ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Statistical Methodology
Series B

J. R. Statist. Soc. B (2017)
79, *Part* 2, *pp.* 525–545

Compound random measures and their use in Bayesian non-parametrics

Jim E. Griffin and Fabrizio Leisen

University of Kent, Canterbury, UK

[Received October 2014. Final revision January 2016]

Summary. A new class of dependent random measures which we call *compound random measures* is proposed and the use of normalized versions of these random measures as priors in Bayesian non-parametric mixture models is considered. Their tractability allows the properties of both compound random measures and normalized compound random measures to be derived. In particular, we show how compound random measures can be constructed with gamma, σ-stable and generalized gamma process marginals. We also derive several forms of the Laplace exponent and characterize dependence through both the Lévy copula and the correlation function. An augmented Pólya urn scheme sampler and a slice sampler are described for posterior inference when a normalized compound random measure is used as the mixing measure in a non-parametric mixture model and a data example is discussed.

Keywords: Dependent random measures; Lévy copula; Mixture models; Multivariate Lévy measures; Partial exchangeability; Slice sampler

1. Introduction

Bayesian non-parametric mixtures have become a standard tool for inference when a distribution of either observable or unobservable quantities is considered unknown. A more challenging problem, which arises in many applications, is to define a prior for a collection of related unknown distributions. For example, Müller *et al*. (2004) considered informing the analysis of a study with results from previous related studies. They considered the Cancer and Leukemia Group B CALGB 9160 (Budman *et al*., 1998) clinical study which looked at the response over time of patients to different anticancer drug therapies. Müller *et al*. (2004) suggested improving the precision of their inference by using the results of the related study CALGB 8881 (Lichtman *et al*., 1993). Fig. 1 shows bivariate plots of two subject-specific regression parameters (β_0 and β_1) for the two studies. The graphs suggest differences between the joint distribution of β_0 and β_1 which should be included in any analysis which combines these data sets. The results for the CALGB 9160 study also suggest that a non-parametric model is needed to describe the shape of the density fully. A natural Bayesian approach would assume different distributions for each study but construct a dependent prior for these distributions.

In general, suppose that $x \in \mathcal{X}$ denotes the value of covariates; then, in a Bayesian non-parametric analysis, a prior needs to be defined across a collection of correlated distributions $\{\tilde{p}_x | x \in \mathcal{X}\}$. This problem was initially studied in a seminal paper on dependent Dirichlet processes (MacEachern, 1998) where generalizations of the Dirichlet process were proposed. Sub-

Address for correspondence: Jim E. Griffin, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, UK.
E-mail: jeg28@kent.ac.uk

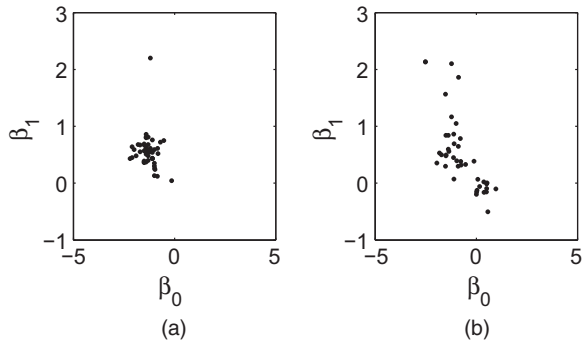

Fig. 1. Scatter plots of the subject-specific regression parameters β_0 and β_1 for the groups in (a) the CALGB 8881 study and (b) the CALGB 9160 study

sequent work used stick breaking constructions of random measures as a basis for defining such a prior. This work is reviewed by Dunson (2010). These priors can usually be represented as

$$\tilde{p}_x = \sum_{i=1}^{\infty} w_i(x)\delta_{\theta_i(x)} \tag{1.1}$$

where $w_1(x), w_2(x), \ldots$ follow a stick breaking process for all $x \in \mathcal{X}$. A drawback with this approach is the stochastic ordering of the $w_i(x)$s for any $x \in \mathcal{A}$ which can lead to strange effects in the prior as x varies.

If $\mathcal{A}$ is countable, several other approaches to defining a prior on a collection of random probability measures have been proposed. The hierarchical Dirichlet process (Teh *et al.*, 2006) assumes that $\tilde{p}_x$ are *a priori* conditionally independent and identically distributed according to a Dirichlet process whose centring measure is itself given a Dirichlet process prior. This construction induces correlation between the elements of $\{\tilde{p}_x | x \in \mathcal{A}\}$ in an analogous way to a parametric hierarchical model. This construction can be extended to more general hierarchical frameworks (see for example Teh and Jordan (2010), for a review). Alternatively, a prior can be defined by using the idea of normalized random measures with independent increments which are defined by normalizing a completely random measure. The prior is defined on a collection of correlated completely random measures $\{\tilde{\mu}_x | x \in \mathcal{A}\}$ which are then normalized for each of x, i.e. $\tilde{p}_x = \tilde{\mu}_x / \tilde{\mu}_x(\mathbb{X})$ where $\mathbb{X}$ is the support of $\tilde{\mu}_x$. Several specific constructions have been proposed including various forms of superposition (Griffin *et al.*, 2013; Lijoi and Nipoti, 2014; Lijoi *et al.*, 2014a, b; Chen *et al.*, 2013; Bassetti *et al.*, 2014), kernel-weighted completely random measures (Foti and Williamson, 2012; Griffin, 2011; Rosinski, 2007; Barndorff-Nielsen *et al.*, 2001) and Lévy copula-based approaches (Leisen and Lijoi, 2011; Leisen *et al.*, 2013; Zhu and Leisen, 2014). In this paper, we develop an alternative method for constructing correlated completely random measures which is tractable, whose properties can be derived and for which sampling methods for posterior inference without truncation can be developed. The construction also provides a unifying framework for previously proposed constructions. Indeed, the σ-stable and gamma vector of dependent random measures that have been studied in the recent works of Leisen and Lijoi (2011), Leisen *et al.* (2013) and Zhu and Leisen (2014) are special cases. Although these references derive useful theoretical results, their application has been limited by the lack of sampling methods for posterior inference. The algorithms that are proposed in this paper can also be used for posterior sampling of models with these non-parametric priors which is another contribution of the paper.

The paper is organized as follows. Section 2 introduces the concepts of completely random

measures, normalized random measures and their multivariate extensions. Section 3 discusses the construction and some properties of a new class of multivariate Lévy process, *compound random measures*, defined by a *score distribution* and a *directing Lévy process*. Section 4 provides a detailed description of compound random measures with a gamma score distribution. Section 5 considers the use of a normalized version of compound random measures in non-parametric mixture models including the description of a Markov chain Monte Carlo scheme for inference. Section 6 provides an illustration of the use of these methods in an example and Section 7 concludes. Additional details are available in the on-line supplementary material. MATLAB code to implement the methods that are described in the paper is available from `http://www.kent.ac.uk/smsas/personal/jeg28/index.htm`.

2. Preliminaries

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathbb{X}, \mathcal{X})$ be a measure space, with $\mathbb{X}$ Polish and $\mathcal{X}$ the Borel σ-algebra of subsets of $\mathbb{X}$. Denote by $\mathbb{M}_{\mathbb{X}}$ the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$, i.e. this means that for any μ in $\mathbb{M}_{\mathbb{X}}$ and any bounded set A in $\mathcal{X}$ we have $\mu(A) < \infty$. Moreover, $\mathcal{M}_{\mathbb{X}}$ stands for the corresponding Borel σ-algebra; see Daley and Vere-Jones (2003) for technical details.

Definition 1. Let $\tilde{\mu}$ be a measurable mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathbb{M}_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ and such that for any $A_1, \ldots, A_n$ in $\mathcal{X}$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent. Then $\tilde{\mu}$ is called a *completely random measure* (CRM).

The concept of a CRM was introduced by Kingman (1967). A CRM can always be represented as a sum of two components

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{i=1}^{M} V_i \delta_{x_i}$$

where the fixed jump points $x_1, \ldots, x_M$ are in $\mathbb{X}$ and the non-negative random jumps $V_1, \ldots, V_M$ are both mutually independent and independent from $\tilde{\mu}_c$. (The representation of a CRM should also take into account a deterministic component; see Cont and Tankov (2004). This consideration also applies later when we shall talk about vectors of CRMs. Since the deterministic part is not relevant for the scope of the paper, we prefer to omit the deterministic component in the representation of a CRM.) $\tilde{\mu}_c$ is a CRM such that

$$\tilde{\mu}_c = \sum_{i=1}^{\infty} J_i \delta_{X_i}$$

where both the positive jump heights J_i and the $\mathbb{X}$-valued jump locations X_i are random. The measure $\tilde{\mu}_c$ is characterized by the *Lévy–Khintchine* representation which states that

$$\mathbb{E}\left[\exp\left\{-\int_X f(x)\,\tilde{\mu}_c(\mathrm{d}x)\right\}\right] = \exp\left(-\int_0^{\infty}\int_X [1-\exp\{-s\,f(x)\}]\,\bar{\nu}(\mathrm{d}s\,\mathrm{d}x)\right)$$

where $f : \mathbb{X} \to \mathbb{R}^+$ is a measurable function such that $\int f\tilde{\mu}_c < \infty$ almost surely and $\bar{\nu}$ is a measure on $\mathbb{R}^+ \times \mathbb{X}$ such that

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

for any B in $\mathcal{X}$. The measure $\bar{\nu}$ is usually called the Lévy intensity of $\tilde{\mu}_c$. Throughout the paper,

we shall consider CRMs without the fixed jump component (i.e. $M = 0$). For our purposes, we shall focus on the *homogeneous* case, i.e. Lévy intensities where the height and location contributions are separated. Formally,

$$\bar{\nu}(\mathrm{d}s, \mathrm{d}x) = \rho(\mathrm{d}s)\,\alpha(\mathrm{d}x)$$

where ρ is a measure on $\mathbb{R}^+$ and α is a non-atomic measure on $\mathbb{X}$, which is usually called the *centring measure*. Some well-known examples are the *gamma* process,

$$\bar{\nu}(\mathrm{d}s, \mathrm{d}x) = s^{-1}\exp(-s)\,\mathrm{d}s\,\alpha(\mathrm{d}x), \qquad s > 0, \tag{2.1}$$

the σ-*stable* process,

$$\bar{\nu}(\mathrm{d}s, \mathrm{d}x) = \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma}\,\mathrm{d}s\,\alpha(\mathrm{d}x), \qquad s > 0, \quad 0 < \sigma < 1, \tag{2.2}$$

and the homogeneous *beta* process,

$$\bar{\nu}(\mathrm{d}s, \mathrm{d}x) = \theta s^{-1}(1-s)^{\theta-1}\,\mathrm{d}s\,\alpha(\mathrm{d}x), \qquad 0 < s < 1, \quad \theta > 0. \tag{2.3}$$

A general class of processes that includes the gamma and σ-stable process is the *generalized gamma* process,

$$\bar{\nu}(\mathrm{d}s, \mathrm{d}x) = \frac{\sigma}{\Gamma(1-\sigma)} s^{-1-\sigma}\exp(-as)\,\mathrm{d}s\,\alpha(\mathrm{d}x), \qquad s > 0, \quad 0 < \sigma < 1, \quad a > 0. \tag{2.4}$$

Random measures are the basis for building Bayesian non-parametric priors.

Definition 2. Let $\tilde{\mu}$ be a random measure in $(\mathbb{M}_{\mathcal{X}}, \mathcal{M}_{\mathbb{X}})$ such that $\tilde{\mu}(\mathbb{X}) \in (0, \infty)$ almost surely. A *normalized random measure* is defined as $\tilde{p} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$.

The definition of a normalized random measure is very general and does not require that the underlying measure is completely random. The Pitman–Yor process (see Pitman and Yor (1997)) is a well-known example of a Bayesian non-parametric prior which cannot be derived by normalizing a CRM. In this particular case, the unnormalized measure is obtained through a change of measure of a σ-stable process. However, many common Bayesian non-parametric priors can be defined as a normalization of a CRM and many other processes can be derived by normalizing processes that are derived from CRMs; see Regazzini *et al.* (2003). For instance, it can be shown that the *Dirichlet process*, which was introduced by Ferguson (1973), is a normalized gamma process. Throughout the paper, we shall assume that the underlying measure is a CRM and use the term normalized random measures with independent increments to emphasize the independence of a CRM on disjoint intervals.

Although non-parametric priors based on normalization are extremely flexible, in many real applications data arise under different conditions and hence assuming a single prior can be too restrictive. For example, data may be divided into different units using covariates. In this case, one would like to consider different distributions for different units instead of a single common distribution for all the units. In these situations, it is more reasonable to consider vectors of dependent random probability measures.

2.1. Vectors of normalized random measures

Suppose that $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are homogeneous CRMs on $(\mathbb{X}, \mathcal{X})$ with respective marginal Lévy intensities

$$\bar{\nu}_j(\mathrm{d}s, \mathrm{d}x) = \nu_j(\mathrm{d}s)\,\alpha(\mathrm{d}x), \qquad j = 1, \ldots, d, \tag{2.5}$$

where ν_j is a measure on $\mathbb{R}^+$ and α is a non-atomic measure on $\mathbb{X}$. Furthermore, $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are dependent (throughout the paper we consider CRMs with no fixed atoms and without deterministic component), and the random vector $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ has independent increments, in the sense that for any $A_1, \ldots, A_n$ in $\mathcal{X}$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$, the random vectors $(\tilde{\mu}_1(A_i), \ldots, \tilde{\mu}_d(A_i))$ and $(\tilde{\mu}_1(A_j), \ldots, \tilde{\mu}_d(A_j))$ are independent. This implies that for any set of measurable functions $\mathbf{f} = (f_1, \ldots, f_d)$ such that $f_j : \mathbb{X} \to \mathbb{R}^+$, $j = 1, \ldots, d$, and $\int |f_j| \mathrm{d}\tilde{\mu}_j < \infty$, we have a multivariate analogue of the Lévy–Khintchine representation (see Sato (1999), Daley and Vere-Jones (2003) and Epifani and Lijoi (2010)):

$$\mathbb{E}[\exp\{-\tilde{\mu}_1(f_1) - \ldots - \tilde{\mu}_d(f_d)\}] = \exp\{-\psi^*_{\rho,d}(\mathbf{f})\} \tag{2.6}$$

where $\tilde{\mu}_j(f_j) = \int f_j \mathrm{d}\tilde{\mu}_j$,

$$\psi^*_{\rho,d}(\mathbf{f}) = \int_{\mathbb{X}} \int_{(0,\infty)^d} [1 - \exp\{-s_1 f_1(x) - \ldots - s_d f_d(x)\}] \rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_d) \alpha(\mathrm{d}x) \tag{2.7}$$

and

$$\int_{(0,\infty)^{d-1}} \rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_{j-1}, A, \mathrm{d}s_{j+1}, \ldots, \mathrm{d}s_d) = \int_A \nu_j(\mathrm{d}s). \tag{2.8}$$

Representation (2.5) implies that the jump heights of $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ are independent from the jump locations. Moreover, these jump locations are common to all the CRMs and are governed by α. It is worth noting that, since $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ has independent increments, its distribution is characterized by a choice of $f_1, \ldots, f_d$ in equation (2.6) such that $f_j = \lambda_j \mathbf{1}_A$ for any set A in $\mathcal{X}$, $\lambda_j \in \mathbb{R}^+$ and $j = 1, \ldots, d$. In this case

$$\psi^*_{\rho,d}(\mathbf{f}) = \alpha(A) \psi_{\rho,d}(\boldsymbol{\lambda})$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ and

$$\psi_{\rho,d}(\boldsymbol{\lambda}) = \int_{(\mathbb{R}^+)^d} \{1 - \exp(-\langle \boldsymbol{\lambda}, \mathbf{s} \rangle)\} \rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_d) \tag{2.9}$$

where $\mathbf{s} = (s_1, \ldots, s_d)$ and $\langle \boldsymbol{\lambda}, \mathbf{s} \rangle = \Sigma_{j=1}^d \lambda_j s_j$.

We close the section with the definition of *vectors of normalized random measures with independent increments*.

Definition 3. Let $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ be a vector of CRMs on $\mathbb{X}$ such that $\tilde{\mu}_j(\mathbb{X}) \in (0, \infty)$ almost surely and let $\tilde{p}_j = \mu_j / \mu_j(\mathbb{X})$ for $j = 1, \ldots, d$. The vector

$$\tilde{p} = (\tilde{p}_1, \ldots, \tilde{p}_d) \tag{2.10}$$

is called a *vector of dependent normalized random measures with independent increments* on $(\mathbb{X}, \mathcal{X})$.

3. Compound random measures

In this section, we shall define a general class of vectors of normalized random measures with independent increments that incorporates many recently proposed priors built by using normalization; see for instance Leisen and Lijoi (2011), Leisen *et al.* (2013), Zhu and Leisen (2014), Griffin *et al.* (2013) and Lijoi *et al.* (2014a). Before introducing the formal definition of compound random measures, we want to provide an intuitive illustration of the model. Consider the following dependent random probability measures:

$$\tilde{p}_1 = \sum_{i \geqslant 1} \pi_{1,i} \delta_{X_i}, \ldots, \tilde{p}_d = \sum_{i \geqslant 1} \pi_{d,i} \delta_{X_i},$$

where

$$\pi_{j,i} = \frac{m_{j,i} J_i}{\sum_l m_{j,l} J_l}. \tag{3.1}$$

The $m_{j,i}$s are perturbation coefficients that identify specific features of the jth random measure and they are independent and identically distributed across the random measures. The shared jumps $(J_i)_{i \geqslant 1}$ lead to dependence among the $\tilde{p}_j$. In the next section, we shall provide a formal definition of compound random measures in terms of their multivariate Lévy intensity.

3.1. Definition

Let $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ be a vector of homogeneous CRMs on $\mathbb{X}$, i.e. the Lévy intensity ν_j of the measure $\tilde{\mu}_j$ is

$$\bar{\nu}_j(\mathrm{d}s, \mathrm{d}x) = \nu_j(\mathrm{d}s)\alpha(\mathrm{d}x), \qquad j = 1, \ldots, d.$$

Following the notation in equation (2.7), we want to define a ρ_d such that

$$\int_{(0,\infty)^{d-1}} \rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_{j-1}, A, \mathrm{d}s_{j+1}, \ldots, \mathrm{d}s_d) = \int_A \nu_j(\mathrm{d}s) \tag{3.2}$$

for any $j = 1, \ldots, d$. In this setting we can define a compound random measure.

Definition 4. A *compound random measure* is a vector of CRMs defined by a *score distribution* h and a *directing Lévy process* with intensity ν^* such that

$$\rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_d) = \int z^{-d} h(s_1/z, \ldots, s_d/z)\, \mathrm{d}s_1 \ldots \mathrm{d}s_d\, \nu^*(\mathrm{d}z) \tag{3.3}$$

where h is the probability mass function or probability density function of the score distribution with parameters z and ν^* is the Lévy intensity of the directing Lévy process which satisfies the condition

$$\int z^{-d} \int \min(1, \| \mathbf{s} \|)\, h(s_1/z, \ldots, s_d/z)\, \mathrm{d}\mathbf{s}\, \nu^*(\mathrm{d}z) < \infty$$

where $\| \mathbf{s} \|$ is the Euclidean norm of the vector $\mathbf{s} = (s_1, \ldots, s_d)$.

The compound Poisson process with jump density h is a compound random measure with a score density h and whose directing Lévy process is a Poisson process. Therefore, compound random measures can be seen as a generalization of compound Poisson processes. It is straightforward to show that $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ can be expressed as

$$\tilde{\mu}_j = \sum_{i=1}^{\infty} m_{j,i} J_i \delta_{X_i} \tag{3.4}$$

where $m_{1,i}, \ldots, m_{d,i} \sim^{\mathrm{IID}} h$ are *scores* and

$$\tilde{\eta} = \sum_{i=1}^{\infty} J_i \delta_{X_i}$$

is a CRM with Lévy intensity $\nu^*(\mathrm{d}s)\alpha(\mathrm{d}x)$. This makes the structure of the prior much more explicit. The random measures share the same jump locations (which have distribution $\alpha/\alpha(\mathbb{X})$)

but the ith jump has a height $m_{j,i}J_i$ in the jth measure and so the jump heights are rescaled by the score (a larger score implies a larger jump height). Clearly, the shared factor J_i leads to dependence between the jump heights in each measure.

To ensure the existence of the vectors of normalized compound random measures as introduced in definition 3, the following condition must be satisfied for each $j = 1, \ldots, d$:

$$\nu_j\{(0,\infty)\} = \int_0^\infty \int z^{-1} h_j(s/z)\, \nu^*(\mathrm{d}z)\, \mathrm{d}s = \infty$$

where $h_j(s/z) = \int h(s_1, \ldots, s_{j-1}, s/z, s_{j+1}, \ldots, s_d)\, \mathrm{d}s_1 \ldots \mathrm{d}s_{j-1}\mathrm{d}s_{j+1} \ldots \mathrm{d}s_d$. If this condition does not hold true, then $\tilde{\mu}_j(\mathbb{X}) = 0$ with positive probability and the normalization does not make sense; see Regazzini *et al.* (2003).

Remark 1. The construction of dependent random measures by superpositions of independent random measures has been considered by many researchers (see Griffin *et al.* (2013), Lijoi *et al.* (2014) and Chen *et al.* (2013) for examples in Bayesian non-parametrics). These can be expressed as compound random measures. For example, suppose that $d = 2$, and let $\mu_1 = \tilde{\mu}_0^* + \tilde{\mu}_1^*$ and $\mu_2 = \tilde{\mu}_0^* + \tilde{\mu}_2^*$ where $\tilde{\mu}_0^*$, $\tilde{\mu}_1^*$ and $\tilde{\mu}_2^*$ are independent CRMs for which $\tilde{\mu}_j^*$ has Lévy intensity $M_j \xi(\mathrm{d}s)\, \alpha(\mathrm{d}x)$ for a suitable choice of ξ. The properties of thinnings of a Poisson process can be used to show that this process is equivalent to a compound random measure with $\nu^*(\mathrm{d}s) = \Sigma_{j=0}^2 M_j \xi(\mathrm{d}s)$ and the score distribution can be $h(m_{1,i} = 1, m_{2,i} = 1) = M_0/(M_0 + M_1 + M_2)$, $h(m_{1,i} = 1, m_{2,i} = 0) = M_1/(M_0 + M_1 + M_2)$ and $h(m_{1,i} = 0, m_{2,i} = 1) = M_2/(M_0 + M_1 + M_2)$. Clearly, this idea can be easily extended to higher dimensions.

In this paper, we shall concentrate on the subclass of compound random measures whose scores are independent and identically distributed so that

$$h(s_1, \ldots, s_d) = \prod_{j=1}^d f(s_j)$$

where f is a continuous univariate density. This implies that each marginal process has the same Lévy intensity of the form

$$\nu_j(\mathrm{d}s) = \nu(\mathrm{d}s) = \int z^{-1} f(s/z)\, \mathrm{d}s\, \nu^*(\mathrm{d}z). \tag{3.5}$$

In Section 5.1, algorithms are introduced to sample from the posterior of a hierarchical mixture model driven by a vector of normalized compound random measures. These samplers depend crucially on knowing the form of the Laplace exponent and its derivatives. Some general results about the Laplace exponent and the dependence are available if we assume that the density $z^{-1} f(s/z)$ admits a moment-generating function.

Theorem 1. Let

$$M_z^f(t) = \int \exp(ts) z^{-1} f(s/z) \mathrm{d}s$$

be the moment-generating function of $z^{-1} f(s/z)$ and suppose that it exists. Then

$$\psi_{\rho,d}(\lambda_1, \ldots, \lambda_d) = \int \left\{ 1 - \prod_{j=1}^d M_z^f(-\lambda_j) \right\} \nu^*(z)\, \mathrm{d}z. \tag{3.6}$$

The proof of theorem 1 is in the on-line supplementary material as well as a further result about the derivatives of the Laplace exponent.

4. Compound random measures with independent gamma-distributed scores

In this paper, we shall focus on exponential or gamma score distributions. Throughout the paper we shall write $\mathrm{Ga}(\phi)$ to be a gamma distribution (or density) with shape ϕ and mean ϕ which has density

$$f(x) = \frac{1}{\Gamma(\phi)} x^{\phi-1} \exp(-x). \tag{4.1}$$

This implies that $z^{-1} f(s/z)$ is the density of a gamma distribution with shape parameter equal to ϕ and mean ϕz. The Lévy intensities ν and ν^* and the score density f are linked by equation (3.5) and a compound random measure can be defined by either deriving ν^* for a fixed choice of f and ν or by directly specifying f and ν^*. In this latter case, it is interesting to consider the properties of the induced ν.

Standard inversion methods can be used to derive the form of ν^*. Equation (3.5) implies that

$$\nu(s) = \int z^{-1} \frac{1}{\Gamma(\phi)} \left(\frac{s}{z}\right)^{\phi-1} \exp\left(-\frac{s}{z}\right) \nu^*(z)\, \mathrm{d}z.$$

The change of variable $t = z^{-1}$ leads to

$$\nu(s) = \frac{s^{\phi-1}}{\Gamma(\phi)} \int \exp(-st)\, t^{\phi-2} \nu^*\left(\frac{1}{t}\right) \mathrm{d}t.$$

This integral can be seen as the classical Laplace transform of the function $S(t) = t^{\phi-2}\nu^*(1/t)$. If we denote by $\mathcal{L}$ the Laplace transform then

$$\nu(s) = \frac{s^{\phi-1}}{\Gamma(\phi)} \mathcal{L}\{S(t)\}(s).$$

This means that

$$\nu^*\left(\frac{1}{t}\right) = t^{2-\phi} \mathcal{L}^{-1}\left\{\frac{\Gamma(\phi)}{s^{\phi-1}} \nu(s)\right\}(t)$$

where $\mathcal{L}^{-1}$ is the inverse Laplace transform. This ensures the unicity of ν^*. The forms for some particular choices of marginal process are shown in Table 1. The results are surprising. A gamma marginal process arises when the directing Lévy process is a beta process and a σ-stable marginal process arises when the directing Lévy process is also a σ-stable process. Generalized gamma marginal processes lead to a directing Lévy process which is a generalization of the beta process (with a power of z which is less than 1) and rescaled to the interval $(0, 1/a)$. In fact, if we use a gamma score distribution with shape ϕ and mean $a\phi$ which has density

$$f(x) = \frac{1}{a^{\phi}\Gamma(\phi)} x^{\phi-1} \exp\left(-\frac{x}{a}\right), \tag{4.2}$$

the directing Lévy intensity is a stable beta distribution (Teh and Görür, 2009) of the form

$$\nu^*(z) = \frac{a^{\sigma+1}\sigma}{\phi} \frac{\Gamma(\phi+1)}{\Gamma(\phi+\sigma)\Gamma(1-\sigma)} z^{-\sigma-1}(1-z)^{\sigma+\phi-1}, \qquad 0 < z < 1.$$

Remark 2. This paper is focused on gamma scores but the class of compound random measures is very wide and other choices can be considered. For instance, if beta$(\alpha, 1)$ scores are selected, i.e.

$$f(x) = \alpha x^{\alpha-1} \qquad \alpha > 0, \quad 0 < x < 1,$$

Table 1. Form of directing Lévy intensity in a compound random measure which leads to particular marginal processes

$\nu^*(z)$	*Support*	*Marginal process*
$z^{-1}(1-z)^{\phi-1}$	$0<z<1$	Gamma
$\frac{\sigma\Gamma(\phi)}{\Gamma(\sigma+\phi)\Gamma(1-\sigma)}z^{-\sigma-1}$	$z>0$	σ-stable
$\frac{\sigma\Gamma(\phi)}{\Gamma(\sigma+\phi)\Gamma(1-\sigma)}z^{-\sigma-1}(1-az)^{\sigma+\phi-1}$	$0<z<1/a$	Generalized gamma

then it is possible to introduce a multivariate version of the beta process. Let $\nu(s)=\theta s^{-1}(1-s)^{\theta-1}$, $0<s<1$ and $\theta>1$, i.e. the Lévy intensity of the jumps of a beta process, then $\nu^*(z)$ is the solution of the integral equation

$$\nu(s)=\int_s^1 f(s/z)z^{-1}\,\nu^*(z)\,\mathrm{d}z, \qquad 0<s<1.$$

A simple application of the fundamental theorem of calculus leads to

$$\nu^*(z)=\theta z^{-1}(1-z)^{\theta-1}+\frac{\theta(\theta-1)}{\alpha}(1-z)^{\theta-2}$$

which is the sum of $\nu(\cdot)$, the Lévy intensity of the original beta process, and a compound Poisson process (if $\theta>1$) with intensity θ/α and jump distribution beta$(1,\theta-1)$.

It is interesting to derive the resulting multivariate Lévy intensities which can be compared with similar results in Leisen and Lijoi (2011), Leisen *et al.* (2013) and Zhu and Leisen (2014).

Theorem 2. Consider a compound random measure process with independent Ga$(\phi,1)$ distributed scores. If the compound random measure process has gamma process marginals then

$$\rho_d(s_1,\ldots,s_d)=\frac{\left(\prod_{j=1}^d s_j\right)^{\phi-1}}{\Gamma(\phi)^{d-1}}|\mathbf{s}|^{-(d\phi+1)/2}\exp\left(-\frac{|\mathbf{s}|}{2}\right)W_{\{(d-2)\phi+1\}/2,\,-d\phi/2}(|\mathbf{s}|) \tag{4.3}$$

where $|\mathbf{s}|=s_1+\ldots+s_d$ and W is the Whittaker function. If the compound random measure process has σ-stable process marginals then

$$\rho_d(s_1,\ldots,s_d)=\frac{\left(\prod_{j=1}^d s_j\right)^{\phi-1}}{\Gamma(\phi)^{d-1}}\frac{\sigma\,\Gamma(\sigma+d\phi)}{\Gamma(\sigma+\phi)\,\Gamma(1-\sigma)}|\mathbf{s}|^{-\sigma-d\phi}. \tag{4.4}$$

The result is proved in the on-line supplementary material with the following corollary.

Corollary 1. Consider a compound random measure process with independent exponentially distributed scores. If the compound random measure has gamma process marginals we recover the multivariate Lévy intensity of Leisen *et al.* (2013):

$$\rho_d(s_1,\ldots,s_d)=\sum_{j=0}^{d-1}\frac{(d-1)!}{(d-1-j)!}|\mathbf{s}|^{-j-1}\exp(-|\mathbf{s}|).$$

Table 2. Lévy intensity of the marginal process in a compound random measure with the directing Lévy processes defined in equations (2.1)–(2.4)

$\nu(s)$	*Directing Lévy process*
$2\frac{1}{\Gamma(\phi)}s^{\phi/2-1}K_{\phi}(2\sqrt{s})$	Gamma
$\frac{\Gamma(\sigma+\phi)}{\Gamma(\phi)}\frac{\sigma}{\Gamma(1-\sigma)}s^{-1-\sigma}$	σ-stable
$\frac{\Gamma(\theta+1)}{\Gamma(\phi)}s^{-1}\exp(-s)U(\theta-\phi,1-\phi,s)$	Beta
$2\frac{1}{\Gamma(\phi)}\frac{\sigma}{\Gamma(1-\sigma)}s^{(\phi-\sigma)/2-1}a^{(\sigma+\phi)/2}K_{\sigma+\phi}\{2\sqrt{(as)}\}$	Generalized gamma

Otherwise, if σ-stable marginals are considered then we recover the multivariate vector that was introduced in Leisen and Lijoi (2011) and Zhu and Leisen (2014):

$$\rho_d(s_1,\ldots,s_d)=\frac{(\sigma)_d}{\Gamma(1-\sigma)}|\mathbf{s}|^{-\sigma-d}.$$

Alternatively, we can specify ν^* and derive ν. The forms for some particular processes are shown in Table 2 where U is the confluent hypergeometric function of the second kind and K is the modified Bessel function of the second kind.

Remark 3. There are several special cases if ν^* is the Lévy intensity of a beta process. Firstly, $U(\theta-\phi,1-\phi,s)=1$ if $\theta=\phi$ and ν is the Lévy intensity of a gamma process. If $\phi=2\theta-1$,

$$U(\theta-\phi,1-\phi,s)=\pi^{-1/2}\exp(s/2)s^{1/2-\theta+\phi}K_{\theta-1/2}(s/2).$$

When $\theta=1$, $U(1-\phi,1-\phi,s)=\exp(s)\int_s^{\infty}u^{-(1-\phi)}\exp(-u)\mathrm{d}u$. The limits as $s\to 0$ are

$$U(\theta-\phi,1-\phi,s)\to\begin{cases}\Gamma(\phi)/\Gamma(\theta)+O(|s|^{\phi}) & 0<\phi<1,\\ 1/\Gamma(1+\theta-\phi)+O\{|s\log(s)|\} & \phi=1,\\ \Gamma(\phi)/\Gamma(\theta)+O(|s|) & \phi>1.\end{cases}$$

Therefore, these processes have a Lévy intensity that is similar to the Lévy intensity of the gamma process close to zero for any choice of ϕ and θ. The tails of the Lévy intensity are exponential. Therefore, the process has similar properties to those of the gamma process.

Remark 4. The generalized gamma process contains some special cases and the Lévy intensities of the marginal process for these processes are shown in Table 2. With a generalized gamma directing Lévy process, it is straightforward to show that

$$\nu(s)\approx\sigma\frac{\Gamma(\sigma+\phi)}{\Gamma(\phi)\Gamma(1-\sigma)}s^{-\sigma-1}$$

for small s. Therefore, the Lévy intensity close to zero is similar to the Lévy intensity of a σ-stable process with parameter σ. For large s, we have

$$\nu(s)s\propto\sqrt{\pi}\frac{1}{\Gamma(\phi)}\frac{\sigma}{\Gamma(1-\sigma)}(as)^{(\phi+\sigma)/2-1/4}s^{-1-\sigma}\exp\{-2\sqrt{(as)}\}.$$

Therefore, the tails will decay like $\exp(-s^{1/2})$.

The next theorems will provide an expression of the Laplace exponent when the scores are gamma distributed with $\phi\geqslant 1$ such that $\phi\in\mathbb{N}$. We want to stress the importance of the Laplace

transform in the Bayesian non-parametric setting. Indeed, it provides a basis for deriving theoretical results about the prior. The Laplace transform can be used to derive some distributional properties such as correlation, partition structure and mixed moments. Additionally, we shall see that the Laplace transform plays a role in the novel sampler that is proposed in this paper.

Theorem 3. Consider a compound random measure process with independent $\mathrm{Ga}(\phi,1)$-distributed scores. Suppose that $\phi \geqslant 1$ such that $\phi \in \mathbb{N}$. Let $\boldsymbol{\lambda} \in (\mathbb{R}^+)^d$ be a vector such that it consists of $l \leqslant d$ distinct values denoted as $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_l)$ with respective multiplicities $\mathbf{n} = (n_1, \ldots, n_l)$. Then

$$\psi_{\rho,d}(\boldsymbol{\lambda}) = \psi_{\rho,d}(\tilde{\boldsymbol{\lambda}}, \mathbf{n}) = \frac{\Gamma(\phi)^l}{\prod_{i=1}^{l} \{\tilde{\lambda}_i^{\phi-1} \Gamma(n_i\phi)\}} \left(\prod_{i=1}^{l} \frac{\partial^{(n_i-1)\phi}}{\partial^{(n_i-1)\phi} \tilde{\lambda}_i} \right) \Upsilon_l^{\phi}(\tilde{\boldsymbol{\lambda}}) \prod_{i=1}^{l} \tilde{\lambda}_i^{n_i\phi-1},$$

where

$$\Upsilon_l^{\phi}(\tilde{\boldsymbol{\lambda}}) = \int \left\{ 1 - \prod_{i=1}^{l} \frac{1}{(1+z\tilde{\lambda}_i)^{\phi}} \right\} \nu^*(z)\, \mathrm{d}z.$$

The proof of theorem 3 is based on the result that is provided in theorem 1 since the moment-generating function of a gamma distribution exists and it is explicit. To compute the expression of $\Upsilon_l^{\phi}(\tilde{\boldsymbol{\lambda}})$ we need to define the set

$$A_{\phi,j} = \{\mathbf{k} \in \{1, \ldots, \phi\}^j : |\mathbf{k}| = \phi\} \qquad \phi \geqslant j.$$

Theorem 4. Consider a compound random measure process with independent $\mathrm{Ga}(\phi,1)$-distributed scores. Suppose that $\phi \geqslant 1$ such that $\phi \in \mathbb{N}$. Let $\Lambda(\tilde{\boldsymbol{\lambda}}, \mathbf{z}) = (1 - \Sigma_{h=1}^{j-1} z_h)\tilde{\lambda}_{i_j} + \Sigma_{h=1}^{j-1} \tilde{\lambda}_{i_h} z_h$ be a function defined on the $(j-1)$-dimensional simplex

$$\Delta_{j-1} = \{\mathbf{z} \in (0,1)^{j-1} : z_1 + \ldots + z_{j-1} < 1\}$$

with the convention that $\Delta_0 = [0,1]$. Let

$$a_i(\tilde{\boldsymbol{\lambda}}) = \frac{\tilde{\lambda}_i^{l-1}}{\prod_{\substack{j=1 \\ j\neq i}}^{l} (\tilde{\lambda}_i - \tilde{\lambda}_j)} \qquad i = 1, \ldots, l;$$

then

$$\Upsilon_l^{\phi}(\tilde{\boldsymbol{\lambda}}) = \begin{cases} \phi! \sum_{j=1}^{\phi} \sum_{\mathbf{k} \in A_{\phi,j}} \sum_{0<i_1<i_2<\ldots<i_j \leqslant l} \dfrac{a_{i_1}^{k_1}(\tilde{\boldsymbol{\lambda}}) \ldots a_{i_j}^{k_j}(\tilde{\boldsymbol{\lambda}})}{k_1! \ldots k_j!} C(i_1, \ldots, i_j; \mathbf{k}; \tilde{\boldsymbol{\lambda}}) & \text{if } l > 1, \\ \psi(\lambda_1) & \text{if } l = 1 \end{cases}$$

where

$$C(i_1, \ldots, i_j; \mathbf{k}; \tilde{\boldsymbol{\lambda}}) = \Gamma(\phi) \int_{\Delta_{j-1}} \left\{ \left(1 - \sum_{h=1}^{j-1} z_h \right)^{k_j} \prod_{h=1}^{j-1} \frac{z_h^{k_h-1}}{\Gamma(k_h)} \right\} \psi\{\Lambda(\tilde{\boldsymbol{\lambda}}, \mathbf{z})\}\, \mathrm{d}\mathbf{z}.$$

For the above integral we assume the usual convention that $\Sigma_i^j = 0$ and $\Pi_i^j = 1$ whenever $i > j$.

In the following corollary, the expression of the Laplace exponent is recovered for the special case of a compound random measure with independent exponentially distributed scores.

Corollary 2. Consider a compound random measure process with independent exponentially distributed scores. It follows that

$$\psi_{\rho,d}(\boldsymbol{\lambda}) = \psi_{\rho,d}(\tilde{\boldsymbol{\lambda}}, \mathbf{n}) = \left\{ \prod_{i=1}^{l} \frac{1}{\Gamma(n_i)} \frac{\partial^{(n_i-1)}}{\partial^{(n_i-1)} \tilde{\lambda}_i} \right\} \Upsilon_I(\tilde{\boldsymbol{\lambda}}) \prod_{i=1}^{l} \tilde{\lambda}_i^{(n_i-1)},$$

$$\Upsilon_l(\tilde{\boldsymbol{\lambda}}) = \begin{cases} \sum_{i=1}^{l} a_i(\tilde{\boldsymbol{\lambda}}) \psi(\lambda_i) & \text{if } l > 1, \\ \psi(\lambda_1) & \text{if } l = 1. \end{cases}$$

The proof of corollary 2 is omitted since it is a direct application of the results of the previous theorems. Note that, if the vector has gamma process marginals, i.e. $\psi(\lambda_i) = \log(1+\lambda_i)$, then we recover the results in Leisen *et al.* (2013). If the vector has σ-stable process marginals, i.e. $\psi(\lambda_i) = \lambda_i^{\sigma}$, then we recover the result in Leisen and Lijoi (2011) and Zhu and Leisen (2014).

Finally, we close the section with some results about the dependence structure of compound random measure processes. A useful description of the dependence of a vector of CRMs is given by the Lévy copula. A Lévy copula is a mathematical tool that allows the construction of multivariate Lévy intensities with fixed marginals; see the on-line supplementary material. The following theorem displays the underlying Lévy copula of a compound random measure.

Theorem 5. Let ρ_d be the compound random measure defined in equation (3.3) and let F be the the distribution function of f. The underlying Lévy copula of the compound random measure is

$$C(s_1, \ldots, s_d) = \int \nu^*(z) \prod_{j=1}^{d} [1 - F\{z^{-1} U^{-1}(s_j)\}] \mathrm{d}z$$

where U^{-1} is the inverse of the tail integral function $U(x) := \int_x^{\infty} \nu(s)\,\mathrm{d}s$.

Furthermore, it is possible to prove a result that is similar to proposition 5 in Leisen *et al.* (2013). This result gives a close formula for the mixed moments of two dimensions of a compound random measure process. The result is expressed in terms of an ordering on sets $\mathbf{0} \prec \mathbf{s}_1 \prec \ldots \prec \mathbf{s}_j$ which is defined in Constantines and Savits (1996).

Theorem 6. Consider a compound random measure process with independent $\mathrm{Ga}(\phi, 1)$-distributed scores. Let $\mathbf{q} = (q_1, \ldots, q_d)$ and let $p_j(\mathbf{q}, k)$ be the set of vectors $(\boldsymbol{\eta}, \mathbf{s}_1, \ldots, \mathbf{s}_j)$ such that the co-ordinates of $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_j)$ are positive and such that $\Sigma_{i=1}^{j} \eta_i = k$. Moreover, $\mathbf{s}_i = (s_{1,i}, \ldots, s_{d,i})$ are vectors such that $\mathbf{0} \prec \mathbf{s}_1 \prec \ldots \prec \mathbf{s}_j$ and $\Sigma_{i=1}^{j} \eta_i (s_{1,i} + \ldots + s_{d,i}) = k = q_1 + \ldots + q_d$. Then,

$$\mathbb{E}\left[\prod_{i=1}^{d} \tilde{\mu}_i(A)^{q_i} \right] = q_1! \ldots q_d! \sum_{k=1}^{|\mathbf{q}|} \alpha(A)^k \times \sum_{j=1}^{|\mathbf{q}|} \sum_{p_j(\mathbf{q},k)} \prod_{i=1}^{j} \frac{1}{\eta_i!} \left[\left\{ \prod_{l=1}^{d} \frac{(\phi)_{s_{l,i}}}{s_{l,i}!} \right\} \int z^{s_{1,i}+\ldots+s_{d,i}} \nu^*(z)\,\mathrm{d}z \right]^{\eta_i}$$

where $|\mathbf{q}| = q_1 + \ldots + q_d$.

Remark 5. For instance, suppose that the compound random measure process has generalized gamma process marginals. Then,

$$\int z^{s_{1,i}+\ldots+s_{d,i}} \nu^*(z)\,\mathrm{d}z = \frac{\sigma a^{\sigma-(s_{1,i}+\ldots+s_{d,i})}}{\Gamma(1-\sigma)} B(k-\sigma-1, \sigma+\phi).$$

5. Normalized compound random measures

Vectors of correlated random probability measures can be defined by normalizing each dimen-

sion of a compound random measure process. This will be called a normalized compound random measure and is defined by a score distribution, a directing Lévy process and a centring measure of the compound random measure. The results derived in Table 1 can be used to define a normalized compound random measure with a particular marginal process. For example, a normalized compound random measure with Dirichlet process marginals arises by normalizing each dimension of a compound random measure with gamma process marginals.

In specifying a normalized compound random measure prior, it is useful to have a method of choosing the parameters of the score distribution to give a particular level of dependence. We describe two possible methods. It is possible to compute the covariance of two dimensions of a normalized compound random measure process. Indeed, following Leisen *et al.* (2013),

$$\operatorname{cov}\{\tilde{p}_1(A), \tilde{p}_2(B)\} = \left\{ \alpha(A \cap B) - \frac{\alpha(A)\,\alpha(B)}{\alpha(\mathbb{X})} \right\} \times \int_{(\mathbb{R}^+)^2} g_\rho(1,1;\lambda_1,\lambda_2) \exp\{-\alpha(\mathbb{X})\,\psi_\rho(\lambda_1,\lambda_2)\} \mathrm{d}\lambda_1 \mathrm{d}\lambda_2 \tag{5.1}$$

where

$$g_\rho(1,1;\lambda_1,\lambda_2) = \int_{(0,\infty)^d} s_1, s_2 \exp\{-\psi_{\rho,2}(\lambda_1,\lambda_2)\}\, \rho_2(s_1,s_2)\, \mathrm{d}s_1\, \mathrm{d}s_2. \tag{5.2}$$

(An expression for g_ρ for a compound random measure is provided in the on-line supplementary material.)

This result can be used to specify any parameters of the score distribution (or a prior for those parameters). Alternatively, if a compound random measure with independent scores is used, the ratio of a jump's height in the ith and jth dimensions has the same distribution as the ratio of two independent random variables following the score distribution. For example, if the scores are independent and follow a gamma distribution with shape ϕ, this ratio follows an F-distribution with ϕ and ϕ degrees of freedom.

5.1. Computational methods

We describe methods for fitting a non-parametric mixture model where the mixing measure is given a normalized compound random measure prior. We assume that the data can be divided into d groups and $y_{j,1}, \ldots, y_{j,n_j}$ are the observations in the jth group. The data are modelled as

$$y_{j,i} \overset{\text{ind.}}{\sim} k(y_{j,i}|\zeta_{j,i}), \qquad \zeta_{j,i} \sim \tilde{p}_j, \quad i = 1,2,\ldots,n_j, \quad j = 1,\ldots,d,$$

where $k(y|\theta)$ is a probability density function for y with parameter θ and $\tilde{p}_1, \ldots, \tilde{p}_d$ are given a normalized compound random measure prior. Using the notation of equation (3.4), we write

$$\tilde{p}_j = \frac{\tilde{\mu}_j}{\tilde{\mu}_j(\mathbb{X})} = \frac{\sum_{k=1}^{\infty} m_{j,k} J_k \delta_{\theta_k}}{\sum_{k=1}^{\infty} m_{j,k} J_k}.$$

Direct simulation from the posterior distribution is impossible since there are an infinite number of parameters. Several Markov chain Monte Carlo (MCMC) methods have been introduced which circumvent this problem in the class of normalized random measure mixtures. Favaro and Teh (2013) described an auxiliary variable method which involves integrating out the unnormalized random measure whereas Griffin and Walker (2011) introduced a slice sampling method. We consider extending both methods to normalized compound random measure mixtures.

We use the notation $m=(m_{j,k})$, $J=(J_1, J_2, \ldots)$ and $\theta=(\theta_1, \theta_2, \ldots)$. The posterior distribution can be expressed in a suitable form for MCMC sampling by introducing latent variables. Firstly, latent allocation variables $c=(c_{j,i})$ (for which $\zeta_{j,i}=\theta_{c_{j,i}}$) are introduced to give

$$p(y, c|m, J, \theta) = \prod_{j=1}^{d}\prod_{i=1}^{n_j}\left\{k(y_{j,i}|\theta_{c_{j,i}})m_{j,c_{j,i}}J_{c_{j,i}}\Big/\sum_{k=1}^{\infty}m_{j,k}J_k\right\}$$
$$= \prod_{j=1}^{d}\prod_{i=1}^{n_j}k(y_{j,i}|\theta_{c_{j,i}})m_{j,c_{j,i}}J_{c_{j,i}}\Big/\left(\sum_{k=1}^{\infty}m_{j,k}J_k\right)^{n_j}. \tag{5.3}$$

Secondly, latent variables $v=(v_1, \ldots, v_d)$ are introduced to define

$$p(y, c, v|m, J, \theta) = \prod_{j=1}^{d}\left\{\prod_{i=1}^{n_j}k(y_{j,i}|\theta_{c_{j,i}})m_{j,c_{j,i}}J_{c_{j,i}}\right\}\prod_{j=1}^{d}\left\{\frac{1}{\Gamma(n_j)}v_j^{n_j-1}\right\}$$
$$\times\exp\left(-\sum_{j=1}^{d}v_j\sum_{k=1}^{\infty}m_{j,k}J_k\right).$$

Integrating over v (using the identity $\{1/\Gamma(n)\}v^{n-1}\exp(-vx)=x^{-n}$) gives expression (5.3).

5.1.1. *Marginal method*

The approach of Favaro and Teh (2013) relies on an analytical form for $p(y, v, c)$ which is available for the normalized random measures with independent increments mixtures by using results of James *et al*. (2009). Suppose that $\{c_{j,i}\}$ takes K distinct values, that $a_{j,k}$ is the number of observations in the jth group allocated to the kth distinct value (i.e. $a_{k,j}=\sum_{i=1}^{n_j}I(c_{j,i}=k)$) and define $a_k=(a_{k,1}, \ldots, a_{k,d})$. Extending the results of James *et al*. (2009) and Favaro and Teh (2013) to vectors of normalized random measures (as in Section 2.1) leads to

$$p(y, v, c) = \prod_{j=1}^{d}\frac{1}{\Gamma(n_j)}v_j^{n_j-1}\exp\{-\psi_{\rho,d}(v)\}\prod_{k=1}^{K}\kappa_{a_k}(v)\prod_{k=1}^{K}g(\{y_{j,i}|c_{j,i}=k\})$$

where

$$\psi_{\rho,d}(v) = \int\left\{1-\exp\left(-\sum_{i=1}^{d}v_is_i\right)\right\}\rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_d),$$
$$\kappa_a(v) = \int\left(\prod_{j=1}^{d}s_j^{a_j}\right)\exp\left(-\sum_{i=1}^{d}v_is_i\right)\rho_d(\mathrm{d}s_1, \ldots, \mathrm{d}s_d)$$

and

$$g(y) = \int\prod k(y_{j,i}|\theta)\,\alpha(d\theta).$$

If the vector of the normalized random measures is chosen to be a normalized compound random measure with independent gamma scores then

$$\kappa_a(v) = \int\left(\prod_{j=1}^{d}s_j^{a_j}\right)\exp\left(-\sum_{i=1}^{d}v_is_i\right)z^{-d}\prod_{j=1}^{d}f(s_j/z)\mathrm{d}s_1\ldots\mathrm{d}s_d\,\nu^*(\mathrm{d}z)$$
$$= \int z^{\sum_{j=1}^{d}a_j}\prod_{j=1}^{d}\int\{s_j^{a_j}\exp(-v_jzs_j)\,f(s_j)\,\mathrm{d}s_j\}\,\nu^*(\mathrm{d}z)$$
$$= \int z^{\sum_{j=1}^{d}a_j}\prod_{j=1}^{d}\tau_{a_j}(z, v_j)\,\nu^*(\mathrm{d}z)$$

where

$$\tau_a(z,v)=\int s^a\exp(-vzs)\,f(s)\,\mathrm{d}s.$$

and theorem 1 provides the expression

$$\psi_{\rho,d}(v)=\int\left\{1-\prod_{j=1}^{d}M_z^f(-s_j)\right\}\nu^*(z)\,\mathrm{d}z.$$

If f is chosen to be a gamma distribution with shape parameter ϕ,

$$\tau_a(z,v)=\int s^a\exp(-vzs)\,f(s)\,\mathrm{d}s=\frac{\Gamma(a+\phi)}{\Gamma(\phi)}(1+vz)^{-a-\phi}.$$

Two algorithms can be defined. One is suitable for conjugate mixtures where $g(y)$ can be calculated analytically and a second algorithm is suitable for non-conjugate mixtures where $g(y)$ cannot be calculated analytically.

In the case of a conjugate mixture model, the steps of the algorithm are as follows.

5.1.1.1. Updating $c_{j,i}$. Let $C_k^{-(j,i)}=\{y_{l,m}|c_{l,m}=k,(l,m)\neq(j,i)\}$ and $K^{-(j,i)}$ be the number of distinct values of $\{c_{l,m}|(l,m)\neq(j,i)\}$. The parameter $c_{j,i}$ is updated from the discrete distribution

$$p(c_{j,i}=k)\propto\begin{cases}\dfrac{\kappa_{a_k+r}(v)\,g(C_k^{-(j,i)}\cup\{y_{j,i}\})}{\kappa_{a_k}(v)\,g(C_k^{-(j,i)})} & 1\leqslant k\leqslant K^{-(j,i)},\\ \kappa_r(v)\,g(y_{j,i}) & k=K^{-(j,i)}+1\end{cases}$$

where r is a d-dimensional vector with $r_m=1$ if $m=j$ and $r_m=0$ otherwise. For independent $\mathrm{Ga}(\phi,1)$ scores,

$$\frac{\kappa_{a_k+r}(v)}{\kappa_{a_k}(v)}=(a_{j,k}+\phi)\frac{\displaystyle\int z^{\sum_{m=1}^d a_{m,k}+1}(1+v_jz)^{-a_{j,k}-1-\phi}\prod_{m=1;m\neq j}^{d}(1+v_mz)^{-a_{m,k}-\phi}\nu^*(z)\,\mathrm{d}z}{\displaystyle\int z^{\sum_{m=1}^d a_{m,k}}\prod_{m=1}^{d}(1+v_mz)^{-a_{m,k}-\phi}\nu^*(z)\,\mathrm{d}z}$$

and

$$\kappa_r(v)=\phi\int z(1+v_jz)^{-1-\phi}\prod_{m=1;m\neq j}^{d}(1+v_mz)^{-\phi}\nu^*(z)\,\mathrm{d}z.$$

5.1.1.2. Updating v_j. The full conditional distribution of v_j is proportional to

$$v_j^{n_j-1}\exp\{-\psi_{\rho,d}(v)\}\prod_{k=1}^{K}\kappa_{a_k}(v).$$

This parameter can be updated by using an adaptive Metropolis–Hastings random walk (Atchadé and Rosenthal, 2005).

5.1.1.3. Updating parameters of f. The full conditional distribution of the parameters of f is proportional to

$$\exp\{-\psi_{\rho,d}(v)\}\prod_{k=1}^{K}\kappa_{a_k}(v).$$

This parameter can be updated by using an adaptive Metropolis–Hastings random walk (Atchadé and Rosenthal, 2005).

In the case of non-conjugate mixtures, Favaro and Teh (2013) defined an auxiliary variable method which introduces the distinct values $\theta_1,\ldots,\theta_K$ into the sampler and M potential distinct values for empty clusters $\theta'_1,\ldots,\theta'_M$.

5.1.1.4. Updating $c_{j,i}$. A set of values $\theta'_1,\ldots,\theta'_M$ is formed. If $c_{j,i}$ is a singleton (i.e. $c_{j,i}\neq c_{k,m}$ for $(j,i)\neq(k,m)$), set $\theta'_1=\theta_{c_{j,i}}$ and sample $\theta'_j\sim\alpha/\alpha(\mathbb{X})$ for $j=2,\ldots,M$. Otherwise, sample $\theta'_j\sim\alpha/\alpha(\mathbb{X})$ for $j=1,\ldots,M$. Sample $c^*_{j,i}$ from the following discrete distribution:

$$p(c^*_{j,i}=k)\propto\begin{cases}\dfrac{\kappa_{a_k+r}(v)}{\kappa_{a_k}(v)}k(y_{j,i}|\theta_k) & 1\leqslant k\leqslant K^{-(j,i)},\\ \dfrac{\alpha(\mathbb{X})}{M}\kappa_r(v)\,k(y_{j,i}|\theta'_{k-K^{-(j,i)}}) & k=K^{-(j,i)}+1,\ldots,K^{-(j,i)}+M.\end{cases}$$

If $c^*_{j,i}\leqslant K^{-(j,i)}$, set $c_{j,i}=c^*_{j,i}$. Otherwise, set $c_{j,i}=K^{-(j,i)}+1$ and $\theta_{K^{-(j,i)}+1}=\theta'_{c^*_{j,i}-K^{-(j,i)}}$.

5.1.1.5. Updating θ_k. The full conditional density of θ_k is proportional to

$$\alpha(\theta_k)\prod_{\{(j,i)|c_{j,i}=k\}}k(y_{j,i}|\theta_k).$$

The full conditional distributions of v_j and any parameters of f are unchanged from the algorithm for conjugate mixture models.

5.1.2. Slice sampling method

We introduce $u=(u_{j,i})$ and define

$$p(y,c,v,u|m,J,\theta)=\prod_{j=1}^{d}\left\{\prod_{i=1}^{n_j}k(y_{j,i}|\theta_{c_{j,i}})m_{j,c_{j,i}}I(u_{j,i}<J_{c_{j,i}})\right\}\prod_{j=1}^{d}\left\{\frac{1}{\Gamma(n_j)}v_j^{n_j-1}\right\}$$
$$\times\exp\left(-\sum_{j=1}^{d}v_j\sum_{k=1}^{\infty}m_{j,k}J_k\right).$$

Integrating over u and v gives expression (5.3). A similar form was derived in Griffin and Walker (2011). This form of the likelihood is still not suitable for MCMC sampling since it involves all jumps. To avoid this, we define $L=\min_{i=1,\ldots,n_j;j=1,\ldots,d}\{u_{j,i}\}$ and divide the jumps into two disjoints sets: $A^\dagger=\{(J^\dagger_k,m^\dagger_{1,k},\ldots,m^\dagger_{d,k})|J^\dagger_k>L\}$ and $A^*=\{(J^*_k,m^*_{1,k},\ldots,m^*_{d,k})|J^*_k\leqslant L\}$. The set $A^\dagger$ has a finite number of elements which are denoted K and A^* has an infinite number of elements. Integrating over A^* leads to a posterior which is suitable for MCMC sampling and has the form

$$\prod_{j=1}^{d}\left\{\prod_{i=1}^{n_j}k(y_{j,i}|\theta_{c_{j,i}})m^\dagger_{j,c_{j,i}}I(u_{j,i}<J_{c_{j,i}})\right\}\prod_{j=1}^{d}\left\{\frac{1}{\Gamma(n_j)}v_j^{n_j-1}\right\}\times\exp\left(-\sum_{j=1}^{d}v_j\sum_{k=1}^{K}m^\dagger_{j,k}J^\dagger_k\right)$$
$$\times\mathbb{E}\left[\exp\left(-\sum_{j=1}^{d}v_j\sum_{k=1}^{\infty}m^*_{j,k}J^*_k\right)\right]. \tag{5.4}$$

An MCMC scheme using this form of likelihood leads to a random truncation of the normalized compound random measures process at each iteration but does not introduce a truncation error since integrating over the latent variables leads to the correct marginal posterior.

The expectation in expression (5.4) can be expressed in terms of a univariate integral by using a variation on theorem 1, giving

$$-\log\left\{\mathbb{E}\left[\exp\left(-\sum_{j=1}^{d} v_j \sum_{k=1}^{\infty} m^*_{j,k} J^*_k\right)\right]\right\} = \int_0^L \left\{1 - \prod_{j=1}^{d} M_z^f(-v_j)\right\} \nu^*(z)\,\mathrm{d}z.$$

The full conditional distributions, a general discussion of the methods for updating parameters and details of the implementation for specific processes are given in the on-line supplementary material.

6. Illustrations

The clinical studies CALGB 8881 (Lichtman *et al.*, 1993) and CALGB 9160 (Budman *et al.*, 1998) looked at the response of patients to different anticancer drug therapies. The response was white blood cell count and patients had between four and 25 measurements taken over the course of the trial. The data were previously analysed by Müller *et al.* (2004), who fitted a non-linear random-effects model for the patient's response over time. The model assumes that the mean response at time t with parameters $\theta = (z_1, z_2, z_3, \tau_1, \tau_2, \beta_0, \beta_1)$ is given by

$$f(\theta, t) = \begin{cases} z_1 & t < \tau_1, \\ rz_1 + (1-r)\, g(\theta, \tau_2) & \tau_1 \leqslant t < \tau_2, \\ g(\theta, t) & t \geqslant \tau_2 \end{cases}$$

where $r = (\tau_2 - t)/(\tau_2 - \tau_1)$ and $g(\theta, t) = z_2 + z_3/[1 + \exp\{\beta_0 - \beta_1(t - \tau_2)\}]$. There were nine different combinations of the anticancer agent cyclophosphamide, the drug granulocyte macrophage colony stimulating factor and amifostine which are summarized in Table 3.

Summaries of the data are available as part of the `DPpackage` in R (Jara *et al.*, 2011) where a non-linear regression model is fitted with $f(\theta_{j,i}, t)$ as the mean for the ith patient in the jth

Table 3. Levels of cyclophosphamide, granulocyte macrophage colony stimulating factor and amifostine across the nine groups†

Group	*Cyclophosphamide* ($g\,m^{-2}$)	*Granulocyte macrophage colony stimulating factor* ($\mu g\,kg^{-1}$)	*Amifostine*	*Study*	*Number of patients*
1	1.5	10.0	0	1	6
2	3.0	5.0	0	2	28
3	3.0	5.0	1	2	18
4	3.0	2.5	0	1	6
5	3.0	5.0	0	1	6
6	3.0	10.0	0	1	6
7	4.5	5.0	0	1	12
8	4.5	10.0	0	1	10
9	6.0	5.0	0	1	6

†The CALGB 8881 study is indicated as study 1 and the CALGB 9160 study as study 2.

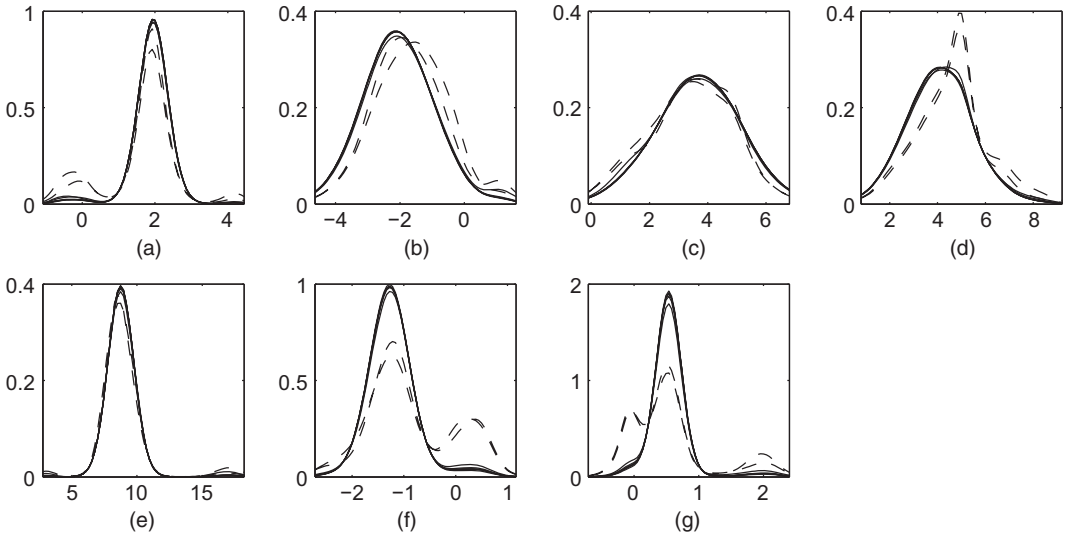

Fig. 2. Posterior mean marginal densities of each parameter in the CALGB example (———, group in the CALGB 8881 study; - - - - - -, group in the CALGB 9160 study): (a) z_1; (b) z_2; (c) z_3; (d) t_1; (e) t_2; (f) β_0; (g) β_1

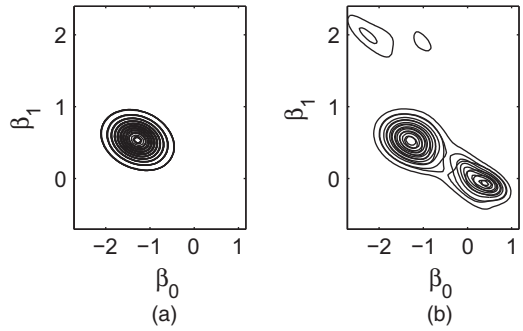

Fig. 3. Posterior mean joint densities of β_0 and β_1 in the CALGB example for groups in (a) the CALGB 8881 study and (b) the CALGB 9160 study

group. We shall consider the differences in the distribution of the estimated values $\hat{\theta}_{j,i}$ across the nine studies. It is assumed that

$$\hat{\theta}_{j,i} \sim N(\mu_{j,i}, \Sigma_{j,i}), \qquad (\mu_{j,j}, \Sigma_{j,j}) \sim \tilde{p}_j$$

where $\tilde{p}_1, \ldots, \tilde{p}_9$ are given a normalized compound random measures process prior with independent $\Gamma(\phi, 1)$-distributed scores and Dirichlet process marginals. The centring measure α is $N(\mu|\hat{\theta}, 100\Sigma)\,\mathrm{IW}(\Sigma|14, 4/9 \times \hat{\Sigma})$ where $\hat{\theta}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance matrix of $\hat{\theta}$. This implies a prior mean of $1/9 \times \hat{\Sigma}$. The parameter ϕ is given an exponential prior with mean 1.

The results of the analysis are illustrated in Fig. 2 which shows the posterior mean marginal density of each parameter. The results within each study are very similar with the main difference occurring between the two studies. All densities are very similar for the parameters z_1, z_2, z_3 and t_2. There is a slight difference in the distribution for t_1 but much bigger differences for parameters β_0 and β_1. The results for the CALGB 8881 study are unimodal whereas those for

CALGB 9160 includes additional modes at 0.5 for β_0 and -0.5 and 2 for β_1. Fig. 3 shows the posterior mean joint density of β_0 and β_1, which shows a bimodal distribution for the CALGB 9160 study with one mode at roughly $(-1.5, 0.5)$ (which is the mode for the CALGB 8881 study) and a second mode at roughly $(-0.5, 0)$. This suggests that the CALGB 9160 data may contain two groups who responded differently. The posterior median of ϕ was 1.03 with a 95% highest posterior density region of $(0.46, 2.36)$.

7. Discussion

The modelling of dependent random measures has been an extremely active area of research for more than 15 years beginning with the seminal work of MacEachern (1999). Much of the work has concentrated on dependent random probability measures with several general approaches developed in the literature. Using the notation of equation (1.1), initial work considered approaches where $w_i(x) = w_i$ and dependence is modelled through the atom location $\theta_i(x)$. This implies that cluster sizes will be similar for all values of x and so leads to a specific form of dependence. Alternatively, many researchers have used $\theta_i(x) = \theta$ for all x with dependence modelled through the weights, often by using a stick breaking construction where $w_i(x) = V_i(x)\Pi_{j<i}\{1 - V_j(x)\}$; see for example Dunson (2010) for a review. This usually leads to computationally tractable methods which either extend random truncation methods such as retrospective sampling (Papaspiliopoulos and Roberts, 2008), or slice sampling (Kalli *et al.*, 2011) or develop truncation ideas for Dirichlet process mixtures (Ishwaran and James, 2001). However, stick breaking approaches have some limitations for modelling. The construction implies a stochastic ordering so that $w_1(x)$ will tend to be the largest weight for all x. This can be inappropriate for some regression problems where we would like different components to have large weights for different values of x. The correlation is usually built on $V_j(x)$ and so $w_i(x)$ is a non-linear function of many correlated processes. This can lead to a dependence structure on $w_i(x)$ which is difficult to interpret. Analytical results such as generalizations of the exchangeable partition probability function are usually impossible to derive for these priors. These methods can often be applied to problems where $\mathcal{X}$ is continuous or discrete. Other priors are restricted to a discrete $\mathcal{X}$. One approach builds a hierarchy of non-parametric processes (see Teh and Jordan (2010) for a review) leading from the seminal work of Teh *et al.* (2006) on hierarchical Dirichlet processes. For example, a two-level hierarchical model could be constructed by assuming that the distributions for each group are conditionally independent draws from a non-parametric prior which is centred on a process which is itself given a non-parametric prior. This leads to the same correlation *a priori* between the distribution for each value in $\mathcal{X}$ (although more complicated hierarchical structures could be introduced to allow different correlation within subsets of $\mathcal{X}$). Posterior simulation is usually implemented by using the Chinese restaurant franchise algorithm.

Compound random measures are defined using a CRM and a finite dimensional score distribution. For a given marginal process, the dependence between the distributions is controlled by the choice of a finite dimensional score distribution. The compound random measure in its most general form is very flexible and allows both hierarchical and regression models. The class of normalized compound random measures includes many previously described priors and provides a useful framework for understanding the links between these priors clearer. This paper has concentrated on priors where the dimensions of the scores are independent, which allows their properties to be derived. If the moment-generating function of the marginal score distributions is available analytically, posterior computation for a normalized compound random measures mixture model can be carried out by using an augmented Pólya urn scheme or a

slice sampler and several useful analytical expressions can be derived. More general, compound random measures type models where the scores are given by a regression were discussed by Ranganath and Blei (2015) who used a truncation of the infinite dimensional parameter and variational Bayes methods to make inference. In future work, we intend to extend both the Pólya urn scheme and the slice sampler to regression models.

In this paper, we have concentrated on the case where the scores are independent and gamma distributed. This allows the dependence between the measures in different dimensions to be modelled by the shape parameter of the gamma distribution. In this case, we show how compound random measures can be constructed with gamma, σ-stable and generalized gamma process marginals. Importantly, the modelling of dependence between random measures can be achieved by the modelling of dependence between random variables and so greatly reduces the difficulty of specifying a prior for a particular problem. Future work will consider studying these classes of compound random measures.

Acknowledgements

Fabrizio Leisen was supported by the European Community's seventh framework programme (FP7/2007-2013) under grant agreement 630677 and Jim E. Griffin was supported by Engineering and Physical Sciences Research Council 'Novel technologies for cross-disciplinary research' grant EP/I036575/I. The authors acknowledge the Cancer and Leukemia Group B for the data that were used in the illustration.

References

Atchadé, Y. F. and Rosenthal, J. S. (2005) On adaptive Markov Chain Monte Carlo algorithms. *Bernoulli*, **11**, 815–828.

Barndorff-Nielsen, O. E., Pedersen, J. and Sato, K. (2001) Multivariate subordination, self-decomposability and stability. *Adv. Appl. Probab.*, **33**, 160–187.

Bassetti, F., Casarin, R. and Leisen, F. (2014) Beta-Product dependent Pitman-Yor Processes for Bayesian inference. *J. Econmetr.*, **180**, 49–72.

Budman, D., Rosner, G., Lichtman, S., Miller, A., Ratain, M. and Schilsky, R. (1998) A randomized trial of wr-2721 (amifostine) as a chemoprotective agent in combination with high-dose cyclophosphamide and molgramostim (GM-CSG). *Cancer Therpeut.*, **1**, 164–167

Chen, C., Rao, V. A., Buntine, W. and Teh, Y. W. (2013) Dependent normalized random measures. In *Proc. Int. Conf. Machine Learning*. Cambridge: MIT Press.

Constantines, G. M. and Savits, T. H. (1996) A multivariate version of the Faa di Bruno formula. *Trans. Am. Math. Soc.*, **348**, 503–520.

Cont, R. and Tankov, P. (2004) *Financial Modelling with Jump Processes*. Boca Raton: Chapman and Hall–CRC.

Daley, D. J. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes*, vol. 1. New York: Springer.

Dunson, D. B. (2010) Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics* (eds N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker). Cambridge: Cambridge University Press.

Epifani, I. and Lijoi, A. (2010) Nonparametric priors for vectors of survival functions. *Statist. Sin.*, **20**, 1455–1484.

Favaro, S. and Teh, Y. W. (2013) MCMC for normalized random measure mixture models. *Statist. Sci.*, **28**, 335–359.

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

Foti, N. and Williamson, S. (2012) Slice sampling normalized kernel-weighted completely random measure mixture models. In *Advances in Neural Information Processing Systems*, vol. 25 (eds F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger), pp. 2240–2248. Red Hook: Curran Associates.

Griffin, J. E. (2011) The Ornstein-Uhlenback Dirichlet process and other time-varying processes for Bayesian non-parametric inference. *J. Statist. Plann Inf.*, **141**, 3648–3664.

Griffin, J. E., Kolossiatis, M. and Steel, M. F. J. (2013) Comparing distributions by using dependent normalized random-measure mixtures. *J. R. Statist. Soc.* B, **75**, 499–529.

Griffin, J. E. and Walker, S. G. (2011) Posterior simulation of normalized random measure mixtures. *J. Computnl Graph. Statist.*, **20**, 241–259.

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

James, L. F., Lijoi, A. and Prünster, I. (2009) Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.*, **36**, 76–97.
Jara, A., Hansen, T. E., Quintana, F. A., Müller, P. and Rosner, G. L. (2011) DPpackage: Bayesian semi- and nonparametric modelling in R. *J. Statist. Softwr*, **40**, no. 5.
Kalli, M., Griffin, J. E. and Walker, S. G. (2011) Slice sampling mixture models. *Statist. Comput.*, **21**, 93–105.
Kingman, J. F. C. (1967) Completely random measures. *Pacif. J. Math.*, **21**, 59–78.
Leisen, F. and Lijoi, A. (2011) Vectors of Poisson-Dirichlet processes. *J. Multiv. Anal.*, **102**, 482–495.
Leisen, F., Lijoi, A. and Spano, D. (2013) A vector of Dirichlet processes. *Electron. J. Statist.*, **7**, 62–90.
Lichtman, S. M., Ratain, M. J., Echo, D. A., Rosner, G., Egorin, M. J., Budman, D. R., Vogelzang, N. J., Norton, L. and Schilsky, R. L. (1993) Phase I trial and granulocyte-macrophage colony-stimulating factor plus high-dose cyclophosphamide given every 2 weeks: a Cancer and Leukemia Group B study. *J. Natn. Cancer Inst.*, **85**, 1319–1326.
Lijoi, A. and Nipoti, B. (2014) A class of hazard rate mixtures for combining survival data from different experiments. *J. Am. Statist. Ass.*, **109**, 802–814.
Lijoi, A., Nipoti, B. and Prünster, I. (2014a) Bayesian inference with dependent normalized completely random measures. *Bernoulli*, **20**, 1260–1291.
Lijoi, A., Nipoti, B. and Prünster, I. (2014b) Dependent mixture models: clustering and borrowing information. *Computnl Statist. Data Anal.*, **71**, 417–433.
MacEachern, S. N. (1999) Dependent non-parametric processes. *Proc. Baysn Statist. Sci. Sect. Am. Statist. Ass.*
Müller, P., Quintana, F. and Rosner, G. (2004) A method for combining inference across related nonparametric Bayesian models. *J. R. Statist. Soc.* B, **66**, 735–749.
Papaspiliopoulos, O. and Roberts, G. O. (2008) Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, **95**, 169–186.
Pitman, J. and Yor, M. (1997) The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.*, **25**, 855–900.
Ranganath, R. and Blei, D. M. (2015) Correlated random measures. *Preprint arXiv:1507.00720*. Princeton University, Princeton.
Regazzini, E., Lijoi, A. and Prünster, I. (2003) Distributional results for means of normalized random measures with independent increments. *Ann. Statist.*, **31**, 560–585.
Rosinski, J. (2007) Tempered stable processes. *Stoch. Processes Appl.*, **117**, 677–707.
Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.
Teh, Y. W. and Görür, D. (2009) Indian buffet proceses with power-law behavior. In *Advances in Neural Information Processing Systems*, vol. 22 (eds Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams and A. Culotta), pp. 1838–1846.
Teh, Y. W. and Jordan, M. I. (2010) Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics* (eds N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker), pp. 158–207. Cambridge: Cambridge University Press.
Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006) Hierarchical Dirichlet processes. *J. Am. Statist. Ass.*, **101**, 1566–1581.
Zhu, W. and Leisen, F. (2014) A multivariate extension of a vector of Poisson-Dirichlet processes. *J. Nonparam. Statist.*, **27**, 89–105.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for "Compound random measures and their use in Bayesian nonparametrics"'.