On the stick–breaking representation of normalized inverse Gaussian priors

S. Favaro[1], A. Lijoi[2] and I. Prünster[3]

[1] Università degli Studi di Torino and Collegio Carlo Alberto.
E-mail: stefano.favaro@unito.it

[2] Università degli Studi di Pavia and Collegio Carlo Alberto.
E-mail: lijoi@unipv.it

[3] Università degli Studi di Torino and Collegio Carlo Alberto.
E-mail: igor@econ.unito.it

Abstract

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 (in distribution) equivalent, representations. Some of these are convenient if one wishes to thoroughly analyze the theoretical properties of the priors being used, others are more useful for modeling dependence and for addressing computational issues. As for the latter purpose, so–called *stick–breaking* constructions certainly stand out. In this paper we focus on the recently introduced normalized inverse Gaussian process and provide a completely explicit stick–breaking representation for it. Such a new result is of interest both from a theoretical viewpoint and for statistical practice.

Key words and phrases: Bayesian Nonparametrics; Dirichlet process; Normalized Inverse Gaussian process; Random Probability Measures; Stick–breaking representation.

1 Introduction

1.1 Bayesian Nonparametrics and the stick–breaking construction

Bayesian nonparametric inference has recently undergone strong development. See [12] 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 [8]. 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 [21], 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 [19], 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; it completes the study of the normalized inverse Gaussian process: most of its properties are known by now but a stick-breaking representation was missing. From a modeling and computational point of view: it paves the way for the definition of complex models based on the normalized inverse Gaussian process by simply replacing the stick-breaking constructed Dirichlet process, most notably within dependent models for nonparametric regression; it allows to extend recent simulation algorithms, based on stick-breaking constructions, to cover also the normalized inverse Gaussian process.

There are several different ways to define the Dirichlet process: each one has the merit of highlighting one of its peculiar aspects. The original definition of [8] 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 of the Dirichlet process, still due to [8], 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 [24] (recalled as Theorem 1 in [28]) under the assumption of non-atomic P_0, and that has been extended to any P_0 in [31]. Let $(V_i)_{i\geq 1}$ be a sequence of independent and identically distributed random variables, with $V_i \sim \text{beta}(1, c)$ and $c > 0$, and define random probability weights $(\tilde{p}_j)_{j\geq 1}$ as

$$p_1 = V_1, \qquad p_j = V_j \prod_{i=1}^{j-1}(1 - V_i) \quad j = 2, 3, \dots \ . \tag{1}$$

If $(Y_i)_{i\geq 1}$ 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 $\sum_{j\geq 1} \tilde{p}_j \, \delta_{Y_j}(\,\cdot\,) = \mathscr{D}_{c,P_0}(\,\cdot\,)$ in distribution, where δ_a denote the point unit mass at a.

Another prominent nonparametric prior is the two-parameter Poisson-Dirichlet process [26], also known, according to terminology introduced in [13], as Pitman-Yor process. It also admits a simple stick-breaking representation: let $(V_i)_{i\geq 1}$ be a sequence of independent random variables, with $V_i \sim \text{Beta}(1 - \sigma, \theta + i\sigma)$, $\sigma \in (0, 1)$ and $\theta > -\sigma$, define the random probability masses as in (1) and let $(Y_i)_{i\geq 1}$ be as above with the only difference of requiring a non-atomic P_0. The corresponding discrete random probability measure $\sum_{j\geq 1} \tilde{p}_j \, \delta_{Y_j}(\,\cdot\,)$ coincides, in distribution, with the two-parameter Poisson-Dirichlet process. See [29] for a detailed account on theoretical properties. Bayesian nonparametric applications include mixture models [13], linguistics [32], species sampling [20] and survival analysis [16].

The extreme flexibility of stick-breaking representations originated a vast literature concerning both modeling and computation. In terms of modeling, the proposal and investigation of

dependent processes, initiated by [22, 23], heavily rely on a stick-breaking construction and have proved to be effective prior specifications within regression problems. Important contributions to this area include, among others, [3, 4, 5, 6, 10, 11, 27]. From a computational point of view significant progress, especially in designing efficient simulation algorithms for hierarchical mixtures, has been made using stick-breaking representations. Among the most relevant contributions, which devise algorithms working in principle for any random probability with explicit stick–breaking representation, are the blocked Gibbs sampler [13], the retrospective sampler [25], the slice sampler [33] and a very efficient synthesis [34] 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 [13] limited to beta distributed V_i. However, their results are readily extendible to 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, to date no other instances are known and this clearly affects their appeal in terms of modeling 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 permutations and is recalled in Section 2. For the moment 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 [24], the Dirichlet process is the only random probability measure admitting stick-breaking representation with independent and identically distributed weights V_i which is invariant under size biased permutations. On the other hand, the two-parameter Poisson-Dirichlet process is essentially the only invariant under size biased permutations random probability measure that admits a stick-breaking representation with independent weights V_i [28]. These considerations hint towards the fact that if one would like to identify further random probability measures both enjoying a sufficient degree of tractability and admitting a sufficiently simple stick-breaking representation one has to focus on dependent stick-breaking weights V_i.

Here we consider the normalized inverse Gaussian process [19]. By now many properties such as finite-dimensional, predictive and posterior distributions are known [19, 14, 15]. 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 in this gap. It is important to anticipate that the stick-breaking weights V_i will be dependent: this fact is not surprising and actually necessary. To see why this is the case first note that any discrete random probability measure admits stick-breaking representation if one allows any possible distribution and form of dependence for the V_i [28]. Moreover, from [26] it is immediate to deduce that the normalized inverse Gaussian process, and more generally any homogeneous normalized random measure with independent increments [30, 14], is invariant under size biased permutations. Therefore, by the above mentioned characterizations provided in [24, 28], 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 necessarily have to 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, achieving a random probability whose properties can still be analyzed in more or less explicit form. If this is not the case, 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 $a > 0$, $b > 0$ and $p \in \mathbb{R}$, in symbols $X \sim \mathrm{GIG}(a, b, p)$, whose probability density function is of the form

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

with $\mathbb{1}_A$ denoting the indicator function of set A and K_ν the modified Bessel function of the third type. Moreover, let Z be a positive 1/2-stable random variable with scale parameter $b > 0$, $Z \sim \mathrm{St}_{1/2}(b)$, whose density is given by

$$f_Z(z) = \frac{b^{1/2}}{(2\,\pi)^{1/2}}\, z^{-3/2} \exp\left\{-\frac{1}{2}\frac{b}{z}\right\} \mathbb{1}_{(0,\infty)}(z).$$

Note that 1/2-stable random variables can also be seen as a reciprocal gamma random variables with shape parameter 1/2 and scale parameter $b/2$. Introduce, now, a sequence $(Z_i)_{i\geq 1}$ of inde-

pendent and identically distributed random variables with $Z_i \sim \mathrm{St}_{1/2}(1)$ and define a sequence of dependent $(0,1)$-valued random variables $(V_i)_{i\geq 1}$ as follows

$$V_1 = \frac{X_1}{X_1 + Z_1} \quad \text{s.t.} \quad X_1 \sim \mathrm{GIG}\left(a, 1, -\frac{1}{2}\right), \tag{3}$$
$$(V_i \mid V_1, \ldots, V_{i-1}) = \frac{X_i}{X_i + Z_i} \quad \text{s.t.} \quad X_i \sim \mathrm{GIG}\left(\frac{a}{\prod_{j=1}^{i-1}(1-V_j)}, 1, -\frac{i}{2}\right), \qquad i \geq 2,$$

where the sequences $(X_i)_{i\geq 1}$ and $(Z_i)_{i\geq 1}$ are independent. Sampling the V_i is straightforward [1]. It will be shown in Section 2 that the distribution of V_1 and of $V_i \mid V_1 \ldots, V_{i-1}$, for $i \geq 2$, in (3), is a special case of normalized generalized inverse Gaussian distribution and admits closed form density

$$f_{V_1}(v) = \frac{a^{\frac{1}{4}}\,(v)^{-\frac{1}{2}}\,(1-v)^{-1}}{(2\pi)^{\frac{1}{2}}\,K_{-1/2}(a^{\frac{1}{2}})} K_{-1}\left\{\left(\frac{a}{1-v}\right)^{\frac{1}{2}}\right\} \mathbb{1}_{(0,1)}(v) \tag{4}$$
$$f_{(V_i|V_1\ldots,V_{i-1})}(v) = \frac{\left(\frac{a}{\prod_{j=1}^{i-1}(1-V_j)}\right)^{\frac{1}{4}} (v)^{-\frac{1}{2}}\,(1-v)^{-\frac{5}{4}+\frac{i}{4}}}{(2\pi)^{\frac{1}{2}}\,K_{-\frac{i}{2}}\left\{\left(\frac{a}{\prod_{j=1}^{i-1}(1-V_j)}\right)^{\frac{1}{2}}\right\}} K_{-\frac{1}{2}-\frac{i}{2}}\left\{\left(\frac{\frac{a}{\prod_{j=1}^{i-1}(1-V_j)}}{1-v}\right)^{\frac{1}{2}}\right\} \mathbb{1}_{(0,1)}(v).$$

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

$$V_1 \sim \text{N-GIG}^*(a, -1/2) \;\;\&\;\; (V_i \mid V_1 \ldots, V_{i-1}) \sim \text{N-GIG}^*\left(\frac{a}{\prod_{j=1}^{i-1}(1-V_j)}, -\frac{i}{2}\right) \; i \geq 2. \tag{3'}$$

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

Proposition 1.1 *Let $(V_i)_{i\geq 1}$ be a sequence of dependent random variables as in (3), or equivalently (3′), and define the random probability weights $(\tilde{p}_j)_{j\geq 1}$ via stick-breaking as in (1). Let $(Y_i)_{i\geq 1}$ be a sequence of independent and identically distributed random variables, independent of the $\tilde{p}_i$ and with non-atomic distribution P_0. Then, for $c = a^{1/2}$,*

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

The proof, together with the necessary background and auxiliary results, is given in Section 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 Section 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 [17].

Definition 2.1 *Let X_1 and X_2 be two independent random variables such that $X_1 \sim GIG(a_1, b_1, p_1)$ and $X_2 \sim GIG(a_2, b_2, p_2)$. 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*

$$f_V(v) = \frac{\left(\frac{a_1}{b_1}\right)^{p_1/2}\left(\frac{a_2}{b_2}\right)^{p_2/2}}{2K_{p_1}\{(b_1a_1)^{1/2}\}K_{p_2}\{(b_2a_2)^{1/2}\}}\, v^{p_1-1}(1-v)^{p_2-1}\left(\frac{\frac{b_1}{v}+\frac{b_2}{1-v}}{a_1v+a_2(1-v)}\right)^{(p_1+p_2)/2}$$
$$\times K_{p_1+p_2}\left[\left\{\left(\frac{b_1}{v}+\frac{b_2}{1-v}\right)(a_1v+a_2(1-v))\right\}^{1/2}\right]\mathbb{1}_{(0,1)}(v). \tag{5}$$

The density displayed in (5) is obtained by application of a simple change of variable and formula 3.471.9 in [9]. Given that generalized inverse Gaussian random variables are infinitely divisible [2], the normalized generalized inverse Gaussian distribution represents another example of the class of normalized infinitely divisible distributions studied in [7].

For our purposes, two special cases are of particular interest. The first one is the normalized inverse Gaussian distribution [19], which corresponds to a N-GIG$(1, b_1, -1/2, 1, b_2, -1/2)$ distribution or, in other terms, to (5) with X_1 and X_2 being inverse Gaussian. Its density simplifies to

$$f_V(v) = \frac{(b_1b_2)^{1/2}e^{{b_1}^{1/2}+{b_2}^{1/2}}}{\pi}\,\frac{K_{-1}\left\{\left(\frac{b_1}{v}+\frac{b_2}{1-v}\right)^{1/2}\right\}}{v^{3/2}(1-v)^{3/2}\left(\frac{b_1}{v}+\frac{b_2}{1-v}\right)^{1/2}}, \tag{6}$$

which is seen to coincide with Eq. (5) in [19] by setting $\alpha_i = {b_i}^{1/2}$, for $i = 1,2$.

The second subclass of Definition 2.1 to be considered corresponds to the random variable (3′) dictating the form of the stick-breaking weights in Proposition 1.1 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

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

which reduces to (4) by setting $a_i = a\,(\prod_{j=1}^{i-1}(1-V_j))^{-1}$, $a > 0$ and $p_i = -i/2$.

2.2 The normalized inverse Gaussian process

The normalized inverse Gaussian process [19] enjoys a good degree of mathematical tractability and it is particularly effective when drawing inference on the clustering structure featured by the data. As for the Dirichlet process its original definition has been given in terms of a consistent system of finite-dimensional distributions, which is important to recall concisely. Consider n independent inverse Gaussian random variables X_i, which admit density as in (2) with $p = -1/2$. For our purpose, we can assume, without loss of generality, $a = 1$ 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 $W_i = X_i(\sum_{j=1}^n X_j)^{-1}$ for $i = 1, \ldots, n$, which admits density on the $(n-1)$-dimensional simplex Δ_{n-1} (w.r.t. the Lebesgue measure on $\mathbb{R}^{n-1}$) coinciding with

$$f(w_1, \ldots, w_{n-1}) = \frac{\mathrm{e}^{\sum_{i=1}^n {b_i}^{1/2}} \prod_{i=1}^n {b_i}^{1/2}}{2^{n/2-1}\pi^{n/2}} w_1^{-3/2} \cdots w_{n-1}^{-3/2} \left(1 - \sum_{i=1}^{n-1} w_i\right)^{-3/2} \\ \times \left(\mathcal{A}_n(w_1, \ldots, w_{n-1})\right)^{-n/4} K_{-\frac{n}{2}}\left(\left(\mathcal{A}_n(w_1, \ldots, w_{n-1})\right)^{1/2}\right), \tag{7}$$

where $\mathcal{A}_n(w_1, \ldots, w_{n-1}) := \sum_{i=1}^{n-1} b_i (w_i)^{-1} + b_n(1 - \sum_{j=1}^{n-1} w_j)^{-1}$. Clearly, (7) reduces to the marginal distribution (6) if $n = 2$. Based on the class of distributions in (7), in [19] a family of consistent finite-dimensional distribution is defined and it is shown that there exists a random probability measure P, termed normalized inverse Gaussian process with parameter measure $\alpha = c\, P_0$ and denoted by $\text{N-IG}_{c,\, P_0}$, having (7) as finite-dimensional distributions.

The second construction of a normalized inverse Gaussian process we will need is the one via normalization of an inverse Gaussian process, which corresponds to the definition of a Dirichlet process as normalized gamma process [8]. To this end first recall the important concept of completely random measure [18]: 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 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

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

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 completely random measure μ is representable as $\mu(\cdot) = \sum_{j \geq 1} J_j \delta_{Y_j}(\cdot)$. 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 jumps J_j, Moreover, without loss of generality, the locations $(Y_j)_{j\geq 1}$ can be assumed to be independent and identically distributed from a non-atomic 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 is *homogeneous*. And we further assume that ρ admits density and $\rho(\mathbb{R}^+) = \infty$. Note that if ρ has infinite total mass, then the completely random measure is strictly positive so that 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 [30, 14], 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

$$P(\cdot) = \frac{\mu(\cdot)}{T} = \sum_{j\geq 1} p_j \delta_{Y_j}(\cdot), \tag{8}$$

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

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

$$\rho(\mathrm{d}s)P_0(\mathrm{d}y) = \frac{b^{1/2}}{(2\pi)^{1/2}} s^{-3/2} \mathrm{e}^{-\frac{1}{2}a\,s} \mathrm{d}s\, P_0(\mathrm{d}y), \tag{9}$$

with $b > 0$ and where, without loss of generality for our scope, we can set $a = 1$. One then obtains the N-IG$_{b^{1/2},P_0}$ process as a homogeneous normalized random measure with independent increments (8) characterized by the Lévy intensity (9) with $a = 1$. The Dirichlet process is obtained by replacing the inverse Gaussian with a gamma completely random measure or, in other terms, (9) 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 $P = \sum_{j\geq 1} p_j \delta_{Y_j}$, the only constraint being that the locations $(Y_j)_{j\geq 1}$, which are are independent and identically distributed from a non-atomic probability measure P_0, are independent from the random probabilities $(p_j)_{j\geq 1}$. Note that homogeneous normalized random measures with independent increments defined in Section 2.3 fit into this general framework. An interesting rearrangement of the elements of $(p_j)_{j\geq 1}$ 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 $(X_n)_{n\geq 1}$ directed by the discrete random probability measure P. Correspondingly, define $(N_i)_{i\geq 1}$ as the successive times at which new values of the sequence $(X_n)_{n\geq 1}$ appear, namely $N_1 = 1$

and $N_j := \inf\{i > N_{j-1} : X_i \notin \{X_1, \ldots, X_{i-1}\}\}$ for any $j \geq 2$. Note that $\mathbb{P}[N_n > n] > 0$ since, due to the discreteness of P, ties will be recorded with positive probability. Moreover, let $(\xi_i)_{i\geq 1}$ be an integer valued sequence such that $\mathbb{P}[\xi_i = n \mid (p_j)_{j\geq 1}] = p_n$ and $\mathbb{P}[X_n = Y_{\xi_n} \mid (p_j)_{j\geq 1}, (Y_j)_{j\geq 1}, (\xi_i)_{i\geq 1}] = 1$. 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

$$\tilde{p}_i = p_{\xi_{N_i}} \tag{10}$$

for $i \geq 1$ with the convention $\tilde{p}_i = 0$ if the distinct values in $(X_n)_{n\geq 1}$ are fewer than i. The sequence $(\tilde{p}_j)_{j\geq 1}$ is termed size-biased permutation of $(p_j)_{j\geq 1}$. Moreover, $(p_j)_{j\geq 1}$, or the corresponding random probability measure P, is termed invariant under size-biased permutation if $(\tilde{p}_j)_{j\geq 1}$, whose coordinates are defined according to (10), has the same finite dimensional distributions as $(p_j)_{j\geq 1}$. See [28] and references therein for details.

The importance of invariance under size-biased permutations is best illustrated by the following considerations. Clearly stick–breaking priors can be defined whatever the choice of $[0, 1]$–valued random weights $(V_i)_{i\geq 1}$ provided they ensure $\sum_{i\geq 1} p_i = 1$, 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 permutations is essential. Indeed, if one aims at analyzing the clustering structure induced by P or at making predictions about the outcomes of future observations, one needs an expression for the exchangeable partition probability function and this can be hardly recovered unless P is invariant under size-biased permutations. To make this point clear, let $X_1, \ldots, X_n$ be a sample from an exchangeable sequence $(X_i)_{i\geq 1}$, directed by P, that features $K_n \leq n$ 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

$$\begin{aligned} p_k^{(n)}(n_1, \ldots, n_k) &:= \mathbb{P}(K_n = k, N_{1,n} = n_1, \ldots, N_{K_n,n} = n_k) \\ &= \sum_{i_1 \neq \cdots \neq i_k} \mathbb{E}(p_{i_j}^{n_1} \cdots p_{i_k}^{n_k}). \end{aligned} \tag{11}$$

From an operational standpoint the expression in (11) is not useful: for certain specifications of the stick-breaking random probabilities p_i's one may be able to compute the expected value $\mathbb{E}(p_{i_j}^{n_1} \cdots p_{i_k}^{n_k})$ but the sum over the indices $i_1, \cdots, i_k$ cannot be evaluated explicitly. Even numerically it is a highly demanding task already for moderately small values of k. Importantly, (11) can be re-expressed in much simpler form in terms of the size-biased permutation $(\tilde{p}_i)_{i\geq 1}$

of the sequence $(p_i)_{i\geq 1}$, namely

$$p_k^{(n)}(n_1,\ldots,n_k) = \mathbb{E}\left(\prod_{i=1}^{k}\tilde{p}_i^{n_i}\prod_{j=1}^{k-1}(1-\sum_{r=1}^{j}\tilde{p}_r)\right) \tag{12}$$

See Equation (8) in Pitman (1996). Now, if and only if P is invariant under size-biased permutations, the $\tilde{p}_i$'s in (12) can be replaced by the stick-breaking random probabilities p_i's. Consequently, $p_k^{(n)}$ can be more easily evaluated by using (12), instead of (11), with the the simple stick-breaking p_i's in place of their sized 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 (12) the well-known Ewens' sampling formula

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

with $(c)_n = c(c+1)\ldots(c+n-1)$ 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-sized biased permutations.

Now consider the class of homogeneous normalized random measures with independent increments (8) and denote by $(J_{(j)})_{j\geq 1}$ the sequence of jumps of μ rearranged in decreasing order. The corresponding sequence of random probability weights is then given by $p_{(j)} = J_{(j)}T^{-1}$ for any $j \geq 1$. [26] generalized the result in [24]. Indeed, they derived a stick-breaking characterization for the size-biased permutation $(\tilde{p}_j)_{j\geq 1}$ of the sequence of ranked random probabilities $(p_{(j)})_{j\geq 1}$ by providing a detailed description of the distribution of the sequence $(V_i)_{i\geq 1}$ within (1) in terms of the measure ρ and the distribution of the total mass T. Now turn attention back to the random probability measures. Since the locations $(Y_j)_{j\geq 1}$ are assumed to be independent and identically distributed from a non-atomic probability measure P_0 independent of the random probabilities $(p_j)_{j\geq 1}$ and given the sequences $(p_{(j)})_{j\geq 1}$ and $(\tilde{p}_j)_{j\geq 1}$ represent two specific rearrangements of the original sequence $(p_j)_{j\geq 1}$, one clearly has

$$P(\cdot) = \sum_{j\geq 1} p_{(j)}\delta_{Y_j}(\cdot) = \sum_{j\geq 1}\tilde{p}_j\delta_{Y_j}(\cdot) \tag{13}$$

in distribution. By combining this with the definition of the Dirichlet process as normalized Gamma process and the identity (13), one recovers the its 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 1.1 is now a quite straightforward application of Theorem 2.1 in [26]. Consider an inverse Gaussian completely random measure: let f_T be the density function of the corresponding total mass T

$$f_T(t) = \frac{e^{b^{1/2}} b^{1/2}}{(2\pi)^{1/2}} t^{-3/2} \exp\left\{-\frac{1}{2}\left(t + \frac{b}{t}\right)\right\}. \tag{14}$$

and denote by λ the density function of ρ in (9) given by

$$\lambda(s) = \frac{b^{1/2}}{(2\pi)^{1/2}} s^{-3/2} \exp\left\{-\frac{1}{2}s\right\}. \tag{15}$$

The normalized inverse Gaussian process, by the identity in (13) and its construction via normalization $P = \mu/T$ recalled in Section 2.2, can be represented as

$$P(\cdot) = \sum_{j\geq 1} \tilde{p}_j \delta_{Y_j}(\cdot) \tag{16}$$

where $(\tilde{p}_j)_{j\geq 1}$ is the size-biased permutation of the ranked random probabilities of P and $(Y_j)_{j\geq 1}$ 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 non-atomic.

Now, by Theorem 2.1 in [26], the sequence $(\tilde{p}_j)_{j\geq 1}$ in (16) has stick-breaking representation in terms of some sequence of dependent random variables $(V_i)_{i\geq 1}$. In particular, they provide a structural expression for the joint distribution of the random variables $(V_1, \ldots, V_i)$, for any $i \geq 1$, in terms of the density functions f_T and λ. We start by deriving the distribution of V_1. According to Eq. (2.d) in [26], the density function of the random variable V_1 is of the form

$$\begin{aligned} f_{V_1}(v_1) &= v_1 \int_0^{+\infty} t\lambda(v_1 t) f_T((1-v_1)t) dt \\ &= \frac{e^{b^{1/2}}\, b^{1/2}}{2\pi}\, v_1^{-1/2}(1-v_1)^{-3/2} \int_0^{+\infty} t^{-1-1} \exp\left\{-\frac{\frac{b}{2(1-v_1)}}{t} - \frac{t}{2}\right\} dt \end{aligned}$$

where the second identity is obtained by inserting (14) and (15). The density displayed in (4) is obtained by combining the identity $\mathrm{K}_{-1/2}(b^{1/2}) = \pi^{1/2} e^{-b^{1/2}} (2b)^{-1/2}$ with Formula 3.471.9 in [9] and setting $b = a$. The representation in (3), or (3′), follows by the definition of normalized generalized inverse Gaussian distribution given in Definition 2.1 and the corresponding density (5). Now consider the case $i = 2$: according to Eq. (2.d) in [26], the density function of the random variables (V_1, V_2) is

$$f_{V_1,V_2}(v_1, v_2) = v_1 v_2 (1-v_1) \int_0^{+\infty} t^2 \lambda(v_1 t)\lambda(v_2(1-v_1)t) f_T((1-v_1)(1-v_2)t) dt$$

$$= \frac{e^{b^{1/2}} b}{(2\pi)^{3/2}} v_1^{-1/2}(1-v_1)^{-2} v_2^{-1/2}(1-v_2)^{-3/2} \int_0^{+\infty} t^{-3/2-1} \exp\left\{ \frac{\frac{b}{2(1-v_1)(1-v_2)}}{t} - \frac{t}{2} \right\} dt$$

where the second identity is obtained by inserting (14) and (15). The density of $(V_2 \mid V_1)$ in (4) is obtained by using formula 3.471.9 in [9] and by dividing by the density of V_1, provided a in (4) is set equal to b appearing in the expression above. The representations in (3) or (3′) follow again by Definition 2.1. Proceeding along the same lines one obtains the density of $(V_i \mid V_1, \ldots, V_{i-1})$ displayed in (4) for any $i \geq 3$.

3 Concluding remarks

For drawing posterior inferences in complex models based on some stick-breaking prior, the knowledge of a posterior representation of the random probability is not necessary. 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 [15] and show how the stick-breaking construction as well as the normalized GIG distribution appear in it.

Consider a N-IG$_{b^{1/2},P_0}$ prior and suppose 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 $f_{U_n}(u) \propto u^{n-1}(u+1/2)^{k/2-n} \exp[-\{b(1+2u)\}^{1/2}]$. It can be shown that conditionally on $X_1, \ldots, X_n$ and on U_n the N-IG$_{b^{1/2},P_0}$ 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$_{\{b(1+2u)\}^{1/2},\, P_0}$ random probability measure. Hence, the stick-breaking weights giving rise to P_u are identified by a dependent sequence $(V_{i,u})_{i\geq 1}$ such that

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

$$V_{i,u} \mid V_{1,u}, \ldots, V_{i-1,u} \sim \text{N-GIG}^*\left(\frac{b(1+2u)}{\prod_{j=1}^{i-1}(1-V_{j,u})},\, -\frac{i}{2}\right).$$

Note, further, that

$$w_{0,u} = \frac{T_u}{T_u + \sum_{i=1}^k J_i} \qquad\qquad w_{j,u} = \frac{J_j}{T_u + \sum_{i=1}^k J_i} \quad j \geq 2$$

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

Acknowledgements

The authors are partially supported by MIUR, Grant 2008MK3AFZ, and Regione Piemonte.

References

[1] Atkinson, A. C. (1982). The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM J. Sci. Statist. Comput.* **3**, 502-515.

[2] Barndorff-Nielsen, O. E. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **38**, 309-311.

[3] De Iorio, M., Müller, P., Rosner, G. and MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *J. Amer. Statist. Assoc.* **99**, 205-215.

[4] Duan, J. A., Guindani, M. and Gelfand, A. E. (2007). Generalized spatial Dirichlet process models. *Biometrika* **94**, 809825.

[5] Dunson, D. B. and Park, J-H. (2008). Kernel stick breaking processes. *Biometrika* **95**, 307-323.

[6] Dunson, D. B., Xue, Y. and Carin, L. (2008). The matrix stick breaking process: Flexible Bayes meta analysis. *J. Amer. Statist. Assoc.* **103**, 317-327.

[7] Favaro, S., Hadjicharalambous, G. and Prünster, I. (2011). On a class of distributions on the simplex. *J. Statist. Plann. Inference* **141**, 2987-3004.

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

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

[10] Griffin, J. and Steel, M. F. J. (2006). Order-Based Dependent Dirichlet Processes. *J. Amer. Stat. Assoc.* **101**, 179-194.

[11] Griffin, J. and Steel, M. F. J. (2011). Stick-Breaking Autoregressive Processes. *J. Econometrics* **162**, 383-396.

[12] Hjort, N. L., Holmes, C. C. Müller, P., Walker, S. G. (Eds.) (2010). *Bayesian Nonparametrics.* Cambridge University Press, Cambridge.

[13] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Ass.* **96**, 161-173.

[14] James, L. F., Lijoi, A. and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Statist.* **33**, 105-120.

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

[16] Jara, A., Lesaffre, E., De Iorio, M. and Quintana F. (2010). Bayesian Semiparametric Inference for Multivariate Doubly-Interval-Censored Data. *Ann. Appl. Stat.* **4**, 2126-2149.

[17] Jørgensen, B. (1981). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer.

[18] Kingman, J. F. C. (1993). *Poisson Processes*. Oxford University Press, Oxford.

[19] Lijoi, A., Mena, R. H. and Prünster, I. (2005). Hierarchical mixture modelling with normalized inverse Gaussian priors. *J. Amer. Stat. Assoc.* **100**, 1278-1291.

[20] Lijoi, A., Mena, R.H. and Prünster, I. (2007). Bayesian nonparametric estimation of the probability of discovering a new species. *Biometrika* **94**, 769-786.

[21] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (Hjort, N. L., Holmes, C. C. Müller, P., Walker, S. G. Eds.), pp. 80-136. Cambridge University Press, Cambridge.

[22] MacEachern, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. Alexandria: American Statistical Association, 50-55.

[23] MacEachern, S.N. (2000). Dependent Dirichlet processes. *Technical Report*. Department of Statistics, Ohio State University.

[24] McCloskey, J. W. (1965). *A model for the distribution of individuals by species in an environment*. Ph.D. thesis, Michigan State University.

[25] Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169-186.

[26] Perman, M., Pitman, J. and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92**, 21-39.

[27] Petrone, S., Guindani, M., and Gelfand, A. E. (2009). Hybrid Dirichlet mixture models for functional data. *J. R. Stat. Soc. Ser. B* **71**, 755-782.

[28] Pitman, J. (1996). Random discrete distribution invariant under sized-biased permutation. *Adv. Appl. Probab.* **28**, 525-539.

[29] Pitman, J. (2006). *Combinatorial Stochastic Processes.* Ecole d'Eté de Probabilités de Saint-Flour XXXII. Springer, New York.

[30] Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560-585.

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

[32] Teh, Y. W. (2006). A Hierarchical Bayesian Language Model based on Pitman-Yor Processes. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 985-992. Association for Computational Linguistics, Stroudsburg.

[33] Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Comm. Statist. Simul. Computat.* **36**, 45-54.

[34] Yau, C., Papaspiliopoulos, O., Roberts, G.O. & Holmes, C. (2011). Bayesian non-parametric hidden Markov models with applications in genomics., *J. Roy. Statist. Soc. Ser. B* **73**, 37-57.