

A Class of Normalized Random Measures with an Exact Predictive Sampling Scheme

LORENZO TRIPPA

Dana-Faber Cancer Institute and Harvard School of Public Health, Harvard University

STEFANO FAVARO

Dipartimento di Statistica e Matematica Applicata, Università di Torino and Collegio Carlo Alberto

ABSTRACT. In this article, we define and investigate a novel class of non-parametric prior distributions, termed the class \mathcal{C} . Such class of priors is dense with respect to the homogeneous normalized random measures with independent increments and it is characterized by a richer predictive structure than those arising from other widely used priors. Our interest in the class \mathcal{C} is mainly motivated by Bayesian non-parametric analysis of some species sampling problems concerning the evaluation of the species relative abundances in a population. We study both the probability distribution of the number of species present in a sample and the probability of discovering a new species conditionally on an observed sample. Finally, by using the coupling from the past method, we provide an exact sampling scheme for the system of predictive distributions characterizing the class \mathcal{C} .

Key words: Bayesian non-parametrics, completely random measures, coupling from the past method, Dirichlet process, Gibbs-type random probability measures, normalized random measures with independent increments, predictive distributions, species sampling problems

1. Introduction

Since the introduction of the Dirichlet process by Ferguson (1973) various approaches for constructing random probability measures, whose distributions act as non-parametric priors, have been proposed in the literature. This article deals with random probability measures obtained by means of the so-called ‘normalization’ approach. In Bayesian non-parametrics the ‘normalization’ approach was first rigorously introduced in Regazzini *et al.* (2003) where a rich class of non-parametric priors, namely normalized random measures with independent increments (NRMIs), is defined according to the normalization of suitably time-changed increasing processes with independent increments. Structural properties of the class of NRMIs have been further investigated in a series of subsequent papers by Nieto-Barajas *et al.* (2004), James *et al.* (2006) and James *et al.* (2009). In particular, James *et al.* (2009) stated a slightly more general definition of NRMIs in terms of the normalization of completely random measures (CRMs), a notion introduced by Kingman (1967). See Lijoi & Prünster (2010) for a comprehensive and stimulating review on this topic.

The idea of normalizing CRMs to define non-parametric priors is clearly inspired by the definition of the Dirichlet process in terms of the normalization of a Gamma CRM. Let α be a non-negative finite measure on $(\mathbb{X}, \mathcal{X})$ with \mathbb{X} being a complete and separable metric space equipped with the Borel σ -algebra \mathcal{X} . A Dirichlet process on \mathbb{X} with parameter α is obtained by normalizing a Gamma CRM on \mathbb{X} with parameter α , i.e. a CRM characterized by a Lévy intensity measure ν of the form $\nu(ds, dx) = s^{-1} e^{-s} ds \alpha(dx)$. More generally, an NRMI is defined as a random probability measure obtained when the Gamma CRM is replaced by any CRM with finite total mass and characterized by some Lévy intensity measure

$v(ds, dx) = \rho(ds|x)\alpha(dx)$. If v is of the form $v(ds, dx) = \rho(ds)\alpha(dx)$ then it is termed homogeneous and the corresponding CRM (NRMI) is termed homogeneous CRM (homogeneous NRMI). Several examples of homogeneous NRMI have been introduced in the recent literature. The most notable in terms of a sufficient mathematical tractability is the so-called normalized generalized Gamma process (Lijoi *et al.*, 2007a) which includes as special cases the Dirichlet process, the normalized stable process (Kingman, 1975) and the normalized inverse-Gaussian process (Lijoi *et al.*, 2005b).

Our interest in NRMI is mainly motivated by their application as non-parametric priors for Bayesian analysis of species sampling problems. In recent years, there has been an enormous growth in the proposal of Bayesian non-parametric methods for several applied problems; see Müller & Quintana (2004), Müller & Quintana (2010) and Dunson (2008) for interesting reviews. One of such recent application concerns precisely species sampling problems which gained a renewed interest due to their importance in genomics and ecology. In species sampling problems one is interested in determining the species composition of a certain population containing an unknown number of species and only a sample drawn from it is available. Specifically, given a sample (X_1, \dots, X_n) consisting of a collection of $k \leq n$ species (X_1^*, \dots, X_k^*) with frequencies (n_1, \dots, n_k) , interest typically lies in evaluating the species variety of the underlying population by estimating both the number of new species to be observed in an additional sample $(X_{n+1}, \dots, X_{n+m})$ and the probability of discovering a new species at the $(n+m+1)$ th draw. A Bayesian non-parametric approach for evaluating the species variety in a population has been set forth in Lijoi *et al.* (2007b); see also Lijoi *et al.* (2008a) for a theoretical study and Lijoi *et al.* (2007c) for a practitioner oriented illustration. The process $(X_i)_{i \geq 1}$ is modelled as an \mathbb{X} -valued exchangeable sequence and, by de Finetti's theorem, the X_i s can be represented as independent samples from a distribution \tilde{p} on $(\mathbb{X}, \mathcal{X})$ having prior distribution Π :

$$\begin{aligned} X_i | \tilde{p} &\stackrel{\text{i.i.d.}}{\sim} \tilde{p}, \\ \tilde{p} &\sim \Pi. \end{aligned} \tag{1}$$

In Lijoi *et al.* (2007b) Bayesian non-parametric estimators for quantities of interest in species sampling problems have been derived under the hypothesis that \tilde{p} is an almost sure discrete random probability measure on \mathbb{X} characterized by a system of predictive distributions of the type

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, \dots, X_n) = g_0(n, k) \mathbb{E}[\tilde{p}(\cdot)] + g_1(n, k) \sum_{j=1}^k (n_j - \sigma) \delta_{X_j^*}(\cdot), \tag{2}$$

where $\sigma \in [0, 1)$ and the weights g_0 and g_1 are functions satisfying the relation $g_0(n, k) + g_1(n, k)(n - \sigma k) = 1$ for any $n \geq 1$ and $k \leq n$. A random probability measure whose predictive distributions satisfy the representation (2) is termed Gibbs-type random probability measure; see Gnedin & Pitman (2005) and Pitman (2006). Further investigations related to Gibbs-type random probability measures, Bayesian non-parametrics and NRMI can be found in Ho *et al.* (2007) and Lijoi *et al.* (2008b).

According to the predictive distribution in (2), for any Gibbs-type random probability measure the number of species k in the observed sample (X_1, \dots, X_n) turns out to be a sufficient statistic for predicting the number of new species to be observed in an additional sample. In principle one would like priors which lead to richer predictive structures, in which the probability of sampling a new species depends explicitly on both k and (n_1, \dots, n_k) . However, by dropping the Gibbs structure, serious issues of mathematical tractability arise. In this paper we consider random probability measures which are not of Gibbs-type. In particular, we define and investigate a class of homogeneous NRMI characterized by a system of predictive distributions of the type

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, \dots, X_n) = w_0(n, k, (n_1, \dots, n_k))\mathbb{E}[\tilde{p}] + \sum_{j=1}^k w_j(n, k, (n_1, \dots, n_k))\delta_{X_j^*}(\cdot), \quad (3)$$

where the weights $w_0(n, k, (n_1, \dots, n_k))$ and $w_j(n, k, (n_1, \dots, n_k))$ are functions explicitly depending on the number of species k and on the entire vector of frequencies (n_1, \dots, n_k) . We term this novel class of homogeneous NRMI, the class \mathcal{C} . To our knowledge, the only example currently known in the literature of a random probability measure characterized by a system of predictive distributions of the type (3) is the so-called generalized Dirichlet process introduced by Regazzini *et al.* (2003). We refer to Lijoi *et al.* (2005a) and Favaro *et al.* (2010) for further developments of the generalized Dirichlet process with applications to mixture models and species sampling problems.

Besides the richer predictive structure the class \mathcal{C} is characterized by two other appealing properties. Any homogeneous NRMI can be arbitrarily well approximated by an NRMI in \mathcal{C} , i.e. the class \mathcal{C} is dense with respect to the homogeneous NRMI and, for any NRMI in \mathcal{C} , an exact sampling scheme for the predictive distributions can be implemented by exploiting the coupling from the past method in Propp & Wilson (1996). The denseness of the class \mathcal{C} implies, from a Bayesian standpoint, that, for any NRMI prior, one can select a random distribution from \mathcal{C} such that the differences between the two prior models are negligible. Under the model assumption (1) with \tilde{p} being an NRMI in the class \mathcal{C} , we derive explicit expressions for quantities of interest in species sampling problems such as the probability distribution of the number of species in a sample (X_1, \dots, X_n) drawn from \tilde{p} and the probability distribution of the number of new species to be observed in an additional sample $(X_{n+1}, \dots, X_{n+m})$, conditionally on an observed sample. However, it is worth pointing out that the evaluation of such explicit expressions is a difficult task since it requires the evaluation of partial Bell polynomials, which is known to be feasible only for small sample sizes. In this respect, the proposed exact sampling scheme represents a valid alternative to the direct evaluation of the partial Bell polynomials and, consequently, it provides a useful tool for making inference in species sampling problems.

The article is structured as follows. In section 2 we provide a concise account on NRMI. In section 3 we define and investigate the class \mathcal{C} . In particular, we focus on the system of predictive distributions characterizing the class \mathcal{C} and we derive some quantities of interest in species sampling problems. In section 4 we describe the exact sampling scheme for the system of predictive distributions characterizing the class \mathcal{C} . The Appendix contains a short review on Bell polynomials. The proofs of the results stated in section 3 can be found online in the Supporting Information.

2. NRMI

We start with the definition of CRM as introduced by Kingman (1967). Let \mathbb{X} be a complete and separable metric space endowed with the Borel σ -algebra \mathcal{X} . We let $\mathcal{M}_{\mathbb{X}}$ stand for the space of finite measures on $(\mathbb{X}, \mathcal{X})$ endowed with the Borel σ -algebra $\mathcal{M}_{\mathbb{X}}$.

Definition 1. Let $\tilde{\mu}$ be a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathcal{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ such that for any B_1, \dots, B_n in \mathcal{X} , with $B_i \cap B_j = \emptyset$ for any $i \neq j$, the random variables $\tilde{\mu}(B_1), \dots, \tilde{\mu}(B_n)$ are mutually independent. The measurable map $\tilde{\mu}$ is called CRM.

An important property of CRMs is their almost sure discreteness (see Kingman, 1993; James, 2003). A CRM $\tilde{\mu}$ can be represented as the sum of two components: a CRM with non-negative random jumps $(S_i)_{i \geq 1}$ at random \mathbb{X} -valued locations $(\bar{X}_i)_{i \geq 1}$ and a CRM with

non-negative random jumps $(J_i)_{i \geq 1}$ at fixed \mathbb{X} -valued locations $(\bar{x}_i)_{i \geq 1}$. Accordingly, if we set $\tilde{\mu}_c = \sum_{i \geq 1} S_i \delta_{\bar{x}_i}$, then

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i \geq 1} J_i \delta_{\bar{x}_i}.$$

The CRM $\tilde{\mu}_c$ can be characterized by using the so-called Lévy–Khintchine representation; for any non-negative function f

$$\mathbb{E} \left[e^{-\int_{\mathbb{X}} f(x) \tilde{\mu}_c(dx)} \right] = \exp \left\{ - \int_{\mathbb{R}^+ \times \mathbb{X}} (1 - e^{-sf(x)}) \nu(ds, dx) \right\}, \tag{4}$$

where ν is a measure on $\mathbb{R}^+ \times \mathbb{X}$, termed Lévy intensity measure, which satisfies the inequality $\int \min(s, 1) d\nu < +\infty$. It is useful to factorize the Lévy intensity measure ν as $\nu(ds, dx) = \rho(ds | x) \alpha(dx)$, where α is a non-negative finite measure on $(\mathbb{X}, \mathcal{X})$ and ρ is a kernel on $\mathbb{X} \times \mathcal{B}(\mathbb{R}^+)$, i.e. $\rho(\cdot | x)$ is a measure for any x in \mathbb{X} and $x \mapsto \rho(B | x)$ is \mathcal{X} -measurable for any B in $\mathcal{B}(\mathbb{R}^+)$. If $\rho(\cdot | x) = \rho(\cdot)$ then the non-negative random jumps $(S_i)_{i \geq 1}$ are independent with respect to the random locations $(\bar{X}_i)_{i \geq 1}$ and both the Lévy intensity measure ν and the CRM $\tilde{\mu}_c$ are termed homogeneous. As one may observe from (4), CRMs are closely connected to the so-called Poisson random measures. Indeed, the CRM $\tilde{\mu}_c$ can be represented as a functional of a Poisson random measure $\tilde{\Pi}$ on $\mathbb{R}^+ \times \mathbb{X}$ with mean intensity measure ν . It can be shown that

$$\tilde{\mu}_c(B) \stackrel{d}{=} \int_B \int_{\mathbb{R}^+} s \tilde{\Pi}(ds, dx)$$

for any $B \in \mathcal{X}$. See Kingman (1993) for a survey on this topic. We can now state the definition of NRMI.

Definition 2. Let $\tilde{\mu}$ be a CRM such that $0 < \tilde{\mu}(\mathbb{X}) < +\infty$ almost surely. The random probability measure $\tilde{p} = \tilde{\mu} / \tilde{\mu}(\mathbb{X})$ is termed NRMI.

Both finiteness and positiveness of the total mass $\tilde{\mu}(\mathbb{X})$ are necessary for normalizing $\tilde{\mu}$. We can use a Lévy intensity measure ν , such that $\nu(\mathbb{R}^+ \times \mathbb{X}) = +\infty$, for defining a CRM $\tilde{\mu}$, without fixed points of discontinuity, such that $\mathbb{P}(0 < \tilde{\mu}(\mathbb{X}) < +\infty) = 1$. This condition on the Lévy measure ν is equivalent to requiring that the CRM $\tilde{\mu}$ has infinitely many jumps.

Let us consider model (1), with $\tilde{p} = \tilde{\mu} / \mu(\mathbb{X})$ being an NRMI without fixed points of discontinuity, and a sample (X_1, \dots, X_n) consisting of a collection of $k \leq n$ distinct observations (X_1^*, \dots, X_k^*) with frequencies (n_1, \dots, n_k) . To characterize the posterior distribution of the NRMI \tilde{p} , James *et al.* (2009) introduced a latent random variable U_n . Let $G_{1,n}$ be a Gamma random variable, independent of the total mass $T = \tilde{\mu}(\mathbb{X})$, having mean n and variance n ; the density function of the latent variable $U_n = G_{1,n} / T$ is

$$f_{U_n}(u) = \frac{u^{n-1}}{\Gamma(n)} \int_{\mathbb{R}^+} t^n e^{-ut} f_T(t) dt,$$

where Γ denotes the Gamma function and f_T denotes the density function of the random variable T . Moreover, the conditional density function of U_n , given (X_1, \dots, X_n) , is

$$f_{U_n}^{(X_1, \dots, X_n)}(u) \propto u^{n-1} e^{-\Psi(u)} \prod_{i=1}^k \int_{\mathbb{R}^+} s^{n_i} e^{-us} \rho(ds | X_i^*),$$

where $\Psi(u) = \int_{\mathbb{R}^+ \times \mathbb{X}} (1 - e^{-us}) \nu(ds, dx)$. The following important result by James *et al.* (2009) states that the posterior distribution of \tilde{p} , given the sample (X_1, \dots, X_n) , is a mixture of

NRMI. This result is stated using the notation $\tilde{\mu}^{(U_n, X_1, \dots, X_n)}$ for denoting a CRM having a distribution identical to the conditional distribution of $\tilde{\mu}$ given (U_n, X_1, \dots, X_n) .

Theorem 1. (Cf. James et al. (2009)) *Let \tilde{p} be an NRMI obtained by normalizing a CRM $\tilde{\mu}$ with Lévy intensity measure $\rho(ds|x)\alpha(dx)$. Then*

$$\tilde{\mu}^{(U_n, X_1, \dots, X_n)} \stackrel{d}{=} \tilde{\mu}^{(U_n)} + \sum_{i=1}^k J_i^{(U_n, X_1, \dots, X_n)} \delta_{X_i^*},$$

where

- (i) $\tilde{\mu}^{(U_n)}$ is a CRM with Lévy intensity measure $\nu^{(U_n)}(ds, dx) = e^{-U_n s} \rho(ds|x)\alpha(dx)$;
- (ii) X_i^* , for $i=1, \dots, k$, are fixed locations and $J_i^{(U_n, X_1, \dots, X_n)}$, for $i=1, \dots, k$, are random jumps with density functions proportional to $s^{n_i} e^{-U_n s} \rho(ds|X_i^*)$;
- (iii) $\tilde{\mu}^{(U_n)}$ and $J_i^{(U_n, X_1, \dots, X_n)}$, $i=1, \dots, k$, are independent.

Moreover, the conditional distribution of \tilde{p} , given (U_n, X_1, \dots, X_n) , coincides with the distribution of the random probability measure

$$\tilde{p}^{(U_n, X_1, \dots, X_n)} \stackrel{d}{=} w \frac{\tilde{\mu}^{(U_n)}}{T^{(U_n)}} + (1-w) \frac{\sum_{i=1}^k J_i^{(U_n, X_1, \dots, X_n)} \delta_{X_i^*}}{\sum_{i=1}^k J_i^{(U_n, X_1, \dots, X_n)}},$$

where $T^{(U_n)} = \tilde{\mu}^{(U_n)}(\mathbb{X})$ and $w = T^{(U_n)}(T^{(U_n)} + \sum_{i=1}^k J_i^{(U_n, X_1, \dots, X_n)})^{-1}$.

3. The class \mathcal{C}

In this section, we define and study a novel class of NRMI that we will call the class \mathcal{C} . Let $\tilde{\mu}$ be a homogeneous CRM characterized by a Lévy intensity measure ν of the form

$$\nu(ds, dx) = \rho(ds)\alpha(dx) = \sum_{i=0}^{\gamma} \beta_i e^{-\tau s} s^{i-1} ds\alpha(dx), \tag{5}$$

where $\gamma \in \mathbb{N}_0$, τ and β_0 are strictly positive, $\beta = (\beta_0, \dots, \beta_\gamma)$ is a non-negative vector and α is a non-atomic measure on $(\mathbb{X}, \mathcal{X})$ with total mass $a = \alpha(\mathbb{X})$. We can write $\alpha = a\alpha_0$ for some probability measure α_0 . The kernel ρ in (5) is a weighted sum of functions belonging to the class

$$\mathcal{K} := \{s \mapsto e^{-\tau s} s^{i-1} : \tau > 0 \text{ and } i \in \mathbb{N}_0\}.$$

Weighted sums of kernels belonging to \mathcal{K} have important applications in the Bayesian literature. These weighted sums can be used, as proposed in Dalal & Hall (1983), for specifying mathematically tractable prior distributions for exchangeable Poisson random variables. More generally, Dalal & Hall (1983) introduced the idea of using mixtures of natural conjugate prior distributions for parametric models belonging to the exponential family because of two major reasons: these mixtures allow one to consider a large class of prior distributions, much larger than the conjugate priors, and, at the same time, the resulting posterior distributions have attractive closed form representations. Similarly, we use the kernels \mathcal{K} because they allow us to define a large class of NRMI and because posterior distributions of these random probability measures, in light of the results in James et al. (2009), have convenient representations. We can now state the definition of an NRMI belonging to the class \mathcal{C} .

Definition 3. *Let $\tilde{\mu}$ be a CRM with Lévy intensity measure of the form (5). A random probability measure \tilde{p} in the class \mathcal{C} with parameter $(\alpha, \gamma, \tau, \beta)$ is defined as $\tilde{p} = \tilde{\mu}l(\mathbb{X})$.*

3.1. Denseness of \mathcal{C} in the class of homogeneous NRMI

An interesting feature of the class \mathcal{C} is the denseness with respect to the homogeneous NRMI, i.e. any homogeneous NRMI can be arbitrarily well approximated by an NRMI in the class \mathcal{C} . In this subsection, given a homogeneous NRMI $\tilde{\rho}$ with intensity measure ν , we formalize the idea of specifying a sequence $(\nu_i)_{i \geq 1}$ of Lévy measures representable as in (5) approximating ν . Then, this sequence is used for defining NRMI $(\tilde{\rho}_i)_{i \geq 1}$ in the class \mathcal{C} that approximate the random probability measure $\tilde{\rho}$. In the next paragraphs we describe a constructive definition of the approximating sequence $(\nu_i)_{i \geq 1}$.

Let $\tilde{\mu}$ be a CRM with a Lévy intensity measure $\nu(ds, dx) = \rho(ds)\alpha(dx)$ such that $\nu(\mathbb{R}^+ \times \mathbb{X}) < +\infty$. The finiteness of ν implies that $\tilde{\mu}$ has a finite number of jumps. The number of jumps M is a Poisson random variable and we can write

$$\tilde{\mu} = \sum_{i \leq M} S_i \delta_{\tilde{X}_i},$$

where (S_1, \dots, S_M) are the random jumps and $(\tilde{X}_1, \dots, \tilde{X}_M)$ are the corresponding \mathbb{X} -valued random locations. To define an approximating sequence of Lévy intensity measures $(\nu_i)_{i \geq 1}$, we consider a collection of non-negative random variables $(S_{i,j})_{i \geq 1, j \geq 1}$. These random variables, conditionally on M and (S_1, \dots, S_M) , are independently distributed and

$$S_{i,j} \mid M, S_1, \dots, S_M \stackrel{d}{=} \begin{cases} 0 & \text{if } M < j, \\ G_{i, \min\{i^2, I(i, S_j)\}} & \text{if } M \geq j, \end{cases} \tag{6}$$

where $I(i, S_j) := \inf\{l \in \mathbb{N} : l \geq i S_j\}$ and $G_{a,b}$ denotes a Gamma random variable with mean b/a and variance b/a^2 . It can be easily verified that

$$\lim_{i \rightarrow +\infty} \mathbb{E} \left[\sum_{j \leq M} (S_j - S_{i,j})^2 \right] = 0.$$

The collection of random variables $(S_{i,j})_{i \geq 1, j \geq 1}$ allows us to define a sequence of homogeneous CRMs $(\tilde{\mu}_i)_{i \geq 1}$ such that, for every $i \geq 1$, the CRM $\tilde{\mu}_i$, conditionally on $\tilde{\mu}$, has M non-negative random jumps $(S_{i,1}, \dots, S_{i,M})$ at the \mathbb{X} -valued locations $(\tilde{X}_1, \dots, \tilde{X}_M)$. The random measure $\tilde{\mu}_i$, for $i \geq 1$, has a random number of identically distributed jumps and expression (6) implies that the Lévy measures of this CRM is

$$\nu_i(ds, dx) = \rho_i(ds)\alpha(dx) = \sum_{j=1}^{i^2} \beta_{i,j} e^{-is} s^{j-1} ds\alpha(dx),$$

where

$$\beta_{i,j} := \frac{i^j}{\Gamma(j)} \rho(s \in \mathbb{R}^+ : \min\{i^2, I(i, s)\} = j).$$

The next proposition states that the sequence of CRMs $(\tilde{\mu}_i)_{i \geq 1}$ approximates the CRM $\tilde{\mu}$.

Proposition 1. *Let $\tilde{\mu}$ and $(\tilde{\mu}_i)_{i \geq 1}$ be homogeneous CRMs defined on the same probability space. Assume that $\tilde{\mu}$ has a finite Lévy measure ν . Denote by (S_1, \dots, S_M) and $(\tilde{X}_1, \dots, \tilde{X}_M)$ the random jumps and the discontinuity points of $\tilde{\mu}$. If $\tilde{\mu}_i$, for every $i \geq 1$, has discontinuity points $(\tilde{X}_1, \dots, \tilde{X}_M)$ and, conditionally on $\tilde{\mu}$, has independent random jumps $(S_{i,1}, \dots, S_{i,M})$ distributed as in (6), then, for any $\varepsilon > 0$*

$$\lim_{i \rightarrow +\infty} \mathbb{P}(\sup_{B \in \mathcal{X}} |\tilde{\mu}(B) - \tilde{\mu}_i(B)| > \varepsilon) = 0. \tag{7}$$

In the next paragraphs, we extend the described construction of a sequence of finite Lévy measures to approximate a homogeneous CRM $\tilde{\mu}$ with intensity measure $\nu(ds, dx)$ such that $\nu(\mathbb{R}^+ \times \mathbb{X}) = +\infty$. That is, we consider a CRM $\tilde{\mu}$ such that the NRMI $\tilde{p} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$ is a well defined random distribution with an infinite number of discontinuity points. For any $\zeta > 0$, the CRM $\tilde{\mu}$ can be decomposed into the sum of two independent CRMs,

$$\tilde{\mu} = \tilde{\mu}^{(0, \zeta]} + \tilde{\mu}^{(\zeta, +\infty)}, \tag{8}$$

where $\tilde{\mu}^{(0, \zeta]}$ is a homogeneous CRM with Lévy intensity measure $\nu^{(0, \zeta]}$ on $(0, \zeta] \times \mathbb{X}$ and $\tilde{\mu}^{(\zeta, +\infty)}$ is a homogeneous CRM with Lévy intensity measure $\nu^{(\zeta, +\infty)}$ on $(\zeta, +\infty) \times \mathbb{X}$. The decomposition (8) is directly suggested by the representation of the CRM $\tilde{\mu}$ as a functional of a Poisson random measure with intensity ν . Indeed, expression (8) exploits the fact that the sum of two independent Poisson random measures characterized by mean intensity measures $\nu^{(0, \zeta]}$ and $\nu^{(\zeta, +\infty)}$ is a Poisson random measure with intensity measure $\nu^{(0, \zeta]} + \nu^{(\zeta, +\infty)}$. The decomposition (8) implies that $\nu = \nu^{(0, \zeta]} + \nu^{(\zeta, +\infty)}$. This decomposition has a clear interpretation: the CRM $\tilde{\mu}^{(\zeta, +\infty)}$ includes the jumps in $\tilde{\mu}$ having magnitude larger than ζ while the CRM $\tilde{\mu}^{(0, \zeta]}$ includes all the remaining jumps in $\tilde{\mu}$.

The CRM $\tilde{\mu}^{(\zeta, +\infty)}$ has a finite number of non-negative random jumps. Therefore, according to proposition 1, we can define a sequence of homogeneous CRMs $(\tilde{\mu}_i^*)_{i \geq 1}$ such that for any $i \geq 1$ the CRM $\tilde{\mu}_i^*$ is characterized by a Lévy intensity measure ν_i^* of the form

$$\nu_i^*(ds, dx) = \rho_i^*(ds)\alpha(dx) = \sum_{j=1}^{\gamma_i} \beta_{i,j} e^{-\tau_i s} s^{j-1} ds\alpha(dx)$$

and

$$\mathbb{P} \left(\sup_{B \in \mathcal{X}} |\tilde{\mu}^{(1/i, +\infty)}(B) - \tilde{\mu}_i^*(B)| > \frac{1}{i} \right) < \frac{1}{i}. \tag{9}$$

Proposition 1 states that, given a CRM having finite intensity measure, for any pair of positive numbers (ε, ζ) , we can define a CRM, whose intensity measure can be represented as a weighted sum of Gamma densities, such that the probability of a distance in total variation between the two CRMs larger than ε is less than ζ . Analogously, the next proposition considers the CRMs $(\tilde{\mu}^{(1/i, +\infty)})_{i \geq 1}$ and each of these CRMs is approximated by a CRM $\tilde{\mu}_i^*$ satisfying (9). We also define, for every $i \geq 1$, a CRM

$$\tilde{\mu}_i := \tilde{\mu}_i^* + \eta_i, \tag{10}$$

where η_i is a Gamma CRM, with Lévy measure $\nu_{\eta_i}(ds, dx) = (\tau_i/is) e^{-\tau_i s} ds\alpha(dx)$, and is independent of $\tilde{\mu}_i^*$. Observe that $\tilde{\mu}_i$ is a CRM having Lévy measure

$$\nu_i(ds, dx) = \rho_i(ds)\alpha(dx) = \sum_{j=0}^{\gamma_i} \beta_{i,j} e^{-\tau_i s} s^{j-1} ds\alpha(dx), \tag{11}$$

where $\beta_{i,0} = \tau_i/i$. Therefore the NRMI $\tilde{p}_i = \tilde{\mu}_i/\tilde{\mu}(\mathbb{X})$ belongs to the class \mathcal{C} . The next proposition points out that the sequence of homogeneous NRMI $(\tilde{p}_i)_{i \geq 1}$ approximates the random distribution \tilde{p} .

Proposition 2. *Let $\tilde{\mu}$ and $\tilde{\mu}^{(0, 1/i]}$, $\tilde{\mu}^{(1/i, \infty)}$, $\tilde{\mu}_i^*$, $\tilde{\mu}_i$, η_i , for $i \geq 1$, be CRMs with intensity measures ν , $\nu^{(0, 1/i]}$, $\nu^{(1/i, \infty)}$, ν_i^* , ν_i and ν_{η_i} . If expressions (8), (9), (10) and (11) hold, $\nu(ds, dx) = \rho(ds)\alpha(dx)$, $\nu_{\eta_i}(ds, dx) = (\tau_i/is) e^{-\tau_i s} ds\alpha(dx)$, $\tilde{\mu}_i^*$ and η_i are independent, $\tilde{\mu}(\mathbb{X}) > 0$ almost surely, $\tilde{p} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$ and, for $i \geq 1$, $\tilde{p}_i = \tilde{\mu}_i/\tilde{\mu}_i(\mathbb{X})$ then, for every $\varepsilon > 0$*

$$\lim_{i \rightarrow +\infty} \mathbb{P}(\sup_{B \in \mathcal{X}} |\tilde{p}(B) - \tilde{p}_i(B)| > \varepsilon) = 0. \tag{12}$$

3.2. Predictive distributions of \mathcal{C} and applications to species sampling problems

Consider model (1), with \tilde{p} belonging to the class \mathcal{C} . By the almost sure discreteness of \tilde{p} , the random vector (X_1, \dots, X_n) induces a distribution on the partitions of the set $\{1, \dots, n\}$. The underlying map, for obtaining this distribution, given (X_1, \dots, X_n) , indicates the partition such that, for any pair of integers i and j in $\{1, \dots, n\}$, the two integers belong to the same subset if and only if $X_i = X_j$. We use K_n for denoting the number of subsets. Let $(X_1^*, \dots, X_{K_n}^*)$ be the points of \mathbb{X} observed in (X_1, \dots, X_n) . We use N_j to denote the cardinality $\sum_{i=1}^n \mathbb{1}(X_j^* = X_i)$, for $j = 1, \dots, K_n$. The probabilities of the partitions of $\{1, \dots, n\}$ can be represented by means of a function depending only on the number of subsets K_n and the subsets' cardinalities N_1, \dots, N_{K_n} . This function is called exchangeable partition probability function (EPPF). We use $p_k^{(n)}(n_1, \dots, n_k)$ to denote the probability of a specific partition of $\{1, \dots, n\}$ with $K_n = k$ subsets having cardinalities $(N_1, \dots, N_{K_n}) = (n_1, \dots, n_k)$. We refer to Pitman (1995) and Pitman (2006) for insightful and comprehensive discussions on this topic.

A simple application of proposition 3 in James *et al.* (2009) allows us to obtain an explicit expression for the EPPF when the exchangeable random variables $(X_i)_{i \geq 1}$ are sampled from an NRMI \tilde{p} belonging to the class \mathcal{C} with parameter $(\alpha, \gamma, \tau, \beta)$:

$$p_k^{(n)}(n_1, \dots, n_k) = \frac{\alpha^k \tau^\alpha \beta_0 e^{-a \sum_{i=1}^\gamma \beta_i \Gamma(i) \tau^{-i}}}{\Gamma(n)} \times \sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma, k}} \prod_{l=1}^k \beta_{r_l} \Gamma(r_l + n_l) g_{n, k}(\gamma, \beta, \tau, |r|, a), \tag{13}$$

where

$$g_{n, k}(\gamma, \tau, \beta, |r|, a) := \int_0^{+\infty} \frac{t^{n-1} e^{a \sum_{i=1}^\gamma \beta_i \Gamma(i) (\tau+t)^{-i}}}{(\tau+t)^{a\beta_0 + |r| + n}} dt,$$

$$\mathcal{G}_{\gamma, k} := \{0, 1, \dots, \gamma\}^k \text{ and } |r| := \sum_{i=1}^k r_i.$$

Remark 1. It is worth observing an alternative representation of the EPPF (13). It can be verified by simple algebra that

$$p_k^{(n)}(n_1, \dots, n_k) = \sum_{j_1 \geq 0} \sum_{j_2 \geq 0} \dots \sum_{j_\gamma \geq 0} \left(\prod_{i=1}^\gamma \frac{(a\beta_i \Gamma(i) \tau^{-i})^{j_i} e^{-a\beta_i \Gamma(i) \tau^{-i}}}{j_i!} \right) \times \sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma, k}} \frac{\prod_{i=1}^k \Gamma(r_i + n_i) a \beta_{r_i} \tau^{-r_i}}{(\sum_{i=1}^\gamma i j_i + \sum_{i=1}^k r_i + a\beta_0)_{n, 1}}, \tag{14}$$

where $(a)_{n, b} = \prod_{i=0}^{n-1} (a + ib)$ denotes the Pochhammer symbol for the n th factorial power of a with increment b . Expression (14) can be used to compute upper and lower bounds of $p_k^{(n)}(n_1, \dots, n_k)$ by interpreting the quantity $\sum_{i=1}^\gamma i j_i$ in (14) as the sum of independent random variables.

Expression (13) allows us to verify that an NRMI in the class \mathcal{C} with parameter $(\alpha, \gamma, \tau, \beta)$ is characterized by a system of predictive distributions of the form (3);

$$\begin{aligned}
 &w_0(n, k, (n_1, \dots, n_k)) \\
 &= \frac{a \sum_{(r_1, \dots, r_{k+1}) \in \mathcal{G}_{\gamma, k+1}} \prod_{l=1}^{k+1} \beta_{r_l} \Gamma(r_l + n_l) g_{n+1, k+1}(\gamma, \beta, \tau, |r|, a)}{n \sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma, k}} \prod_{l=1}^k \beta_{r_l} \Gamma(r_l + n_l) g_{n, k}(\gamma, \tau, \beta, |r|, a)}
 \end{aligned}$$

and

$$\begin{aligned}
 &w_j(n, k, (n_1, \dots, n_k)) \\
 &= \frac{\sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma, k}} \prod_{l=1}^k \beta_{r_l} \Gamma(r_l + n_l + \mathbb{1}(l=j)) g_{n+1, k}(\gamma, \beta, \tau, |r|, a)}{n \sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma, k}} \prod_{l=1}^k \beta_{r_l} \Gamma(r_l + n_l) g_{n, k}(\gamma, \tau, \beta, |r|, a)}
 \end{aligned}$$

with the proviso that $n_{k+1} = 1$. Before proceeding we compare the predictive structures of the Gibbs-type random probability measures and the introduced class \mathcal{C} . For the Gibbs-type random probability measures, the predictive distribution (2) is a linear combination of the prior guess $\mathbb{E}[\tilde{p}(\cdot)] = \alpha_0(\cdot)$ and a weighted empirical distribution. The random variable X_{n+1} is not identical to any of the observed variables (X_1, \dots, X_n) with probability $g_0(n, k)$, whereas it coincides with X_j^* with probability $g_1(n, k)(n_j - \sigma)$, for $j = 1, \dots, k$. In the limiting case of the Dirichlet process, the dependence on k disappears; this, which is a characterizing property of the Dirichlet process (see Zabell, 1982), represents a severe limitation for predictive purposes. The predictive distributions associated with a random probability measure in the class \mathcal{C} are characterized by a more elaborate structure. The predictive distribution is still a linear combination of the prior guess α_0 and a weighted empirical distribution, but now the conditional probability $w_0(n, k, (n_1, \dots, n_k))$ that X_{n+1} is a new species depends on the observed frequencies (n_1, \dots, n_k) .

In the next proposition, we derive the distribution of the number of distinct observations K_n to be observed in a sample (X_1, \dots, X_n) drawn from an NRMI \tilde{p} in \mathcal{C} . The proposition is stated using the definition of the (n, k) th partial Bell polynomial associated with a non-negative sequence of real numbers $w_\bullet := \{w_i, i \geq 1\}$. A brief account on partial Bell polynomials is given in the Appendix.

Proposition 3. *Let $(X_n)_{n \geq 1}$ be an exchangeable sequence governed by an NRMI \tilde{p} in the class \mathcal{C} with parameter $(\alpha, \gamma, \tau, \beta)$. Then*

$$\begin{aligned}
 \mathbb{P}(K_n = k) &= \frac{a^k \tau^{\alpha \beta_0} e^{-a \sum_{i=1}^{\gamma} \beta_i \Gamma(i) \tau^{-i}}}{\Gamma(n)} \\
 &\quad \times \int_0^{+\infty} \frac{t^{n-1} e^{a \sum_{i=1}^{\gamma} \beta_i \Gamma(i) (\tau+t)^{-i}}}{(\tau+t)^{\alpha \beta_0}} B_{n, k}(w_\bullet(t; \gamma, \tau, \beta)) dt, \tag{15}
 \end{aligned}$$

where $B_{n, k}(w_\bullet(t; \gamma, \tau, \beta))$ is the (n, k) th partial Bell polynomial characterized by the sequence $w_\bullet(t; \gamma, \tau, \beta)$ such that $w_l(t; \gamma, \tau, \beta) = \sum_{i=0}^{\gamma} \beta_i (\tau+t)^{-i-l} \Gamma(i+l)$.

The next proposition shows that the expected value and variance of the random variable K_n can be easily derived.

Proposition 4. *Let $(X_n)_{n \geq 1}$ be an exchangeable sequence governed by an NRMI \tilde{p} in the class \mathcal{C} with parameter $(\alpha, \gamma, \tau, \beta)$. Then, for any $n > 1$*

$$\mathbb{P}(K_n = 1 + K_{n-1}) = \sum_{j \geq 0} \mathbb{P} \left(\sum_{i=1}^{\gamma} i Y_i = j \right) \sum_{i=0}^{\gamma} \frac{(a\beta_0 + j)_{(n-1),1} \Gamma(i+1) \beta_i \tau^{-i} a}{(a\beta_0 + j + i)_{n,1}}, \tag{16}$$

where Y_1, \dots, Y_{γ} is a collection of independent Poisson random variables with parameters $\Gamma(1)\beta_1\tau^{-1}, \dots, \Gamma(\gamma)\beta_{\gamma}\tau^{-1}$, respectively. Moreover, for any $1 < m < n$

$$\begin{aligned} &\mathbb{P}(K_m = 1 + K_{m-1}, K_n = 1 + K_{n-1}) \\ &= a^2 \sum_{j \geq 0} \mathbb{P} \left(\sum_{i=1}^{\gamma} i Y_i = j \right) (a\beta_0 + j)_{m-1,1} \\ &\quad \times \sum_{i_1=0}^{\gamma} \sum_{i_2=0}^{\gamma} \frac{(a\beta_0 + j + i_1 + m)_{n-m-1,1} \Gamma(i_1+1) \beta_{i_1} \tau^{-i_1} \Gamma(i_2+1) \beta_{i_2} \tau^{-i_2}}{(a\beta_0 + j + i_1 + i_2)_{n,1}}. \end{aligned} \tag{17}$$

In the next two propositions we denote by $X^{(1,n)} := (X_1, \dots, X_n)$ a sample of size n and, analogously, we use $X^{(n+1,m)} := (X_{n+1}, \dots, X_{n+m})$ to denote an additional sample of size m , whose characteristics can be predicted based on $X^{(1,n)}$. We also define the random variable $K_m^{(n)} := K_{n+m} - K_n$. Note that $K_m^{(n)}$ indicates the number of new distinct observations in the additional sample $X^{(n+1,m)}$. The probability distribution of the random variable K_n in (15) can be interpreted as the prior distribution for the number of species to be observed in the sample $X^{(1,n)}$. This distribution represents the starting point for deriving, given $X^{(1,n)}$, the conditional distribution of $K_m^{(n)}$ and the conditional probability of observing a new species at the $(n + m + 1)$ th draw; we use $K_1^{(n+m)} = 1$ to denote this event. The next two propositions provide explicit expressions for these two quantities.

Proposition 5. *Let $(X_n)_{n \geq 1}$ be an exchangeable sequence governed by an NRM \tilde{p} in the class C with parameter $(\alpha, \gamma, \tau, \beta)$. Then*

$$\begin{aligned} \mathbb{P}(K_m^{(n)} = j | X^{(1,n)}) &= \frac{a^{k+j} \tau^{a\beta_0} e^{-a \sum_{i=1}^{\gamma} \beta_i \Gamma(i) \tau^{-i}}}{p_k^{(n)}(n_1, \dots, n_k) \Gamma(n+m)} \\ &\quad \times \sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma,k}} \prod_{l=1}^k \beta_{r_l} \Gamma(r_l + n_l) \sum_{s=j}^m \binom{m}{s} (n + |r|)_{m-s,1} \\ &\quad \times \int_0^{+\infty} \frac{t^{n+m-1} e^{a \sum_{i=1}^{\gamma} \beta_i \Gamma(i) (\tau+t)^{-i}}}{(\tau+t)^{a\beta_0 + |r| + n + m - s}} B_{s,j}(w_{\bullet}(t; \gamma, \tau, \beta)) dt, \end{aligned} \tag{18}$$

where $B_{n,k}(w_{\bullet}(t; \gamma, \tau, \beta))$ is the (n, k) th partial Bell polynomial characterized by the sequence $w_{\bullet}(t; \gamma, \tau, \beta)$ such that $w_i(t; \gamma, \tau, \beta) = \sum_{l=0}^{\gamma} \beta_l (\tau + t)^{-i-l} \Gamma(i+l)$ and $p_k^{(n)}(n_1, \dots, n_k)$ is the EPPF in (13).

The conditional distribution (18) can be interpreted as the predictive distribution of the number of new species to be observed in the additional sample $X^{(n+1,m)}$. This conditional distribution exhibits dependence on both the number of species k and the frequencies (n_1, \dots, n_k) observed in $X^{(1,n)}$. In contrast, under a Gibbs-type model, as shown in Lijoi *et al.* (2007b) the conditional distribution of $K_m^{(n)}$, given $X^{(1,n)}$, depends only on the number of species k in $X^{(1,n)}$.

Proposition 6. *Let $(X_n)_{n \geq 1}$ be an exchangeable sequence governed by an NRM \tilde{p} in the class C with parameter $(\alpha, \gamma, \tau, \beta)$. Then*

$$\begin{aligned} \mathbb{P}(K_1^{(n+m)} = 1 \mid X^{(1,n)}) &= \frac{a^{k+1} \tau^{\alpha \beta_0} e^{-a \sum_{i=1}^{\gamma} \beta_i \Gamma(i) \tau^{-i}}}{\Gamma(n+m+1) p_k^{(n)}(n_1, \dots, n_k)} \sum_{i=0}^{\gamma} \beta_i \Gamma(i+1) \\ &\times \sum_{(r_1, \dots, r_k) \in \mathcal{G}_{\gamma, k}} \prod_{l=1}^k \beta_{r_l} \Gamma(r_l + n_l) \sum_{j=0}^m \sum_{s=j}^m \binom{m}{s} a^j \left(n + \sum_{h=1}^k r_h \right)_{m-s, 1} \\ &\times \int_0^{+\infty} \frac{t^{n+m} e^{a \sum_{v=1}^{\gamma} \beta_v \Gamma(v) (\tau+t)^{-v}}}{(\tau+t)^{\alpha \beta_0 + i + 1 + |r| + n + m - s}} B_{s,j}(w_{\bullet}(t; \gamma, \tau, \beta)) dt, \end{aligned}$$

where $B_{n,k}(w_{\bullet}(t; \gamma, \tau, \beta))$ is the (n, k) th partial Bell polynomial characterized by the sequence $w_{\bullet}(t; \gamma, \tau, \beta)$ such that $w_i(t; \gamma, \tau, \beta) = \sum_{l=0}^{\gamma} \beta_l (\tau+t)^{-i-l} \Gamma(i+l)$ and $p_k^{(n)}(n_1, \dots, n_k)$ is the EPPF in (13).

4. An exact sampling scheme

In this section, we propose a coupling from the past algorithm (Propp & Wilson, 1996) for sampling from the predictive distributions characterizing NRMI in the class \mathcal{C} .

If $\tilde{p} = \tilde{\mu} / \tilde{\mu}(\mathbb{X})$, then the CRM $\tilde{\mu}$ can be represented as the sum of $(\gamma + 1)$ independent CRMs $(\tilde{\mu}_0, \tilde{\mu}_1, \dots, \tilde{\mu}_{\gamma})$ such that $\tilde{\mu}_i$, for $i=0, \dots, \gamma$, has Lévy intensity measure $\nu_i(dx, ds) = \beta_i e^{-\tau s} s^{i-1} ds \alpha(dx)$. The number of jumps in $\tilde{\mu}_i$, for $i=1, \dots, \gamma$, has a Poisson distribution. For any $i=1, \dots, \gamma$ let $H_i(B)$ stand for the number of jumps in $\tilde{\mu}_i$ contained in $B \in \mathcal{X}$; that is, H_i is a Poisson random measure with intensity $\alpha \Gamma(i) \tau^{-i} \beta_i$. The jumps in $\tilde{\mu}_i$ are independent Gamma random variables with mean (i/τ) . We observe, by directly exploiting the fact that a normalized Gamma CRM is a Dirichlet process, that the conditional distribution of \tilde{p} given H_1, \dots, H_{γ} is the law of a Dirichlet process with parameter $\beta_0 \alpha + H$, where $H = \sum_{i=1}^{\gamma} i H_i$. The conditional distribution of $X^{(1,n)}$ given H_1, \dots, H_{γ} has a simple closed form. We also use the latent variable U_n introduced in James *et al.* (2009); this random variable, given $(\tilde{\mu}_0, \tilde{\mu}_1, \dots, \tilde{\mu}_{\gamma})$ and $X^{(1,n)}$, has a conditional Gamma distribution with mean $n/\tilde{\mu}(\mathbb{X})$.

The posterior characterization of an NRMI \tilde{p} discussed in James *et al.* (2009) allows us to derive the conditional distribution of $H(\mathbb{X})$ given $(U_n, X^{(1,n)})$. The random variable $H_i(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\})$, for $i=1, \dots, \gamma$, given $(U_n, X^{(1,n)})$, has a conditional Poisson distribution with mean

$$\int \int \beta_i e^{-U_n s - \tau s} s^{i-1} ds \alpha(dx) = \alpha(\mathbb{X}) \Gamma(i) (\tau + U_n)^{-i} \beta_i, \tag{19}$$

and

$$\begin{aligned} \mathbb{P}((H(X_1^*), \dots, H(X_{K_n}^*)) = (i_1, \dots, i_k) \mid U_n, X^{(1,n)}) \\ = \mathbb{1}(K_n = k) \prod_{j=1}^{K_n} \frac{\beta_{i_j} \int e^{-U_n s - \tau s} s^{i_j - 1 + N_j} ds}{\sum_{h=0}^{\gamma} \beta_h \int e^{-U_n s - \tau s} s^{h-1 + N_j} ds} \end{aligned} \tag{20}$$

for every $k=1, \dots, n$ and $(i_1, \dots, i_k) \in \{0, \dots, \gamma\}^k$. Moreover, the random variables

$$H_1(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}), \dots, H_{\gamma}(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}), \sum_{i=1}^{K_n} H(X_i^*),$$

given $(U_n, X^{(1,n)})$, are conditionally independent. It follows that, for $i \geq 0$, the conditional probability $\mathbb{P}(H(\mathbb{X}) = i \mid U_n, X^{(1,n)})$ can be computed by using expressions (19) and (20). It is known that, if \tilde{p} is defined by normalizing a Gamma CRM, then $X^{(1,n)}$ and U_n are independent. This fact implies that, if \tilde{p} belongs to the \mathcal{C} class, then

$$\begin{aligned} \mathbb{P}(U_n \in du \mid X^{(1,n)}, H(\mathbb{X})) &= \mathbb{P}(U_n \in du \mid H(\mathbb{X})) \\ &= \int \frac{e^{-tu} t^n u^{n-1} e^{-\tau t} \tau^{H(\mathbb{X}) + a\beta_0} t^{H(\mathbb{X}) + a\beta_0 - 1}}{\Gamma(n)\Gamma(H(\mathbb{X}) + a\beta_0)} dt. \end{aligned} \tag{21}$$

The reported conditional distributions illustrate that it is simple to construct a Gibbs sampling algorithm that iteratively generates random variables from the conditional distribution of $H(\mathbb{X})$ given $(U_n, X^{(1,n)})$, and from the full conditional of U_n given $(H(\mathbb{X}), X^{(1,n)})$. This Gibbs sampler constitutes the skeleton of our algorithm.

Observe that if U_n can be exactly generated conditionally on $X^{(1,n)}$, then we can sample from the predictive distribution. Indeed, for every $B \in \mathcal{X}$

$$\begin{aligned} \mathbb{P}(X_{n+1} \in B \mid U_n, X^{(1,n)}) &= \sum \mathbb{P}(H(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}), H(X_1^*), \dots, H(X_{K_n}^*) \mid X^{(1,n)}, U_n) \\ &\quad \times \mathbb{P}(X_{n+1} \in B \mid H(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}), H(X_1^*), \dots, H(X_{K_n}^*), X^{(1,n)}) \\ &= \sum \mathbb{P}(H(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}), H(X_1^*), \dots, H(X_{K_n}^*) \mid X^{(1,n)}, U_n) \\ &\quad \times \frac{(H(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}) + a\beta_0)\alpha_0(B) + \sum_{j=1}^{K_n} (H(X_j^*) + N_j)\mathbb{1}(X_j^* \in B)}{a\beta_0 + H(\mathbb{X}) + n}, \end{aligned}$$

where the sum is over the possible values of $(H(\mathbb{X} \setminus \{X_1^*, \dots, X_{K_n}^*\}), H(X_1^*), \dots, H(X_{K_n}^*))$. Similarly, the sequence $(X_i)_{i>n}$, conditionally on $(U_n, X^{(1,n)})$ and H , is identical in distribution to an exchangeable sequence sampled from a Dirichlet process parameterized by the measure $\beta_0\alpha + H + \sum_{i=1}^n \delta_{X_i}$.

Before stating the algorithm we briefly review the coupling from the past method introduced by Propp & Wilson (26). Consider a space \mathbf{Z} with a partial order \preceq . The space \mathbf{Z} has a maximal element \bar{z} and a minimal element \underline{z} ; i.e. the relations $z \preceq \bar{z}$ and $\underline{z} \preceq z$ hold for every $z \in \mathbf{Z}$. Let π be a distribution on \mathbf{Z} . We define two classes of random processes

$$L := (L_{-j})_{j \geq 0} \quad \text{and} \quad V := (V_{-j})_{j \geq 0}.$$

The trajectories of $L_{-j} := (L_{-j}^i)_{i \geq -j}$ and $V_{-j} := (V_{-j}^i)_{i \geq -j}$ start at time $-j$ and, for every $i \geq -j$, both L_{-j}^i and V_{-j}^i are \mathbf{Z} -valued variables. The starting values are $L_{-j}^{-j} = \underline{z}$ and $V_{-j}^{-j} = \bar{z}$. The random processes L_{-j} and V_{-j} are Markov chains having identical transition probabilities and could be used for implementing a Markov Chain Monte Carlo algorithm for approximate sampling π distributed variables. The definition of V and L satisfies a monotonicity condition and a coalescence condition. The former is verified if, for every pair (j_1, j_2) and $i \geq \max(-j_1, -j_2)$, the relation $V_{-j_1}^i \succeq L_{-j_1}^i$ holds almost surely, the event $V_{-j_1}^i \succeq V_{-j_2}^i$ implies $V_{-j_1}^{i+1} \succeq V_{-j_2}^{i+1}$ and, similarly, the event $L_{-j_1}^i \succeq L_{-j_2}^i$ implies $L_{-j_1}^{i+1} \succeq L_{-j_2}^{i+1}$. The latter is verified if the binary sequence $(\mathbb{1}(L_0^i = V_0^i))_{i \geq 0}$ converges almost surely to 1. Propp & Wilson (1996) proved that, under these assumptions, the limits $\lim_j V_{-j}^0$ and $\lim_j L_{-j}^0$ are well defined and almost surely identical. Moreover, the equality $\mathbb{P}(\lim_j V_{-j}^0 \in B) = \pi(B)$ holds for every measurable set B . Finally, the variable $C := \min\{j \geq 1 : V_{-j}^0 = L_{-j}^0\}$ is well defined, for every $j \geq C$, the equalities $V_{-C}^0 = L_{-C}^0 = V_{-j}^0 = L_{-j}^0$ hold and, if it is feasible to generate V_{-C}^0 , then the π distributed random variable $\lim_j V_{-j}^0$ can be exactly generated.

We can now start describing an exact algorithm for generating $(U_n, H(\mathbb{X}))$ conditionally on $X^{(1,n)}$ by defining the transition probabilities of the Markov chains L_{-j} and V_{-j} . We define $\mathbf{Z} := [0, \infty]^2$ and introduce the notation $L_{-j}^i = (L_{-j,1}^i, L_{-j,2}^i)$ and $V_{-j}^i = (V_{-j,1}^i, V_{-j,2}^i)$ for emphasizing that L_{-j}^i and V_{-j}^i are random vectors. The transitions $L_{-j}^i \rightarrow L_{-j}^{i+1}$ have strict similarities

with the previously outlined Gibbs sampler. The random variable $L_{-j,1}^{i+1}$, conditioned on L_{-j}^i , can be generated by sampling from the conditional distribution of U_n given $(H(\mathbb{X}), X^{(1,n)})$, assuming $H(\mathbb{X}) = L_{-j,2}^i$. The random variable $L_{-j,2}^{i+1}$, conditioned on $(L_{-j}^i, L_{-j,1}^{i+1})$, can be generated by sampling from the conditional distribution of $H(\mathbb{X})$ given $(U_n, X^{(1,n)})$, assuming $U_n = L_{-j,1}^{i+1}$. These two steps describe the transition probabilities of L_{-j} and V_{-j} ; we recall that the transition probabilities of these Markov chains are identical. We complete the definitions of L_{-j} and V_{-j} by setting $L_{-j}^{-j} = (0, \infty)$, $V_{-j}^{-j} = (\infty, 0)$ and $\mathbb{P}(L_{-j,1}^{-j+1} = 0 | L_{-j}^{-j} = (0, \infty)) = 1$.

To jointly define L and V , we introduce the independent random elements $\Phi = (\Phi_i)_{i \in \mathbb{Z}}$, $\nabla = (\nabla_i)_{i \in \mathbb{Z}}$ and $\Upsilon = (\Upsilon_i(t))_{t \geq 0, i \in \mathbb{Z}}$, where \mathbb{Z} is the set of the integer numbers. The random variables $(\Phi_i)_{i \in \mathbb{Z}}$ are independently and uniformly distributed on the unit interval, $(\nabla_i)_{i \in \mathbb{Z}}$ is a sequence of independent Gamma random variables with mean n and variance n , and $(\Upsilon_i(t))_{t > 0, i \in \mathbb{Z}}$ are independent increasing Gamma processes such that, for every $t > 0$, $\Upsilon_i(t)$ is a Gamma random variable with mean t and variance t . We can now define, for $j \geq 0$ and $i \geq -j$, the random variables $(L_{-j,1}^i, L_{-j,2}^i)$ and $(V_{-j,1}^i, V_{-j,2}^i)$ by using a system of equations: for every $i \geq -j$

$$L_{-j,1}^{i+1} = \frac{\tau \nabla_i}{\Upsilon_i(a\beta_0 + L_{-j,2}^i)}, \quad V_{-j,1}^{i+1} = \frac{\tau \nabla_i}{\Upsilon_i(a\beta_0 + V_{-j,2}^i)}, \tag{22}$$

and

$$L_{-j,2}^{i+1} = \inf\{h \in \mathbb{N}_0 : F(h, L_{-j,1}^{i+1}) \geq \Phi_i\}, \quad V_{-j,2}^{i+1} = \inf\{h \in \mathbb{N}_0 : F(h, V_{-j,1}^{i+1}) \geq \Phi_i\}, \tag{23}$$

where F is a real function on $\mathbb{N}_0 \times [0, \infty)$ such that $F(h, u)$ is the conditional probability of the event $H(\mathbb{X}) \leq h$ given $X^{(1,n)}$ and the latent variable $U_n = u$. Expression (23) is a direct application of the so-called probability integral transform method. Equations (22) and (23) are consistent with the previously reported definitions of the Markov chains L_{-j} and V_{-j} . Indeed, expression (21) shows that the conditional density of U_n given $(H(\mathbb{X}), X^{(1,n)})$ can be represented as a mixture of Gamma distributions with fixed shape parameter.

We use the following partial order: $(u_1, h_1) \leq (u_2, h_2)$ if and only if $u_1 \leq u_2$ and $h_1 \geq h_2$. Expression (22) shows that $\mathbb{P}(L_{-j,1}^{i+1} \leq L_{-j,2}^{i+1} | L_{-j,1}^i \leq L_{-j,2}^i) = 1$. Moreover, expressions (19) and (20) imply that, for every h , if $u_1 \leq u_2$, then,

$$\mathbb{P}(H(\mathbb{X}) \leq h | X^{(1,n)}, U_n = u_1) \leq \mathbb{P}(H(\mathbb{X}) \leq h | X^{(1,n)}, U_n = u_2).$$

These facts imply that the definition of L and V satisfies the monotonicity requirement. Expressions (22) and (23) imply that L and V satisfy the coalescence condition, indeed, for every δ such that $\mathbb{P}(H(\mathbb{X}) = 0 | X^{(1,n)}, U_n = 0) > \delta > 0$,

$$\mathbb{P}\left(\bigcap_{i \geq l} \{L_{-j}^i = V_{-j}^i\} \mid \min\{\Phi_h; h = -j, -j+1, \dots, l-2\} < \delta\right) = 1.$$

We specify a procedure for generating the variables $(L_{-j}^i, V_{-j}^i)_{1 \leq j \leq C, -j \leq i \leq 0}$. We recall that $C = \min\{j \geq 1 : V_{-j}^0 = L_{-j}^0\}$. The algorithm, for every $1 \leq j \leq C$, sequentially generates the variables $(L_{-j}^i, V_{-j}^i)_{-j \leq i \leq 0}$ after $(L_{-j+1}^i, V_{-j+1}^i)_{-j < i \leq 0}$. At each iteration $1 \leq j \leq C$, for every $-j < i \leq 0$, the vector (L_{-j}^i, V_{-j}^i) is generated after $(L_{-j}^{i-1}, V_{-j}^{i-1})$. We also recall that the distribution π of V_{-C}^0 and the conditional distribution $\mathbb{P}(U_n, H(\mathbb{X}) | X^{(1,n)})$ are identical. At the beginning of the j th iteration Φ_{-j} , ∇_{-j} and $\Upsilon_{-j}(\alpha(\mathbb{X}))$ are generated to compute $(L_{-j}^{-j+1}, V_{-j}^{-j+1})$ by using (22) and (23) and the equalities $L_{-j}^{-j} = (0, \infty)$ and $V_{-j}^{-j} = (\infty, 0)$. We use the proviso $\Upsilon_{-j}(\infty)^{-1} = 0$. After the random variables (L_{-j}^i, V_{-j}^i) have been computed, where $1 \leq j \leq C$ and $-j < i < 0$, consistently with the fact that the variables (Φ_i, ∇_i) as well as some terms of the sequence $\Upsilon_i(t)$, $t = \alpha(\mathbb{X}), \alpha(\mathbb{X}) + 1, \dots$, have already been generated, the algorithm, given these variables,

generates $Y_i(\alpha(\mathbb{X}) + L_{-j,2}^i)$ and $Y_{-j}(\alpha(\mathbb{X}) + V_{-j,2}^i)$ to compute $(L_{-j}^{i+1}, V_{-j}^{i+1})$ by using expressions (22) and (23). That is, at each step the algorithm generates the components of (Φ, ∇, Y) that are necessary for computing a pair of \mathbf{Z} -valued variables (L_{-j}^i, V_{-j}^i) conditionally on the previously generated components. These components can be easily generated from the conditional distributions. The random vector $(\Phi_{-j}, \nabla_{-j}), j=1, 2, \dots$, is independent of the complementary components of (Φ, ∇, Y) ; we only need to generate a Uniform random variable and a Gamma random variable. The definition of (Φ, ∇, Y) also implies that, when we generate $Y_{-j}(t)$, it suffices to condition on those terms of the sequence $Y_i(t), t = \alpha(\mathbb{X}), \alpha(\mathbb{X}) + 1, \dots$, that have already been generated. Moreover, for every $t > 0, t_1 < t_2 < \dots < t_l$ and $j = 1, \dots, l - 1$,

$$\mathbb{P}(Y_i(t) \in du \mid Y_i(t_1), \dots, Y_i(t_l)) = \begin{cases} \mathbb{P}(Y_i(t) \in du \mid Y_i(t_1)) & \text{if } 0 < t \leq t_1, \\ \mathbb{P}(Y_i(t) \in du \mid Y_i(t_j), Y_i(t_{j+1})) & \text{if } t_j < t \leq t_{j+1}, \\ \mathbb{P}(Y_i(t) \in du \mid Y_i(t_l)) & \text{if } t_l \leq t. \end{cases} \quad (24)$$

We only need to generate linear transformations of Beta or Gamma random variables. Importantly, expression (24) implies that for generating $Y_i(\alpha(\mathbb{X}) + L_{-j,2}^i)$ and $Y_i(\alpha(\mathbb{X}) + V_{-j,2}^i)$ given the previously observed components of (Φ, ∇, Y) we only need the values of $L_{-j,2}^i, V_{-j,2}^i, L_{-j+1,2}^i, V_{-j+1,2}^i, Y_i(\alpha(\mathbb{X}) + L_{-j+1,2}^i)$ and $Y_i(\alpha(\mathbb{X}) + V_{-j+1,2}^i)$, indeed, the monotonicity condition guarantees that

$$\dots V_{-j+2,2}^i \leq V_{-j+1,2}^i \leq V_{-j,2}^i \leq L_{-j,2}^i \leq L_{-j+1,2}^i \leq L_{-j+2,2}^i \leq \dots$$

We conclude this section with a few examples representative of the computational efficiency of the proposed algorithm. To evaluate the computational efficiency we considered some alternative scenarios and sampled iteratively from the target distribution π . Figure 1 shows Monte Carlo approximations of the distributions of the times to coalescence C of the algorithm for nine different scenarios. The first line of Fig. 1 represents three different Lévy intensity measures; each graph illustrates a different map $s \rightarrow \rho(ds)\alpha(\mathbb{X})$. The second line of Fig. 1 emphasizes that the three alternative probability models are representative of remarkably different *a priori* beliefs on the unknown discrete distribution \tilde{p} ; each graph shows the expected values of the 12 largest normalized jumps of \tilde{p} . The expected values have been computed using the Monte Carlo method. The third line illustrates the observed times to coalescence when the sample size is $n = 100, k = 10$ and $n_1 = n_1 = \dots = n_{10}$; each histogram is based on 100 iterations of the algorithm. The fourth line represents the times to coalescence when $k = 4, n_1 = 40, n_2 = 30, n_3 = 20$ and $n_4 = 10$. The fifth line considers $k = 7, n_1 = 70$ and $n_2 = n_3 = \dots = n_7 = 5$.

Table 1 illustrates the computational times in seconds necessary for sampling 100 independent variables from the target distribution π . We consider again three alternative models; for all the three models $\tau = \alpha(\mathbb{X}) = 1$ and $\gamma = 9$. The 1st model is parameterized by $\beta_i = 1/(i + 1)!$ for $i = 0, \dots, 9$, the 2nd one by $\beta_i = 0.3/i!$ for $i = 0, \dots, 9$ and the 3rd by $\beta_i = 0.1(i + 1)/i!$ for every $i = 0, \dots, 9$. Table 1 gives the computational times that were necessary for sampling 100 variables from the target distribution in the nine scenarios that we considered. The algorithm has been implemented using the R language.

Table 1. Computational times in seconds for sampling 100 variables from π

Frequencies	Model 1	Model 2	Model 3
$k = 10; n_1 = \dots = n_{10} = 1000$	13.8	11.6	8.9
$k = 4; n_1 = 4000, n_2 = 3000, n_3 = 2000, n_4 = 1000$	4.5	3.8	3.1
$k = 7; n_1 = 7000, n_2 = \dots = n_7 = 500$	4.5	4.2	3.5

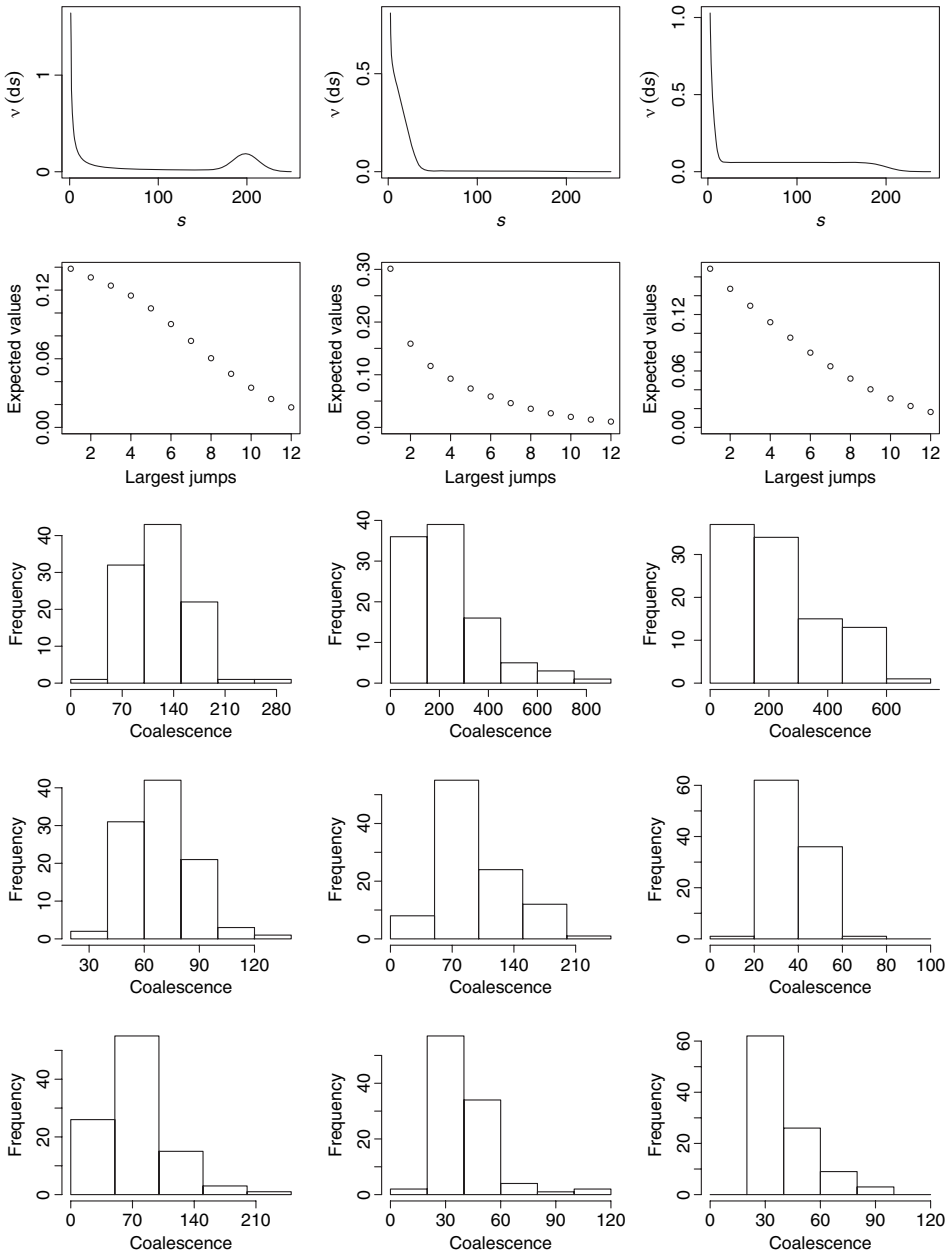


Fig. 1. Times to coalescence for nine alternative scenarios.

Acknowledgements

The authors are grateful to an Associate Editor and two anonymous Referees for their valuable comments and suggestions. Stefano Favaro is partially supported by the Italian Ministry of University and Research, grant 2008MK3AFZ. Lorenzo Trippa has been supported by an American-Italian Cancer Foundation Post-Doctoral Research Fellowship.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Proofs for propositions 1, 2, 3, 4, 5 and 6.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

- Bell, E. T. (1927). Partition polynomials. *Ann. Math.* **29**, 38–46.
- Comtet, L. (1974). *Advanced combinatorics*. D. Reidel Publishing Company, Boston.
- Dalal, S. R. & Hall, W. J. (1983). Approximating priors by mixtures of natural conjugate priors. *J. Roy. Statist. Soc. Ser. B* **45**, 278–286.
- Dunson, D. B. (2008). *Nonparametric Bayes applications to biostatistics*. Technical Report 06, Duke University Department of Statistical Science, Durham, NC.
- Favaro, S., Prünster, I. & Walker, S. G. (2010). On a class of random probability measures with general predictive structure. *Scand. J. Statist.* **38**, 359–376.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.
- Gnedin, A. & Pitman, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S. Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **325**, 83–102.
- Ho, M. W., James, L. F. & Lau, J. W. (2007). Gibbs Partitions (EPPF's) Derived From a Stable Subordinator are Fox H and Meijer G Transforms. Available on <http://arxiv.org/abs/0708.0619>.
- James, L. F. (2003). A simple proof of the almost sure discreteness of a class of random measures. *Statist. Probab. Lett.* **65**, 363–368.
- James, L. F., Lijoi, A. & Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Statist.* **33**, 105–120.
- James, L. F., Lijoi, A. & Prünster, I. (2009). Posterior analysis of normalized random measure with independent increments. *Scand. J. Statist.* **36**, 76–97.
- Kingman, J. F. C. (1967). Completely random measures. *Pacific. J. Math.* **21**, 59–78.
- Kingman, J. F. C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B* **37**, 1–22.
- Kingman, J. F. C. (1993). *Poisson processes*. Oxford University Press, Oxford.
- Lijoi, A. & Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian nonparametrics* (eds. N. L. Hjort, C. C. Holmes, P. Müller & S. G. Walker), 80–136. Cambridge University Press, Cambridge.
- Lijoi, A., Mena, R. H. & Prünster, I. (2005a). Bayesian Nonparametric analysis for a generalized Dirichlet process prior. *Stat. Inference Stoch. Process.* **8**, 283–309.
- Lijoi, A., Mena, R. H. & Prünster, I. (2005b). Hierarchical mixture modelling with normalized inverse-Gaussian priors. *J. Amer. Stat. Assoc.* **100**, 1278–1291.
- Lijoi, A., Mena, R. H. & Prünster, I. (2007a). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B* **69**, 715–740.
- Lijoi, A., Mena, R. H. & Prünster, I. (2007b). Bayesian nonparametric estimation of the probability of discovering a new species. *Biometrika* **94**, 769–786.
- Lijoi, A., Mena, R. H. & Prünster, I. (2007c). A Bayesian nonparametric method for prediction in EST analysis. *BMC Bioinformatics* **8**, 339.
- Lijoi, A., Prünster, I. & Walker, S. G. (2008a). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18**, 1519–1547.
- Lijoi, A., Prünster, I. & Walker, S. G. (2008b). Investigating nonparametric priors with Gibbs structure. *Stat. Sin.* **18**, 1653–1668.
- Müller, P. & Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statist. Sci.* **19**, 95–110.
- Müller, P. & Quintana, F. A. (2010). More nonparametric Bayesian models for biostatistics. In *Bayesian nonparametrics* (eds. N. L. Hjort, C. C. Holmes, P. Müller & S. G. Walker), 277–294. Cambridge University Press, Cambridge.
- Nieto-Barajas, L. E., Prünster, I. & Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Statist.* **32**, 2343–2360.
- Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *PTRF* **102**, 145–158.

Pitman, J. (2006). *Combinatorial stochastic processes*. Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture Notes in Mathematics N. 1875. Springer, New York.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Rand. Str. & Alg.* **9**, 223–252.

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

Zabell, S. L. (1982). W. E. Johnson's 'sufficiency' postulate. *Ann. Statist.* **10**, 1090–1099.

Received June 2010, in final form April 2011

Lorenzo Trippa, Department of Biostatistics, DFCI Biostats & Comp Biology, 44 Binney Street, Boston, Massachusetts 02115, USA.

E-mail: ltrippa@jimmy.harvard.edu

Appendix

Bell Polynomials

The partition polynomials, introduced by Bell (1927), have found many applications in combinatorics, probability theory and statistics. A particular type of partition polynomials are the so-called Bell polynomials (see Comtet, 1974).

Definition 4. Let $w_\bullet := \{w_i, i \geq 1\}$ be a sequence of real numbers. Then the (n, k) th partial Bell polynomial $B_{n,k}(w_\bullet)$ is defined by the expansion

$$\exp\{xw(t)\} = \sum_{n=0}^{+\infty} \sum_{k=0}^{+\infty} B_{n,k}(w_\bullet) x^k \frac{t^n}{n!}$$

where $w(t)$ is the exponential generating function of the sequence w_\bullet and $w_0 = w(0) = 0$.

From definition 4 it is possible to isolate $B_{n,k}(w_\bullet)$ by differentiating the appropriate number of times and then setting $x = t = 0$, i.e.

$$B_{n,k}(w_\bullet) = \left. \frac{\partial^n}{\partial t^n} \frac{1}{k!} \frac{\partial^k}{\partial x^k} \exp\{xw(t)\} \right|_{x=0, t=0}$$

for all $n \geq 0$ and $k \geq 0$. This shows that $B_{n,k}(w_\bullet)$ corresponds to the n th Taylor coefficient of $(1/k!)w^k(t)$ or $w^k(t)/k! = \sum_{n=0}^{+\infty} B_{n,k}(w_\bullet) t^n/n!$. By setting $k = 0$ one gets $B_{0,0} = 1$ and $B_{n,0} = 0$ for $n \geq 1$, whereas for $k = 1$ one has $B_{n,1} = w_n$ for all $n \geq 0$. Also, since $w_0 = 0$, one has

$$\frac{1}{k!} w^k(t) = \frac{1}{k!} \left(w_1 t + w_2 \frac{t^2}{2!} + \dots \right)^k = w_1^k \frac{t^k}{k!} + \dots \tag{25}$$

so that $B_{n,k}(w_\bullet) = 0$ whenever $k > n$ and $B_{n,n}(w_\bullet) = w_1^n$ for all $n \geq 0$. Partial Bell polynomials can be computed by resorting to the following recurrence relation

$$B_{n,k}(w_\bullet) = \sum_{m=1}^{n-k+1} \binom{n}{m} w_m B_{n-m,k-1}(w_\bullet).$$