On the stick-breaking representation of normalized inverse Gaussian priors

Biometrika (2012), **99**, 3, *pp.* 663–674

Printed in Great Britain

doi: 10.1093/biomet/ass023
Advance Access publication 18 June 2012

On the stick-breaking representation of normalized inverse Gaussian priors

By S. FAVARO

Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy
stefano.favaro@unito.it

A. LIJOI

Department of Economics and Business, University of Pavia, Via San Felice 5, 27100 Pavia, Italy
lijoi@unipv.it

and I. PRÜNSTER

Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy
igor@econ.unito.it

Summary

Random probability measures are the main tool for Bayesian nonparametric inference, with their laws acting as prior distributions. Many well-known priors used in practice admit different, though equivalent, representations. In terms of computational convenience, stick-breaking representations stand out. In this paper we focus on the normalized inverse Gaussian process and provide a completely explicit stick-breaking representation for it. This result is of interest both from a theoretical viewpoint and for statistical practice.

Some key words: Bayesian nonparametric inference; Dirichlet process; Normalized inverse Gaussian process; Random probability measure; Stick-breaking representation.

1. Introduction

1·1. *Bayesian nonparametrics and the stick-breaking construction*

Bayesian nonparametric inference has recently undergone strong development. See Hjort et al. (2010) for an up-to-date review. At the heart of the approach lies the concept of random probability measure, whose law acts as a prior for Bayesian nonparametric inference, the most notable example being the Dirichlet process (Ferguson, 1973). There exist different representations for a number of nonparametric priors, which, although equivalent in distribution, may serve different purposes. For example, representations based on completely random measures allow one to study analytically their properties (Lijoi & Prünster, 2010), whereas stick-breaking representations have displayed great potential in addressing modelling and computational issues. The main result of this paper is a stick-breaking representation of the normalized inverse Gaussian process (Lijoi et al., 2005), a tractable alternative to the Dirichlet process. Our result

is of interest from a theoretical point of view since it is the first representation of a random probability measure in terms of dependent and non-beta distributed stick-breaking weights and it completes the study of the normalized inverse Gaussian process, for which a stick-breaking representation was missing. From a modelling and computational point of view it paves the way for the definition of complex models based on the normalized inverse Gaussian process by replacing the stick-breaking constructed Dirichlet process, most notably within dependent models for nonparametric regression, and it allows extensions of recent simulation algorithms, based on stick-breaking constructions, to the normalized inverse Gaussian process.

There are several different ways to define the Dirichlet process, each highlighting one of its aspects. The original definition of Ferguson (1973) constructs the Dirichlet process $\mathscr{D}_{c,P_0}$, with parameter $\alpha = cP_0$ and P_0 a probability measure, in terms of a consistent family of finite-dimensional Dirichlet distributions. An alternative definition, also due to Ferguson (1973), relies on the idea of normalizing a gamma process. A third construction is based on a stick-breaking procedure that follows from a result in J. W. McCloskey's 1965 PhD thesis at Michigan State University, recalled as Theorem 1 in Pitman (1996) under the assumption of nonatomic P_0, and that has been extended to any P_0 in Sethuraman (1994). Let $(V_i)_{i\geqslant 1}$ be a sequence of independent and identically distributed random variables, with $V_i \sim \mathrm{Be}(1, c)$ and $c > 0$, and define random probability weights $(\tilde{p}_j)_{j\geqslant 1}$ as

$$\tilde{p}_1 = V_1, \qquad \tilde{p}_j = V_j \prod_{i=1}^{j-1}(1 - V_i) \quad (j = 2, 3, \ldots). \tag{1}$$

If $(Y_i)_{i\geqslant 1}$ is a sequence of independent and identically distributed random variables, independent of the $\tilde{p}_i$ and whose common probability distribution is P_0, then $\sum_{j\geqslant 1} \tilde{p}_j\, \delta_{Y_j}(\cdot) = \mathscr{D}_{c,P_0}(\cdot)$ in distribution, where δ_a denotes the unit point mass at a.

Another prominent nonparametric prior is the two-parameter Poisson–Dirichlet process (Perman et al., 1992), also known, according to terminology introduced in Ishwaran & James (2001), as the Pitman–Yor process. This admits a simple stick-breaking representation: let $(V_i)_{i\geqslant 1}$ be independent random variables with $V_i \sim \mathrm{Be}(1-\sigma, \theta + i\sigma)$, $\sigma \in (0, 1)$ and $\theta > -\sigma$, and define the random probability masses as in (1) and $(Y_i)_{i\geqslant 1}$ as above, except that we require a nonatomic P_0. The resulting discrete random probability $\sum_{j\geqslant 1}\tilde{p}_j\delta_{Y_j}(\cdot)$ coincides, in distribution, with the two-parameter Poisson–Dirichlet process. See Pitman (2006) for an exhaustive account. Bayesian nonparametric applications include mixtures (Ishwaran & James, 2001), linguistics (Teh, 2006), species sampling (Lijoi et al., 2007) and survival analysis (Jara et al., 2010).

The extreme flexibility of stick-breaking representations has originated a vast literature concerning both modelling and computation. In terms of modelling, the dependent processes, initiated by MacEachern (1999), heavily rely on a stick-breaking construction and have proved to be effective prior specifications in regression problems. See Hjort et al. (2010) for a review of recent contributions. From a computational point of view significant progress, especially in designing efficient simulation algorithms for hierarchical mixtures, has been made using such representations. Among the most relevant contributions, which devise algorithms that work in principle for any random probability measure with an explicit stick-breaking representation, are the blocked Gibbs sampler (Ishwaran & James, 2001), the retrospective sampler (Papaspiliopoulos & Roberts, 2008), the slice sampler (Walker, 2007) and a very efficient synthesis (Yau et al., 2011) of these last two.

1·2. *General remarks and motivation*

Starting from the stick-breaking representations of the Dirichlet and the two-parameter Poisson–Dirichlet processes, a general class of stick-breaking priors can be defined by allowing independent stick-breaking weights V_i with an arbitrary distribution on $(0, 1)$. This issue is addressed in Ishwaran & James (2001) limited to beta distributed V_i. However, their results also hold for general distributions. One might wonder whether, besides the Dirichlet and two-parameter Poisson–Dirichlet processes, there are other members of this large class that share a similar degree of mathematical tractability, allowing the investigation of some of their properties, such as the prediction rules, the posterior distribution or the distribution of the random partition they induce. Unfortunately, no other instances are known, and this clearly limits their appeal in terms of modelling and applications beyond the Dirichlet and two-parameter Poisson–Dirichlet cases. The reason for the poor tractability may be traced back to a distributional concept originated in population genetics, which is termed invariance under size-biased permutation and is recalled in § 2. Here it is enough to note that such an invariance property implies significant mathematical simplifications when working out distributional properties of a random probability measure. In particular, as shown in McCloskey's PhD thesis, the Dirichlet process is the only random probability measure admitting a stick-breaking representation with independent and identically distributed weights V_i which is invariant under size-biased permutation. On the other hand, the two-parameter Poisson–Dirichlet process is essentially the only random probability measure invariant under size-biased permutation that admits a stick-breaking representation with independent weights V_i (Pitman, 1996). These considerations suggest that if one would like to identify further random probability measures both enjoying sufficient tractability and admitting a simple stick-breaking representation one has to focus on dependent stick-breaking weights V_i.

Here we consider the normalized inverse Gaussian process (Lijoi et al., 2005). By now many of its properties are known (Lijoi et al., 2005; James et al., 2006, 2009). In some sense, for any distributional property of the Dirichlet process, an analogous property of the normalized inverse Gaussian process is known, with the notable exception of a stick-breaking representation. Our main result fills this gap. It is important to anticipate that the stick-breaking weights V_i will be dependent: this is not surprising and actually is necessary. To see why, first note that any discrete random probability measure admits a stick-breaking representation if one allows any possible distribution and form of dependence for the V_i (Pitman, 1996). Moreover, from Perman et al. (1992) it is immediate that the normalized inverse Gaussian process, and more generally any homogeneous normalized random measure with independent increments (Regazzini et al., 2003; James et al., 2006), is invariant under size-biased permutation. Therefore, by the above-mentioned characterizations provided in McCloskey's PhD thesis and in Pitman (1996), none of them can admit stick-breaking representation with independent, and a fortiori independent and identically distributed, weights V_i. Hence, the weights V_i must be dependent. In this respect, the stick-breaking representation for the normalized inverse Gaussian process represents the first case of a tractable prior with explicit stick-breaking representation based on dependent weights. To avoid misunderstandings it is to be stressed that, in principle, one can define random probability measures by writing down a stick-breaking representation with either independent or dependent weights: the key point is, however, obtaining a measure whose properties can be analysed in more or less explicit form. If not, an arbitrary stick-breaking representation is essentially a vacuous object since the construction itself is not able to provide, on its own, intuition and understanding of the behaviour of the resulting random probability.

1·3. *The stick-breaking representation of the normalized inverse Gaussian process*

Let us first fix some notation and display the distributions that play a key role in the construction. Let X be a generalized inverse Gaussian random variable with parameters $a, b > 0$ and $p \in \mathbb{R}$, in symbols $X \sim \text{GIG}(a, b, p)$, whose probability density function is

$$f_X(x) = \frac{(a/b)^{p/2}}{2K_p\{(ab)^{1/2}\}} x^{p-1} \exp\left\{-\frac{1}{2}\left(ax + \frac{b}{x}\right)\right\} \quad (x > 0), \tag{2}$$

with K_ν indicating the modified Bessel function of the third type. Moreover, let Z be a positive $1/2$-stable random variable with scale parameter $b > 0$, $Z \sim \text{St}_{1/2}(b)$, whose density is

$$f_Z(z) = \frac{b^{1/2}}{(2\pi)^{1/2}} z^{-3/2} \exp\left(-\frac{1}{2}\frac{b}{z}\right) \quad (z > 0).$$

Note that $1/2$-stable random variables can also be seen as reciprocal gamma random variables with shape parameter $1/2$ and scale parameter $b/2$. Introduce a sequence $(Z_i)_{i \geqslant 1}$ of independent and identically distributed random variables with $Z_i \sim \text{St}_{1/2}(1)$ and define a sequence of dependent $(0, 1)$-valued random variables $(V_i)_{i \geqslant 1}$ as follows

$$\begin{aligned} V_1 &= \frac{X_1}{X_1 + Z_1}, \quad X_1 \sim \text{GIG}(\tau_1, 1, -1/2), \\ (V_i \mid V_1, \ldots, V_{i-1}) &= \frac{X_i}{X_i + Z_i}, \quad X_i \sim \text{GIG}(\tau_i, 1, -i/2) \quad (i \geqslant 2), \end{aligned} \tag{3}$$

where $\tau_1 = a$, $\tau_i = a/\{\prod_{j=1}^{i-1}(1 - V_j)\}$, for $i \geqslant 2$, and the sequences $(X_i)_{i \geqslant 1}$ and $(Z_i)_{i \geqslant 1}$ are independent. Sampling the V_i is straightforward (Atkinson, 1982). It will be shown in § 2 that the distribution of V_1 and of $V_i \mid V_1, \ldots, V_{i-1}$, for $i \geqslant 2$, in (3), is a special case of normalized generalized inverse Gaussian distribution and admits closed form density, for $v \in (0, 1)$,

$$\begin{aligned} f_{V_1}(v) &= \frac{a^{1/4} v^{-1/2} (1 - v)^{-1}}{(2\pi)^{1/2} K_{-1/2}\left(a^{1/2}\right)} K_{-1}\left\{\left(\frac{a}{1 - v}\right)^{1/2}\right\}, \\ f_{(V_i \mid V_1, \ldots, V_{i-1})}(v) &= \frac{\tau_i^{1/4} v^{-1/2} (1 - v)^{-5/4 + i/4}}{(2\pi)^{1/2} K_{-i/2}\left(\tau_i^{1/2}\right)} K_{-1/2 - i/2}\left\{\left(\frac{\tau_i}{1 - v}\right)^{1/2}\right\} \quad (i \geqslant 2). \end{aligned} \tag{4}$$

Such a class of distributions is denoted as N-GIG$^*(a, p)$, and (3) can be expressed as

$$V_1 \sim \text{N-GIG}^*(a, -1/2), \qquad (V_i \mid V_1, \ldots, V_{i-1}) \sim \text{N-GIG}^*(\tau_i, -i/2) \quad (i \geqslant 2). \tag{5}$$

Finally, N-IG$_{c, P_0}(\cdot)$ stands for a normalized inverse Gaussian process with parameter $\alpha = c\, P_0$.

PROPOSITION 1. *Let $(V_i)_{i \geqslant 1}$ be a sequence of dependent random variables as in* (3), *or equivalently* (5), *and define the random probability weights $(\tilde{p}_j)_{j \geqslant 1}$ via stick-breaking as in* (1). *Let $(Y_i)_{i \geqslant 1}$ be a sequence of independent and identically distributed random variables, independent*

of the $\tilde{p}_i$ *and with nonatomic distribution* P_0. *Then, for* $c = a^{1/2}$,

$$\sum_{j \geqslant 1} \tilde{p}_j \, \delta_{Y_j}(\cdot) = \text{N-IG}_{c, P_0}(\cdot).$$

The proof, together with the necessary background and auxiliary results, is given in § 2.

2. Background, auxiliary results and proof

2·1. *The normalized generalized inverse Gaussian distribution*

We start by introducing a new distribution on $(0, 1)$, which includes both the normalized inverse Gaussian distribution and the distribution defining the stick-breaking weights (4) as specific examples. To this end first recall that the parameter space of the distribution of a GIG(a, b, p) random variable admitting density (2) is given by $\Theta = \Theta_1 \cup \Theta_2 \cup \Theta_3$ with $\Theta_1 = \{(a, b, p) : a > 0, b > 0, p \in \mathbb{R}\}$, $\Theta_2 = \{(a, b, p) : a > 0, b = 0, p > 0\}$ and $\Theta_3 = \{(a, b, p) : a = 0, b > 0, p < 0\}$. Interesting special cases correspond to each subspace Θ_i $(i = 1, 2, 3)$: from Θ_1, the case considered in § 1, one obtains the inverse Gaussian distributions, $p = -1/2$, among others; Θ_2 corresponds to the class of gamma distributions; Θ_3 identifies the class of reciprocal gamma distributions and hence, in particular, of the positive 1/2-stable distribution, $p = -1/2$. An exhaustive account is provided in Jørgensen (1982).

Definition 1. *Let* X_1 *and* X_2 *be two independent random variables such that* $X_1 \sim \text{GIG}(a_1, b_1, p_1)$ *and* $X_2 \sim \text{GIG}(a_2, b_2, p_2)$. *The random variable* $V = X_1(X_1 + X_2)^{-1}$, *taking values in* $(0, 1)$, *is termed normalized generalized inverse Gaussian,* $V \sim$ N-GIG$(a_1, b_1, p_1, a_2, b_2, p_2)$, *and admits density*

$$f_V(v) = \frac{\left(\frac{a_1}{b_1}\right)^{p_1/2} \left(\frac{a_2}{b_2}\right)^{p_2/2}}{2K_{p_1}\{(b_1 a_1)^{1/2}\} K_{p_2}\{(b_2 a_2)^{1/2}\}} \, v^{p_1 - 1}(1 - v)^{p_2 - 1} \left\{ \frac{\frac{b_1}{v} + \frac{b_2}{1 - v}}{a_1 v + a_2(1 - v)} \right\}^{(p_1 + p_2)/2}$$
$$\times \, K_{p_1 + p_2}\left(\left[\left\{ \frac{b_1}{v} + \frac{b_2}{1 - v} \right\} \{a_1 v + a_2(1 - v)\} \right]^{1/2} \right), \quad v \in (0, 1). \tag{6}$$

The density displayed in (6) is obtained by application of a simple change of variable and formula 3.471.9 in Gradshteyn & Ryzhik (2000). Given that generalized inverse Gaussian random variables are infinitely divisible (Barndorff-Nielsen & Halgreen, 1977), the normalized generalized inverse Gaussian distribution represents another example of the class of normalized infinitely divisible distributions studied in Favaro et al. (2011).

For our purposes, two special cases are of particular interest. The first is the normalized inverse Gaussian distribution (Lijoi et al., 2005), which corresponds to a N-GIG$(1, b_1, -1/2, 1, b_2, -1/2)$ distribution or, in other terms, to (6) with X_1 and X_2 inverse Gaussian. Its density simplifies to

$$f_V(v) = \frac{(b_1 b_2)^{1/2} \, \mathrm{e}^{b_1^{1/2} + b_2^{1/2}}}{\pi} \frac{K_{-1}\left\{ \left(\frac{b_1}{v} + \frac{b_2}{1 - v} \right)^{1/2} \right\}}{v^{3/2}(1 - v)^{3/2} \left(\frac{b_1}{v} + \frac{b_2}{1 - v} \right)^{1/2}}, \quad v \in (0, 1), \tag{7}$$

which is seen to coincide with equation (5) in Lijoi et al. (2005) by setting $\alpha_i = b_i^{1/2}$, for $i = 1, 2$.

The second subclass of Definition 1 to be considered corresponds to the random variable (5) dictating the form of the stick-breaking weights in Proposition 1 and is denoted by N-GIG$^*(a, p)$. Specifically, one has N-GIG$^*(a, p)$ = N-GIG$(a, 1, p, 0, 1, -1/2)$ with density

$$f_V(v) = \frac{a^{1/4}}{(2\pi)^{1/2} K_p(a^{1/2})} v^{-1/2} (1-v)^{-5/4-p/2} K_{p-\frac{1}{2}} \left\{ \left(\frac{a}{1-v} \right)^{1/2} \right\}, \quad v \in (0,1),$$

which reduces to (4) when the parameters a and p are replaced by τ_i and $p_i = -i/2$, respectively, for $i \geqslant 1$.

2·2. *The normalized inverse Gaussian process*

The normalized inverse Gaussian process (Lijoi et al., 2005) enjoys mathematical tractability and it is particularly effective when drawing inference on the clustering structure featured by data. Its original definition was given in terms of a consistent system of finite-dimensional distributions. Consider n independent inverse Gaussian random variables X_i, which admit a density as in (2) with $p = -1/2$. For our purposes we can take $a = 1$ without loss of generality and write $X_i \sim \text{IG}(1, b_i)$, for $i = 1, \ldots, n$. As one defines the Dirichlet distribution via normalization of independent gamma random variables, one can construct the normalized inverse Gaussian distribution with parameter $(b_1, \ldots, b_n)$ as the distribution of the random vector $(W_1, \ldots, W_n)$, where $W_i = X_i / (\sum_{j=1}^n X_j)$ for $i = 1, \ldots, n$, which admits density on the $(n-1)$-dimensional simplex Δ_{n-1}, with respect to the Lebesgue measure on $\mathbb{R}^{n-1}$, coinciding with

$$f(w_1, \ldots, w_{n-1}) = \frac{\exp(\sum_{i=1}^n {b_i}^{1/2}) \prod_{i=1}^n {b_i}^{1/2}}{2^{n/2-1}\pi^{n/2}} w_1^{-3/2} \cdots w_{n-1}^{-3/2} \left(1 - \sum_{i=1}^{n-1} w_i \right)^{-3/2}$$
$$\times \left\{ \mathcal{A}_n(w_1, \ldots, w_{n-1}) \right\}^{-n/4} K_{-\frac{n}{2}} \left[\left\{ \mathcal{A}_n(w_1, \ldots, w_{n-1}) \right\}^{1/2} \right], \quad (8)$$

where $\mathcal{A}_n(w_1, \ldots, w_{n-1}) = \sum_{i=1}^{n-1} (b_i / w_i) + b_n / (1 - \sum_{j=1}^{n-1} w_j)$. Clearly, (8) reduces to the marginal distribution (7) if $n = 2$. In Lijoi et al. (2005) it is shown that there exists a random probability measure P, termed a normalized inverse Gaussian process with parameter measure $\alpha = c\, P_0$ and denoted as N-IG$_{c,\, P_0}(\cdot)$, with (8) defining its family of finite-dimensional distributions.

The second construction of a normalized inverse Gaussian process we will need is obtained via normalization of an inverse Gaussian process, which corresponds to the definition of a Dirichlet process as a normalized gamma process (Ferguson, 1973). To this end recall the concept of a completely random measure (Kingman, 1993): suppose μ is a random measure on some complete and separable metric space $\mathbb{X}$ such that for any measurable $A_1, \ldots, A_n$, with $A_i \cap A_j = \emptyset$ for $i \neq j$, the random variables $\mu(A_1), \ldots, \mu(A_n)$ are mutually independent. Then, μ is termed a completely random measure. A completely random measure μ, without jumps at fixed points of discontinuity, is uniquely identified by its Lévy intensity ν by means of its Lévy–Khintchine representation

$$E\left[\exp\left\{ -\int_{\mathbb{X}} f(x) \mu(\mathrm{d}x) \right\} \right] = \exp\left[-\int_{\mathbb{R}^+ \times \mathbb{X}} \left\{ 1 - \mathrm{e}^{-s f(y)} \right\} \nu(\mathrm{d}s, \mathrm{d}y) \right]$$

for any measurable $\mathbb{R}$-valued function such that $\int |f| \, \mathrm{d}\mu < \infty$ almost surely. Another property to recall is the almost sure discreteness of completely random measures, which implies that any

such measure μ is representable as $\mu(\cdot) = \sum_{j \geqslant 1} J_j \delta_{Y_j}(\cdot)$. For our purposes it is enough to focus on completely random measures μ such that $\mu(\mathbb{X}) < \infty$ almost surely and the locations Y_j are independent of the nonnegative jumps J_j. Moreover, without loss of generality, the locations $(Y_j)_{j \geqslant 1}$ can be assumed to be independent and identically distributed from a nonatomic P_0. This is equivalent to saying that the Lévy intensity factorizes as $\nu(\mathrm{d}s, \mathrm{d}y) = \rho(\mathrm{d}s) P_0(\mathrm{d}y)$ so that the corresponding completely random measure μ is homogeneous. If ρ has infinite total mass then μ is strictly positive, so the normalization we are going to carry out is admissible. On the other hand, homogeneity is motivated by mere technical convenience.

We now recall the definition of homogeneous normalized random measures with independent increments (Regazzini et al., 2003; James et al., 2006), which contain the Dirichlet and normalized inverse Gaussian processes as special cases. In fact, starting from a completely random measure satisfying the above conditions one can always define a homogeneous normalized random measure with independent increments as

$$P(\cdot) = T^{-1} \mu(\cdot) = \sum_{j \geqslant 1} p_j \delta_{Y_j}(\cdot), \tag{9}$$

with $T = \mu(\mathbb{X}) = \sum_{j \geqslant 1} J_j$ and $p_j = J_j / T$ for any $j \geqslant 1$.

To define the normalized inverse Gaussian process via normalization consider an inverse Gaussian completely random measure, which is characterized by the Lévy intensity

$$\rho(\mathrm{d}s) P_0(\mathrm{d}y) = \frac{b^{1/2}}{(2\pi)^{1/2}} s^{-3/2} \mathrm{e}^{-1/2\, a\, s}\, \mathrm{d}s\, P_0(\mathrm{d}y) \quad (s > 0,\, y \in \mathbb{X},\, b > 0), \tag{10}$$

where, without loss of generality for our scope, we can set $a = 1$. One then obtains the N-IG$_{b^{1/2}, P_0}(\cdot)$ process as a homogeneous normalized random measure with independent increments (9) characterized by the Lévy intensity (10) with $a = 1$. The Dirichlet process is obtained by replacing the inverse Gaussian completely random measure with a gamma completely random measure or, in other terms, (10) with $\rho(\mathrm{d}s) P_0(\mathrm{d}y) = a\, s^{-1} \mathrm{e}^{-s}\, \mathrm{d}s\, P_0(\mathrm{d}y)$ for any $a > 0$.

2·3. *Size-biased permutations*

Consider any discrete random probability measure $P = \sum_{j \geqslant 1} p_j \delta_{Y_j}$, the only constraint being that the locations $(Y_j)_{j \geqslant 1}$, which are independent and identically distributed from a nonatomic probability measure P_0, are independent of the random probabilities $(p_j)_{j \geqslant 1}$. Homogeneous normalized random measures with independent increments defined in § 2·2 fit into this general framework. An interesting rearrangement of the elements of $(p_j)_{j \geqslant 1}$ can be obtained by the so-called size-biased permutation, a concept originated in population genetics and defined by the following procedure. Consider an exchangeable sequence $(X_n)_{n \geqslant 1}$ directed by the discrete random probability measure P. Correspondingly, define $(N_i)_{i \geqslant 1}$ as the successive times at which new values of the sequence $(X_n)_{n \geqslant 1}$ appear, namely $N_1 = 1$ and $N_j = \inf\{i > N_{j-1} : X_i \notin (X_1, \ldots, X_{i-1})\}$ for any $j \geqslant 2$. Note that $\mathrm{pr}(N_n > n) > 0$ since, due to the discreteness of P, ties will be recorded with positive probability. Moreover, let $(\xi_i)_{i \geqslant 1}$ be an integer-valued sequence such that $\mathrm{pr}\{\xi_i = n \mid (p_j)_{j \geqslant 1}\} = p_n$ and $\mathrm{pr}\{X_n = Y_{\xi_n} \mid (p_j)_{j \geqslant 1},\ (Y_j)_{j \geqslant 1},\ (\xi_i)_{i \geqslant 1}\} = 1$. Hence ξ_n identifies the specific location X_n coincides with and this clearly entails that $\xi_{N_i} \neq \xi_{N_\ell}$ if $i \neq \ell$. Finally, set

$$\tilde{p}_i = p_{\xi_{N_i}} \quad (i \geqslant 1), \tag{11}$$

with the convention $\tilde{p}_i = 0$ if there are fewer than i distinct values in $(X_n)_{n\geqslant 1}$. The sequence $(\tilde{p}_j)_{j\geqslant 1}$ is termed a size-biased permutation of $(p_j)_{j\geqslant 1}$. Moreover, $(p_j)_{j\geqslant 1}$, or the corresponding random probability measure P, is termed invariant under size-biased permutation if $(\tilde{p}_j)_{j\geqslant 1}$, whose coordinates are defined according to (11), has the same finite-dimensional distributions as $(p_j)_{j\geqslant 1}$. See Pitman (1996) and references therein for details.

The importance of invariance under size-biased permutation is best illustrated by the following considerations. Clearly stick-breaking priors can be defined whatever the choice of [0, 1]-valued random weights $(V_i)_{i\geqslant 1}$ provided they ensure $\sum_{i\geqslant 1} p_i = 1$, almost surely, in (1). Nonetheless for the investigation of certain distributional properties of P, which are of interest in statistical applications, invariance under size-biased permutation is essential. Indeed, if one wishes to analyse the clustering structure induced by P or to make predictions about the outcomes of future observations, one needs an expression for the exchangeable partition probability function and this can hardly be derived unless P is invariant under size-biased permutation. To make this point clear, let $X_1, \ldots, X_n$ be a sample from an exchangeable sequence $(X_i)_{i\geqslant 1}$, directed by P, that features $K_n \leqslant n$ distinct values: these, in turn, define a partition into K_n clusters with respective frequencies $N_{1,n}, \ldots, N_{K_n,n}$. Hence, the exchangeable partition probability function is the probability distribution of the random vector $(K_n, N_{1,n}, \ldots, N_{K_n,n})$, that is

$$\begin{aligned} p_k^{(n)}(n_1, \ldots, n_k) &= \mathrm{pr}(K_n = k, N_{1,n} = n_1, \ldots, N_{K_n,n} = n_k) \\ &= \sum_{i_1 \neq \cdots \neq i_k} E(p_{i_j}^{n_1} \cdots p_{i_k}^{n_k}). \end{aligned} \tag{12}$$

From an operational standpoint the expression in (12) is not useful: for certain specifications of the stick-breaking random probabilities p_i one may be able to compute the expected value $E(p_{i_j}^{n_1} \cdots p_{i_k}^{n_k})$, but the sum over the indices $i_1, \ldots, i_k$ cannot be evaluated explicitly. Even numerically it is a highly demanding task already for moderately small values of k. Importantly, (12) can be re-expressed in much simpler form in terms of the size-biased permutation $(\tilde{p}_i)_{i\geqslant 1}$ of the sequence $(p_i)_{i\geqslant 1}$, namely

$$p_k^{(n)}(n_1, \ldots, n_k) = E\left\{ \prod_{i=1}^{k} \tilde{p}_i^{n_i} \prod_{j=1}^{k-1} \left(1 - \sum_{r=1}^{j} \tilde{p}_r \right) \right\}; \tag{13}$$

see equation (8) in Pitman (1996). Now, if and only if P is invariant under size-biased permutation, the $\tilde{p}_i$s in (13) can be replaced by the stick-breaking random probabilities p_is. Consequently, $p_k^{(n)}$ can be more easily evaluated by using (13), instead of (12), with the simple stick-breaking p_is in place of their size-biased permutations, whose distribution, unless invariance holds, is typically very complicated or not known. For instance, in the Dirichlet case one then immediately obtains from (13) the Ewens sampling formula

$$p_k^{(n)}(n_1, \ldots, n_k) = \frac{c^k}{(c)_n} \prod_{i=1}^{k} (n_i - 1)!$$

with $(c)_n = c(c+1)\cdots(c+n-1)$ denoting the ascending factorial. Similarly one obtains the Pitman sampling formula in the two-parameter Poisson–Dirichlet case. In light of the above considerations it is apparent why no exchangeable partition probability function is known for stick-breaking priors other than those invariant under size-biased permutation. Now consider the class of homogeneous normalized random measures with independent increments (9) and

denote by $(J_{(j)})_{j\geqslant 1}$ the sequence of jumps of μ rearranged in decreasing order. The corresponding sequence of random probability weights is then given by $p_{(j)} = J_{(j)}/T$ for any $j \geqslant 1$. Perman et al. (1992) generalized the result in McCloskey's PhD thesis. Indeed, they derived a stick-breaking characterization for the size-biased permutation $(\tilde{p}_j)_{j\geqslant 1}$ of the sequence of ranked random probabilities $(p_{(j)})_{j\geqslant 1}$ by providing a detailed description of the distribution of the sequence $(V_i)_{i\geqslant 1}$ within (1) in terms of the measure ρ and the distribution of the total mass T. Now focus on the random probability measures. Since the locations $(Y_j)_{j\geqslant 1}$ are assumed to be independent and identically distributed from a nonatomic probability measure P_0 independent of the random probabilities $(p_j)_{j\geqslant 1}$ and given that the sequences $(p_{(j)})_{j\geqslant 1}$ and $(\tilde{p}_j)_{j\geqslant 1}$ represent two specific rearrangements of the original sequence $(p_j)_{j\geqslant 1}$, one has

$$P(\cdot) = \sum_{j\geqslant 1} p_{(j)}\delta_{Y_j}(\cdot) = \sum_{j\geqslant 1} \tilde{p}_j\delta_{Y_j}(\cdot) \tag{14}$$

in distribution. By combining this with the definition of the Dirichlet process as a normalized gamma process and the identity (14), one recovers the stick-breaking representation. The same strategy is followed for the derivation of the stick-breaking representation in the normalized inverse Gaussian case.

2·4. *Proof*

Given the material provided in the previous sections, the proof of Proposition 1 is now a quite straightforward application of Theorem 2.1 in Perman et al. (1992). Consider an inverse Gaussian completely random measure: let f_T be the density function of the corresponding total mass T,

$$f_T(t) = \frac{e^{b^{1/2}}b^{1/2}}{(2\pi)^{1/2}}t^{-3/2}\exp\left\{-\frac{1}{2}\left(t+\frac{b}{t}\right)\right\} \quad (t, b > 0), \tag{15}$$

and denote by λ the density function of ρ in (10) given by

$$\lambda(s) = \frac{b^{1/2}}{(2\pi)^{1/2}}s^{-3/2}\exp\left(-\frac{1}{2}s\right) \quad (s > 0). \tag{16}$$

The normalized inverse Gaussian process, by the identity in (14) and its construction via normalization $P = \mu/T$ recalled in § 2·2, can be represented as

$$P(\cdot) = \sum_{j\geqslant 1} \tilde{p}_j\delta_{Y_j}(\cdot) \tag{17}$$

where $(\tilde{p}_j)_{j\geqslant 1}$ is the size-biased permutation of the ranked random probabilities of P and $(Y_j)_{j\geqslant 1}$ is a sequence of independent and identically distributed random variables, which are independent of the $\tilde{p}_j$ and whose common probability distribution P_0 is nonatomic.

By Theorem 2.1 in Perman et al. (1992), the sequence $(\tilde{p}_j)_{j\geqslant 1}$ in (17) has stick-breaking representation in terms of some sequence of dependent random variables $(V_i)_{i\geqslant 1}$. In particular, they provide a structural expression for the joint distribution of the random variables $(V_1, \ldots, V_i)$, for any $i \geqslant 1$, in terms of the density functions f_T and λ. We start by deriving the distribution of V_1. According to equation (2.d) in Perman et al. (1992), the density function of the random variable

V_1 is of the form

$$f_{V_1}(v_1) = v_1 \int_0^{+\infty} t\lambda(v_1 t) f_T\{(1-v_1)t\}\,\mathrm{d}t$$
$$= \frac{\mathrm{e}^{b^{1/2}} b^{1/2}}{2\pi} v_1^{-1/2}(1-v_1)^{-3/2} \int_0^{+\infty} t^{-1-1} \exp\left(-\frac{b}{2t(1-v_1)} - \frac{t}{2}\right)\mathrm{d}t \quad v_1 \in (0,1),$$

where the second identity is obtained by inserting (15) and (16). The density displayed in (4) is obtained by combining the identity $K_{-1/2}(b^{1/2}) = \pi^{1/2}(2b)^{-1/2}e^{-b^{1/2}}$ with Formula 3.471.9 in Gradshteyn & Ryzhik (2000) and setting $b = a$. The representation in (3), or (5), follows from the definition of normalized generalized inverse Gaussian distribution given in Definition 1 and the corresponding density (6). Now consider the case $i = 2$: according to equation (2.d) in Perman et al. (1992), the joint density function of the random variables (V_1, V_2) is

$$f_{V_1,V_2}(v_1, v_2) = v_1 v_2 (1-v_1) \int_0^{+\infty} t^2 \lambda(v_1 t)\lambda\{v_2(1-v_1)t\} f_T\{(1-v_1)(1-v_2)t\}\,\mathrm{d}t$$
$$= \frac{\mathrm{e}^{b^{1/2}} b}{(2\pi)^{3/2}} v_1^{-1/2}(1-v_1)^{-2} v_2^{-1/2}(1-v_2)^{-3/2}$$
$$\times \int_0^{+\infty} t^{-3/2-1} \exp\left\{\frac{b}{2t(1-v_1)(1-v_2)} - \frac{t}{2}\right\}\mathrm{d}t \quad v_1, v_2 \in (0,1),$$

where the second identity is obtained by inserting (15) and (16). The density of $(V_2 \mid V_1)$ in (4) is obtained by using formula 3.471.9 in Gradshteyn & Ryzhik (2000) and by dividing by the marginal density of V_1, provided a in (4) is set equal to b appearing in the expression above. The representations in (3) or (5) follow again by Definition 1. Proceeding along the same lines one obtains the density of $(V_i \mid V_1, \ldots, V_{i-1})$ displayed in (4) for any $i \geqslant 3$.

3. CONCLUDING REMARKS

The knowledge of a posterior representation of the random probability is not necessary for drawing posterior inferences in complex models based on some stick-breaking prior. Nonetheless, the derivation of a posterior representation is important for understanding the distributional structure of the model conditional on observed data. Here we provide such a structural description starting from the general result provided in Theorem 1 in James et al. (2009), and show how the stick-breaking construction as well as the normalized generalized inverse gaussian distribution appear in it.

Consider a N-IG$_{b^{1/2},P_0}(\cdot)$ prior, suppose that the observed sample $X_1, \ldots, X_n$ has displayed k distinct values $X_1^*, \ldots, X_k^*$ with respective frequencies $n_1, \ldots, n_k$, and introduce a latent random variable U_n whose density function, conditionally on $X_1, \ldots, X_n$, is such that $f_{U_n}(u) \propto u^{n-1}(u+1/2)^{k/2-n} \exp[-\{b(1+2u)\}^{1/2}]$ with $u > 0$. It can be shown that conditionally on $X_1, \ldots, X_n$ and on U_n the N-IG$_{b^{1/2},P_0}(\cdot)$ process coincides in distribution with

$$w_{0,u}\, P_u + \sum_{i=1}^{k} w_{i,u}\, \delta_{X_i^*}$$

where $P_u = \mu_u / T_u$ is a N-IG$_{\{b(1+2u)\}^{1/2},\, P_0}(\cdot)$ random probability measure. Hence, the stick-breaking weights giving rise to P_u are identified by a dependent sequence $(V_{i,u})_{i \geqslant 1}$ such that

$$V_{1,u} \sim \text{N-GIG}^* \left(\tau_{1,u}, -\frac{1}{2} \right),$$

$$V_{i,u} \mid V_{1,u}, \ldots, V_{i-1,u} \sim \text{N-GIG}^* \left(\tau_{i,u}, -\frac{i}{2} \right) \quad (i \geqslant 2),$$

where $\tau_{1,u} = b(1+2u)$ and $\tau_{i,u} = b(1+2u)/\{\prod_{j=1}^{i-1}(1 - V_{j,u})\}$, for $i \geqslant 2$. Note, further, that

$$w_{0,u} = \frac{T_u}{T_u + \sum_{i=1}^{k} J_i}, \quad w_{j,u} = \frac{J_j}{T_u + \sum_{i=1}^{k} J_i} \quad (j \geqslant 1),$$

where J_i, for $i = 1, \ldots, k$, and T_u are independent and $J_i \sim \text{Ga}(n_i - 1/2,\ u + 1/2)$.

Acknowledgement

Detailed comments and suggestions by the editor, an associate editor and two referees are gratefully acknowledged. All authors are also affiliated to the Collegio Carlo Alberto and were supported by grants of the Italian Ministry of University and Research and Regione Piemonte.

References

Atkinson, A. C. (1982). The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM J. Sci. Statist. Comp.* **3**, 502–15.

Barndorff-Nielsen, O. E. & Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahr. verw. Geb.* **38**, 309–11.

Favaro, S., Hadjicharalambous, G. & Prünster, I. (2011). On a class of distributions on the simplex. *J. Statist. Plan. Infer.* **141**, 2987–3004.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–30.

Gradshteyn, L. S. & Ryzhik, L. M. (2000). *Table of Integrals, Series, and Products*. New York: Academic Press.

Hjort, N. L., Holmes, C. C., Müller, P. & Walker, S. G., Ed., (2010). *Bayesian Nonparametrics*. Cambridge: Cambridge University Press.

Ishwaran, H. & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Am. Statist. Assoc.* **96**, 161–73.

James, L. F., Lijoi, A. & Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Statist.* **33**, 105–20.

James, L. F., Lijoi, A. & Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.* **36**, 76–97.

Jara, A., Lesaffre, E., De Iorio, M. & Quintana F. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Ann. Appl. Statist.* **4**, 2126–49.

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. New York: Springer.

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

Lijoi, A., Mena, R. H. & Prünster, I. (2005). Hierarchical mixture modelling with normalized inverse Gaussian priors. *J. Am. Statist. Assoc.* **100**, 1278–91.

Lijoi, A., Mena, R. H. & Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering a new species. *Biometrika* **94**, 769–86.

Lijoi, A. & Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics,* Ed. N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker, pp. 80–136. Cambridge: Cambridge University Press.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria: American Statistical Association, 50–5.

Papaspiliopoulos, O. & Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169–86.

Perman, M., Pitman, J. & Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Prob. Theory Rel. Fields* **92**, 21–39.

PITMAN, J. (1996). Random discrete distribution invariant under sized-biased permutation. *Adv. Appl. Prob.* **28**, 525–39.

PITMAN, J. (2006). *Combinatorial Stochastic Processes.* Ecole d'Été de Probabilités de Saint-Flour XXXII. New York: Springer.

REGAZZINI, E., LIJOI, A. & PRÜNSTER, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–85.

SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–50.

TEH, Y. W. (2006). A hierarchical Bayesian language model based on Pitman-Yor processes. In *Proc. 21st Int. Conf. Comp. Linguistics and 44th Ann. Meeting Assoc. Comp. Linguistics*, pp. 985–92. Stroudsburg: Association for Computational Linguistics.

WALKER, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Commun. Statist.* B **36**, 45–54.

YAU, C., PAPASPILIOPOULOS, O., ROBERTS, G. O. & HOLMES, C. C. (2011). Bayesian non-parametric hidden Markov models with applications in genomics. *J. R. Statist. Soc.* B **73**, 37–57.

[*Received August* 2011. *Revised February* 2012]