On the Stick-Breaking Representation for Homogeneous NRMIs

S. Favaro[*], A. Lijoi[†], C. Nava[‡], B. Nipoti[§], I. Prünster[¶], and Y. W. Teh[‖]

Abstract. In this paper, we consider homogeneous normalized random measures with independent increments (hNRMI), a class of nonparametric priors recently introduced in the literature. Many of their distributional properties are known by now but their stick-breaking representation is missing. Here we display such a representation, which will feature dependent stick-breaking weights, and then derive explicit versions for noteworthy special cases of hNRMI. Posterior characterizations are also discussed. Finally, we devise an algorithm for slice sampling mixture models based on hNRMIs, which relies on the representation we have obtained, and implement it to analyze real data.

Keywords: Bayesian Nonparametrics, generalized Dirichlet process, normalized generalized gamma process, normalized random measures with independent increments, normalized stable process, normalized inverse-Gaussian process, random probability measure, stick-breaking representation.

1 Introduction

Normalized random measures with independent increments (NRMIs) have been introduced in Regazzini et al. (2003) and represent a broad class of nonparametric priors, which includes the Dirichlet process prior as a special case. Other notable members of this class are the σ-stable NRMI (Kingman, 1975), the normalized inverse-Gaussian process (Lijoi et al., 2005b), the generalized gamma NRMI (Lijoi et al., 2007; Pitman, 2003), and a generalized Dirichlet process obtained via normalization of superposed gamma processes (Regazzini et al., 2003; Lijoi et al., 2005a). Distributional properties of general NRMI have been studied in great detail so far. For example, characterizations of their predictive structure and of their posterior distribution are provided in James et al. (2009). See also Lijoi and Prünster (2010) for a review.

The probability distribution of an NRMI $\tilde{p}$ is a discrete nonparametric prior thus implying that

$$\tilde{p} = \sum_{j \geq 1} \tilde{p}_j \, \delta_{Y_j} \tag{1}$$

where the $\tilde{p}_j$'s are random weights such that $\sum_{j \geq 1} \tilde{p}_j = 1$, almost surely, the Y_j's are random locations, and δ_c is a unit point mass at c. An important open problem

[*]University of Torino and Collegio Carlo Alberto, Italy, stefano.favaro@unito.it
[†]University of Pavia and Collegio Carlo Alberto, Italy, lijoi@unipv.it
[‡]University of Torino and Collegio Carlo Alberto, Italy, consuelo.nava@carloalberto.org
[§]University of Torino and Collegio Carlo Alberto, Italy, bernardo.nipoti@unito.it
[¶]University of Torino and Collegio Carlo Alberto, Italy, igor.pruenster@unito.it
[‖]University of Oxford, UK, y.w.teh@stats.ox.ac.uk

concerning NRMIs is to provide their stick-breaking representation. This amounts to identifying a sequence of $[0, 1]$-valued random variables $(V_i)_{i \geq 1}$ such that the random weights of an NRMI in (1) coincide in distribution with

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

The rationale of the term "stick-breaking" is apparent. Suppose one has a unit length stick and breaks it into two bits of length V_1 and $1 - V_1$. The first bit represents $\tilde{p}_1$ and in order to obtain $\tilde{p}_2$ it is enough to split the remaining part, of length $1 - V_1$, into two parts having respective lengths $V_2(1 - V_1)$ and $(1 - V_2)(1 - V_1)$. The former will coincide with $\tilde{p}_2$ and the latter will be split to generate $\tilde{p}_3$, and so on. Stick-breaking priors are very popular in Bayesian Nonparametrics practice, mainly because of the availability of computational algorithms that allow sampling both a priori and a posteriori; see Ishwaran and James (2001); Papaspiliopoulos and Roberts (2008); Walker (2007); Yau et al. (2011). Moreover, the stick-breaking construction has proved to be very fruitful for defining more general models capable of dealing with partially exchangeable observations; see the seminal contributions by MacEachern (1999, 2000), whereas a recent account can be found in the monograph edited by Hjort et al. (2010).

A description of the sequence $(V_i)_{i \geq 1}$ that characterizes various well-known nonparametric priors has been provided in the literature. The first important contribution is Sethuraman (1994) who has shown that the Dirichlet process admits a representation as in (1) with independent and identically distributed (i.i.d.) stick-breaking weights (2) of the form

$$V_i \sim \text{beta}(1, a), \tag{3}$$

namely a beta distribution with parameters 1 and a, where a stands for the total mass parameter of the Dirichlet process. Other members of the class of NRMIs for which a stick-breaking representation is known are the σ-stable NRMI (Perman et al., 1992) and recently the normalized inverse-Gaussian process (Favaro et al., 2012). It is worth pointing out that in the former case the V_i's are still independent but, unlike in the Dirichlet process, not identically distributed, whereas in the latter case the V_i's are neither identically distributed nor independent. Outside the class of NRMIs, Perman et al. (1992) identify the stick-breaking representation of the two-parameter Poisson–Dirichlet process which admits a particularly simple stick-breaking representation with the V_i's in (2) being independent and

$$V_i \sim \text{beta}(1 - \sigma, \theta + i\sigma), \qquad \sigma > 0, \ \theta > -\sigma. \tag{4}$$

Moreover, Favaro and Walker (2013) provide a stick-breaking representation for Gibbs-type priors, a class of nonparametric priors due to Gnedin and Pitman (2006). Finally, Teh et al. (2007) derive a stick-breaking representation of the beta process: although not a random probability measure, this is a key model in various applications involving the Bayesian nonparametric approach.

In this paper, we establish the stick-breaking representation for general *homogeneous* NRMIs, which essentially correspond to NRMIs whose weights $\tilde{p}_j$'s and locations Y_j's in

(1) are independent. The results will then be specialized to two cases, namely the generalized Dirichlet process and the generalized gamma NRMI: these have already found applications in several contexts, most notably mixture modeling and species sampling. Similarly to what happens in the case of the normalized inverse-Gaussian process, the weights V_i are dependent. We shall then focus on the insights that the stick-breaking representation yields for these processes in terms of their posterior characterizations. It is important to stress that characterizations for NRMIs are useful also beyond the exchangeable setting where a discrete nonparametric is specified directly on the data. In general, when a discrete random probability measure is used at a latent level within complex nonparametric models, its posterior structure, conditional on the latent variables it generates and the other model components, is analogous to that of the simple exchangeable case considered here. This remark clearly applies also to NRMIs, which are large but still quite tractable class of random probability measures. It then comes to no surprise that, in addition to mixture models and dependent processes, some other nice uses of NRMIs appeared recently. For instance, complex models featuring NRMIs as key components were considered in Caron (2012), Caron and Fox (2015) and Caron et al. (2014) for analyzing data related to rankings and graphs.

The paper is structured as follows. In Section 2, we recall some basic elements on completely random measures, define NRMIs and the two special cases we will consider in detail. In Section 3, we establish their stick-breaking representation. Their posterior characterization is discussed in Section 4. In Section 5, we devise an algorithm for sampling mixture models based on hNRMIs that, in turn, is implemented in Section 5.3 to analyze real data. Section 6 is dedicated to the concluding remarks. All proofs are gathered in Section 7.

2 Normalized random measures with independent increments

We start by recalling the notions of completely random measure and of NRMI. The former is a popular tool for defining, under suitable transformations, a random probability measure whose distribution acts as a prior for Bayesian inference; see Lijoi and Prünster (2010) for a review of discrete nonparametric priors using completely random measures as a unifying concept.

Let $\mathbb{X}$ be a separable and complete metric space with $\mathscr{X}$ denoting its Borel σ-algebra. Moreover, $M_{\mathbb{X}}$ is the space of boundedly finite measures on $\mathbb{X}$ which is endowed with a suitable topology that allows one to introduce the associated Borel σ-algebra $\mathscr{M}_{\mathbb{X}}$. For details on the definition of $(M_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ see Daley and Vere-Jones (2008).

Definition 1. *A measurable function $\tilde{\mu}$ defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $(M_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ is a* completely random measure *(CRM) if for any $d \geq 2$ and any pairwise disjoint sets $A_1, \ldots, A_d$ in $\mathscr{X}$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_d)$ are mutually independent.*

CRMs have been introduced in Kingman (1967) and a detailed treatment can be found in Kingman (1993). For our purposes it is important to recall that CRMs are

almost surely discrete, i.e., $\tilde{\mu}$ can always be represented as $\sum_{i \geq 1} J_i \, \delta_{Y_i}$, and that the Laplace transform of $\tilde{\mu}(B)$, for any B in $\mathscr{X}$, admits the following representation:

$$\mathbb{E}\left[\mathrm{e}^{-\lambda \tilde{\mu}(B)}\right] = \exp\left\{-\int_{\mathbb{R}^+ \times B}\left[1 - \mathrm{e}^{-\lambda v}\right] \nu(\mathrm{d}v, \mathrm{d}x)\right\} \tag{5}$$

for any $\lambda > 0$, with ν being a measure on $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\int_B \int_{\mathbb{R}^+} \min\{v, 1\} \, \nu(\mathrm{d}v, \mathrm{d}x) < \infty \tag{6}$$

for any B in $\mathscr{X}$. The measure ν is known as the *Lévy intensity* of $\tilde{\mu}$ and regulates the intensity of the jumps of a CRM and their locations. By virtue of (5), it characterizes the CRM $\tilde{\mu}$. In the following, we shall assume that ν can be factorized as

$$\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v) \, \alpha(\mathrm{d}x) \tag{7}$$

where ρ is some measure on $\mathbb{R}^+$ and α a measure on $\mathbb{X}$. This factorization assumption is equivalent to dealing with CRMs $\tilde{\mu} = \sum_{i \geq 1} J_i \delta_{Y_i}$ for which the jumps J_i and the locations Y_i are independent. When $\tilde{\mu}$ has an intensity for which (7) holds true, it is termed *homogeneous* CRM.

As mentioned above, one can define random probability measures via suitable transformations of a CRM. A natural transformation to consider is "normalization". For such an operation to be well-defined, one needs the total (random) mass $\tilde{\mu}(\mathbb{X})$ to be both positive and finite, almost surely. Necessary and sufficient conditions that ensure that $0 < \tilde{\mu}(\mathbb{X}) < \infty$, with probability 1, are

$$\rho(\mathbb{R}^+) = \infty \quad \text{and} \quad \alpha(\mathbb{X}) \in (0, \infty); \tag{8}$$

see Regazzini et al. (2003). The first condition corresponds to requiring the CRM to jump infinitely often on any finite set and essentially boils down to ruling out compound Poisson processes. The second condition allows us to write $\alpha = a P_0$ with $a := \alpha(\mathbb{X}) > 0$ and P_0 a probability measure, a notation often used in the sequel.

Definition 2. *Let $\tilde{\mu}$ be a homogeneous CRM on $\mathbb{X}$ such that the conditions in* (8) *hold true. Then a* homogeneous normalized random measure with independent increments *(hNRMI) $\tilde{p}$ on $(\mathbb{X}, \mathscr{X})$ is defined as*

$$\tilde{p}(\cdot) = \frac{\tilde{\mu}(\cdot)}{\tilde{\mu}(\mathbb{X})}.$$

The discreteness of CRMs clearly implies that an hNRMI $\tilde{p}$ is discrete, almost surely, and it can be represented as in (1) with $\tilde{p}_i = J_i / \sum_{l \geq 1} J_l$ for any $i \geq 1$ and the random locations Y_i i.i.d. with distribution $P_0 = \alpha/a$. Moreover, by virtue of the homogeneity assumption (7), the random probability masses $\tilde{p}_i$ are independent from the Y_i's. In view of this remark, and with the additional assumption of α being non-atomic, hN-RMIs represent a subclass of *Poisson–Kingman models* (Pitman, 2003) and, a fortiori,

of *species sampling models* (Pitman, 1996b). Clearly, nonhomogeneous NRMI do not belong to these classes of models.

We now close this section by recalling the two special cases of homogeneous NRMIs for which a stick-breaking representation will be discussed. They can be seen as different generalizations of the Dirichlet process.

2.1   Generalized Dirichlet process

Consider a CRM $\tilde{\mu}$ characterized by the Lévy intensity (7) with

$$\rho(\mathrm{d}v) = \sum_{j=1}^{\gamma} \frac{\mathrm{e}^{-j\,v}}{v}\,\mathrm{d}v. \tag{9}$$

where γ is a positive integer. The corresponding Laplace transform (5) admits a closed form expression given by

$$\mathbb{E}\left[\mathrm{e}^{-\lambda\tilde{\mu}(B)}\right] = \frac{(\gamma!)^{\alpha(B)}}{\left\{(\lambda+1)_{\gamma}\right\}^{\alpha(B)}} \tag{10}$$

where, for any integer n and real $c > 0$, $(c)_n = \Gamma(c+n)/\Gamma(c)$ is the n–th ascending factorial of c. The expression in (10) also suggests an interesting interpretation of $\tilde{\mu}$. Indeed, for any B in $\mathscr{X}$ such that $\alpha(B) > 0$, the random variable $\tilde{\mu}(B)$ is distributed as the sum of γ independent gamma random variables with shape and scale parameters $\alpha(B)$ and j, respectively, for $j = 1, \ldots, \gamma$. In other terms, $\tilde{\mu}(B)$ is a generalized gamma convolution with Thorin measure $t_B = \alpha(B) \sum_{j=1}^{\gamma} \delta_j$. See Bondesson (1992) for this noteworthy class of infinitely divisible distributions. Incidentally, note that when $\mathbb{X} = \mathbb{R}^+$, such a CRM coincides with a Lévy process that belongs to a special class for which quasi–invariance properties have been studied in Von Renesse et al. (2008).

The hNRMI obtained by normalizing a homogeneous CRM $\tilde{\mu}$ with Lévy intensity as in (9) is termed *generalized Dirichlet process* with parameters (γ, a, P_0), where $\alpha = aP_0$. In symbols $\tilde{p} \sim \mathrm{GD}(\gamma, a, P_0)$. Note that when $\gamma = 1$, $\tilde{\mu}$ reduces to a gamma CRM and the resulting $\tilde{p}$ is the Dirichlet process with base measure α introduced in Ferguson (1973). The present generalization of the Dirichlet process has first appeared in Regazzini et al. (2003) where the probability distribution of its mean functional is studied. Lijoi et al. (2005a) determined the finite–dimensional distributions of $\mathrm{GD}(\gamma, a, P_0)$ and expressions for their predictive distributions and used them for density estimation within mixture models. Moreover, in Favaro et al. (2011) $\mathrm{GD}(\gamma, a, P_0)$ models have been explored in relation to species sampling problems.

2.2   Normalized generalized gamma process

For any $\sigma \in (0,1)$ and $\tau > 0$, the Lévy intensity (7) identified by

$$\rho(\mathrm{d}v) = \frac{1}{\Gamma(1-\sigma)}\, v^{-1-\sigma}\, \mathrm{e}^{-\tau v}\,\mathrm{d}v \tag{11}$$

gives rise to a CRM $\tilde{\mu}$ that takes on the name of *generalized gamma process*; see Brix (1999) where this CRM has been originally proposed. The Laplace transform of $\tilde{\mu}(B)$ coincides with

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(B)}\right] = \exp\left\{-\frac{\alpha(B)}{\sigma}\left[(\lambda+\tau)^{\sigma} - \tau^{\sigma}\right]\right\}. \tag{12}$$

When $\tau \to 0$, the Laplace transform in (12) reduces to $\exp\{-\alpha(B)\lambda^{\sigma}/\sigma\}$ and $\tilde{\mu}$ is a σ-stable CRM. Its normalization yields the σ-stable hNRMI introduced by Kingman (1975). On the other hand, if $\sigma \to 0$, then the Laplace transform in (12) tends to $[\tau/(\lambda+\tau)]^{\alpha(B)}$ which is the Laplace transform of a gamma random variable with shape and scale parameter equal to $\alpha(B)$ and τ, respectively. Hence, the normalization of $\tilde{\mu}$ in this case identifies a Dirichlet process with base measure α. Another noteworthy instance is obtained by setting $\sigma = 1/2$ in (11) which leads to the inverse Gaussian CRM and, after normalization, to the normalized inverse-Gaussian process (Lijoi et al., 2005b). The random probability measure obtained by normalizing $\tilde{\mu}$ characterized by (11) encompasses all this special cases and is termed *normalized generalized gamma process*. Its parameters are (σ, β, P_0) where $\beta = a\tau^{\sigma}/\sigma$ with $\alpha = aP_0$ and it will be denoted as $\tilde{p} \sim \mathrm{NGG}(\sigma, \beta, P_0)$; see Lijoi et al. (2007). Alternatively, such a random probability measure can also be constructed as exponentially tilted σ-stable Poisson–Kingman model as shown in Pitman (2003).

3   Stick-breaking representation

As mentioned in Section 1, stick-breaking representations are available for some hNRMIs commonly used in Bayesian nonparametric modeling such as, e.g., the Dirichlet process and the σ-stable hNRMI. For both processes, the V_i's in (2) are independent, with identity in distribution holding true only for the Dirichlet process. More recently, Favaro et al. (2012) have obtained a stick-breaking representation of the normalized inverse-Gaussian process and the ith weight V_i, conditional on $(V_1, \ldots, V_{i-1})$, has a normalized generalized inverse Gaussian distribution with parameters depending on $(V_1, \ldots, V_{i-1})$. Here we go beyond these three cases and provide a general result that holds true for any hNRMI.

Before proceeding, first note that the dependence among the V_i weights in the normalized inverse-Gaussian case is not surprising and actually will be the general rule for hNRMIs. This can be explained by the fact that hNRMIs are invariant under size-biased permutation. Indeed, by virtue of a result in Pitman (1996a), the only invariant under size-biased permutation random probability measure whose random masses admit a stick-breaking representation with independent weights is the two-parameter Poisson–Dirichlet process. All other random probability measures enjoying the invariance under size-biased permutation property will necessarily have dependent stick-breaking weights (2). Now, as far as the class of hNRMIs is concerned, the only two members that can be seen as instances of the two-parameter Poisson–Dirichlet process are the Dirichlet and the normalized σ-stable processes. Hence, beyond these two cases, the stick-breaking representation of hNRMIs will have dependent weights and, for our purposes, it is enough to record this fact; see Pitman (1996b) and Favaro et al. (2012) for a discussion

of the implications of invariance under size-biased permutation for the computation of quantities of interest in Bayesian inference.

The following result, which essentially is an adaptation of Theorem 2.1 in Perman et al. (1992), displays the conditional distributions of each V_i, given $V_1, \ldots, V_{i-1}$, in terms of the Lévy intensity of the CRM $\tilde{\mu}$ underlying the hNRMI at issue and of the density function f_T of the total mass $\tilde{\mu}(\mathbb{X})$. In order to simplify notation, henceforth we set $w_1 = 1$ and $w_j = \prod_{i=1}^{j-1}(1 - v_i)$, for $j = 2, 3, \ldots$.

Proposition 1. *Suppose $\tilde{p} = \sum_{i \geq 1} \tilde{p}_i \delta_{Y_i}$ is an hNRMI characterized by a Lévy intensity ν as in (7) where $\rho(\mathrm{d}v) = \rho(v)\mathrm{d}v$. Then the random probability masses $(\tilde{p}_i)_{i \geq 1}$ admit stick-breaking representation (2) with weights sequence $(V_i)_{i \geq 1}$ such that the distribution of V_1 has support on $(0, 1)$ and density (with respect to the Lebesgue measure) given by*

$$f_{V_1}(v_1) = a\, v_1 \int_0^{+\infty} t\, \rho(tv_1)\, f_T(tw_2)\mathrm{d}t \tag{13}$$

and the conditional distribution of $V_i | V_1, \ldots, V_{i-1}$ has support on $(0, 1)$ and density (with respect to the Lebesgue measure)

$$f_{V_i}(v_i | v_1, v_2, \ldots, v_{i-1}) = a\, v_i\, w_i\, \frac{\int_0^{+\infty} t^i \prod_{j=1}^i \rho(tv_j w_j) f_T(tw_{i+1})\mathrm{d}t}{\int_0^{+\infty} t^{i-1} \prod_{j=1}^{i-1} \rho(tv_j w_j) f_T(tw_i)\mathrm{d}t}. \tag{14}$$

Note that the stick-breaking representation (3) of the Dirichlet process can be easily recovered from Proposition 1. In this case, $f_T(t) = t^{a-1}\mathrm{e}^{-t}/\Gamma(a)$ and $\rho(v) = \mathrm{e}^{-v}\, v^{-1}$. If these are plugged into (13) and (14), the integrals can be computed quite straightforwardly and the representation (3) easily follows.

Relying on this general result, which characterizes the stick-breaking weights in terms of the jump part of the Lévy intensity ρ and the law of the (random) total mass $\tilde{\mu}(\mathbb{X})$, we now provide descriptions of the stick-breaking representations of the two special cases of hNRMIs sketched in Section 2.

3.1 Generalized Dirichlet process

An application of Proposition 1, combined with the availability of the density function f_T of $\tilde{\mu}(\mathbb{X})$, leads to a description of the distribution of the stick-breaking weights of a generalized Dirichlet process $\tilde{p} \sim \mathrm{GD}(\alpha, \gamma)$. Note that, when $\gamma > 1$, from expression (7.6.4) in Exton (1976) one has

$$f_T(t) = \frac{(\gamma!)^a}{\Gamma(\gamma a)}\mathrm{e}^{-\gamma t} t^{\gamma a - 1} \Phi_2^{(\gamma-1)}(a^{(\gamma-1)}; \gamma a; t, 2t, \ldots, (\gamma-1)t) \tag{15}$$

where $\Phi_2^{(n)}$ is the confluent form of the fourth Lauricella hypergeometric function and $a^{(\gamma-1)}$ is a vector of size $\gamma - 1$ with all the components equal to a.

Proposition 2. *If $\tilde{p} \sim GD(\gamma, a, P_0)$, then its stick-breaking representation has dependent weights such that V_1 has density function given by*

$$f_{V_1}(v_1) = a\,(\gamma!)^a w_2^{-1} \sum_{j_1=0}^{\gamma-1} z_1^{-a} \left\{ (-1)^{\gamma-1}\,(1-z_1)_{\gamma-1} \right\}^{-a} \tag{16}$$

where $z_1 = (\bar{v}_1 + \gamma w_2)/w_2$ with $\bar{v}_1 = v_1(1+j_1)$. Moreover, for any $i \geq 2$, the conditional distribution of V_i, given $V_1, \ldots, V_{i-1}$ has density

$$f_{V_i}(v_i|v_1,\ldots,v_{i-1}) = a\,\frac{v_i w_i (1-v_i)^{-1}}{\sum_{C_{i-1}} z_{i-1}^{-a} \left\{ (1-z_{i-1})_{\gamma-1} \right\}^{-a}} \sum_{C_i} z_i^{-a} \left\{ (1-z_i)_{\gamma-1} \right\}^{-a} \tag{17}$$

where C_i is the set of all indices $(j_1,\ldots,j_i) \in \{0,\ldots,\gamma-1\}^i$, $z_i = (\bar{v}_i + \gamma w_{i+1})/w_{i+1}$ and $\bar{v}_i = \sum_{\ell=1}^{i} v_\ell w_\ell (1+j_\ell)$.

3.2 The normalized generalized gamma process

As already noted, the Dirichlet, the normalized σ-stable and the normalized inverse-Gaussian processes, for which a stick-breaking representation is known, are special instances of normalized generalized gamma processes. It is therefore of interest to derive an explicit stick-breaking representation for this rich subclass of hNRMI. By resorting to Proposition 1, we are now in a position to provide such a result.

Proposition 3. *Suppose $\tilde{p}$ is a normalized generalized gamma process obtained by normalizing a CRM having intensity (11). Then $\tilde{p}$ admits stick-breaking representation in terms of a dependent stick-breaking sequence $(V_i)_{i\geq 1}$ such that V_1 has density*

$$f_{V_1}(v_1) = \frac{v_1^{-\sigma}(1-v_1)^{\sigma-1}\,e^\beta}{\Gamma(\sigma)\Gamma(1-\sigma)} \sum_{j\geq 0} \frac{(1-\sigma)_j}{j!} \frac{\beta^{\frac{j}{\sigma}}}{(1-v_1)^j}\,\Gamma\left(1-\frac{j}{\sigma};\frac{\beta}{(1-v_1)^\sigma}\right) \tag{18}$$

where $\beta = a\tau^\sigma/\sigma$ and, for any $i \geq 2$ the conditional density of V_i, given $V_1,\ldots,V_{i-1}$, is

$$f_{V_i}(v_i|v_1,\ldots,v_{i-1})$$

$$= \frac{\sigma\,\Gamma((i-1)\sigma)}{\Gamma(1-\sigma)\,\Gamma(i\sigma)}\,v_i^{-\sigma}(1-v_i)^{i\sigma-1}\,\frac{\sum_{j\geq 0} \frac{(1-i\sigma)_j}{j!} \frac{\beta^{\frac{j}{\sigma}}}{w_{i+1}^j}\Gamma\left(i-\frac{j}{\sigma};\frac{\beta}{w_{i+1}^\sigma}\right)}{\sum_{j\geq 0} \frac{(1-(i-1)\sigma)_j}{j!} \frac{\beta^{\frac{j}{\sigma}}}{w_i^j}\Gamma\left(i-1-\frac{j}{\sigma};\frac{\beta}{w_i^\sigma}\right)}, \tag{19}$$

where the w_i's are defined immediately before the statement of Proposition 1.

The stick-breaking representation for the σ-stable NRMI can be recovered as a special case, by letting $\beta \to 0$. In fact, the sum in (18) and the ratio of sums in (19) converge to 1 and $i-1$, respectively. This leads to an independent sequence $(V_i)_{i\geq 1}$, such that $V_i \sim \text{beta}(1-\sigma, i\sigma)$, for $i \geq 1$. A second remarkable special case, which can be recovered via Proposition 3, is the normalized inverse-Gaussian process. The stick-breaking

representation of such process is studied in Favaro et al. (2012) where closed-form expressions for the distributions of the V_i's are provided. The same expressions can be, in principle, recovered from (18) and (19). Nonetheless, a more direct approach relies on the availability of a closed form expression for f_T, namely a simple inverse Gaussian density. This can be plugged into Proposition 1 to obtain the result.

Remark. In addition to being a subclass of hNRMIs (and of general NRMIs), normalized generalized gamma processes belong also to a different class of nonparametric priors, namely Gibbs-type priors which have been introduced by Gnedin and Pitman (2006). As shown in Lijoi et al. (2008), NGG(σ, β, P_0) processes are the only NRMIs being also of Gibbs-type. Note that in the Gibbs framework the NGG(σ, β, P_0) process is obtained as a suitable mixture of normalized σ-stable CRM rather than by normalizing a generalized gamma CRM but the representations are equivalent in distribution. Now, a stick-breaking representation for Gibbs-type priors has been derived in Favaro and Walker (2013), and therefore one can alternatively use their result, instead of Proposition 1, as the starting point for deriving a stick-breaking characterization of the NGG(σ, β, P_0) process given in Proposition 3.

4 Posterior representation

In this section, we provide an explicit representation of the posterior distribution of an hNRMI represented in stick-breaking form. This has the merit of shedding light on the structure underlying such processes.

Start by considering a sequence of $\mathbb{X}$-valued exchangeable random elements $(X_n)_{n \geq 1}$, defined on $(\Omega, \mathscr{F}, \mathbb{P})$. Further assume that the sequence is directed by an hNRMI or, equivalently, that the de Finetti measure of the exchangeable sequence is the law of an hNRMI. Then, the X_i's are conditionally i.i.d. given a hNRMI $\tilde{p}$, i.e.,

$$
\begin{aligned}
X_i \,|\, \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p}, \qquad i = 1, \ldots, n, \\
\tilde{p} &\sim \text{hNRMI}.
\end{aligned}
$$

Given the discreteness of hNRMI, a sample $X^{(n)} := (X_1, \ldots, X_n)$ will feature ties with positive probability, and one observes $k \leq n$ distinct observations, $(X_1^*, \ldots, X_k^*)$, with frequencies $(n_1, \ldots, n_k)$ such that $\sum_{j=1}^k n_j = n$.

As shown in James et al. (2006) the only conjugate hNRMI is the Dirichlet process. However, a sort of conditional conjugacy still holds for all NRMIs in the sense that, conditionally on a suitable latent variable, the posterior distribution is still an NRMI with updated Lévy intensity and fixed points of discontinuity in correspondence of the observations X_j^*'s. This posterior characterization in terms of CRMs, proven in James et al. (2009), forms the basis for obtaining the posterior stick-breaking characterization of an hNRMI. Before stating the result, let us introduce a latent random variable U_n whose density function, conditionally on $X^{(n)}$, is given by

$$
f_{U_n}(u) \propto u^{n-1} \mathrm{e}^{-\psi(u)} \prod_{j=1}^k \kappa_{n_j}(u) \tag{20}
$$

where $\psi(u)$ is the Laplace exponent of a homogeneous CRM $\tilde{\mu}$, namely $\psi(u) = a \int_{\mathbb{R}^+}(1-\mathrm{e}^{-uv})\rho(\mathrm{d}v)$ and $\kappa_{n_i}(u) = \int_0^\infty v^{n_i}\mathrm{e}^{-uv}\rho(\mathrm{d}v)$, the n_ith moment of the exponentially tilted jump part of the Lévy intensity of $\tilde{\mu}$.

Proposition 4. *Let $\tilde{p}$ be an hNRMI with Lévy intensity (7). Then, conditionally on $X^{(n)}$ and U_n, $\tilde{p}$ coincides in distribution with the random probability measure*

$$\varphi_{0,u}\,\tilde{p}_u + \sum_{j=1}^k \varphi_{j,u}\,\delta_{X_j^*} \tag{21}$$

where $\tilde{p}_u$ is an hNRMI admitting stick-breaking representation (2) with dependent weights $(V_{i,u})_{i\geq 1}$. Specifically, $V_{1,u}$ has density function given by

$$f_{V_{1,u}}(v_{1,u}) = a\,v_{1,u}\,\mathrm{e}^{\psi(u)} \int_0^\infty t^i\,\mathrm{e}^{-ut}\,\rho(tv_{1,u})\,f_T(tw_{2,u})\,\mathrm{d}t \tag{22}$$

and, for any $i \geq 2$, the conditional density of $V_{i,u}$ given $V_{1,u},\ldots,V_{i-1,u}$ is

$$f_{V_{i,u}}(v_{i,u}|v_{1,u}, v_{2,u}, \ldots, v_{i-1,u})$$

$$= a\,v_{i,u}\,w_{i,u}\,\frac{\int_0^{+\infty} t^i \mathrm{e}^{-ut}\left(\prod_{j=1}^i \rho(tv_{j,u}w_{j,u})\right) f_T(tw_{i+1,u})\mathrm{d}t}{\int_0^{+\infty} t^{i-1}\mathrm{e}^{-ut}\left(\prod_{j=1}^{i-1}\rho(tv_{j,u}w_{j,u})\right) f_T(tw_{i,u})\mathrm{d}t}, \tag{23}$$

where $w_{i,u} = \prod_{j=1}^{i-1}(1 - v_{j,u})$, for $i = 2, 3, \ldots$, $w_{1,u} = 1$ and f_T denotes the density function of the (prior) total mass $\tilde{\mu}(\mathbb{X})$. Furthermore,

$$\varphi_{0,u} = \frac{T_u}{T_u + \sum_{i=1}^k J_{i,u}} \qquad \varphi_{j,u} = \frac{J_{j,u}}{T_u + \sum_{i=1}^k J_{i,u}} \quad j \geq 2 \tag{24}$$

where the random variables T_u and $J_{i,u}$, for $i = 1,\ldots,k$, are independent and admit densities of the form $f_{T_u}(t) = \mathrm{e}^{\psi(u)-ut}f_T(t)$ and $f_{J_{i,u}}(y) \propto y^{n_i}\mathrm{e}^{-uy}\rho(y)$, for $i = 1,\ldots,k$, respectively.

We now focus again on the two special cases considered before, namely the generalized Dirichlet and the normalized generalized gamma processes.

4.1 Generalized Dirichlet process

For the generalized Dirichlet process the posterior stick-breaking representation can be deduced from (21) with $\tilde{p}_u$ of the form (2) and the dependent weights $(V_{i,u})_{i\geq 1}$ turn out to be such that $V_{1,u}$ has density function

$$f_{V_{1,u}}(v_{1,u}) = a\,\mathrm{e}^{\psi(u)}\,(\gamma!)^a\,w_{2,u}^{-1}\sum_{i_1=0}^{\gamma-1} z_{1,u}^{-a}\left\{(-1)^{\gamma-1}\,(1-z_{1,u})_{\gamma-1}\right\}^{-a} \tag{25}$$

where we recall that $\psi(u) = a \sum_{j=1}^{\gamma} \log(1 + u/j)$, $z_{1,u} = (\bar{v}_{1,u} + \gamma w_{2,u})/w_{2,u}$ and $\bar{v}_{1,u} = v_{1,u}(1 + i_1)$. On the other hand, for any $i \geq 2$, the conditional density of $V_{i,u}$, given $V_{1,u}, \ldots, V_{i-1,u}$, is

$$f_{V_{i,u}}(v_{i,u}|v_{1,u}, v_{2,u}, \ldots, v_{i-1,u})$$

$$= \frac{a\, v_{i,u} w_{i,u} (1 - v_{i,u})^{-1}}{\sum_{C_{i-1}} z_{i-1,u}^{-a} \left\{ (1 - z_{i-1,u})_{\gamma-1} \right\}^{-a}} \sum_{C_i} z_{i,u}^{-a} \left\{ (1 - z_{i,u})_{\gamma-1} \right\}^{-a} \tag{26}$$

where, as before, $\bar{v}_{i,u} = \sum_{\ell=1}^{i} v_{\ell,u} w_{\ell,u}(1 + j_\ell)$ and $z_{i,u} = (u + \bar{v}_{i,u} + \gamma w_{i+1,u})/w_{i+1,u}$ for any $i \geq 2$. Details for the determination of the two densities in (25) and (26) are displayed in Section 7.5. The description of the posterior is then completed by providing the distribution of the other quantities involved in Proposition 4: the independent jumps $J_{i,u}$ at the distinct values of the observations have density

$$f_{Ji,u}(s) = \frac{s^{n_i-1} \sum_{j=1}^{\gamma} e^{-(j+u)s}}{\Gamma(n_i)\, [\zeta(n_j, u+1) - \zeta(n_j, u+\gamma+1)]}$$

where $\zeta(s, q) = \sum_{l=0}^{\infty} (l+q)^{-s}$ is the generalized Riemann Zeta (or Hurwitz) function, and the probability distribution of the total mass T_u is characterized by a density $f_{T_u}(t) = e^{\psi(u)}\, e^{-ut}\, f_T(t)$ where f_T is the density of the prior total mass in (15). Finally, the latent random variable U_n has density function

$$f_{U_n}(u) \propto u^{n-1} \left\{ \prod_{i=1}^{\gamma} (i+u)^{-a} \right\} \prod_{j=1}^{k} [\zeta(n_j, u+1) - \zeta(n_j, u+\gamma+1)].$$

4.2   The normalized generalized gamma process

When $\tilde{p}$ is a normalized generalized gamma process with Lévy intensity as in (11), from Proposition 4, one can work out a particularly simple stick-breaking representation of $\tilde{p}_u$. Indeed, the density functions of $V_{1,u}$ and $V_{i,u}|V_{1,u}, \ldots, V_{i-1,u}$ are exactly as in (18) and (19), respectively, with β replaced by $\beta_u = a(u+\tau)^\sigma/\sigma$. Therefore, $\tilde{p}_u$ coincides with the prior but for an updated parameter which depends on the latent variable U_n; see Section 7.6 for details. For the special case where $\sigma = 1/2$, this is in accordance with the posterior representation obtained for the normalized inverse-Gaussian process in Favaro et al. (2012).

The description of the posterior in Proposition 4 is completed by noting that the independent jumps $J_{i,u}$ at the distinct observations are gamma distributed, with scale and shape parameters coinciding with $(u+\tau)$ and $n_i - \sigma$, respectively. Moreover, T_u has distribution $f_{T_u}(t) = e^{\frac{a\tau^\sigma}{\sigma}}\, e^{-(u+\tau)t}\, f_\sigma(t)$. Finally, the latent random variable U_n admits density function

$$f_{U_n}(u) \propto \frac{u^{n-1}}{(u+\tau)^{n-k\sigma}}\, e^{-\frac{a}{\sigma}(u+\tau)^\sigma}.$$

5   Simulation algorithm

In this section, we illustrate how Proposition 1, and in turn Propositions 2 and 3, can be exploited in order to devise a slice sampling algorithm for mixture models built on hNRMIs. We start by considering a random density function $f_{\tilde{p}}$ defined as

$$f_{\tilde{p}}(x) = \int_{\mathbb{X}} k(x|y)\mathrm{d}\tilde{p}(y),$$

where $k(x|y)$ is a continuous kernel density function and $\tilde{p}$ is an hNRMI. This is equivalent to considering the following hierarchical model

$$
\begin{aligned}
X_1,\ldots,X_n \mid Y_1,\ldots,Y_n &\overset{\text{ind}}{\sim} k(X_i|Y_i), \\
Y_i \mid \tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \qquad\qquad \text{for } i=1,\ldots,n, \\
\tilde{p} &\sim \text{hNRMI}.
\end{aligned}
\qquad (27)
$$

Moreover, by exploiting (1), we can write

$$f_{\tilde{p}}(x) = \sum_{j \geq 1} \tilde{p}_j k(x|Y_j). \qquad (28)$$

The infinite-dimensionality of (28) can be tackled by following the slice sampling approach introduced in Walker (2007). Specifically, if $X \sim f_{\tilde{p}}$, we introduce a random variable U such that the joint distribution of (X, U) is given by

$$f_{\tilde{p},U}(x,u) = \sum_{j \geq 1} \mathbb{1}_{\{u < \tilde{p}_j\}} k(x|Y_j).$$

We observe that U is assumed uniformly distributed and that, conditionally on U, the distribution of X reduces to a finite mixture

$$f_{\tilde{p}|U}(x|u) = \frac{1}{N_u} \sum_{j \in A_u} k(x|Y_j),$$

where $A_u = \{j : \tilde{p}_j > U\}$ has finite size and $N_u = \sum_{j \in A_u} \tilde{p}_j$. At this point, an additional latent variable D can be introduced, referring to the component of the mixture from which X is taken. This leads to the joint density function

$$f_{\tilde{p},U,D}(x,u,d) = \mathbb{1}_{\{u < \tilde{p}_d\}} k(x|Y_d). \qquad (29)$$

For improving the efficiency of the algorithm, here we implement the so-called slice-efficient sampler by Kalli et al. (2011); see also Papaspiliopoulos (2008). Specifically, we replace (29) with

$$f_{\tilde{p},U,D}(x,u,d) = \mathrm{e}^{\xi d}\mathbb{1}_{\{u < \mathrm{e}^{-\xi d}\}}\tilde{p}_d k(x|Y_d),$$

for some $\xi > 0$. The complete likelihood then becomes

$$\prod_{i=1}^{n} \mathrm{e}^{\xi d_i}\mathbb{1}_{\{u_i < \mathrm{e}^{-\xi d_i}\}}\tilde{p}_{d_i} k(x_i|Y_{d_i}).$$

The variables that need to be updated at each iteration of the sampler are $\{(Y_j, \tilde{p}_j), j = 1, 2, \ldots; (D_i, U_i), i = 1, \ldots, n\}$. It can be easily checked that if, for every $i \geq 1$, conditionally on $U_i = u_i$, we set $N_i = \lfloor -(1/\xi) \log u_i \rfloor$, with $\lfloor x \rfloor$ being the integer part of x, then it is enough to sample $\tilde{p}_j$ and Y_j for $j = 1, \ldots, N = \max N_i$. As for the conditional sampling of the Y_j's, the D_i's and the U_i's, we refer to Walker (2007) and Kalli et al. (2011). The focus of this section is on the sampling of the random probabilities $\tilde{p}_j$ for the class of hNRMIs.

Equations (13) and (14) represent the starting point for deriving the joint distribution of $V^{(N)} = (V_1, \ldots, V_N)$, for every $N \geq 1$, conditionally on $D^{(n)} = (D_1, \ldots, D_n)$. It is easy to show that

$$f_{V^{(N)}|D^{(n)}}(v_1, \ldots, v_N \,|\, d_1, \ldots, d_n) \propto$$
$$\prod_{i=1}^{N} v_i^{n_i}(1 - v_i)^{m_i} \int_0^\infty t^N \prod_{j=1}^{N} \rho\left(tv_j \prod_{i=1}^{j-1}(1 - v_i)\right) f_T\left(t \prod_{i=1}^{N}(1 - v_i)\right) \mathrm{d}t, \quad (30)$$

where $n_i = \sum_{j=1}^n \mathbb{1}_{\{d_j=i\}}$ and $m_i = \sum_{j=1}^n \mathbb{1}_{\{d_j>i\}}$. Moreover, the joint distribution of $V^{(N)}$ and the total mass T, conditionally on $D^{(n)}$, has density function

$$f_{T,V^{(N)}|D^{(n)}}(t, v_1, \ldots, v_N \,|\, d_1, \ldots, d_n) \propto$$
$$t^N \prod_{j=1}^{N} v_j^{n_j}(1 - v_j)^{m_j} \rho\left(tv_j \prod_{i=1}^{j-1}(1 - v_i)\right) f_T\left(t \prod_{i=1}^{N}(1 - v_i)\right). \quad (31)$$

Equation (31) easily leads to the full conditional distributions for T and V_i, for every $i = 1, \ldots, N$, that are needed for the sampler. Specifically, we have

$$f_{T|\text{rest}}(t \,|\, \text{rest}) \propto t^N \prod_{j=1}^{N} \rho\left(tv_j \prod_{i=1}^{j-1}(1 - v_i)\right) f_T\left(t \prod_{i=1}^{N}(1 - v_i)\right) \quad (32)$$

and

$$f_{V_i|\text{rest}}(v_i \,|\, \text{rest}) \propto v_i^{n_i}(1 - v_i)^{m_i} f_T\left(t \prod_{l=1}^{N}(1 - v_l)\right) \prod_{j=1}^{N} \rho\left(tv_j \prod_{l=1}^{j-1}(1 - v_l)\right). \quad (33)$$

Remark. Within conditional samplers, a natural competitor for the slice sampler in mixture models is the algorithm implied by the series representation of increasing Lévy processes in Ferguson and Klass (1972). Though the Ferguson & Klass sampler is attractive since it generates the jumps in decreasing order, on the other hand, it requires truncating the series at a certain threshold and is therefore only approximate. On the contrary, the slice sampler is exact, but it does not generate the probability weights in decreasing order since it is based on the stick-breaking representation. When the weights' distribution is heavy tailed, this might imply the necessity of simulating a very

large number of weights. If the stick-breaking representation is not simple or the overall model is complex, it turns out to be cumbersome to simulate the required number of jumps and hence one ends up truncating. And, among approximate algorithms, the Ferguson & Klass is preferable, given the weights arranged in decreasing order and the possibility to control the error of approximation, for instance, by moment-matching techniques as proposed in Arbel and Prünster (2015). Hence, there is no algorithm which is preferable in general, but rather one has to make an informed choice based on the model (and, in particular, the effort needed to invert the Lévy measure in the Ferguson & Klass algorithm and to simulate the stick-breaking weights within the slice sampler) and the application at hand.

Next, we investigate how the full conditional distributions (32) and (33) specialize when we consider a generalized Dirichlet process or a normalized generalized gamma process as mixing measures $\tilde{p}$. To simplify the notation, we introduce the random variables $W := \prod_{i=1}^{N}(1 - V_i)$ and $W^{(i)} := \prod_{j \neq i}(1 - V_j)$.

5.1 Generalized Dirichlet process

For the generalized Dirichlet process, the joint conditional distribution (31) reduces to

$$f_{T,V^{(N)}|D^{(n)}}(t, v_1, \ldots, v_N \,|\, d_1, \ldots, d_n) \propto t^{\gamma a - 1} \prod_{j=1}^{N} v_j^{n_j - 1}(1 - v_j)^{m_j + j - N}$$

$$\times \Phi_2^{(\gamma - 1)}(a^{(\gamma - 1)}; \gamma a; tw, 2tw, \ldots, (\gamma - 1)tw) \sum_{l=1}^{\gamma} \mathrm{e}^{-t(\gamma + l(1 - w))}.$$

Accordingly, we have

$$f_{T|\mathrm{rest}}(t \,|\, \mathrm{rest}) \propto t^{\gamma a - 1} \Phi_2^{(\gamma - 1)}(a^{(\gamma - 1)}; \gamma a; tw, 2tw, \ldots, (\gamma - 1)tw) \sum_{l=1}^{\gamma} \mathrm{e}^{-t(\gamma + l(1 - w))} \quad (34)$$

and, for every $i = 1, \ldots, N$,

$$f_{V_i|\mathrm{rest}}(v_i \,|\, \mathrm{rest}) \propto$$

$$v_i^{n_i - 1}(1 - v_i)^{m_i + i - N} \Phi_2^{(\gamma - 1)}(a^{(\gamma - 1)}; \gamma a; tw, 2tw, \ldots, (\gamma - 1)tw) \sum_{l=1}^{\gamma} \mathrm{e}^{-t(\gamma + l(1 - w))},$$

$$(35)$$

where we recall that w depends on v_i. Devising a random variate generator for the full conditional distribution (34) is not straightforward and goes beyond the goal of the paper. Nonetheless, it is important to observe that recent developments on the approximate evaluation of the confluent form of the fourth Lauricella hypergeometric function $\Phi_2^{(n)}$ (see Butler and Wood (2015) and reference therein to the code available online) suggest that a suitable Metropolis–Hastings algorithm can be implemented. Sampling from the full conditional in (35) is less cumbersome since, due to its compact support, one can either rely on a suitable discretized approximation or on an acceptance/rejection algorithm.

5.2   The normalized generalized gamma process

For the normalized generalized gamma process, the joint conditional distribution (31) reduces to

$$f_{T,V^{(N)}|D^{(n)}}(t, v_1, \ldots, v_N \,|\, d_1, \ldots, d_n) \propto t^{N(1-\sigma)}\mathrm{e}^{-t\tau} \prod_{j=1}^{N} v_j^{n_j-\sigma}(1-v_j)^{m_j-(N-j)\sigma} f_\sigma(tw).$$

Accordingly, we have

$$f_{T|\mathrm{rest}}(t \,|\, \mathrm{rest}) \propto t^{N(1-\sigma)}\mathrm{e}^{-t\tau} f_\sigma(tw) \tag{36}$$

and, for every $i = 1, \ldots, N$,

$$f_{V_i|\mathrm{rest}}(v_i \,|\, \mathrm{rest}) \propto v_i^{n_i-\sigma}(1-v_i)^{m_i-(N-i)\sigma} f_\sigma((1-v_i)tw^{(i)}). \tag{37}$$

If a random variable T' is distributed according to the density in (36), then the random variable $S = T'W$ has density function proportional to

$$g(s) \propto s^{N(1-\sigma)}\mathrm{e}^{-s\tau/w} f_\sigma(s), \tag{38}$$

that corresponds to the so-called gamma tilted stable distribution. In the following illustration, we rely on an efficient exact random variate generator for (38), recently developed in Favaro et al. (2015). As for the full conditionals (37) for the V_i's, we resort to a discrete approximation of (37) as it turns out to be a good compromise in terms of accuracy and ease of implementation. Specifically, we consider a fine partition of the domain $[0, 1]$ made of intervals of the same size (100 intervals in the illustration of Section 5.3) and approximate (37) with a uniform distribution on each interval.

Finally, notice that a different approach to slice sampling mixture models based on the normalized generalized gamma process can be found in Favaro and Walker (2013), where the Kanter's representation for f_σ (Kanter, 1975) is exploited.

5.3   Illustration on acidity data

In order to appreciate the practical impact of the distributional results we have been describing so far, we display an illustrative application based on the implementation of the slice-efficient sampler, for the normalized generalized gamma case, discussed in Sections 5 and 5.2. We investigate the performance of the algorithm by analyzing a dataset consisting of measurements of an acid neutralizing capacity (ANC) index in a sample of 155 lakes in North-Central Wisconsin, USA. A low value of ANC can lead to a loss of biological resources. The identification of clusters of lakes is important for the determination of lake characteristics which can be used to predict higher acidification. These data were studied by several authors and were considered on a log-scale as we do. The estimates for the number of clusters K_{155}, under different approaches, typically range between 2 and 5; see, for example, Richardson and Green (1997) where a fairly equal support for 3–5 components is obtained. For illustrative purposes, we focus on

two values of σ, namely $\sigma = 1/2$ (corresponding to the normalized inverse-Gaussian process) and $\sigma = 1/4$. The comparison between these two models is interesting as it gives some insight on the feasibility, for $\sigma \neq 1/2$, of the computational approach that we propose. The case corresponding to $\sigma = 1/2$ is used as a basis for comparison as it stands out for its ease of implementation: the stable density involved in (36) and (37) can be written in closed form and the resulting full conditional distributions can be evaluated with negligible computational burden.

For the sake of simplicity, and without loss of generality, we set $a = 1$. In order to compare the normalized generalized gamma models with different specifications of σ, we fix the other parameter τ in way that the induced prior distributions on K_{155} have the same mode; see Lijoi et al. (2007) for more detail on this centering strategy, where the distribution of K_{155} for the normalized generalized gamma priors is also provided. Moreover, in order to compare the performance of the normalized generalized gamma models as σ varies, we opt for a misspecification of the prior centering of K_{155}, namely in 15 which is significantly larger than the number of components found in previous studies of the same dataset.

Since for a given $\sigma \in (0, 1)$ there is an interval (τ', τ'') satisfying this requirement, then we pick $\tau^* \in (\tau', \tau'')$ such that $P_{(\tau^*, \sigma)}(K_{155} = 15) = \max_{\tau \in (\tau', \tau'')} P_{(\tau, \sigma)}(K_{155} = 15)$. For example, as shown in the left panel of Figure 1, when $\sigma = 1/2$, for every $\tau \in [0.093, 0.126]$ the prior mode for K_{155} is in 15. In this domain, we pick the value $\tau^* = 0.110$ since it maximizes $P_{(\tau, 1/2)}(K_{155} = 15)$. Similarly, when $\sigma = 1/4$ we obtain $\tau^* = 6.800$. From Figure 1 it is apparent that, for a given σ, the mode shifts to the right as τ increases. This is in accordance with the interpretation given in Lijoi et al. (2007) of τ as a location parameter. Moreover, σ controls the variability, but, at the same time, larger values of σ imply also a shift to the right of the distribution of K_{155}. Consequently, for large σ it is not always possible to center the distribution on the desired value if too small. For instance, in the example we consider with $n = 155$, the largest value of σ compatible with a mode in 15 is 0.606. Larger values of σ lead to a
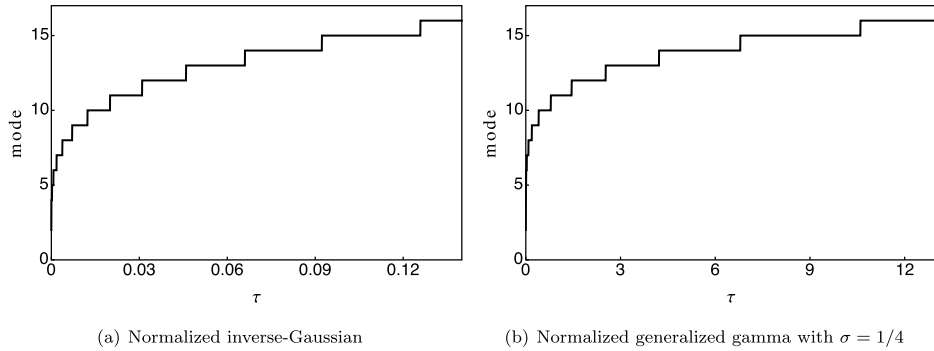

(a) Normalized inverse-Gaussian   (b) Normalized generalized gamma with $\sigma = 1/4$

Figure 1: Analysis of the role of τ in determining the mode of the prior distribution of K_{155}.

mode in values larger than 15 even if τ is set equal to its lowest possible value, namely 0.

The illustration is integrated by comparing the performance of these two models with the simpler model based on the Dirichlet process, case corresponding to $\sigma \to 0$, for which we have set the total mass parameter $a = 3.912$ in order to achieve the same prior guess for K_{155}.

We complete model (27) with a quite standard specification, similar to the one adopted by Walker (2007). Namely, we assume that $k(\cdot|\cdot)$ is a Gaussian kernel and that, for every $i \geq 1$, Y_i is the parameter vector (M_i, S_i^2) whose components are mean and variance of the Gaussian kernel. The base measure P_0 is assumed normal-inverse-gamma, that is, a priori, the M_i's and the S_i^2's are independent i.i.d. sequences such that each M_i is normally distributed with mean m and variance $1/s$, whereas each S_i^2 is such that $1/S_i^2$ has gamma distribution with both shape and scale parameters equal to ϵ. Finally, we set $m = \bar{X}$, $s = 0.05$, $\epsilon = 0.5$ and $\xi = 0.5$.

We ran the MCMC algorithm for 100,000 iterations after 20,000 of burn-in. After thinning we have stored a sample of 2,000 iterations that, in turn, were used for estimating the sampling density function. For each model, posterior estimates, together with samples from the predictive density function, are plotted in Figure 2. Given one can always fit a mixture with more components than needed, it is not surprising that there are no significant differences between posterior estimates obtained by means of the three models.

In terms of number of clusters, the considered models show different features. In Table 1, the posterior distributions for the number of components K_{155} are reported. While the two models based on the normalized inverse-Gaussian and the normalized generalized gamma with $\sigma = 1/4$, lead to posterior distributions for K_{155} with mode in 4, the model based on the Dirichlet process has posterior mode in 5. The difference between the three models is more apparent if we look at the posterior mean $\mathbb{E}[K_{155}|X^{(n)}]$ that is equal to 3.817, 4.862 and 5.511 for the normalized inverse-Gaussian, the normalized generalized gamma with $\sigma = 1/4$, and Dirichlet process, respectively. Moreover, for the same three models, the posterior variance is equal to 0.849, 1.673 and 2.526, and the posterior probability that $K_{155} \leq 5$ is equal to 0.955, 0.717 and 0.534, respectively. These results clearly show that a larger value of σ makes the model more robust with respect to misspecifications of the a priori number of components. In contrast, if the a priori centering is close to the "correct" value, the posterior distributions of K_{155} essentially coincide. This is in line with the findings of Lijoi et al. (2007), where the role of σ is studied in detail.

The considered diagnostic tools suggest, for the three models, convergence of the chain. Moreover, after thinning, the autocorrelation of the chain is smoothed out significantly. It is worth noting that while the choice of σ does not seem to influence the mixing behavior of the corresponding MCMC algorithms, it has an impact on the computational time required to run the algorithm. Among the three models, the Dirichlet process leads to the fastest algorithm due to the simple updating mechanism for the stick-breaking weights that characterizes it. The algorithm based on the normalized generalized gamma process with $\sigma = 1/4$ is significantly slower than the one built on the normalized inverse-Gaussian process. This is due to the updating of the V_i's in the
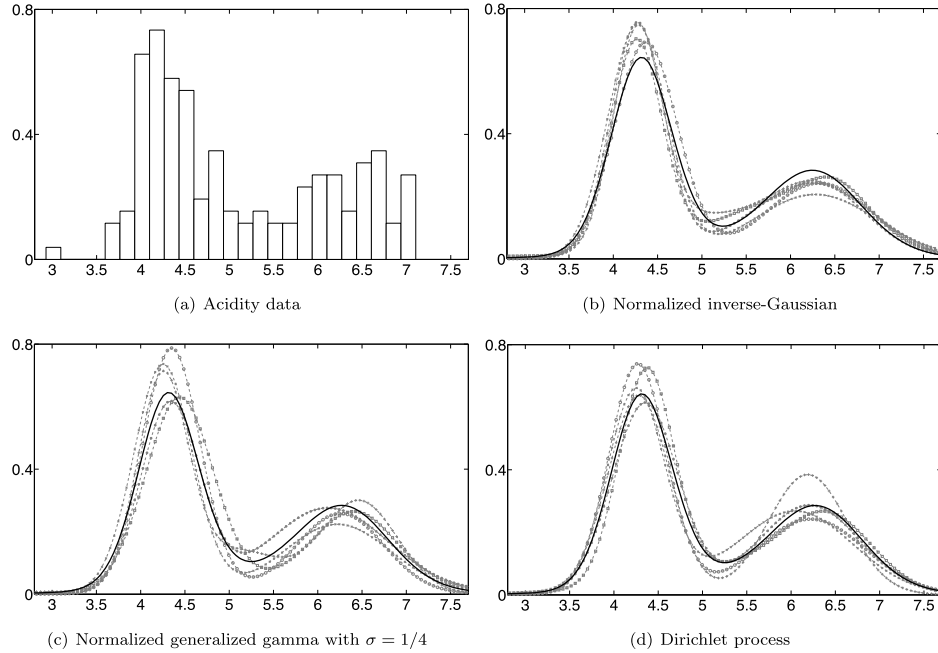
Figure 2: (a) Histogram of the acidity dataset; (b)–(d) Dashed lines are samples of predictive densities whereas continuous lines are the posterior expected values of $f_{\tilde{p}}$ that are used as posterior density estimates, obtained through the MCMC algorithm under three different prior specifications.

normalized generalized gamma process case by means of the approach described in Section 5.2, which requires the evaluation, over a grid of values, of the density of a positive σ-stable random variable at every iteration. This task is generally computationally quite demanding when $\sigma \neq 1/2$ and leads to a slower algorithm.

6 Concluding remarks

We have provided some new distributional results for hNRMIs that are relevant for achieving a stick-breaking representation. Indeed, one can rely on Proposition 1 to provide the densities of the stick-breaking weights V_i, once ρ and f_T have been specified. This has been illustrated for two examples, namely the generalized Dirichlet process and the normalized generalized gamma process. Despite the expressions we obtain might look somehow involved in these two cases, they still shed some light on their distributional structure, which is the main goal of the paper. Moreover, they also allow the formulation of simulation algorithms for hierarchical mixture models as discussed in Section 5. As for the actual implementation of the suggested MCMC sampler we have confined ourselves

	NIG $\tau = 0.110$	NGG ($\sigma = 1/4, \tau = 6.800$)	Dirichlet $a = 3.912$
mode	4	4	5
mean	3.817	4.862	5.511
$K_{155} \leq 2$	0.032	0.007	0.007
$K_{155} = 3$	0.374	0.128	0.077
$K_{155} = 4$	0.392	0.292	0.196
$K_{155} = 5$	0.157	0.291	0.254
$K_{155} = 6$	0.037	0.177	0.219
$K_{155} = 7$	0.006	0.074	0.137
$K_{155} = 8$	0.001	0.024	0.069
$K_{155} = 9$	< 0.001	0.006	0.028
$K_{155} \geq 10$	< 0.001	0.001	0.013

Table 1: Posterior mode, posterior mean and posterior probabilities for the number of components K_{155} for the normalized inverse-Gaussian (NIG), the normalized generalized gamma (NGG) and the Dirichlet process.

to the normalized generalized gamma case. In principle, the same can be done for any hNRMI, the only difficulties lying in simulating from the full conditionals of T and V_i.

Finally, it should be noted that it would be interesting to have distributional representations of the stick-breaking weights in terms of some random variables that can be easily simulated for hNRMIs in general. This would allow to truncate the series at a fixed threshold and obtain approximate realizations of $\tilde{p}$ in a straightforward way. For the normalized inverse-Gaussian case, i.e., a normalized generalized gamma process with $\sigma = 1/2$ such a representation has been achieved in Favaro et al. (2012) and crucially depends on the availability of a closed form expression of the density function f_T. Hence, the route taken in Favaro et al. (2012) cannot be extended to the normalized generalized gamma and, a fortiori, to the general hNRMI case. A new strategy has to be devised for the derivation of simple random variate representations of the V_i's.

7 Proofs

7.1 Proof of Proposition 1

The proof follows by first noting that hNRMIs are invariant under size-biased permutation and then rephrasing the result displayed in Theorem 2.1 of Perman et al. (1992) for subordinators, namely when $\mathbb{X} = \mathbb{R}$ and $\alpha(\mathrm{d}x) = \mathrm{d}x$, in the context of CRMs with intensity (7). Indeed, a subordinator on $[0,1]$, as the one in Perman et al. (1992), has jumps independent from locations that, in turn, are distributed according to a uniform distribution on $[0,1]$. Given this independence, one can simply take the sequence of jumps, indexed, say, by the integers, and attach labels generated independently by an arbitrary (non-atomic) probability measure on a Polish space. This then corresponds to a CRM and the stated result follows.   □

7.2   Proof of Proposition 2

By virtue of Proposition 1, one has

$$
\begin{aligned}
f_{V_1}(v_1) &= \frac{a\,(\gamma!)^a}{\Gamma(\gamma a)} \int_0^{+\infty} \sum_{j_1=0}^{\gamma-1} \mathrm{e}^{-(j_1+1)tv_1}\mathrm{e}^{-\gamma t(1-v_1)}(t(1-v_1))^{\gamma a-1} \\
&\qquad \times \Phi_2^{(\gamma-1)}(a^{(\gamma-1)};\gamma a;t(1-v_1),\ldots,(\gamma-1)t(1-v_1))\mathrm{d}t \\
&= \frac{a\,(\gamma!)^a}{\Gamma(\gamma a)}(1-v_1)^{\gamma a-1}\sum_{j_1=0}^{\gamma-1}\int_0^{+\infty}\mathrm{e}^{-t(v_1(j_1+1)+\gamma(1-v_1))}t^{\gamma a-1} \\
&\qquad \times \Phi_2^{(\gamma-1)}(a^{(\gamma-1)};\gamma a;t(1-v_1),\ldots,(\gamma-1)t(1-v_1))\,\mathrm{d}t.
\end{aligned}
$$

The change of variable $z = t(v_1(j_1+1)+\gamma(1-v_1))$ and (2.4.10) in Exton (1976) lead to

$$
\begin{aligned}
f_{V_1}(v_1) &= \frac{a\,(\gamma!)^a}{\Gamma(\gamma a)} w_2^{\gamma a-1} \sum_{j_1=0}^{\gamma-1} \frac{1}{(\bar{v}_1+\gamma w_2)^{\gamma a}} \int_0^{+\infty} \mathrm{e}^{-z} z^{\gamma a-1} \\
&\qquad \times \Phi_2^{(\gamma-1)}\left(a^{(\gamma-1)};\gamma a;\frac{zw_2}{\bar{v}_1+\gamma w_2},\ldots,\frac{z(\gamma-1)w_2}{\bar{v}_1+\gamma w_2}\right)\,\mathrm{d}z \\
&= a\,(\gamma!)^a\,w_2^{\gamma a-1}\sum_{i_1=0}^{\gamma-1}\frac{1}{(\bar{v}_1+\gamma w_2)^{\gamma a}} \\
&\qquad \times F_D^{(\gamma-1)}\left(\gamma a;a^{(\gamma-1)};\gamma a;\frac{w_2}{\bar{v}_1+\gamma w_2},\ldots,\frac{(\gamma-1)w_2}{\bar{v}_1+\gamma w_2}\right).
\end{aligned}
$$

Here $F_D^{(n)}$ is the Lauricella function of the fourth type (see Section 2 in Exton (1976)) and, since $F_D^{(n)}(c;b_1,\ldots,b_n;c;x_1,\ldots,x_n) = \prod_{i=1}^n(1-x_i)^{-b_i}$, the following simplification occurs:

$$
\begin{aligned}
f_{V_1}(v_1) &= a\,(\gamma!)^a\,w_2^{\gamma a-1}\sum_{j_1=0}^{\gamma-1}\frac{1}{(\bar{v}_1+\gamma w_2)^{\gamma a}}\prod_{i=1}^{\gamma-1}\left(1-\frac{iw_2}{\bar{v}_1+\gamma w_2}\right)^{-a} \\
&= a\,(\gamma!)^a\,w_2^{\gamma a-1}\sum_{j_1=0}^{\gamma-1}(\bar{v}_1+\gamma w_2)^{-a}\prod_{i=1}^{\gamma-1}(\bar{v}_1+\gamma w_2-iw_2)^{-a},
\end{aligned}
$$

and some simple algebra yields (16). As for the conditional distribution of V_1, given $V_1,\ldots,V_{i-1}$, we shall resort to (14). In particular, similarly to the determination of the density of V_1, the integral in the numerator of (14) is

$$
\begin{aligned}
\int_0^\infty t^i f_T(tw_{i+1})\prod_{j=1}^i \rho(tv_jw_j)\,\mathrm{d}t &= (\gamma!)^a\,w_{i+1}^{\gamma a-1}\sum_{C_i}\frac{1}{(\bar{v}_i+\gamma w_{i+1})^{\gamma a}} \\
&\qquad \times F_D^{(\gamma-1)}\left(\gamma a;a^{(\gamma-1)};\gamma a;\frac{w_{i+1}}{\bar{v}_i+\gamma w_{i+1}},\ldots,\frac{(\gamma-1)w_{i+1}}{\bar{v}_i+\gamma w_{i+1}}\right)
\end{aligned}
$$

and the same simplification of $F_D^{(\gamma-1)}$ used before now applies so that

$$\int_0^\infty t^i f_T(tw_{i+1}) \prod_{j=1}^i \rho(tv_jw_j)\, dt$$

$$= (\gamma!)^a\, w_{i+1}^{a-1} \sum_{C_i} \frac{1}{(\bar{v}_i + \gamma w_{i+1})^a} \left\{ (-1)^{\gamma-1} \left(1 - \frac{\bar{v}_i + \gamma w_{i+1}}{w_{i+1}} \right)_{\gamma-1} \right\}^{-a}.$$

In a similar fashion, one works out the integral in the denominator of (14), and the result in (17) follows. □

7.3 Proof of Proposition 3

Note that in this case the density of the total mass T is not known in closed form. Anyhow, since the distribution of the generalized gamma CRM can be obtained as an exponential tilting of the distribution of the σ-stable CRM (Pitman, 2003), one then has

$$f_T(t) = e^{\frac{a\tau^\sigma}{\sigma}}\, e^{-\tau t}\, f_\sigma(t)$$

where f_σ is the density of a positive σ-stable random variable such that $\int_0^\infty e^{-\lambda s} f_\sigma(s)\, ds = \exp\{-a\lambda^\sigma/\sigma\}$. This implies that

$$f_{V_1}(v_1) = a \int_0^\infty tv_1 \frac{1}{\Gamma(1-\sigma)} t^{-1-\sigma} v_1^{-1-\sigma}\, e^{-\tau tv_1} f_T(t(1-v_1))\, dt$$

$$= \frac{a\, v_1^{-\sigma}\, e^{\frac{a\tau^\sigma}{\sigma}}}{\Gamma(1-\sigma)} \int_0^\infty t^{-\sigma}\, e^{-\tau tv_1}\, e^{-\tau t(1-v_1)}\, f_\sigma(t(1-v_1))\, dt$$

$$= \frac{a\, v_1^{-\sigma}\, e^{\frac{a\tau^\sigma}{\sigma}}}{\Gamma(1-\sigma)\Gamma(\sigma)} \int_0^\infty u^{\sigma-1} \int_0^\infty e^{-(u+\tau)t}\, f_\sigma(t(1-v_1))\, dt\, du$$

where we have used the fact that $t^{-\sigma}\Gamma(\sigma) = \int_0^\infty u^{\sigma-1}\, e^{-ut} du$. The inner integral above is nothing but the Laplace transform of f_σ, and then one has

$$f_{V_1}(v_1)$$

$$= \frac{a\, v_1^{-\sigma}\, e^{\frac{a\tau^\sigma}{\sigma}}}{\Gamma(1-\sigma)\Gamma(\sigma)} \frac{1}{1-v_1} \int_0^\infty u^{\sigma-1}\, e^{-\frac{a(u+\tau)^\sigma}{\sigma(1-v_1)^\sigma}}\, du$$

$$= \frac{a\, v_1^{-\sigma}\, e^{\frac{a\tau^\sigma}{\sigma}}}{\Gamma(1-\sigma)\Gamma(\sigma)} \frac{1}{1-v_1} \frac{\sigma^{\frac{1}{\sigma}-1}(1-v_1)}{a^{\frac{1}{\sigma}}} \int_{\frac{a\tau^\sigma}{\sigma(1-v_1)^\sigma}}^\infty s^{\frac{1}{\sigma}-1} \left\{ (1-v_1)\frac{s^{\frac{1}{\sigma}}\sigma^{\frac{1}{\sigma}}}{a^{\frac{1}{\sigma}}} - \tau \right\}^{\sigma-1}\, e^{-s}\, ds$$

$$= \frac{e^{\frac{a\tau^\sigma}{\sigma}}}{\Gamma(1-\sigma)\Gamma(\sigma)}\, v_1^{-\sigma}(1-v_1)^{\sigma-1} \sum_{j\geq 0} \frac{(1-\sigma)_j}{j!} \frac{\tau^j\, a^{\frac{j}{\sigma}}}{\sigma^{\frac{j}{\sigma}}(1-v_1)^j} \int_{\frac{a\tau^\sigma}{\sigma(1-v_1)^\sigma}}^\infty s^{-\frac{j}{\sigma}}\, e^{-s}\, ds$$

which coincides with (18), after setting $\beta := a\tau^\sigma/\sigma$. As for the determination of the density of V_i, conditional on $V_1, \ldots, V_{i-1}$, one can proceed in a similar fashion as for f_{V_1}. Indeed, note that the integral appearing in the numerator of (14) now coincides with

$$\frac{\mathrm{e}^\beta \, (\prod_{l=1}^i v_l w_l)^{-1-\sigma}}{(\Gamma(1-\sigma))^i} \int_0^\infty t^{-i\sigma} \, \mathrm{e}^{-\tau t \sum_{l=1}^i v_l w_l - \tau t w_{i+1}} \, f_\sigma(t w_{i+1}) \, \mathrm{d}t$$

$$= \frac{\mathrm{e}^\beta \, (\prod_{l=1}^i v_l w_l)^{-1-\sigma}}{(\Gamma(1-\sigma))^i} \int_0^\infty t^{-i\sigma} \, \mathrm{e}^{-\tau t} \, f_\sigma(t w_{i+1}) \, \mathrm{d}t$$

$$= \frac{\mathrm{e}^\beta \, (\prod_{l=1}^i v_l w_l)^{-1-\sigma}}{(\Gamma(1-\sigma))^i \, \Gamma(i\sigma)} \int_0^\infty y^{i\sigma-1} \int_0^\infty \mathrm{e}^{-(y+\tau)t} \, f_\sigma(t w_{i+1}) \, \mathrm{d}t \, \mathrm{d}y$$

$$= \frac{\mathrm{e}^\beta}{(\Gamma(1-\sigma))^i \, \Gamma(i\sigma)} \frac{(\prod_{l=1}^i v_l w_l)^{-1-\sigma}}{w_{i+1}} \int_0^\infty y^{i\sigma-1} \, \mathrm{e}^{-\frac{a}{\sigma} \frac{(y+\tau)^\sigma}{w_{i+1}^\sigma}} \, \mathrm{d}y$$

where the first equality above follows from the identity $\sum_{l=1}^i v_l w_l + w_{i+1} = 1$. A similar change of variable and power series expansion as the one used for V_1 leads then to determining the numerator as

$$\frac{\mathrm{e}^\beta \, \sigma^{i-1}}{(\Gamma(1-\sigma))^i \, \Gamma(i\sigma) \, a^i} \frac{w_{i+1}^{i\sigma-1}}{(\prod_{l=1}^i v_l w_l)^{1+\sigma}} \sum_{j\geq 0} \frac{(1-i\sigma)_j}{j!} \frac{\beta^{\frac{j}{\sigma}}}{w_{i+1}^\sigma} \, \Gamma\left(i - \frac{j}{\sigma}; \, \frac{\beta}{w_{i+1}^\sigma}\right).$$

The expression for the integral in the denominator is similar, with i replaced by $i-1$, and the result in (19) follows. □

7.4   Proof of Proposition 4

The proof follows immediately by adapting the posterior characterization of NRMIs given in Theorems 1 and 2 of James et al. (2009) and combining it with the stick-breaking representation given in Proposition 1. □

7.5   Details for the determination of (25) and (26)

In view of (22), one has

$$f_{V_{1,u}}(v_{1,u}) = a \, v_{1,u} \, \mathrm{e}^{\psi(u)} \int_0^\infty t \, \mathrm{e}^{-ut} \, \rho(t v_{1,u}) \, f_T(t w_{2,u}) \, \mathrm{d}t$$

$$= \mathrm{e}^{\psi(u)} \frac{a \, (\gamma!)^a}{\Gamma(\gamma a)} \, w_{2,u}^{\gamma a-1} \sum_{j_1=0}^{\gamma-1} \int_0^{+\infty} \mathrm{e}^{-t(u+\bar{v}_{1,u}+\gamma w_{2,u})} \, t^{\gamma a-1}$$

$$\times \Phi_2^{(\gamma-1)}(a^{(\gamma-1)}; \gamma a; t w_{2,u}, \ldots, (\gamma-1) t w_{2,u}) \, \mathrm{d}t.$$

The change of variable $z = t(u + \bar{v}_{1,u} + \gamma w_{2,u})$ and (2.4.10) in Exton (1976) yield

$$f_{V_{1,u}}(v_{1,u}) = a\,\mathrm{e}^{\psi(u)}\,(\gamma!)^a\,w_{2,u}^{\gamma a-1}\sum_{j_1=0}^{\gamma-1}\frac{1}{(u+\bar{v}_{1,u}+\gamma w_{2,u})^{\gamma a}}$$

$$\times F_D^{(\gamma-1)}\left(\gamma a; a^{(\gamma-1)}; \gamma a; \frac{w_{2,u}}{u+\bar{v}_{1,u}+\gamma w_{2,u}},\dots,\frac{(\gamma-1)w_{2,u}}{u+\bar{v}_{1,u}+\gamma w_{2,u}}\right).$$

Finally, the usual simplification of the Lauricella function above and some simple algebra lead to (25). Similarly, one recovers (26) upon noting that the numerator of (23) can be rewritten as

$$\int_0^{+\infty} t^i e^{-ut}\,\{\prod_{j=1}^i \rho(tv_{j,u}w_{j,u})\}\,f_T(tw_{i+1,u})\mathrm{d}t$$

$$=\frac{(\gamma!)^a}{\Gamma(\gamma a)}\frac{w_{i+1,u}^{\gamma a-1}}{\prod_{j=1}^i v_j w_j}\sum_{C_i}\int_0^{+\infty}\mathrm{e}^{-(u+\bar{v}_{i,u}+\gamma w_{i+1,u})t}\,t^{\gamma a-1}$$

$$\times\;\Phi_2^{(\gamma-1)}(a^{(\gamma-1)};\gamma a;tw_{i+1,u},\dots,(\gamma-1)tw_{i+1,u})\;\mathrm{d}t$$

$$=\frac{(\gamma!)^a}{\Gamma(\gamma a)}\frac{w_{i+1,u}^{\gamma a-1}}{\prod_{j=1}^i v_j w_j}\sum_{C_i}\frac{1}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})^{\gamma a-1}}$$

$$\times\int_0^{+\infty}\mathrm{e}^{-z}\,z^{\gamma a-1}$$

$$\times\Phi_2^{(\gamma-1)}\left(a^{(\gamma-1)};\gamma a;\frac{zw_{i+1,u}}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})},\dots,\frac{z(\gamma-1)w_{i+1,u}}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})}\right)\;\mathrm{d}z$$

$$=\frac{(\gamma!)^a\,w_{i+1,u}^{\gamma a-1}}{\prod_{j=1}^i v_j w_j}\sum_{C_i}\frac{1}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})^{\gamma a-1}}$$

$$\times F_D^{(\gamma-1)}\left(\gamma a; a^{(\gamma-1)};\gamma a;\frac{w_{i+1,u}}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})},\dots,\frac{(\gamma-1)\,w_{i+1,u}}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})}\right),$$

which, by virtue of the simplification of $F_D^{(n)}(c;b_1,\dots,b_n;c;x_1,\dots,x_n)$, reduces to

$$\frac{(\gamma!)^a\,(-1)^{\gamma a-1}\,w_{i+1,u}^{a-1}}{\prod_{j=1}^i v_j w_j}\sum_{C_i}\frac{1}{(u+\bar{v}_{i,u}+\gamma w_{i+1,u})^a}\left\{\left(1-\frac{u+\bar{v}_{i,u}+\gamma w_{i+1,u}}{w_{i+1,u}}\right)_{\gamma-1}\right\}^{-a}.$$

The integral in the denominator can be evaluated in a similar fashion, and, combining the two, one obtains (26). $\square$

7.6 Details of the posterior stick-breaking weights for the normalized generalized gamma process

If one recalls that $\psi(u) = a[(u+\tau)^\sigma - \tau^\sigma]/\sigma$ and sets $\beta = a\tau^\sigma/\sigma$, one has

$$
f_{V_{1,u}}(v_{1,u}) = \frac{a\,\mathrm{e}^{\psi(u)+\beta}\,v_1}{\Gamma(1-\sigma)} \int_0^\infty t\,\mathrm{e}^{-ut}\,t^{-1-\sigma}v_1^{-1-\sigma}\,\mathrm{e}^{-\tau t v_1}\,\mathrm{e}^{-\tau t(1-v_1)}\,f_\sigma(t(1-v_1))\,\mathrm{d}t
$$

$$
= \frac{a\,\mathrm{e}^{\beta_u}\,v_1^{-\sigma}}{\Gamma(1-\sigma)} \int_0^\infty t^{-\sigma}\,\mathrm{e}^{-(u+\tau)t}\,f_\sigma(t(1-v_1))\,\mathrm{d}t
$$

$$
= \frac{a\,\mathrm{e}^{\beta_u}\,v_1^{-\sigma}}{\Gamma(1-\sigma)\,\Gamma(\sigma)} \int_0^\infty y^{\sigma-1} \int_0^\infty \mathrm{e}^{-(y+u+\tau)t}\,f_\sigma(t(1-v_1))\,\mathrm{d}t\,\mathrm{d}y,
$$

and this basically coincides with the same integral evaluated for the determination of f_{V_1} with τ replaced by $u+\tau$. Hence, the density function of $V_{1,u}$ coincides with (18) with β replaced by $\beta_u = a(u+\tau)^\sigma/\sigma$; see proof of Proposition 3. For the determination of the density of $V_{i,u}$, conditional on $V_{1,u}, \ldots, V_{i-1,u}$, for any $i \geq 2$, one proceeds as in the proof of (19) and computes separately the integrals in the numerator and denominator of (23). The numerator is

$$
\int_0^{+\infty} t^i \mathrm{e}^{-ut}\,\Big\{\prod_{j=1}^i \rho(t v_{j,u} w_{j,u})\Big\}\,f_T(t w_{i+1,u})\,\mathrm{d}t
$$

$$
= \frac{\mathrm{e}^\beta\,\prod_{j=1}^i (v_{j,u} w_{j,u})^{-1-\sigma}}{(\Gamma(1-\sigma))^i} \int_0^\infty t^{i\sigma}\,\mathrm{e}^{-ut}\,\mathrm{e}^{-\tau t \sum_{j=1}^i v_{j,u} w_{j,u}}\,\mathrm{e}^{-\tau t w_{i+1,u}}\,f_\sigma(t w_{i+1,u})\,\mathrm{d}t
$$

$$
= \frac{\mathrm{e}^\beta\,\prod_{j=1}^i (v_{j,u} w_{j,u})^{-1-\sigma}}{(\Gamma(1-\sigma))^i\,\Gamma(i\sigma)} \int_0^\infty y^{i\sigma-1} \int_0^\infty \mathrm{e}^{-(y+u+\tau)t}\,f_\sigma(t w_{i+1,u})\,\mathrm{d}t\,\mathrm{d}y
$$

$$
= \frac{\mathrm{e}^\beta\,\sigma^{i-1}}{(\Gamma(1-\sigma))^i\,\Gamma(i\sigma)\,a^i}\,\frac{w_{i+1,u}^{i\sigma-1}}{(\prod_{j=1}^i v_{j,u} w_{j,u})^{1+\sigma}} \sum_{j\geq 0} \frac{(1-i\sigma)_j}{j!}\,\frac{\beta_u^{\frac{j}{\sigma}}}{w_{i+1,u}^\sigma}\,\Gamma\!\left(i - \frac{j}{\sigma}; \frac{\beta_u}{w_{i+1,u}^\sigma}\right).
$$

An analogous expression can be established for the denominator, and the ratio will provide an expression for the density of $V_{i,u}$, conditional on $V_{1,u}, \ldots, V_{i-1,u}$, that coincides with (19) where β is replaced by β_u.   □

References

Arbel, J. and Prünster, I. (2015). "A moment-matching Ferguson & Klass algorithm." *Technical Report.*   710

Bondesson, L. (1992). *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*, volume 76 of *Lecture Notes in Statistics*. Springer-Verlag, New York. MR1224674. doi: http://dx.doi.org/10.1007/978-1-4612-2948-3.   701

Brix, A. (1999). "Generalized gamma measures and shot-noise Cox processes." *Advances in Applied Probability*, 31(4): 929–953. MR1747450. doi: http://dx.doi.org/10.1239/aap/1029955251. 702

Butler, R. W. and Wood, A. T. (2015). "Laplace approximation of Lauricella functions F_A and F_D." *Advances in Computational Mathematics*. To appear. doi: http://dx.doi.org/10.1007/s10444-014-9397-5. 710

Caron, F. (2012). "Bayesian nonparametric models for bipartite graphs." In: *Advances in Neural Information Processing Systems*, 2051–2059. 699

Caron, F. and Fox, E. B. (2015). "Sparse graphs using exchangeable random measures." arXiv:1401.1137. 699

Caron, F., Teh, Y. W., and Murphy, T. B. (2014). "Bayesian nonparametric Plackett–Luce models for the analysis of preferences for college degree programmes." *The Annals of Applied Statistics*, 8(2): 1145–1181. MR3262549. doi: http://dx.doi.org/10.1214/14-AOAS717. 699

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II*. Probability and Its Applications. Springer, New York, second edition. MR2371524. doi: http://dx.doi.org/10.1007/978-0-387-49835-5. 699

Exton, H. (1976). *Multiple Hypergeometric Functions and Applications*. Ellis Horwood Ltd., Chichester. MR0422713. 703, 716, 719

Favaro, S., Lijoi, A., and Pruenster, I. (2012). "On the stick-breaking representation of normalized inverse Gaussian priors." *Biometrika*, 99(3): 663–674. MR2966776. doi: http://dx.doi.org/10.1093/biomet/ass023. 698, 702, 705, 707, 715

Favaro, S., Nipoti, B., and Teh, Y. W. (2015). "Random variate generation for Laguerre-type exponentially tilted α-stable distributions." *Electronic Journal of Statistics*, 9: 1230–1242. MR3355756. doi: http://dx.doi.org/10.1214/15-EJS1033. 711

Favaro, S., Prünster, I., and Walker, S. G. (2011). "On a class of random probability measures with general predictive structure." *Scandinavian Journal of Statistics*, 38(2): 359–376. MR2829605. doi: http://dx.doi.org/10.1111/j.1467-9469.2010.00702.x. 701

Favaro, S. and Walker, S. G. (2013). "Slice sampling σ-stable Poisson–Kingman mixture models." *Journal of Computational and Graphical Statistics*, 1(2): 830–847. MR3173745. doi: http://dx.doi.org/10.1080/10618600.2012.681211. 698, 705, 711

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 22(4): 209–230. MR0350949. 701

Ferguson, T. S. and Klass, M. J. (1972). "A representation of independent increment processes without Gaussian components." *The Annals of Mathematical Statistics*, 43(5): 1634–1643. MR0373022. 709

Gnedin, A. and Pitman, J. (2006). "Exchangeable Gibbs partitions and Stirling triangles." *Journal of Mathematical Sciences*, 138(3): 5674–5685. MR2160320. doi: http://dx.doi.org/10.1007/s10958-006-0335-z. 698, 705

Hjort, N., Holmes, C., Müller, P., and Walker, S. (2010). *Bayesian Nonparametrics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge. MR2722987. doi: http://dx.doi.org/10.1017/CBO9780511802478. 698

Ishwaran, H. and James, L. F. (2001). "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96(453): 161–173. MR1952729. doi: http://dx.doi.org/10.1198/016214501750332758. 698

James, L. F., Lijoi, A., and Prünster, I. (2006). "Conjugacy as a distinctive feature of the Dirichlet process." *Scandinavian Journal of Statistics*, 33(1): 105–120. MR2255112. doi: http://dx.doi.org/10.1111/j.1467-9469.2005.00486.x. 705

— (2009). "Posterior analysis for normalized random measures with independent increments." *Scandinavian Journal of Statistics*, 36(1): 76–97. MR2508332. doi: http://dx.doi.org/10.1111/j.1467-9469.2008.00609.x. 697, 705, 718

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). "Slice sampling mixture models." *Statistics and Computing*, 21(1): 93–105. MR2746606. doi: http://dx.doi.org/10.1007/s11222-009-9150-y. 708, 709

Kanter, M. (1975). "Stable densities under change of scale and total variation inequalities." *The Annals of Probability*, 3(4): 697–707. MR0436265. 711

Kingman, J. F. C. (1967). "Completely random measures." *Pacific Journal of Mathematics*, 21(1): 59–78. MR0210185. 699

— (1975). "Random discrete distributions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 37(1): 1–22. MR0368264. 697, 702

— (1993). *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford Univ. Press, New York. MR1207584. 699

Lijoi, A., Mena, R. H., and Prünster, I. (2005a). "Bayesian nonparametric analysis for a generalized Dirichlet process prior." *Statistical Inference for Stochastic Processes*, 8(3): 283–309. MR2177315. doi: http://dx.doi.org/10.1007/s11203-005-6071-z. 697, 701

— (2005b). "Hierarchical mixture modeling with normalized inverse-Gaussian priors." *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: http://dx.doi.org/10.1198/016214505000000132. 697, 702

— (2007). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 715–740. MR2370077. doi: http://dx.doi.org/10.1111/j.1467-9868.2007.00609.x. 697, 702, 712, 713

Lijoi, A. and Prünster, I. (2010). "Models beyond the Dirichlet process." In *Bayesian nonparametrics*, Camb. Ser. Stat. Probab. Math., 80–136. Cambridge Univ. Press, Cambridge. MR2730661. 697, 699

Lijoi, A., Prunster, I., and Walker, S. G. (2008). "Investigating nonparametric priors with Gibbs structure." *Statistica Sinica*, 18(4): 1653–1668. MR2469329. 705

MacEachern, S. N. (1999). "Dependent nonparametric processes." In: *ASA Proceedings of the Section on Bayesian Statistical Science*, 50–55. American Statistical Association, Alexandria, VA. 698

— (2000). "Dependent Dirichlet processes." *Unpublished manuscript, Department of Statistics, The Ohio State University.* 698

Papaspiliopoulos, O. (2008). "A note on posterior sampling from Dirichlet mixture models." *CRiSM Working Paper*, 08–20. 708

Papaspiliopoulos, O. and Roberts, G. O. (2008). "Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models." *Biometrika*, 95(1): 169–186. MR2409721. doi: http://dx.doi.org/10.1093/biomet/asm086. 698

Perman, M., Pitman, J., and Yor, M. (1992). "Size-biased sampling of Poisson point processes and excursions." *Probability Theory and Related Fields*, 92(1): 21–39. MR1156448. doi: http://dx.doi.org/10.1007/BF01205234. 698, 703, 715

Pitman, J. (1996a). "Random discrete distributions invariant under size-biased permutation." *Advances in Applied Probability*, 28(2): 525–539. MR1387889. doi: http://dx.doi.org/10.2307/1428070. 702

— (1996b). "Some developments of the Blackwell–MacQueen urn scheme." In: *Statistics, Probability and Game Theory*, volume 30 of *IMS Lecture Notes – Monograph Series*, 245–267. Institute of Mathematical Statistics, Hayward, CA. MR1481784. doi: http://dx.doi.org/10.1214/lnms/1215453576. 701, 702

— (2003). "Poisson–Kingman partitions." In: *Statistics and Science: a Festschrift for Terry Speed*, volume 40 of *IMS Lecture Notes – Monograph Series*, 1–34. Institute of Mathematical Statistics, Beachwood, OH. MR2004330. doi: http://dx.doi.org/10.1214/lnms/1215091133. 697, 700, 702, 717

Regazzini, E., Lijoi, A., and Prünster, I. (2003). "Distributional results for means of random measures with independent increments." *The Annals of Statistics*, 31(2): 560–585. MR1983542. doi: http://dx.doi.org/10.1214/aos/1051027881. 697, 700, 701

Richardson, S. and Green, P. J. (1997). "On Bayesian analysis of mixtures with an unknown number of components (with discussion)." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4): 731–792. MR1483213. doi: http://dx.doi.org/10.1111/1467-9868.00095. 711

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors." *Statistica Sinica*, 4(2): 639–650. MR1309433. 698

Teh, Y. W., Görür, D., and Ghahramani, Z. (2007). "Stick-breaking construction for the Indian buffet process." In: *International Conference on Artificial Intelligence and Statistics*, 556–563. 698

Von Renesse, M.-K., Yor, M., and Zambotti, L. (2008). "Quasi-invariance properties of a class of subordinators." *Stochastic Processes and their Applications*, 118(11): 2038–2057. MR2462287. doi: http://dx.doi.org/10.1016/j.spa.2007.11.008. 701

Walker, S. G. (2007). "Sampling the Dirichlet mixture model with slices." *Communications in Statistics – Simulation and Computation*, 36(1): 45–54. MR2370888. doi: http://dx.doi.org/10.1080/03610910601096262.   698, 708, 709, 713

Yau, C., Papaspiliopoulos, O., Roberts, G. O., and Holmes, C. (2011). "Bayesian non-parametric hidden Markov models with applications in genomics." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1): 37–57. MR2797735. doi: http://dx.doi.org/10.1111/j.1467-9868.2010.00756.x.   698