Original Articles

A marginal sampler for σ-stable Poisson-Kingman mixture models

María Lomelí, Stefano Favaro & Yee Whye Teh

Page 0 | Received 14 Jul 2014, Accepted 01 Oct 2015, Accepted author version posted online: 21 Jan 2016

Download citation http://dx.doi.org/10.1080/10618600.2015.1110526 Crossmark

Abstract

We investigate the class of σ-stable Poisson-Kingman random probability

A marginal sampler for σ-stable Poisson-Kingman mixture models

María Lomelí
Gatsby Computational Neuroscience Unit,
University College London

Stefano Favaro
Department of Economics and Statistics,
University of Torino and Collegio Carlo Alberto

Yee Whye Teh
Department of Statistics,
University of Oxford

Abstract

We investigate the class of σ-stable Poisson-Kingman random probability measures (RPMs) in the context of Bayesian nonparametric mixture modeling. This is a large class of discrete RPMs which encompasses most of the popular discrete RPMs used in Bayesian nonparametrics, such as the Dirichlet process, Pitman-Yor process, the normalized inverse Gaussian process and the normalized generalized Gamma process. We show how certain sampling properties and marginal characterisations of σ-stable Poisson-Kingman RPMs can be usefully exploited for devising a Markov chain Monte Carlo (MCMC) algorithm for performing posterior inference with a Bayesian nonparametric mixture model. Specifically, we introduce a novel and efficient MCMC sampling scheme in an augmented space that has a fixed number of auxiliary variables per iteration. We apply our sampling scheme to a density estimation and clustering tasks with unidimensional and multidimensional datasets, and compare it against competing MCMC sampling schemes.

Keywords: Bayesian nonparametrics; Mixture models; MCMC posterior sampling; Normalized generalized Gamma process; Pitman-Yor process; σ-stable Poisson-Kingman random probability measures.

1 Introduction

Flexibly modeling the distribution of continuous data is an important concern in Bayesian nonparametrics and it requires the specification of a prior model for continuous distributions. A fruitful and general approach for defining such a prior model was first suggested by Lo (1984) in terms of an infinite dimensional mixture model, nowadays, it is the subject of a rich and active literature. Specifically, let P be a discrete random probability measure (RPM) with distribution $\mathcal{P}$. Given a collection of continuous and possibly multivariate observations $X_1, \ldots, X_n$, the infinite dimensional mixture model is defined hierarchically by means of a corresponding collection of latent random variables $Y_1, \ldots, Y_n$ from an exchangeable sequence directed by P, i.e.,

$$\begin{aligned} P &\sim \mathcal{P} \\ Y_i|P &\overset{\text{iid}}{\sim} P \\ X_i|Y_i &\overset{\text{ind}}{\sim} F(\cdot|Y_i) \end{aligned} \tag{1}$$

where $F(\cdot|Y_i)$ is a continuous distribution parameterized by Y_i. The distribution $F(\cdot|Y_i)$ is referred to as the kernel, whereas P is the mixing measure. The nonparametric hierarchical model (1) defines a mixture model with a potentially countably infinite number of components. By the discreteness of P, each pair (Y_i, Y_j) takes on the same value with positive probability, this value identifies a mixture component. In such a way, the Y_i allocates the X_i to a component from a random number of mixture components, thus providing a model for the unknown number of clusters within the data. The formulation of (1) presented in Lo (1984) sets P to be the Dirichlet process introduced by Ferguson (1973), hence the name of Dirichlet process mixture model'.

It is apparent that one can replace a Dirichlet process mixing measure with any other discrete RPM. Ishwaran & James (2001) first replaced the Dirichlet process with stick-breaking RPMs. As a notable example they focussed on the two parameter Poisson-Dirichet process, also known as Pitman-Yor process, which is a discrete RPM introduced in Perman *et al.* (1992) and further investigated in Pitman & Yor (1997) and James (2002). Nieto-Barajas *et al.* (2004) replaced the Dirichlet process with normalized random measures (NRMs) and Lijoi *et al.* (2007) focused on the normalized generalized Gamma process. See also James (2002) Lijoi *et al.* (2005), James *et al.*

(2009) and James (2013). Both the Pitman-Yor process and the normalized generalized Gamma process are valid alternatives to the Dirichlet process: they preserve almost the same mathematical tractability but they provide clustering properties that make use of all of the information contained in the sample. Indeed, the Dirichlet process allocates observations to the mixture model components with a probability depending solely on the number of times that the mixture component occurs. In contrast, under the Pitman-Yor process and the normalized generalized Gamma process, the allocation probability depends heavily on the number of distinct mixture components. This more flexible allocation mechanism turns out to be a distinguishing feature if we pick any of this priors for the mixture model (1). See Lijoi *et al*. (2005) and Lijoi *et al*. (2007) for details.

Several Markov chain Monte Carlo (MCMC) methods have been proposed for posterior sampling from the Dirichlet process mixture model. Marginal MCMC methods remove the infinite dimensionality of the problem by exploiting the tractable marginalization with respect to the Dirichlet process. See Escobar (1994), MacEachern (1994) and Escobar & West (1995) for early works, and Neal (2000) for an overview with some noteworthy developments such as the celebrated Algorithm 8. Conversely, conditional MCMC methods maintain the infinite dimensional part and find appropriate ways for sampling a sufficient but finite number of the atoms of the Dirichlet process. See Ishwaran & James (2001), Walker (2007), Papaspiliopoulos & Roberts (2008). Recently, marginal and conditional MCMC methods have been developed under more general classes of mixing measures, such as stick-breaking RPMs and NRMs, among others. See Ishwaran & James (2001), Griffin & Walker (2011), Barrios *et al*. (2013), Favaro & Teh (2013) and Favaro *et al*. (2014) for details.

In this paper we introduce a marginal MCMC method for posterior sampling from (1) with P belonging to the class of σ-stable Poisson-Kingman RPMs introduced in Pitman (2003). We refer to such a model as a σ-stable Poisson-Kingman mixture model. A conditional MCMC method for σ-stable Poisson-Kingman mixture model has been recently introduced in Favaro & Walker (2012). The class of σ-stable Poisson-Kingman RPMs forms a large class of discrete RPMs which encompasses most of the popular discrete RPMs used in Bayesian nonparametrics, e.g., the Pitman-Yor and the normalized generalized Gamma processes. It also includes the Dirichlet process as a special limiting case. Our main contribution is to provide a general purpose framework

for performing posterior inference with any member of this class of priors. In contrast to Favaro & Walker (2012), we exploit marginal characterisations of σ-stable Poisson-Kingman RPMs in order to remove the infinite dimensionality of the sampling problem. Efficient algorithms often rely upon simplifying properties of the priors just as inference algorithms for graphical models rely upon the conditional independencies encoded by the graph. Thus, in our experiments, we found this improved the algorithmic performance. We applied our algorithm for a density estimation and clustering tasks with unidimensional and multidimensional datasets, and compare it against competing MCMC sampling schemes.

The paper is structured as follows. In Section 2, we recall the definition of σ-stable Poisson-Kingman RPM, as well as some of its marginal properties which are fundamental for devising our marginal MCMC method. In Section 3, we present the marginal MCMC method for posterior sampling σ-stable Poisson-Kingman mixture models. Section 4 contains unidimensional and multidimensional experiments and Section 5 concludes with a brief discussion.

2 Preliminaries on σ-stable Poisson-Kingman RPMs

We start by recalling the definition of completely random measures (CRMs), the reader is referred to Kingman (1967) for a detailed account on CRMs. Let $\mathbb{X}$ be a complete and separable metric space endowed with the Borel σ-field $\mathcal{B}(\mathbb{X})$. A CRM μ is a random element taking values on the space of boundedly finite measures on $\mathbb{X}$ such that, for any $A_1, \ldots, A_n$ in $\mathcal{B}(\mathbb{X})$, with $A_i \cap A_j = \varnothing$ for $i \neq j$, the random variables $\mu(A_1), \ldots, \mu(A_n)$ are mutually independent. Kingman (1967) showed that μ is discrete almost surely so it can be represented in terms of nonnegative random masses $(u_k)_{k\geqslant 1}$ at $\mathbb{X}$-valued random locations $(\phi_k)_{k\geqslant 1}$, that is $\mu = \sum_{k\geqslant 1} u_k \delta_{\phi_k}$. The distribution of μ is characterized in terms of the distribution of the random point set $(u_k, \phi_k)_{k\geqslant 1}$ as a Poisson random measure on $\mathbb{R}^+ \times \mathbb{X}$ with mean measure ν, which is referred to as the Lévy intensity measure. In this paper we focus on homogeneous CRMs, namely, CRMs such that $\nu(ds, dy) = \rho(ds)H_0(dy)$ for some Lévy measure ρ on $\mathbb{R}^+$ and some non-atomic base distribution H_0 on $\mathbb{X}$. Homogeneity implies independence between $(u_k)_{k\geqslant 1}$ and $(\phi_k)_{k\geqslant 1}$, where $(\phi_k)_{k\geqslant 1}$ are independent and identically distributed as H_0 while the law of $(u_k)_{k\geqslant 1}$ is governed by ρ and we denote by $\mathrm{CRM}(\rho, H_0)$ the

distribution of a homogeneous CRM.

Homogeneous CRMs provide a fundamental tool for defining almost surely discrete random probability measures (RPMs) via the normalization approach. Specifically, let μ be a homogeneous CRM with Lévy measure ρ and base distribution H_0. Furthermore, let $T = \mu(\mathbb{X}) = \sum_{k\geqslant 1} u_k$ be the total mass of μ. Both positiveness and finiteness of the random variable T are ensured by the following conditions: $\int_{\mathbb{R}^+} \rho(ds) = +\infty$ and $\int_{\mathbb{R}^+} (1 - \mathrm{e}^{-s})\rho(ds) < +\infty$. Once these conditions are satisfied, one can define an almost surely discrete RPM as

$$P = \frac{\mu}{T} = \sum_{k\geqslant 1} p_k \delta_{\phi_k} \tag{2}$$

with $p_k = u_k/T$. Since μ is homogeneous, the law of the random probabilities $(p_k)_{k\geqslant 1}$ is governed by the Lévy measure ρ, and the atoms $(\phi_k)_{k\geqslant 1}$ are random variables independent of $(p_k)_{k\geqslant 1}$ and independent and identically distributed according to H_0. The RPM displayed in (2) is known from James *et al*. (2009) as the normalized random measure (NRM) with Lévy measure ρ and base distribution H_0. We refer to James (2002) and Regazzini *et al*. (2003) for a comprehensive account on homogeneous NRMs and denote by $\mathrm{NRM}(\rho, H_0)$ the distribution of P.

Since $P = \mu/T$ is almost surely discrete, there is a positive probability of $Y_i = Y_j$ for each pair of indexes $i \neq j$. This induces a random partition Π on $\mathbb{N}$, where i and j are in the same block in Π if and only if $Y_i = Y_j$. Kingman (1978) showed that Π is exchangeable and its distribution, the so-called exchangeable partition probability function (EPPF), can be deduced from the law of the NRM. See Pitman (2006) for a comprehensive account of EPPFs. A second object induced by $(Y_i)_{i\geqslant 1}$ is a size-biased permutation of the atoms in μ. Specifically, order the blocks in Π by increasing order of the least element in each block, and for each $k \in \mathbb{N}$ let Z_k be the least element of the kth block. Z_k is the index among $(Y_i)_{i\geqslant 1}$ of the first appearance of the kth unique value in the sequence. Let $V_k = \mu(\{Y_{Z_k}\})$ be the mass of the corresponding atom in μ. Then $(V_k)_{k\geqslant 1}$ is a size-biased permutation of the masses of atoms in μ, with larger masses tending to appear earlier in the sequence. It is easy to see that $\sum_{k\geqslant 1} V_k = T$, and that the sequence can be understood as a stick-breaking construction: starting with a stick of length $S_0 = T$; break off the first piece of length V_1; the stick's leftover length is $S_1 = S_0 - V_1$; then, the second piece with length V_2 is broken off, etc.

Theorem 2.1 of Perman *et al.* (1992) states that the sequence of surplus masses $(S_k)_{k\geqslant 0}$ forms a Markov chain and gives the corresponding initial distribution and transition kernels. See the supplementary material for a proof of this theorem. Let us denote by $f_\rho(t)$ the density function of T. The EPPF of the random partition Π can be derived from this theorem by enriching the generative process for the sequence $(Y_i)_{i\geqslant 1}$ as follows, where we simulate parts of the CRM as and when required.

i) Start with drawing the total mass from its distribution $\mathbb{P}_{\rho,H_0}(T \in dt) = f_\rho(t)dt$.

ii) The first draw Y_1 from μ/T is a size-biased pick from the masses of μ. The actual value of Y_1 is simply $Y_1^* \sim H_0$, while the mass of the corresponding atom in μ is V_1, with conditional distribution given by

$$\mathbb{P}_{\rho,H_0}(V_1 \in dv_1 | T \in dt) = \frac{v_1}{t}\rho(dv_1)\frac{f_\rho(t-v_1)}{f_\rho(t)}.$$

The leftover mass is $S_1 = T - V_1$.

iii) For subsequent draws $i \geqslant 2$:

- Let K be the current number of distinct values among $Y_1, \ldots, Y_{i-1}$, and $Y_1^*, \ldots, Y_K^*$ the unique values, i.e., atoms in μ. The masses of these first K atoms are denoted $V_1, \ldots, V_K$ and the leftover mass is $S_K = T - \sum_{k=1}^K V_k$.
- For each $k \leqslant K$, with probability V_k/T, we set $Y_i = Y_k^*$.
- With probability S_K/T, Y_i takes on the value of an atom in μ besides the first K atoms. The actual value Y_{K+1}^* is drawn from H_0, while its mass is drawn from

$$\mathbb{P}_{\rho,H_0}(V_{K+1} \in dv_{K+1} | S_K \in ds_K) = \frac{v_{K+1}}{s_K}\rho(dv_{K+1})\frac{f_\rho(s_K - v_{K+1})}{f_\rho(s_K)}.$$

The leftover mass is again $S_{K+1} = S_K - V_{K+1}$.

By multiplying the above infinitesimal probabilities one obtains the joint distribution of the random elements T, Π, $(V_i)_{i\geqslant 1}$ and $(Y_i^*)_{i\geqslant 1}$. Such a joint distribution was first obtained in Pitman (2003) and it is recalled in the next proposition, see also Pitman (2006) for details.

Proposition 1. *Let Π_n be the exchangeable random partition of $[n] := \{1,\dots,n\}$ induced by a sample $(Y_i)_{i\in[n]}$ from $P \sim \mathrm{NRM}(\rho, H_0)$. Let $(Y_j^*)_{j\in[K]}$ be the K distinct values in $(Y_i)_{i\in[n]}$ with masses $(V_j)_{j\in[K]}$. Then*

$$\begin{aligned}\mathbb{P}_{\rho,H_0}(T \in dt, \Pi_n = (c_k)_{k\in[K]}, Y_k^* \in dy_k^*, V_k \in dv_k \text{ for } k \in [K]) \\ = t^{-n} f_\rho(t - \textstyle\sum_{k=1}^K v_k) dt \displaystyle\prod_{k=1}^K v_k^{|c_k|} \rho(dv_k) H_0(dy_k^*),\end{aligned} \tag{3}$$

where $(c_k)_{k\in[K]}$ denotes a particular partition of $[n]$ with K blocks, $c_1,\dots,c_K$, ordered by increasing least element and $|c_k|$ is the cardinality of block c_k. The distribution (3) is invariant to the size-biased order.

2.1 σ-stable Poisson-Kingman RPMs

Poisson-Kingman RPMs have been introduced in Pitman (2003) as a generalization of homogeneous NRMs. Let $\mu \sim \mathrm{CRM}(\rho, H_0)$ and let $T = \mu(\mathbb{X})$ be finite, positive almost surely, and absolutely continuous with respect to Lebesgue measure. For any $t \in \mathbb{R}^+$, let us consider the conditional distribution of μ/t given that the total mass $T \in dt$. This distribution is denoted by $\mathrm{PK}(\rho, \delta_t, H_0)$. It is the distribution of a RPM and δ_t denotes the Dirac delta function. Poisson-Kingman RPMs form a class of RPMs whose distributions are obtained by mixing $\mathrm{PK}(\rho, \delta_t, H_0)$, over t, with respect to some distribution γ on the positive real line. Specifically, a Poisson-Kingman RPM has the hierarchical representation

$$\begin{aligned} T &\sim \gamma \\ P|T = t &\sim \mathrm{PK}(\rho, \delta_t, H_0). \end{aligned} \tag{4}$$

The RPM P is referred to as the Poisson-Kingman RPM with Lévy measure ρ, base distribution H_0 and mixing distribution γ. Throughout the paper we denote by $\mathrm{PK}(\rho, \gamma, H_0)$ the distribution of P. If $\gamma(dt) = f_\rho(t)dt$ then the distribution $\mathrm{PK}(\rho, f_\rho, H_0)$ coincides with $\mathrm{NRM}(\rho, H_0)$. Since μ is homogeneous, the atoms $(\phi_k)_{k\geqslant 1}$ of P are independent of their masses $(p_k)_{k\geqslant 1}$. They form a sequence of independent random variables identically distributed according to H_0. Finally, the masses of P have distribution governed by the Lévy measure ρ and the distribution γ.

In this paper we focus on the class of σ-stable Poisson-Kingman RPMs. This is a noteworthy subclass of Poisson-Kingman RPMs which encompasses most of the popular discrete RPMs used in Bayesian nonparametrics, e.g., the Pitman-Yor process and the normalized generalized Gamma process. For any $\sigma \in (0,1)$, $f_\sigma(t) = \frac{1}{\pi}\sum_{j=0}^{\infty} \frac{(-1)^{j+1}}{j!}\sin(\pi\sigma j)\frac{\Gamma(\sigma j+1)}{t^{\sigma j+1}}$ is the density function of a positive σ-stable random variable, let us denote it by f_σ. A σ-stable Poisson-Kingman RPMs is a Poisson-Kingman RPM with Lévy measure

$$\rho(dx) = \rho_\sigma(dx) := \frac{\sigma}{\Gamma(1-\sigma)} x^{-\sigma-1} dx, \tag{5}$$

base distribution H_0 and mixing distribution $\gamma(dt) = h(t) f_\sigma(dt) / \int_0^{+\infty} h(t) f_\sigma(t) dt$, for any non-negative measurable function h. Accordingly, σ-stable Poisson-Kingman RPMs form a class of discrete RPMs indexed by the parameters (σ, h). The Dirichlet process can be recovered as a limiting σ-stable Poisson-Kingman RPM if $\sigma \to 0$, for some choices of h. Throughout the paper we denote by $\mathrm{PK}(\rho_\sigma, h, H_0)$ the distribution of a σ-stable Poisson-Kingman RPM with parameters (σ, h).

Examples of σ-stable Poisson-Kingman RPMs are obtained by specifying the tilting function h. The *normalized σ-stable process* (NS) in Kingman (1975) corresponds to $h(t) = 1$. The *normalized generalized gamma process* (NGG) in James (2002) and Pitman (2003) corresponds to $h(t) = \exp\{\tau - \tau^{1/\sigma} t\}$, for any $\tau > 0$. See also Lijoi *et al.* (2005), Lijoi *et al.* (2007), Lijoi *et al.* (2008), James *et al.* (2009) and James (2013). The *Pitman-Yor process* (PY) in Perman *et al.* (1992) corresponds to $h(t) = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\sigma+1)} t^{-\theta}$ with $\theta \geqslant -\sigma$, see Pitman & Yor (1997). The *Gamma-tilted* process (GT) corresponds to $h(t) = t^{-\theta}\exp\{-\eta t\}$, for any $\eta > 0$ or $\eta = 0$ and $\theta > -\sigma$. The *Poisson-Gamma class* ($\mathcal{PG}$) in James (2013) corresponds to $h(t) = \int_{\mathbb{R}^+} \exp\{\tau - \tau^{1/\sigma} t\} F(d\tau)$, for any distribution function F over the positive real line. See also Pitman & Yor (1997) and James (2002). Let $\mathcal{T}$ a positive random variable, the *composition of classes class* (CC) in Ho *et al.* (2008) corresponds to $h(t) = \frac{\mathbb{E}[f(t\mathcal{T}^{1/\sigma})]}{\int_0^{+\infty} \mathbb{E}[f(t\mathcal{T}^{1/\sigma})] f_\sigma(t) \mathrm{d}t}$, where f is any positive function, see James (2002) for details. Let S_σ a positive σ-stable random variable, the *Lamperti class* (LT) in Ho *et al.* (2008) corresponds to the choice

$$h(t) = L_\sigma \mathbb{E}\left(g(S_\sigma t^{-1})\right), \tag{6}$$

where $\frac{1}{L_\sigma} = \frac{\sin(\pi\sigma)}{\pi}\int_0^{+\infty} \frac{f(y) y^{\sigma-1}}{y^{2\sigma} + 2y^\sigma \cos(\pi\sigma) + 1} \mathrm{d}y$ and g is any positive function such that (6) is well-defined,

see James (2002) for details. The Mittag-Leffler class (ML) in Ho *et al.* (2008) corresponds to the choice of $g(x) = \exp\{-x^\sigma\}$ in the tilting function (6). Figure 1 shows the relationships between these examples of σ-stable Poisson-Kingman RPMs.

The distribution of the exchangeable random partition induced by a sample from a σ-stable Poisson-Kingman RPMs is obtained by a direct application of Proposition 1. See the supplementary material for examples of EPPFs obtained from this proposition. The next proposition specializes Proposition 1 to σ-stable Poisson-Kingman RPMs.

Proposition 2. *Let Π_n be the exchangeable random partition of $[n]$ induced by a sample $(Y_i)_{i\in[n]}$ from $P \sim \mathrm{PK}(\rho_\sigma, h, H_0)$. Then,*

$$\mathbb{P}_{\rho_\sigma,h,H_0}(\Pi_n = (c_k)_{k\in[K]}) = V_{n,K}\prod_{k=1}^{K} W_{|c_k|} \tag{7}$$

where we set $V_{n,K} = \int_{\mathbb{R}^+}\int_0^t t^{-n}(t-s)^{n-1-k\sigma}h(t)f_\sigma(s)dtds$ and $W_m = \Gamma(m-\sigma)/\Gamma(1-\sigma) = [1-\sigma]_{m-1} := \prod_{i=0}^{m-2}(1-\sigma+i)$.

Proposition 2 provides one of the main tools for deriving the marginal MCMC sampler in Section 3. We refer to Gnedin & Pitman (2006) and Pitman (2006) for a comprehensive study of exchangeable random partitions with distribution of the form (7). These random exchangeable partitions are typically referred to as Gibbs-type with parameter $\sigma \in (0,1)$.

3 Marginal samplers for σ-stable Poisson-Kingman mixture models

In this section we develop a marginal sampler that can be effectively applied to all members of the σ-stable Poisson-Kingman process family. Our sampler does not require any numerical integrations, nor evaluations of special functions, e.g. the density f_σ of the positive σ-stable distribution as in Wuertz *et al.* (2013). It applies to non-conjugate hierarchical mixture models based on σ-stable Poisson-Kingman RPMs by extending the Reuse data augmentation scheme of Favaro & Teh (2013).

3.1 Effective representation with data augmentation

If $\gamma(dt) \propto h(t) f_\sigma(t)$, the joint distribution over the induced partition Π_n, total mass T and surplus mass S is given by:

$$\begin{aligned}&\mathbb{P}_{\rho_\sigma,\gamma,H_0}(T \in dt, S \in ds, \Pi_n = (c_k)_{k\in[K]})\\ =&t^{-n}(t-s)^{n-1-K\sigma} f_\sigma(s) h(t) dt \frac{\sigma^K}{\Gamma(n-K\sigma)} \prod_{k=1}^{K} \frac{\Gamma(|c_k|-\sigma)}{\Gamma(1-\sigma)} \mathbb{I}_{(0,t)}(s) \mathbb{I}_{(0,\infty)}(t).\end{aligned}$$

With the exception of two difficulties, this joint distribution easily allows us to construct marginal samplers. The first difficulty is that it is necessary to compute f_σ if working with an MCMC scheme using the above representation. Current software packages compute this density using numerical integration techniques, which can be unnecessarily expensive. The following is an integral representation (Kanter, 1975, Zolotarev, 1966) with a view to introducing an auxiliary variable into our system thus, removing the need to evaluate the integral numerically. Let $\sigma \in (0,1)$. Then,

$$f_\sigma(t) = \frac{1}{\pi} \frac{\sigma}{1-\sigma} \left(\frac{1}{t}\right)^{\frac{1}{1-\sigma}} \int_0^\pi A(z) \exp\left(-\left(\frac{1}{t}\right)^{\frac{\sigma}{1-\sigma}} A(z)\right) dz \tag{8}$$

where Zolotarev's function is

$$A(z) = \left[\frac{\sin(\sigma z)}{\sin(z)}\right]^{\frac{1}{1-\sigma}} \left[\frac{\sin((1-\sigma)z)}{\sin(\sigma z)}\right], \qquad z \in (0,\pi).$$

Zolotarev's representation has been used by Devroye (2009) to construct a random number generator for polynomially and exponentially tilted σ-stable random variables and in a rejection sampling scheme by Favaro & Walker (2012). Our proposal here is to introduce an auxiliary variable Z using a data augmentation scheme (Tanner & Wong (1987)), with conditional distribution given $T \in dt$ described by the integrand in (8).

The second difficulty is that the variables T and S are dependent and that computations with small values of T and S might not be numerically stable. To address these problems, we propose the following reparameterization: $W = \frac{\sigma}{1-\sigma} \log T$, and $R = S/T$ where $W \in \mathbb{R}$ and $R \in (0,1)$. This

gives our final representation:

$$\begin{aligned}
&\mathbb{P}_{\rho_\sigma,\gamma,H_0}(W \in dw, R \in dr, Z \in dz, \Pi_n = (c_k)_{k\in[K]}) \\
= \quad &\frac{1}{\pi} e^{-w(1+(1-\sigma)K)}(1-r)^{n-1-K\sigma} r^{-\frac{1}{1-\sigma}} h(e^{\frac{1-\sigma}{\sigma}w}) A(z) e^{-e^{-w} r^{-\frac{\sigma}{1-\sigma}} A(z)} \frac{\sigma^K}{\Gamma(n-\sigma K)} \prod_{k=1}^{K} [1-\sigma]_{|c_k|-1}.
\end{aligned}$$

3.2 σ-stable Poisson-Kingman mixture models

To make the derivation of our sampler explicit, we will consider a σ-stable Poisson-Kingman RPM as the random mixing distribution in a Bayesian nonparametric mixture model:

$$\begin{aligned}
T &\sim \gamma \\
P|T &\sim \mathrm{PK}(\rho_\sigma, \delta_T, H_0) \\
Y_i \mid P &\overset{\text{iid}}{\sim} P \\
X_i \mid Y_i &\overset{\text{ind}}{\sim} F(\cdot \mid Y_i).
\end{aligned}$$

$F(\cdot \mid Y)$ is the observation's distribution and our dataset consists of n observations $(x_i)_{i\in[n]}$ of the corresponding variables $(X_i)_{i\in[n]}$. We will assume that $F(\cdot \mid Y)$ is smooth.

In the following we will derive two marginal samplers for our nonparametric mixture models. As opposed to conditional samplers, which maintain explicit representations of the random probability measure P, marginal samplers marginalize out P, retaining only the induced partition Π_n of the dataset. In our case, we will include as well the auxiliary variables W, R, Z in the final representation. Denoting the unique values (component parameters) by $(Y_k^*)_{k\in[K]}$, the joint distribution over all variables is given by:

$$\begin{aligned}
&\mathbb{P}(W \in dw, R \in dr, Z \in dz, \Pi_n = (c_k)_{k\in[K]}, Y_k^* \in dy_k^* \text{ for } k \in [K], X_i \in dx_i \text{ for } i \in [n]) \\
=&\frac{1}{\pi} e^{-w(1+(1-\sigma)K)}(1-r)^{n-1-K\sigma} r^{-\frac{1}{1-\sigma}} h(e^{\frac{1-\sigma}{\sigma}w}) A(z) e^{-r^{-\frac{\sigma}{1-\sigma}} e^{-w} A(z)} dw\, dr\, dz \\
&\times \frac{\sigma^K}{\Gamma(n-\sigma K)} \prod_{k=1}^{K} [1-\sigma]_{n_k-1} H_0(dy_k^*) \prod_{i\in c_k} F(dx_i|y_k^*). \qquad (9)
\end{aligned}$$

The system of predictive distributions governing the distribution over partitions given the other variables can be read from the joint distribution (9). Specifically, the conditional distribution of a

new variable Y_{n+1} is:

$$\mathbb{P}(Y_{n+1} \in dy_{n+1} \,|\, \Pi_n = (c_k)_{k\in[K]}, Y_k^* \in dy_k^* \text{ for } k \in [K], W \in dw, R \in dr, Z \in dz)$$
$$\propto \sigma e^{(\sigma-1)w}(1-r)^{-\sigma}\frac{\Gamma(n+1-\sigma K)}{\Gamma(n+1-\sigma(K+1))}H_0(dy_{n+1}) + \sum_{k=1}^{K}(|c_k|-\sigma)\delta_{y_k^*}(dy_{n+1}).$$

The conditional probability of the next observation joining an existing cluster c_k is proportional to $|c_k|-\sigma$, which is the same for all exchangeable random probability measures based on the σ-stable CRM. The conditional probability of joining a new cluster is more complex and is dependent upon the auxiliary variables. Such system of predictive distributions were first studied by Blackwell & McQueen (1973) for the chinese restaurant process, see also Aldous (1985) for details and Ewens (1972) for an early account in population genetics.

3.3 Sampler updates

In this section we will first describe Gibbs updates to the partition Π_n, conditioned on the auxiliary variables W, R, Z before describing updates to the auxiliary variables. We describe a non-conjugate case where the component parameters $(Y_k^*)_{k\in[K]}$ cannot be marginalized out, and we derive an extension of Favaro & Teh (2013). In the case where the base distribution H_0 is conjugate to the observation's distribution $F(\cdot)$, the component parameters can be marginalized out as well, which leads to an extension to Algorithm 3 of Neal (2000).

3.3.1 Non-Conjugate Marginalized Sampler

In general, the base distribution H_0 might not be conjugate to the observation's distribution F, and the cluster parameters cannot be marginalized out tractably. In this case the state space of our Markov chain consists of $(c_k)_{k\in[K]}$, W, R, Z, as well as the cluster parameters $(Y_k^*)_{k\in[K]}$. The Gibbs updates to the partition involve updating the cluster assignment of one observation, say the ith one, at a time. We can adapt the Reuse algorithm of Favaro & Teh (2013) to update our partitions.

In this algorithm a fixed number $M > 0$ of potential new clusters is maintained along with those in the current partition $(c_k)_{k\in[K]}$. The parameters for each of these potential new clusters are denoted by $(Y_\ell^{\text{new}})_{\ell\in[M]}$. When updating the cluster assignment of the ith observation, we consider

the potential new clusters as well as those in $(c_k^{\neg i})_{k\in[K^{\neg i}]}$. If one of the potential new clusters is chosen, it is moved into the partition, and in its place a new cluster is generated by drawing a new parameter from H_0. Conversely, when a cluster in the partition is emptied, it is moved into the list of potential new clusters, displacing a randomly chosen one. The parameters of the potential new clusters are refreshed by iid draws from H_0 after a full iteration through the dataset, see Algorithm 1 for details. The conditional probability of the cluster assignment of the ith observation is:

$$\begin{aligned}\mathbb{P}(i \text{ joins cluster } c_k^{\neg i} \mid \text{rest}) &\propto (|c_k^{\neg i}| - \sigma) F(X_i \in dx_i | y_k^*) \\ \mathbb{P}(i \text{ joins new cluster } \ell \mid \text{rest}) &\propto \frac{1}{M}\sigma e^{(\sigma-1)w}(1-r)^{-\sigma}\frac{\Gamma(n-\sigma K^{\neg i})}{\Gamma(n-\sigma(K^{\neg i}+1))} F(X_i \in dx_i | y_\ell^{\text{new}}).\end{aligned}$$

If H_0 is conjugate, we can replace the likelihood term in the cluster assignment rule by the conditional density of x under cluster c, denoted by $F(X \in \mathrm{d}x|\mathbf{x}_c)$, given the observations $\mathbf{X}_c = (X_j)_{j\in c}$ currently assigned to that cluster:

$$F(X \in \mathrm{d}x|\mathbf{x}_c) = \frac{\int H_0(dy) F(X \in \mathrm{d}x|y) \prod_{j\in c} F(X \in \mathrm{d}x_j|y)}{\int H_0(dy) \prod_{j\in c} F(X \in \mathrm{d}x_j|y)}.$$

3.3.2 Updates to Auxiliary Variables

The updates to the auxiliary variables W, R and Z are straightforward. Their conditional densities can be read off the joint density (9):

$$\begin{aligned}\mathbb{P}(W \in dw \mid \text{rest}) &\propto e^{-w(1+(1-\sigma)K)} h(e^{\frac{1-\sigma}{\sigma}w}) e^{-r^{-\frac{\sigma}{1-\sigma}}e^{-w}A(z)} dw, \quad w \in \mathbb{R} \\ \mathbb{P}(R \in dr \mid \text{rest}) &\propto (1-r)^{n-1-K\sigma} r^{-\frac{1}{1-\sigma}} e^{-r^{-\frac{\sigma}{1-\sigma}}e^{-w}A(z)} dr, \quad r \in (0,1) \\ \mathbb{P}(Z \in dz \mid \text{rest}) &\propto A(z) e^{-r^{-\frac{\sigma}{1-\sigma}}e^{-w}A(z)} dz, \quad z \in (0,\pi).\end{aligned}$$

Although these are not in standard form, their states can be updated easily using generic MCMC methods. We used slice sampling by Neal (2003) in our implementation, see the supplementary material for details. If we have a prior on the index parameter σ, it can be updated as well. Due to the heavy tailed nature of the conditional distribution, we recommend transforming σ to $\log \frac{\sigma}{1-\sigma}$.

Algorithm 1 ReUse$(\Pi_n, M, \{X_i\}_{i\in[n]}, \{Y^*_c\}_{c\in\Pi_n}, H_0 \mid \text{rest})$

Draw $\{Y^e_j\}_{j=1}^M \overset{\text{i.i.d.}}{\sim} H_0$
for $i = 1 \to n$ **do**
  Let $c \in \Pi_n$ be such that $i \in c$
  $c \leftarrow c \backslash \{i\}$
  if $c = \varnothing$ **then**
    $k \sim \text{UniformDiscrete}(\frac{1}{M})$
    $Y^e_k \leftarrow Y^*_c$
    $\Pi_n \leftarrow \Pi_n \backslash \{c\}$
  end if
  Set c' according to $\Pr[i \text{ joins cluster } c' \mid \{X_i\}_{i\in c}, Y^*_c, \text{rest}]$
  if $c' \in [M]$ **then**
    $\Pi_n \leftarrow \Pi_n \cup \{\{i\}\}$
    $Y^*_{\{i\}} \leftarrow Y^e_{c'}$
    $Y^e_{c'} \sim H_0$
  else
    $c' \leftarrow c' \cup \{i\}$
  end if
end for

3.4 Differences with Favaro & Walker (2012) conditional sampler

If we start with Proposition 1, we can do the following 2 changes of variables: $P_j = \frac{V_j}{T - \sum_{\ell<j} V_\ell}$ and $U_j = \frac{P_j}{1-\sum_{\ell<j} P_\ell}$. Then, we obtain the corresponding joint in terms of N $(0,1)$-valued stick-breaking weights $\{U_j\}_{j=1}^N$ which corresponds to equation (19) of Favaro & Walker (2012). The truncation level N needs to be randomised to have an exact MCMC scheme. The authors propose to do so with Kalli *et al.* (2011)'s efficient slice sampler. The number of auxiliary variables, $N + 2$, is a random quantity as opposed to keeping it fixed when using our marginal sampler. This could potentially lead to slower running times and larger memory requirements to store these quantities when the number of data points is large. Furthermore, in our implementation of this conditional sampler, we found that some of the auxiliary variables are highly correlated which leads to slow mixing of the chain. A quantitative comparison is presented in Table 2 in terms of running times in seconds and effective sample size (ESS).

Algorithm 2 MarginalSamplerNonConj(h_T, σ, M, H_0)

for $t = 2 \to iter$ **do**
  Update $z^{(t)}$: Slice sample $\tilde{\mathbb{P}}\left(Z \in \mathrm{d}z \mid \text{rest}\right)$
  Update $p^{(t)}$: Slice sample $\tilde{\mathbb{P}}\left(P \in \mathrm{d}p \mid \text{rest}\right)$
  Update $w^{(t)}$: Slice sample $\tilde{\mathbb{P}}\left(W \in \mathrm{d}w \mid \text{rest}\right)$
  Update $\pi^{(t)}, \{x_c^*\}_{c\in\pi}^{(t)}$: **ReUse**$(\Pi_n, M, \{Y_i\}_{i\in[n]}, \{X_c^*\}_{c\in\Pi_n}, H_0 \mid \text{rest})$
end for

4 Numerical illustrations

In this section, we illustrate the algorithm on unidimensional and multidimensional data sets. We applied our MCMC sampler for density estimation of a σ-stable PK $(\sigma, H_0, h_T(t))$ mixture model, for various choices of $h(t)$. In the experiments where we used a conjugate prior for the mixture component's parameter we sampled the parameters rather than integrating them out. We kept the hyperparameters of each h-tilting function fixed.

4.1 Unidimensional experiment

The dataset from Roeder (1990) consists of measurements of velocities in km/sec of 82 galaxies from a survey of the Corona Borealis region. We chose the base distribution H_0 and the corresponding likelihood F for the kth cluster:

$$\begin{aligned} H_0(\mathrm{d}\mu_k, \mathrm{d}\tau_k) &= F_\mu\left(\mathrm{d}\mu_k \mid \mu_0, \tau_0\tau_k\right) F_\tau(\mathrm{d}\tau_k \mid \alpha_0, \beta_0) \\ F(\mathrm{d}x_1, \ldots, \mathrm{d}x_{n_k} \mid \mu_k, \tau_k) &= \prod_{i=1}^{n_k} \mathcal{N}\left(x_i \mid \mu_k, \tau_k^{-1}\right) \end{aligned}$$

where $X_1, \ldots, X_{n_k}$ are the observations currently assigned to cluster k. $\mathcal{N}$ denotes a Normal distribution with given mean μ_k and variance τ_k^{-1}. In the first sampler (*Marg-Conj I*), we used $H_0(\mathrm{d}\mu_k, \mathrm{d}\tau_k) = \mathcal{N}\left(\mathrm{d}\mu_k \mid \mu_0, \tau_0^{-1}\right) \delta\{\tau_k = \tau\}$ with a common precision parameter τ among all clusters and set it to $\frac{1}{4}$ of the range of the data. In the second sampler (*Marg-Conj II*), we used $H_0(\mathrm{d}\mu_k, \mathrm{d}\tau_k) = \mathcal{N}\left(\mathrm{d}\mu_k \mid \mu_0, \tau_0^{-1}\tau_k^{-1}\right) \mathrm{Gamma}(\tau_k \mid \alpha_0, \beta_0)$. In the third sampler (*Marg-NonConj*), we used a non conjugate distribution for the mean per cluster, $\mu = \log\varphi$ where $\varphi \sim \mathrm{Gamma}(\varphi \mid a_0, b_0)$, and for the precision $\tau_k \sim \mathrm{Gamma}(\tau_k \mid \alpha_0, \beta_0)$.

In Table 1, we reported a modest increase in the running times if we use a non-conjugate prior (*Marg-NonConj*) versus a conjugate prior (*Marg-Conj II*) . In Table 2, the algorithm's sensitivity to the number of new components M was tested and compared against the conditional sampler of Favaro & Walker (2012). As we increase the marginal sampler's number of new components per iteration, the ESS increases. Intuitively, the computation time increases but also leads to a potentially better mixing of the algorithm. In contrast, we found that the conditional sampler was not performing very well due to high correlations between the auxiliary variables. Finally, in Table 3 we present that different values for σ can be effectively chosen without modifying the algorithm as opposed to Favaro & Walker (2012), which is only available for $\sigma = 0.5$.

After assessing the algorithmic performance, we used it for posterior inference with a non-parametric mixture model where the top level is a prior from the σ-Stable Poisson-Kingman class. Since any prior in this class can be chosen, one possible criterion for model selection is predictive performance. In Table 4, we reported the average leave-one-out predictive probabilities. See the supplementary material for details on how to compute these quantities. We can see that all priors in this class have similar average predictive probabilities but the NGG slightly outperforms the rest.

In Figure 2, the mode of the posterior distribution is reported, it is around 10 clusters, and there are 6 clusters in the coclustering probability matrix. Indeed, a good estimate of the density might include superfluous components having vanishingly small weights as explained in Miller & Harrison (2013). The third plot shows the corresponding density estimate which is consistent under certain conditions as shown in De Blasi *et al.* (2015).

4.2 Multidimensional experiment

The dataset from de la Mata-Espinosa *et al.* (2011) consists of n D-dimensional triacylglyceride profiles of different types of oils where $n = 120$ and $D = 4000$. The observations consist of profiles of olive, monovarietal vegetable and blends of oils. Within each type there could be several subtypes so we cannot know the number of varieties a priori. We preprocessed the data by applying Principal Component Analysis (PCA) (Jolliffe, 2002) to get the relevant dimensions in it; a useful technique when the signal to noise ratio is small. We used the first $d = 8$ principal components which explained 97% of the variance and encoded sufficient information for the mixture model to

recover distinct clusters.

Then a σ-stable Poisson-Kingman mixture of multivariate Normals with unknown covariance matrix and mean vector was chosen for different h-tilting functions. A multivariate Normal-Inverse Wishart was chosen as a base measure and the corresponding likelihood F for the kth cluster is:

$$H_0(\mathrm{d}\mu_k, \mathrm{d}\Sigma_k) = \mathcal{N}_d\left(\mathrm{d}\mu_k \mid \mu_0, r_0\Sigma_k^{-1}\right) \mathcal{IW}_d\left(\mathrm{d}\Sigma_k \mid \nu_0, S_0\right)$$
$$F(\mathrm{d}x_1, \ldots, \mathrm{d}x_{n_k} \mid \mu_k, \Sigma_k) = \prod_{i=1}^{n_k} \mathcal{N}_d\left(\mathrm{d}x_i \mid \mu_k, \Sigma_k\right)$$

where $X_1, \ldots, X_{n_k}$ are the observations currently assigned to cluster k. $\mathcal{N}_d$ denotes a d-variate Normal distribution with given mean vector μ_k and covariance matrix Σ_k^{-1}, $\mathcal{IW}_d$ denotes an inverse Wishart over $d \times d$ positive definite matrices with given degrees of freedom ν_0 and scale matrix S_0. The Inverse Wishart is parameterised as in Gelman *et al.* (1995). S_0 was chosen to be a diagonal matrix with each diagonal element given by the maximum range of the data across all dimensions and degrees of freedom $\nu_0 = d + 3$, a weakly informative case.

In Table 5, the average 5-fold predictive probabilities are reported, see supplementary material for details. Again, we observe that all priors in this class have similar average predictive probabilities. In Figure 3, the mean curve per cluster and the coclustering probability matrix are reported. This mean curve reflects the average triacylglyceride profile per oil type. The coclustering probability matrix was used as an input to an agglomerative clustering algorithm to obtain a hierarchical clustering representation as in Medvedovic & Sivaganesan (2002). In certain contexts, it is useful to think of a hierarchical clustering rather than a flat one since it may be natural to think of superclasses.

In Figure 3, the mean curves per cluster are shown. These were found by thresholding the hierarchy and ignoring clusters of size one. The first plot corresponds to the olive oil cluster, it is well represented by the mean curve. The last two plots correspond to data that belongs to non-olive blends of oil. The second and third plots correspond to non-olive monovarietal oil clusters. We could interpret these two clusters as different varieties of vegetable oil since their corresponding mean curves are indeed different. In the dendrogram it is clear that most of the data belongs to 3 large clusters and that 60% of the triacylglycerides are olive oil.

5 Discussion

A completely random measure completely specifies its total mass, but if we allow the total mass to come from a different distribution, we obtain the larger class of Poisson-Kingman RPMs introduced by Pitman (2003). Successively, if pick the σ-stable Lévy measure we obtain the σ-stable Poisson Kingman class. This class of random probability measures is natural, but certain intractabilities have hindered its use. For instance, the intractability associated to the σ-stable density, as noted by Lijoi *et al*. (2008). The aim of this paper was to review this class of priors and some characterisations which allowed us to build a novel algorithm for posterior simulation that is efficient and easy to use.

Our algorithm is the first sampler of marginal type for mixture models with σ-stable Poisson-Kingman priors. Previously, a conditional sampler has been proposed by Favaro & Walker (2012). One of the advantages of our approach is that the number of auxiliary variables per iteration is smaller than the conditional sampler's, hence, it has smaller memory and storage requirements. It also has better ESS and running times as shown in our experiments. Both conditional and marginal samplers for this class are general purpose: they do not depend on a particular characterisation of a specific Bayesian nonparametric prior as opposed to previous approaches. This makes them very useful and should be added to our Bayesian nonparametrics toolbox. Our approach could be used as a building-block in a more complex model using the proposed algorithm. This is an interesting avenue for future research.

6 Acknowledgements

We thank Lancelot James, Charles Blundell and the anonymous referees for useful comments. María Lomelí thanks the Gatsby Charitable Foundation for generous funding. Stefano Favaro is supported by the European Research Council through StG N-BNP 306406. Yee Whye Teh is supported by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013) ERC grant agreement no. 617071.

References

Aldous, D. 1985. Exchangeability and Related Topics. *Pages 1–198 of: École d'Été de Probabilités de Saint-Flour XIII–1983*. Springer, Berlin.

Barrios, E., Lijoi, A., Nieto-Barajas, L. E., & Prüenster, I. 2013. Modeling with Normalized Random Measure mixture models. *Statistical Science*, **28**, 283–464.

Blackwell, D., & McQueen, J. B. 1973. Ferguson distributions via Pólya urn schemes. *Annals of Statistics*, **1**, 353–355.

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prüenster, I., & Ruggiero, M. 2015. Are Gibbs-type priors the most natural generalization of the Dirichlet process? *Pages 212–229 of: IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37.

de la Mata-Espinosa, P., Bosque-Sendra, J. M., Bro, R., & Cuadros-Rodriguez, L. 2011. Discriminating olive and Non-olive oils using HPLC-CAD and chemometrics. *Analytical and Bioanalytical Chemistry*, **399**, 2083–2092.

Devroye, L. 2009. Random variate generation for exponentially and polynomially tilted Stable distributions. *ACM Transactions on Modelling and Computer Simulation*, **19**, 1–20.

Escobar, M. D. 1994. Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, **89**, 268–277.

Escobar, M. D., & West, M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577–588.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 3 87–112.

Favaro, S., & Teh, Y. W. 2013. MCMC for Normalized Random Measure mixture models. *Statistical Science*, **28**, 335–359.

Favaro, S., & Walker, S. G. 2012. Slice sampling σ-Stable Poisson-Kingman mixture models. *Journal of Computational and Graphical Statistics*, **22**, 830–847.

Favaro, S., Lomeli, M., & Teh, Y. W. 2014. On a class of σ-stable Poisson-Kingman models and an effective marginalized sampler. *Statistics and Computing*, **25**, 67–78.

Ferguson, T. S. 1973. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. 1995. *Bayesian data analysis*. Chapman & Hall, London.

Gnedin, A., & Pitman, J. 2006. Exchangeable Gibbs partitions and Stirling triangles. *Journal of Mathematical Sciences*, **138**, 5674–5684.

Griffin, J. E., & Walker, S. G. 2011. Posterior simulation of Normalized Random Measure mixtures. *Journal of Computational and Graphical Statistics*, **20**, 241–259.

Ho, M. W., James, L. F., & Lau, J. W. 2008. *Gibbs partitions (EPPFs) derived from a Stable subordinator are Fox H and Meijer G transforms*. ArXiv:0708.0619.

Ishwaran, H., & James, L. F. 2001. Gibbs sampling methods for Stick-Breaking priors. *Journal of the American Statistical Association*, **96**, 161–173.

James, L. F. 2002. *Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics*. ArXiv:math/0205093.

James, L. F. 2013. *Stick-Breaking PG(α, ψ)-Generalized Gamma processes*. ArXiv:1308.6570.

James, L. F., Lijoi, A., & Prüenster, I. 2009. Posterior analysis for Normalized Random Measures with Independent Increments. *Scandinavian Journal of Statistics*, **36**, 76–97.

Jolliffe, I.T. 2002. *Principal component analysis*. 2nd. edn. Springer Verlag, Berlin.

Kalli, M., Griffin, J. E., & Walker, S. G. 2011. Slice sampling mixture models. *Statistics and Computing*, **21**, 93–105.

Kanter, M. 1975. Stable densities under change of scale and total variation inequalities. *Annals of Probability*, **3**, 697–707.

Kingman, J. F. C. 1967. Completely Random Measures. *Pacific Journal of Mathematics*, **21**, 59–78.

Kingman, J. F. C. 1975. Random discrete distributions. *Journal of the Royal Statistical Society*, **37**, 1–22.

Kingman, J. F. C. 1978. The representation of partition structures. *Journal of the London Mathematical Society*, **18**, 374–380.

Lijoi, A., Mena, R. H., & Prüenster, I. 2005. Hierarchical mixture modelling with Normalized Inverse-Gaussian priors. *Journal of the American Statistical Association*, **100**, 1278–1291.

Lijoi, A., Mena, R. H., & Prüenster, I. 2007. Controlling the reinforcement in Bayesian nonparametric mixture models. *Journal of the Royal Statistical Society B*, **69**, 715–740.

Lijoi, A., Prünster, I., & Walker, S. G. 2008. Investigating nonparametric priors with Gibbs structure. *Statistica Sinica*, **18**, 1653 – 1668.

Lo, A.Y. 1984. On a class of Bayesian nonparametric estimates: I. density estimates. *Annals of Statistics*, **12**, 351–357.

MacEachern, S. N. 1994. Estimating Normal means with a conjugate Style Dirichlet process prior. *Communications in Statistics: Simulation and Computation*, **23**, 727–741.

Medvedovic, M., & Sivaganesan, S. 2002. Bayesian infinite mixture model based clustering of gene expression profiles. *Bioinformatics*, **18**, 1194–1206.

Miller, J. W., & Harrison, T. M. 2013. A simple example of Dirichlet process mixture inconsistency for the number of components. *In: Neural Information Processing Systems*.

Neal, R. M. 2000. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

Neal, R. M. 2003. Slice sampling. *Annals of Statistics*, **31**, 705–767.

Nieto-Barajas, L. E., Pruenster, I., & Walker, S. G. 2004. Normalized Random Measures driven by increasing additive processes. *Annals of Statistics*, **32**, 2343–2360.

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. *Probability Theory and Related Fields*, **92**, 21–39.

Pitman, J. 2003. Poisson-Kingman partitions. *Pages 1–34 of:* Goldstein, D. R. (ed), *Statistics and Science: a Festschrift for Terry Speed*. Institute of Mathematical Statistics.

Pitman, J. 2006. *Combinatorial Stochastic Processes*. Lecture Notes in Mathematics. Springer-Verlag, Berlin.

Pitman, J., & Yor, M. 1997. The two parameter Poisson-Dirichlet distribution derived from a Stable subordinator. *Annals of Probability*, **25**, 855–900.

Regazzini, E., Lijoi, A., & Prüenster, I. 2003. Distributional results for means of random measures with independent increments. *Annals of Statistics*, **31**, 560–585.

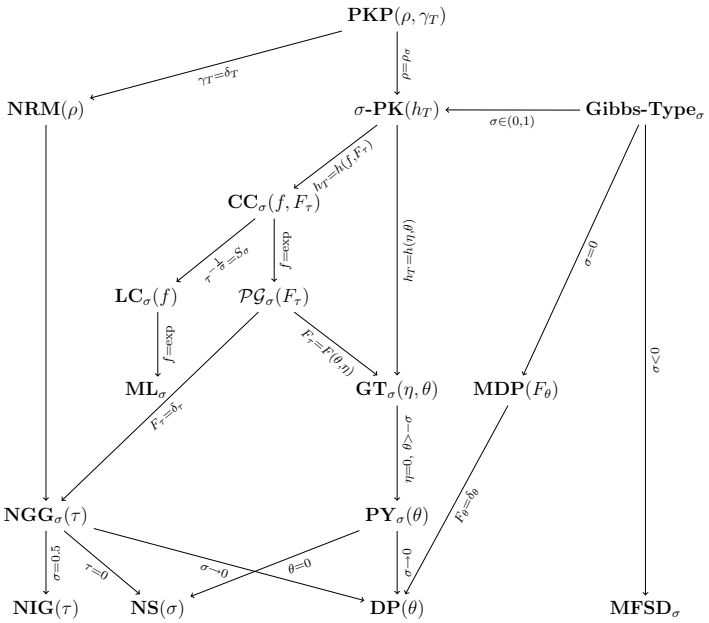
Roeder, K. 1990. Density estimation with confidence sets exemplified by super-clusters and voids in the galaxies. *Journal of the American Statistical Association*, **85**, 617–624.

Tanner, M., & Wong, W. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528–550.

Walker, S. G. 2007. Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, **36**, 45–54.

Wuertz, A., Maechler, M., & Rmetrics core team members. 2013. *stabledist*. CRAN R Package Documentation.

Zolotarev, V. M. 1966. On the representation of Stable laws by integrals. *Selected Translations Math Statist. and Probability*, **6**, 84–88.



1

Figure 1: Mixture of DPs (MDP), Mixture of finite symmetric Dirichlet (MFSD).

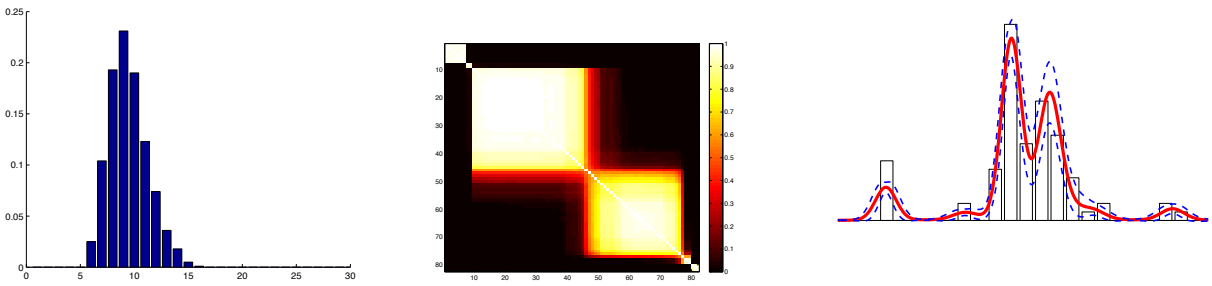


Figure 2: 210,000 iterations, 10,000 burn in and 20 thinning factor.

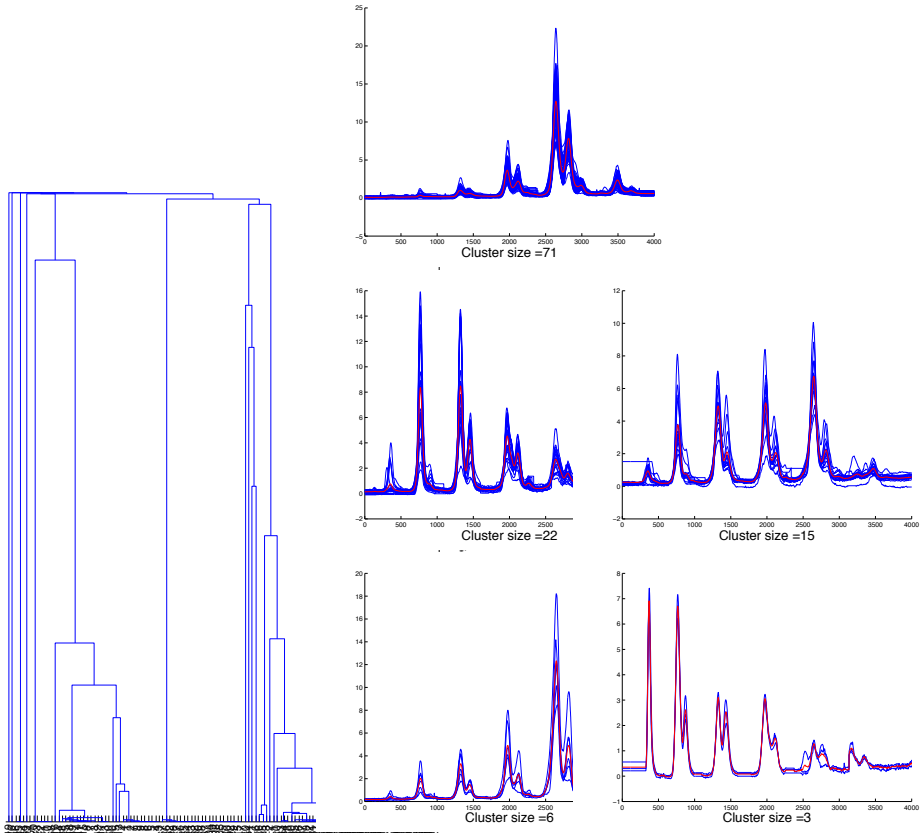


Figure 3: Dendrogram and mean profile per cluster (in red), profiles in each cluster (blue) using a NGG prior.

Table 1: Running times in seconds and number of cluster's ESS averaged over 5 chains. Unidimensional dataset, 30,000 iterations, 10,000 burn in.

Algorithm	σ	M	**Running time**	**ESS($\pm$std)**
Pitma-Yor process ($\theta = 10$)				
Marginal-Conj II	0.5	4	23770.3(2098.22)	4857.644(447.583)
Marginal-NonConj	0.5	4	46352.4(252.27)	5663.696(89.264)
Normalized Generalized Gamma process ($\tau = 1$)				
Marginal-Conj II	0.5	4	22434.1(78.191)	3400.855(110.420)
Marginal-NonConj	0.5	4	28933.5(133.97)	5361.945(88.521)

Table 2: Running times in seconds and number of cluster's ESS averaged over 10 chains. Unidimensional dataset, 50,000 iterations per chain, 20,000 burn in.

Algorithm	σ	M	**Running time**	**ESS($\pm$std)**
Pitman-Yor process ($\theta = 50$)				
Marginal-Conj I	0.5	2	1.1124e+04	4121.94(821.562)
Marginal-Conj I	0.5	6	1.1216e+04	11215.55(596.249)
Marginal-Conj I	0.5	10	1.1355e+04	12469.87(548.981)
Marginal-Conj I	0.5	15	1.1385e+04	13087.92(504.595)
Marginal-Conj I	0.5	20	1.1415e+04	12792.78(391.123)
Conditional-Conj I	0.5	-	1.5659e+04	707.82 (95.754)
Normalized Generalized Gamma process ($\tau = 50$)				
Marginal-Conj I	0.5	2	1.1617e+04	4601.63(574.339)
Marginal-Conj I	0.5	6	1.1650e+04	10296.85(425.333)
Marginal-Conj I	0.5	10	1.1692e+04	11415.41(377.418)
Marginal-Conj I	0.5	15	1.1795e+04	11473.44(374.031)
Marginal-Conj I	0.5	20	1.1875e+04	11461.08(506.744)
Conditional-Conj I	0.5	-	1.5014e+04	848.73 (135.138)

Table 3: Running times in seconds and number of cluster's ESS averaged over 5 chains. Unidimensional dataset, 30,000 iterations per chain, 10,000 burn in.

Algorithm	σ	M	**Running time**	**ESS($\pm$std)**
Pitma-Yor process ($\theta = 10$)				
Marginal-Conj I	0.3	4	4685.7(84.104)	2382.799(169.359)
Marginal-Conj I	0.5	4	4757.2(37.077)	2944.065(195.011)
Marginal-Conj I	0.7	4	4655.2(52.514)	2726.232(132.828)
Conditional-Conj I	0.5	-	10141.6(237.735)	905.444(41.475)
Normalized Stable process				
Marginal-Conj I	0.3	4	7658.3(193.773)	2630.264(429.877)
Marginal-Conj I	0.5	4	8203.1(106.798)	3139.412(351.788)
Marginal-Conj I	0.7	4	8095.7(85.2640)	2394.756(295.923)
Conditional-Conj I	0.5	-	10033.1(22.647)	912.382(167.089)
Normalized Generalized Gamma process ($\tau = 1$)				
Marginal-Conj I	0.3	4	7685.8(208.98)	3587.733(569.984)
Marginal-Conj I	0.5	4	8055.6(93.164)	4443.905(367.297)
Marginal-Conj I	0.7	4	8117.9(113.188)	4936.649(411.568)
Conditional-Conj I	0.5	-	10046.9(206.538)	1000.214(70.148)

Table 4: Unidimensional experiment's average leave-one-out predictive probabilities.

Examples	Average predictive probability
PY	0.1033(0.052)
NGG	**0.1228(0.065)**
Gamma Tilted	0.1186(0.065)
NS	0.1123(0.057)

Table 5: Multidimensional experiment's average 5-fold predictive probabilities.

Examples	Average predictive probability
DP	5.5484e-12 (7.6848e-13)
PY	4.1285e-12 (7.5549e-13)
NGG	**9.6266e-12 (3.4035e-12)**
Gamma tilted	6.7099e-12(1.5767e-12)
NS	8.3328e-12(9.7106e-13)
Lamperti tilted	5.4251e-12 (1.0538e-12)