. Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I Density Estimates," *The Annals of Statistics*, 12, 351–357.

Mauldin, R. D., Sudderth, W. D., and Williams, S. C. (1992), "Pólya Trees and Random Distributions," *The Annals of Statistics*, 20, 1203–1221.

MacEachern, S. N. (1994), "Estimating Normal Means With a Conjugate Style Dirichlet Process Prior," *Communications in Statistics Simulation and Computation*, 23, 727–741.

MacEachern, S. N., and Müller, P. (1998), "Estimating Mixture of Dirichlet Process Models," *Journal of Computational and Graphical Statistics*, 7, 223–239.

Nieto-Barajas, L. E., Prünster, I., and Walker, S. G. (2004), "Normalized Random Measures Driven by Increasing Additive Processes," *The Annals of Statistics*, 32, 2343–2360.

Paddock, S. M., Ruggeri, F., Lavine, M., and West, M. (2003), "Randomized Polya Tree Models for Nonparametric Bayesian Inference," *Statistica Sinica*, 13, 443–460.

Perman, M., Pitman, J., and Yor, M. (1992), "Size-Biased Sampling of Poisson Point Processes and Excursions," *Probability Theory and Related Fields*, 92, 21–39.

Petrone, S., and Veronese, P. (2002), "Nonparametric Mixture Priors Based on an Exponential Random Scheme," *Statistical Methods and Applications*, 11, 1–20.

Pitman, J. (1995), "Exchangeable and Partially Exchangeable Random Partitions," *Probability Theory and Related Fields*, 102, 145–158.

——— (1996), "Some Developments of the Blackwell–MacQueen Urn Scheme," in *Statistics, Probability and Game Theory. Papers in Honor of David Blackwell*, eds. T. S. Ferguson et al., Hayward, CA: Institute of Mathematical Statistics, pp. 245–267.

——— (2003), "Poisson–Kingman Partitions," in *Science and Statistics: A Festschrift for Terry Speed*, ed. D. R. Goldstein, Hayward, CA: Institute of Mathematical Statistics, pp. 1–35.

Prünster, I. (2002), "Random Probability Measures Derived From Increasing Additive Processes and Their Application to Bayesian Statistics," unpublished doctoral thesis, University of Pavia.

Regazzini, E. (2001), "Foundations of Bayesian Statistics and Some Theory of Bayesian Nonparametric Methods," lecture notes, Stanford University.

Regazzini, E., Guglielmi, A., and Di Nunno, G. (2002), "Theory and Numerical Analysis for Exact Distribution of Functionals of a Dirichlet Process," *The Annals of Statistics*, 30, 1376–1411.

Regazzini, E., Lijoi, A., and Prünster, I. (2003), "Distributional Results for Means of Random Measures With Independent Increments," *The Annals of Statistics*, 31, 560–585.

Roeder, K. (1990), "Density Estimation With Confidence Sets Exemplified by Superclusters and Voids in the Galaxies," *Journal of the American Statistical Association*, 85, 617–624.

Roeder, K., and Wasserman, L. (1997), "Practical Bayesian Density Estimation Using Mixtures of Normals," *Journal of the American Statistical Association*, 92, 894–902.

Seshadri, V. (1993), *The Inverse Gaussian Distribution*, New York: Oxford University Press.

Walker, S., and Damien, P. (1998), "A Full Bayesian Non-Parametric Analysis Involving a Neutral to the Right Process," *Scandinavia Journal of Statistics*, 25, 669–680.

Walker, S., Damien, P., Laud, P. W., and Smith, A. F. M. (1999), "Bayesian Nonparametric Inference for Random Distributions and Related Functions" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 61, 485–527.