doi: 10.1111/j.1467-9469.2005.00486.x
 Vol 33: 105–120, 2006

Conjugacy as a Distinctive Feature of the Dirichlet Process

LANCELOT F. JAMES

Department of Information Systems and Management, Hong Kong University of Science and Technology

ANTONIO LIJOI

Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia and CNR-IMATI, Milano

IGOR PRÜNSTER

Dipartimento di Statistica e Matematica Applicata and ICER, Università degli Studi di Torino

ABSTRACT. Recently the class of normalized random measures with independent increments, which contains the Dirichlet process as a particular case, has been introduced. Here a new technique for deriving moments of these random probability measures is proposed. It is shown that, *a priori*, most of the appealing properties featured by the Dirichlet process are preserved. When passing to posterior computations, we obtain a characterization of the Dirichlet process as the only conjugate member of the whole class of normalized random measures with independent increments.

Key words: Bayesian non-parametrics, conjugacy, Dirichlet process, increasing Lévy process, normalized random measure with independent increments, predictive distribution

1. Introduction

In Ferguson (1973) it is shown that the Dirichlet process prior 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. In the same way, even though not motivated by applications to Bayesian inference, Kingman (1975) derived a random probability measure by normalizing the so-called γ-stable subordinator. These two examples clearly point out a natural approach for defining priors on spaces of probability measures which is based on the normalization of increasing processes having independent increments. However, the Bayesian non-parametric literature has mainly focussed on alternative representations of priors on spaces of probability distributions such as, e.g. neutral to the right priors (Doksum, 1974; Ferguson & Phadia, 1979; Walker & Muliere, 1997; Epifani *et al*., 2003) and Pólya tree processes (Mauldin *et al*., 1992; Lavine, 1992). Only recently the ‘normalization approach’ has gained new interest. In Regazzini *et al*. (2003), it has been successfully employed leading to the definition of 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 non-parametric setting, are constructed via the normalization of suitably reparameterized increasing additive processes, i.e. increasing processes with independent but not necessarily stationary increments. For an exhaustive account of the theory of increasing additive processes (see, e.g. Sato, 1999). This paper focusses attention on a suitable subclass of normalized RMI, namely those generated by increasing Lévy processes also known as ‘subordinators’. To fix ideas we briefly recall the construction of such a random probability measure.

Let $\xi=\{\xi_t : t\geq 0\}$ be a subordinator (without drift) defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. It is well known that any subordinator ξ is uniquely determined by its *Lévy measure* v on $\mathscr{B}(\mathbb{R}^+)$ which satisfies $\int_{\mathbb{R}^+}\min(1,x)v(\mathrm{d}x)<+\infty$. Indicate by α a non-null and finite measure on R. The time change $t=\alpha((-\infty,x])$, yields a new process $\xi_\alpha=\{\xi_{\alpha((-\infty,x])}:x\in\mathbb{R}\}$ which is almost surely finite. Intuitively, one can think of such an operation as observing ξ on the finite interval $(0, \alpha(\mathbb{R}))$ with a time deformation dictated by α. Note that ξ_α shares all the properties of the subordinator ξ but the stationarity of the increments (unless α is a uniform measure on a bounded interval). By virtue of the celebrated Lévy–Khintchine representation, the Laplace transform of ξ_α is given by

$$\mathbb{E}\left[\mathrm{e}^{-\lambda\xi_{\alpha((-\infty,\,x])}}\right]=\mathrm{e}^{-\alpha((-\infty,\,x])\psi(\lambda)}\quad\text{for every } x\in\mathbb{R}$$

where $\psi(\lambda)=\int_{\mathbb{R}^+}(1-\mathrm{e}^{-\lambda v})v(\mathrm{d}v)$ is known as the 'Laplace exponent' and $\mathbb{E}[\cdot]$ stands for expectation with respect to $\mathbb{P}$. The reader interested in time changed subordinators is referred to Barndorff-Nielsen & Shephard (2006) where such type of processes are dealt with in detail. In the following we will use the same symbol ξ_α for denoting both the reparameterized process and the corresponding random measure. At this point, a random probability measure can be easily determined by normalization. The latter operation requires an extra condition to ensure the normalizing random quantity to be positive and finite almost surely. Finiteness is assured by the reparameterization, while for positiveness it suffices to require ξ to be an *infinite activity* process, i.e. a process that jumps infinitely often on any finite time interval. In terms of the Lévy measure infinite activity is equivalent to imposing $v(\mathbb{R}^+)=+\infty$. The random probability measures, obtained by normalization,

$$\tilde{P}(\cdot)=\frac{\xi_{\alpha(\cdot)}}{\xi_{\alpha(\mathrm{R})}}\tag{1}$$

are a subclass of the normalized RMIs, introduced in Regazzini *et al.* (2003), and will be termed 'normalized homogeneous random measures with independent increments' (normalized HRMI). A few interesting facts about $\tilde{P}$ can be quickly pointed out. First, it should be noted that the realizations of $\tilde{P}$ are discrete probability distributions on $\mathbb{R}$, almost surely with respect to $\mathbb{P}$. A proof of this result is given, e.g. in James (2003). Also, observe that x being an atom of α implies that $\tilde{P}(\{x\})>0$ with $\mathbb{P}$–probability 1. Moreover, $\tilde{P}$ is supported by a countable dense subset of $\mathbb{R}$ if and only if $x\mapsto A(x)=\alpha((-\infty,x])$ is strictly increasing.

It seems worth recalling how to obtain the Dirichlet process as a particular case of normalized HRMI as defined in (1). One starts with a Gamma process, i.e. a subordinator characterized by the Lévy measure $v(\mathrm{d}v)=v^{-1}\,\mathrm{e}^{-v}\,\mathrm{d}v$. The corresponding Laplace exponent is given by $\log(1+\lambda)$. Such a process clearly enjoys the infinite activity property. The time change yields an (a.s.) finite Gamma process ξ_α with parameter α and the random probability measure obtained via normalization in (1) is well defined and coincides with the Dirichlet process.

An allied class of random probability measures, the Poisson–Kingman models, independently proposed in a stimulating paper of Pitman (2003) not concerned with Bayesian applications, essentially coincides with the just defined normalized HRMIs under the additional assumption of diffuseness of α (see also James, 2002).

The aim of this paper was to study the distinctive features of the Dirichlet process within the class of normalized HRMI. On the one hand, it is shown that, *a priori*, most of the appealing properties featured by the Dirichlet process are preserved. On the other hand, when passing to posterior computations, we derive a characterization of the Dirichlet process as the only conjugate member of the whole class of normalized HRMI. The reader may note that we obtain opposite results with respect to the ones well known for neutral to the right priors. Indeed, it is known that in generalizing the Dirichlet process to a neutral to the right process one preserves conjugacy, but loses a simple moment structure both *a priori* and *a posteriori*.

To carry out our analysis a new technique for deriving moments of normalized HRMI is introduced. Section 2 investigates some distributional properties normalized HRMI share with the Dirichlet process. Attention is devoted to the expectation, the variance and correlation structure of a normalized HRMI. In section 3 it is shown that the Dirichlet process is the only conjugate normalized HRMI. Finally, predictive distributions are considered. In order to ease the flow of ideas, proofs are given in the appendix.

2. Moments of a normalized HRMI

The Dirichlet process has undoubtedly been playing a prominent role in Bayesian nonparametric inference for the last 30 yr. Its success is to be attributed to a number of properties it enjoys, which can be easily derived by exploiting the knowledge of its finite-dimensional distributions. When considering moments of normalized HRMI, the major difficulty to be overcome is represented by the fact that the finite–dimensional distribution are, in general, unknown even for a non-normalized HRMI.

As far as *a priori* moments are concerned, the first issue to be considered is the determination of the expected value of $\tilde{P}$ which takes the interpretation of prior guess at the shape of $\tilde{P}$. As a normalized HRMI is a species sampling model, i.e. it admits a representation as $\Sigma_{i\geq 1} P_i \delta_{X_i}$, where the sequences of weights $(P_i)_{i\geq 1}$ and locations $(X_i)_{i\geq 1}$ are independent, it follows immediately that

$$P_0(B) := \mathbb{E}\left[\tilde{P}(B)\right] = \frac{\alpha(B)}{a} \quad \text{for every } B \in \mathscr{B}(\mathbb{R}) \tag{2}$$

having set $a := \alpha(\mathbb{R})$ (see Pitman, 2003 for details). Thus, a well-known property of the Dirichlet process prior extends to the more general class of normalized HRMI. The fact that just in the Gamma case the normalized increments are independent of the total mass, might lead to conjecture that (2) is peculiar to the Dirichlet process. Thus, one has that, for any normalized HRMI, the marginal distribution of the first observation does not depend upon the process employed to define it.

To our purposes it is useful to prove (2) by means of a new technique, which does not resort to the theory of species sampling models, as most of the following results heavily rely on this type of arguments. The key point for obtaining (2) is the equality

$$\mathbb{E}[\tilde{P}(B)] = \mathbb{E}\left[\frac{\xi_{\alpha(B)}}{\xi_a}\right] = \int_0^{+\infty} \mathbb{E}\left[\mathrm{e}^{-u\xi_a}\xi_{\alpha(B)}\right] \mathrm{d}u$$

which follows from Fubini's theorem. If one writes $\xi_a = \xi_{\alpha(B)} + \xi_{\alpha(B^c)}$, with B^c denoting the complement of a set B, and exploits independence of the increments of ξ_α, it is easy to show that

$$\begin{aligned}\mathbb{E}\left[\tilde{P}(B)\right] &= \int_0^{+\infty} \mathbb{E}\left[\xi_{\alpha(B)}\,\mathrm{e}^{-u\xi_{\alpha(B)}}\right] \mathbb{E}\left[\mathrm{e}^{-u\xi_{\alpha(B^c)}}\right] \mathrm{d}u \\ &= \int_0^{+\infty} \mathbb{E}\left[-\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{e}^{-u\xi_{\alpha(B)}}\right] \mathbb{E}\left[\mathrm{e}^{-u\xi_{\alpha(B^c)}}\right] \mathrm{d}u.\end{aligned}$$

An application of theorem 16.8 in Billingsley (1995) leads to

$$\begin{aligned}\mathbb{E}\left[\tilde{P}(B)\right] &= \int_0^{+\infty}\left\{-\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}\left[\mathrm{e}^{-u\xi_{\alpha(B)}}\right]\right\}\mathbb{E}\left[\mathrm{e}^{-u\,\xi_{\alpha(B^c)}}\right]\mathrm{d}u\\ &= \alpha(B)\int_0^{+\infty}\mathrm{e}^{-a\,\psi(u)}\left(\frac{\mathrm{d}}{\mathrm{d}u}\psi(u)\right)\mathrm{d}u\\ &= \frac{\alpha(B)}{a}\int_0^{+\infty}\left\{-\frac{\mathrm{d}}{\mathrm{d}u}\,\mathrm{e}^{-a\,\psi(u)}\right\}\mathrm{d}u = \frac{\alpha(B)}{a}\left(1-\mathrm{e}^{-a\,\nu(\mathrm{R}^+)}\right)\end{aligned}$$

and (2) follows as $a>0$ and $\nu(\mathbb{R}^+)=+\infty$.

Now we move on in analysing various properties of normalized HRMIs summarized by suitable descriptive indices. These quantities, once the prior guess at the shape of $\tilde{P}$ is set, are usually fixed in such a way to incorporate additional qualitative prior knowledge into the model. For instance, Walker & Damien (1998) suggest to control the variance of $\tilde{P}$, while Dickey & Jiang (1998) are more concerned with the correlation structure. Refer to Walker *et al.* (1999) for a comprehensive discussion of prior specification in a non-parametric setting. Before stating the results, which are expressed in terms of the prior guess P_0, it is useful to introduce some quantities which will be frequently used throughout. For any $\rho>0$, set

$$V_\rho^{(n)}(u) := \left\{(-1)^n\frac{\mathrm{d}^n}{\mathrm{d}u^n}\,\mathrm{e}^{-\rho\,\psi(u)}\right\}\mathrm{e}^{\rho\,\psi(u)} \quad n=1,2,\ldots, \tag{3}$$

$V_\rho^{(0)}(u)\equiv 1$ and, after defining

$$\mu_n(u) := \int_{\mathbb{R}^+} v^n\,\mathrm{e}^{-uv}\,\nu(\mathrm{d}v) = (-1)^{n-1}\frac{\mathrm{d}^n}{\mathrm{d}u^n}\,\psi(u) \quad \forall u>0,$$

let

$$\Delta_\rho^{(n)}(u) := \sum_{i=0}^{n-1}\binom{n-1}{i}\mu_{n-i}(u)\,V_\rho^{(i)}(u).$$

Finally, we introduce

$$\mathcal{I}_a := a\int_{\mathbb{R}^+} u\,\mathrm{e}^{-a\psi(u)}\,\mu_2(u)\,\mathrm{d}u \quad \text{and} \quad \mathcal{J}_a := a\int_{\mathbb{R}^+} u^2\,\mathrm{e}^{-a\psi(u)}\,\mu_3(u)\,\mathrm{d}u.$$

We are now in a position to state the following proposition.

Proposition 1

Let $B, B_1, B_2\in\mathscr{B}(\mathbb{R})$ and set $C := B_1\cap B_2$. If $\tilde{P}$ is a normalized HRMI, then

$$\mathrm{var}\left[\tilde{P}(B)\right] = P_0(B)(1-P_0(B))\,\mathcal{I}_a \tag{4}$$

$$\mathrm{cov}(\tilde{P}(B_1),\tilde{P}(B_2)) = [P_0(C)-P_0(B_1)\,P_0(B_2)]\,\mathcal{I}_a \tag{5}$$

$$\gamma_1(\tilde{P}(B)) = \frac{1-2\,P_0(B)}{2\sqrt{P_0(B)(1-P_0(B))}}\,\frac{\mathcal{J}_a}{(\mathcal{I}_a)^{3/2}} \tag{6}$$

where γ_1 denotes the skewness coefficient.

At this point, a comparison with the corresponding properties of the Dirichlet process is in order. One easily verifies that for the Dirichlet process $\mathcal{I}_a = 1/(a+1)$ and $\mathcal{J}_a = 4[(a+1)\times(a+2)]^{-1}$. Inserting them in (4)–(6) leads to the desired quantities for the Dirichlet case. It seems worth noting that the particular subordinator, on which $\tilde{P}$ is based, acts on (4)–(6) just as a multiplicative constant (by means of $\mathcal{I}_a$ and $\mathcal{J}_a$) without affecting the deep structure of $\tilde{P}$.

In particular, with reference to the correlation structure, as a straightforward consequence of (5), one has that

$$\mathrm{cov}(\tilde{P}(B_1), \tilde{P}(B_2)) = -P_0(B_1)\, P_0(B_2)\, \mathcal{I}_a$$

whenever B_1 and B_2 are disjoint, thus yielding the following expression for the correlation coefficient

$$\rho(\tilde{P}(B_1), \tilde{P}(B_2)) = -\sqrt{\frac{P_0(B_1)\, P_0(B_2)}{[1 - P_0(B_1)][1 - P_0(B_2)]}}$$

which obviously is the same whatever normalized HRMI has been chosen.

According to the inferential problem at issue, one can alternatively be interested in the expected value of some dispersion indices of the random probability measure. This leads to considering, for instance, random functionals of $\tilde{P}$ such as the variance, $V_{\tilde{P}} = \int x^2 \tilde{P}(\mathrm{d}x) - (\int x \tilde{P}(\mathrm{d}x))^2$, or Gini's mean difference, $\Delta_{\tilde{P}} = \int |x - y|\, \tilde{P}(\mathrm{d}x)\, \tilde{P}(\mathrm{d}y)$. It turns out that also in this case the Dirichlet process does not play a particular role within the class of normalized HRMI. Indeed, provided $\int x^2\, \alpha(\mathrm{d}x) < +\infty$, one has

$$\mathbb{E}\left[V_{\tilde{P}}\right] = (1 - \mathcal{I}_a)\, V_{P_0} \tag{7}$$

$$\mathbb{E}\left[\Delta_{\tilde{P}}\right] = (1 - \mathcal{I}_a)\, \Delta_{P_0} \tag{8}$$

where V_{P_0} and Δ_{P_0} are the variance and Gini's mean difference, respectively, of the probability measure P_0. Details about the determination of (7) and (8) are given in the appendix.

These results allow us to state that some properties, which were thought to be distinctive of the Dirichlet process, are shared by any normalized HRMI. Hence, it seems that such properties crucially depend upon the normalization procedure itself and not upon the particular process to be normalized. In moving on to posterior computations, the real distinctive feature of the Dirichlet process will become apparent. Before proceeding, a small illustrative example might help clarifying ideas.

Example. A remarkable example of Lévy process that can be used for defining a random probability measure is the γ-stable subordinator. Its Lévy measure is given by

$$v(\mathrm{d}v) = v^{-\gamma-1}\, \mathrm{d}v \quad \gamma \in (0, 1).$$

The resulting normalized HRMI was first considered in Kingman (1975) with α equal to the uniform distribution on the unit interval. Note that it also arises as a limiting case of the two parameter Poisson–Dirichlet family due to Pitman (1995). In this case, one has $\mathcal{I}_a = 1 - \gamma$ and $\mathcal{J}_a = (2 - \gamma)(1 - \gamma)$ thus yielding, e.g.

$$\mathrm{var}\left[\tilde{P}(B)\right] = P_0(B)(1 - P_0(B))\,(1 - \gamma)$$

$$\gamma_1(\tilde{P}(B_1)) = \frac{1 - 2\, P_0(B)}{2\sqrt{P_0(B)(1 - P_0(B))}}\ \frac{2 - \gamma}{\sqrt{(1 - \gamma)}}.$$

Analogously, one can easily determine the covariance structure of the prior and expectations of dispersion indices.

3. Characterization of the Dirichlet process

In this section, we look at posterior quantities under the usual assumption of exchangeability of the observations. Suppose that a sequence $(X_n)_{n \geq 1}$ of exchangeable observations is defined on $(\Omega, \mathscr{F}, \mathbb{P})$ in such a way that, given $\tilde{P}$, the X_is are independent and identically distributed (i.i.d.) with distribution $\tilde{P}$. It is well known that the Dirichlet process is conjugate, i.e. given $(X_1, \ldots, X_n)$ its posterior distribution is again a Dirichlet process with parameter measure $\alpha + \Sigma_{i=1}^n \delta_{X_i}$. The main result of this paper is given in the next theorem which provides a characterization of the Dirichlet process in the sense that it identifies conjugacy as the only distinctive feature of the Dirichlet process within the class of normalized HRMI.

Theorem 1

Let $\mathscr{P}$ be the class of normalized HRMI and let $\tilde{P}$ be in $\mathscr{P}$. Then, the posterior distribution of $\tilde{P}$, given a sample $X_1, \ldots, X_n$, is in $\mathscr{P}$ if and only if $\tilde{P}$ is the Dirichlet process.

An important goal in inferential procedures is the prediction of future values of a random quantity based on its past outcomes. One of the reasons of the success of the Dirichlet process is certainly the appealing form of its predictive distributions that are given by

$$\mathbb{P}(X_{n+1} \in B | X_1, \ldots, X_n) = \frac{a}{a+n} \frac{\alpha(B)}{a} + \frac{n}{a+n} \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(B). \tag{9}$$

Although normalized HRMI are not conjugate, it is possible to determine expressions for their predictive distributions. The following proposition provides the predictive distributions corresponding to grouped data, i.e given that, among the past n observations, n_j fall into C_j, for $j = 1, \ldots, k$, with $C_1, \ldots, C_k$ disjoint subsets of $\mathscr{B}(\mathbb{R})$ and $\Sigma_j n_j = n$.

Proposition 2

Let $\tilde{P}$ be a normalized HRMI. The predictive distribution of X_{n+1}, given $(X_1, \ldots, X_n) \in \times_{j=1}^k C_j^{n_j}$, is of the form

$$\begin{aligned}
&\mathbb{P}\left(X_{n+1} \in B \,\middle|\, (X_1, \ldots, X_n) \in \times_{j=1}^k C_j^{n_j} \right) \\
&\quad = \frac{\alpha(B \cap C_{k+1})}{n} \frac{\int_0^{+\infty} u^n \mathrm{e}^{-a\psi(u)} \mu_1(u) \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)}(u)\,\mathrm{d}u}{\int_0^{+\infty} u^{n-1} \mathrm{e}^{-a\psi(u)} \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)}(u)\,\mathrm{d}u} \\
&\quad\quad + \frac{1}{n} \sum_{j=1}^{k} \frac{\alpha(B \cap C_j)}{\alpha(C_j)} \sum_{i=0}^{n_j} \binom{n_j}{i} \\
&\quad\quad \times \frac{\int_0^{+\infty} u^n \mathrm{e}^{-a\psi(u)} \left(\prod_{l \neq j} \Delta_{\alpha_l}^{(n_l)}(u) \right) \Delta_{\alpha(B \cap C_j)}^{(i+1)}(u)\, V_{\alpha(B^c \cap C_j)}^{(n_j - i)}(u)\,\mathrm{d}u}{\int_0^{+\infty} u^{n-1} \mathrm{e}^{-a\psi(u)} \prod_{j=1}^k \Delta_{\alpha_j}^{(n_j)}(u)\,\mathrm{d}u}
\end{aligned} \tag{10}$$

having set $\alpha_i = \alpha(C_i)$ for $i = 1, \ldots, k$.

The predictive distribution for grouped data in the Dirichlet case, first given in Regazzini (1978), can be obtained by setting v as the Lévy measure of the Gamma process in (10). Moreover, from the previous result one can easily deduce the expression for exact data, but unfortunately it remains rather complicated. On the contrary, if one makes the additional assumption of diffuseness of α, it is possible to recover a nice result due to Pitman (2003) (see also Prünster, 2002; James, 2002).

Corollary 1

If $\tilde{P}$ is a normalized HRMI and α is diffuse, then the predictive distributions are of the form

$$\begin{aligned}
&\mathbb{P}(X_{n+1}\in B \mid X_1,\ldots,X_n)\\
&\quad=\frac{a\int_{\mathbb{R}^+}u^n\mathrm{e}^{-a\psi(u)}\mu_{n_1}(u)\ldots\mu_{n_k}(u)\,\mu_1(u)\,\mathrm{d}u}{n\int_{\mathbb{R}^+}u^{n-1}\mathrm{e}^{-a\psi(u)}\mu_{n_1}(u)\ldots\mu_{n_k}(u)\,\mathrm{d}u}\,\frac{\alpha(B)}{a}\\
&\qquad+\frac{1}{n}\sum_{j=1}^{k}\frac{\int_{\mathbb{R}^+}u^n\mathrm{e}^{-a\psi(u)}\mu_{n_1}(u)\ldots\mu_{n_j+1}(u)\ldots\mu_{n_k}(u)\,\mathrm{d}u}{\int_{\mathbb{R}^+}u^{n-1}\mathrm{e}^{-a\psi(u)}\mu_{n_1}(u)\ldots\mu_{n_k}(u)\,\mathrm{d}u}\,\delta_{X_j^*}(B),
\end{aligned}$$

having denoted by $X_1^*,\ldots,X_k^*$ the k distinct observations within the sample, $n_j>0$ terms being equal to X_j^*, for $j=1,\ldots,k$.

Thus, in the case of diffuse α, the predictive distributions have quite intuitive forms, as they consist of a linear combination of the prior guess and of a weighted version of the empirical distribution.

Example (continued). By proposition 2, it is clear that the normalized stable HRMI is not conjugate. As far as the structure of predictive distributions is concerned, one easily finds that

$$\Delta_\rho^{(n)}(u)=\sum_{j=0}^{n-1}w_j(u;\rho)$$

where $w_0(u;\rho)=\rho\,\mu_n(u)$ and, for $j=1,\ldots,n-1$,

$$w_j(u;\rho)=\rho^{j+1}\sum_{(i_1,\ldots,i_j)\in\Lambda_j}\zeta(i_1,\ldots,i_j;n)\,\mu_{n-i_1}(u)\ldots\,\mu_{i_{j-1}-i_j}(u)\mu_{i_j}(u)$$

having set $\Lambda_j=\{(i_1,\ldots,i_j): 1\le i_j<i_{j-1}<\cdots<i_1\le n-1\}$, and $\zeta(i_1,\ldots,i_j;n):=\binom{n-1}{i_1}\binom{i_1-1}{i_2}\ldots\binom{i_{j-1}-1}{i_j}$. An explicit expression for $\mu_m(u)$, with $m\ge1$, is available, so that one easily gets $\mu_{n-i_1}(u)\ldots\mu_{i_{j-1}-i_j}(u)\mu_{i_j}(u)=u^{-n+(j+1)\gamma}\,\Gamma(n-i_1-\gamma)\,\ldots\,\Gamma(i_{j-1}-i_j-\gamma)\,\Gamma(i_j-\gamma)$, for $j=1,\ldots,n-1$, and $\mu_n(u)=u^{-n+\gamma}\Gamma(n-\gamma)$ if $j=0$.

In the case of α being non-atomic with exact observations, the predictive distribution reduces to the simple expression

$$\mathbb{P}(X_{n+1}\in B\mid X_1,\ldots,X_n)=\frac{k}{n}\gamma\,\frac{\alpha(B)}{a}+\frac{1}{n}\sum_{j=1}^{k}(n_j-\gamma)\delta_{X_j^*}(B) \tag{11}$$

obtained first by Pitman (1995) in considering the two parameter Poisson–Dirichlet family.

4. Concluding remarks

In this study, we have investigated the role played by the Dirichlet process within the class of normalized HRMIs and identified conjugacy as its distinctive feature. Although not being conjugate, normalized HRMIs address some interesting issues: as they allow for a relatively simple moment structure, they may represent a useful alternative to the Dirichlet process within the context of hierarchical mixture modelling. To this end it seems necessary to find special normalized HRMIs for which the relevant quantities can be derived in explicit form and, moreover, to adapt simulation techniques such as those proposed in Ishwaran & James (2001).

Acknowledgements

A. Lijoi and I. Prünster were partially supported by the Italian Ministry of University and Research (MIUR), research project 'Bayesian nonparametric methods and their applications'. The authors are grateful to an Associate Editor for her/his comments and suggestions which led to a substantial improvement of this paper. The authors also thank Eugenio Regazzini for some helpful discussions.

References

Barndorff-Nielsen, O. & Shephard, N. (2006). *Financial volatility in continuous time*. Cambridge University Press, Cambridge, to appear.

Billingsley, P. (1995). *Probability and measure*, 3rd edn. Wiley Series of Probability Statistics. John Wiley & Sons, New York, NY.

Dickey, J. M. & Jiang, T. J. (1998). Filtered-variate prior distributions for histogram smoothing. *J. Amer. Statist. Assoc.* **93**, 651–662.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2**, 183–201.

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

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

Ferguson, T. S. & Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist.* **7**, 163–186.

Ishwaran, H. & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173.

James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. *Technical Report Mathematics ArXiv*. Available at http://arxiv.org/abs/math.PR/0205093

James, L. F. (2003). A simple proof of the almost sure discreteness of a class of random measures. *Statist. Probab. Lett.* **65**, 363–368.

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

Lavine, M. (1992). Some aspects of Pólya tree distributions for statistical modelling. *Ann. Statist.* **20**, 1222–1235.

Lo, A. Y. (1991). A characterization of the Dirichlet process. *Statist. Probab. Lett.* **12**, 185–187.

Mauldin, R. D., Sudderth, W. D. & Williams, S. C. (1992). Pólya trees and random distributions. *Ann. Statist.* **20**, 1203–1221.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theor. Rel. Fields* **102**, 145–158.

Pitman, J. (2003). Poisson-Kingman partitions. In *Science and statistics: a Festschrift for terry speed*. IMS Lecture Notes Monogr. Ser. (ed. D. R. Goldstein), Vol. 40, 1–34. *Instit. Math. Statist.*, Beachwood.

Prünster, I. (2002). *Random probability measures derived from increasing additive processes and their application to Bayesian statistics*. PhD Thesis, University of Pavia, Pavia.

Regazzini, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *G. Ist. Ital. degli Attuari* **41**, 77–89.

Regazzini, E. (1996). Impostazione non parametrica di problemi d'inferenza statistica bayesiana. *Technical Report IAMI–CNR*, 96.21. Available at http://www.mi.imati.cnr.it/iami/abstracts/96-21.html

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

Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge Studies Advanced Mathematics Vol. 68. Cambridge University Press, Cambridge.

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

Walker, S. G., Damien, P., Laud, P. W. & Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions. With discussion and a reply by the authors. *J. R. Statist. Soc. Ser. B* **61**, 485–527.

Walker, S. & Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.* **25**, 1762–1780.

Received November 2003, in final form September 2005

Igor Prünster, Dipartimento di Statistica e Matematica Applicata, Università degli Studi di Torino, Piazza Arbarello 8, 10122 Torino, Italy.
E-mail: igor@econ.unito.it

Appendix A

Moment calculations

We begin this appendix with a couple of technical points that play a crucial role in proving propositions 1 and 2. First, we deal with the computation of the following moments

$$\mathbb{E}\left[\tilde{P}(B_1)^{n_1}\dots\tilde{P}(B_k)^{n_k}\right]$$

where $\{B_1,\dots,B_k\}$ is a family of pairwise disjoint sets in $\mathscr{B}(\mathbb{R})$ and the integers n_j are such that $\sum_{j=1}^k n_j=n$. To do so, exploit the definition of $\tilde{P}$ in (1) and notice that

$$\begin{aligned}\mathbb{E}\left[\tilde{P}(B_1)^{n_1}\dots\tilde{P}(B_k)^{n_k}\right]&=\frac{1}{\Gamma(n)}\int_0^{+\infty}u^{n-1}\,\mathbb{E}\left[\mathrm{e}^{-u\xi_a}\prod_{j=1}^k\xi_{\alpha(B_j)}^{n_j}\right]\mathrm{d}u\\&=\frac{1}{\Gamma(n)}\int_0^{+\infty}u^{n-1}\prod_{j=1}^{k+1}\mathbb{E}\left[\mathrm{e}^{-u\xi_{\alpha(B_j)}}\,\xi_{\alpha(B_j)}^{n_j}\right]\mathrm{d}u\\&=\frac{1}{\Gamma(n)}\int_0^{+\infty}u^{n-1}\left\{\prod_{j=1}^{k+1}(-1)^{n_j}\frac{\mathrm{d}^{n_j}}{\mathrm{d}u^{n_j}}\mathrm{e}^{-\alpha(B_j)\psi(u)}\right\}\mathrm{d}u\end{aligned}$$

where we set $B_{k+1}=(\cup_{j=1}^k B_j)^c$ and $n_{k+1}=0$. By virtue of the definition of $V_\rho^{(n)}$ as provided in (3) we then have

$$\mathbb{E}\left[\tilde{P}(B_1)^{n_1}\dots\tilde{P}(B_k)^{n_k}\right]=\frac{1}{\Gamma(n)}\int_0^{+\infty}u^{n-1}\,\mathrm{e}^{-a\psi(u)}\left\{\prod_{j=1}^k V_{\alpha(B_j)}^{(n_j)}(u)\right\}\mathrm{d}u. \tag{12}$$

In order to determine the quantities $V_{\alpha(B_j)}^{(n_j)}$ it is possible to resort to the following recursion formula

$$V_\rho^{(n)}(u)=\rho\sum_{i=0}^{n-1}\binom{n-1}{i}\mu_{n-i}(u)V_\rho^{(i)}(u)=\rho\Delta_\rho^{(n)}(u) \tag{13}$$

which holds true for any $\rho>0$ and can be proved by induction.

The second issue we face here is proving that

$$\lim_{u\to 0}u^n\,\mu_n(u)=0. \tag{14}$$

Note that $\int_{(0,1]}v^n\,\mathrm{e}^{-uv}\nu(\mathrm{d}v)\le\int_{(0,1]}v\,\nu(\mathrm{d}v)<+\infty$, so we have to care only about the integral on $(1,+\infty)$. Let $W(v):=\nu((1,v])$, for any $v>1$ and integrate by parts to obtain

$$\int_{(1,+\infty)}v^n\,\mathrm{e}^{-uv}\,\nu(\mathrm{d}v)=\left[v^n\,\mathrm{e}^{-uv}\,W(v)\right]_{v=1}^{+\infty}-\int_1^{+\infty}(nv^{n-1}\mathrm{e}^{-uv}-uv^n\mathrm{e}^{-uv})\,W(v)\,\mathrm{d}v.$$

As $\lim_{v\to+\infty}W(v)$ is finite, the first summand above vanishes. As far as the second summand is concerned, the change of variable $z=g(v)=uv$ yields

$$\int_1^{+\infty}(nv^{n-1}\mathrm{e}^{-uv}-uv^n\mathrm{e}^{-uv})\,W(v)\,\mathrm{d}v=\frac{1}{u^{n-1}}\int_u^{+\infty}W\left(\frac{z}{u}\right)\left\{nz^{n-1}\mathrm{e}^{-z}-z^n\mathrm{e}^{-z}\right\}\mathrm{d}z.$$

Hence

$$\lim_{u\to 0} u^n \int_{(1,+\infty)} v^n\,\mathrm{e}^{-uv}\,\nu(\mathrm{d}v) = \lim_{u\to 0} u \int_u^{+\infty} W\left(\frac{z}{u}\right)\left\{z^n\mathrm{e}^{-z} - nz^{n-1}\mathrm{e}^{-z}\right\}\mathrm{d}z = 0$$

as the integral is a bounded function of u. This implies the validity of (14).

Proof of proposition 1

Let us start by determining the variance of $\tilde{P}(B)$. To compute the second moment, we can resort to (12), with $k=1$ and $n_1=n=2$, and to the recursion formula in (13) thus getting

$$\begin{aligned}
\mathbb{E}\left[\tilde{P}(B)^2\right] &= \int_0^{+\infty} u\,\mathrm{e}^{-a\psi(u)}\,V^{(2)}_{\alpha(B)}(u)\,\mathrm{d}u \\
&= \alpha(B)\int_0^{+\infty} u\,\mathrm{e}^{-a\psi(u)}\left\{\mu_2(u)+\mu_1(u)\,V^{(1)}_{\alpha(B)}(u)\right\}\mathrm{d}u \\
&= \alpha(B)\int_0^{+\infty} u\,\mathrm{e}^{-a\psi(u)}\left\{\mu_2(u)+\alpha(B)\,\mu_1(u)^2\right\}\mathrm{d}u \\
&= \frac{\alpha(B)}{a}\mathcal{I}_a + \alpha(B)^2\int_0^{+\infty} u\,\mathrm{e}^{-a\psi(u)}\,(\psi'(u))^2\,\mathrm{d}u.
\end{aligned}$$

As for the integral appearing in the second summand above, one can apply integration by parts and obtain

$$\begin{aligned}
&\int_0^{+\infty} u\,\mathrm{e}^{-a\psi(u)}\,(\psi'(u))^2\,\mathrm{d}u \\
&\quad = -\frac{1}{a}\left\{\left[u\,\psi'(u)\,\mathrm{e}^{-a\psi(u)}\right]_0^{+\infty} - \int_0^{+\infty}(\psi'(u)+u\,\psi''(u))\,\mathrm{e}^{-a\psi(u)}\,\mathrm{d}u\right\}.
\end{aligned}$$

It can be immediately checked that $\lim_{u\to+\infty} u\,\psi'(u)\,\mathrm{e}^{-a\psi(u)}=0$. Moreover, by virtue of (14), with $n=1$, one has $\lim_{u\to 0} u\,\psi'(u)\,\mathrm{e}^{-a\psi(u)} = \lim_{u\to 0} u\,\psi'(u) = 0$. Hence

$$\int_0^{+\infty} u\,\mathrm{e}^{-a\psi(u)}\,(\psi'(u))^2\,\mathrm{d}u = \frac{1}{a^2}\left\{1-\mathcal{I}_a\right\}$$

and

$$\mathbb{E}\left[\tilde{P}(B)^2\right] = \frac{\alpha(B)^2}{a^2} + \frac{\alpha(B)(a-\alpha(B))}{a^2}\mathcal{I}_a.$$

At this point, (4) can be easily deduced.

As far as the covariance is concerned, set $B_1^* := B_1 \setminus C$, $B_2^* := B_2 \setminus C$. Then

$$\begin{aligned}
\mathbb{E}\left[\tilde{P}(B_1)\,\tilde{P}(B_2)\right] &= \mathbb{E}\left[\tilde{P}(B_1^*)\tilde{P}(B_2^*)\right] + \mathbb{E}\left[\tilde{P}(C)\tilde{P}(B_2^*)\right] \\
&\quad + \mathbb{E}\left[\tilde{P}(B_1^*)\tilde{P}(C)\right] + \mathbb{E}\left[\tilde{P}^2(C)\right].
\end{aligned}$$

The last summand can be computed as performed above, whereas the first three summands can still be determined by using the general formula (12) for moment measures with $k=2$ and $n_1=n_2=1$. Simple algebra leads to

$$\mathbb{E}\left[\tilde{P}(B_1)\,\tilde{P}(B_2)\right] = \frac{\alpha(B_1)\alpha(B_2)}{a^2} + \frac{a\,\alpha(C)-\alpha(B_1)\alpha(B_2)}{a^2}\mathcal{I}_a \tag{15}$$

and (5) easily follows.

At this point we are left with showing that (6) holds. To this end let us compute the third centred moment of $\tilde{P}(B)$ for any $B \in \mathscr{B}(\mathbb{R})$. Resorting to (2) and (15) one has

$$\mathbb{E}\left[\left(\tilde{P}(B) - \frac{\alpha(B)}{a}\right)^3\right] = \mathbb{E}\left[\tilde{P}(B)^3\right] - 3\,\frac{\alpha^2(B)(a-\alpha(B))}{a^3}\mathcal{I}_a - \frac{\alpha^3(B)}{a^3}. \tag{16}$$

With reference to the computation of $\mathbb{E}[\tilde{P}(B)^3]$ one can resort to (12). Set $k=1$ and $n_1=3$ to obtain

$$\mathbb{E}[\tilde{P}(B)^3] = \frac{1}{\Gamma(3)} \int_0^{+\infty} u^2\,\mathrm{e}^{-a\psi(u)}\,V^{(3)}_{\alpha(B)}(u)\,\mathrm{d}u.$$

Integration by parts, together with the application of (14), with $n=2$, leads to

$$\mathbb{E}[\tilde{P}(B)^3] = \frac{\alpha(B)\,a^2 - 3\alpha^2(B) + 2\alpha^3(B)}{2a^3}\,\mathcal{J}_a + \frac{3\alpha^2(B)a - 3\alpha^3(B)}{a^3}\mathcal{I}_a + \frac{\alpha(B)^3}{a^3}. \tag{17}$$

By inserting (17) into (16), one obtains

$$\mathbb{E}\left[\left(\tilde{P}(B) - \frac{\alpha(B)}{a}\right)^3\right] = \frac{\alpha(B)(a-\alpha(B))(a-2\alpha(B))}{2a^3}\,\mathcal{J}_a. \tag{18}$$

Combining (4) and (18) the result follows.

Proof of proposition 2

Letting $C_1, \ldots, C_k$ denote k disjoint subsets of R, our aim consists first in determining the probability that the $(n+1)$th observation lies in a set B, given that, among the past n observations, n_j fall into C_j, for $j=1, \ldots, k$, i.e.

$$\mathbb{P}(X_{n+1} \in B | (X_1, \ldots, X_n) \in \times_{j=1}^k C_j^{n_j}) = \frac{\mathbb{E}\left[\tilde{P}(B)\,\tilde{P}(C_1)^{n_1} \ldots \tilde{P}(C_k)^{n_k}\right]}{\mathbb{E}\left[\tilde{P}(C_1)^{n_1} \ldots \tilde{P}(C_k)^{n_k}\right]}.$$

As far as the numerator is concerned, set $C_{k+1} = (\cup_{j=1}^k C_j)^c$ and observe that

$$\mathbb{E}\left[\tilde{P}(B)\,\tilde{P}(C_1)^{n_1} \ldots \tilde{P}(C_k)^{n_k}\right] = \sum_{j=1}^{k+1} \mathbb{E}\left[\tilde{P}(B \cap C_j) \prod_{i=1}^k \tilde{P}(C_i)^{n_i}\right].$$

Now, one can directly compute $\mathbb{E}\left[\tilde{P}(B \cap C_{k+1}) \prod_{i=1}^k \tilde{P}(C_i)^{n_i}\right]$ via (12). For any other $j=1, \ldots, k$, one has to observe that

$$\mathbb{E}\left[\tilde{P}(B \cap C_j) \prod_{i=1}^k \tilde{P}(C_i)^{n_i}\right]$$
$$= \sum_{r=0}^{n_j} \binom{n_j}{r} \mathbb{E}\left[\tilde{P}(B \cap C_j)^{r+1}\,\tilde{P}(B^c \cap C_j)^{n_j - r} \prod_{i \neq j} \tilde{P}(C_i)^{n_i}\right].$$

Hence, by virtue of (12),

$$\mathbb{E}\left[\tilde{P}(B \cap C_j)^{r+1}\,\tilde{P}(B^c \cap C_j)^{n_j - r} \prod_{i \neq j} \tilde{P}(C_i)^{n_i}\right]$$
$$= \frac{\alpha(B \cap C_j)\,\alpha(B^c \cap C_j)\,\prod_{i \neq j} \alpha(C_j)}{\Gamma(n+1)}$$
$$\times \int_0^{+\infty} u^n\,\mathrm{e}^{-a\psi(u)}\,\Delta^{(r+1)}_{\alpha(B\cap C_j)}(u)\Delta^{(n_j - r)}_{\alpha(B^c \cap C_j)}(u) \prod_{i \neq j} \Delta^{(n_i)}_{\alpha(C_i)}(u)\,\mathrm{d}u.$$

If we let, for shortness, $\alpha_i = \alpha(C_i)$, $\alpha'_j = \alpha(B \cap C_j)$ and $\alpha'' = \alpha(B^c \cap C_j)$, we can definitely write

$$\mathbb{E}\left[\tilde{P}(B)\,\tilde{P}(C_1)^{n_1} \dots \tilde{P}(C_k)^{n_k}\right]$$

$$= \frac{1}{\Gamma(n+1)}\left\{\alpha'_{k+1}\left(\prod_{i=1}^{k}\alpha_i\right)\int_0^{+\infty} u^n \mathrm{e}^{-a\psi(u)}\,\mu_1(u)\prod_{j=1}^{k}\Delta_{\alpha_j}^{(n_j)}(u)\,\mathrm{d}u\right.$$

$$\left. + \sum_{j=1}^{k}\left(\prod_{i\neq j}\alpha_i\right)\alpha'_j\alpha''_j\sum_{r=0}^{n_j}\binom{n_j}{r}\int_0^{+\infty} u^n \mathrm{e}^{-a\psi(u)}\left(\prod_{i\neq j}\Delta_{\alpha_i}^{(n_i)}(u)\right)\Delta_{\alpha'_j}^{(r+1)}(u)\,\Delta_{\alpha''_j}^{(n_j-r)}(u)\,\mathrm{d}u\right\}.$$

The denominator can be determined by using formula (12) thus yielding (10).

Proof of corollary 1

Start by setting $C_j = C_{j,\varepsilon} = (X_j^* - \varepsilon,\ X_j^* + \varepsilon)$ in (10). By the hypothesis of diffuseness of α, one has $\Delta_{\alpha_i}^{(n_i)} = \mu_{n_i} + o(\alpha_i)$ as $\varepsilon \downarrow 0$ and $\delta_{X_j^*}(B)\,V_{\alpha(B^c\cap C_j)}^{(n_j-i)} = 0$ $i = 0, 1, \dots, n_j - 1$ so that the claimed results follows by taking the limit as $\varepsilon \downarrow 0$.

Details for the determination of (7) and (8). The main trick in proving these relations consists in taking a suitable discretization of the measure parameter α. More precisely, for any $m \geq 1$ fix a partition of R into intervals, say $B_{m,0}, \dots, B_{m,k_m+1}$, where $B_{m,0} = (-\infty, -R_m]$, $B_{m,k_m+1} = (R_m, +\infty)$ and $\{B_{m,i} : i = 1, \dots, k_m\}$ is a partition of $[-R_m, R_m]$. Moreover, let R_m be such that $\lim_m R_m = +\infty$ and let $\max_{1\leq i\leq k_m}|B_{m,i}| \to 0$ as $m \to +\infty$, where $|B|$ is the length of the interval B. The discretized measure is defined as

$$\alpha_m = \sum_{i=0}^{k_m+1}\alpha(B_{m,i})\delta_{y_{m,i}}$$

where $y_{m,i}$ is any inner point in $B_{m,i}$ for $i = 0, \dots, k_m + 1$. This implies that the normalized HRMI with parameter measure α_m and Lévy measure v is given by

$$\tilde{P}_m = \sum_{i=0}^{k_m+1}\tilde{P}(B_{m,i})\,\delta_{y_{m,i}}.$$

At this stage, computation of

$$\mathbb{E}\left[V_{\tilde{P}_m}\right] = \sum_{i=0}^{k_m+1} y_{m,i}^2\,\mathbb{E}\left[\tilde{P}(B_{m,i})\right] - \mathbb{E}\left[\sum_{i=0}^{k_m+1} y_{m,i}\,\tilde{P}(B_{m,i})\right]^2$$

is straightforward, yielding

$$\mathbb{E}\left[V_{\tilde{P}_m}\right] = \sum_{i=0}^{k_m+1} y_{m,i}^2\,\frac{\alpha(B_{m,i})}{a} - (1-\mathcal{I}_a)\left\{\sum_{i=0}^{k_m+1} y_{m,i}\,\frac{\alpha(B_{m,i})}{a}\right\}^2 - \mathcal{I}_a\sum_{i=0}^{k_m+1} y_{m,i}^2\,\frac{\alpha(B_{m,i})}{a}.$$

Taking the limit, as $m \to +\infty$, completes the proof. One proceeds in a similar fashion in order to compute $\mathbb{E}\left[\Delta_{\tilde{P}}\right]$

Appendix B

Characterization of the Dirichlet process

We first prove a lemma providing the one-step predictive distribution.

Lemma 1

If $\tilde{P}$ is a normalized HRMI and the random variables X_1, X_2 are, conditional on $\tilde{P}$, i.i.d. with distribution $\tilde{P}$, then

$$\mathbb{P}(X_2 \in B \mid X_1) = \frac{\alpha(B)}{a}[1 - \mathcal{I}_a] + \delta_{X_1}(B)\mathcal{I}_a \quad \textit{for every } B \in \mathscr{B}(\mathbb{R}).$$

Proof of lemma 1

Suppose $X_1 = x_1$ and set $C_\varepsilon := (x_1 - \varepsilon, x_1 - \varepsilon)$ with $\varepsilon > 0$. Note that, if D denotes the support of α, then $\mathbb{P}(X_1 \in D) = \mathbb{E}[\tilde{P}(D)] = 1$. For this reason we can assume that x_1 lies in D. Hence, the infinite activity condition $v(\mathbb{R}^+) = +\infty$ ensures $\tilde{P}(C_\varepsilon) > 0$ a.s.–$\mathbb{P}$. To prove the result, it is enough to show that

$$\mathbb{P}(X_2 \in B \mid X_1 \in C_\varepsilon) \to \frac{\alpha(B)}{a}(1 - \mathcal{I}_a) + \mathcal{I}_a\, \delta_{x_1}(B) \quad \text{as}\, \varepsilon \to 0$$

for any set $B = (-\infty, y]$, with $y \in \mathbb{R}$. To this end, choose $\varepsilon > 0$ in such a way that either $C_\varepsilon \subset B$ or $C_\varepsilon \cap B$ is empty. As

$$\mathbb{P}(X_2 \in B \mid X_1 \in C_\varepsilon) = \frac{\mathbb{E}\left[\tilde{P}(B)\tilde{P}(C_\varepsilon)\right]}{\mathbb{E}\left[\tilde{P}(C_\varepsilon)\right]},$$

combination of (2) with (5) provides the desired result.

Before proving theorem 1, we set some notation and recall some basic facts. Let $m_n(\cdot) = \mathbb{E}[\tilde{P}^n(\cdot)]$ be the (marginal) distribution of the random vector $X^{(n)} = (X_1, \ldots, X_n)$. Introduce a family $\{P_{x^{(n)}} : x^{(n)} \in \mathbb{R}^n, n \geq 1\}$ of probability measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $x^{(n)} \mapsto P_{x^{(n)}}(B)$ is $\mathscr{B}(\mathbb{R}^n)$-measurable, for any B in $\mathscr{B}(\mathbb{R})$ and

$$P_{X^{(n)}}(B) = P\left[X_{n+1} \in B \,\middle|\, X^{(n)}\right] \quad \text{a.s.} - \mathbb{P}.$$

Hence, $P_{X^{(n)}}$ is a version of the conditional distribution of X_{n+1}, given $X^{(n)}$ and $P_{x^{(n)}}$ can be interpreted as the predictive distribution of X_{n+1} given that we have observed $X^{(n)} = x^{(n)}$. We will also write $P_{x^{(n)}}(B) = P[X_{n+1} \in B | X^{(n)} = x^{(n)}]$. Moreover, denote by $\tilde{P}_{X^{(n)}}$ a random probability measure whose distribution coincides with the posterior distribution of $\tilde{P}$, given $X^{(n)}$. With these definitions one can show that there exists a set $\mathcal{N} \in \mathscr{B}(\mathbb{R}^n)$ such that $m_n(\mathcal{N}^c) = 0$ and for any vector $x^{(n)} \in \mathcal{N}$

$$\mathbb{E}\left[\tilde{P}(B) \,\middle|\, X^{(n)} = x^{(n)}\right] = \mathbb{E}\left[\tilde{P}_{x^{(n)}}(B)\right] = P_{x^{(n)}}(B) \quad \forall B \in \mathscr{B}(\mathbb{R}).$$

For simplicity, and with no loss of generality, in the following we will suppose that $x^{(n)} \in \mathcal{N}$.

Let us also recall two properties shared by the predictive distributions $P_{x^{(n)}}$. First note that $\tilde{P}_{x^{(n)}}$ can be obtained from conditioning $\tilde{P}_{x^{(n-1)}}$ to $X_n = x_n$. For a rigorous proof of this fact see, e.g. Regazzini (1996). As a consequence, one has that for any measurable set B

$$P_{x^{(n)}}(B) = \mathbb{E}\left[\tilde{P}_{x^{(n-1)}}(B) \,\middle|\, X_n = x_n\right], \tag{19}$$

i.e. the n-step prediction on the left-hand side can be recovered as a one-step prediction based on $\tilde{P}_{x^{(n-1)}}$. Secondly, because of exchangeability, the mapping

$$x^{(n)} \mapsto P_{x^{(n)}}(B)$$

is invariant with respect to permutations of the components of the vector $x^{(n)}=(x_1,\dots,x_n)$. For instance, if $n\geq 2$ and $y^{(n)}=(x_1,\dots,x_{n-2},x_n,x_{n-1})$ then $\mathbb{E}\left[\tilde P_{y^{(n)}}(B)\right]$ can be understood as

$$\mathbb{E}\left[\tilde P(B)\,\middle|\, X^{(n-2)}=x^{(n-2)}, X_{n-1}=x_n, X_n=x_{n-1}\right]$$

and the above property entails

$$\mathbb{E}\left[\tilde P_{x^{(n)}}(B)\right]=\mathbb{E}\left[\tilde P_{y^{(n)}}(B)\right]\quad \forall B\in\mathscr{B}(\mathbb{R}). \tag{20}$$

We now proceed with proving theorem 1.

Proof of theorem 1

The 'if' part is trivial. Let us focus, then, on the 'only if' part. Assume that, for any $n\geq 1$, the law of $\tilde P_{x^{(n)}}$ coincides with the law of some normalized HRMI. In other words there exists a subordinator, $\xi^{(n)}=\{\xi_t^{(n)}: t\geq 0\}$ say, with Lévy measure v_n such that the posterior distribution of $\tilde P$ equals the distribution of the random probability measure $\tilde P_{x^{(n)}}$ defined by

$$\tilde P_{x^{(n)}}((-\infty,y])=\frac{\xi_{A_n(y)}}{\xi_{a_n}}\quad \forall\, y\in\mathbb{R} \tag{21}$$

for some finite measure α_n, where $A_n(y)=\alpha_n((-\infty,y])$ and $a_n=\alpha_n(\mathbb{R})$. A simple application of lemma 1 to (19) yields

$$\mathbb{E}\left[\tilde P_{x^{(n)}}(B)\right]=P_{x^{(n)}}(B)=\frac{\alpha_{n-1}(B)}{a_{n-1}}\left[1-\mathcal{I}(x^{(n-1)})\right]+\delta_{x_n}(B)\mathcal{I}(x^{(n-1)})$$

where for any $n\geq 2$

$$\mathcal{I}(x^{(n-1)}):=a_{n-1}\int_0^{+\infty}u\,\mathrm{e}^{-a_{n-1}\psi_{n-1}(u)}\,\mu_{2,\,n-1}(u)\,\mathrm{d}u,$$

$$\psi_{n-1}(u)=\int_{\mathbb{R}^+}(1-\mathrm{e}^{-us})\,v_{n-1}(\mathrm{d}s),\qquad \mu_{2,\,n-1}(u)=\int_{\mathbb{R}^+}s^2\,\mathrm{e}^{-us}\,v_{n-1}(\mathrm{d}s).$$

Let us also agree that $\mathcal{I}(x^{(0)}):=\mathcal{I}_a$. As $\alpha_{n-1}(\cdot)/a_{n-1}=P_{x_{n-1}}(\cdot)$, combination of (19) and of lemma 1 yields

$$\frac{\alpha_{n-1}(B)}{a_{n-1}}=\frac{\alpha(B)}{a}\prod_{j=0}^{n-2}\left[1-\mathcal{I}(x^{(j)})\right]+\sum_{k=1}^{n-1}\delta_{x_k}(B)\mathcal{I}(x^{(k-1)})\prod_{j=k}^{n-2}\left[1-\mathcal{I}(x^{(j)})\right] \tag{22}$$

with the proviso that $\prod_{j=n-1}^{n-2}\left[1-\mathcal{I}(x^{(j)})\right]\equiv 1$. We now claim that, for any $n\geq 2$, one has

$$\mathcal{I}(x^{(n-1)})=\frac{\mathcal{I}_a}{1+(n-1)\mathcal{I}_a}. \tag{23}$$

By virtue of (22)

$$\mathbb{E}\left[\tilde P_{x^{(n)}}(B)\right]=\frac{\alpha(B)}{a}\prod_{j=0}^{n-1}\left[1-\mathcal{I}(x^{(j)})\right]+\sum_{k=1}^{n}\delta_{x_k}(B)\mathcal{I}(x^{(k-1)})\prod_{j=k}^{n-1}\left[1-\mathcal{I}(x^{(j)})\right],$$

so that, if $\mathcal{I}(\cdot)$ is replaced by the expression in (23), one has

$$\mathbb{E}\left[\tilde P_{x^{(n)}}(B)\right]=\frac{1}{1+(n-1)\mathcal{I}_a}\left\{\frac{\alpha(B)}{a}(1-\mathcal{I}_a)+\mathcal{I}_a\sum_{i=1}^{n}\delta_{x_i}(B)\right\}.$$

It can be seen that $\mathcal{I}_a$ is a constant in $[0,1]$, so that there exist a constant $c\geq 0$ such that $\mathcal{I}_a=\frac{1}{c+1}$. One, then gets to the conclusion that $\mathbb{E}[\tilde P_{x^{(n)}}(B)]=P_{x^{(n)}}(B)$ has the form (9). As it has been shown in Regazzini (1978) and Lo (1991), (9) also characterizes the Dirichlet process. Thus $c=a$.

In order to complete the proof, we need to show that (23) holds true. To this end, we proceed by induction. Let us first show that (23) is true for $n=2$. By virtue of the assumption in theorem 1, both $\tilde{P}_{x_1}$ and $\tilde{P}_{x_2}$ are normalized HRMIs with parameters (α_{x_1}, v_1) and (α_{x_2}, v_2) respectively. This fact, together with lemma 1, implies that

$$\mathbb{E}\left[\tilde{P}_{x_j}(B)\right] = \frac{\alpha_{x_j}(B)}{a_{x_j}} = \frac{\alpha(B)}{a}(1-\mathcal{I}_a) + \delta_{x_j}(B)\mathcal{I}_a \quad j=1,2.$$

Next, resort to (20), with $x^{(2)}=(x_1, x_2)$ and $y^{(2)}=(x_2, x_1)$, and to (19) to state the validity of the following equalities

$$\begin{aligned}
\mathbb{E}\left[\tilde{P}_{x^{(2)}}(B)\right] &= \frac{\alpha_{x_1}(B)}{a_{x_1}}(1-\mathcal{I}(x_1)) + \delta_{x_2}(B)\mathcal{I}(x_1) \\
&= \frac{\alpha(B)}{a}(1-\mathcal{I}_a)(1-\mathcal{I}(x_1)) + \delta_{x_1}(B)\mathcal{I}_a(1-\mathcal{I}(x_1)) + \delta_{x_2}(B)\mathcal{I}(x_1) \\
&= \mathbb{E}\left[\tilde{P}_{y^{(2)}}(B)\right] = \frac{\alpha_{x_2}(B)}{a_{x_2}}(1-\mathcal{I}(x_2)) + \delta_{x_1}(B)\mathcal{I}(x_2) \\
&= \frac{\alpha(B)}{a}(1-\mathcal{I}_a)(1-\mathcal{I}(x_2)) + \delta_{x_1}(B)\mathcal{I}(x_2) + \delta_{x_2}(B)\mathcal{I}_a(1-\mathcal{I}(x_2)).
\end{aligned}$$

Here, $\mathcal{I}(x_2)$ is defined in a similar fashion as $\mathcal{I}(x_1)$ and it refers to the normalized HRMI $\tilde{P}_{x_2}$. As the previous equalities do hold for any set $B \in \mathscr{B}(\mathbb{R})$, it must be

$$\mathcal{I}(x_1) = \mathcal{I}(x_2) = \frac{\mathcal{I}_a}{1+\mathcal{I}_a},$$

which proves (23) for $n=2$. Suppose, now that $\mathcal{I}(x^{(j)})$ is of the form (23) for each $j=1, \dots, n-2$, with $n>2$. Let

$$\mathcal{I}(x^{(n-2)}, x_n) := a^* \int_0^{+\infty} u\, \mathrm{e}^{-a^*\psi^*(u)} \left(\int_{\mathbb{R}^+} v^2 \mathrm{e}^{-uv} v^*(\mathrm{d}v)\right)\mathrm{d}u$$

where $a^* = \alpha^*(\mathbb{R})$ and α^* and v^* are the parameter and the Lévy measure, respectively, associated to the normalized HRMI $\tilde{P}_{x^{(n-2)}, x_n}$. Thus, lemma 1 combined with (19) provides

$$\mathbb{E}\left[\tilde{P}_{y^{(n)}}(B)\right] = \frac{\alpha^*(B)}{a^*}\left[1-\mathcal{I}(x^{(n-2)}, x_n)\right] + \delta_{x_{n-1}}(B)\mathcal{I}(x^{(n-2)}, x_n).$$

As $P_{x_{n-2}}$ is an HRMI with parameters (α_{n-2}, v_{n-2}), one has

$$\begin{aligned}
\frac{\alpha^*(B)}{a^*} &= \mathbb{E}\left[\tilde{P}_{x^{(n-2)},\, x_n}(B)\right] = \frac{\alpha_{n-2}(B)}{a_{n-2}}\left[1-\mathcal{I}(x^{(n-2)})\right] + \delta_{x_n}(B)\mathcal{I}(x^{(n-2)}) \\
&= \left\{\frac{\alpha(B)}{a}\prod_{j=1}^{n-2}\left[1-\mathcal{I}(x^{(j-1)})\right] + \sum_{k=1}^{n-2}\delta_{x_k}(B)\mathcal{I}(x^{(k-1)})\prod_{j=k}^{n-3}\left[1-\mathcal{I}(x^{(j)})\right]\right\} \\
&\quad \times \left[1-\mathcal{I}(x^{(n-2)})\right] + \delta_{x_n}(B)\mathcal{I}(x^{(n-2)}) \\
&= \frac{\alpha(B)}{a}\,\frac{1-\mathcal{I}_a}{1+(n-2)\mathcal{I}_a} + \left(\sum_{k=1}^{n-2}\delta_{x_k}(B) + \delta_{x_n}(B)\right)\frac{\mathcal{I}_a}{1+(n-2)\mathcal{I}_a}
\end{aligned}$$

where the last two equalities have been deduced by virtue of (22) and by the hypothesis on $\mathcal{I}(x^{(j)})$, for $j=1,\dots,n-2$. Finally

$$\mathbb{E}\left[\tilde{P}_{y^{(n)}}(B)\right]=\frac{\alpha(B)}{a}\,\frac{(1-\mathcal{I}_a)[1-\mathcal{I}(x^{(n-2)},x_n)]}{1+(n-2)\mathcal{I}_a}$$
$$+\left(\sum_{k=1}^{n-2}\delta_{x_k}(B)+\delta_{x_n}(B)\right)\frac{\mathcal{I}_a[1-\mathcal{I}(x^{(n-2)},x_n)]}{1+(n-2)\mathcal{I}_a}$$
$$+\delta_{x_{n-1}}(B)\mathcal{I}(x^{(n-2)},x_n).$$

In a similar fashion one gets

$$\mathbb{E}\left[\tilde{P}_{x^{(n)}}(B)\right]=\frac{\alpha(B)}{a}\,\frac{(1-\mathcal{I}_a)[1-\mathcal{I}(x^{(n-1)})]}{1+(n-2)\mathcal{I}_a}$$
$$+\left(\sum_{k=1}^{n-1}\delta_{x_k}(B)\right)\frac{\mathcal{I}_a[1-\mathcal{I}(x^{(n-1)})]}{1+(n-2)\mathcal{I}_a}+\delta_{x_n}(B)\mathcal{I}(x^{(n-1)}).$$

As (19) holds true, then $\mathcal{I}(x^{(n-2)},x_n)=\mathcal{I}(x^{(n-1)})$ and

$$\mathcal{I}(x^{(n-1)})=\frac{\mathcal{I}_a[1-\mathcal{I}(x^{(n-1)})]}{1+(n-2)\mathcal{I}_a}.$$

This proves that (23) is valid for any $n\geq 2$.