Slice sampling σ-stable Poisson-Kingman mixture models

Slice Sampling σ-Stable Poisson-Kingman Mixture Models

Stefano FAVARO and Stephen G. WALKER

The article is concerned with the use of Markov chain Monte Carlo methods for posterior sampling in Bayesian nonparametric mixture models. In particular, we consider the problem of slice sampling mixture models for a large class of mixing measures generalizing the celebrated Dirichlet process. Such a class of measures, known in the literature as σ-stable Poisson-Kingman models, includes as special cases most of the discrete priors currently known in Bayesian nonparametrics, for example, the two-parameter Poisson-Dirichlet process and the normalized generalized Gamma process. The proposed approach is illustrated on some simulated data examples. This article has online supplementary material.

Key Words: Bayesian nonparametrics; MCMC posterior sampling; Normalized random measures; Size-biased random permutation; Stick-breaking representation for the σ-Stable Poisson-Kingman models.

1. INTRODUCTION

Density estimation is a standard issue in Bayesian nonparametrics and requires the specification of priors selecting almost surely distributions admitting a density function. A useful and general device for defining a prior on densities has been introduced by Lo (1984) in terms of the so-called infinite dimensional mixture model. The basic idea consists in introducing an almost surely discrete random probability measure $\tilde{P}$ on $\mathbb{X}$, a Polish space endowed with the usual Borel σ-field $\mathcal{X}$, which is convoluted with a suitable kernel k. Specifically, let $\tilde{P}$ be an almost surely discrete random probability measure on $\mathbb{X}$, that is,

$$\tilde{P} = \sum_{j \geq 1} \tilde{P}_j \delta_{X_j} \tag{1}$$

for some sequence $(X_j)_{j \geq 1}$ of $\mathbb{X}$-valued random locations and some sequence $(\tilde{P}_j)_{j \geq 1}$ of nonnegative random masses that sum to one, almost surely. A random density function can

Stefano Favaro, Department of Economics and Statistics, University of Torino, Turin, and Collegio Carlo Alberto, Moncalieri, Italy (E-mail: *stefano.favaro@unito.it*). Stephen G. Walker, Institute of Mathematics, Statistics and Actuarial Sciences, University of Kent, Canterbury, Kent CT2 7NZ, UK (E-mail: *S.G.Walker@kent.ac.uk*).

Journal of Computational and Graphical Statistics, Volume 22, Number 4, Pages 830–847
DOI: 10.1080/10618600.2012.681211

then be defined as follows:

$$f_{\tilde{P}}(y) = \int_{\mathbb{X}} k(y|x) d\tilde{P}(x), \tag{2}$$

where the $k(y|x)$ is a continuous and possibly multivariate density function for each fixed $x \in \mathbb{X}$. The original formulation by Lo (1984) set the mixing measure $\tilde{P}$ to coincide with the Dirichlet process introduced by Ferguson (1973). Hence, it takes on the name of Dirichlet process mixture model. However, it is apparent that one can replace the Dirichlet process with any almost surely discrete random probability measure $\tilde{P}$. See Lijoi and Prünster (2010) for a detailed overview on some classes of mixing measures alternative to the Dirichlet process.

Various Markov chain Monte Carlo (MCMC) methods for sampling from the posterior distribution of $f_{\tilde{P}}$ have been proposed in the literature. Early work in this direction exploited the integration with respect to the underlying mixing measure $\tilde{P}$, thus removing the infinite dimensional aspect of the problem. The main references in this area are represented by the algorithms devised by Escobar (1994) and Escobar and West (1995), and originally developed for the Dirichlet process mixture model (see also the subsequent variants proposed by MacEachern 1994; MacEachern 1998; MacEachern and Müller 1998; Neal 2000). These sampling methods are usually referred to as marginal methods and, as noted by Ishwaran and James (2001), they can be applied to any mixing measure $\tilde{P}$ for which the induced system of predictive distributions is known in explicit form. An alternative family of sampling methods rely on the simulation of the whole posterior mixture model and, hence, of the underlying mixing measure $\tilde{P}$ as well. These methods, usually referred to as conditional methods, do not remove the infinite dimensional aspect of the problem and they focus on finding appropriate ways for sampling a sufficient but finite number of random masses of $\tilde{P}$ in Equation (1). Recently, for the Dirichlet process mixture model, there has been interest in developing conditional methods that only use a finite number of random masses and allow inference according to the true infinite dimensional mixing measure $\tilde{P}$. The retrospective sampler by Papaspiliopoulos and Roberts (2008) used a carefully constructed Metropolis-Hastings update, whereas Walker (2007) used slice sampling ideas (see also Griffin and Walker 2011; Kalli, Griffin, and Walker 2011 for some recent developments on the slice sampling approach).

In the present article, we focus on the mixture model in Equation (2) with mixing measure $\tilde{P}$ in a large class of almost surely discrete random probability measures generalizing the Dirichlet process and known in the literature as σ-stable Poisson-Kingman models. Such a class of models has been introduced by Pitman (2003) and includes as special cases most of the discrete priors currently known in Bayesian nonparametrics. Indeed, apart from the Dirichlet process, it includes the two-parameter Poisson-Dirichlet process by Perman, Pitman, and Yor (1992) (see also Pitman 1995; Pitman and Yor 1997) and, consequently, the normalized σ-stable process introduced by Kingman (1975) in the context of optimal storage problems. It also includes the normalized generalized Gamma process by Pitman (2003) (see also Lijoi, Mena, and Prünster 2007) and, consequently, the normalized inverse Gaussian process introduced by Lijoi, Mena, and Prünster (2005) in the context of Bayesian nonparametric mixture modeling. We aim to extend the slice sampling ideas by Walker (2007), originally introduced for the Dirichlet process mixture

model, to the σ-stable Poisson-Kingman mixture model, that is, a mixture model with an underlying σ-stable Poisson-Kingman mixing measure. The proposed slice sampling is achieved by resorting to a collection of results on size-biased sampling of Poisson point processes, introduced by Perman, Pitman, and Yor (1992) and here applied to derive a stick-breaking representation for the σ-stable Poisson-Kingman models. Indeed, the proposed stick-breaking representation of the underlying mixing measure $\tilde{P}$ turns out to be a key tool for MCMC posterior sampling via the slice sampling approach by Walker (2007).

The article is structured as follows. Section 2 reviews the general definition of Poisson-Kingman model. Section 3 presents the subclass of σ-stable Poisson-Kingman models, the corresponding stick-breaking representation, and the detailed slice sampling algorithm for posterior sampling σ-stable Poisson-Kingman mixture models. Section 4 contains illustrations of the proposed algorithm and Section 5 concludes with a brief discussion.

2. POISSON-KINGMAN MODELS

To review the general definition of the Poisson-Kingman model, we start by providing a succinct description of the class of completely random measures (CRMs) introduced by Kingman (1967). The reader is referred to the monograph by Kingman (1993) for a detailed treatment on the subject of CRMs.

Let $\tilde{\mu}$ be a random element defined on some probability space and taking values on the space of bounded finite measures on $(\mathbb{X}, \mathcal{X})$. Kingman (1967) termed the random element $\tilde{\mu}$ CRM if for any $n \geq 1$ and $A_1, \ldots, A_n$ in $\mathcal{X}$, with $A_i \cap A_j = \oslash$ for $i \neq j$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent. Kingman (1967) showed that any CRM can be decomposed into the sum of three independent components: a nonrandom measure, a countable collection of nonnegative random masses at nonrandom $\mathbb{X}$-valued locations, and a countable collection of nonnegative random masses at $\mathbb{X}$-valued random locations. We consider CRMs with the sole component of random masses $(\tilde{J}_j)_{j\geq 1}$ at random locations $(X_j)_{j\geq 1}$, that is,

$$\tilde{\mu}(\cdot) = \sum_{j\geq 1} \tilde{J}_j \delta_{X_j}(\cdot). \tag{3}$$

The distribution of $\tilde{\mu}$ in Equation (3) is determined by its Laplace functional transform having the Lévy-Khintchine representation

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

for any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $\int_{\mathbb{X}} |f(x)|\tilde{\mu}(dx) < +\infty$ almost surely. The measure ν on $\mathbb{R}^+ \times \mathbb{X}$ is usually referred to as Lévy intensity measure and it characterizes uniquely $\tilde{\mu}$. For our purpose, it is enough to focus on Lévy intensity measures ν factorizing as $\nu(ds, dy) = \rho(ds)P_0(dy)$ for some Lévy measure ρ absolutely continuous with respect to the Lebesgue measure and some nonatomic probability measure P_0 on $(\mathbb{X}, \mathcal{X})$. Such a factorization implies the independence between the random locations and the random masses so that, without loss of generality, the random locations $(X_j)_{j\geq 1}$ can be assumed to be independent and identically distributed according to P_0, while the

distribution of the corresponding random masses $(\tilde{J}_j)_{j\geq 1}$ is governed by the Lévy measure ρ.

The definition of Poisson-Kingman model has been proposed by Pitman (2003) in terms of a suitable transformation of CRMs. Such a definition is strictly related to the definition of homogeneous normalized random measure with independent increments (NRMI), introduced in Bayesian nonparametrics by James, Lijoi, and Prünster (2008). Let $\tilde{\mu}$ be a CRM with Lévy intensity measure $\nu(ds, dy) = \rho(ds)P_0(dy)$ and let $T = \sum_{j\geq 1} \tilde{J}_j$ be the corresponding total mass that is assumed to be absolutely continuous with respect to the Lebesgue measure. Now, if $\text{card}(\{\tilde{J}_j : j \geq 1\} \cap (0, \epsilon)) = \int_0^\epsilon \rho(s)ds = +\infty$ for any $\epsilon > 0$, one can define an almost surely discrete random probability measure $\tilde{P}$ on $\mathbb{X}$ as

$$\tilde{P}(\cdot) = \frac{\tilde{\mu}(\cdot)}{T} = \sum_{j\geq 1} \tilde{P}_j \delta_{X_j}(\cdot), \tag{4}$$

where

$$\tilde{P}_j = \frac{\tilde{J}_j}{T} \tag{5}$$

and $(X_j)_{j\geq 1}$ is a sequence of random variables, independent of $(\tilde{P}_j)_{j\geq 1}$, and independent and identically distributed according to P_0. In the article by James, Lijoi, and Prünster (2008), the random probability measure $\tilde{P}$ in Equation (4) is termed homogeneous NRMI and it is denoted by NRMI(ρ, P_0). Intuitively, a Poisson-Kingman model can be defined as a generalization of a NRMI(ρ, P_0) obtained by suitably deforming the distribution of the normalizing total mass T. A formal definition of the class of Poisson-Kingman models is given below.

Consider a CRM $\tilde{\mu}$ with Lévy intensity measure $\nu(ds, dy) = \rho(ds)P_0(dy)$ and denote by $(J_{(j)})_{j\geq 1}$ the decreasing rearrangement of the sequence of random masses of $\tilde{\mu}$. Moreover, denote by $(P_{(j)})_{j\geq 1}$ the sequence of decreasing ordered random probabilities obtained by normalizing the sequence $(J_{(j)})_{j\geq 1}$ with respect to the total mass T, that is,

$$P_{(j)} = \frac{J_{(j)}}{T}. \tag{6}$$

In other terms, the sequence $(P_{(j)})_{j\geq 1}$ represents the decreasing rearrangement of the random probabilities of $\tilde{P}$ defined in Equation (5).

Definition 2.1 Let γ be a distribution on $\mathbb{R}^+$ and let $P_{\rho,t}$ be the regular conditional distribution of the sequence $(P_{(j)})_{j\geq 1}$ of decreasing ordered random probabilities defined in Equation (6), given the total mass $T = t$. Then, the distribution

$$\int_{\mathbb{R}^+} P_{\rho,t}\gamma(dt)$$

is termed Poisson-Kingman distribution with Lévy measure ρ and mixing distribution γ. Such a distribution is denoted by PK(ρ, γ).

The definition of Poisson-Kingman distribution can be applied for defining an almost surely discrete random probability measure P on $\mathbb{X}$ as

$$P(\cdot) = \sum_{j \geq 1} P_{(j)} \delta_{X_j}(\cdot), \tag{7}$$

where $(P_{(j)})_{j \geq 1}$ is distributed according to a PK(ρ, γ) distribution and $(X_j)_{j \geq 1}$ is a sequence of random variables, independent of $(P_{(j)})_{j \geq 1}$, and independent and identically distributed according to P_0. In the article by Pitman (2003), the random probability measure P in Equation (7) is termed Poisson-Kingman model and it is denoted by PK(ρ, γ, P_0). Clearly, the definition of NRMI(ρ, P_0) can be recovered as a special case of the definition of PK(ρ, γ, P_0) model by setting γ to coincide with the distribution of the total mass T.

3. SAMPLING σ-STABLE POISSON-KINGMAN MIXTURE MODELS

The class of σ-stable Poisson-Kingman models, also referred to as the class of Gibbs-type random probability measures, has been introduced by Pitman (2003) and further investigated by Gnedin and Pitman (2005). Consider a sequence $(P_{(j)})_{j \geq 1}$ of decreasing ordered random probabilities distributed according to a PK$(\rho^{(\sigma)}, \gamma)$ distribution where

$$\rho^{(\sigma)}(ds) = \frac{\sigma}{\Gamma(1-\sigma)} s^{-\sigma-1} ds,$$

for some $\sigma \in (0, 1)$, and where γ is a distribution absolutely continuous with respect to the Lebesgue measure with density function g_γ. In the article by Pitman (2003), the PK$(\rho^{(\sigma)}, g_\gamma, P_0)$ model is termed σ-stable Poisson-Kingman model. Without loss of generality, we assume $g_\gamma(t) = h(t) f_\sigma(t)$ for some nonnegative measurable function h such that g_γ is a proper density function, and where f_σ denotes the positive σ-stable density function, that is,

$$f_\sigma(t) = \frac{1}{\pi} \sum_{i \geq 0} \frac{(-1)^{i+1}}{i!} \sin(\pi \sigma i) \frac{\Gamma(\sigma i + 1)}{t^{\sigma i + 1}}. \tag{8}$$

It is worth pointing out that the class of PK$(\rho^{(\sigma)}, h f_\sigma, P_0)$ models includes most of the discrete priors currently known in Bayesian nonparametrics. Indeed, as we will see in the examples below, by a suitable specification of the function h and the parameter σ, one can recover as special cases the Dirichlet process, the normalized σ-stable process, the two-parameter Poisson-Dirichlet process, the normalized inverse Gaussian process, and the normalized generalized Gamma process.

For our purpose, it is useful to consider a different rearrangement of the sequence of random probabilities $(P_{(j)})_{j \geq 1}$ distributed according to a PK$(\rho^{(\sigma)}, \gamma)$ distribution. Such a rearrangement is obtained via the so-called size-biased random permutation approach. Specifically, this means that

$$P_j = P_{(\pi_j)},$$

where for any positive integer $k \geq 1$ and for all the finite sets $\{i_1, \ldots, i_k\}$ of distinct positive integers, the conditional probability of the event $\{\pi_j = i_j \text{ for all } 1 \leq i \leq k\}$ given

$(P_{(j)})_{j\geq 1}$ coincides with

$$P_{(i_1)}\frac{P_{(i_2)}}{1-P_{(i_1)}}\cdots\frac{P_{(i_k)}}{1-P_{(i_1)}-\cdots-P_{(i_k)}}.$$

The sequence of random probabilities $(P_j)_{j\geq 1}$ is termed as the size-biased random permutation of the sequence $(P_{(j)})_{j\geq 1}$. The reader is referred to the literature by Perman, Pitman, and Yor (1992) and Pitman (1996) for a detailed study on the interplay between size-biased random permutations and almost surely discrete random probability measures. In particular, Perman, Pitman, and Yor (1992) provided a collection of results that can be usefully applied to obtain a stick-breaking representation for the size-biased random permutation of a sequence decreasing ordered random probabilities distributed according to a PK$(\rho^{(\sigma)}, \gamma)$ distribution. The next proposition is an application of Theorem 2.1 in the article by Perman, Pitman, and Yor (1992) and it represents the key result for slice sampling σ-stable Poisson-Kingman mixture models.

Proposition 3.1 Let $(P_j)_{j\geq 1}$ be the size-biased random permutation of a sequence $(P_{(j)})_{j\geq 1}$ of decreasing ordered random probabilities distributed according to a PK$(\rho^{(\sigma)}, g_\gamma)$ distribution. Then,

$$P_j = V_j\prod_{i=1}^{j-1}(1-V_i) \tag{9}$$

for a sequence of random variables $(V_j)_{j\geq 1}$ such that $V_j|T, V_1, \ldots, V_{j-1}$ is absolute continuous with respect to the Lebesgue measure, and its density function on $(0, 1)$ coincides with

$$g(v_j|t, v_1, \ldots, v_{j-1}) = \frac{\sigma(tz_j)^{-\sigma}}{\Gamma(1-\sigma)f_\sigma(tz_j)}v_j^{-\sigma}f_\sigma(tz_j(1-v_j)) \tag{10}$$

for any $j \geq 1$, where $z_j = \prod_{i=1}^{j-1}(1-v_i)$ with the proviso $z_1 = 1$. The corresponding PK$(\rho^{(\sigma)}, g_\gamma, P_0)$ model P can be represented as

$$P(\cdot) = \sum_{i\geq 1} P_j\delta_{X_j}(\cdot), \tag{11}$$

where $(X_j)_{j\geq 1}$ is a sequence of random variables, independent of $(P_j)_{j\geq 1}$, and independent and identically distributed according to P_0.

The following examples discuss the stick-breaking representations of two PK$(\rho^{(\sigma)}, hf_\sigma, P_0)$ models well known in Bayesian nonparametrics: the two-parameter Poisson-Dirichlet process and the normalized generalized Gamma process. Both the stick-breaking representations include as a special case the celebrated stick-breaking representation of the Dirichlet process introduced by Sethuraman (1994).

Example 3.1 Let $(P_{(j)})_{j\geq 1}$ be a sequence of decreasing ordered random probabilities distributed according to a PK$(\rho^{(\sigma)}, w^{(\theta,\sigma)})$ distribution with

$$w^{(\theta,\sigma)}(t) = \frac{\Gamma(\theta+1)}{\Gamma(\theta/\sigma+1)}t^{-\theta}f_\sigma(t)$$

for some $\sigma \in (0, 1)$ and $\theta > -\sigma$. The corresponding PK($\rho^{(\sigma)}$, $w^{(\theta,\sigma)}$, P_0) model P is known in the literature as two-parameter Poisson-Dirichlet process. It includes as special cases the normalized σ-stable process recovered by setting $\theta = 0$, and the Dirichlet process recovered by setting $\sigma \to 0$. An application of Proposition 3.1 leads to the stick-breaking representation of the two-parameter Poisson-Dirichlet process

$$P(\cdot) = \sum_{j\geq 1} V_j \prod_{i=1}^{j-1}(1 - V_i)\delta_{X_j}(\cdot), \tag{12}$$

where $(V_j)_{j\geq 1}$ is a sequence of independent random variables such that V_j is distributed according to a Beta distribution with parameter $(1 - \sigma, \theta + j\sigma)$, for any $j \geq 1$. Moreover, $(X_j)_{j\geq 1}$ is a sequence of random variables, independent of $(V_j)_{j\geq 1}$, and independent and identically distributed according to P_0. The stick-breaking representation in Equation (12) was originally obtained by Perman, Pitman, and Yor (1992) and further investigated by Pitman (1996).

Example 3.2 Let $(P_{(j)})_{j\geq 1}$ be a sequence of decreasing ordered random probabilities distributed according to a PK($\rho^{(\sigma)}$, $v^{(b,\sigma)}$) distribution with

$$v^{(b,\sigma)}(t) = \mathrm{e}^{b^\sigma - bt} f_\sigma(t)$$

for some $\sigma \in (0, 1)$ and $b > 0$. The corresponding PK($\rho^{(\sigma)}$, $v^{(b,\sigma)}$, P_0) model P is known in the literature as normalized generalized Gamma process. It includes as special cases the normalized inverse Gaussian process recovered by setting $\sigma = 1/2$ and the normalized σ-stable process recovered by setting $b = 0$. An application of Proposition 3.1 leads to the stick-breaking representation of the normalized generalized Gamma process

$$P(\cdot) = \sum_{j\geq 1} V_j \prod_{i=1}^{j-1}(1 - V_i)\delta_{X_j}(\cdot), \tag{13}$$

where $(V_j)_{j\geq 1}$ is a sequence of random variables such that $V_j|V_1, \ldots, V_{j-1}$ is absolute continuous with respect to the Lebesgue measure, and its density function on $(0, 1)$ coincides with

$$g(v_j|v_1, \ldots, v_{j-1}) = \frac{\sigma}{\Gamma(1 - \sigma)}(v_j z_j)^{-\sigma} \frac{\int_0^{+\infty} t^{-j\sigma}\mathrm{e}^{-bt} f_\sigma(tz_{j+1})dt}{\int_0^{+\infty} t^{-(j-1)\sigma}\mathrm{e}^{-bt} f_\sigma(tz_j)dt}$$

for any $j \geq 1$, where we set $z_j = \prod_{i=1}^{j-1}(1 - v_i)$ with the proviso $z_1 = 1$. Moreover, $(X_j)_{j\geq 1}$ is a sequence of random variables, independent of $(V_j)_{j\geq 1}$, and independent and identically distributed according to P_0. See Favaro, Lijoi, and Prünster (2012) for a detailed analysis of the stick-breaking representation in Equation (13) with $\sigma = 1/2$.

3.1 The Slice Sampling Algorithm

The stick-breaking representation in Proposition 3.1 can be usefully applied to slice sample σ-stable Poisson-Kingman mixture models. The starting point is the mixture model f_P in Equation (2) with a mixing measure P being a PK($\rho^{(\sigma)}$, hf_σ, P_0) model. In particular,

according to the series representation in Equation (11) of P, we can write

$$f_P(y) = \sum_{j \geq 1} P_j k(y|X_j). \tag{14}$$

Using the slice sampling idea introduced by Walker (2007), the infinite dimensional aspect of f_P in Equation (14) is tackled by introducing a suitable latent random variable U, which has joint density function with Y given by

$$f_{(P,U)}(y, u) = \sum_{j \geq 1} \mathbb{1}(u < P_j) k(y|X_j). \tag{15}$$

Given the latent variable U, the number of components of the mixture model f_P is finite, the indices being $A_u = \{j : P_j > U\}$. In particular, given the latent variable U, one has the finite mixture model

$$f_{(P|U)}(y|u) = \frac{1}{N_u} \sum_{j \in A_u} k(y|X_j)$$

where the size of the set A_u is determined by $\sum_{j \geq 1} \mathbb{1}(P_j > u)$ and $N_u = \sum_{j \in A_u} P_j$. One can then introduce a further latent random variable D, which indicates the components of the mixture model from which y is taken and gives the joint density function

$$f_{(P,U,D)}(y, u, d) = \mathbb{1}(u < P_d) k(y|X_d). \tag{16}$$

Clearly, without the latent variable U, the latent variable D can take an infinite number of values that would make the implementation of MCMC algorithms problematic.

The latent random variable U introduced in the augmented mixture model in Equation (15) can be easily updated because it is assumed to be uniformly distributed. Moreover, given the latent variable U, the set A_u can be easily found using the stick-breaking representation of the mixing measure P. Here, to improve the efficiency in the search for the set A_u, we consider the more general class of slice sampler recently introduced by Kalli, Griffin, and Walker (2011), the so-called slice-efficient sampler. Specifically, we replace Equation (16) with the joint density function

$$f_{(P,U,D)}(y, u, d) = \mathrm{e}^{\xi d} \mathbb{1}(u < \mathrm{e}^{-\xi d}) P_d k(y|X_d) \tag{17}$$

for some $\xi > 0$ (see Kalli, Griffin, and Walker (2011) for details). From Equation (17), the complete data likelihood based on a sample of size n is then easily seen to be

$$\prod_{i=1}^{n} \mathrm{e}^{\xi d_i} \mathbb{1}(u_i < \mathrm{e}^{-\xi d_i}) P_{d_i} k(y_i|X_{d_i}). \tag{18}$$

In particular, it can be easily verified that for any $i \geq 1$, conditioning on the latent variable $U_i = u_i$, it is that $1 \leq D_i \leq N_i$ where $N_i = \lfloor -(1/\xi) \log u_i \rfloor$ with $\lfloor x \rfloor$ being the integer part of x. Therefore, to present a valid MCMC algorithm we only need to record the N_i's at each iteration so as to adequately sample the D_i's. Moreover, if we set $N = \max\{N_i\}$, note that we only need the P_j's and the X_j's for $j = 1, \ldots, N$. We conclude this section by describing the sampling of the random probabilities P_j's defined via the stick-breaking representation in Equation (9). The sampling of the D_i's given the P_j's and the sampling of the X_j's given the $D_i = d_i$'s will be discussed in the next section together with some illustrations of the proposed algorithm.

Using the stick-breaking representation of the mixing measure P, for any $N \geq 1$ we need to sample the random variables $(T, V_1, \ldots, V_N)$ from their conditional distributions, given the allocation random variables $(D_1, \ldots, D_n)$. For such purpose, by an application of Equation (10) it can be easily verified that the conditional density function of $(T, V_1, \ldots, V_N)$, given the allocation variables $(D_1, \ldots, D_n)$, is

$$f(t, v_1, \ldots, v_N | d_1, \ldots, d_n) \propto h(t) t^{-N\sigma} f_\sigma \left(t \prod_{j=1}^{N} (1 - v_j) \right) \prod_{j=1}^{N} v_j^{n_j - \sigma} (1 - v_j)^{m_j - (N-j)\sigma}, \tag{19}$$

where

$$n_j = \sum_{i=1}^{n} \mathbb{1}(d_i = j) \tag{20}$$

and

$$m_j = \sum_{i=1}^{n} \mathbb{1}(d_i > j). \tag{21}$$

To sample from Equation (19), we apply the representation by Kanter (1975) of the positive σ-stable density function f_σ. Specifically, setting

$$K_\sigma(z) = \left(\frac{\sin(\pi\sigma z)}{\sin(\pi z)} \right)^{-\frac{1}{1-\sigma}} \left(\frac{\sin((1-\sigma)\pi z)}{\sin(\pi\sigma z)} \right)$$

for $0 < z < 1$, it follows from the article by Kanter (1975) that

$$f_\sigma(s) = \frac{\sigma}{1-\sigma} s^{-\frac{1}{1-\sigma}} \int_0^1 \mathrm{e}^{-s^{-\frac{\sigma}{1-\sigma}} K_\sigma(z)} K_\sigma(z) dz. \tag{22}$$

Therefore, using the representation in Equation (22) for a positive σ-stable density function f_σ, we can introduce an augmented version of $f(t, v_1, \ldots, v_N | d_1, \ldots, d_n)$ given by

$$f^*(t, z, v_1, \ldots, v_N | d_1, \ldots, d_n) \propto h(t) t^{-N\sigma} f_\sigma^* \left(t \prod_{j=1}^{N} (1 - v_j), z \right) \times \prod_{j=1}^{N} v_j^{n_j - \sigma} (1 - v_j)^{m_j - (N-j)\sigma},$$

where

$$f_\sigma^*(s, z) \propto s^{-\frac{1}{1-\sigma}} \mathrm{e}^{-s^{-\frac{\sigma}{1-\sigma}} K_\sigma(z)} K_\sigma(z).$$

To simplify this part of the model and to facilitate an easy MCMC implementation, we introduce a further latent random variable V that interacts with the model providing an augmented version of the density function $f_\sigma^*(s, z)$ given by

$$f_\sigma^*(s, z, v) \propto s^{-\frac{1}{1-\sigma}} K_\sigma(z) \mathbb{1}\left(v < \mathrm{e}^{-s^{-\frac{\sigma}{1-\sigma}} K_\sigma(z)} \right).$$

Finally, we are now able to describe the full conditional distributions of the random variables $(T, V, Z, V_1, \ldots, V_N)$ that are required to be sampled. Specifically, for any $N \geq 1$ and for

any $j = 1, \ldots, N$, one has the following conditional density function for the V_j's

$$f^*(v_j|\text{ rest }) \propto (1-v_j)^{m_j-(N-j)\sigma-\frac{1}{1-\sigma}} v_j^{n_j-\sigma} \mathbb{1}\left(1-v_j > \frac{\left(\frac{K_\sigma(z)}{-\log v}\right)^{\frac{1-\sigma}{\sigma}}}{t\prod_{l\neq j}(1-v_l)}\right).$$

Moreover, for any $N \geq 1$

$$f^*(z|\text{ rest }) \propto K_\sigma(z)\mathbb{1}\left(K_\sigma(z) < -(\log v)\, s^{\frac{\sigma}{1-\sigma}}\right)$$

and

$$f^*(t|\text{ rest}) \propto h(t)\, t^{-N\sigma-\frac{1}{1-\sigma}} \mathbb{1}\left(t > \frac{\left(\frac{K_\sigma(z)}{-\log v}\right)^{\frac{1-\sigma}{\sigma}}}{\prod_{l=1}^{N}(1-v_l)}\right).$$

According to Kanter (1975), the function $K_\sigma(\cdot)$ is strictly decreasing on $(0, 1)$ and hence a value z_0 can be obtained easily whereby $K_\sigma(z_0) > -(\log v)\, s^{\frac{\sigma}{1-\sigma}}$ yet z_0 is close to z_1, where $K_\sigma(z_1) = -(\log v)\, s^{\sigma/1-\sigma}$. We can now use rejection sampling by sampling a proposal z^* as a uniform random variable over the set $(z_0, 1)$ and accepting as a sample from $f^*(z|\text{ rest })$ with probability

$$\mathbb{1}(z^* > z_1)\frac{K_\sigma(z^*)}{K_\sigma(z_0)}.$$

In particular, we know that $z^* > z_1$ when $K_\sigma(z^*) < -(\log v)s^{\sigma/1-\sigma}$. This algorithm works very well for all $\sigma \in (0, 1)$ and z_1. An illustration is given based on the choice of $\sigma = 0.4$ and $z_1 = 0.89$; Figure 1 presents the histogram of 1000 samples based on uniform rejection sampling.

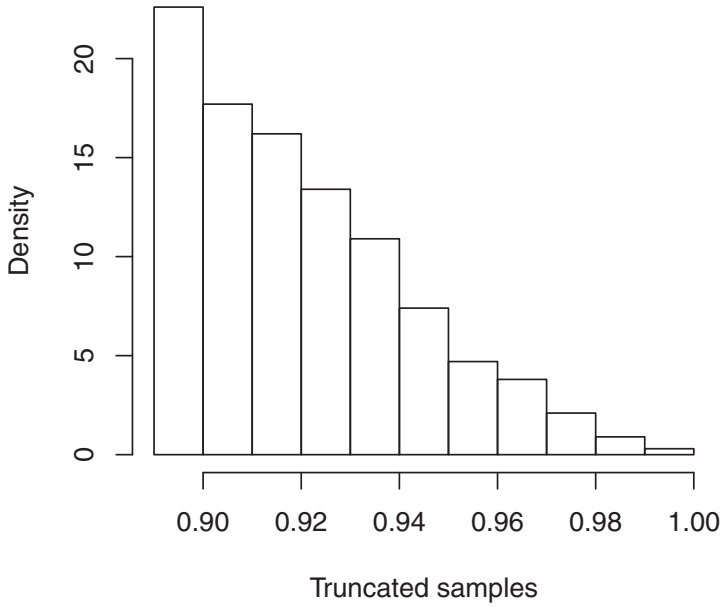

Figure 1. Samples from $K_{0.4}(z)$ restricted to lie in $(0.89, 1)$.

4. ILLUSTRATIONS

In this section, we illustrate on simulated data and on a real dataset—the slice sampling algorithm for σ-stable Poisson-Kingman mixture models. To start, for a $n = 50$ sample size, we generate independent and identically distributed random variables Y_i's as follows: with probability $1/2$, $Y_i \sim \mathcal{N}(-10, 2^2)$ and with probability $1/2$, $Y_i \sim \mathcal{N}(10, 2^2)$, where we denoted by $\mathcal{N}(\mu, \sigma^2)$ the Gaussian distribution with mean μ and variance σ^2. We employ the standard Gaussian mixture model assuming a known variance $\tau^2 = 2^2$ for each component and assuming a mixing measure P being a PK($\rho^{(1/2)}$, $hf_{1/2}$, P_0). Hence, according to Equation (14), the corresponding σ-stable Poisson-Kingman Gaussian mixture model is

$$f_P(y) = \sum_{j\geq 1} P_j \mathcal{N}(y|\mu_j, \tau^2),$$

where $(P_j)_{j\geq 1}$ is the sequence of stick-breaking random probabilities described in Proposition 3.1. Finally, for any $j \geq 1$, we denote by $\pi(\mu_j)$ the prior distribution for μ_j, which is taken to be a Gaussian distribution with mean 0 and variance 10^2. From Equation (18), the complete data likelihood based on the sample of size $n = 50$ is given by

$$\prod_{i=1}^{50} \mathrm{e}^{\xi d_i} \mathbb{1}(u_i < \mathrm{e}^{-\xi d_i}) P_{d_i} \mathcal{N}(y_i|\mu_{d_i}, \tau^2)$$

for some choice of $\xi > 0$. In the applications that follow, this is fixed at $\xi = 1$. For any $i \geq 1$, given the latent variable U_i, the choice of the latent variable D_i is finite and hence can be sampled. Moreover, once the latent variable D_i have been sampled, then it is possible to sample the μ_j. To know how many of the μ_j's and the V_j's we need to sample, we use the values of $N_i = \lfloor(-\log w_i)/\xi\rfloor$ and $N = \max\{N_i\}$. Specifically, if we sample the μ_j's and the V_j's for $j = 1, \ldots, N$, then we can implement a precise sampler since the D_i's are bounded by the N_i's, respectively.

We start illustrating the sampling algorithm by describing the full conditional distribution of $(T, V_1, \ldots, V_N)$, for any $N \geq 1$. In particular, from Equation (19), the conditional density function of the random variables $(T, V_1, \ldots, V_N)$, given the allocation random variables $(D_1, \ldots, D_n)$, is proportional to

$$h(t)\, f_{1/2}\left(t \prod_{j=1}^{N}(1 - v_j)\right) t^{-N/2} \prod_{j=1}^{N} v_j^{n_j - 1/2}(1 - v_j)^{m_j - (N-j)/2}$$

with n_j and m_j specified by Equations (20) and (21). Also, for $\sigma = 1/2$, the σ-stable density function f_σ in Equation (8) reduces to

$$f_{1/2}(t) \propto t^{-3/2}\mathrm{e}^{-\frac{1}{4t}}$$

(see, e.g., Section 0.3 in the article by Pitman 2006). To make the sampler easier to cope with, for the exponential term appearing in the $f_{1/2}$ density function, we included a latent random variable V that enters the latent model via the term

$$\mathbb{1}\left(v < \exp\left\{-\frac{1}{4t(1 - v_1)\ldots(1 - v_N)}\right\}\right).$$

Now the sampling of each V_j's is straightforward and the corresponding full conditional density function is proportional to

$$v_j^{n_j-1/2}(1-v_j)^{m_j-(N-j)/2-3/2}\,\mathbb{1}\left((1-v_j) > \frac{1}{(-\log v)\,4t\prod_{l\neq j}(1-v_l)}\right).$$

Finally, we need to consider the full conditional density function of the total mass variable T, which is proportional to

$$h(t)\,t^{-N/2-3/2}\,\mathbb{1}\left(t > \frac{1}{(-\log v)\,4\prod_j(1-v_j)}\right).$$

Both of these densities pose no real problems in being sampled. Once we have sampled $(V_1,\ldots,V_N)$, then we have the corresponding random probabilities $(P_1,\ldots,P_N)$. Hence, we can now sample the allocation variables $(D_1,\ldots,D_n)$ according to the full conditional distribution

$$\mathbb{P}(D_i = j|\text{ rest }) \propto P_j\, e^{\xi j}\mathcal{N}(y_i|\mu_j,\tau^2)\mathbb{1}(j\leq N_i)$$

and, finally, we can also sample the μ_j's according to the following full conditional density function

$$\prod_{d_i=j}\mathcal{N}(y_i|\mu_j,\tau^2)\pi(\mu_j).$$

In the specific illustrations, we took the prior π to be a Gaussian distribution with mean 0 and variance 10^2. We consider the sampling algorithm for two choices of the function h: $h(t)=1$ corresponding to the normalized $1/2$-stable process (see Example 3.1 for details), and $h(t)=\mathrm{e}^{1-t}$ corresponding to the normalized inverse Gaussian process with parameter $b=1$ (see Example 3.2 for details). We ran the algorithm for 50,000, iterations keeping the final 9000 for estimating the sampling density function. The algorithm took a matter of seconds to run. To sample from the predictive at each iteration, we sample the weights P_j's. If the sample $d_{n+1}\leq N$, then we take the predictive sample from $\mathcal{N}(\mu_{d_{n+1}},\tau^2)$; if $d_{n+1}>N$, then we can take $\mu_{d_{n+1}}$ from the prior π and take y_{n+1} from $\mathcal{N}(\mu_{d_{n+1}},\tau^2)$. The predictive plot is presented in Figure 2 for the choice $h(t)=1$ and in Figure 3 for the choice $h(t)=\mathrm{e}^{1-t}$.

For the latter case, that is, for $h(t)=e^{1-t}$, we provide further information. The estimate of the total mass parameter is 1.70 and the convergence of the running mean throughout the chain is given in Figure 4. The chain was run for 100,000 iterations, and every 100th iteration was taken to construct the running average. In this case, the estimated correlation between V_1 and V_2 is 0.74.

Including and updating τ, if a prior is provided, is straightforward. Suppose the prior is $\pi(\lambda)$, where $\lambda=\tau^{-2}$, then

$$\pi(\lambda|\text{ rest }) \propto \pi(\lambda)\,\lambda^{n/2}\,\exp\left\{-0.5\lambda\sum_{i=1}^{n}(y(i)-\mu_{d(i)})^2\right\}.$$

Hence, a Gamma prior is most appropriate here and in the following example we take it to be a standard Exponential, that is, $\pi(\lambda)=\exp(-\lambda)$. Now, we simulated $n=50$ data points Y_i's from a mixture of two Gaussian distributions defined as follows: with probability $1/2$,

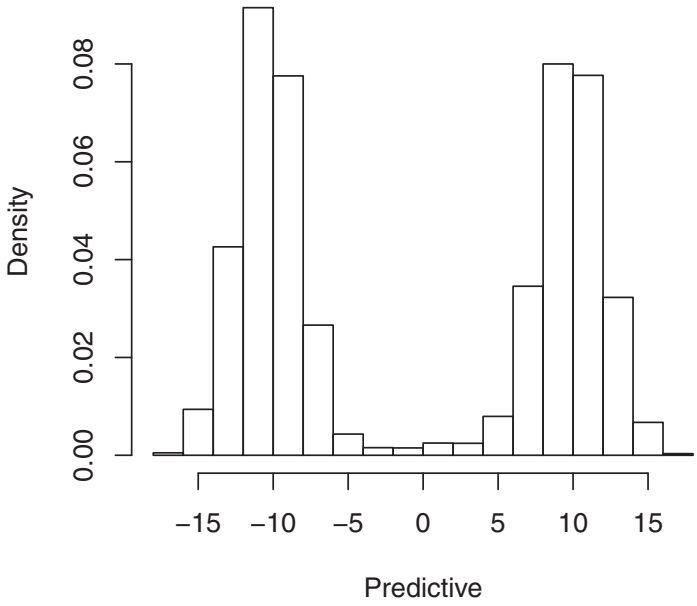

Figure 2. Mixture of Gaussian distributions, $Y_i \sim \frac{1}{2}\mathcal{N}(-10, 2^2) + \frac{1}{2}\mathcal{N}(10, 2^2)$: samples from the predictive density function for $h(t) = 1$.

$Y_i \sim \mathcal{N}(0, 1)$ and with probability $1/2$ $Y_i \sim \mathcal{N}(3, 0.5^2)$. The data are provided in Figure 5, and 10 posterior samples using the same MCMC sampler as described above were collected throughout the chain, separated by 1000 iterations. The samples are shown in Figure 6. These density functions were constructed using the weights, means, and variance, obtained at a particular iteration, and the notion of the density function being a mixture of Gaussian

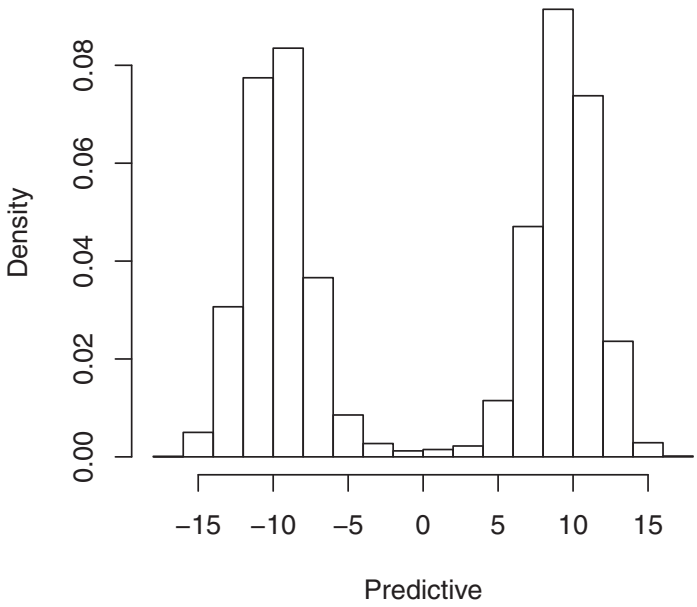

Figure 3. Mixture of Gaussian distributions, $Y_i \sim \frac{1}{2}\mathcal{N}(-10, 2^2) + \frac{1}{2}\mathcal{N}(10, 2^2)$: samples from the predictive density function for $h(t) = e^{1-t}$.

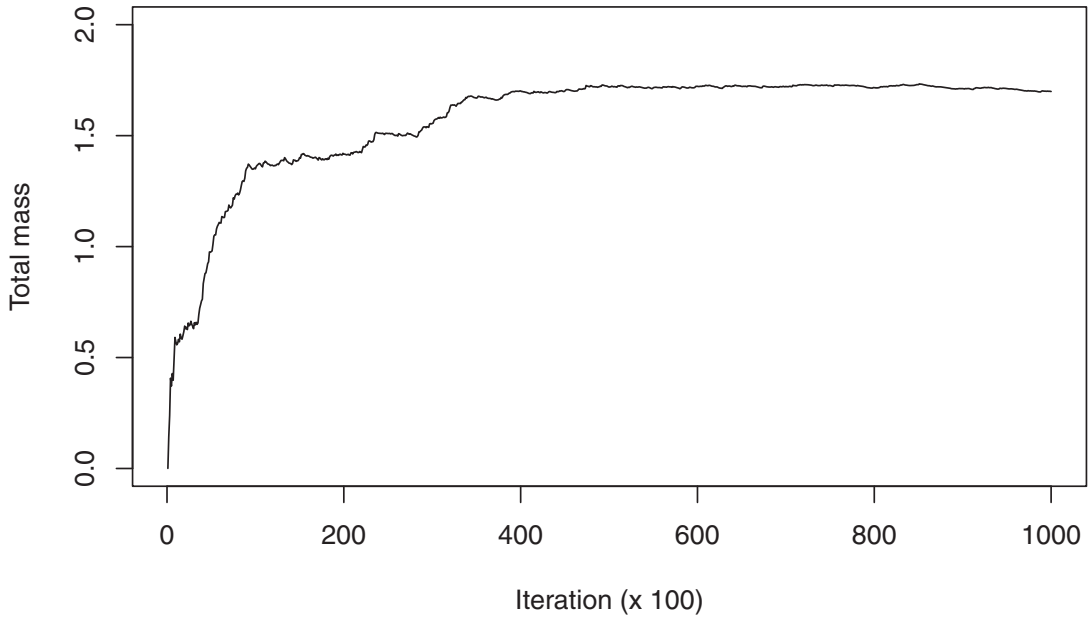

Figure 4. Mixture of Gaussian distributions, $Y_i \sim \frac{1}{2}\mathcal{N}(-10, 2^2) + \frac{1}{2}\mathcal{N}(10, 2^2)$: running average for total mass parameter with $h(t) = e^{1-t}$.

density functions. We computed the mixtures up to the value of N, and the remaining mass, typically very small, was added to a Gaussian density function with the prior mean and the sampled variance.

We also run the model, with prior on the variance term, with $\sigma = 1/2$ and $h(t) = e^{1-t}$ (see Example 3.2 for details) on the Galaxy dataset (see Roeder 1990), and which consists of 82 velocities of galaxies—an obligatory exercise when working with infinite mixture models.

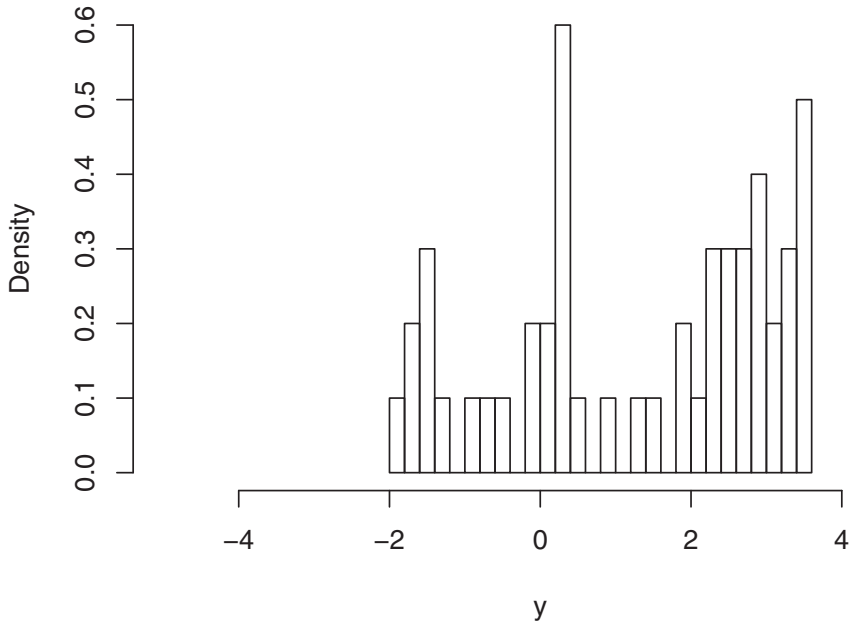

Figure 5. Mixture of Gaussian distributions, $Y_i \sim \frac{1}{2}\mathcal{N}(0, 1) + \frac{1}{2}\mathcal{N}(3, 0.5^2)$: data.

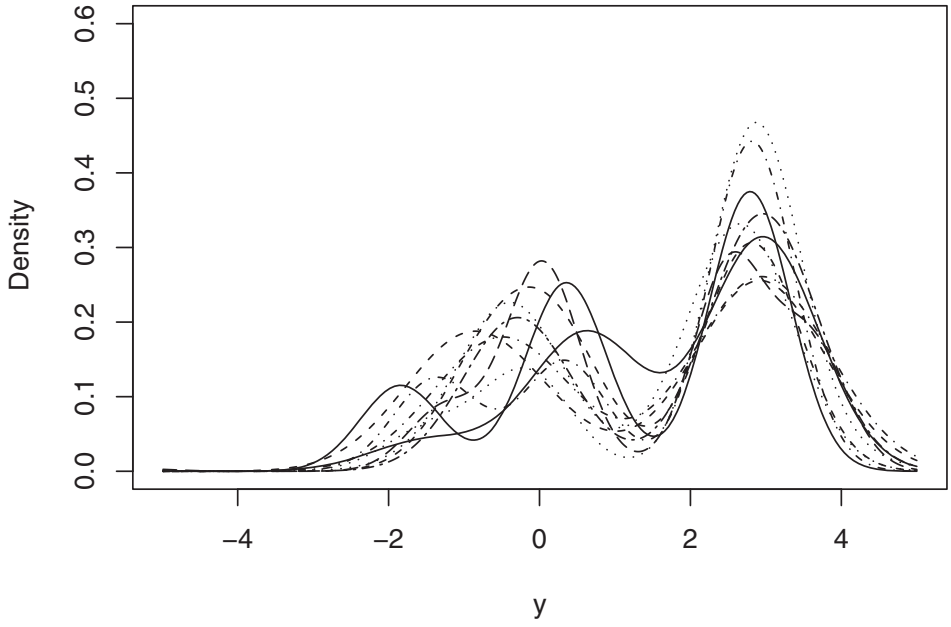

Figure 6. Mixture of Gaussian distributions, $Y_i \sim \frac{1}{2}\mathcal{N}(0, 1) + \frac{1}{2}\mathcal{N}(3, 0.5^2)$: posterior samples of densities.

We ran our model with 30,000 iterations and every 50th iteration was used to sample from the predictive. So, at the relevant iterations, we sample from the density function given all the parameters. The density function estimate is given in Figure 7. Note that we have transformed the data ($Y = Y/1000 - 20$) and the prior settings were precisely as those for the example of the mixture of two Gaussian distributions. In Figure 8, we provide five

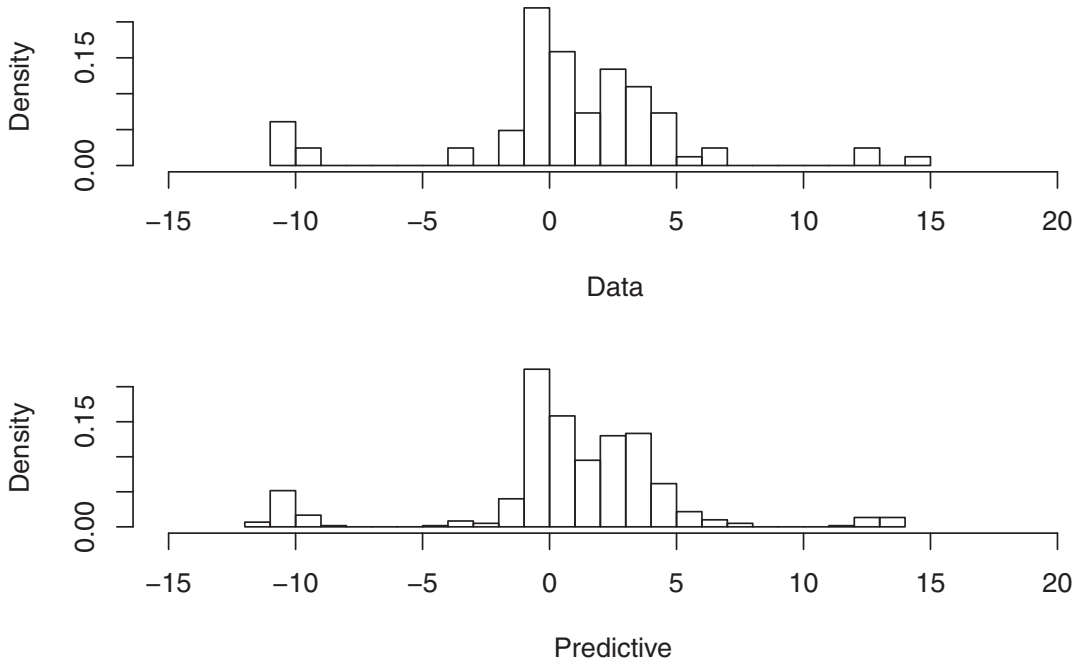

Figure 7. Predictive density for Galaxy dataset using $h(t) = e^{1-t}$.

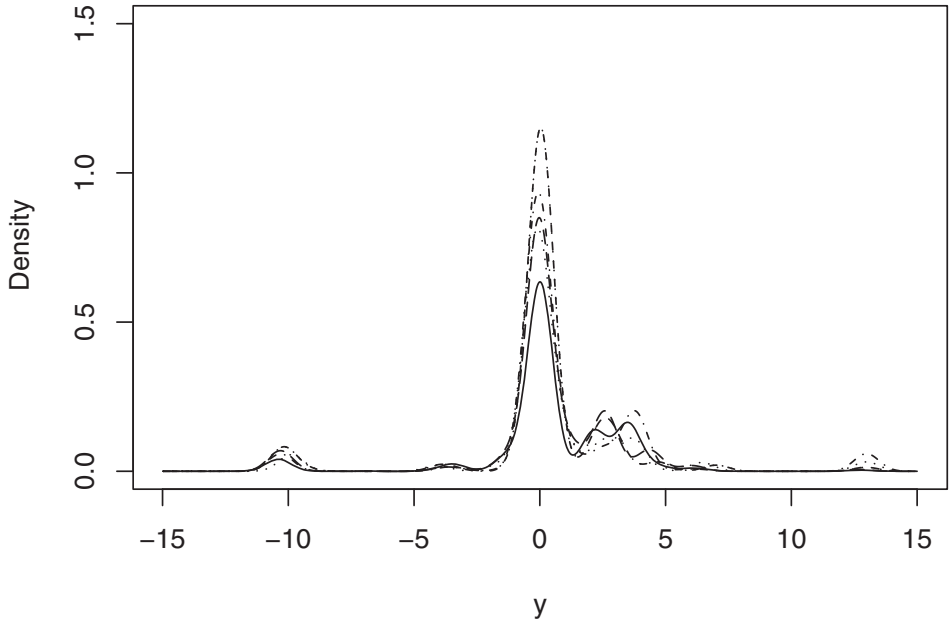

Figure 8. Posterior samples of densities for Galaxy dataset using $h(t) = e^{1-t}$.

posterior samples taken from the chain and separated by 1000 iterations. This was done in exactly the same manner as described in the previous example.

5. DISCUSSION

We have shown how to slice sample a class of mixture models, which includes all of the current popular choices of mixing measures. With "standard" stick-breaking models the stick-breaking variables $(V_j)_{j\geq 1}$ are independent, even as they appear in the full conditional distribution sampled in the posterior MCMC algorithm.

On the other hand, the models we are considering in this article have corresponding full conditional distribution given by Equation (19). This only leads to independent (v_j) variables in special cases, which depend on the choice of σ and $h(t)$. Nevertheless, we have shown how to sample this joint distribution and hence implement a valid MCMC algorithm. The joint density function arises as a consequence of using a σ-stable Poisson-Kingman based mixture model and specifically due to the representation of such random probability measures as stick-breaking processes, with not necessarily independent stick-breaking variables.

The present article is different from the work by Griffin and Walker (2011) in the following way. In the article by Griffin and Walker (2011), normalized random measure mixture models are handled using the normalized representation in terms of Lévy processes. Whereas in the present article, we remove the need to consider Lévy processes by using a representation of normalized random measures as stick-breaking processes with dependent stick-breaking variables. We argue this latter representation is in fact the simpler to deal with.

Finally, we note that the class of σ-stable Poisson-Kingman models is indexed by a function $h(t)$, in addition to what is usually encountered in discrete random probability measure. This function and its role need to be properly investigated. This might not have happened to date due to the lack of tools to do posterior inference with this class of model. However, this is now not the case.

SUPPLEMENTARY MATERIALS

The datasets and the Scilab code that implements the σ-stable Poisson-Kingman mixture models described in Section 4 are available as a zip file (slicePK.zip).

ACKNOWLEDGMENT

Detailed comments and suggestions by an Associate Editor and two Referees are gratefully acknowledged. Stefano Favaro is supported by grants from the Italian Ministry of University and Research and Regione Piemonte.

[Received June 2011. Revised February 2012.]

REFERENCES

Escobar, M. D. (1994), "Estimating Normal Means With a Dirichlet Process Prior," *Journal of the American Statistical Association*, 89, 268–277. [831]

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588. [831]

Favaro, S., Lijoi, A., and Prünster, I. (2012), "On the Stick-Breaking Representation of Normalized Inverse Gaussian Priors," *Biometrika*, 99, 663–674. [836]

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230. [831]

Gnedin, A., and Pitman, J. (2005), "Exchangeable Gibbs Partitions and Stirling Triangles," *Zapiski Nauchnych Seminarov POMI*, 325, 83–102. [834]

Griffin, J. E., and Walker, S. G. (2011), "Posterior Simulation of Normalized Random Measure Mixtures," *Journal of Computational and Graphical Statistics*, 20, 241–259. [831,845]

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96, 161–173. [831]

James, L. F., Lijoi, A., and Prünster, I. (2009), "Posterior Analysis of Normalized Random Measure With Independent Increments," *Scandinavian Journal of Statistics*, 36, 76–97. [833]

Kalli, M., Griffin, J. E., and Walker, S. G. (2011), "Slice Sampling Mixture Models," *Statistics and Computing*, 21, 93–105. [831,837]

Kanter, M. (1975), "Stable Densities Under Total Variation Inequalities," *Annals of Probability*, 3, 697–707. [838,839]

Kingman, J. F. C. (1967), "Completely Random Measures," *Pacific Journal of Mathematics*, 21, 59–78. [832]

——— (1975), "Random Discrete Distributions," *Journal of the Royal Statistical Society*, Series B, 37, 1–22. [831]

——— (1993), *Poisson Processes*, Oxford: Oxford University Press.

Lijoi, A., Mena, R. H., and Prünster, I. (2005), "Hierarchical Mixture Modelling With Normalized Inverse-Gaussian Priors," *Journal of the American Statistical Association*, 100, 1278–1291. [831]

Lijoi, A., Mena, R. H., and Prünster, I. (2007), "Controlling the Reinforcement in Bayesian Nonparametric Mixture Models," *Journal of the Royal Statistical Society*, Series B, 69, 715–740. [831]

Lijoi, A., and Prünster, I. (2010), "Models Beyond the Dirichlet Process," in *Bayesian Nonparametrics*, eds. N. L. Hjort, C. C. Holmes, P. Müller, and S. G. Walker, Cambridge: Cambridge University Press, pp. 80–136.

Lo, A. Y. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates," *The Annals of Statistics*, 12, 351–357. [830,831]

MacEachern, S. N. (1994), "Estimating Normal Means With a Conjugate Style Dirichlet Process Prior," *Communications in Statistics—Simulation and Computation*, 23, 727–741. [831]

——— (1998), "Computational Methods for Mixture of Dirichlet Process Models," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer-Verlag, pp. 23–44.

MacEachern, S. N., and Müller, P. (1998), "Estimating Mixture of Dirichlet Process Models," *Journal of Computational and Graphical Statistics*, 7, 223–238. [831]

Neal, R. (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249–265. [831]

Papaspiliopoulos, O., and Roberts, G. O. (2008), "Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models," *Biometrika*, 95, 169–186. [831]

Perman, M., Pitman, J., and Yor, M. (1992), "Size-Biased Sampling of Poisson Point Processes and Excursions," *Probability Theory and Related Fields*, 92, 21–39. [831,835,836]

Pitman, J. (1995), "Exchangeable and Partially Exchangeable Random Partitions," *Probability Theory and Related Fields*, 102, 145–158. [831]

——— (1996), "Random Discrete Distribution Invariant Under Sized-Biased Permutation," *Advances in Applied Probability*, 28, 525–539. [835,836]

——— (2003), "Poisson-Kingman Partitions," in *Science and Statistics: A Festschrift for Terry Speed,* ed. D. R. Goldstein, Beachwood, OH: Institute of Mathematical Statistics, pp. 1–34.

——— (2006), *Combinatorial Stochastic Processes: Ecole d'Eté de Probabilités de Saint-Flour XXXII*, New York: Springer.

Pitman, J., and Yor, M. (1997), "The Two Parameter Poisson-Dirichlet Distribution Derived From a Stable Subordinator," *Annals of Probability*, 27, 1870–1902. [831]

Roeder, K. (1990), "Density Estimation With Confidence Sets Exemplified by Superclusters and Voids in the Galaxies," *Journal of the American Statistical Association*, 85, 617–624. [843]

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650. [835]

Walker, S. G. (2007), "Sampling the Dirichlet Mixture Model with Slices," *Communications in Statistics—Simulation and Computation*, 36, 45–54. [831,837]