Statistical Inference for Stochastic Processes (2005) 8:283–309
DOI 10.1007/s11203-005-6071-z

Bayesian Nonparametric Analysis for a Generalized Dirichlet Process Prior

ANTONIO LIJOI[1,*], RAMSÉS H. MENA[2] and IGOR PRÜNSTER[3]
[1]*Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia, Via San Felice 5, 27100 Pavia, Italy*
[2]*Departimento de Probabilidad y Estadistica, IIMAS-UNAM-México, 04510 Mexico D.F., Mexico*
[3]*Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia, Via San Felice 5, 27100 Pavia, Italy and ICER, Villa Gualino, viale Settimio Severo 63, 10133 Torino, Italy*

Abstract. This paper considers a generalization of the Dirichlet process which is obtained by suitably normalizing superposed independent gamma processes having increasing integer-valued scale parameter. A comprehensive treatment of this random probability measure is provided. We prove results concerning its finite-dimensional distributions, moments, predictive distributions and the distribution of its mean. Most expressions are given in terms of multiple hypergeometric functions, thus highlighting the interplay between Bayesian Nonparametrics and special functions. Finally, a suitable simulation algorithm is applied in order to compute quantities of statistical interest.

AMS 2000 Mathematics Subject Classifications: 62F15, 60G57.

Key words: Bayesian nonparametric inference, Dirichlet process, generalized gamma convolutions, Lauricella hypergeometric functions, means of random probability measures, predictive distributions.

1. Introduction

Since the introduction of the Dirichlet process by Ferguson (1973), there have been several attempts at generalizations. For a review of the literature in the area the reader is referred to Walker et al. (1999). Here we study a particular generalized Dirichlet process obtained by the so called "normalization approach". It is well-known that the Dirichlet process can be constructed by suitably normalizing the increments of a gamma process, i.e. a process having independent increments and whose marginal distribution is gamma. (See Ferguson, 1973). In the same spirit, even though not motivated by applications to Bayesian inference, Kingman (1975) derived

*Author for correspondence: e-mail: lijoi@unipv.it

a random probability measure by normalizing the so-called γ-stable subordinator. Recently, the "normalization approach" has gained new interest in a Bayesian context. In Regazzini et al. (2003), this approach has been used to define a novel class of priors: the normalized random measures with independent increments (normalized RMI). Such random probability measures, whose laws act as priors in a Bayesian nonparametric setting, are constructed via the normalization of suitably reparameterized increasing additive processes, i.e. increasing processes with independent but not necessarily stationary increments. For the purposes of the present paper, it is enough to recall the definition of subordinator rather than the more general one of increasing additive process. An exhaustive account of the theory of both subordinators and increasing additive processes can be found, e.g., in Sato (1999).

A stochastic process $\xi = \{\xi_t : t \geqslant 0\}$, defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$, is an *increasing Lévy process* or *subordinator* if:

(i) for any choice of $n \geqslant 1$ and $0 \leqslant t_0 < t_1 < \cdots < t_n$ the random variables $\xi_{t_0},\ \xi_{t_1} - \xi_{t_0},\ \ldots,\ \xi_{t_n} - \xi_{t_{n-1}}$ are independent;

(ii) the distribution of $\xi_{s+t} - \xi_s$ does not depend on s;

(iii) $\xi_0 = 0$ a.s.-$\mathbb{P}$;

(iv) ξ is stochastically continuous;

(v) there exists $\Omega_0 \in \mathscr{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that $t \mapsto \xi_t(\omega)$ is increasing and right continuous for each $\omega \in \Omega_0$.

Moreover, the Laplace transform of ξ_t is given by $\mathrm{E}[\mathrm{e}^{-\lambda \xi_t}] = \exp[-t \int_{[0+\infty)} (1 - \mathrm{e}^{-\lambda v}) \nu(\mathrm{d}v)]$ for any $\lambda \geqslant 0$, where $\mathrm{E}[\,\cdot\,]$ denotes expectation with respect to $\mathbb{P}$ and ν stands for the Lévy measure corresponding to ξ. It is well-known that a subordinator is uniquely determined by its Lévy measure ν, whose support coincides with $\mathbb{R}^+$ and which satisfies $\int_{(0,+\infty)} (v \wedge 1) \nu(\mathrm{d}v) < +\infty$.

In this paper we study a normalized RMI first considered in Regazzini et al. (2003). In order to define this random probability measure, let ξ be the subordinator identified by the Lévy measure

$$\nu(\mathrm{d}v) = \frac{(1 - \mathrm{e}^{-\gamma v})}{(1 - \mathrm{e}^{-v})} \frac{\mathrm{e}^{-v}}{v} \mathrm{d}v \quad \gamma > 0.$$

Moreover, let α be a finite measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with $\alpha(\mathbb{R}) = a > 0$, and set $A(\cdot) = \alpha(-\infty, \cdot]$. Correspondingly, the time–change $t = A(x)$ yields a reparameterized process $\xi_\alpha = \{\xi_{A(x)}: x \in \mathbb{R}\}$, which still preserves the monotonicity property of the original process. Moreover, since $\nu(0, +\infty) = +\infty$, one has $0 < \xi_a < +\infty$ a.s. –$\mathbb{P}$. Hence

$$\tilde{P}(-\infty, x] = \frac{\xi_{A(x)}}{\xi_a}, \tag{1}$$

defines a random probability distribution function on $\mathbb{R}$ and its law can be employed as a prior for Bayesian nonparametric inference. Notice that when $\gamma = 1$, ξ is a gamma process and $\tilde{P}$ is the Dirichlet process with parameter measure α. Hence, following Regazzini et al. (2003) we term such a random probability measure a *generalized Dirichlet process* with parameters (α, γ).

Here we restrict attention to the case in which γ is a positive integer. This assumption allows us to look at the generalized Dirichlet process as a random probability measure obtained by normalization of superposed independent gamma processes with increasing integer-valued scale parameter. Indeed, according to such a position, the Laplace transform of ξ_α is

$$\mathrm{E}\left[\mathrm{e}^{-\lambda \xi_{A(x)}}\right] = \prod_{j=1}^{\gamma} \left(1 + \frac{\lambda}{j}\right)^{-A(x)} \qquad \forall x \in \mathbb{R},\ \forall \lambda \in \mathbb{R}^+.$$

Thus $\xi_{A(x)}$ is distributed as the convolution of γ independent gamma random variables with parameters $(j, A(x))$, $j = 1, \ldots, \gamma$, for every $x \in \mathbb{R}$. In other terms, the distribution of $\xi_{A(x)}$ is a member of the Thorin class $\mathcal{T}$, also known as the class of generalized gamma convolutions, with Thorin measure $\mu_{A(x)} = A(x) \sum_{j=1}^{\gamma} \delta_j$, where δ_y denotes the unit point mass at y. (See Bondesson, 1992) for a detailed treatment of this and other related classes of distributions.

Before proceeding, it seems useful to relate the generalized Dirichlet process to other classes of random probability measures studied in the literature. First of all, it has to be remarked that the previous construction can be carried out for any subordinator satisfying $\nu(0, +\infty) = +\infty$. This yields a subclass of the family of normalized RMI, which, under the additional assumption of α being non-atomic, essentially coincides with the Poisson-Kingman models, independently proposed by Pitman (2003). Indeed, it turns out that Poisson–Kingman models are also a subclass of species sampling models, another class of random probability measures due to Pitman (1996), which are defined as

$$\tilde{P}(\cdot) = \sum_{i \geqslant 1} P_i \ \delta_{X_i}(\cdot) + \left(1 - \sum_{i \geqslant 1} P_i\right) H(\cdot), \tag{2}$$

where $0 < P_i < 1$ are random weights such that $\sum_{i>1} P_i \leqslant 1$, independent of the locations X_i, which are i.i.d. with some nonatomic distribution H. Although the definition of a species sampling model provides an intuitive and quite general framework, it leaves a difficult open problem, namely, the concrete assignment of the random weights P_i. It is clear that such an issue is crucial for applicability of these models. The most popular approach for achieving this goal is the so-called "stick-breaking" procedure already adopted by Halmos (1944) and Freedman (1963). Stick–breaking priors, as defined in Ishwaran and James (2001) and further developed in Ishwaran and James (2003), contain, among others, the Dirichlet process (defined in a stick–breaking fashion by Sethuraman (1994)), the two-parameter Poisson–Dirichlet process (Pitman, 1995; Pitman and Yor, 1997) and the Dirichlet multinomial process (Muliere and Secchi, 1995). Another possible approach for constructing random weights in (2) is to resort to Poisson–Kingman models. Since the generalized Dirichlet process, defined as in (1) with a nonatomic α, is a Poisson–Kingman model, this paper can also be seen as an attempt to derive explicit expressions for quantities of statistical relevance of a particular species sampling model. Another noteworthy example is the so-called normalized inverse Gaussian process studied in Lijoi et al. (2004). However it is worth remarking that, in general, a normalized RMI is not a species sampling model.

Throughout the paper *exchangeable* observations are considered. Assume that the sequence of observations $(X_n)_{n\geqslant 1}$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in such a way that, conditional on the generalized Dirichlet process $\tilde{P}$, they are independent and identically distributed (i.i.d.) with distribution $\tilde{P}$.

In this work we provide a comprehensive treatment of the generalized Dirichlet process. Most results are expressed in terms of multiple hypergeometric functions highlighting the interplay between Bayesian Nonparametrics and the theory of special functions. Other examples of this close connection can be found in Regazzini (1998) and Lijoi and Regazzini (2004), where functionals of the Dirichlet process are considered. Section 2 is devoted to the derivation of the marginal and the finite-dimensional distributions of the generalized Dirichlet process. Section 3 investigates its structure providing expressions for its expected value, variance and covariance. Moreover its predictive distributions are obtained and, by means of a numerical example, the role of the parameter (a, γ) is investigated. Section 4 is devoted to the study of means of the generalized Dirichlet process. Results concerning the posterior density of the mean, given in Regazzini et al. (2003) are improved and the issue of the symmetry of the distribution of a mean is considered. Section 5 shows how to simulate from such a random probability measure and provides an illustrative example. Proofs are deferred to the Appendix A.

2. Finite-Dimensional Distributions

An important issue when dealing with a random probability measure concerns the determination of its finite-dimensional distributions. Such a task is commonly a hard one to achieve for nonparametric priors, a notable exception being the Dirichlet process. One of the merits of the generalized

Dirichlet process is that its finite-dimensional distributions can be explicitly derived in terms of multivariate hypergeometric functions.

Before stating the main result of the present section, some notations are to be introduced. For any vector $\boldsymbol{x} = (x_1, \dots, x_n)' \in \mathbb{R}^n$, let $|\boldsymbol{x}| := \sum_{i=1}^n x_i$. Denote by $B_1, \dots, B_n$ any partition of $\mathbb{R}$ and set $\alpha_i := \alpha(B_i)$, $i = 1, \dots, n$. Moreover $\Delta_{n-1} = \{\boldsymbol{v} = (v_1, \dots, v_{n-1})' : v_i \geqslant 0,\ i = 1, \dots, n-1,\ |\boldsymbol{v}| \leqslant 1\}$ is the $(n-1)$-dimensional simplex. Finally, using the same notation as in Exton (1976)

$$\Phi_2^{(N)}(\boldsymbol{b}; c; \boldsymbol{x}) = \sum_{m_1, \dots, m_N} \frac{(b_1)_{m_1} \cdots (b_N)_{m_N}}{(c)_{|\boldsymbol{m}|}\, m_1! \cdots m_N!} x_1^{m_1} \cdots x_N^{m_N} \tag{3}$$

stands for the confluent form of the fourth Lauricella hypergeometric function. Here $(b)_m$ is the Pochhammer symbol;

$$(b)_m = b(b+1) \cdots (b+m-1) \quad \text{for any } b > 0 \text{ and positive integer } m,$$

where we assume $(b)_0 = 1$. Moreover, let $\mathbf{1}_{\gamma-1} = (1, \dots, 1)'$, $\boldsymbol{J}_{\gamma-1} = (1, \dots, \gamma-1)'$ denote $(\gamma-1)$-dimensional vectors.

PROPOSITION 1. *Let $\tilde{P}$ be a generalized Dirichlet process with parameter (α, γ). The random vector $(\tilde{P}(B_1), \dots, \tilde{P}(B_{n-1}))$ admits probability density, with respect to the Lebesgue measure on the simplex Δ_{n-1}, given by*

$$\begin{aligned} f(\sigma_1, \dots, \sigma_{n-1}) = {} & \frac{(\gamma!)^a}{\prod_{j=1}^n \Gamma(\gamma\alpha_j)} \sigma_1^{\gamma\alpha_1 - 1} \cdots \sigma_{n-1}^{\gamma\alpha_{n-1}-1} (1 - |\boldsymbol{\sigma}|)^{\gamma\alpha_n - 1} \int_0^{+\infty} v^{\gamma a - 1} e^{-\gamma v} \times \\ & \times \left\{ \prod_{j=1}^{n-1} \Phi_2^{(\gamma-1)}(\alpha_j \mathbf{1}_{\gamma-1}, \gamma\alpha_j; v\sigma_j \boldsymbol{J}_{\gamma-1}) \right\} \times \\ & \times \Phi_2^{(\gamma-1)} \big(\alpha_n \mathbf{1}_{\gamma-1}, \gamma\alpha_n; v(1 - |\boldsymbol{\sigma}|) \boldsymbol{J}_{\gamma-1}\big)\, dv. \end{aligned}$$

In particular, the distribution of $\tilde{P}(B)$, for any $B \in \mathscr{B}(\mathbb{R})$, has density function on $[0, 1]$ *coinciding with*

$$\begin{aligned} f(\sigma) = {} & \frac{(\gamma!)^a}{\Gamma(\gamma\alpha(B))\Gamma(\gamma\alpha(B^c))} \sigma^{\gamma\alpha(B)-1} (1-\sigma)^{\gamma\alpha(B^c)-1} \int_0^{+\infty} v^{\gamma a - 1} e^{-\gamma v} \times \\ & \times \Phi_2^{(\gamma-1)} \big(\alpha(B) \mathbf{1}_{\gamma-1}, \gamma\alpha(B); v\sigma\, \boldsymbol{J}_{\gamma-1}\big) \times \\ & \times \Phi_2^{(\gamma-1)} \big(\alpha(B^c)\, \mathbf{1}_{\gamma-1}, \gamma\alpha(B^c); v(1-\sigma)\, \boldsymbol{J}_{\gamma-1}\big)\, dv. \end{aligned} \tag{4}$$

From Proposition 1, it can be seen that some features of the Dirichlet process carry over, with some modifications, to this more general case. Looking at the density in (4), one immediately recognizes the beta density as the first factor of the product. In contrast to the Dirichlet case, its parameters differ from those of the Dirichlet process by the multiplicative constant γ.

The following corollary highlights some particular cases. The results can be deduced from Proposition 1 using a well-known integral representation of the first Lauricella multiple hypergeometric function

$$F_A^{(N)}(a, \boldsymbol{b}; \boldsymbol{c}; \boldsymbol{x}) = \sum_{m_1,\dots,m_N} \frac{(a)_{|\boldsymbol{m}|}(b_1)_{m_1} \cdots (b_N)_{m_N}}{(c_1)_{m_1} \cdots (c_N)_{m_N}\, m_1! \cdots m_N!} x_1^{m_1} \cdots x_N^{m_N}, \quad (5)$$

with $|x_i| < 1$, for each $i = 1, \dots, N$.

COROLLARY 2. *Let $\tilde{P}$ be a generalized Dirichlet process with parameter $(\alpha, 2)$. Then the probability distribution of the vector $(\tilde{P}(B_1), \dots, \tilde{P}(B_{n-1}))$ admits density function on the simplex Δ_{n-1}*

$$f(\sigma_1, \dots, \sigma_{n-1}) = \frac{\Gamma(2a)\,\sigma_1^{2\alpha_1-1}\sigma_2^{2\alpha_2-1} \dots (1-|\boldsymbol{\sigma}|)^{2\alpha_n-1}}{2^a \prod_{j=1}^n \Gamma(2\alpha_j)} \times$$
$$\times F_A^{(n)}\left(2a, \boldsymbol{\alpha}; 2\boldsymbol{\alpha}; \frac{\sigma_1}{2}, \dots, \frac{1-|\boldsymbol{\sigma}|}{2}\right),$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_{n-1})$.

Remark. Notice that also for any other γ the finite-dimensional distributions can be expressed in terms of a multiple power series, which can be seen as a sort of generalization of Lauricella's F_A. For instance, the marginal density function in (4) can be rewritten as follows

$$f(\sigma) = \frac{\sigma^{\gamma\alpha(B)-1}(1-\sigma)^{\gamma\alpha(B^c)-1}}{B(\gamma\alpha(B), \gamma\alpha(B^c))} \frac{(\gamma!)^a}{\gamma^{\gamma a}} \times$$
$$\times \sum_{i_1,\dots,i_{\gamma-1}; j_1,\dots,j_{\gamma-1}} \frac{(\gamma a)_{|\boldsymbol{i}|+|\boldsymbol{j}|}(\alpha(B^c))_{i_1} \dots (\alpha(B^c))_{i_{\gamma-1}}(\alpha(B))_{j_1} \dots (\alpha(B))_{j_{\gamma-1}}}{i_1! \dots i_{\gamma-1}! j_1! \dots j_{\gamma-1}! (\gamma\alpha(B^c))_{|\boldsymbol{i}|}(\gamma(\alpha(B)))_{|\boldsymbol{j}|}} \times$$
$$\times\ 2^{i_2+j_2} \cdots (\gamma-1)^{i_{\gamma-1}+j_{\gamma-1}}\,\sigma^{|\boldsymbol{i}|}(1-\sigma)^{|\boldsymbol{j}|}.$$

Remark. The random probability measure we are considering could have been constructed by resorting to a different line of reasoning. Indeed, one can start from the finite-dimensional distributions given in Proposition 1 and check they satisfy suitable consistency conditions such as those given in Regazzini (2001).

3. Moments and Predictive Distributions

It is commonly agreed that the nonparametric approach guarantees greater flexibility to inferential procedures. However, such flexibility has to be constrained in order to incorporate real qualitative prior knowledge into the model. This is usually done by tuning some moments according to one's prior opinion. Walker and Damien (1998) suggest controlling the mean and variance of $\tilde{P}$.

The expected value of $\tilde{P}$ takes on the interpretation of a prior guess at the shape of $\tilde{P}$ and is a crucial quantity in terms of prior specification. Indeed, if $\tilde{P}$ is a generalized Dirichlet process with parameter (α, γ), then

$$\mathrm{E}[\tilde{P}(B)] = \frac{\alpha(B)}{a}$$

for any measurable set B. This follows immediately from Pitman (2003).

Having set the prior guess at the shape of $\tilde{P}$ through the choice of α, one has still two degrees of freedom in order to complete the prior specification: the total mass a and the parameter γ. At this point it is useful to introduce the fourth Lauricella multiple hypergeometric function

$$F_D^{(N)}(a, \boldsymbol{b}; c; \boldsymbol{x}) = \sum_{m_1, \dots, m_N} \frac{(a)_{|\boldsymbol{m}|}\,(b_1)_{m_1} \cdots (b_N)_{m_N}}{(c)_{|\boldsymbol{m}|}\, m_1! \cdots m_N!} x_1^{m_1} \cdots x_N^{m_N},$$

where $|x_i| < 1$ for any $i = 1, \dots, N$. Indeed, the variance of $\tilde{P}$ can be expressed as

$$\mathrm{Var}[\tilde{P}(B)] = \frac{\alpha(B)(a - \alpha(B))}{a^2} \mathcal{I}_{a,\gamma}, \tag{6}$$

with

$$\mathcal{I}_{a,\gamma} := \frac{a\,(\gamma!)^a \Gamma(\gamma a)}{\gamma^{\gamma a}\Gamma(\gamma a + 2)} \sum_{k=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a, \boldsymbol{a}_k^*; \gamma a + 2; \frac{1}{\gamma} \boldsymbol{J}_{\gamma-1}\right) \tag{7}$$

having set $\boldsymbol{a}_k^* = (a, \dots, a+2, \dots, a)' \in \mathbb{R}^\gamma$, with $a+2$ being the kth element of the vector. Details of the derivation of (6) are given in the Appendix A.

It is also interesting to determine the covariance structure of a generalized Dirichlet process. It is well known that the Dirichlet process is such that the correlation between disjoint sets is negative. Here we show that this property extends to the generalized Dirichlet process. Let $B_1, B_2 \in \mathscr{B}(\mathbb{R})$ and set $C = B_1 \cap B_2$, then

$$\mathrm{Cov}(\tilde{P}(B_1), \tilde{P}(B_2)) = \frac{a\,\alpha(C) - \alpha(B_1)\,\alpha(B_2)}{a^2} \mathcal{I}_{a,\gamma} \tag{8}$$

and, as a straightforward consequence, one has that

$$\mathrm{Cov}(\tilde{P}(B_1), \tilde{P}(B_2)) = -\frac{\alpha(B_1)\,\alpha(B_2)}{a^2} \mathcal{I}_{a,\gamma},$$

whenever B_1 and B_2 are disjoint. See the Appendix A for the details about the derivation of (8).

Finally, in order to complete the prior specification, we can consider the skewness coefficient which is given by

$$\mathrm{sk}[\tilde{P}(B)] = \frac{a - 2\alpha(B)}{2\sqrt{\alpha(B)(a-\alpha(B))}}\,\mathcal{K}_{a,\gamma} \tag{9}$$

and

$$\mathcal{K}_{a,\gamma} := \frac{4\,\gamma^{\gamma a/2}\,[\Gamma(\gamma a+2)]^{1/2}}{a^{1/2}\,(\gamma!)^{a/2}\,(\gamma a+2)\,[\Gamma(\gamma a)]^{1/2}} \times \\ \times \frac{\sum_{k=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a, \boldsymbol{a}_k^{**}; \gamma a+3; \frac{1}{\gamma}\,\boldsymbol{J}_{\gamma-1}\right)}{\left[\sum_{k=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a, \boldsymbol{a}_k^{*}; \gamma a+2; \frac{1}{\gamma}\,\boldsymbol{J}_{\gamma-1}\right)\right]^{3/2}}, \tag{10}$$

where $\boldsymbol{a}_k^{**} = (a, \dots, a+3, \dots, a)$ with $a+3$ being the kth element of the vector.

An important goal in inferential procedures is the prediction of future values of a random quantity based on its past outcomes. The intuitive structure of the predictive distributions associated with the Dirichlet process

$$\mathbb{P}(X_{n+1} \in \cdot\,|X_1, \dots, X_n) = \frac{a}{a+n}\,\frac{\alpha(\cdot)}{a} + \frac{n}{a+n}\,\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(\cdot)$$

has considerably contributed to its success. Here we show that the predictive distributions of the generalized Dirichlet process still have an intuitive closed form expression: they consist of a linear combination of the prior guess and of a weighted version of the empirical distribution, as shown by Pitman (2003) for Poisson–Kingman models. Moreover, the weights at issue are expressible in terms of fourth type Lauricella multiple hyergeometric functions.

Denote by $X_1^*, \dots, X_k^*$ the k distinct observations within the sample $X_1, \dots, X_n$, $n_j > 0$ terms being equal to X_j^*, for $j = 1, \dots, k$. Consider, now, the following notation

$$\boldsymbol{n} = (n_1, \dots, n_k)', \quad \boldsymbol{n}^+ = (n_1, \dots, n_k, 1)', \quad \boldsymbol{n}_j^+ = (n_1, \dots, n_j+1, \dots, n_k)'$$

and, for $q = 1, \dots, \gamma$, introduce the measures, corresponding to a generalization of the Dirichlet updating mechanism,

$$\alpha_q^*(\cdot\,; \boldsymbol{n}, \boldsymbol{r}^k) := \alpha(\cdot) + \sum_{i=1}^{k} n_i\,\delta_{X_i^*}(\cdot)\,\delta_q(\{r_i\}), \tag{11}$$

where $\boldsymbol{r}^k = (r_1, \dots, r_k)' \in \{1, \dots, \gamma\}^k$. Clearly, $a_q^*(\boldsymbol{n}, \boldsymbol{r}^k) = a + \sum_{i=1}^{k} n_i\,\delta_q(\{r_i\})$ denotes the total mass.

PROPOSITION 3. *Let $\tilde{P}$ be a generalized Dirichlet process with parameter (α,γ). If α is non-atomic, then the predictive distribution, given $X_1,\dots,X_n$ is of the form*

$$\mathbb{P}(X_{n+1}\in\cdot\mid X_1,\dots,X_n)=\frac{a}{\gamma a+n}\,\frac{\alpha(\cdot)}{a}\,w(\boldsymbol{n}^+)+ \\ +\frac{n}{\gamma a+n}\,\frac{1}{n}\sum_{j=1}^{k}n_j\,w(\boldsymbol{n}_j^+)\,\delta_{X_j^*}(\cdot),$$

and the weights are given by

$$w(\boldsymbol{n}^+)=\frac{\sum_{\boldsymbol{r}^{k+1}}F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}^*(\boldsymbol{n}^+,\boldsymbol{r}^{k+1});\gamma a+n+1;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)}{\sum_{\boldsymbol{r}^k}F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}^*(\boldsymbol{n},\boldsymbol{r}^k);\gamma a+n;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)} \tag{12}$$

and

$$w(\boldsymbol{n}_j^+)=\frac{\sum_{\boldsymbol{r}^k}F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}^*(\boldsymbol{n}_j^+,\boldsymbol{r}^k);\gamma a+n+1;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)}{\sum_{\boldsymbol{r}^k}F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}^*(\boldsymbol{n},\boldsymbol{r}^k);\gamma a+n;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)} \tag{13}$$

with $\boldsymbol{a}^*(\boldsymbol{n},\boldsymbol{r}^k):=(a_1^*(\boldsymbol{n},\boldsymbol{r}^k),\dots,a_{\gamma-1}^*(\boldsymbol{n},\boldsymbol{r}^k))$.

It is worth noting that the predictive distributions of the generalized Dirichlet process, besides being interesting from a theoretical point of view, are also useful from a computational perspective, since they are the key ingredient for the simulation algorithm employed in Section 5 for drawing samples from the random probability measure itself.

Let us close this section with some considerations concerning the role of the parameters (a,γ) in the context of prior specification. The prior opinion on the unknown $\tilde{P}$ is reflected by the choice of the parameter measure $P_0=\alpha/a$. Further information can be taken into account by suitably choosing (a,γ). We suggest looking at the variance and skewness of $\tilde{P}$. It can be seen from (6) and (9) that these quantities can be tuned by acting on $\mathcal{I}_{a,\gamma}$ and on $\mathcal{K}_{a,\gamma}$. Here, we provide two tables containing values of $\mathcal{I}_{a,\gamma}$ and $\mathcal{K}_{a,\gamma}$ corresponding to different parameter values (a,γ). Table I shows that the prior variance decreases as a or γ increase. Since the prior variance can also be seen as a measure of the weight given to the prior guess P_0, both a and γ represent the degree of belief in P_0: the bigger a and γ the greater is confidence in P_0. Furthermore, Table I points out that different pairs of (a,γ) can lead to very similar variance structures. Table II shows that also the skewness decreases as a and γ increase and, moreover, that pairs, which yield to similar a priori variances, differ significantly with respect to the skewness. For instance, $\mathcal{I}_{a,\gamma}$ is equal to 0.334 and 0.34 for the the pairs $(2,1)$ and $(1,2)$,

Table I. Values of $\mathcal{I}_{a,\gamma}$ for different choices of (a, γ)

a	γ			
	1	2	5	10
1	0.500	0.340	0.193	0.128
2	0.334	0.209	0.112	0.074
5	0.167	0.098	0.051	0.033
10	0.091	0.052	0.026	0.018

Table II. Values of $\mathcal{K}_{a,\gamma}$ for different choices of (a, γ)

a	γ			
	1	2	5	10
1	0.9428	0.8777	0.0324	2.042e-06
2	0.4396	0.1917	0.0001	2.354e-13
5	0.1488	0.0073	2.067e-11	1.352e-33
10	0.0561	0.0017	3.035e-22	7.331e-67

respectively, but, for $(a, \gamma) = (1, 2)$, the skewness of $\tilde{P}(B)$, for any B, is twice the skewness associated to the case $(2, 1)$. Thus, within the structural constraints, it seems reasonable to set (a, γ) simultaneously in such a way as to incorporate a priori opinions on the variance and skewness.

4. Means

The study of means of random probability measures has been an important area of research in Bayesian Nonparametrics in the past decade. For the Dirichlet case fundamental results have been given in the pioneering works of Cifarelli and Regazzini (1979, 1990) and, more recently, Regazzini et al. (2002) provided a comprehensive treatment of the topic. Currently attention is devoted to the derivation of results for means of random probability measures different from the Dirichlet process: among other contributions, we mention Allaart (2003), Bloomer and Hill (2002), Epifani et al. (2003), Hjort (2003), James (2002), Nieto–Barajas et al. (2004).

Regazzini et al. (2003) obtain conditions for existence and study the distribution of means of normalized RMI under prior and posterior conditions. In particular, with reference to the generalized Dirichlet process at issue, it is shown that, given a real-valued measurable function f, the mean $\int_{\mathbb{R}} f(x)\, \tilde{P}(\mathrm{d}x)$ is finite if and only if

$$\int_{\mathbb{R}} \log(1+\lambda|f(x)|)_{\gamma}\, \alpha(\mathrm{d}x) < +\infty \quad \text{for every } \lambda > 0. \tag{14}$$

One immediately sees that (14) is equivalent to the condition derived in Feigin and Tweedie (1989) for means of the Dirichlet process. Moreover, in Regazzini et al. (2003) the exact prior distribution of the mean is determined. Its expression is given by

$$\mathbb{F}(\sigma)=\frac{1}{2}-\frac{(\gamma!)^a}{\pi}\int_0^{+\infty}\frac{1}{s}\exp\left\{-\frac{1}{2}\sum_{k=1}^{\gamma}\int_{\mathbb{R}}\log(k^2+s^2(f(x)-\sigma)^2)\alpha(\mathrm{d}x)\right\}\times$$
$$\times\sin\left(\sum_{k=1}^{\gamma}\int_{\mathbb{R}}\arctan\frac{s(f(x)-\sigma)}{k}\alpha(\mathrm{d}x)\right)\mathrm{d}s, \tag{15}$$

having denoted by $\mathbb{F}$ the probability distribution function of $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$. As far as the posterior distribution of $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$ is concerned, Regazzini et al. (2003) show that this is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and obtain its expression, by means of a limiting argument, in terms of a Radon–Nikodým derivative. Here we aim at improving this result in the sense that, under the weak additional assumption of absolute continuity of the parameter measure α, we derive a completely explicit expression of the posterior density function of the mean, which is much simpler than the one provided in Regazzini et al. (2003). Before stating the result, we recall that $I_{c^+}^n h(\sigma)=\int_c^{\sigma}\frac{(\sigma-u)^{n-1}}{(n-1)!}h(u)\mathrm{d}u$ is the *Liouville-Weyl fractional integral*, for $n\geqslant 1$, and $I_{c^+}^0$ represents the identity operator. (See, e.g., Oldham and Spanier, 1974).

PROPOSITION 4. *Let $\tilde{P}$ be a generalized Dirichlet process with parameters (α,γ). Suppose α is nonatomic and* (14) *is satisfied. Then, for every $\sigma\in\mathbb{R}$, the posterior density function of $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$, given $(X_1,\ldots,X_n)=(x_1,\ldots,x_n)$, is of the form*

$$\rho(\sigma)=\begin{cases}\frac{1}{\pi}I_{c^+}^{n-1}\,\mathrm{Im}\,\psi(\sigma) & \text{if } n \text{ is even,}\\ \frac{-1}{\pi}I_{c^+}^{n-1}\,\mathrm{Re}\,\psi(\sigma) & \text{if } n \text{ is odd}\end{cases} \tag{16}$$

with

$$\psi(\sigma)=\frac{\Gamma(\gamma a+n)\,\gamma^{\gamma a}\sum_{\boldsymbol{r}^k}\int_0^{+\infty}t^{n-1}\mathrm{e}^{\,-\sum_{q=1}^{\gamma}\int_{\mathbb{R}}\log(q+it(f(x)-\sigma))\,\alpha_q^*(\mathrm{d}x;\boldsymbol{n},\boldsymbol{r}^k)}\mathrm{d}t}{(\gamma!)^a\,\Gamma(\gamma a)\sum_{\boldsymbol{r}^k}F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}^*(\boldsymbol{n},\boldsymbol{r}^k);\gamma a+n;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)}, \tag{17}$$

having defined $\alpha_q^*(\cdot\,;\boldsymbol{n},\boldsymbol{r}^k)$ as in (11)*, and* $c=\inf\operatorname{supp}(\alpha)$.

As far as the Bayes estimate of a mean, under quadratic loss function, is concerned, from Proposition 3, provided α is nonatomic and (14) is satisfied, one has

$$\mathrm{E}\left[\int_{\mathbb{R}} f(x)\,\tilde{P}(\mathrm{d}x)\mid X_1,\ldots,X_n\right]=\frac{w(\boldsymbol{n}^+)}{\gamma a+n}\int_{\mathbb{R}} f(x)\alpha(\mathrm{d}x)+ \\ +\frac{1}{\gamma a+n}\sum_{j=1}^{k} n_j\, w(\boldsymbol{n}_j^+)\, f(X_j^*),$$

where $w(\boldsymbol{n}^+)$ and $w(\boldsymbol{n}_j^+)$ are given by (12) and (13), respectively.

Studying some qualitative properties of the distribution of the mean is also an interesting topic. In particular, we study the symmetry of the distribution of $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$. Such an issue has been considered by a number of authors in the Dirichlet case. In particular, Regazzini et al. (2002) show that if the measure $\alpha\circ f^{-1}$ is symmetric, then the distribution of the mean of Dirichlet process $\int_{\mathbb{R}} f(x)\mathscr{D}_\alpha(\mathrm{d}x)$ is symmetric, having denoted, as usual, by $\alpha\circ f^{-1}$ the distribution of $f(X)$ when X has distribution α. Here we aim at extending such a property to the generalized Dirichlet process. Moreover, we provide an expression for a measure of symmetry such as the skewness coefficient.

PROPOSITION 5. *Let $\tilde{P}$ be a generalized Dirichlet process with parameters (α,γ). If $\alpha\circ f^{-1}$ is symmetric, then the distribution of the mean of a generalized Dirichlet process $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$ is symmetric as well. Further, if $\int_{\mathbb{R}} |f(x)|^3\alpha(\mathrm{d}x)<+\infty$ then the skewness coefficient of $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$ is given by*

$$\mathrm{sk}\left(\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)\right)=c\,\mathrm{sk}(Y),$$

where Y has distribution $(\alpha\circ f^{-1})/a$ and c is a constant equal to

$$\frac{2}{\gamma a+2}\sqrt{\frac{\gamma^{\gamma a+1}(\gamma a+1)}{(\gamma!)^a}}\;\frac{\sum_{r=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}_r^{**}\mathbf{1}_{\gamma-1};\gamma a+3;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)}{\left[\sum_{r=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}_r^{*};\gamma a+2;\frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)\right]^{3/2}}$$

5. Simulation Algorithm

Despite the availability of explicit expressions for many quantities of statistical interest, their practical use is somehow limited by the fact that they involve multiple hypergeometric functions. For this reason, one needs to resort to some computational scheme that allows to draw samples from generalized Dirichlet process priors. In such a framework, knowledge of the predictive distributions, determined in Proposition 3, is crucial. Indeed, most of the simulation algorithms developed in Bayesian Nonparametrics rely upon variations on the Blackwell–MacQueen Pólya urn scheme and upon the development of appropriate Gibbs sampling procedures. (See Escobar, 1994; Mac Eachern, 1994; Escobar and West, 1995) for the Dirichlet case and Pitman (1996) and Ishwaran and James (2001) for extensions to more general random probability measures.

In our case, a sample $\boldsymbol{X}=(X_1,\dots,X_n)$ from a generalized Dirichlet process with parameters (α,γ) is characterized through the following generalized Pólya urn scheme. Let $Z_1,\dots,Z_n$ be an i.i.d. sample from $\alpha(\cdot)/a$ and, for any $i=1,\dots,n$, let $\boldsymbol{n}_i=(n_{i,1},\dots,n_{i,k_i})$ be the frequencies of k_i distinct observations, $X^*_{i,1},\dots,X^*_{i,k_i}$, in $\boldsymbol{X}_{i-1}=(X_1,\dots,X_{i-1})$. Moreover, $\boldsymbol{n}_i^+=(n_{i,1},\dots,n_{i,k_i},1)$, $\boldsymbol{n}^+_{i,j}=(n_{i,1},\dots,n_{i,j}+1,\dots,n_{i,k_i},1)$. Then, the sample $\boldsymbol{X}$ can be generated as follows: $X_1=Z_1$ and for $i=2,\dots,n$

$$(X_i\mid \boldsymbol{X}_{i-1})=\begin{cases} Z_i, & \text{with prob } \left\{a\,w(\boldsymbol{n}_i^+)\right\}/(\gamma\,a+n)\\ X^*_{i,j}, & \text{with prob } \left\{n_j\,w(\boldsymbol{n}^+_{i,j})\right\}/(\gamma\,a+n) \quad j=1,\dots,k_i,\end{cases} \tag{18}$$

where $w(\boldsymbol{n}_i^+), w(\boldsymbol{n}^+_{i,j})$ are given in Proposition 2.

Some other Pólya urn Gibbs sampler methods can be based on the above scheme in order to fit general semiparametric settings. For example, we can consider the following hierarchical model

$$\begin{aligned}(Y_i\mid X_i,\theta)&\overset{\text{ind}}{\sim}\mathscr{L}(Y_i\mid X_i,\theta),\quad i=1,\dots,n\\ (X_i\mid P)&\overset{\text{iid}}{\sim}P\\ \theta&\sim\mathscr{L}(\theta)\\ P&\sim\tilde{P}(\alpha,\gamma),\end{aligned} \tag{19}$$

where $\tilde{P}(\alpha,\gamma)$ denotes a generalized Dirichlet process with prior guess $\alpha(\cdot)/a$. Model (19), with $\tilde{P}$ corresponding to the Dirichlet process, has been popularized by Escobar and West (1995). Further implications of its partially exchangeable structure on the partitioning of the observations are carefully examined in Petrone and Raftery (1997). Integrating out over P in (19) we get

$$\begin{aligned}(Y_i\mid X_i,\theta)&\overset{\text{ind}}{\sim}\mathscr{L}(Y_i\mid X_i,\theta),\quad i=1,\dots,n\\ (X_1,\dots,X_n)&\sim\mathscr{L}(X_1,\dots,X_n)\\ \theta&\sim\mathscr{L}(\theta),\end{aligned} \tag{20}$$

where $\mathscr{L}(X_1,\dots,X_n)$ denotes the joint law of $\boldsymbol{X}=(X_1,\dots,X_n)$ characterized by the Pólya urn (18). Following Ishwaran and James (2001), a Gibbs sampler algorithm can be used to sample from the posterior distribution $\mathscr{L}(\boldsymbol{X}, \theta \mid \boldsymbol{Y})$ of (20). Having denoted by $\boldsymbol{X}_{-i}$ the vector $\boldsymbol{X}$ with the ith coordinate deleted, we proceed by iteratively drawing samples from the distributions of $(X_i \mid \boldsymbol{X}_{-i}, \theta, \boldsymbol{Y})$, for $i = 1, \ldots, n$, and the distribution of $(\theta \mid \boldsymbol{X}, \boldsymbol{Y})$. Each iteration of the algorithm consists of two steps:

(i) for each $i = 1, \ldots, n$, generate X_i values from the conditional distribution

$$\mathbb{P}(X_i \in \cdot \mid \boldsymbol{X}_{-i}, \theta, Y) = q^*_{i,0}\, \mathbb{P}(X_i \in \cdot \mid \theta, Y_i) + \sum_{j=1}^{k_i} q^*_{i,j}\, \delta_{X^*_j}(\cdot), \tag{21}$$

where the mixing proportions are given by

$$q^*_{i,0} \propto \left\{ \frac{a w(\boldsymbol{n}^+_i)}{\gamma a + n} \right\} \int_{\mathcal{X}} f(Y_i \mid X, \theta) \overline{\alpha}(\mathrm{d}X) \quad \text{and}$$
$$q^*_{i,j} \propto \left\{ \frac{n_{i,j} w(\boldsymbol{n}^+_{i,j})}{\gamma a + n} \right\} f(Y_i \mid X^*_j, \theta),$$

subject to the constraint $\sum_{j=0}^{k_i} q^*_{i,j} = 1$. Here f denotes the density corresponding to $\mathscr{L}(Y \mid X, \theta)$ and $\overline{\alpha}(\cdot)$ the distribution corresponding to $\alpha(\cdot)/a$.

(ii) generate θ from a probability distribution whose density function is given by

$$\mathscr{L}(\mathrm{d}\theta \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \mathscr{L}(\mathrm{d}\theta) \prod_{i=1}^{n} f(Y_i \mid X_i, \theta). \tag{22}$$

Some other Gibbs sampler methods can be also generalized such as the acceleration method described in Ishwaran and James (2001).

The added computational difficulty when using a generalized Dirichlet process ($\gamma > 1$) as a nonparametric prior, rather than a Dirichlet process, derives from the complexity of the weight functions. These functions have to be computed $k + 1$ times within each step of the Pólya urn scheme, which makes the algorithms slow. However, if $\gamma = 2$ or $\gamma = 3$ then $F_D^{(\gamma-1)}$ reduces to the Gauss or the Appell hypergeometric function, respectively. These functions can be quickly computed by means of commonly used software packages. If $\gamma \geqslant 4$, one can determine a numerical approximation of the weights by relying on formula (A3) given in the Appendix. Further simplification is due to the fact that the weights are invariant with respect to permutations of the components of the vector $\boldsymbol{n} = (n_1, \ldots, n_k)$. Therefore, we only need to compute the weights corresponding to the unique set of partitions of the integer n, representing the sample size in the Pólya urn. For example, if $n=4$ then we have $n_{(1)}=4$, $(n_{(1)}, n_{(2)})=(1,3)$ or $(n_{(1)}, n_{(2)})=(2,2)$, $(n_{(1)}, n_{(2)}, n_{(3)})=(1,1,2)$ and $(n_{(1)}, \ldots, n_{(4)})=(1,1,1,1)$ as unique partitions, from which, for instance, the weights corresponding to $(n_1, n_2, n_3)=(1,2,1)$ are expressed as $(w_1, w_2, w_3)=(w_{(1)}, w_{(3)}, w_{(2)})$. Hence, we can construct a table with all the possible weights to be read within the sampling algorithms instead of computing them each time is needed. The number $\pi(n)$ of partitions of an integer n as a sum of positive integers can be found by the recursion formula

$$\pi(n)=\sum_{j=1}^{n}(-1)^{j+1}\left[\pi\left(n-\frac{1}{2}j(3j-1)\right)+\pi\left(n-\frac{1}{2}j(3j+1)\right)\right].$$

For other recursion formulas as well as algorithms to find the partitions we refer to Skiena (1990). It is also worth noticing that the weights are independent from the sample, which implies that the same weights-database works for any sample of the same size.

EXAMPLE. Let us consider a data set $\mathrm{Y}=(Y_1, \ldots, Y_{40})$, where the first 20 observations are drawn from a normal distribution $\mathrm{N}(2,1)$ and $Y_{21}, \ldots, Y_{40}$ from a normal distribution $\mathrm{N}(7,1)$. We then consider the following hierarchical model to describe such data:

$$\begin{aligned}
(Y_i \mid X_i, \theta) &\overset{\text{ind}}{\sim} \mathrm{N}(Y_i \mid X_i; 1), \quad i=1,\ldots,n\\
(X_i \mid P) &\overset{\text{iid}}{\sim} P\\
P &\sim \tilde{P}(a,\gamma).
\end{aligned}$$

We compare the performance of different random probability measures with the same prior guess at the shape of $\tilde{P}$ coinciding with a normal distribution with mean $\overline{Y}$ and variance two times the range of the data. When considering the generalized Dirichlet process, we fix $a=1$ and consider the two random probability measures corresponding to $\gamma=2$ and $\gamma=3$. For the sake of comparison with the Dirichlet process, we set the total mass a^* in order to match its variance with the variance of each of the two generalized Dirichlet processes above. Hence, if we notice from Equation (7) that in the Dirichlet case $\mathcal{I}_{a^*,1}=1/(a^*+1)$, then the Dirichlet process matches the variance of $\tilde{P}(\alpha,\gamma)$ if $a^*=\mathcal{I}_{a,\gamma}^{-1}-1$. Hence, for $a=1$ and $\gamma=2$ we have $a^*=1.93$, whereas for $a=1$ and $\gamma=3$ we have $a^*=2.77$.

Given this, we construct the database with the 31185 different weights corresponding to the partition of $n=39$ and then carry out 10,000 iterations from the generalized Pólya urn Gibbs sampler. Table III features the weights for a specific partition of $39=19+9+8+3$ and for different values of γ. In general, increasing the value of γ leads to an increase in the

Table III. Weights from a generalized Dirichlet process with parameter $a=1$ and $\gamma=1,2,\ldots,10$. The weights correspond to the partition of 39 given by $(n_1,n_2,n_3,n_4)=(19,9,8,3)$

γ	w_0	w_1 $(n_1=19)$	w_2 $(n_2=9)$	w_3 $(n_3=8)$	w_4 $(n_4=3)$	w_1/w_4
1	0.025	0.475	0.225	0.200	0.075	6.333
2	0.048	0.463	0.219	0.194	0.073	6.376
3	0.070	0.455	0.213	0.189	0.070	6.462
4	0.090	0.449	0.208	0.185	0.068	6.594
5	0.107	0.443	0.204	0.180	0.066	6.763
6	0.122	0.439	0.200	0.177	0.063	6.952
7	0.134	0.435	0.197	0.174	0.061	7.145
8	0.145	0.431	0.195	0.171	0.059	7.328
9	0.154	0.427	0.193	0.169	0.057	7.494
10	0.161	0.424	0.191	0.168	0.056	7.641

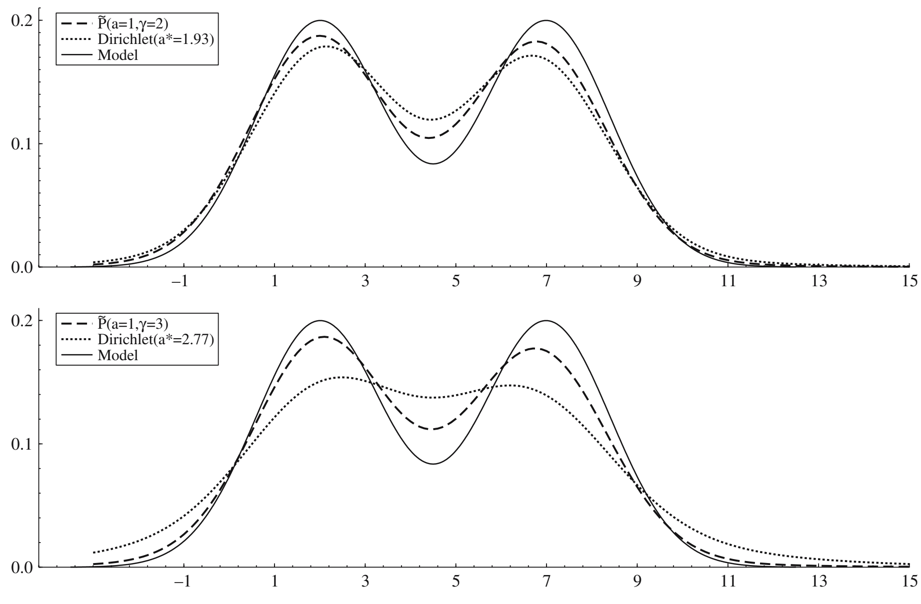

Figure 1. Posterior density estimates corresponding to 10,000 MCMC drawings.

weight given to the prior guess at the shape (w_0), but at the same time it also increases the relative weight of the groups with higher frequencies. As shown in the last column of Table III, the relative weight given to $n_1=19$ with respect to $n_4=3$ increases along with γ. This suggests a sort of reinforcing mechanism working for the generalized Dirichlet process. Figure 1 shows the smoothed MC estimates for the posterior densities resulting from the generalized Pólya urn Gibbs sampler. Note that, with the same number of iterations, the use of a generalized Dirichlet process provides a more accurate fitting than the corresponding Dirichlet process. This suggests that the generalized Dirichlet process prior may represent a fruitful alternative to the more commonly used Dirichlet process and might stimulate further investigation on specific applications with real datasets.

Appendix A

Proof (of Proposition 1). The argument we shall use basically relies upon the representation of the density function of the sum of independent gamma r.v.'s in terms of the confluent form $\Phi_2^{(n)}$, of the fourth Lauricella function. (See Exton, 1976). Hence, for any $B \in \mathscr{B}(\mathbb{R})$, the density function of $\xi_{\alpha(B)}$ is

$$f_{\xi_{\alpha(B)}}(v) = \frac{(\gamma!)^{\alpha(B)}}{\Gamma(\gamma\alpha(B))} \mathrm{e}^{-\gamma v} v^{\gamma\alpha(B)-1} \ \Phi_2^{(\gamma-1)}(\alpha(B)\mathbf{1}_{\gamma-1}; \gamma\alpha(B); v\,\boldsymbol{J}_{\gamma-1}). \tag{A1}$$

In order to determine the density function of the vector

$$\left(\frac{\xi_{\alpha(B_1)}}{\xi_a}, \ldots, \frac{\xi_{\alpha(B_{n-1})}}{\xi_a}\right) = (\tilde{P}(B_1), \ldots, \tilde{P}(B_{n-1})),$$

we use (A1) with the fact that $\xi_a \stackrel{d}{=} \sum_{i=1}^n \xi_{\alpha_i}$, having set $\alpha_i = \alpha(B_i)$, for $i = 1, \ldots, n$. The density function of $(\xi_{\alpha_1}, \ldots, \xi_{\alpha_n})$ is given by

$$f(v_1, \ldots, v_n) = \frac{(\gamma!)^a \,\mathrm{e}^{-\gamma|\boldsymbol{v}|} \left\{\prod_{j=1}^n v_j^{\gamma\alpha_j-1}\right\}}{\prod_{j=1}^n \Gamma(\gamma\alpha_j)} \prod_{j=1}^n \Phi_2^{(\gamma-1)}(\alpha_j\,\mathbf{1}_{\gamma-1}; \gamma\alpha_j; v_j\,\boldsymbol{J}_{\gamma-1})$$

and, considering the transformed random vector $(Y_1, \ldots, Y_{n-1}) = (\xi_{\alpha_1}, \ldots, \xi_{\alpha_{n-1}}, \xi_a)$, it is easy to determine the corresponding density function which coincides with

$$\begin{aligned} f(y_1, \ldots, y_{n-1}, y) = {} & \frac{(\gamma!)^a \,\mathrm{e}^{-\gamma y} \left\{\prod_{j=1}^{n-1} y_j^{\gamma\alpha_j-1}\right\} (y - |\boldsymbol{Y}|)^{\gamma\alpha_n-1}}{\prod_{j=1}^n \Gamma(\gamma\alpha_j)} \\ & \times \prod_{j=1}^{n-1} \Phi_2^{(\gamma-1)}(\alpha_j\,\mathbf{1}_{\gamma-1}; \gamma\alpha_j; y_j\,\boldsymbol{J}_{\gamma-1}) \Phi_2^{(\gamma-1)} \\ & \times (\alpha_n\,\mathbf{1}_{\gamma-1}; \gamma\alpha_n; (y - |\boldsymbol{Y}|)\,\boldsymbol{J}_{\gamma-1}). \end{aligned}$$

Finally, the required density is obtained by the variable transformation $\sigma_j = y_j/y$, for $j=1,\dots,n-1$, and $\sigma = y$, and integration over σ. □

Proof (of Corollary 2). Since in the case $\gamma=2$, Φ_2 reduces to the confluent hypergeometric function ${}_1F_1$, one simply needs to interchange the operation of integration and multiple summation to obtain

$$
\begin{aligned}
f(\sigma_1,\dots,\sigma_{n-1}) = & \frac{2^a\,\sigma_1^{2\alpha_1-1}\cdots\sigma_{n-1}^{\alpha_{n-1}-1}(1-|\boldsymbol{\sigma}|)^{\alpha_n-1}}{\prod_{j=1}^n\Gamma(2\alpha_j)}\times\\
& \times\sum_{m_1,\dots,m_n}\frac{(\alpha_1)_{m_1}\cdots(\alpha_n)_{m_n}\sigma_1^{m_1}\cdots\sigma_{n-1}^{m_{n-1}}(1-|\boldsymbol{\sigma}|)^{m_n}}{(2\alpha_1)_{m_1}\cdots(2\alpha_n)_{m_n}m_1!\cdots m_n!}\times\\
& \times\int_0^{+\infty}\sigma^{2a+|\boldsymbol{m}|-1}\mathrm{e}^{-2\sigma}\,\mathrm{d}\sigma
\end{aligned}
$$

and the result follows by solving the gamma integral. □

Details for the determination of (6)*,* (8) *and* (9)*.* From Proposition 1 in James et al. (2004), variance, covariance and skewness of a normalized RMI based upon a subordinator are given by (6), (8) and (9), respectively, with $\mathcal{I}_{a,\gamma}$ defined as

$$
\mathcal{I}_{a,\gamma}=a\int_0^{+\infty}u\,\mathrm{e}^{-a\,\psi(u)}\int_0^{+\infty}v^2\,\mathrm{e}^{-uv}\,\nu(\mathrm{d}v)\,\mathrm{d}u,
$$

where ψ stands for the Laplace exponent of the random measure at issue and ν is the Lévy measure and $\mathcal{K}_{a,\gamma}$ given by

$$
\mathcal{K}_{a,\gamma}=\frac{a\int_0^{+\infty}u^2\,\mathrm{e}^{-a\,\psi(u)}\int_0^{+\infty}v^3\,\mathrm{e}^{-uv}\,\nu(\mathrm{d}v)\,\mathrm{d}u}{\mathcal{I}_{a,\gamma}^{3/2}}.
$$

Hence, in the case of a generalized Dirichlet process with parameter (α,γ) one has

$$
\begin{aligned}
\mathcal{I}_{a,\gamma} &= a\,(\gamma!)^a\sum_{k=0}^{\gamma-1}\int_0^{+\infty}\frac{u}{\prod_{j=1}^{\gamma}(j+u)^a}\int_0^{+\infty}v\,\mathrm{e}^{-(u+k+1)v}\,\mathrm{d}v\,\mathrm{d}u\\
&= a\,(\gamma!)^a\sum_{k=1}^{\gamma}\int_0^{+\infty}\frac{u}{(u+k)^2\prod_{j=1}^{\gamma}(j+u)^a}\,\mathrm{d}u\\
&= a\sum_{k=1}^{\gamma}k^{-2}\int_0^{+\infty}u\,\mathrm{E}\left[\mathrm{e}^{-u(\eta_k+\zeta_a^{(\gamma)})}\right]\mathrm{d}u = a\sum_{k=1}^{\gamma}k^{-2}\,\mathrm{E}\left[\frac{1}{(\eta_k+\zeta_a^{(\gamma)})^2}\right],
\end{aligned}
$$

where η_k is a Gamma$(k,2)$ random variable and $\zeta_a^{(\gamma)} \stackrel{d}{=} \sum_{j=1}^{\gamma} Z_j^*$, the Z_j^*s being independent and Gamma(j,a) r.v.s. Moreover, η_k and $\zeta_a^{(\gamma)}$ are independent. If $Y_k=\eta_k+\zeta_a^{(\gamma)}$, the density function of Y_k coincides with

$$f_k(y)=\frac{(\gamma!)^a\,k^2}{\Gamma(\gamma a)}\int_0^y (y-x)\,\mathrm{e}^{-k(y-x)}\,x^{\gamma a-1}\,\mathrm{e}^{-\gamma x}\,\Phi_2^{(\gamma-1)}(a\,\mathbf{1}_{\gamma-1};\gamma a;x\,\boldsymbol{J}_{\gamma-1})\,\mathrm{d}x$$

and the change of variable $z=g_1(x)=x/y$ yields

$$f_k(y)=\frac{(\gamma!)^a\,k^2}{\Gamma(\gamma a)}\,y^{\gamma a+1}\,\mathrm{e}^{-ky}\times$$
$$\times\int_0^1 z^{\gamma a-1}(1-z)\,\mathrm{e}^{-(\gamma-k)yz}\,\Phi_2^{(\gamma-1)}(a\,\mathbf{1}_{\gamma-1};\gamma a;yz\,\boldsymbol{J}_{\gamma-1})\,\mathrm{d}z.$$

Hence, turning back to the computation of $\mathcal{I}_{a,\gamma}$, use the Fubini theorem to obtain

$$\begin{aligned}
\mathcal{I}_{a,\gamma}&=a\sum_{k=1}^{\gamma}k^{-2}\int_0^{+\infty}y^{-2}\,f_k(y)\,\mathrm{d}y\\
&=\frac{a(\gamma!)^a}{\Gamma(\gamma a)}\sum_{k=1}^{\gamma}\int_0^1 z^{\gamma a-1}(1-z)\times\\
&\quad\times\int_0^{+\infty}y^{\gamma a-1}\mathrm{e}^{-[k+(\gamma-k)z]y}\,\Phi_2^{(\gamma-1)}(a\,\mathbf{1}_{\gamma-1};\gamma a;zy\,\boldsymbol{J}_{\gamma-1})\mathrm{d}y\,\mathrm{d}z\\
&=a(\gamma!)^a\sum_{k=1}^{\gamma}\int_0^1\frac{z^{\gamma a-1}(1-z)}{[k+(\gamma-k)z]^{\gamma a}}\times\\
&\quad\times F_D^{(\gamma-1)}\left(\gamma a,a\,\mathbf{1}_{\gamma-1};\gamma a;\frac{z}{k+(\gamma-k)z}\,\boldsymbol{J}_{\gamma-1}\right)\mathrm{d}z
\end{aligned}$$

and the last equality follows from (2.4.10) in Exton (1976). Now use the fact that, for any $\boldsymbol{x}\in[0,1)^n$, $a>0$ and $b_i>0$ $(i=1,\dots,n)$, $F_D^{(n)}(a,\boldsymbol{b};a;\boldsymbol{x})=\prod_{i=1}^n(1-x_i)^{-b_i}$ in order to obtain

$$\begin{aligned}
\mathcal{I}_{a,\gamma}&=\frac{a\,(\gamma!)^a}{\gamma^{\gamma a}}\sum_{k=1}^{\gamma}\int_0^1\frac{v^{\gamma a-1}(1-v)}{[\gamma-(\gamma-k)v]^2}\,F_D^{(\gamma-1)}\left(\gamma a,a\,\mathbf{1}_{\gamma-1};\gamma a;\frac{v}{\gamma}\,\boldsymbol{J}_{\gamma-1}\right)\mathrm{d}v\\
&=\frac{a\,(\gamma!)^a}{\gamma^{\gamma a}}\sum_{k=1}^{\gamma}\int_0^1 v^{\gamma a-1}(1-v)\left[1-\frac{(\gamma-k)v}{\gamma}\right]^{-2}\prod_{j=1}^{\gamma-1}\left[1-\frac{jv}{\gamma}\right]^{-a}\mathrm{d}v\\
&=\frac{a\,(\gamma!)^a\Gamma(\gamma a)}{\gamma^{\gamma a}\Gamma(\gamma a+2)}\sum_{k=1}^{\gamma}F_D^{(\gamma-1)}\left(\gamma a,\boldsymbol{a}_k^*;\gamma a+2;\frac{1}{\gamma}\,\boldsymbol{J}_{\gamma-1}\right)
\end{aligned}$$

having employed the change of variable $z=g_2(v)=kv/[\gamma-(\gamma-k)v]$, (A3) in Lauricella (1893) and having set $\boldsymbol{a}_k^*=(a,\ldots,a+2,\ldots,a)$ with $a+2$ being the kth element of the vector.

Using arguments analogous to those employed for determining $\mathcal{I}_{a,\gamma}$ one has

$$\begin{aligned}\mathcal{J}_a &:= a\int_0^{+\infty} u^2 \mathrm{e}^{-a\psi(u)} \int_0^{+\infty} v^3\, \mathrm{e}^{-uv}\, \nu(\mathrm{d}v)\, \mathrm{d}u \\ &= \frac{4a(\gamma!)^a\Gamma(\gamma a)}{\Gamma(\gamma a+3)\,\gamma^{\gamma a}} \sum_{r=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a, \boldsymbol{a}_r^{**}; \gamma a+3; \frac{1}{\gamma}\,\boldsymbol{J}_{\gamma-1}\right), \qquad \text{(A2)}\end{aligned}$$

where $\boldsymbol{a}_r^{**}=(a,\ldots,a+3,\ldots,a)$ with $a+3$ being the rth element of the vector. Using (A2) and some simple algebra, one obtains $\mathcal{K}_{a,\gamma}$.

Proof (of Proposition 3). Under the hypothesis of diffuseness of α, Pitman (2003) shows that the predictive distributions corresponding to normalized RMIs based upon subordinators is of the form

$$\frac{\Pi_{k+1}^{(n+1)}(n_1,\ldots,n_k,1)}{n\,\Pi_k^{(n)}(n_1,\ldots,n_k)}\,\alpha(\cdot)+\frac{1}{n}\sum_{j=1}^{k}\frac{\Pi_k^{(n+1)}(n_1,\ldots,n_j+1,\ldots,n_k)}{\Pi_k^{(n)}(n_1,\ldots,n_k)}\,\delta_{X_j^*}(\cdot), \qquad \text{(A3)}$$

having defined $\Pi_k^{(n)}(n_1,\ldots,n_k)=\int_0^{+\infty} u^{n-1}\,\mathrm{e}^{-a\psi(u)}\prod_{j=1}^k \mu_{n_j}(u)\,\mathrm{d}u$ and, for $j=1,\ldots,k$, $\mu_{n_j}(u)=\int_{(0,+\infty)} v^{n_j}\mathrm{e}^{-uv}\,\nu(\mathrm{d}v)$. (See also James, 2002; Prünster, 2002) for different derivations of this result. Since, by construction, the generalized Dirichlet process prior lies within this subclass of normalized RMI, let us first determine $\Pi_k^{(n)}(n_1,\ldots,n_k)$. In our case

$$\mu_{n_j}(u)=\sum_{r=0}^{\gamma-1}\int_0^{+\infty} v^{n_j-1}\mathrm{e}^{-(u+r+1)v}\,\mathrm{d}v=\Gamma(n_j)\sum_{r=1}^{\gamma}(u+r)^{-n_j} \qquad \text{(A4)}$$

so that

$$\begin{aligned}\Pi_k^{(n)}(n_1,\ldots,n_k) &= \prod_{j=1}^k \Gamma(n_j)\sum_{\boldsymbol{r}^k}\int_0^{+\infty} u^{n-1}\left\{\prod_{j=1}^{\gamma}(j+u)^{-a}\right\}\left\{\prod_{i=1}^{k}(r_i+u)^{-n_i}\right\}\mathrm{d}u \\ &= \prod_{j=1}^k \Gamma(n_j)\sum_{\boldsymbol{r}^k}\frac{1}{r_1^{n_1}\cdots r_k^{n_k}}\int_0^{+\infty} u^{n-1}\,\mathrm{E}\left[\mathrm{e}^{-uV_{a,k}}\right]\mathrm{d}u \\ &= \Gamma(n)\left\{\prod_{j=1}^k \Gamma(n_j)\right\}\sum_{\boldsymbol{r}^k}\frac{1}{r_1^{n_1}\cdots r_k^{n_k}}\,\mathrm{E}\left[V_{a,k}^{-n}\right],\end{aligned}$$

where $V_{a,k}\overset{d}{=}\sum_{q=1}^{\gamma} Y_q^*$ with the Y_q^*s being independent Gamma$(q, a_q^*(\boldsymbol{n},\boldsymbol{r}^k))$ random variables. Recall the notation $\boldsymbol{a}^*(\boldsymbol{n},\boldsymbol{r}^k):=(a_1^*(\boldsymbol{n},\boldsymbol{r}^k),\ldots,a_{\gamma-1}^*(\boldsymbol{n},\boldsymbol{r}^k))$.

The density function of $V_{a,k}$ can be expressed in terms of the confluent form of the fourth Lauricella function thus yielding

$$\Pi_k^{(n)}(n_1,\dots,n_k)=\frac{(\gamma!)^a\Gamma(n)\prod_{j=1}^k\Gamma(n_j)}{\Gamma(\gamma a+n)}\times$$
$$\times\sum_{\boldsymbol{r}^k}\int_0^{+\infty}z^{\gamma a-1}\mathrm{e}^{-\gamma z}\Phi_2^{(\gamma-1)}(\boldsymbol{a}^*(\boldsymbol{n},\boldsymbol{r}^k);\gamma a+n;z\boldsymbol{J}_{\gamma-1})\mathrm{d}z$$
$$=\frac{(\gamma!)^a\,\Gamma(n)\prod_{j=1}^k\Gamma(n_j)}{\gamma^{\gamma a}\,\Gamma(\gamma a+n)}\,\Gamma(\gamma a)\times$$
$$\times\sum_{\boldsymbol{r}^k}F_D^{(\gamma-1)}\left(\gamma a,a^*(\boldsymbol{n},\boldsymbol{r}^k);\gamma a+n;\frac{1}{\gamma}\,\boldsymbol{J}_{\gamma-1}\right)$$

where we have exploited the variable transformation $z=g(v)=v/\gamma$ and (2.4.10) in Exton (1976). Now one has to compute (A3) and rearrange the terms appropriately in order to obtain the desired result. □

Proof (*of Proposition 4*). In order to derive the posterior distribution we start by discretizing the generalized Dirichlet process according to the procedure proposed in Regazzini et al. (2003). It essentially consists in discretizing the random probability measure and the sample along a tree of nested partitions, which, at level m, is made up of sets $B_{m,0},B_{m,1},\dots,B_{m,k_m+1}$ with $B_{m,0}=(-\infty,-R_m)$ and $B_{m,k_m+1}=(R_m,+\infty)$. Moreover, let R_m be such that $\lim_m R_m=+\infty$ and let $\max_{1\leqslant i\leqslant k_m}|B_{m,i}|\to 0$ as $m\to+\infty$, where $|B|$ is the length of the interval B. The discretized random mean, at level m of the tree, will be of the form $\int_{\mathbb{R}}f(x)\tilde{P}_m(\mathrm{d}x)=\sum_{j=0}^{k_m+1}f(b_{m,j})\tilde{P}(B_{m,j})$, where $b_{m,j}$ is any point in $B_{m,j}$ for $j=1,\dots,k_m$ and $b_{m,0}=-R_m$, $b_{m,k_m+1}=R_m$. Denote, as previously, by $x_1^*,\dots,x_k^*$ the k distinct observations within the sample, $n_j>0$ terms being equal to x_j^*, for $j=1,\dots,k$. Whenever the j-th distinct element, x_j^*, lies in $B_{m,i}$, it is as if we had observed $b_{m,i}$. Note that, whatever tree of partitions has been chosen, there always exists m^* such that for every $m>m^*$ the k distinct observations within the sample fall in k distinct sets of the partition. In Proposition 3 of Regazzini et al. (2003) it is proved that a posterior density function of the discretized mean is given by

$$\frac{(-1)^n}{\mu_m^{(n)}(\boldsymbol{x})}\frac{\partial^n}{\partial r_{m,s_1}^{n_{s_1}}\dots\partial r_{m,s_k}^{n_{s_k}}}I_{a^+}^{n-1}\mathbb{F}(\sigma;r_{m,0},\dots,r_{m,k_m+1})\bigg|_{(r_{m,0},\dots,r_{m,k_m+1})=(f(b_{m,0}),\dots,f(b_{m,k_m+1}))},\tag{A5}$$

where $\mu_m^{(n)}$ indicates the discretized marginal distribution of $(X_1,\dots,X_n)$ and $\mathbb{F}$ the prior distribution function of the discretized mean. In the case of generalized Dirichlet process (A5) becomes

$$\frac{(-1)^n}{\pi\,\mu_m^{(n)}(\boldsymbol{x})} I_{a^+}^{n-1} \mathrm{Im} \int_0^{+\infty} \frac{1}{t} \frac{\partial^n}{\partial r_{m,s_1}^{n_{s_1}} \ldots \partial r_{m,s_k}^{n_{s_k}}} \exp \left\{ -\sum_{q=1}^{\gamma} \sum_{j=0}^{k_{m+1}} \log(q + it(r_{m,j} - \sigma))\, \alpha_{m,j} \right\} \mathrm{d}t. \tag{A6}$$

Now it is useful to rewrite the derivative inside (A6) as

$$\begin{aligned} &\frac{\partial^n}{\partial r_{m,s_1}^{n_{s_1}} \ldots \partial r_{m,s_k}^{n_{s_k}}} \exp \left\{ -\sum_{q=1}^{\gamma} \sum_{j=0}^{k_{m+1}} \log(q + it(r_{m,j} - \sigma))\, \alpha_{m,j} \right\} \\ &= \left(\prod_{l=1}^{k} \alpha_{m,s_l} \right) \exp \left\{ -\sum_{q=1}^{\gamma} \sum_{j=0}^{k_{m+1}} \log(q + it(r_{m,j} - \sigma))\, \alpha_{m,j} \right\} \prod_{l=1}^{k} \Lambda_{\alpha_{m,s_l}}^{n_{s_l}}(t), \end{aligned} \tag{A7}$$

where Λ is defined as

$$\begin{aligned} \Lambda_{\alpha_{m,s_l}}^{n_{s_l}}(t) := &\frac{\mathrm{e}^{\sum_{q=1}^{\gamma} \log(q + it(r_{m,s_l} - \sigma))\, \alpha_{m,s_l}}}{\alpha_{m,s_l}} \times \\ &\times \left\{ \frac{\partial^{n_{s_l}}}{\partial r_{m,s_l}^{n_{s_l}}} \exp \left\{ -\sum_{q=1}^{\gamma} \log(q + it(r_{m,s_l} - \sigma))\, \alpha_{m,s_l} \right\} \right\}. \end{aligned} \tag{A8}$$

Now we move on to computing an explicit expression for $\mu_m^{(n)}$. Indeed, one has

$$\begin{aligned} \mu_m^{(n)}(\boldsymbol{x}) &= \mathrm{E}\left[\xi_a^{-n}\, \xi_{\alpha_{m,s_1}}^{n_{s_1}} \cdots \xi_{\alpha_{m,s_k}}^{n_{s_k}} \right] \\ &= \frac{1}{\Gamma(n)} \int_0^{+\infty} u^{n-1}\, \mathrm{E}\left[\exp(-u\, \xi_a) \prod_{l=1}^{k} \xi_{\alpha_{m,s_l}}^{n_{s_l}} \right] \mathrm{d}u \\ &= \frac{\left(\prod_{l=1}^{k} \alpha_{m,s_l} \right)}{\Gamma(n)} \int_0^{+\infty} u^{n-1} \prod_{q=1}^{\gamma} \left(1 + \frac{u}{q} \right)^{-a} \prod_{l=1}^{k} \Delta_{\alpha_{m,s_l}}^{n_{s_l}}(u) \mathrm{d}u \end{aligned} \tag{A9}$$

having set

$$\Delta_{\alpha_{m,s_l}}^{n_{s_l}}(u) = \frac{\prod_{q=1}^{\gamma} \left(1 + \frac{u}{q} \right)^{\alpha_{m,s_l}}}{\alpha_{m,s_l}} \left\{ \frac{\partial^{n_{s_l}}}{\partial u^{n_{s_l}}} \prod_{q=1}^{\gamma} \left(1 + \frac{u}{q} \right)^{-\alpha_{m,s_l}} \right\}. \tag{A10}$$

Inserting (A7) and (A9) into (A6), one obtains

$$\frac{(-1)^n(n-1)!}{\pi} I_{a^+}^{n-1} \mathrm{Im} \frac{\int_0^{+\infty} \frac{1}{t} \exp\left\{-\sum_{q=1}^{\gamma} \sum_{j=0}^{k_{m+1}} \log(q+it(r_{m,j}-\sigma))\,\alpha_{m,j}\right\} \prod_{l=1}^{k} \Lambda_{\alpha_{m,s_l}}^{-n_{s_l}} \mathrm{d}t}{\int_0^{+\infty} u^{n-1} \prod_{q=1}^{\gamma} \left(1+\frac{u}{q}\right)^{-a} \prod_{l=1}^{k} \Delta_{\alpha_{m,s_l}}^{n_{s_l}}(u)\mathrm{d}u}.$$

Now we have to derive the limiting posterior density. Before proceeding, it is worth to check that, by virtue of the diffuseness of α, for (A8) and (A10) the following relations hold true

$$\Lambda_{\alpha_{m,s_l}}^{n_{s_l}} = (-1)^{-n_{s_l}} (n_{s_l}-1)!\,(it)^{n_{s_l}} \sum_{p=1}^{\gamma} \left(p+it(r_{m,s_l}-\sigma)\right)^{-n_{s_l}} + o(\alpha_{m,s_l})$$

$$\Delta_{\alpha_{m,s_l}}^{n_{s_l}} = (-1)^{-n_{s_l}} (n_{s_l}-1)! \sum_{p=1}^{\gamma} \left(1+\frac{u}{p}\right)^{-n_{s_l}} + o(\alpha_{m,s_l}).$$

Given this, by letting m tend to $+\infty$, Theorem 35.7 in Billingsley (2003) and dominated convergence lead to write the limiting posterior density as (16) with

$$\psi(y) = \frac{(n-1)! \int_0^{+\infty} t^{n-1} \mathrm{e}^{-\sum_{q=1}^{\gamma} \int_{\mathbb{R}} \log(q+it(f(x)-\sigma))\alpha(\mathrm{d}x)} \prod_{l=1}^{k} \sum_{p=1}^{\gamma} \left(p+it(f(x_l^*)-\sigma)\right)^{-n_l} \mathrm{d}t}{\int_0^{+\infty} u^{n-1} \prod_{q=1}^{\gamma} \left(1+\frac{u}{q}\right)^{-a} \prod_{l=1}^{k} \sum_{p=1}^{\gamma} \left(1+\frac{u}{p}\right)^{-n_l} \mathrm{d}u},$$

$$= \frac{(n-1)! \sum_{\boldsymbol{r}^k} \int_0^{+\infty} t^{n-1} \mathrm{e}^{-\sum_{q=1}^{\gamma} \int_{\mathbb{R}} \log(q+it(f(x)-\sigma))\alpha(\mathrm{d}x)} \prod_{l=1}^{k} \left(r_i+it(f(x_l^*)-\sigma)\right)^{-n_l} \mathrm{d}t}{\sum_{\boldsymbol{r}^k} \int_0^{+\infty} u^{n-1} \left\{\prod_{q=1}^{\gamma} \left(1+\frac{u}{q}\right)^{-a}\right\} \prod_{l=1}^{k} \left(1+\frac{u}{r_i}\right)^{-n_l} \mathrm{d}u}. \tag{A11}$$

The result follows by the definition of $\alpha_q^*(\cdot\,;\boldsymbol{n},\boldsymbol{r}^k)$ given in (11) and by simplifying the denominator as in the proof of Proposition 3. □

Proof (of Proposition 5). From the results obtained in Regazzini et al. (2003) it follows immediately that $\int_{\mathbb{R}} f(x)\tilde{P}(\mathrm{d}x)$ is equal in distribution to $\int_{\mathbb{R}} x\tilde{P}'(\mathrm{d}x)$, where $\tilde{P}'$ is a generalized Dirichlet process with parameter $(\alpha \circ f^{-1}, \gamma)$. Hence, we confine ourselves, with no loss of generality, to studying the symmetry of the distribution of $\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)$. If α is symmetric, then for proving symmetry one can exploit the expression of the probability distribution function $\mathbb{F}$ of $\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)$ in (15) derived in Regazzini et al. (2003). Indeed, note that the symmetry of α implies that $\mathbb{F}(-\sigma) = 1-\mathbb{F}(\sigma)$, for any $\sigma \in \mathbb{R}$, by a simple change of variable.

Let us now consider the problem of determining the skewness coefficient of the random mean $\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)$. Indicate by m_k and $\bar{m}_k$, respectively, the kth moment and the kth centered moment of a random variable with distribution $\alpha(\cdot)/a$. Our aim is to show that $\mathrm{E}\big[\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)-m_1\big]^3=c\,\bar{m}_3$ for some constant c. One immediately verifies that

$$\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)-m_1\right]^3=\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)\right]^3-3m_1\,\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)\right]^2+2m_1^3. \tag{A12}$$

Let us start by computing $\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)\right]^3$. To this end we discretize the generalized Dirichlet process according to the same procedure employed in the proof of Proposition 3. Refer to that proof for the notation and set $\tilde{P}(B_{m,i})=\tilde{P}_{m,i}$ and $\alpha(B_{m,i})=\alpha_{m,i}$. Thus, at level m of the tree of partitions one has

$$\begin{aligned}\mathrm{E}\left[\sum_{i=0}^{k_m+1} b_{m,i}\,\tilde{P}_{m,i}\right]^3 &= \sum_{i=0}^{k_m+1} b_{m,i}^3\,\mathrm{E}[\tilde{P}_{m,i}^3]+6\sum_{i_1<i_2} b_{m,i_1}^2 b_{m,i_2}\,\mathrm{E}[\tilde{P}_{m,i_1}^2\tilde{P}_{m,i_2}]+\\ &\quad +6\sum_{i_1<i_2<i_3} b_{m,i_1}b_{m,i_2}b_{m,i_3}\,\mathrm{E}[\tilde{P}_{m,i_1}\tilde{P}_{m,i_2}\tilde{P}_{m,i_3}]\end{aligned} \tag{A13}$$

As far as the computation of $\mathrm{E}[\tilde{P}_{m,i}^3]$ is concerned one has that

$$\begin{aligned}\mathrm{E}[\tilde{P}_{m,i}^3]&=\mathrm{E}\left[\frac{1}{\Gamma(3)}\int_0^{+\infty} u^2\mathrm{e}^{-u\xi_a}\xi_{\alpha_{m,i}}^3\mathrm{d}u\right]\\ &=\frac{1}{2}\int_0^{+\infty} u^2\mathrm{E}\left[-\frac{\mathrm{d}^3}{\mathrm{d}u^3}\mathrm{e}^{-u\xi_{\alpha_{m,i}}}\right]\mathrm{E}\left[\mathrm{e}^{-u\xi_{(a-\alpha_{m,i})}}\right]\mathrm{d}u\\ &=\frac{1}{2}\int_0^{+\infty} u^2\left[-\frac{\mathrm{d}^3}{\mathrm{d}u^3}\mathrm{e}^{-\alpha_{m,i}\psi(u)}\right]\mathrm{e}^{-(a-\alpha_{m,i})\psi(u)}\mathrm{d}u\\ &=\frac{\alpha_{m,i}}{2}\int_0^{+\infty} u^2\mathrm{e}^{-a\psi(u)}\big[\mu_3(u)+3\alpha_{m,i}\mu_1(u)\mu_2(u)\big]\mathrm{d}u+\\ &\quad+\frac{\alpha_{m,i}}{2}\int_0^{+\infty} u^2\mathrm{e}^{-a\psi(u)}\alpha_{m,i}^2\mu_1^3(u)\mathrm{d}u,\end{aligned}$$

where μ_j is defined as in (A4) and ψ stands, as previously, for the characteristic exponent of the reparameterized subordinator used for defining the generalized Dirichlet process. By means of repeated integration by parts one obtains

$$\begin{aligned}&\frac{\alpha_{m,i}}{2}\int_0^{+\infty} u^2\mathrm{e}^{-a\psi(u)}\alpha_{m,i}^2\mu_1^3(u)\mathrm{d}u\\ &=\frac{\alpha_{m,i}^3}{a^3}[1-\mathcal{I}_{a,\gamma}]-\frac{\alpha_{m,i}^3}{a^3}\int_0^{+\infty} u^2\mathrm{e}^{-a\psi(u)}\mu_1(u)\mu_2(u)\mathrm{d}u,\end{aligned} \tag{A14}$$

where the expression of $\mathcal{I}_{a,\gamma}$ is given in (7). Exploitation of (A14), some algebra, integration by parts of $\int_0^{+\infty} u^2 \mathrm{e}^{-a\psi(u)} \mu_1(u)\mu_2(u)\mathrm{d}u$ and (A2) lead to write

$$\mathrm{E}[\tilde{P}^3_{m,i}] = \frac{\alpha_{m,i}a^2 - 3\alpha^2_{m,i} + 2\alpha^3_{m,i}}{2a^3}\, a\,\mathcal{J}_a + \frac{3\alpha^2_{m,i}a - 3\alpha^3_{m,i}}{a^3}\mathcal{I}_{a,\gamma} + \frac{\alpha^3_{m,i}}{a^3}.$$

Now we are left with calculating $\mathrm{E}[\tilde{P}^2_{m,i_1}\tilde{P}_{m,i_2}]$ and $\mathrm{E}[\tilde{P}_{m,i_1}\tilde{P}_{m,i_2}\tilde{P}_{m,i_3}]$. This can be achieved proceeding as for $\mathrm{E}[\tilde{P}^3_{m,i}]$ and one obtains

$$\begin{aligned}\mathrm{E}[\tilde{P}^2_{m,i_1}\tilde{P}_{m,i_2}] = {} & \frac{2\alpha^2_{m,i_1}\alpha_{m,i_2} - \alpha_{m,i_1}\alpha_{m,i_2}a}{2a^3}\,\mathcal{J}_a + \\ & + \frac{\alpha_{m,i_1}\alpha_{m,i_2}a - 3\alpha^2_{m,i_1}\alpha_{m,i_2}}{a^3}\mathcal{I}_{a,\gamma} + \frac{\alpha^2_{m,i_1}\alpha_{m,i_2}}{a^3}\end{aligned}$$

$$\begin{aligned}\mathrm{E}[\tilde{P}_{m,i_1}\tilde{P}_{m,i_2}\tilde{P}_{m,i_3}] = {} & \frac{\alpha_{m,i_1}\alpha_{m,i_2}\alpha_{m,i_3}}{a^3}\,\mathcal{J}_a + \frac{3\alpha_{m,i_1}\alpha_{m,i_2}\alpha_{m,i_3}}{a^3}\mathcal{I}_{a,\gamma} + \\ & + \frac{\alpha_{m,i_1}\alpha_{m,i_2}\alpha_{m,i_3}}{a^3}.\end{aligned}$$

Now inserting these expressions into (A13), letting m tend to infinity and, finally, rearranging the terms appropriately we obtain

$$\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}_m(\mathrm{d}x)\right]^3 = \frac{\mathcal{J}_a}{2}\bar{m}^3 + 3\mathcal{I}_{a,\gamma}\left[m_1 m_2 - m_1^3\right] + m_1^3. \tag{A15}$$

The derivation of $\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}(\mathrm{d}x)\right]^2$ can be carried out in a similar fashion, yielding

$$\mathrm{E}\left[\int_{\mathbb{R}} x\tilde{P}_m(\mathrm{d}x)\right]^2 = m_1^2 - \mathcal{I}_{a,\gamma}(m_1^2 + m_2). \tag{A16}$$

Inserting now (A15) and (A16) into (A12) leads to

$$\begin{aligned}&\mathrm{E}\left[\int_R x\tilde{P}(\mathrm{d}x) - m_1\right]^3 \\ &\quad = \frac{2a(\gamma!)^a\Gamma(\gamma a)\sum_{r=1}^{\gamma} F_D^{(\gamma-1)}\left(\gamma a, \boldsymbol{a}_r^{**}\mathbf{1}_{\gamma-1}; \gamma a + 3; \frac{1}{\gamma}\boldsymbol{J}_{\gamma-1}\right)}{\Gamma(\gamma a + 3)\,\gamma^{\gamma a}}\,\bar{m}_3,\end{aligned}$$

where, as previously, $\boldsymbol{a}_r^{**} = (a, \ldots, a+3, \ldots, a)' \in \mathbb{R}^\gamma$ with $a+3$ being the rth element of the vector. Similar arguments lead to show that $\mathrm{Var}[\int x\,\tilde{P}(\mathrm{d}x)] = \mathcal{I}_{a,\gamma}\,\bar{m}_2$ and the conclusion follows. □

Acknowledgements

The authors are grateful to two anonymous referees for their valuable comments. A. Lijoi and I. Prünster were partially supported by the Italian Ministry of University and Research (MIUR), PRIN 2002015111 and PRIN 2003138887.

References

Allaart, P.: Moments of the mean of Dubins-Freedman random probability distributions, *J. Theor. Probab.* **16** (2003), 471–488.

Billingsley, P.: *Probability and Measure*, 3rd ed. Wiley, New York, 2003.

Bloomer, L. and Hill, T. P.: Random probability measures with given mean and variance, *J. Theor. Probab.* **14** (2002), 919–937.

Bondesson, L.: *Generalized gamma convolutions and related classes of distributions and densities*, Lecture Notes in Statistics Vol. **76**, Springer, New York, 1992.

Cifarelli, D. M. and Regazzini, E.: Considerazioni generali sull'impostazione bayesiana di problemi non parametrici. Le medie associative nel contesto del processo aleatorio di Dirichlet. Parte II, *Rivista di matematica per le scienze economiche e sociali* **1** (1979), 95–111.

Cifarelli, D. M. and Regazzini, E.: Distribution functions of means of a Dirichlet process, *Ann. Stat.* **18** (1990), 429–442 (Correction in *Ann. Statist.* **22** (1994), 1633–1634).

Epifani, I., Lijoi, A. and Prünster, I.: Exponential functionals and means of neutral-to-the-right priors, *Biometrika* **90** (2003), 791–808.

Escobar, M. D.: Estimating normal means with a Dirichlet process prior, *J. Am. Stat. Ass.* **89** (1994), 268–277.

Escobar, M. D. and West, M.: Bayesian density estimation and inference using mixtures, *J. Am. Stat. Ass.* **90** (1995), 577–588.

Exton, H.: *Multiple hypergeometric functions and applications*, Ellis Horwood, Chichester, 1976.

Ferguson, T. S.: A Bayesian analysis of some nonparametric problems, *Ann. Stat.* **1** (1973), 209–230.

Freedman, D. A.: On the asymptotic behavior of Bayes' estimates in the discrete case, *Ann. Math. Stat.* **34** (1963), 1386–1403.

Halmos, P.: Random alms, *Ann. Math. Stat.* **15** (1944), 182–189.

Hjort, N. L.: Topics in non–parametric Bayesian statistics, In: P. J. Green et al. (eds), *Highly Structured Stochastic Systems*, Oxford University Press, Oxford, Oxford Statistical Science Series **27** (2003), 455–478.

Ishwaran, H. and James, L. F.: Gibbs sampling methods for stick-breaking priors, *J. Am. Stat. Ass.* **96** (2001), 161–173.

Ishwaran, H. and James, L. F.: Some further developments for stick–breaking priors: finite and infinite clustering and classification, *Sankhya* **65** (2003), 577–592.

James, L. F.: Poisson Process Partition Calculus with applications to exchangeable models and Bayesian Nonparametrics, *Mathematics ArXiv math.PR/0205093* (2002).

James, L. F., Lijoi, A. and Prünster, I.: Conjugacy as a distinctive feature of the Dirichlet process, *Scandinavian Journal of Statistics* (2004), to appear.

Kingman, J. F. C.: Random discrete distributions, *J. Roy. Stat. Soc. Ser. B* **37** (1975), 1–22.

Lauricella, G.: Sulle funzioni ipergeometriche a più variabili, *Rend. Circ. Mat. Palermo* **7** (1893), 111–158.

Lijoi, A., Mena, R. H. and Prünster, I.: Hierarchical mixture modelling with normalized inverse Gaussian priors, *ICER Working Papers* **12/04** (2004).

Lijoi, A. and Regazzini, E.: Means of a Dirichlet process and multiple hypergeometric functions, *Ann. Probab.* **32** (2004), 1469–1495.

MacEachern, S. N.: Estimating normal means with a conjugate style Dirichlet process prior, *Commun. Statist. Simulation Comp.* **23** (1994), 727–741.

Muliere, P. and Secchi, P.: A note on a proper Bayesian bootstrap, *Quaderni del Dipartimento di Economia Politica e Metodi Quantitativi* **18** (1995), Universtià degli Studi di Pavia.

Nieto-Barajas, L. E., Prünster, I. and Walker, S. G.: Normalized random measures driven by increasing additive processes, *Ann. Stat.* **32** (2004), to appear.

Petrone, S. and Raftery, A. E.: A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability, *Stat. Prob. Lett.* **36** (1997), 69–83.

Pitman, J.: Exchangeable and partially exchangeable random partitions, *Prob. Theory Related Fields* **102** (1995), 145–158.

Pitman, J.: Some developments of the Blackwell-MacQueen urn scheme, In: T. S. Ferguson et al. (eds), *Statistics, probability and game theory. Papers in honor of David Blackwell*, Institute of Mathematical Statistics, Hayward, Lecture Notes, Monograph Series Vol 30, 1996, pp. 245–267.

Pitman, J.: Poisson-Kingman partitions. In: Goldstein, D.R. (ed), *Science and Statistics: A Festschrift for Terry Speed*, Institute of Mathematical Statistics, Hayward, Lecture Notes, Monograph Series, Vol 40, 2003, pp. 1–35.

Pitman, J. and Yor, M.: The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator, *Ann. Prob.* **25** (1997), 855–900.

Prünster, I.: *Random probability measures derived from increasing additive processes and their application to Bayesian statistics*, Ph.d dissertation, Università degli Studi di Pavia, Pavia, 2002.

Oldham, K.B. and Spanier, J.: *The fractional calculus*, Academic Press, New York, 1974.

Regazzini, E.: An example of the interplay between statistics and special functions, In: *Tricomi's ideas and contemporary applied mathematics*, Acc. Naz. Lincei e Acc. Scienze Torino, Roma, Atti dei Convegni Lincei Vol. 147, 1998, pp. 303–320.

Regazzini, E.: *Foundations of Bayesian statistics and some theory of Bayesian nonparametric methods*, Lecture Notes, Stanford University, Stanford, 2001.

Regazzini, E., Guglielmi, A. and Di Nunno, G.: Theory and numerical analysis for exact distribution of functionals of a Dirichlet process, *Ann. Stat.* **30** (2002), 1376–1411.

Regazzini, E., Lijoi, A. and Prünster, I.: Distributional results for means of random measures with independent increments, *Ann. Stat.* **31** (2003), 560–585.

Sato, K.: *Lévy processes and infinitely divisible distributions*, Cambridge University Press, Cambridge, 1999.

Sethuraman, J.: A constructive definition of Dirichlet priors, *Stat. Sinica* **4** (1994), 639–650.

Skiena, S.: *Implementing discrete mathematics: combinatorics and graph theory with Mathematica*, Addison-Wesley, Reading MA, 1990.

Walker, S. G. and Damien, P.: A full Bayesian non–parametric analysis involving a neutral to the right process, *Scand. J. Stat.* **25** (1998), 669–680.

Walker, S. G., Damien, P., Laud, P. W. and Smith, A. F. M.: Bayesian nonparametric inference for random distributions and related functions, *J. Roy. Stat. Soc. Ser. B* **61** (1999), 485–527.