A conditional algorithm for Bayesian finite mixture models via normalized point process.

Un algoritmo per la stima bayesiana di misture finito dimensionali costruite mediante normalizzazione di processi di punto

Raffaele Argiento

Abstract Modelling via finite mixtures is one of the most fruitful Bayesian approach, particularly useful when there is unobserved heterogeneity in the data. The most popular algorithm under this model is the reversible jump MCMC, that can be nontrivial to design, especially in high-dimensional spaces. In this work, we first introduce a class of finite discrete random probability measures obtained by normalization of finite point processes. Then, we use the new class as the mixing measure of a mixture model and derive its posterior characterization. The resulting new class encompasses the popular finite Dirichlet mixture model; here, in order to compute posterior, we propose an alternative to the reversible jump. In particular, borrowing notation from the nonparametric Bayesian literature, we set up a conditional MCMC algorithm based on the posterior characterization of the unnormalized point process. In order to show the performance of our algorithm and the flexibility of the model, we illustrate some examples on the popular Galaxy dataset.

Abstract *La classe dei modelli mistura è frequentemente utilizzata come strumento per l'analisi di popolazioni eterogenee. Per ottenere delle stime bayesiane dei parametri di questi modelli, sono comunemente utilizzati gli algoritmi MCMC di tipo "Reversible Jump". Tuttavia, questi ultimi sono molto difficili da configurare, in special modo quando i dati appartengono a spazi di dimensione elevata. In questo lavoro, come primo passo, introdurremo una classe di misure di probabilità aleatorie. Tali misure saranno costruite come normalizzazione di processi di punto finito dimensionali di cui daremo una caratterizzazione a posteriori. Come secondo passo, utilizzeremo gli elementi della nuova classe come misure miscelanti in modelli mistura, generalizzando, così, la ben nota famiglia di misture di Dirichlet finito dimensionali. Proporremo un campionatore di tipo Gibbs in alternativa all'usuale algoritmo a salti reversibili. In particolare, prendendo in prestito la nomenclatura dalla letteratura bayesiana nonparametrica, costruiremo un agoritmo di tipo condizionale basandoci sulla caratterizzazione a posteriori del processo di punto finito*

Raffaele Argiento
CNR-IMATI, Via Bassini 15, Milano (Italy), and School of Mathematics, Statistics and Actuarial Science (SMSAS) at the University of Kent, Canterbury (UK), e-mail: raffaele@mi.imati.cnr.it

dimensionale non normalizzato. Per illustrare le prestazioni del nostro algoritmo e la flessibilità del modello, illustreremo due esempi di mistura considerando il popolare set di dati Galaxy.

Key words: Mixture Models, Reversible Jump MCMC, Point Processes, Density Estimation

1 Introduction

Mixture models are very powerful and natural statistical tools to fit data from an heterogeneous population. Under such a model each observation is assumed to have arisen from one of $0 < M \leq \infty$ groups. Let $Y \in \mathscr{Y} \subset \mathbb{R}^r$ be the population variable, a mixture model for Y is given by

$$f_Y(y;P) = \int_\Theta f(y;\theta)P(d\theta) = \sum_{j=1}^{M} P_j f(y;\tau_j), \tag{1}$$

where $\{f(\cdot;\theta), \theta \in \Theta \subset \mathbb{R}^d\}$ is a parametric family of densities on $\mathscr{Y}$, while $P(\cdot)$ is an almost sure discrete density on Θ, and it is called *mixing measure*.

For each $j = 1,\dots,M$, the density $f(y,\tau_j)$ is referred to as a kernel of the mixture, and is weighted by P_j, the relative frequency of the group in the population. Model (1) provides a framework by which observations may be clustered together into groups for discrimination or classification, so that conditionally to the groups data are independent and identically distributed within the groups and independent between groups.

In Bayesian Nonparametrics framework (i.e., $M = \infty$) the Dirichlet process mixture model ([8]), i.e. model (1) were the mixing measure is indeed the Dirichlet process, plays a pivotal role. This popularity is mainly due to its high flexibility in density estimation problems as well as in clustering and for its mathematical and computational tractability. However, in some statistical applications, clustering induced by the Dirichlet process may be restrictive (see for instance [2]), so that many alternative mixing measures have been proposed. In particular, [7] replaces the Dirichlet process with a huge and flexible class of random probability measures obtained by *normalization* of random (infinite dimensional) measures with independent increments. On the other hand, in a Bayesian parametric context (i.e., $M < \infty$ almost surely) the most natural approach is to choose a prior on M, then, conditionally to $M = m$, the mixture weights $(P_1,\dots,P_m)$ are chosen according to a $m-1$ dimensional $Dirichlet_m$ distribution. We denote the latter model the *finite Dirichlet mixture model* (FDMM) and refer, among the other, to Nobile [10], Richardson and Green [11], Stephens [13] and Miller and Harrison [9] for more details on it.

In this work we put ourselves in a parametric context and we introduce a new class of random measures obtained by normalization of a finite point processes. Then, as in [7] in the nonparametric case, we aim at using elements of the new

class as mixing measures in Model (1). We call the new class of random probability measures the class of *normalized independent finite point processes* and we derive the prior the family of prior distributions it induces on the data partition by giving a general formula for the corresponding exchangeable partition probability functions [12]. As a further result, we characterize the posterior distribution of a normalized independent finite point process given a sample from it.

It is well known that a finite Dirichlet vector can be obtained by normalization of Gamma random variables with common scale parameter, so that our normalized class encompasses the finite Dirichlet process; we specialized our theoretical result for the Dirichlet measure when the number of jumps is chosen according to a shifted Poisson on the positive integers $1,2,\ldots$. We refer to this simple process as the *normalized finite Poisson-Dirichlet* measure and set up the relative mixture model.

Several inference methods have been proposed for the finite Dirichlet mixture model, the most commonly-used method being reversible jump Markov chain Monte Carlo ([11]). Reversible jump is a very general technique, and has been successfully applied in many contexts, but it can be difficult to use since applying it to new situations requires one to design good reversible jump moves, which is often nontrivial, particularly in high-dimensional parameter spaces.

Among the main achievements of this work, there is the construction of (conjugate) Gibbs sampler scheme to simulate from the posterior distribution of a normalize finite Poisson-Dirichlet; in particular, borrowing notation from the nonparametric Bayesian literature, we set up a conditional MCMC algorithm based on the posterior characterization of the unnormalized point process.

For illustration purposes, we apply the finite Poisson-Dirichlet model to a popular dataset, namely the Galaxy dataset, since it is nowadays the favorite test dataset for any new nonparametric model in a density estimation context. We illustrate the performances of our algorithm for different sets of hyperparameters and compare our results with the one obtained under the reversible jump MCMC.

2 The General Finite Point Process

Let $\mathscr{X}$ be a complete separable metric space, a *finite point process* X is a random countable subset of $\mathscr{X}$. In this paper attention is restricted to processes whose realizations are finite subset of $\mathscr{X}$. Formally, for any subset $x \subseteq \mathscr{X}$, let $m(x)$ denote the cardinality of x, realization of X are constrained on $N_f = \{x \subseteq \mathscr{X} : m(x) < \infty\}$. Elements of N_f are called *finite point configurations*.

We refer to the book of Daley and Vere-Jones ([4]) for a complete treatment of finite point processes. We mention here that the law of a finite point process is identified given the following:

1. *A discrete probability density* $\{q_m, m = 0,1,\ldots\}$ *determining the law of the total number* M *of points of the process,*

2. *for each integer $m \geq 1$, a probability distribution $\Pi_m(\cdot)$ on the Borel sets of $\mathscr{X}^m$, that determines the joint distribution of the positions of the points of the process, given that their total number is m.*

In particular $\{q_m\}$ and $\{\Pi_m\}$ provide a constructive definition that could be used to simulate the process: first, generate a random number M according to the distribution $\{q_m\}$, and then, supposing $M = m$ and, excepting the case $m = 0$ in which there is nothing else to do, generate a random set $X = \{\xi_1, \dots, \xi_m\}$, where $(\xi_1, \dots, \xi_m)$ is distributed according to $\Pi_m(\cdot)$.

We point out that, a point process $X = \{\xi_1, \dots, \xi_m\}$ is a set of unordered points. This fact, is implicit in Conditions 1. and 2. above: to be consistent with treating point processes as unordered sets, one as to stipulate that the distributions $\Pi_m(\cdot)$ should give equal weight to all $m!$ permutations of the elements in the vector $(\xi_1, \dots, \xi_n)$, i.e. $\Pi_m(\cdot)$ should be symmetric. If this is not already the case, it is easily achieved by introducing the symmetrized form for any partition $(A_1, \dots, A_m)$ of $\mathscr{X}$: $\Pi_m^{sym}(A_1 \times \cdots \times A_m) = \frac{1}{m!} \sum_{perm} \Pi_m(A_{i_1} \times \cdots \times A_{i_m})$, where the summation *perm* is taken over all $m!$ permutations $(i_1, \dots, i_m)$ of the integers $(1, \dots, m)$. An alternative notation to identify the law of X, which has some advantages in simplifying combinatorial formulae, utilizes the nonprobability measures

$$J_m(A_1 \times \cdots \times A_m) = q_m \sum_{perm} \Pi_m(A_{i_1} \times \cdots \times A_{i_m}) = m! q_m \Pi_m^{sym}(A_1 \times \cdots \times A_m).$$

This latter is referred as *Janossy measure* ([4]). The Jannossy measure has a simple interpretation. If $\mathscr{X} = \mathbb{R}^d$ and $j_m(\xi_1, \dots, \xi_m)$ denotes the density of $J_m(\cdot)$ with respect to Lesbegue measure and $\xi_i \neq \xi_j$ for $i \neq j$, then $j_m(\xi_1, \dots, \xi_m) d\xi_1 \dots d\xi_m = \mathbb{P}($ *there are exactly m points in the process, one in each of the distinct infinitesimal regions* $(\xi_i, \xi_i + d\xi_i))$. Janossy densities play a fundamental role in the study of finite point processes and spatial point patterns, we refer to [4] for more details. Here, we will use this mathematical object to characterize the posterior distribution of a new class of finite discrete random probability measures. In fact, in this work we will deal with a simple family of finite point processes, namely the family of *independent finite point processes* (IFPP):

Definition 1. Let $\nu(\cdot)$ be a density on $\mathscr{X}$, and $\{p_m, m = 0, 1, \dots\}$ a probability density, we will say that X is an *independent finite point process*, or $X \sim IFPP(\nu, p_m)$, if its Janossy density can be written as

$$j_m(\xi_1, \dots, \xi_m) = m! p_m \prod_{j=1}^{m} \nu(x_j).$$

3 Normalized independent finite point processes

Let $\Theta \subset \mathbb{R}^d$, for some positive integer d and let $\mathscr{X}$ be $\mathbb{R}^+ \times \Theta$. We denote with $\xi = (s, \tau)$ a point of $\mathscr{X}$. Let $\nu(s, \tau)$ be a density on $\mathscr{X}$ so that $\nu(s, \tau) = h(s) p_0(\tau)$,

where $h(\cdot)$ is a density on $\mathbb{R}^+$ and $p_0(\cdot)$ is a density on Θ. Finally, let $\{q_m, m = 0,1,\dots\}$ be a discrete density on the natural number such that $q_0 = 0$. We consider the independent finite point process $\tilde{P} = \{(S_1,\tau_1),\dots,(S_m,\tau_M)\}$ with parameters ν and $\{q_m\}$, i.e. $\tilde{P} \sim IFPP(\nu, q_m, p_0)$. Let $\mathscr{J} := \{1,\dots,M\}$ be the set of point's indexes, since we are assuming $q_0 = 0$, the random variable $T := \sum_{j\in\mathscr{J}} S_j$ is almost surely different from 0, so that we can give the following

Definition 2. Let $\tilde{P} \sim IFPP(h, q_m, p_0)$, with $q_0 = 0$. A normalized independent finite point process (Norm-IFPP) with parameter h and $\{q_m\}$ is the discrete probability measure on Θ defined by

$$P(\cdot) = \sum_{j\in\mathscr{J}} P_j \delta_{\tau_j}(\cdot) \stackrel{d}{=} \sum_{j\in\mathscr{J}} \frac{S_j}{T} \delta_{\tau_j}(\cdot), \tag{2}$$

where $T = \sum_{j\in\mathscr{J}} S_j$. We will use the notation $P \sim Norm - IFPP(h, q_m, p_0)$.

The finite dimensional processes defined in (2) belong to the wide class of species sampling models, deeply investigated in [12], and we use some of the results there to derive ours. Let $(\theta_1,\dots,\theta_n)$ be a sample from a Norm-IFPP (or more generally, from a species sampling model); since it is a sample from a discrete probability, it induces a random partition $\boldsymbol{\rho}_n := \{C_1,\dots,C_k\}$ on the set $\mathbb{N}_n := \{1,\dots,n\}$ where $C_j = \{i : \theta_i = \theta_j^*\}$ for $j = 1,\dots,k$. If $\#C_i = n_i$ for $1 \le i \le k$, the marginal law of $(\theta_1,\dots,\theta_n)$ has unique characterization: $\mathscr{L}(\boldsymbol{\rho}_n, \theta_1^*,\dots,\theta_k^*) = p(n_1,\dots,n_k)\prod_{j=1}^k \mathscr{L}(\theta_j^*)$, where p is the exchangeable partition probability function (eppf) associated to the random probability. The eppf p is a probability law on the set of the partitions of $\mathbb{N}_n$. The following proposition provides an expression for the eppf of a Norm-HFPP measure.

Proposition 1. *Let $(n_1,\dots,n_k)$ be a vector of positive integers such that $\sum_{i=1}^k n_i = n$. Then, the eppf associated with a Norm-IFPP$(h,\{q_n\},p_0)$ is*

$$\pi(n_1,\dots,n_k) = \int_0^{+\infty} \frac{u^{n-1}}{\Gamma(n)} \left\{ \sum_{m=0}^{\infty} \frac{(m+k)!}{m!} \psi(u)^m q_{m+k} \right\} \prod_{i=1}^k \kappa(n_i,u)du$$

where $\psi(u)$ is the Laplace transform of the density $h(s)$, i.e.

$$\psi(u) := \int_0^{\infty} e^{-us}h(s)ds, \quad \textit{and} \quad \kappa(n_j,u) := \int_0^{\infty} u^{n_i} e^{-us}h(s)ds = (-1)^{n_j}\frac{d}{du^{n_j}}\psi(u).$$

We denote by $\mathscr{J}_a$ the indexes of *allocated* jumps of the process, i.e., the indexes $j \in \mathscr{J}$ corresponding to some S_j such that there exists a location for which $\tau_j = \theta_i^*$, $i = 1,\dots,k$. The remaining values are *non-allocated* jumps. We use the superscript (na) for random variables related to *non-allocated* jumps. We also introduce the random variable $U := \Gamma_n/T$, where $\Gamma_n \sim gamma(n,1)$, being Γ_n and T independent.

Proposition 2. *If P is an Norm-IFPP$(h,\{q_n\},p_0)$, then the unnormalized process $\tilde{P}$, given $\boldsymbol{\theta}^* = (\theta_1^*,\dots,\theta_k^*)$, $\boldsymbol{n} = (n_1,\dots,n_k)$ and $U = u$, is the superposition of two processes $\tilde{P} \stackrel{d}{=} \tilde{P}^{(na)} \cup \tilde{P}^{(a)}$ where*

1. *The process of non-allocated jumps $\tilde{P}^{(na)}$, is an independent finite point process with Janossy density given by*

$$j_m((s_1,\tau_1),\dots,(s_m,\tau_m)) = m!p^\star_m \prod_{j=1}^m h^\star(s_j)p_0(\tau_j),$$

   where, denoting with $\psi(u)$ is the Laplace transform of h, $h^\star_u(s) \propto \mathrm{e}^{-us}h(s)$ and $q^\star_m \propto \frac{(m+k)!}{m!}\psi(u)^m q_{m+k}$, $m = 0,1,2,\dots$.
2. *The process of allocated jumps $\tilde{P}^{(a)}$ is the unordered set of the points $(S_1,\tau_1),\dots,(S_k,\tau_k)$, such that, for $j = 1,\dots,k$, $\tau_j = \theta^*_j$ and the distribution of S_j is proportional to $s^{n_j}\mathrm{e}^{-us}h(s)$.*
3. *Conditionally to $\mathscr{J}^{(a)}$, $\tilde{P}^{(a)}$ and $\tilde{P}^{(na)}$ are independent.*
4. *the posterior law of U given $\boldsymbol{\theta}^*$ has density on the positive real given by*

$$f_{U|\boldsymbol{\theta}^*}(u|\boldsymbol{\theta}^*) \propto \frac{u^{n-1}}{\Gamma(n)}\left\{\sum_{m=0}^\infty \frac{(m+k)!}{m!}\psi(u)^m q_{m+k}\right\}\prod_{i=1}^k \kappa(n_i,u)$$

4 A simple mixture model

Let h to be the density of a $gamma(\gamma,1)$ distribution (being γ the shape parameter) and let $q_m = e^{-\Lambda}\Lambda^{m-1}/(m-1)!$ $(m = 1,2,\dots)$ be the density of a shifted Poisson distribution. It is simple to realize that under this choice of h and $\{q_m\}$ the Norm-IFPP is a finite Dirichlet measure, that is an almost sure discrete probability measure as in (2), where conditionally to $M = m > 0$ the jumps size $(P_1,\dots,P_m)$ of P are a sample from the $(m-1)$-dimensional Dirichlet$_m(\gamma,\dots,\gamma)$ distribution. We are going to call this measure a *normalized finite Poisson-Dirichlet* (NFPD) process with parameters $\gamma > 0$, $\Lambda > 0$ and $p_0(\cdot)$; We will use the notation $P \sim NFPD(\gamma,\Lambda,p_0)$. Observe that the Laplace transform and its derivatives for a $gamma(\gamma,1)$ density are given by $\psi(u) = \frac{1}{(u+1)^\alpha}$, and $\kappa(n,u) = \frac{1}{(u+1)^{n_j+\alpha}}\frac{\Gamma(\gamma+n)}{\Gamma(\gamma)}$, $u > 0$, $n = 1,2,\dots$, so that, by applying Proposition 1, we obtain that the eppf of a NFPD(γ,Λ,p_0) is given by

$$p(n_1,\dots,n_k) = V(n,k)\prod_{j=1}^k \Lambda\frac{\Gamma(\gamma+n_j)}{\Gamma(\gamma)},$$

where $V(n,k) = \int_0^\infty \frac{u^{n-1}}{\Gamma(n)}\frac{\Lambda+k(u+1)^\gamma}{\Lambda(u+1)^{n+\gamma(k+1)}}\mathrm{e}^{\frac{\Lambda-(u+1)^\gamma}{(u+1)^\gamma}}du$.

Let $Y_1,\dots,Y_n$ a set of data in an Euclidean space $\mathscr{Y}$, we consider the following mixture model

$$\begin{aligned} &Y_1,\ldots,Y_n|\theta_1,\ldots,\theta_n \overset{ind}{\sim} f(y;\theta_i) \\ &\theta_1,\ldots,\theta_n|P \overset{iid}{\sim} P \\ &P \sim NFPD(\gamma,\Lambda,p_0). \end{aligned} \tag{3}$$

where $f(\cdot;\theta_i)$ is a parametric density on $\mathscr{Y}$, for all $\theta \in \Theta \subset \mathbb{R}^d$. We point out that p_0 is the density of a non-atomic probability measure P_0 on Θ, such that $\mathbb{E}(P(A)) = P_0(A)$ for all $A \in \mathscr{B}(\Theta)$. Model (3) will be addressed here as *NFPD hierarchical mixture* model. It is well known that this model is equivalent to assume that the Y_i's, conditionally on P, are independently distributed according to the random density (1). We point out that model (3) is equivalent to the popular finite Dirichlet mixture model (see [10, 11, 13, 9]). Essentially, conditionally to $M = m > 0$ we can think at the jump size $(P_1,\ldots,P_m)$ of the mixing measure P in (2) as a sample from the $(m-1)$-dimensional Dirichlet$_m(\gamma,\ldots,\gamma)$ distribution.

Thanks to the posterior characterization given in Proposition 2, we are able to build a blocked Gibbs sampler to update blocks of parameters, which are drawn from multivariate distributions. The Gibbs sampler we propose here, has the same structure of that ones provided in [3] and [1] for the finite dimensional approximation of Normalized Generalized gamma and Normalized Completely random mixtures, respectively. In particular the parameter of our Gibbs sampler is $(\tilde{P},\boldsymbol{\theta},U)$, where $\tilde{P}$ is the unnormalized finite point process and U is an auxiliary augmenting variable prescribed by the posterior characterization in Proposition 2.

Description of the full-conditionals is below.

1. **Sampling from** $\mathscr{L}(U|\boldsymbol{Y},\boldsymbol{\theta},P)$: By its construction, conditionally on $\tilde{P}$, the random variable U is distributed as gamma with parameters (n,T).
2. **Sampling from** $\mathscr{L}(\boldsymbol{\theta}|u,\boldsymbol{Y},\tilde{P})$: each θ_i, for $i = 1,\ldots,n$, has discrete law with support $\{\tau_1,\ldots,\tau_M\}$, and probabilities $\mathbb{P}(\theta_i = \tau_j) \propto S_j f(Y_i;\tau_j)$.
3. **Sampling from** $\mathscr{L}(\tilde{P}|u,\boldsymbol{\theta},\boldsymbol{Y})$: first we observe that conditionally on $\boldsymbol{\theta}$ the process $\tilde{P}$ does not depend on the data $\boldsymbol{Y}$. So that we have to sample from $\mathscr{L}(\tilde{P}|u,\boldsymbol{\theta})$; then, thanks to Proposition 2 we can split this step into two substeps:

   3.a **Sampling from** $\mathscr{L}(\tilde{P}^{(a)}|u,\boldsymbol{\theta},\boldsymbol{Y})$: the allocated process is a set of independent pairs of vaiable $((S_1^{(a)},\tau_1^{(a)}),\ldots(S_k^{(a)},\tau_k^{(a)})$ such that, for $j = 1,\ldots,k$:

   $$S_j^{(a)} \sim gamma(n_j+\gamma,u+1) \quad \text{independent from} \quad \tau_j^{(a)} \sim \prod_{i\in C_j} f(y_i,\tau_j)p_0(\tau_j)$$

   where C_j is the cluster of data identified by τ_j.

   3.b **Sampling from** $\mathscr{L}(\tilde{P}^{(na)}|,u,\boldsymbol{\theta},\boldsymbol{Y})$: the process of non allocate jumps is a independent finite process with Janossy density:

   $$j_m((s_1,\tau_1),\ldots,(s_m,\tau_m)) = m!p_m^\star \prod_{j=1}^m h^\star(s_j)p_0(\tau_j),$$

where, $h^{\star}_u$ is the density of a $gamma(\gamma, u+1)$ and p_m is the discrete density of the following two-components mixture

$$\frac{(u+1)^{\gamma}k}{(u+1)^{\gamma}k+\Lambda}\mathscr{P}_0(\Lambda/(u+1)^{\gamma}) + \frac{\Lambda}{(u+1)^{\gamma}k+\Lambda}\mathscr{P}_0(\Lambda/(u+1)^{\gamma}),$$

where $\mathscr{P}_i(\lambda)$ the shifted Poisson distribution on $\{i, i+1, i+2, \ldots\}$ with mean $i+\lambda$, $i = 0, 1$;

We point out that, if the density p_0 and the family of kernel $f(y;\theta)$ are conjugate, then all the full conditionals of the latter algorithm belong to standard distributions, so that its implementation is quite straightforward.

5 Galaxy data

This super-popular dataset contains $n = 82$ measured velocities of different galaxies from six well-separated conic sections of space. Values are expressed in Km/s, scaled by a factor of 10^{-3}. We report some of the posterior estimates (mainly density and number of components estimates) for two different sets of hyperparameters of model (3) when $f(\cdot;\theta)$ is the Gaussian density on $\mathbb{R}$ and $\theta = (\mu, \sigma^2)$ stands for its mean and variance, and $P_0(d\mu, d\sigma^2) = \mathscr{N}(d\mu; m_0, \sigma^2/\kappa_0) \times inv-gamma(d\sigma^2; a, b)$; here $\mathscr{N}(m_0, \sigma^2/\kappa_0)$ is the Gaussian distribution with m_0 mean and σ^2/κ_0 variance, while $inv-gamma(a,b)$ is the inverse-gamma distribution with mean $b/(a-1)$ (if $a > 1$). We have fixed $m_0 = \bar{x}_n = 20.8315$, $\kappa_0 = 0.01$, $a = 2$, $b = 1$. In the fist set of experiments we aim at comparing the performances of our algorithm with the reversible jump sampler of Richardson and Green [11]. In particular, we have used the `mixAK` R-package ([6]). This package implements a reversible jump MCMC algorithm for univariate data in C++ and link it to R. As far as implementation of our algorithm is concerned, we implemented a C code, and linked it to R for post processing. For each experiment we have ran, we have fixed 5000 iteration of burn-inn, a thinning of 10, and a final sample size of 5000. We have fixed the hyperparameters γ and Λ in (3) in such a way that $\mathbb{E}(K_n) = 6$ (this mean was computed via a Monte Carlo method); in particular we have chosen $(\gamma, \Lambda) \in \{(1000, 0.0013), (100, 0.0136), (10, 0.21), (5, 5)\}$.

Figure 1 shows the density estimates (posterior mean of the random density "parameter") under the different values of the hyperparameters: all the estimates are quite similar and detect the "right" number of clusters. In order to compare the two algorithms, we computed the integrated autocorrelation time τ for the number M of components of the mixture (1) under the four experiments; τ controls the accuracy of Monte Carlo estimates computed using the MCMC chain and provides a measure of the efficiency of the method. The same indexes have been also used in [1] and [5] to assess the performance of their methods. We refer to the latter two papers for details on how to compute the integrated autocorrelation time and mention here that a small value of τ implies good mixing and hence an effi-

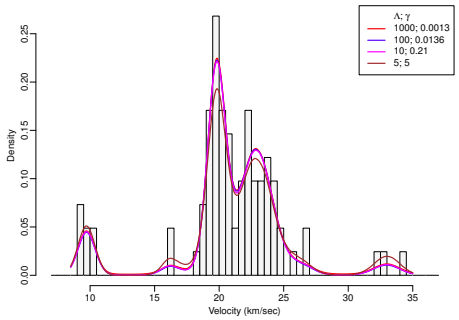

Fig. 1 Density estimation under the firs set of hyperparameters with different values of Λ and γ.

cient method. Result are summarized in in Table 1, in terms of running time ours Gibbs sampler and the reversible jump are comparable, but our algorithm strongly outperforms the reversible jump in term of index τ. In the second experiment we

(Λ,γ)	time Gibbs	$\mathbb{E}(M\|data)$ Gibbs	$\hat{\tau}$ Gibbs	time RJ	$\mathbb{E}(M\|data)$ RJ	$\hat{\tau}$ RJ
(1000,0.0013)	15.13 min.	1003.47	1.53	22.69 min.	669.33	864.44
(100, 0.0136)	1.51 min.	103.19	1.51	2.12 min.	98.16	138.40
(10, 0.21)	12.50 sec.	13.18	1.33	12.03 sec.	10.31	3.45
(5,5)	9.60 sec.	9.34	22.26	9.25 sec.	7.10	6.29

Table 1 Running times, posterior mean of M and estimation of integrated autocorrelation times $\hat{\tau}$ under the Gibbs sampler of Section 4 and the Reversible Jump MCMC implemented in the R-package `mixAK`.

have added a level of hierarchy to model (3) by choosing prior distributions for both γ and Λ, we considered $\gamma \sim gamma(a_1, b_1)$ and $\Lambda \sim gamma(a_2, b_2)$. We have updated the Gibbs sampler illustrated at the end of Section 4, by computing the full conditionals of this two parameters. In particular, from one hand $\Lambda|rest \sim \frac{\psi(u)}{1+b_2}\ gamma(k+a_2+1, 1-\psi(u)+b_2)\ \frac{1-\psi(u)+b_2}{1+b_2} gamma(k+a_2, 1-\psi(u)+b_2)$, where $\psi(u) = \frac{1}{(u+1)^\gamma}$ is the Laplace transform of a $gamma(\gamma,1)$ density; on the other, $\pi(\gamma|rest) \propto (\Lambda\psi(u)+k)\, e^{\Lambda\psi(u)} \frac{1}{\psi(u)^k} \prod_{j=1}^{k} \frac{\Gamma(\gamma+n_j)}{\Gamma(\gamma)}$, and we have to resort to a Metropolis-Hasting step to sample from this non standard full conditional. We set $a_1 = 0.01$ and $b_1 = 0.01$ resulting in vague prior on the parameter Λ, while we was more informative on γ by choosing $a_2 = 2$ and $b_2 = 1$. The running time under this setting was quite short: 8.287 sec.; this is justified by the fact that the number of non allocated jumps was small $\mathbb{E}(M^{(na)}|data) = 0.86$. The integrated autocorrelation time is 3.98, while the posterior distribution the autocorrelation plot and the scatter plot for the parameter Λ and γ are reported in Figure 2.

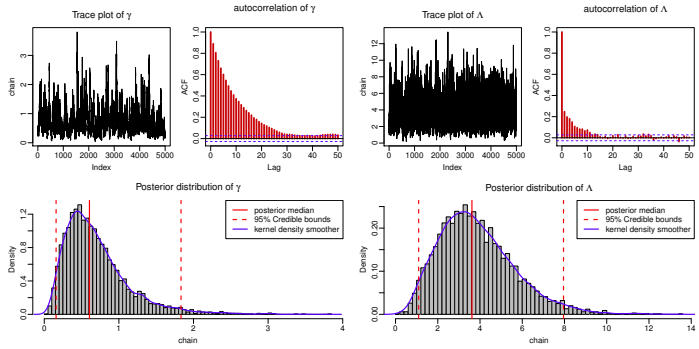

Fig. 2 Trace plots, autocorrelations and posterior distributions, of γ and Λ

References

1. Argiento, R., Bianchini, I. Guglielmi, A.: A blocked Gibbs sampler for NGG-mixture models via a priori truncation. Stat. Comput. **26(3)**, pp 641–661 (2016)
2. Argiento, R., Guglielmi, A., Hsiao, C.K., Ruggeri, F., Wang C.: Modelling the association between clusters of SNPs and disease responses. In: Nonparametric Bayesian Methods in Biostatistics and Bioinformatics (R. Mitra, P. Mueller Eds.), pp. 115-134, Springer, international publishing (2015)
3. Argiento, R., Bianchini, I., Guglielmi, A: A priori truncation method for posterior sampling from homogeneous normalized completely random measure mixture models. arXiv preprint arXiv:1507.04528. (2015)
4. Daley, D.J., Vere-Jones D.: An Introduction to the Theory of Point Processes Volume I: Elementary Theory and Methods (Second Edition). Springer-Verlag New York (2003)
5. Kalli, M., Griffin, J.E., Walker, S.G.: Slice sampling mixture models. Stat. Comput. **21(1)**, pp. 93-105 (2011)
6. Komárek, A.: A new R package for Bayesian estimation of multivariate normal mixtures allowing for selection of the number of components and interval-censored data. Comput. Stat. Data An., **53(12)**, pp. 3932–3947, (2009)
7. Lijoi, A., Mena, R. H. and Prünster, I.: Controlling the reinforcement in Bayesian nonparametric mixture models. J. R. Stat. Soc. Series B Stat Methodol. **69(4)** 715-740 (2007)
8. Lo, A. J.: On a class of Bayesian nonparametric estimates: I. density estimates. Ann. Stat. **12(1)**, 351–357 (1984)
9. Miller, J. W. and Harrison M. T.: Mixture models with a prior on the number of components. arXiv preprint arXiv:1502.06241 (2015)
10. Nobile, A.: Bayesian Analysis of Finite Mixture Distributions. PhD thesis, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, (1994)
11. Richardson, S. and Green, P. J.: On Bayesian analysis of mixtures with an unknown number of components. J. R. Stat. Soc. Series B Stat Methodol. **59(4)**, 731–792 (1997)
12. Pitman, J.: Some developments of the Blackwell-Macqueen urn scheme. In T. S. Ferguson, L. S. Shapley, and M. J. B. (Eds.), Statistics, Probability and Game Theory: Papers in Honor of David Blackwell, Volume 30 of IMS Lecture Notes-Monograph Series, pp. 245–267. Hayward (USA), Institute of Mathematical Statistics, (1996)
13. Stephens, M.: Bayesian analysis of mixture models with an unknown number of components – An alternative to reversible jump methods. Ann. Stat. **28(1)** 40–74 (2000)