Scandinavian Journal of Statistics, Vol. 36: 76–97, 2009
doi: 10.1111/j.1467-9469.2008.00609.x

Posterior Analysis for Normalized Random Measures with Independent Increments

LANCELOT F. JAMES

Department of Information and Systems 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

IGOR PRÜNSTER

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

ABSTRACT. One of the main research areas in Bayesian Nonparametrics is the proposal and study of priors which generalize the Dirichlet process. In this paper, we provide a comprehensive Bayesian non-parametric analysis of random probabilities which are obtained by normalizing random measures with independent increments (NRMI). Special cases of these priors have already shown to be useful for statistical applications such as mixture models and species sampling problems. However, in order to fully exploit these priors, the derivation of the posterior distribution of NRMIs is crucial: here we achieve this goal and, indeed, provide explicit and tractable expressions suitable for practical implementation. The posterior distribution of an NRMI turns out to be a mixture with respect to the distribution of a specific latent variable. The analysis is completed by the derivation of the corresponding predictive distributions and by a thorough investigation of the marginal structure. These results allow to derive a generalized Blackwell–MacQueen sampling scheme, which is then adapted to cover also mixture models driven by general NRMIs.

Key words: Bayesian Nonparametrics, Dirichlet process, normalized random measure, Poisson random measure, posterior distribution, predictive distribution

1. Introduction

The starting problem in Bayesian non-parametric inference is the definition of a prior distribution on the space of all probability measures. After the introduction of the Dirichlet process by Ferguson (1973), various approaches for constructing random probability measures, whose distribution acts as a non-parametric prior, have been undertaken with the aim of overcoming some of the drawbacks of the Dirichlet process (see Müller and Quintana, 2004, for a recent review). In the present paper, we focus on priors derived by a suitable normalization procedure. To this end, it is worth recalling that the Dirichlet process can be defined by normalizing the increments of a Gamma process (see Ferguson, 1973). Indeed, the idea of constructing random probability measures by means of a normalization procedure has been exploited and developed in a variety of contexts not closely related to Bayesian inference. An early example is Kingman (1975), where a random discrete distribution generated by the stable subordinator is considered in connection with optimal storage problems. Other interesting applications of the 'normalization' approach can be found in various areas such as computer science, population genetics, statistical physics, excursion theory, combinatorics and number theory. Further details and references on this are found in Pitman (2006).

Even though the analysis of Kingman (1975) is developed without any reference to possible implications for Bayesian inference, these are effectively pointed out by A.F.M. Smith in the discussion of Kingman (1975): '… Ferguson's Dirichlet process is a special case of a rather more general class of processes. The question of interest to a Bayesian statistician is whether there are any other processes in this class which are tractable'. In Regazzini *et al.* (2003) the class of normalized random measures with independent increments (NRMI) is formally introduced as the normalization of suitably time-changed independent increment processes and distributional results for their means derived: this work shows that, at least in terms of means, such processes are indeed tractable (see also James, 2002). In Lijoi *et al.* (2005) attention is focused on a special case of NRMI, namely the normalized-inverse Gaussian (N-IG) process: the quantities relevant for its implementation in the context of mixture models are derived and it is shown that such a prior exhibits an interesting and useful clustering behaviour, quite different from that of the Dirichlet process. The N-IG process is then embedded in a larger subclass of NRMI in Lijoi *et al.* (2007a), which allows for an additional parameter which greatly influences the clustering structure. Special NRMIs turn out to be useful also in relation to species sampling problems, in particular, for the analysis of expressed sequence tags (ESTs) in genomics as shown in Lijoi *et al.* (2007b). In order to both understand better the structural properties of and go beyond the specific processes dealt with in the above mentioned papers, it is clear that the knowledge of the posterior distribution of an NRMI is required. Here we fill this gap and provide a complete and implementable description of the posterior distribution: this addresses the issue of tractability raised by A. F. M. Smith, which in a Bayesian setting, necessarily coincides with the tractability of the posterior distribution.

Before proceeding, the important contributions in Perman *et al.* (1992), Pitman & Yor (1997) and Pitman (2003) related to Kingman's construction, albeit not directly in Bayesian Nonparametrics, are to be noted. In Pitman (1996, 2003) a thorough analysis of the two parameter Poisson-Dirichlet family, which can be generated by a stable subordinator, is provided. The utility of this family for Bayesian mixture models is discussed in Ishwaran & James (2001, 2003).

1.1. Preliminaries

The results achieved in the paper are heavily based on the notion of completely random measure. Hence, it is worth providing a brief preliminary description of the main concepts involved in the next sections.

For any topological space $\mathcal{T}$, $\mathscr{B}(\mathcal{T})$ will denote the Borel σ-field of subsets of $\mathcal{T}$. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be some probability space and $\mathbb{X}$ be complete, separable and endowed with a metric $d_{\mathbb{X}}$. Define on $(\Omega, \mathscr{F}, \mathbb{P})$ a Poisson random measure $\tilde{N}$ on $\mathbb{S} = \mathbb{R}^+ \times \mathbb{X}$ with intensity measure ν. This means that

(i) for any C in $\mathscr{B}(\mathbb{S})$ such that $\nu(C) = \mathbb{E}[\tilde{N}(C)] < \infty$, the probability distribution of the random variable $\tilde{N}(C)$ is Poisson($\nu(C)$);
(ii) for any finite collection of pairwise disjoint sets, $A_1, \dots, A_k$, in $\mathscr{B}(\mathbb{S})$, the random variables $\tilde{N}(A_1), \dots, \tilde{N}(A_k)$ are mutually independent.

Moreover, the measure ν must satisfy the following conditions,

$$\int_{(0,1)} s\nu(\mathrm{d}s, \mathbb{X}) < \infty, \quad \nu([1, \infty) \times \mathbb{X}) < \infty.$$

We refer to Daley & Vere-Jones (1988) for an exhaustive account on Poisson random measures.

If $(\mathbb{M}, \mathscr{B}(\mathbb{M}))$ is the space of boundedly finite measures on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$, denote by $\tilde{\mu}$ a random element defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and with values in $(\mathbb{M}, \mathscr{B}(\mathbb{M}))$ which can be represented as a linear functional of the Poisson random measure $\tilde{N}$, with intensity v, as follows,

$$\tilde{\mu}(B) = \int_{\mathbb{R}^+ \times B} s\tilde{N}(\mathrm{d}s, \mathrm{d}x) \quad \forall B \in \mathscr{B}(\mathbb{X}). \tag{1}$$

It can be easily seen from the properties of $\tilde{N}$ that $\tilde{\mu}$ is, in the terminology of Kingman (1967), a *completely random measure* on $\mathbb{X}$, i.e. for any collection of disjoint sets in $\mathscr{B}(\mathbb{X})$, $A_1, A_2, \ldots$, the random variables $\tilde{\mu}(A_1)$, $\tilde{\mu}(A_2), \ldots$ are mutually independent and $\tilde{\mu}(\cup_{j \geq 1} A_j) = \sum_{j \geq 1} \tilde{\mu}(A_j)$ holds true a.s.-$\mathbb{P}$. It is well known that $\tilde{\mu}$ is uniquely characterized by its *Laplace functional*

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}(\mathrm{d}x)}\right] = \mathrm{e}^{-\int_{\mathbb{S}}[1-\mathrm{e}^{-sh(x)}]v(\mathrm{d}s,\, \mathrm{d}x)},$$

where $h : \mathbb{X} \to \mathbb{R}^+$ is a measurable function. For a proof of such a representation, see theorem 2 in Kingman (1967). Details and further references on completely random measures can be found in Kingman (1993).

From this preliminary illustration, it is apparent that both the Poisson random measure $\tilde{N}$ and the completely random measure $\tilde{\mu}$ are identified by the corresponding intensity measure v. This suggests a simple and useful distinction of the random measures we deal with according to the decomposition of v. Letting H be a non-atomic and σ-finite measure on $\mathbb{X}$, we have:

(a) if $v(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)H(\mathrm{d}x)$, for some measure ρ on $\mathbb{R}^+$, we say that the corresponding $\tilde{N}$ and $\tilde{\mu}$ are *homogeneous*;

(b) if $v(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s \,|\, x)H(\mathrm{d}x)$, where $\rho : \mathscr{B}(\mathbb{R}^+) \times \mathbb{X} \to \mathbb{R}^+$ is a kernel, i.e. $x \mapsto \rho(C \,|\, x)$ is $\mathscr{B}(\mathbb{X})$ measurable for any $C \in \mathscr{B}(\mathbb{R}^+)$ and $\rho(\cdot \,|\, x)$ is a σ-finite measure on $\mathscr{B}(\mathbb{R}^+)$ for any x in $\mathbb{X}$, we say that the corresponding $\tilde{N}$ and $\tilde{\mu}$ are *non-homogeneous*.

Recall that in our framework v always admits a disintegration as in (b); this follows, e.g. from theorem 15.3.3 in Kallenberg (1986).

Remark 1. Note that the construction which led us to define a random measure via (1) can be extended by considering more general linear functionals of the Poisson measure $\tilde{N}$. For example, James (2002), using an approach closely connected to Perman *et al.* (1992), considers the so-called h-biased random measures, that is $\int_{S \times \mathbb{X}} h(s)\tilde{N}(\mathrm{d}s,\, \mathrm{d}x)$, where $h : S \to \mathbb{R}^+$ and S is any complete and separable metric space. The results we provide in the next sections can also be extended to h-biased random measures.

1.2. Construction of NRMI

Since the aim is to define random probability measures by means of normalization of completely random measures, the total mass $T := \tilde{\mu}(\mathbb{X})$ needs to be finite and positive, almost surely. This happens if $v(\mathbb{S}) = +\infty$ and the Laplace exponent

$$\psi(\lambda) := \int_{\mathbb{S}} \left[1 - \mathrm{e}^{-\lambda s}\right] v(\mathrm{d}s,\, \mathrm{d}x) \tag{2}$$

is finite for any positive λ. A proof of this fact can be found, e.g. in Regazzini *et al.* (2003, p. 563 and proposition 1, respectively). When these conditions hold true, a normalized random measure with independent increments (NRMI) on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is given by

$$\tilde{P}(\cdot) = \frac{\tilde{\mu}(\cdot)}{T}. \tag{3}$$

Note that, when $\mathbb{X}=\mathbb{R}$, this definition coincides with the one given in Regazzini *et al.* (2003) in terms of increasing additive processes. Indeed, it is worth remarking that an increasing additive process can always be seen as the càdlàg distribution function induced by a completely random measure on $\mathbb{R}$. Moreover, as shown in James (2003), NRMIs select, almost surely, discrete distributions. Before proceeding, we recall that T is assumed to be a random variable whose distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and denote its density as f_T. Such a regularity assumption allows to avoid some technical difficulties and it is commonly adopted in this framework. The interested reader is referred to section 3 of Pitman (2003) for further details.

It is worth noting that some priors that are used in Bayesian non-parametric inference can be defined as in (3). For instance, consider the Dirichlet process with parameter measure $H=\theta P_0$. Then, as already noted by Ferguson (1973), such a prior can be recovered by considering a Gamma random measure with Laplace functional

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}(\mathrm{d}x)}\right]=\mathrm{e}^{-\theta\int_{\mathbb{S}}\left[1-\mathrm{e}^{-sh(x)}\right]\frac{\mathrm{e}^{-s}}{s}\,\mathrm{d}sP_0(\mathrm{d}x)}=\mathrm{e}^{-\theta\int_{\mathbb{X}}\log[1+h(x)]P_0(\mathrm{d}x)}$$

for any $h:\mathbb{X}\to\mathbb{R}^+$ such that $\int\log[1+h(x)]P_0(\mathrm{d}x)<\infty$. Other examples are the normalized stable process (Kingman, 1975); the normalized inverse-Gaussian process (Lijoi *et al.*, 2005); the generalized Gamma process (James, 2002; Lijoi *et al.*, 2007a). It is interesting to note that the two latter models as well as the two parameter Poisson–Dirichlet process are derivable from a stable subordinator by a change of measure (see Pitman, 2003).

We close this subsection by pointing out that $\tilde{P}$ in (3) admits a series representation of the kind $\Sigma_{i\geq1}\tilde{p}_i\delta_{X_i}(\cdot)$, where δ_x denotes the point mass at x. The most notable example is the Sethuraman (1994) representation of the Dirichlet process. In the case of a general NRMI, if the underlying intensity ν is homogeneous, then the weights $\tilde{p}_i$s are independent from the locations X_i and $\tilde{P}$ is a species sampling model (see Pitman, 1996, 2003). On the other hand, when ν is non-homogeneous, the weights and the locations are no longer independent and $\tilde{P}$ is not a species sampling model.

1.3. Outline of the paper

In this paper we consider Bayesian inference by exploiting the law of an NRMI as a non-parametric prior distribution. Under the usual assumption of exchangeability of the observation process, we derive in section 2 a representation for the posterior distribution of $\tilde{P}$ in terms of a mixture with respect to the distribution of a suitable latent variable. In section 3 we determine the prediction rule and thoroughly study the marginal distribution of the observations. Relying on these results, a generalization of the Blackwell–MacQueen sampling scheme is also provided. In section 4 the results are adapted to cover mixture models driven by NRMIs and the corresponding simulation algorithm is described in detail. Finally, section 5 provides some concluding remarks. In order to ease the flow of ideas, proofs are given in the Appendix.

2. Posterior distributions for NRMIs

In this section we aim at deriving a tractable expression for the posterior distribution of an NRMI. This represents a challenging issue since, with the exception of the Dirichlet process, NRMIs are not conjugate as shown in James *et al.* (2006). Indeed, apart from its simplicity and ease of interpretation, the popularity of the Dirichlet process is also due to its conjugacy property which makes posterior inferences more tractable from an analytic point of view. However, we are able to show that, conditional on a specific latent variable, the posterior distribution of an NRMI coincides with the distribution of another NRMI having a rescaled intensity and fixed points of discontinuity. This can be seen as a kind of conditional conjugacy.

Let us first introduce a sequence $(X_n)_{n\geq 1}$ of exchangeable observations defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and with values in $\mathbb{X}$ in such a way that, given $\tilde{P}$, the X_is are i.i.d. with distribution $\tilde{P}$, i.e.

$$\mathbb{P}[X_1 \in C_1, \ldots, X_n \in C_n \mid \tilde{P}] = \prod_{i=1}^{n} \tilde{P}(C_i). \tag{4}$$

Moreover, set $\boldsymbol{X} = (X_1, \ldots, X_n)$. It is clear that one can always represent $\boldsymbol{X}$ as $(\boldsymbol{Y}, \boldsymbol{\pi})$, where $\boldsymbol{Y} = (Y_1, \ldots, Y_{n(\boldsymbol{\pi})})$ denotes the distinct observations within the sample and $\boldsymbol{\pi}$ stands for a partition of $\{1, \ldots, n\}$ of size $n(\boldsymbol{\pi})$ recording which observations within the sample are equal. The number of elements in the jth set of the partition is indicated by n_j, for $j = 1, \ldots, n(\boldsymbol{\pi})$, so that $\Sigma_{j=1}^{n(\boldsymbol{\pi})} n_j = n$. The partition mechanism is ideally suited to carry out posterior analysis when data contain ties: this is certainly the case for discrete random probability measures and thus, in particular, for NRMIs.

Before stating the main theorem, we define a positive random variable U_n as follows. Let Γ_n be a Gamma random variable with scale parameter 1 and shape parameter n, which is independent from the total mass T. Then, set $U_n = \Gamma_n / T$. It is immediate to show that, for any $n \geq 1$, the density function of U_n is given by

$$f_{U_n}(u) = \frac{u^{n-1}}{\Gamma(n)} \int_{\mathbb{R}^+} t^n \, \mathrm{e}^{-ut} f_T(t) \, \mathrm{d}t, \tag{5}$$

where f_T is the density function of T. It will be shown that the posterior distribution of U_n, given $\boldsymbol{X}$, is of great importance for our analysis.

Proposition 1
Let $\tilde{P}$ be an NRMI. Then, the conditional distribution of U_n, given $\boldsymbol{X}$, admits a density function coinciding with

$$f_{U_n}^{\boldsymbol{X}}(u) \propto u^{n-1} \prod_{i=1}^{n(\boldsymbol{\pi})} \tau_{n_i}(u \mid Y_i) \, \mathrm{e}^{-\psi(u)},$$

where $\tau_{n_i}(u \mid Y_i) = \int_{\mathbb{R}^+} s^{n_i} \, \mathrm{e}^{-us} \rho(\mathrm{d}s \mid Y_i)$ for $i = 1, \ldots, n(\boldsymbol{\pi})$.

Even though its proof is based on the result of the next theorem 1, it is worth introducing it in advance because of the key role played by this latent random variable U_n for developing the posterior analysis of NRMIs. In what follows, for any pair of random elements Z and W defined on $(\Omega, \mathscr{F}, \mathbb{P})$, we use the symbol $Z^{(W)}$ to denote a random element on $(\Omega, \mathscr{F}, \mathbb{P})$ whose distribution coincides with a regular conditional distribution of Z, given W. Let us provide the main result concerning a posterior characterization of the completely random measure itself.

Theorem 1
Let $\tilde{P}$ be an NRMI with intensity $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s \mid x) H(\mathrm{d}x)$. Then

$$\tilde{\mu}^{(U_n, \boldsymbol{X})} \stackrel{d}{=} \tilde{\mu}^{(U_n)} + \sum_{i=1}^{n(\boldsymbol{\pi})} J_i^{(U_n, \boldsymbol{X})} \delta_{Y_i},$$

where

(i) $\tilde{\mu}^{(U_n)}$ *is a completely random measure with intensity*

$$\nu^{(U_n)}(\mathrm{d}s, \mathrm{d}x) = \mathrm{e}^{-U_n s}\rho(\mathrm{d}s \mid x)H(\mathrm{d}x),$$

(ii) Y_i*, for* $i = 1, \ldots, n(\boldsymbol{\pi})$*, are the fixed points of discontinuity and the* $J_i^{(U_n, \boldsymbol{X})}$*s are the corresponding jumps whose density is proportional to* $s^{n_i}\,\mathrm{e}^{-U_n s}\rho(\mathrm{d}s \mid Y_i)$*,*

(iii) $\tilde{\mu}^{(U_n)}$ *and* $J_i^{(U_n, \boldsymbol{X})}$ $(i = 1, \ldots, n(\boldsymbol{\pi}))$ *are independent.*

Given the importance of theorem 1, we provide two alternative proofs in the Appendix, which rely on different general techniques for deriving posterior distributions. The first works with the underlying Poisson random measure, which constitutes the core of many discrete random measures, and is due to James (2002, 2005a). The second proof exploits the approach set forth in Prünster (2002) and works directly at the level of the completely random measure $\tilde{\mu}$.

The result in theorem 1 sheds some light on the deep structure of the random measures at issue. It essentially shows that, given some latent variable, a posteriori $\tilde{\mu}$ is still a completely random measure with fixed points of discontinuity corresponding to the locations of the observations. The reader may note that this characterization is somehow reminiscent of the posterior characterization of neutral to the right priors provided by Ferguson (1974). Recall that the class of neutral to the right priors, introduced in Doksum (1974) and of great popularity in the context of survival analysis, is defined via an exponential transformation of increasing additive processes. Indeed, Ferguson's characterization studies the posterior distribution of the increasing additive process (instead of its transformation) and identifies it as a process with updated Poisson intensity and with fixed points of discontinuity at the location of the observations (see also Hjort, 1990; Walker & Muliere, 1997; Kim, 1999; James, 2006). Besides the analogy, it is worth remarking two substantial differences. The first is due to the non-conjugacy of NRMIs: in contrast to the neutral to the right case, here we first have to identify an appropriate latent variable and then, conditionally on it, determine a posterior characterization of $\tilde{\mu}$. The second is due to the type of transformation of $\tilde{\mu}$ employed for defining the random probability measures: NRMIs are obtained via normalization while neutral to the right measures via an exponential transformation. This clearly affects the updating mechanism of the intensity measure and the distribution of the jumps which are very different. The previous result is also essential for deriving the posterior distribution for the class of NRMIs. In the following, by posterior distribution of $\tilde{P}$, given U_n, we always refer to the distribution of $\tilde{P}$ given the data $\boldsymbol{X}$ and U_n.

Theorem 2

If $\tilde{P}$ *is an NRMI with intensity* $\nu(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s \mid x)H(\mathrm{d}x)$*, then the posterior distribution of* $\tilde{P}$*, given* U_n*, is again an NRMI (with fixed points of discontinuity). In particular, it coincides in distribution with the random measure*

$$w\,\frac{\tilde{\mu}^{(U_n)}}{T^{(U_n)}} + (1-w)\,\frac{\sum_{i=1}^{n(\boldsymbol{\pi})} J_i^{(U_n,\boldsymbol{X})}\delta_{Y_i}}{\sum_{i=1}^{n(\boldsymbol{\pi})} J_i^{(U_n,\boldsymbol{X})}},$$

where $T^{(U_n)} = \tilde{\mu}^{(U_n)}(\mathbb{X})$*,* $w = T^{(U_n)}\{T^{(U_n)} + \sum_{i=1}^{n(\boldsymbol{\pi})} J_i^{(U_n,\boldsymbol{X})}\}^{-1}$*. The distributions of* $\tilde{\mu}^{(U_n)}$ *and* $J_i^{(U_n,\boldsymbol{X})}$ $(i = 1, \ldots, n(\boldsymbol{\pi}))$ *and the distribution of* U_n*, given* $\boldsymbol{X}$*, are those specified in theorem 1.*

We close the present section by introducing two examples of NRMIs; thus, pointing out how the results obtained so far can be applied in order to determine the posterior distributions. It is worth remarking that other examples can be easily obtained by simply plugging into theorem 1 any Poisson intensity leading to a well-defined NRMI (3).

Example 1. We first consider an NRMI based on the homogeneous intensity measure

$$\nu\,(\mathrm{d}s, \mathrm{d}x) = \frac{1}{\sqrt{2\pi}} \frac{\mathrm{e}^{-\frac{1}{2}s}}{s^{\frac{3}{2}}} \mathrm{d}s H(\mathrm{d}x). \tag{6}$$

Since $\nu(\mathbb{R}^+ \times \mathbb{X}) = \infty$, then T is positive almost surely and finiteness of (2) is equivalent to requiring H to be a finite measure. Hence, H can be represented as $H = \theta P_0$, where $\theta > 0$ and P_0 is a probability distribution on $\mathbb{X}$. The resulting prior $\tilde{P}$, obtained through (3), is also known as normalized inverse-Gaussian (N-IG) process. Note that for this process a description of the family of finite-dimensional distributions has been provided in Lijoi *et al.* (2005). Here, based on theorem 1, we provide a characterization of the posterior distribution of this useful prior. It can be easily checked that $\tau_j(u \mid x) = \tau_j(u) = 2^{j-1}\Gamma(j-\frac{1}{2})(\sqrt{\pi}[2u+1]^{j-1/2})^{-1}$, for any $j \geq 1$. Moreover, $\psi(u) = \theta(\sqrt{2u+1} - 1)$. From proposition 1, one then gets

$$f^{X}_{U_n}(u) \propto \frac{u^{n-1}\,\mathrm{e}^{-\theta\sqrt{2u+1}}}{(2u+1)^{n-n(\boldsymbol{\pi})/2}}.$$

Given U_n, the posterior distribution of $\tilde{\mu}$ coincides with the distribution of $\tilde{\mu}^{(U_n)} + \sum_{i=1}^{n(\boldsymbol{\pi})} J_i^{(U_n,X)} \delta_{Y_i}$ where $\tilde{\mu}^{(U_n)}$ is a completely random measure with intensity

$$\frac{1}{\sqrt{2\pi}} \frac{\mathrm{e}^{-s\left(\frac{1}{2}+U_n\right)}}{s^{\frac{3}{2}}} \mathrm{d}s H(\mathrm{d}x),$$

and the jumps $J_i^{(U_n,X)}$ are Gamma distributed with scale parameter $U_n + 1/2$ and shape parameter $n_i - 1/2$, for $i = 1, \dots, n(\boldsymbol{\pi})$. By replacing (6) with the intensity corresponding to generalized Gamma random measures, which include the inverse Gaussian process as a special case, one obtains the class of NRMI considered in Lijoi *et al.* (2007a) and applied, within a hierarchical model, to clustering problems. Theorem 1 allows to derive their posterior distribution in a straightforward way.

Example 2. Let us now consider an NRMI based on the non-homogeneous intensity

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{e}^{-\beta(x)s}}{s} \mathrm{d}s H(\mathrm{d}x),$$

where $\beta : \mathbb{X} \to \mathbb{R}^+$. Dykstra & Laud (1981) discussed such a random measure for the case $\mathbb{X} = \mathbb{R}$ and termed it *extended Gamma process* with parameters (H, β). This model described on more abstract spaces is discussed in Lo (1982) and is termed a *weighted Gamma process*. Much attention has been paid to extended Gamma processes in the Bayesian literature, with particular emphasis on problems related to survival analysis. In order to exploit the extended Gamma process for defining an NRMI, we need to ensure that T is positive and finite almost surely. Since $\nu(\mathbb{S}) = \infty$, positiveness follows. Moreover, finiteness is equivalent to the requirement that H and β are such that $\int_{\mathbb{X}} \log(1 + \lambda\beta(x)^{-1}) H\,(\mathrm{d}x) < \infty$, for every $\lambda \geq 0$. Given these, the corresponding *extended Gamma NRMI* with parameter (H, β) is well defined. Since $\tau_j(u \mid x) = \Gamma(j)[\beta(x) + u]^{-j}$, for any $j \geq 1$ and x in $\mathbb{X}$, and $\psi(u) = \int_{\mathbb{X}} \log[\beta(x) + u] H(\mathrm{d}x)$, from proposition 1 it is possible to deduce that

$$f^{X}_{U_n}(u) \propto u^{n-1} \exp\left\{ -\int_{\mathbb{X}} \log[\beta(x) + u] H^*\,(\mathrm{d}x) \right\}, \tag{7}$$

where $H^*(\cdot) = H(\cdot) + \sum_{i=1}^{n(\boldsymbol{\pi})} n_i \delta_{Y_i}(\cdot)$. As for the posterior distribution, by theorem 1 one has that, conditionally on U_n, the posterior distribution of $\tilde{\mu}$ coincides with the distribution of the sum of an extended Gamma process with parameter $(H, \beta + U_n)$ and $n(\boldsymbol{\pi})$ jumps corresponding to the distinct observations $\boldsymbol{Y}$. Conditionally on U_n and $\boldsymbol{X}$, the ith jump is Gamma distributed with parameters $(\beta(Y_i)+U_n, n_i)$, for $i=1,\ldots,n(\boldsymbol{\pi})$. Thus, for any function $h:\mathbb{X}\to\mathbb{R}^+$ such that $\int_{\mathbb{X}}\log[h(x)+\beta(x)]H(\mathrm{d}x)<\infty$,

$$\mathbb{E}[\mathrm{e}^{-\int_{\mathbb{X}}h(x)\tilde{\mu}(\mathrm{d}x)}\,|\,U_n,\boldsymbol{X}]=\mathrm{e}^{-\int_{\mathbb{X}}\log[h(x)+\beta(x)+U_n]H(\mathrm{d}x)-\sum_{i=1}^{n(\boldsymbol{\pi})}n_i\log[h(Y_i)+\beta(Y_i)+U_n]}$$
$$=\mathrm{e}^{-\int_{\mathbb{X}}\log[h(x)+\beta(x)+U_n]H^*(\mathrm{d}x)}$$

and one easily concludes that the extended Gamma NRMI, given U_n and $\boldsymbol{X}$, is still an extended Gamma NRMI with parameter $(H^*,\beta+U_n)$. It is worth noting that priors based on non-homogeneous measures have always played an important role in the Bayesian non-parametric inference for modelling survival data and spatial phenomena (see, e.g. Ferguson, 1974; Lo, 1982; Hjort, 1990; Walker & Muliere, 1997; Wolpert & Ickstadt, 1998). Up to now NRMI based on non-homogeneous intensities appeared to be untractable, but thanks to theorem 1 this seems not to be the case anymore.

3. Predictive and marginal distributions

Apart from the posterior distribution, a Bayesian can be also interested in a rule for predicting future values of the observations, given those already observed, and a sampling scheme for generating observations governed by an NRMI. When $\tilde{P}$ is a Dirichlet process, with parameter measure θP_0, it is well known that the predictive distribution has the following simple form,

$$\mathbb{P}[X_{n+1}\in C\,|\,\boldsymbol{X}]=\frac{\theta}{\theta+n}P_0(C)+\frac{n}{\theta+n}\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(C) \tag{8}$$

for any C in $\mathscr{B}(\mathbb{X})$. Moreover, the marginal distribution of the observations can be expressed in terms of the celebrated Ewens sampling formula. More precisely, given that the distribution of $\boldsymbol{X}$ is characterized by the joint distribution of $(\boldsymbol{Y},\boldsymbol{\pi})$, one has that the latter coincides with

$$\left(\prod_{i=1}^{n(\boldsymbol{\pi})}P_0(\mathrm{d}Y_i)\right)\frac{\theta^{n(\boldsymbol{\pi})}}{(\theta)_n}\prod_{i=1}^{n(\boldsymbol{\pi})}\Gamma(n_i), \tag{9}$$

where $(\theta)_n=\Gamma(\theta+n)/\Gamma(\theta)$ is the Pochhammer symbol. The Ewens sampling formula is the best-known case of exchangeable partition probability function (EPPF) and it basically represents the marginal distribution of the partition $\boldsymbol{\pi}$. A detailed illustration of the EPPF concept can be found in Pitman (2006). Its role in a Bayesian context, for the homogeneous case, can be deduced from Pitman (1996) and Ishwaran & James (2003), whereas for the non-homogeneous case one can refer to James (2006). In this section we provide the analogues of (8) and (9) for the more general class of NRMIs.

3.1. The prediction rule

Once we have derived the posterior distribution of an NRMI, the determination of the corresponding predictive distributions is quite straightforward.

Proposition 2
Let $\tilde{P}$ be an NRMI with intensity $\nu(\mathrm{d}s,\mathrm{d}x)=\rho(\mathrm{d}s\,|\,x)H(\mathrm{d}x)$. Then the predictive distribution for X_{n+1} given $\boldsymbol{X}$ coincides with

$$\mathbb{P}\left[X_{n+1}\in \mathrm{d}x \,|\, \boldsymbol{X}\right]=w^{(n)}H(\mathrm{d}x)+\frac{1}{n}\sum_{j=1}^{n(\boldsymbol{\pi})} w_j^{(n)}\delta_{Y_j}(\mathrm{d}x), \tag{10}$$

where, for $j=1,\ldots,n(\boldsymbol{\pi})$,

$$w^{(n)}=\frac{1}{n}\int_{\mathbb{R}^+} u\tau_1(u\,|\,x) f_{U_n}^{\mathbf{X}}(u)\,\mathrm{d}u, \quad w_j^{(n)}=\int_{\mathbb{R}^+} u\frac{\tau_{n_j+1}(u\,|\,Y_j)}{\tau_{n_j}(u\,|\,Y_j)} f_{U_n}^{X}(u)\,\mathrm{d}u.$$

These predictive distributions have quite intuitive forms, since they consist of a linear combination of H and of a *weighted* version of the empirical distribution. Note that the prediction rule reduces to the one provided by Pitman (2003) in the homogeneous case (see also James, 2002; Prünster, 2002).

3.2. *The marginal distribution*

It is apparent from the previous results on the posterior and the predictive distributions that the use of partitions is of great help. The same can be said when facing the issue of characterizing the marginal distribution of the vector of (exchangeable) observations $\boldsymbol{X}=(X_1,\ldots,X_n)$, for any $n\geq 1$. Indeed, the marginal distribution of $\boldsymbol{X}$ can be described in terms of the distribution of $(\boldsymbol{Y},\boldsymbol{\pi})$, where, as before, $\boldsymbol{\pi}$ is a partition of the n integers $\{1,\ldots,n\}$ into $n(\boldsymbol{\pi})$ sets, $Y=(Y_1,\ldots,Y_{n(\boldsymbol{\pi})})$ is the vector of distinct values among the X_is. Note that $n(\boldsymbol{\pi})\in\{1,\ldots,n\}$ since, as was mentioned before, NRMIs select discrete distributions on $\mathbb{X}$ with probability 1. This allows us to confine ourselves to the determination of the distribution of $(Y,\boldsymbol{\pi})$. Before describing the distribution $\mathscr{M}$ of $\boldsymbol{X}$, let us introduce the following quantity,

$$\kappa_{n_j}(u)=\int_{\mathbb{X}}\tau_{n_j}(u\,|\,x)H\,(\mathrm{d}x),$$

which is the cumulant of order n_j of the conditional distribution of the total mass T, given $U_n=u$.

Proposition 3
Let $\tilde{P}$ *be an NRMI. Then the distribution of* $(\boldsymbol{Y},\boldsymbol{\pi})$ *coincides with*

$$\frac{1}{\Gamma(n)}\left\{\int_{\mathbb{R}^+} u^{n-1}\,\mathrm{e}^{-\psi(u)}\left[\prod_{j=1}^{n(\boldsymbol{\pi})}\tau_{n_j}(u\,|\,Y_j)\right]\,\mathrm{d}u\right\}\prod_{j=1}^{n(\boldsymbol{\pi})}H\,(\mathrm{d}Y_i). \tag{11}$$

Moreover, the marginal distribution of $\boldsymbol{\pi}$ *yields the EPPF and it is given by*

$$\Pi^{(n)}(\boldsymbol{\pi})=\frac{1}{\Gamma(n)}\int_{\mathbb{R}^+} u^{n-1}\,\mathrm{e}^{-\psi(u)}\left[\prod_{j=1}^{n(\boldsymbol{\pi})}\kappa_{n_j}(u)\right]\,\mathrm{d}u. \tag{12}$$

The EPPF given in (12) was first obtained by Pitman (2003). For a concrete use of the marginal distribution of the X_is, we will generally need a simpler description of $\mathscr{M}$ and of the corresponding EPPF. This can be achieved by working conditionally on the latent variable U_n. As for the EPPF, a tractable form we wish to obtain is of the kind

$$\Pi^{(n)}(\boldsymbol{\pi})=V_{n,n(\boldsymbol{\pi})}\prod_{i=1}^{n(\boldsymbol{\pi})}W_{n_i},$$

where $V_{n,n(\boldsymbol{\pi})}$ is a positive quantity not depending on the specific $(n_1,\ldots,n_{n(\boldsymbol{\pi})})$ and each W_{n_i} is a positive number depending solely on the corresponding n_i. A random partition having such an EPPF is said to be of a *Gibbs type* (see Pitman, 2006, for the notion of *infinite* and *finite*

Gibbs partitions). However, it is worth recalling that the only infinite EPPF admitting such a representation are the EPPFs derived from a Dirichlet process and those derived from a stable law of index $0<\alpha<1$ (see Pitman, 2006). Among them, we mention the two parameter Poisson–Dirichlet process and the generalized Gamma class of processes.

Now, by examining (11), an augmentation and an application of Bayes' rule makes it apparent that, for fixed $u>0$ and $\boldsymbol{\pi}$,

$$\mathbb{P}\left[Y_i \in \mathrm{d}y \,|\, U_n = u, \boldsymbol{\pi}\right] = \frac{\tau_{n_i}(u\,|\,y)H(\mathrm{d}y)}{\kappa_{n_i}(u)} =: H_{i,n}(\mathrm{d}y\,|\,u) \tag{13}$$

for any $i=1,\ldots,n(\boldsymbol{\pi})$. At this point, we can provide a characterization of $\mathscr{M}$, conditional on U_n.

Proposition 4

Let $\tilde{P}$ be an NRMI. Conditional on U_n and on the partition $\boldsymbol{\pi}$, the $n(\boldsymbol{\pi})$ distinct values $Y_1,\ldots,Y_{n(\boldsymbol{\pi})}$ among the X_is are independent and the distribution of Y_i is given by (13)*, for any $i=1,\ldots,n(\boldsymbol{\pi})$. Moreover, the conditional distribution of the random partition $\boldsymbol{\pi}$, given $U_n=u$, coincides with*

$$\Pi^{(n)}(\boldsymbol{\pi}\,|\,u) = \frac{\mathrm{e}^{-\psi(u)}\prod_{i=1}^{n(\boldsymbol{\pi})}\kappa_{n_i}(u)}{\int_{\mathbb{R}^+} t^n\,\mathrm{e}^{-ut} f_T(t)\,\mathrm{d}t}. \tag{14}$$

Hence, conditional on U_n, $\boldsymbol{\pi}$ is a finite Gibbs partition.

Note that in the homogeneous case the distinct observations are independent and identically distributed (i.i.d.) with common distribution P_0.

In the light of proposition 4, an interesting quantity to consider is the number $n(\boldsymbol{\pi})$ of distinct observations in a sample $\boldsymbol{X}$ of size n. For example, in non-parametric mixture models, $n(\boldsymbol{\pi})$ stands for the number of clusters in the sample of observations. Because of this, the literature has devoted much attention to it. In the Dirichlet case, the distribution of $n(\boldsymbol{\pi})$ has been investigated by Korwar & Hollander (1973) and exploited in the context of mixture models by Antoniak (1974) and Lo (1984), where it takes on the interpretation of prior distribution on the number of components. In Pitman (2003, 2006) this distribution is described for the case of the two parameter Poisson–Dirichlet process. More recently, the distribution of $n(\boldsymbol{\pi})$ for NIG and generalized Gamma mixture models has been studied in Lijoi *et al.* (2005, 2007a). We also refer to Lijoi *et al.* (2007b), where such distributions are used for devising a Bayesian non-parametric estimator of the discovery probability in genomics problems. In our case, using the fact that, conditionally on U_n, $\boldsymbol{\pi}$ is a finite Gibbs partition one can determine the distribution of $n(\boldsymbol{\pi})$, given U_n, as follows,

$$\mathbb{P}\left[n(\boldsymbol{\pi})=k\,|\,U_n=u\right] = \frac{\mathrm{e}^{-\psi(u)}}{\int_{\mathbb{R}^+} t^n\,\mathrm{e}^{-ut} f_T(t)\,\mathrm{d}t}\,\frac{n!}{k!}\sum_{(n_1,\ldots,n_k)}\prod_{j=1}^{k}\frac{\kappa_{n_j}(u)}{n_j!}$$

for $k=1,\ldots,n$. The sum above runs over all vectors of positive integers $(n_1,\ldots,n_k)$ such that $\sum_{i=1}^k n_i = n$.

An important related issue to consider in this setting, is the distribution of the random vector $(|\Pi_{1,n}|,\ldots,|\Pi_{n,n}|)$, where $|\Pi_{i,n}|$ denotes the number of clusters of size i. According to this definition, one obviously has $\sum_{i=1}^n |\Pi_{i,n}| = n(\boldsymbol{\pi})$ and $\sum_{i=1}^n i|\Pi_{i,n}| = n$. Combination of proposition 4 and of formula (52) in Pitman (2006, chapter 1) yields

$$\mathbb{P}\left[|\Pi_{j,n}|=m_j, 1\le j\le n\,|\,U_n=u\right] = \frac{n!\,\mathrm{e}^{-\psi(u)}}{\int_{\mathbb{R}^+} t^n\,\mathrm{e}^{-ut} f_T(t)\,\mathrm{d}t}\prod_{j=1}^{n}\left(\frac{\kappa_j(u)}{j!}\right)^{m_j}\frac{1}{m_j!} \tag{15}$$

where $\sum_{j=1}^n m_j = k$ and $\sum_{j=1}^n jm_j = n$. Equivalently (15) is the conditional distribution, given U_n, of the number of values of $(X_1,\ldots,X_n)$ appearing one time, two times, etc. corresponding to the numbers $(m_1, \dots, m_n)$. Moreover, (15) is a generalization of the well-known Ewens sampling formula.

Remark 2. It is interesting to note that all our results conditioned on U_n, contain the known unconditional results for the Dirichlet process. This is because the Dirichlet process is independent of U_n. To see this, recall that the Dirichlet process, with total mass $\theta > 0$, corresponds to the choice of $\rho(\mathrm{d}s) = \theta s^{-1} \mathrm{e}^{-s} \mathrm{d}s$. It follows that for each j, $\kappa_j(u) = \theta(1+u)^{-j}\Gamma(j)$ and $\mathbb{E}[T^{(U_n)} \mid \boldsymbol{X}] = \mathbb{E}[T^{(U_n)}] = (1+U_n)^{-n}[\Gamma(\theta)/\Gamma(\theta+n)]$. Additionally

$$f_{U_n}^{\boldsymbol{X}}(u) := f_{U_n}(u) \propto u^{n-1}(1+u)^{-(n+\theta)},$$

that is $U_n = \Gamma_n/T$ is a Gamma–Gamma random variable independent of $\boldsymbol{X}$. Or, equivalently, $1/(1+U_n)$ is a Beta(θ, n) random variable. Hence, (15), specializes to

$$\mathbb{P}\left[|\Pi_{j,n}| = m_j, 1 \le j \le n \mid U_n = u\right] = \frac{n!}{\prod_{i=1}^{n}(\theta+i-1)} \prod_{j=1}^{n} \left(\frac{\theta}{j}\right)^{m_j} \frac{1}{m_j!}.$$

This coincides with the Ewens sampling formula derived by Ewens (1972), which is equivalent to an important result in Antoniak (1974). Finally, note that (14) becomes

$$\Pi^{(n)}(\boldsymbol{\pi} \mid u) := \frac{\theta^{n(\boldsymbol{\pi})} \prod_{j=1}^{n(\boldsymbol{\pi})}(e_j - 1)!}{\prod_{i=1}^{n}(\theta+i-1)},$$

which is the variant of Ewens sampling formula, often called the Chinese restaurant process (see Ishwaran & James, 2003; Pitman, 2006). The calculations for the Dirichlet process involving U_n may be found in James (2005b), where it is shown that U_n and its variants still play a significant role.

Let us illustrate the results concerning the predictive and marginal distributions provided in this section by referring to the two examples initiated in section 2.

Example 1 (continued). With reference to the N-IG process, as shown in Lijoi *et al.* (2005), an application of proposition 2 leads to a predictive distribution of the form (10) with

$$w^{(n)} = \frac{\sum_{r=0}^{n} \binom{n}{r} (-\theta^2)^{-r+1} \Gamma(n(\boldsymbol{\pi}) + 1 + 2r - 2n; \theta)}{2n \sum_{r=0}^{n-1} \binom{n-1}{r} (-\theta^2)^{-r} \Gamma(n(\boldsymbol{\pi}) + 2 + 2r - 2n; \theta)} \tag{16}$$

$$w_j^{(n)} = \left(n_j - \frac{1}{2}\right) \frac{\sum_{r=0}^{n} \binom{n}{r} (-\theta^2)^{-r+1} \Gamma(n(\boldsymbol{\pi}) + 2r - 2n; \theta)}{\sum_{r=0}^{n-1} \binom{n-1}{r} (-\theta^2)^{-r} \Gamma(n(\boldsymbol{\pi}) + 2 + 2r - 2n; \theta)}, \tag{17}$$

where $\Gamma(a,b) = \int_b^\infty x^{a-1} \mathrm{e}^{-x} \, \mathrm{d}x$ is the incomplete Gamma function. The EPPF corresponding to the N-IG process turns out to be

$$\frac{\mathrm{e}^{\theta}(-\theta^2)^{n-1}}{2^{n(\boldsymbol{\pi})-1}\Gamma(n)} \sum_{r=0}^{n-1} \binom{n-1}{r} (-\theta^2)^{-r} \Gamma(n(\boldsymbol{\pi}) + 2 + 2r - 2n; \theta) \left\{ \prod_{j=1}^{n(\boldsymbol{\pi})} \left(\frac{1}{2}\right)_{n_j - 1} \right\}.$$

With reference to the conditional representations, we have, for instance, that the conditional distribution of the random partition $\boldsymbol{\pi}$, given $U_n = u$, coincides with

$$\Pi^{(n)}(\boldsymbol{\pi} \mid u) = \frac{\sqrt{\pi}\, \mathrm{e}^{-\theta} 2^{n - n(\boldsymbol{\pi}) - 1/2}}{(\theta\sqrt{1+2u})^{n - n(\boldsymbol{\pi}) + 1/2} K_{n-1/2}(\theta\sqrt{1+2u})} \prod_{i=1}^{n(\boldsymbol{\pi})} (1-\sigma)_{n_i - 1},$$

where K_v denotes the modified Bessel function of second kind with index v.

Example 2 (continued). According to proposition 2, the weights of the predictive distribution are given by

$$w^{(n)} = \frac{1}{nK_X}\int_{\mathbb{R}^+} u^n \, \mathrm{e}^{-\int_{\mathbb{R}^+}\log[u+\beta(y)]H_x^*(\mathrm{d}y)}\,\mathrm{d}u,$$
$$w_j^{(n)} = \frac{n_j}{K_X}\int_{\mathbb{R}^+} u^n \, \mathrm{e}^{-\int_{\mathbb{R}^+}\log[u+\beta(y)]H_{Y_j}^*(\mathrm{d}y)}\,\mathrm{d}u,$$

where $H_v^*(\mathrm{d}y) = H(\mathrm{d}y) + \sum_{i=1}^{n(\pi)} n_i \delta_{Y_i}(\mathrm{d}y) + \delta_v(\mathrm{d}y)$, for any v in $\mathbb{X}$, and K_X is the normalizing constant in (7). The predictive distribution can now be given a simplified representation as

$$P\left[X_{n+1}\in \mathrm{d}x \mid \boldsymbol{X}\right] = \frac{1}{nK_X}\int_{\mathbb{R}^+} u^n \, \mathrm{e}^{-\int_{\mathbb{X}}\log[u+\beta(y)]H_x^*(\mathrm{d}y)}\,\mathrm{d}u H^*(\mathrm{d}x).$$

If one exploits proposition 4, it is possible to describe the partition structure induced by the normalized extended Gamma prior through its conditional EPPF

$$\Pi^{(n)}(\boldsymbol{\pi}\mid u) = \frac{\mathrm{e}^{-\int_{\mathbb{X}}\log[u+\beta(x)]H(\mathrm{d}x)}}{\zeta_n}\prod_{i=1}^{n(\boldsymbol{\pi})}\int_{\mathbb{X}}\frac{\Gamma(n_i)}{[u+\beta(y)]^{n_i}}H(\mathrm{d}y),$$

where $\zeta_n := \int_{\mathbb{R}^+} t^n \mathrm{e}^{-ut} f_T(t)\,\mathrm{d}t$. It is worth remarking the nice and simple Gibbs structure featured by the above conditional EPPF. Moving to the unconditional EPPF, we resort to proposition 3 and obtain

$$\frac{\prod_{i=1}^{n(\boldsymbol{\pi})}\Gamma(n_i)}{\Gamma(n)}\int_{\mathbb{X}^{n(\boldsymbol{\pi})}}\int_{\mathbb{R}^+} u^{n-1}\,\mathrm{e}^{-\int_{\mathbb{X}}\log[u+\beta(x)]H^*(\mathrm{d}x)}\,\mathrm{d}u H(\mathrm{d}y_1)\cdots H(\mathrm{d}y_{n(\boldsymbol{\pi})}),$$

where $H^*(\mathrm{d}x) = H(\mathrm{d}x) + \sum_{i=1}^{n(\boldsymbol{\pi})} n_i \delta_{y_i}(\mathrm{d}x)$.

3.3. A generalized Blackwell–MacQueen sampling scheme

Proposition 2, combined with the representation of the latent variable U_n in proposition 1, suggests a simple scheme for sampling from the marginal distribution of the observations governed by a general NRMI. This yields an extension of the celebrated Blackwell–MacQueen sampling scheme for the Dirichlet process. Let us provide a description of the algorithm. Firstly, introduce a sequential formulation for partitions and related functions: for $r = 1, \ldots, n$, let $\boldsymbol{\pi}_r = \{C_{1,r}, \ldots, C_{n(\boldsymbol{\pi}_r),r}\}$ denote a partition of the integers $\{1, \ldots, r\}$ into $n(\boldsymbol{\pi}_r) \le r$ distinct sets. For each $j = 1, \ldots, n(\boldsymbol{\pi}_r)$, one now has $C_{j,r} = \{i \in \{1, \ldots, r\} : X_i = Y_j\}$ and the size of each set $C_{j,r}$ is denoted by $n_{j,r}$. Note that $\boldsymbol{\pi}_n = \boldsymbol{\pi}$. The main idea of the algorithm is to exploit the simple structure of the predictive conditional on the latent variable U_n. Indeed, such a predictive distribution can be represented as follows,

$$m(\mathrm{d}X_i \mid X_1, \ldots, X_{i-1}, U_{i-1}) = m(\mathrm{d}X_i \mid \boldsymbol{Y}, \boldsymbol{\pi}_{i-1}, U_{i-1})$$
$$\propto \kappa_1(U_{i-1})H_{1,1}(\mathrm{d}X_i \mid U_{i-1}) + \sum_{j=1}^{n(\boldsymbol{\pi}_{i-1})}\frac{\tau_{n_{j,i-1}+1}(U_{i-1}\mid Y_j)}{\tau_{n_{j,i-1}}(U_{i-1}\mid Y_j)}\delta_{Y_j}(\mathrm{d}X_i) \qquad (18)$$

for any $i \ge 2$ and $m(\mathrm{d}X_1 \mid u) \propto \kappa_1(u)H_{1,1}(\mathrm{d}X_1 \mid u)$. The computational recipe works, then, as follows,

(1) sample U_0 from $q_0(u) = \mathrm{e}^{-\psi(u)}\int_{\mathbb{X}}\tau_1(u \mid x)\eta(\mathrm{d}x)$,
(2) sample X_1 from $m(\mathrm{d}X_1 \mid U_0)$,
(3) for any $i \ge 2$,
  (3a) sample U_{i-1} from $f_{U_{i-1}}^{\boldsymbol{X}_{i-1}}(u)$ where $\boldsymbol{X}_{i-1} = (X_1, \ldots, X_{i-1})$,
  (3b) sample X_i from $m(\mathrm{d}X_i \mid X_1, \ldots, X_{i-1}, U_{i-1})$,
(4) go to (3).

The sampling scheme can be applied once the Poisson intensity of the underlying completely random measure is assigned: indeed all relevant densities from which to sample are known at least up to a proportionality constant. For instance, in the N-IG case, (18) reduces to

$$m(\mathrm{d}X_i \mid X_1,\ldots,X_{i-1},U_{i-1}) = \frac{\theta(1+2U_{i-1})^{\frac{1}{2}}}{\theta(1+2U_{i-1})^{\frac{1}{2}}+2(i-1)-n(\boldsymbol{\pi}_{i-1})}P_0(\mathrm{d}X_i) + \frac{2}{\theta(1+2U_{i-1})^{\frac{1}{2}}+2(i-1)-n(\boldsymbol{\pi}_{i-1})}\sum_{i=1}^{n(\boldsymbol{\pi}_{i-1})}\left(n_i-\frac{1}{2}\right)\delta_{Y_i}(\mathrm{d}X_i), \tag{19}$$

which is straightforward to compute in contrast to the unconditional predictive which is of the form (10) with weights (16)–(17). Having established a computational scheme for generating from the marginal distribution of the observations, the most natural application to think of is Bayesian non-parametric inference within hierarchical mixtures. This is the topic of the next section.

4. Hierarchical mixture models

In terms of statistical applications, owing to the success of the Dirichlet process, one of the most fruitful ways for exploiting NRMIs is their potential use as basic building blocks in hierarchical mixture models. In this setting, $\boldsymbol{X}$ are missing values which capture the clustering structure within the data. This class of models was first introduced, for the Dirichlet process, by Lo (1984) and later popularized by the development of suitable Markov Chain Monte Carlo (MCMC) techniques in Escobar & West (1995). Recently, mixtures of Dirichlet processes have been generalized to mixtures of stick-breaking priors (Ishwaran & James 2001, 2003) and of particular NRMIs (Lijoi *et al.*, 2005, 2007a).

We first recall the model as set up by Lo (1984). Suppose $\{f(\cdot \mid x) : x \in \mathbb{X}\}$ is a family of non-negative kernels defined on a complete and separable metric space $\mathbb{W}$ such that $\int_{\boldsymbol{Y}} f(w \mid x)\lambda(\mathrm{d}w) = 1$ for any x in $\mathbb{X}$ and for some σ-finite measure λ. Next, let $\boldsymbol{W} = (W_1,\ldots,W_n)$ be a vector of $\mathbb{W}$-valued random elements such that

$$\begin{aligned} &W_i \mid X_i \overset{\text{i.n.d.}}{\sim} f(\cdot \mid X_i),\\ &X_i \mid \tilde{P} \overset{\text{i.i.d.}}{\sim} \tilde{P},\\ &\tilde{P} \sim \text{NRMI}. \end{aligned}$$

This is the same as supposing that $W_1,\ldots,W_n$ are exchangeable draws from the random density $\tilde{f}(\cdot) = \int_{\mathbb{X}} f(\cdot \mid x)\tilde{P}(\mathrm{d}x)$. One is naturally interested in the determination of the distribution of the posterior density $\tilde{f}$, given the observations $\boldsymbol{W}$, which coincides with the distribution of the random density

$$\int_{\mathbb{X}} f(\cdot \mid x)\tilde{P}^{\boldsymbol{W}}(\mathrm{d}x),$$

where $\tilde{P}^{\boldsymbol{W}}$ is the (posterior) random probability measure whose distribution is

$$\int \mathbb{P}(\mathrm{d}p \mid \boldsymbol{X})\mathbb{P}(\mathrm{d}\boldsymbol{X} \mid \boldsymbol{W}). \tag{20}$$

Notice that in the previous integral $\mathbb{P}(\mathrm{d}p \mid \boldsymbol{X})$ is the posterior distribution of the NRMI $\tilde{P}$, given $\boldsymbol{X}$, which is provided by theorem 2 and $\mathbb{P}(\mathrm{d}\boldsymbol{X} \mid \boldsymbol{W})$ is the distribution of the latent variables, given the data $\boldsymbol{W}$, which can be determined via Bayes' theorem as

$$\frac{\{\prod_{i=1}^n f(W_i \mid X_i)\}\, m(\mathrm{d}\boldsymbol{X})}{\int \{\prod_{i=1}^n f(W_i \mid X_i)\}\, m(\mathrm{d}\boldsymbol{X})'}$$

where $m(\mathrm{d}\boldsymbol{X})$ is the marginal distribution of the latent variables as described in (11) (see Ishwaran & James, 2003). It is apparent that the main difficulties arise from the evaluation of the integral in (20). In fact, one has to integrate with respect to all possible partitions of the n latent variables $\boldsymbol{X}$. The impossibility of achieving an exact analytical evaluation of the posterior distribution of $\tilde{f}$, given $\boldsymbol{W}$, makes it necessary to devise a computational scheme for drawing samples from the posterior. To this end, the generalization of the Blackwell–MacQueen urn scheme as described in subsection 3.3 is important. As a first step, generate a sample $X_{1,0},\ldots,X_{n,0}$ of i.i.d. values of the latent variable from $\mathbb{E}[\tilde{P}(\mathrm{d}x)]=\int_0^\infty \tau_1(u\,|\,x)\,\mathrm{e}^{-\psi(u)}\,\mathrm{d}uH(\mathrm{d}x)$. Then, for any $t\geq 1$, proceed as follows

(1) draw U_n^t from $f_{U_n}^{\boldsymbol{X}^{t-1}}(u)$ where $\boldsymbol{X}^{t-1}=(X_{1,t-1},\ldots,X_{n,t-1})$ is the vector of latent variables sampled in the previous step $t-1$;

(2) draw the latent $X_{1,t},\ldots,X_{n,t}$ from the Pólya urn scheme as follows: for any i sample X_i from

$$P(X_{i,t}\in\cdot\,|\,\boldsymbol{X}_{-i}^t,\boldsymbol{W},U_n^t)=q_{i,0}^*(U_n^t)H_{1,1}(\mathrm{d}X_{i,t}\,|\,U_n^t)f(W_i\,|\,X_{i,t})+\sum_{j=1}^{k_{i,t}}q_{i,j}^*(U_n^t)\delta_{Y_j}(\cdot),$$

where $\boldsymbol{X}_{-i}^t=(X_{1,t},\ldots,X_{i-1,t},X_{i+1,t-1},\ldots,X_{n,t-1})$, Y_j are the $k_{i,t}$ distinct values in the vector $\boldsymbol{X}_{-i}^t$. The mixing proportions are given by

$$q_{i,0}^*(U_n^t)\propto\kappa_1(U_n^t)\int_{\mathbb{X}}f(W_i\,|\,x)H_{1,1}(\mathrm{d}x\,|\,U_n^t)\qquad q_{i,j}^*(U_n^t)\propto\frac{\tau_{n_j+1}(U_n^t\,|\,Y_j)}{\tau_{n_j}(U_n^t\,|\,Y_j)}f(W_i\,|\,Y_j),$$

subject to the constraint $\sum_{j=0}^{k_{i,t}}q_{i,j}^*(U_n^t)=1$.

This represents a generalization of the Escobar & West (1995) algorithm and, by resorting to the latent variable U_n, allows the generation of a sample from a mixture model governed by any NRMI.

It is well-known that the performance, in terms of mixing speed, of the Escobar & West (1995) algorithm can be improved by implementing an acceleration step which basically consists in adding a further iteration to the algorithm we have just described. Such a variation of the MCMC algorithm for Mixture of Dirichlet Process (MDP) models has been proposed by MacEachern (1994, 1998; see also Ishwaran & James, 2001). Indeed, step (2) above is used in order to fix the number of clusters and the cluster memberships for the latent variables. In order to generate the representative of each cluster, i.e. the unique distinct values Y_j, one proceeds as follows. Suppose that from step (2) one has k_t clusters with memberships identified by the sets of indices $I_{1,t},\ldots,I_{k_t,t}$. Then

(3) draw the unique values $Y_{1,t},\ldots,Y_{k_t,t}$ from the full conditional

$$P\left(Y_{j,t}\in\mathrm{d}x\,|\,\boldsymbol{W},\boldsymbol{X}^t,U_n^t\right)\propto\prod_{i\in I_{j,t}}f(y_i\,|\,x)H_{1,1}(\mathrm{d}x\,|\,U_n^t).$$

One can see that an important point of the algorithm is the evaluation of the weights $q_{i,0}^*$. In order to obtain an explicit form for them, one can choose a conjugate pair $\{f(\cdot\,|\,\cdot),P_0\}$.

Numerical example. As an illustration we analyse a data set concerning the environmental problem of acidification, which consists of measurements of an acid neutralizing capacity (ANC) index in a sample of 155 lakes in North-Central Wisconsin, USA. A low value of ANC can lead to a loss of biological resources. The identification of clusters of lakes is important for the determination of lake characteristics which can be used to predict higher acidification. Also these data were studied by several authors and were considered on a log-scale as we do. Most of the previous studies support the existence of two to three clusters (see, e.g. Crawford, 1994; McGrory & Titterington, 2007).

Here we compare Dirichlet and N-IG mixtures in terms of the posterior distribution on the number of components. The model we adopt is a normal mixture where both means and variances are random and chosen according to either a Dirichlet or a N-IG process, i.e.

$$(W_i \mid m_i, V_i) \overset{\text{i.n.d}}{\sim} \mathrm{N}(Y_i \mid m_i, V_i), \quad i = 1, \ldots, n,$$
$$(m_i, V_i \mid \tilde{P}) \overset{\text{i.i.d}}{\sim} \tilde{P},$$
$$\tilde{P} \sim \text{Dir or N-IG},$$

where N is a normal kernel. In order to appreciate the different behaviours, we fix the prior parameters for both mixtures so that the prior distribution on the number of components $n(\boldsymbol{\pi})$ has mode in 20, thus far away from the low number of components estimated in previous studies. This is achieved by setting the total mass parameter θ equal to 5.9 in the Dirichlet case and equal to 1.29 in the N-IG case. Figure 1 displays the corresponding prior distributions for $n(\boldsymbol{\pi})$.

The general phenomenon of the N-IG process inducing a relatively flat prior, in contrast to the Dirichlet process inducing a highly peaked distribution, is apparent from the plot. For the remaining part concerning P_0, we employ the quite standard semiparametric prior specification of Escobar & West (1995), namely $P_0(\mathrm{d}x\,\mathrm{d}v) = \mathrm{N}(x \mid \mu, \tau v^{-1})\mathrm{Ga}(v \mid 1, 1)\,\mathrm{d}x\,\mathrm{d}v$, where $\mathrm{Ga}(\cdot \mid c, d)$ is the density corresponding to a Gamma distribution with mean c/d. A further hierarchy is assumed for μ and τ, i.e. $\mu \sim N(\cdot \mid 0, .001)$ and $\tau^{-1} \sim \mathrm{Ga}(\cdot \mid 1, 100)$. Simulations for the Dirichlet process mixture were carried out using the usual Blackwell–MacQueen sampling scheme with acceleration step. As for the N-IG mixture, we resorted to the algorithm detailed above: the possibility of using the predictive distributions conditionally on U_n given in (19) reduced the computational burden significantly with respect to the unconditional scheme used in Lijoi *et al.* (2005) where computation of the weights in (16)–(17) were required. All inferences are based on 20,000 iterations after a burn-in period of 5000 sweeps. Table 1 reports the posterior distribution on the number of components in the mixture: the N-IG mixture favours two to three components and, though starting from a prior tuned on 20 components, more than 90% of the mass is concentrated on one to six components. In contrast, the

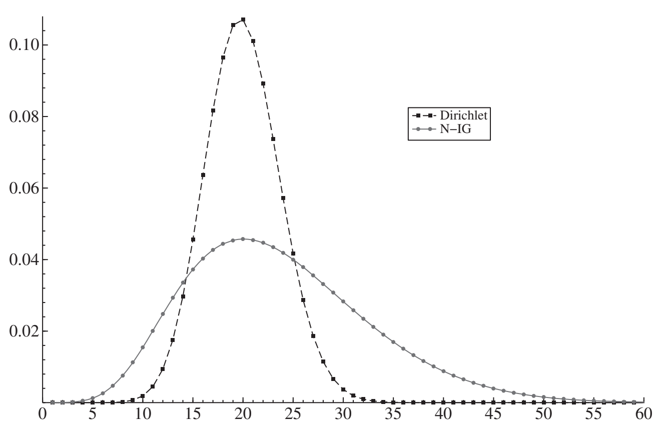

Fig. 1. Prior distributions for the number of components $n(\boldsymbol{\pi})$ corresponding to the Dirichlet and the N-IG mixtures for the 155 acidity data. Their parameters are specified such that the mode of the prior distribution of $n(\boldsymbol{\pi})$ is in 20. The probabilities are connected by lines only for visual simplification.

Table 1. *Posterior probabilities on the number of components $n(\boldsymbol{\pi})$ corresponding to the Dirichlet and the N-IG mixtures for the 155 acidity data*

$n(\boldsymbol{\pi})$	1	2	3	4	5	6	7	8	9	≥ 10
Dirichlet mixture	0.007	0.031	0.074	0.117	0.157	0.160	0.145	0.118	0.078	0.113
N-IG mixture	0.151	0.226	0.212	0.166	0.104	0.063	0.038	0.019	0.011	0.010

Dirichlet mixture is still stuck on a significantly higher number of components: the posterior mode is in six components and the shortest interval cumulating 90% posterior probability is given by [3, 11] components. On the other hand, by tuning the mode of the prior distribution on the number of components on a smaller number of clusters (e.g. $\theta = 1.63$, which corresponds to a median in eight components) also the Dirichlet mixture leads to infer the existence of two to three components. This clearly highlights the fact that N-IG mixtures are more robust with respect to wrong prior specifications.

5. Concluding remarks and computational issues

The present paper has aimed at providing the theoretical framework for a complete Bayesian analysis of NRMIs. Particular cases of these priors have been shown to be useful in various settings such as, e.g. mixture modelling or prediction problems arising when one needs to evaluate the probability of discovering a new species. The main goal is now to study novel concrete examples of NRMIs and evaluate their suitability to the specific applications. Hence, within the class of NRMIs, one has a wide range of non-parametric priors to resort to and does not need to confine herself to the Dirichlet process motivating her choice with the intractability of other options.

Employing the terminology of Papaspiliopoulos & Roberts (2008), one can set up either a conditional or a marginal algorithm and, for both cases, the results of the present paper are essential. As for the latter class of algorithms, one can refer to subsection 3.3, on the generalization of the Blackwell–MacQueen sampling scheme, and to section 4, on application to hierarchical mixture models. As for the former, the representation of the posterior distribution in theorem 1 can be used in order to build a Ferguson–Klass type algorithm (see Ferguson & Klass, 1972; Walker & Damien, 2000): at any iteration of the algorithm one samples a U_n value, given the data $\boldsymbol{X}$, from $f_{U_n}^{\boldsymbol{X}}$ and then simulates a realization of $\tilde{\mu}^{(U_n,\boldsymbol{X})}$. This, combined with a standard Gibbs sampler, allows to exploit any NRMI as a basic building block in complex hierarchical mixture models. A preliminary investigation on this simulation approach is provided in Nieto-Barajas & Prünster (2008): the authors resort to it in order to develop a sensitivity analysis for non-parametric density estimation based on NRMIs.

Finally, it is worth mentioning a recent interesting contribution heavily relying on NRMIs: they are used in order to define time-dependent random probability measures (see Griffin, 2007).

Acknowledgements

The authors are grateful to two anonymous referees for their valuable comments. Moreover, special thanks are due to Ramsés H. Mena for his helpful suggestions and discussions. Lancelot F. James is supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region. Antonio Lijoi and Igor Prünster are partially supported by grants of the Italian Ministry of University and Research.

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–1174.

Crawford, S. L. (1994). An application of the Laplace method to finite mixture distributions. *J. Amer. Statist. Assoc.* **89**, 259–267.

Daley, D. J. & Vere-Jones, D. (1988). *An introduction to the theory of point processes*. Springer, New York.

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

Dykstra, R. L. & Laud, P. W. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–367.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577–588.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**, 87–112.

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

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.

Ferguson, T. S. & Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Statist.* **43**, 1634–1643.

Griffin, J. (2007). The Ornstein–Uhlenbeck Dirichlet process and other time-varying processes for Bayesian nonparametric inference. *CRiSM Working Paper* **07-03**. University of Warwick.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18**, 1259–1294.

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

Ishwaran, H. & James, L. F. (2003). Generalized weighted Chinese restaurant processes for species sampling mixture models. *Statist. Sinica* **13**, 1211–1235.

James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. Unpublished manuscript. *MathArXiv* 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.

James, L. F. (2005a). Poisson process partition calculus with applications to Bayesian Lévy moving averages and shot–noise processes. *Ann. Statist.* **33**, 1771–1799.

James, L. F. (2005b). Functionals of Dirichlet processes, the Cifarelli-Regazzini identity and Beta–Gamma processes. *Ann. Statist.* **33**, 647–660.

James, L. F. (2006). Poisson calculus for spatial neutral to the right processes. *Ann. Statist.* **34**, 416–440.

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

Kallenberg, O. (1986). *Random measures*. Akademie Verlag, Berlin.

Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27**, 562–588.

Kingman, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.

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

Kingman, J. F. C. (1993). *Poisson processes*. Oxford University Press, Oxford.

Korwar, R. M. & Hollander, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1**, 705–711.

Lijoi, A., Mena, R. H. & Prünster, I. (2005). Hierarchical mixture modelling with normalized inverse Gaussian priors. *J. Amer. Statist. 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 Stat. Methodol.* **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.

Lo, A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrsch. Verw. Gebiete* **59**, 55–66.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.* **12**, 351–357.

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

MacEachern, S. N. (1998). Computational methods for mixture of Dirichlet process models. In *Practical nonparametric and semiparametric Bayesian statistics* (eds D. Dey, P. Müller & D. Sinha), 23–43. Springer, New York.
McGrory, C. A., Titterington, D. M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Comput. Statist. Data Anal.* **51**, 5352–5367.
Müller, P. & Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statist. Sci.* **19**, 95–110.
Nieto-Barajas, L. E. & Prünster, I. (2008). A sensitivity analysis for Bayesian nonparametric density estimators. *Statist. Sinica* (in press).
Papaspiliopoulos, O. & Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169–186.
Perman, M., Pitman, J. & Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92**, 21–39.
Pitman, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory. Papers in honor of David Blackwell* (eds T. S. Ferguson, L. S. Shapley & J. B. MacQueen), 245–267. Lecture Notes, Monograph Series, **30**, IMS, Hayward.
Pitman, J. (2003). Poisson-Kingman partitions. In *Science and Statistics: A Festschrift for Terry Speed* (ed. D. R. Goldstein), 1–35. Lecture Notes, Monograph Series, **40**, IMS, Hayward.
Pitman, J. (2006). *Combinatorial stochastic processes*. Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture Notes in Mathematics **1875**. Springer, Berlin.
Pitman, J. & Yor, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.
Prünster, I. (2002). *Random probability measures derived from increasing additive processes and their application to Bayesian statistics*. Ph.d dissertation, University of Pavia.
Regazzini, E., Lijoi, A. & Prünster, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–585.
Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–650.
Walker, S. G. & Damien, P. (2000). Representations of Lévy processes without Gaussian components. *Biometrika* **87**, 477–483.
Walker, S. & Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.* **25**, 1762–1780.
Wolpert, R. & Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85**, 251–267.

Received January 2007, in final form February 2008

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: Proofs

Proof of theorem 1. We can show the validity of the statements in theorem 1 by working directly on the underlying Poisson process $\tilde{N}$. The basic idea is to use the fact that, $\tilde{\mu}$ being a function of $\tilde{N}$ through (1), the posterior distribution of $\tilde{\mu}$ given $\boldsymbol{X}$ can be deduced from the posterior distribution of $\tilde{N}$ given $\boldsymbol{X}$. First, let $\mathbb{P}_v$ denote the distribution of the Poisson random measure $\tilde{N}$ with intensity v and, consequently, $\mathbb{E}_v[\cdot]$ represents the expected value computed with respect to $\mathbb{P}_v$. The proof of theorem 1, then, follows from an application of the approach of James (2005a) in conjunction with the introduction of the latent variable U_n.

First notice that $\tilde{P}(\mathrm{d}y) = T^{-1}\tilde{\mu}(\mathrm{d}y)$ is a special case of the random probability measure described in James (2005a, equation (22), p. 1780) as

$$P_{\tilde{\mu}}(\mathrm{d}y) = q(y, \tilde{\mu})\tilde{\mu}(\mathrm{d}y).$$

That is seen by recalling that $T := \tilde{\mu}(\mathbb{X})$ and setting

$$T^{-1} = q(y, \tilde{\mu}).$$

Note further that in our setting we use the notation (Y_j, J_j) to play the role of the unique points $(Y_j^*, J_{j,n})$ for $j=1,\dots,n(\boldsymbol{\pi})$ described in James (2005a). The $(J_j)=(J_1,\dots,J_{n(\boldsymbol{\pi})})$ now represent the unique values of n latent variables say $\tilde{\mathbf{J}}=(\tilde{J}_1,\dots,\tilde{J}_n)$. The (Y_j) represent the unique values of $\mathbf{X}$. Now, let $N_n^*=N'+\sum_{j=1}^{n(\boldsymbol{\pi})}\delta_{J_j,Y_j}$ and

$$\mu_n^*(\mathrm{d}y)=\int_0^\infty sN_n^*,(\mathrm{d}s,\,\mathrm{d}y):=\mu'(\mathrm{d}y)+\sum_{j=1}^{n(\boldsymbol{\pi})}J_j\delta_{Y_j}(\mathrm{d}y),$$

where μ' and N' are of the same form as $\tilde{\mu}$ and $\tilde{N}$, respectively. Hence, it follows that $\mu_n^*(\mathbb{X})=T'+\sum_{l=1}^{n(\boldsymbol{\pi})}J_l$, where $T'=\mu'(\mathbb{X})$. Furthermore, for $j=1,\dots,n(\boldsymbol{\pi})$,

$$q(Y_j,\mu_n^*)=\frac{1}{(T'+\sum_{l=1}^{n(\boldsymbol{\pi})}J_l)}$$

which does not depend on Y_j or j. This implies that,

$$\prod_{j=1}^{n(\boldsymbol{\pi})}[q(Y_j,\mu_n^*)]^{n_j}=\frac{1}{(T'+\sum_{i=1}^{n(\boldsymbol{\pi})}J_i))^n}.$$

Furthermore, specializing a definition in James (2005a, p. 1781), we have that

$$\phi_n(\tilde{\mathbf{J}},\mathbf{X})=\int\frac{\mathbb{P}_v(\mathrm{d}N)}{(T+\sum_{i=1}^{n(\boldsymbol{\pi})}J_i))^n}=\int_0^\infty\frac{f_T(t)}{[(t+\sum_{j=1}^{n(\boldsymbol{\pi})}J_j)]^n}\,\mathrm{d}t.$$

Now from theorem 3.2 in James (2005a), it can be deduced that the posterior distribution of $\tilde{\mu}\,|\,\mathbf{X}$ is equivalent to that of $\mu_n^*\,|\,\mathbf{X}$ and is determined by the posterior distribution of $N\,|\,\mathbf{X}$ which is equivalent to the distribution of $N_n^*\,|\,\mathbf{X}$. That is, statement (*i*) of theorem 3.2 in James (2005a) shows that the posterior distribution of $\tilde{N}$, given $\boldsymbol{X}$, coincides with the distribution of the random measure $N_n^*=N'+\sum_{i=1}^{n(\boldsymbol{\pi})}\delta_{J_i,Y_i}$, where the joint law of $(N',(J_j))$, given $\boldsymbol{X}$, evaluated at some point $(N,s_1,\dots,s_{n(\boldsymbol{\pi})})$, is proportional to the joint measure

$$\frac{1}{(T+\sum_{i=1}^{n(\boldsymbol{\pi})}s_i))^n}\mathbb{P}_v(\mathrm{d}N)\prod_{i=1}^{n(\boldsymbol{\pi})}[s_i]^{n_i}\rho(\mathrm{d}s_i\,|\,Y_i). \tag{21}$$

This in turn determines the posterior distribution of $\tilde{\mu}\,|\,\boldsymbol{X}$. Additionally, given the form of ϕ_n, statements (*ii*) and (*iii*) of theorem 3.2 in James (2005a) can be exploited in order to provide a preliminary description of the posterior distribution of $\tilde{P}$.

Now to obtain the generally more tractable distributions given $U_n,\boldsymbol{X}$ we first apply the Gamma identity,

$$\frac{1}{(T+\sum_{i=1}^{n(\boldsymbol{\pi})}J_i)^n}=\frac{1}{\Gamma(n)}\int_{\mathbb{R}^+}\mathrm{e}^{-u[T+\sum_{i=1}^{n(\boldsymbol{\pi})}J_i]}u^{n-1}\,\mathrm{d}u.$$

An augmentation of the previous expression combined with (21) yields a joint distribution of $(N',(J_j),U_n,\mathbf{X})$ proportional to

$$u^{n-1}\,\mathrm{e}^{-uT}\,\mathbb{P}_v(\mathrm{d}N)\prod_{i=1}^{n(\boldsymbol{\pi})}s_i^{n_i}\,\mathrm{e}^{-us_i}\rho(\mathrm{d}s_i\,|\,Y_i)H(\mathrm{d}Y_i). \tag{22}$$

Now applying proposition 2.1 of James (2005a), with $uT:=N(f):=\int f(s,y)N(\mathrm{d}s,\,\mathrm{d}y)$, where $f(s,y)=us$, yields the equivalence of measures

$$\mathrm{e}^{-uT}\,\mathbb{P}_v(\mathrm{d}N)=\mathbb{P}_{v_u}(\mathrm{d}N)\,\mathrm{e}^{-\psi(u)}$$

where $\nu_u(\mathrm{d}s, \mathrm{d}x) = \mathrm{e}^{-us}\rho(\mathrm{d}s \mid y)H(\mathrm{d}y)$ and $\mathbb{E}_\nu[\mathrm{e}^{-uT}] = \mathrm{e}^{-\psi(u)}$. Applying this equality to (22) yields a further description of the joint distribution of $(N', (J_j), U_n, \mathbf{X})$ proportional to

$$u^{n-1}\,\mathrm{e}^{-\psi(u)}\mathbb{P}_{\nu_u}(\mathrm{d}N)\prod_{i=1}^{n(\boldsymbol{\pi})} s_i^{n_i}\,\mathrm{e}^{-us_i}\rho(\mathrm{d}s_i \mid Y_i)H(\mathrm{d}Y_i). \tag{23}$$

A description of the distribution of $N', (J_j) \mid U_n, \boldsymbol{X}$ and hence that of $\tilde{N} \mid U_n, \boldsymbol{X}$ and $\tilde{\mu} \mid U_n, \boldsymbol{X}$ follows from application of Bayes' rule to (23). Moreover, conditionally on U_n and $\boldsymbol{X}$, N' and the J_is have the same distribution as $\tilde{N}^{(U_n)}$ and the $J_i^{(U_n,\boldsymbol{X})}$s, respectively.

Alternative Proof of theorem 1. We provide an alternative proof which obtains the posterior Laplace functional via a limiting argument. We first compute the Laplace functional of $\tilde{\mu}$ given $\boldsymbol{X}$. To this end, consider $n(\boldsymbol{\pi})$ disjoint subsets $C_1, \ldots, C_{n(\boldsymbol{\pi})}$ of $\mathbb{X}$ and set $C_{n(\boldsymbol{\pi})+1} = (\cup_{j=1}^{n(\boldsymbol{\pi})} C_j)^c$. Moreover, for notational simplicity, we set $\tilde{\mu}_j = \tilde{\mu}(C_j)$ for $j = 1, \ldots, n(\boldsymbol{\pi})$ and $n(\boldsymbol{\pi}) = k$. If one combines the assumption of exchangeability of the observations as outlined in (4) with the definition of the NRMI $\tilde{P}$ as given in (3), the conditional Laplace functional of $\tilde{\mu}$ is given by

$$\mathbb{E}(\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}(\mathrm{d}x)} \mid \boldsymbol{Y} \in \times_{j=1}^k C_j) = \frac{\mathbb{E}\left(\mathrm{e}^{-\int_{\mathbb{X}} h(x)\mu(\mathrm{d}x)}T^{-n}\tilde{\mu}_1^{n_1}\cdots\tilde{\mu}_k^{n_k}\right)}{\mathbb{E}\left(T^{-n}\tilde{\mu}_1^{n_1}\cdots\tilde{\mu}_k^{n_k}\right)}.$$

Let us first focus on the numerator which can be rewritten as

$$\begin{aligned}
&\frac{1}{\Gamma(n)}\int_{\mathbb{R}^+} u^{n-1}\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}}(h(x)+u)\tilde{\mu}(\mathrm{d}x)}\tilde{\mu}_1^{n_1}\cdots\tilde{\mu}_k^{n_k}\right]\mathrm{d}u \\
&\quad = \frac{1}{\Gamma(n)}\int_{\mathbb{R}^+} u^{n-1}\mathbb{E}\left[\mathrm{e}^{-\int_{C_{k+1}}(h(x)+u)\tilde{\mu}(\mathrm{d}x)}\right]\prod_{i=1}^k(-1)^{n_i}\frac{\mathrm{d}^{n_i}}{\mathrm{d}u^{n_i}}\mathbb{E}\left[\mathrm{e}^{-\int_{C_j}(h(x)+u)\tilde{\mu}(\mathrm{d}x)}\right]\mathrm{d}u.
\end{aligned}$$

Now, introduce the following functions, for any C_j,

$$\begin{aligned}
V_{C_j}^{(n)}(u) = &\left\{(-1)^n\frac{\mathrm{d}^n}{\mathrm{d}u^n}\,\mathrm{e}^{-\int_{\mathbb{R}^+\times C_j}\left(1-\mathrm{e}^{-(h(x)+u)s}\right)\rho(\mathrm{d}s \mid x)H(\mathrm{d}x)}\right\} \\
&\times \mathrm{e}^{\int_{\mathbb{R}^+\times C_j}(1-\mathrm{e}^{-(h(x)+u)s})\rho(\mathrm{d}s \mid x)H(\mathrm{d}x)}
\end{aligned}$$

for any $n \geq 1$ and set $V_{C_j}^{(0)}(u) \equiv 1$. By induction, one observes that

$$V_{C_j}^{(n)}(u) = \int_{C_j}\sum_{i=0}^{n-1}\binom{n-1}{i}\phi_{n-i}(u,x)V_{C_j}^{(i)}(u)H(\mathrm{d}x) = \int_{C_j}\Delta_{H_j}^{(n)}(u,x)H(\mathrm{d}x),$$

where

$$\Delta_{H_j}^{(n)}(u,x) := \sum_{i=0}^{n-1}\binom{n-1}{i}\phi_{n-i}(u,x)V_{C_j}^{(i)}(u),$$

$\phi_{n-i}(u,x) = \int_{\mathbb{R}^+}\mathrm{e}^{-(h(x)+u)s}s^{n-i}\rho(\mathrm{d}s \mid x)$ and $H_j = H(C_j)$. Hence the numerator is equal to

$$\frac{1}{\Gamma(n)}\int_{\mathbb{R}^+} u^{n-1}\,\mathrm{e}^{-\int_{\mathbb{S}}\left(1-\mathrm{e}^{-(h(x)+u)s}\right)\rho(\mathrm{d}s \mid x)H(\mathrm{d}x)}\prod_{i=1}^k\int_{C_j}\Delta_{H_j}^{(n_j)}(u,x)H(\mathrm{d}x)\,\mathrm{d}u.$$

The denominator is determined via similar arguments thus yielding

$$\begin{aligned}
&\mathbb{E}(\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}(\mathrm{d}x)} \mid \boldsymbol{Y} \in \times_{j=1}^k C_j) \\
&\quad = \frac{\int_{\mathbb{R}^+} u^{n-1}\,\mathrm{e}^{-\int_{\mathbb{R}^+\times\mathbb{X}}(1-\mathrm{e}^{-(h(x)+u)s})\rho(\mathrm{d}s \mid x)H(\mathrm{d}x)}\prod_{j=1}^k\int_{C_j}\Delta_{H_j}^{(n_j)}(u,x)H(\mathrm{d}x)\,\mathrm{d}u}{\int_{\mathbb{R}^+} u^{n-1}\,\mathrm{e}^{-\int_{\mathbb{R}^+\times\mathbb{X}}(1-\mathrm{e}^{-us})\rho(\mathrm{d}s \mid x)H(\mathrm{d}x)}\prod_{j=1}^k\int_{C_j}\Delta_{H_j}^{(n_j)}(u,x)H(\mathrm{d}x)\,\mathrm{d}u}.
\end{aligned}$$

If we set $C_j = C_{j,\varepsilon} := \{x \in \mathbb{X} : d_{\mathbb{X}}(x, Y_j) < \varepsilon\}$, where $d_{\mathbb{X}}$ is the distance function on $\mathbb{X}$, non-atomicity of H yields

$$\int_{C_j} \Delta_{H_j}^{(n_j)}(u, x) H(\mathrm{d}x) = H(\mathrm{d}Y_j)(\phi_{n_j}(u, Y_j) + o(H(\mathrm{d}Y_j)) \quad \text{as } \varepsilon \downarrow 0$$

holds true. Hence, as $\varepsilon \downarrow 0$

$$\begin{aligned}
&\mathbb{E}\left(\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}(\mathrm{d}x)} \middle| \boldsymbol{Y} \in \times_{j=1}^{k} C_j\right) \\
&\quad \to \int_{\mathbb{R}^+} \mathrm{e}^{-\int_{\mathbb{R}^+\times\mathbb{X}} \left(1-\mathrm{e}^{-(h(x)+u)s}\right)\rho(\mathrm{d}s\,|\,x)H(\mathrm{d}x)} \\
&\quad\quad \times \prod_{i=1}^{k} \int_{\mathbb{R}^+} \mathrm{e}^{-h(Y_i)s} \frac{s^{n_i}\,\mathrm{e}^{-us}\rho(\mathrm{d}s\,|\,Y_i)}{\tau_{n_i}(u\,|\,Y_i)} \frac{u^{n-1}\prod_{i=1}^{k}\tau_{n_i}(u\,|\,Y_i)\,\mathrm{d}u}{\int_{\mathbb{R}^+} u^{n-1}\prod_{i=1}^{k}\tau_{n_i}(u\,|\,Y_i)\,\mathrm{d}u} \\
&\quad = \int_{\mathbb{R}^+} \mathbb{E}\left(\mathrm{e}^{-\int_{\mathbb{X}} h(x)\tilde{\mu}^{(u)}(\mathrm{d}x)}\right) \prod_{i=1}^{k} \mathbb{E}\left(\mathrm{e}^{-h(Y_i)J_i^{(u,\boldsymbol{X})}}\right) \\
&\quad\quad \times \frac{u^{n-1}\left(\prod_{i=1}^{k}\tau_{n_i}(u\,|\,Y_i)\right)\mathrm{e}^{-\psi(u)}\,\mathrm{d}u}{\int_{\mathbb{R}^+} u^{n-1}\left(\prod_{i=1}^{k}\tau_{n_i}(u\,|\,Y_i)\right)\mathrm{e}^{-\psi(u)}\,\mathrm{d}u}.
\end{aligned}$$

Thus, the proof is complete.

Proof of proposition 1. This easily follows from an application of Bayes' rule to (23). That is by first integrating out the N, and the s_is to first obtain a joint distribution of $U_n, \boldsymbol{X}$. Note how this also gives the τ_{n_i}

Proof of theorem 2. For denoting a linear functional of the completely random measure $\tilde{\mu}$ we use the short notation $\tilde{\mu}(f) = \int_{\mathbb{X}} f(x)\tilde{\mu}(\mathrm{d}x)$ for any measurable $f : \mathbb{X} \to \mathbb{R}$ such that $\tilde{\mu}(|f|) < \infty$ a.s. Now, notice that for any $y_1, \ldots, y_n \in (0, 1)$ and $A_1, \ldots, A_n \in \mathscr{B}(\mathbb{X})$ one has

$$\mathbb{P}[\tilde{P}(A_1) \le y_1, \ldots, \tilde{P}(A_n) \le y_n | U_n, \boldsymbol{X}] = \mathbb{P}\left[\tilde{\mu}(\mathbb{I}_{A_1} - y_1) \le 0, \ldots, \tilde{\mu}(\mathbb{I}_{A_n} - y_n) \le 0 \,|\, U_n, \boldsymbol{X}\right].$$

By definition the latter coincides with

$$\mathbb{P}[\tilde{\mu}^{(U_n,\boldsymbol{X})}(\mathbb{I}_{A_1} - y_1) \le 0, \ldots, \tilde{\mu}^{(U_n,\boldsymbol{X})}(\mathbb{I}_{A_n} - y_n) \le 0],$$

and the result follows since the finite dimensional distributions of $\tilde{\mu}^{(U_n,\boldsymbol{X})}/\tilde{\mu}^{(U_n,\boldsymbol{X})}(\mathbb{X})$ coincide with the finite dimensional distributions of $\tilde{P}$ given U_n and $\boldsymbol{X}$.

Proof of proposition 2. The proof follows from observing that

$$\mathbb{P}[X_{n+1} \in \mathrm{d}x \,|\, \boldsymbol{X}] = \mathbb{E}[\tilde{P}(\mathrm{d}x) \,|\, \boldsymbol{X}] = \int_{\mathbb{R}^+} \mathbb{E}[\tilde{P}(\mathrm{d}x) \,|\, U_n = u, \boldsymbol{X}] f_{U_n}^{\boldsymbol{X}}(u)\,\mathrm{d}u.$$

By virtue of theorem 1,

$$\begin{aligned}
\mathbb{E}\left[\tilde{P}(\mathrm{d}x) \,|\, U_n = u, \boldsymbol{X}\right] &= \mathbb{E}\left[\frac{\tilde{\mu}^{(u)}(\mathrm{d}x)}{T^{(u)} + \sum_{i=1}^{n(\pi)} J_i^{(u,\boldsymbol{X})}}\right] + \mathbb{E}\left[\frac{\sum_{i=1}^{n(\pi)} J_i^{(u,\boldsymbol{X})}\delta_{Y_i}(\mathrm{d}x)}{T^{(u)} + \sum_{i=1}^{n(\pi)} J_i^{(u,\boldsymbol{X})}}\right] \\
&= I_1(u, x, \boldsymbol{X}) + I_2(u, x, \boldsymbol{X}).
\end{aligned}$$

Let us now focus on $I_1(u, x, \boldsymbol{X})$. We are going to prove that

$$\int_{\mathbb{R}^+} I_1(u, x, \boldsymbol{X}) f_{U_n}^{\boldsymbol{X}}(u)\,\mathrm{d}u = w^{(n)} H(\mathrm{d}x). \tag{24}$$

To this end, one can exploit the independence, conditional on $U_n = u$ and on $\boldsymbol{X}$, between the $J_i^{(u,\boldsymbol{X})}$s and μ_u and the independence of the increments of μ_u to show that

$$I_1(u, x, \boldsymbol{X}) = \int_{\mathbb{R}^+} \mathbb{E}[\mathrm{e}^{-v \sum_{i=1}^{n(\boldsymbol{\pi})} J_i^{(u,\boldsymbol{X})}}]\, \mathbb{E}[\tilde{\mu}^{(u)}(\mathrm{d}x)\, \mathrm{e}^{-vT^{(u)}}]\, \mathrm{d}v$$

$$= H(\mathrm{d}x) \int_{\mathbb{R}^+} \left(\prod_{i=1}^{n(\boldsymbol{\pi})} \frac{\tau_{n_i}(u+v \mid Y_i)}{\tau_{n_i}(u \mid Y_i)} \right) \tau_1(u+v \mid x)\, \mathrm{e}^{-\psi^{(u)}(v)}\, \mathrm{d}v,$$

where $\psi^{(u)}(v) = -\log \mathbb{E}[\mathrm{e}^{-v\tilde{\mu}^{(u)}}]$. Now, observe that $\psi^{(u)}(v) + \psi(u) = \psi(u+v)$ so that the right-hand side of (24) reduces to

$$H(\mathrm{d}x) \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} u^{n-1} \mathrm{e}^{-\psi(u+v)} \left(\prod_{i=1}^{n(\boldsymbol{\pi})} \tau_{n_i}(u+v \mid Y_i) \right) \tau_1(u+v \mid x)\, \mathrm{d}u\, \mathrm{d}v.$$

The change of variable $(w, z) = (u+v, u)$ and subsequent integration with respect to z immediately yield $w^{(n)}$. The proof for the remaining weights of the predictive distribution moves along the same lines and it is omitted. Note that one may also use proposition 3 to prove this result.

Proof of proposition 3. This easily follows from (23), if one integrates out N, the s_is and u.

Proof of proposition 4. The conditional distribution of $\boldsymbol{Y}$, given U_n and $\boldsymbol{\pi}$, is obtained by applying Bayes' rule to (23). An application of Bayes' rule also yields readily a description of the conditional distribution of $\boldsymbol{\pi}$ given U_n, the normalizing constant being $\sum_{\boldsymbol{\pi}} \prod_{i=1}^{n(\boldsymbol{\pi})} \kappa_{n_i}(U_n)$. Here $\sum_{\boldsymbol{\pi}}$ stands for the sum over all partitions of the set of integers $\{1, \ldots, n\}$. The simpler form in (14) may be obtained by noting some known relationships between cumulants, partitions and moments. However, for immediate clarity one can use (23) to establish that identity

$$f_{U_n}(u) = \frac{1}{\Gamma(n)} u^{n-1}\, \mathrm{e}^{-\psi(u)} \sum_{\boldsymbol{\pi}} \prod_{i=1}^{n(\boldsymbol{\pi})} \kappa_{n_i}(u).$$

The result then follows by noting the form of $f_{U_n}(u)$ given in (5).