A hybrid sampler for Poisson-Kingman mixture models

Authored by:

Yee Whye Teh
Maria Lomeli
Stefano Favaro

Abstract

This paper concerns the introduction of a new Markov Chain Monte Carlo scheme for posterior sampling in Bayesian nonparametric mixture models with priors that belong to the general Poisson-Kingman class. We present a novel and compact way of representing the infinite dimensional component of the model such that while explicitly representing this infinite component it has less memory and storage requirements than previous MCMC schemes. We describe comparative simulation results demonstrating the efficacy of the proposed MCMC algorithm against existing marginal and conditional MCMC samplers.

1 Paper Body

According to Ghahramani [9], models that have a nonparametric component give us more flexiblity that could lead to better predictive performance. This is because their capacity to learn does not saturate hence their predictions should continue to improve as we get more and more data. Furthermore, we are able to fully consider our uncertainty about predictions thanks to the Bayesian paradigm. However, a major impediment to the widespread use of Bayesian nonparametric models is the problem of inference. Over the years, many MCMC methods have been proposed to perform inference which usually rely on a tailored representation of the underlying process [5, 4, 18, 20, 28, 6]. This is an active research area since dealing with this infinite dimensional component forbids the direct use of standard simulation-based methods for posterior inference. These methods usually require a finite-dimensional representation. There are two main sampling approaches to facilitate simulation in the case of Bayesian nonparametric models: random truncation and marginalization. These two schemes are known in the literature as conditional and marginal samplers. In conditional samplers, the infinite-dimensional prior is replaced by a finite-dimensional representation chosen according to a truncation level. In marginal samplers, the need to represent the infinite-dimensional component can be bypassed by marginalising it out. Marginal samplers have less storage requirements than conditional samplers but could potentially have worst mixing properties. However, not integrating out the infinite dimensional compnent leads to a more comprehensive representation of the random probability measure, useful to compute expectations of interest with respect to the posterior. In this paper, we propose a novel MCMC sampler for Poisson-Kingman mixture models, a very large class of Bayesian nonparametric mixture models that encompass all previously explored ones in the literature. Our approach is based on a hybrid scheme that combines the main strengths of 1

both conditional and marginal samplers. In the flavour of probabilistic programming, we view our contribution as a step towards wider usage of flexible Bayesian nonparametric models, as it allows automated inference in probabilistic programs built out of a wide variety of Bayesian nonparametric building blocks.

2

Poisson-Kingman processes

Poisson-Kingman random probability measures (RPMs) have been introduced in Pitman [23] as a generalization of homogeneous Normalized Random Measures (NRMs) [25, 13]. Let X be a complete and separable metric space endowed with the Borel ?-field BpXq, let ? ? CRMp?, H0 q be a homogeneous Completely Random Measure (CRM) with L?evy measure ? and base distribution H0 , see Kingman [15] for a good overview about CRMs and references therein. Then, the corresponding total mass of ? is T ? ?pXq and let it be finite, positive almost surely, and absolutely continuous with respect to Lebesgue measure. For any t P R‘ , let us consider the conditional distribution of ?{t given that the total mass T P dt. This distribution is denoted by PKp?, ?t , H0 q, it is the distribution of a RPM, where ?t denotes the usual Dirac delta function. Poisson-Kingman RPMs form a class of RPMs whose distributions are obtained by mixing PKp?, ?t , H0 q, over t, with respect to some distribution ? on the positive real line. Specifically, a Poisson-Kingman RPM has following the hierarchical representation T ?? P —T ? t ? PKp?, ?t , H0 q.

(1)

The RPM P is referred to as the Poisson-Kingman RPM with L?evy measure ?, base distribution H0 and mixing distribution ?. Throughout the paper we denote by PKp?, ?, H0 q the distribution of P and, without loss of generality, we will assume that ?pdtq9hptqf? ptqdt where f? is the density of the total mass T under the CRM and h is a non-negative function. Note that, when ?pdtq ? ? f? ptqdt then the distribution PKp?, f? , H0 q coincides with NRMp?, H0 q. The resulting P ? k?1 pk ??k is almost surely discrete and since ? is homogeneous, the atoms p?k qk?1 of P are independent of their masses ppk qk?1 and form a sequence of independent random variables identically distributed according to H0 . Finally, the masses of P have distribution governed by the L?evy measure ? and the distribution ?. n

One nice property is that P is almost surely discrete: if we obtain a sample tYi ui?1 from it, there is a positive probability of Yi ? Yj for each pair of indexes i ? j. Hence, it induces a random partition ? on N, where i and j are in the same block in ? if and only if Yi ? Yj . Kingman [16] showed that ? is exchangeable, this property will be one of the main tools for the derivation of our hybrid sampler. 2.1

Size-biased sampling Poisson-Kingman processes

A second object induced by a Poisson-Kingman RPM is a size-biased permutation of its atoms. Specifically, order the blocks in ? by increasing order of the least element in each block, and for each k P N let Zk be the least element of the kth block. Zk is the index among pYi qi?1 of the first appearance of the kth unique value in the sequence. Let J?k ? ?ptYZk uq be the mass of the corresponding atom in ?. Then pJ?k qk?1 is a size-biased permutation of the masses of atoms in ?, ? with larger masses tending to appear earlier in the sequence. It is easy to see that k?1 J?k ? T , and that the sequence can be understood as a stick-breaking construction: starting with a stick of length T0 ? T ; break off the first piece of length J?1 ; the surplus length of stick is T1 ? T0 ? J?1 ; then the second piece with length J?2 is broken off, etc. Theorem 2.1 of Perman et al. [21] states that the sequence of surplus masses pTk qk?0 forms a Markov chain and gives the corresponding initial distribution and transition kernels. The corresponding generative process for the sequence pYi qi?1 is as follows: i) Start with drawing the total mass from its distribution P?,h,H0 pT P dtq9hptqf? ptqdt. ii) The first draw Y1 from P is a size-biased pick from the masses of ?. The actual value of Y1 is simply Y1? ? H0 , while the mass of the corresponding atom in ? is J?1 , with conditional 2

distribution s1 f? pt ? s1 q P?,h,H0 pJ?1 P ds1 —T P dtq ? ?pds1 q , t f? ptq

with surplus mass

T1 ? T ? J?1 .

iii) For subsequent draws i ? 2: ? Let K be the current number of distinct values among Y1 , . . . , Yi?1 , and Y1? , . . . , YK? the unique values, i.e., atoms in ?. The masses of these first K atoms are denoted by ?K J?1 , . . . , J?K and the surplus mass is TK ? T ? k?1 J?k . ? For each k ? K, with probability J?k {T , we set Yi ? Y ? . k

? With probability TK {T , Yi takes on the value of an atom in ? besides the first K ? atoms. The actual value YK‘1 is drawn from H0 , while its mass is drawn from sK‘1 f? ptK ? sK‘1 q P?,h,H0 pJ?K‘1 P dsK‘1 —TK P dtK q ? ?pdsK‘1 q , tK f? ptK q

TK‘1 ? TK ?J?K‘1 .

By multiplying the above infinitesimal probabilities, one obtains the joint distribution of the random elements T , ?, pJ?i qi?1 and pYi? qi?1 P?,h,H0 p?n ? pck qkPrKs , Yk? P dyk? , J?k P dsk for k P rKs, T P dtq ? t?n f? pt ?

?K

k?1 sk qhptqdt

K ?

(2)

—c —

sk k ?pdsk qH0 pdyk? q,

k?1

where pck qkPrKs denotes a particular partition of rns with K blocks, c1 , . . . , cK , ordered by increasing least element and —ck — is the cardinality of block ck . The distribution (2) is invariant to the size-biased order. Such a joint distribution was first obtained in Pitman [23] , see also Pitman [24] for further details. 2.2

Relationship to the usual Stick-breaking construction

In the generative process above, we mentioned that it is reminiscent of the well known stick breaking construction from Ishwaran & James [12], where you break a stick of length one but it is not the same. However, we can effectively reparameterize the model, starting with Equation (2), due to d d P J? two useful identities in distribution: Pj ? T ?? j J? and Vj ? 1?? j P‘ for j ? 1, . . . , K.

‘?j

‘

‘?j

Indeed, using this reparameterization, we obtain the corresponding joint in terms of K p0, 1q-valued K stick-breaking weights tVj uj?1 which correspond to a stick-breaking representation. Note that this joint distribution is for a general L?evy measure ?, density f? and it is conditioned on the valued of the random variable T . We can recover the well known Stick breaking representations for the Dirichlet and Pitman-Yor processes, for a specific choice of ? and if we integrate out T , see the supplementary material for further details about the latter. However, in general, these stick-breaking random variables form a sequence of dependent random variables with a complicated distribution, except for the two previously mentioned processes, see Pitman [22] for details. 2.3

Poisson-Kingman mixture model

We are mainly interested in using Poisson-Kingman RPMs as a building block for an infinite mixture model. Indeed, we can use Equation (1) as the top level of the following hierarchical specification

T ?? P —T ? PKp?? , ?T , H0 q iid Yi — P ? P

ind

Xi — Yi ? F p? — Yi q 3

(3)

X1

J?1 , Y1?

X4

X1

J?2 , Y2?

X2

J?3 , Y3?

J?4 , Y1e

X6

X8

X5

X3

T? n

P4
‘=1
0
Y1 e , Y2e
J¿ o 9

Figure 1: Varying table size Chinese restaurant representation for observations tXi ui?1 where F p? — Y q is the likelihood term for each mixture component, and our dataset consists of n observations pxi qiPrns of the corresponding variables pXi qiPrns . We will assume that F p? — Y q is smooth. After specifying the model we would like to carry out inference for clustering and/or density estimation tasks. We can do it exactly and more efficiently than with known MCMC samplers with our novel approach. In the next section, we present our main contribution and in the following one we show how it outperforms other samplers.

3
Hybrid Sampler

Equation?s (2) joint distribution is written in terms of the first K size-biased weights. In order to obtain a complete representation of the RPM, we need to size-bias sample from it a countably infinite number of times. Succesively, devise some way of representing this object exactly in a computer with finite memory and storage is needed. We introduce the following novel strategy: starting from equation (2), we exploit the generative process of section 2.1 when reassigning observations to clusters. In addition to this, we reparame?K terize the model in terms of a surplus mass random variable V ? T ? k?1 J?k and end up with the following joint distribution P?,h,H0 p?n ? pck qkPrKs , Yk? P dyk? , J?k P dsk for k P rKs, T ?

K ?
J?k P dv, Xi P dxi for i P rnsq
k?1
(4) K ?
? pv ‘
? sk q?n h v ‘
k?1
K ?
? sk
k?1
f? pvq
K ?
—c —
sk k ?pdsk qH0 pdyk? q
?
F pdxi —yk? q.
iPck
k?1

For this reason, while having a complete representation of the infinite dimensional part of the model we only need to explicitly represent those size-biased weights associated to occupied clusters plus a surplus mass term which is associated to the rest of the empty clusters, as Figure 1 shows. The cluster reassignment step can be seen as a lazy sampling scheme: we explicitly represent and update the weights associated to occupied clusters and create a size-biased weight only when a new cluster appears. To make this possible we use the induced partition and we call Equation (4) the varying table size Chinese restaurant representation because the size-biased weights can be thought as the sizes of the tables in our restaurant. In the next subsection, we compute the complete conditionals of each random variable of interest to implement an overall Gibbs sampling MCMC scheme. 3.1

Complete conditionals

Starting from equation (4), we obtain the following complete conditionals for the Gibbs sampler ? ??n ? ? K K ? ? P pV P dv — Restq9 v ‘ sk f? pvqh v ‘ sk dv (5) k?1

k?1

? ??n ? ? ? ? ? ? —c — P J?i P dsi — Rest 9 v ‘ si ‘ sk h v ‘ si ‘ sk si i ?pdsi qIp0,Surpmassi q psi qdsi k?i

k?i

4

where Surpmassi ? V ‘

?k j?1

? J?j ? j?i J?j .

sc F pdxi — tXj ujPc Yc? q if i is assigned to existing cluster c Ppci ? c — c?i , Restq9 v ? if i is assigned to a new cluster c M F pdxi — Yc q According to the rule above, the ith observation will be either reassigned to an existing cluster or to one of the M new clusters in the ReUse algorithm as in Favaro & Teh [6]. If it is assigned to a new cluster, then we need to sample a new size-biased weight from the following ? ? P J?k‘1 P dsk‘1 — Rest 9f? pv ? sk‘1 q?psk‘1 qsk‘1 Ip0,vq psk‘1 qdsk‘1 . (6) Every time a new cluster is created we need to obtain its corresponding size-biased weight which could happen 1 ? R ? n times per iteration hence, it has a significant contribution to the overall computational cost. For this reason, an independent and identically distributed (i.i.d.) draw from its corresponding complete conditional (6) is highly desirable. In the next subsection we present a way to achieve this. Finally, for updating cluster parameters tYk? ukPrKs , in the case where H0 is non-conjugate to the likelihood, we use an extension of Favaro & Teh [6]?s ReUse algorithm, see Algorithm 3 in the supplementary material for details. The complete conditionals in Equation (5) do not have a standard form but a generic MCMC method can be applied to sample from each within the Gibbs sampler. We use slice sampling from Neal [19] to update the size-biased weights and the surplus mass. However, there is a class of priors where the total mass?s density is intractable so an additional step needs to be introduced to sample the surplus mass. In the next subsection we present two alternative ways to overcome this issue. 3.2

Example of classes of Poisson-Kingman priors

a) ?-Stable Poisson-Kingman processes [23]. For any ? P p0, 1q, let f? ptq ? ?8 p?1qj‘1 ?p?j‘1q 1 be the density function of a positive ?-Stable random variable and sinp??jq j?0 ? j! t?j‘1 ? ???1 dx. This class of RPMs is denoted by PKp?? , hT , H0 q where h ?pdxq ? ?? pdxq :? ?p1??q x is a function that indexes each member of the class. For example, in the experimental section, we picked 3 choices of the h function that index the following processes: Pitman-Yor, Normalized Stable and Normalized Generalized Gamma processes. This class includes all Gibbs type priors with parameter ? P p0, 1q, so other choices of h are possible, see Gnedin & Pitman [10] and De Blasi et al. [1] for a noteworthy account of this class of Bayesian nonparametric priors. In this case, the total mass?s density is intractable and we propose two ways of dealing with this. Firstly, we used Kanter [14]?s integral representation for the ?-Stable density as in Lomeli et al. [17], introduce an auxiliary variable Z and slice sample each variable ? ? ? ??n k k ? ? ? ? ?? ? ? 1?? P pV P dv — Restq9 v ‘ exp ?v 1?? Apzq h v ‘ si dv si v i?1

i?1

? ? ? P pZ P dz — Restq9Apzq exp ?v p? 1?? q Apzq dz, see Algorithm 1 in the supplementary material for details. Alternatively, we can completely bypass the evaluation of the total mass?s density by updating the surplus mass with a Metropolis-Hastings step with an independent proposal from a Stable or from an Exponentially Tilted Stable(?). It is straight forward to obtain i.i.d draws from these proposals, see Devroye [3] and Hofert [11] for an improved rejection sampling method for the Exponentially tilted case. This leads to the following acceptance ratio ? ??n ? ? ?k ?k 1 1 1 s v ‘ s h v ‘ 1 1 i?1 i i?1 i dv exp p?vq P pV P dv — Restq f? pvq exp p??vq ? ? , ??n ? ? ?k ?k P pV P dv — Restq f? pv 1 q exp p??v 1 q v ‘ i?1 si h v ‘ i?1 si dv exp p?v 1 q see Algorithm 2 in the supplementary material for details. Finally, to sample a new size-biased weight ? ? P J?k‘1 P dsk‘1 — Rest 9f? pv ? sk‘1 qs?? Ip0,vq psk‘1 qdsk‘1 . k‘1

5

Fortunately, we can get an i.i.d. draw from the above due to an identity in distribution given by Favaro et al. [8] for the usual stick breaking weights for any prior in this class such that ? ? uv where u ? v are coprime integers. Then we just reparameterize it back to obtain the new size-biased weight, see Algorithm 4 in the supplementary material for details. b)

? logBeta-Poisson-Kingman

?pa‘bq ?paq?pbq

exp p?atq p1 ? expp?tqq

b?1

processes

[25,

27].

Let

f? ptq

?

d

be the density of a positive random variable X ? ? log Y ,

where Y ? Betapa, bq and ?pxq ? expp?axqp1?expp?bxqq . This class of RPMs generalises the xp1?expp?xqq Gamma process but has similar properties. Indeed, if we take b ? 1 and the density function for T is ?ptq ? f? ptq we recover the L?evy measure and total mass?s density function of a Gamma process. Finally, to sample a new size-biased weight b?1 ? ? p1 ? expps p1 ? expp?bsk'1 qq k'1 ? vqq P J?k'1 P dsk'1 — Rest 9 dsk'1 Ip0,vq psk'1 q. 1 ? expp?sk'1 q

If b ? 1, this complete conditional is a monotone decreasing unnormalised density with maximum at b. We can easily get an i.i.d. draw with a simple rejection sampler [2] where the rejection constant is bv and the proposal is U p0, vq. There is no other known sampler for this process. 3.3

Relationship to marginal and conditional MCMC samplers

Starting from equation (2), another strategy would be to reparameterize the model in terms of the usual stick breaking weights. Next, we could choose a random truncation level and represent finitely many sticks as in Favaro & Walker [7]. Alternatively, we could integrate out the random probability measure and sample only the partition induced by it as in Lomeli et al. [17]. Conditional samplers have large memory requirements as often, the number of sticks needed can be very large. Furthermore, the conditional distributions of the stick lengths are quite involved so they tend to have slow running times. Marginal samplers have less storage requirements than conditional samplers but could potentially have worst mixing properties. For example, Lomeli et al. [17] had to introduce a number of auxiliary variables which worsen the mixing. Our novel hybrid sampler exploits marginal and conditional samplers advantages. It has less memory requirements since it just represents the size-biased weights of occupied as opposed to conditional samplers which represent both empty and occupied clusters. Also, it does not integrate out the size-biased weights thus, we obtain a more comprehensive representation of the RPM.

4

Performance assesssment

We illustrate the performance of our hybrid sampler on a range of Bayesian nonparametric mixture models, obtained by different specifications of ? and ?, as in Equation (3). At the top level of this hierarchical specification, different Bayesian nonparametric priors were chosen from both classes presented in the examples section. We chose the base distribution H0 and the likelihood term F for the kth cluster to be ' ? ' ? ?nk N xi — ?k , ?12 , H0 pd?k q ? N d?k — ?0 , ?02 and F pdx1 , . . . , dxnk — ?k , ?1 q ? i?1 n

k are the nk observations assigned to the kth cluster at some iteration. N denotes a where tXj uj?1 Normal distribution with mean ?k and variance ?12 , a common parameter among all clusters. The mean?s prior distribution is Normal, centered at ?0 and with variance ?02 . Although the base distribution is conjugate to the likelihood we treated it as non-conjugate case and sampled the parameters at each iteration rather than integrating them out.

We used the dataset from Roeder [26] to test the algorithmic performance in terms of running time and effective sample size (ESS), as Table 1 shows. The dataset consists of measurements of velocities in km/sec of n ? 82 galaxies from a survey of the Corona Borealis region. For the ?-Stable Poisson-Kingman class, we compared it against our implementation of Favaro & Walker [7]?s conditional sampler and against the marginal sampler of Lomeli et al. [17]. We chose to compare our hybrid sampler against these existing approaches which follow the same general purpose paradigm. 6

Algorithm Pitman-Yor process (? ? 10) Hybrid Hybrid-MH (? ? 0) Conditional Marginal Hybrid Hybrid-MH (? ? 50) Conditional Marginal Normalized Stable process Hybrid Hybrid-MH (? ? 0) Conditional Marginal Hybrid Hybrid-MH (? ? 50) Conditional Marginal Normalized Generalized Gamma process (? ? 1) Hybrid Hybrid-MH (? ? 0) Conditional Marginal Hybrid Hybrid-MH (? ? 50) Conditional Marginal -logBeta (a ? 1, b ? 2) Hybrid Conditional Marginal

?

Running time

ESS(?std)

0.3 0.3 0.3 0.3 0.5 0.5 0.5 0.5

7135.1(28.316) 5469.4(186.066) NA 4685.7(84.104) 3246.9(24.894) 4902.3(6.936) 10141.6(237.735) 4757.2(37.077)

2635.488(187.335) 2015.625(152.030) NA 2382.799(169.359) 3595.508(174.075) 3579.686(135.726) 905.444(41.475) 2944.065(195.011)

0.3 0.3 0.3 0.3 0.5 0.5 0.5 0.5

5054.7(70.675) 7866.4(803.228) NA 7658.3(193.773) 5382.9(57.561) 4537.2(37.292) 10033.1(22.647) 8203.1(106.798)

5324.146(167.843) 5074.909(100.300) NA 2630.264(429.877) 4877.378(469.794) 4454.999(348.356) 912.382(167.089) 3139.412(351.788)

0.3 0.3 0.3 0.3 0.5 0.5 0.5 0.5

4157.8(92.863) 4745.5(187.506) NA 7685.8(208.98) 6299.2(102.853) 4686.4(35.661) 10046.9(206.538) 8055.6(93.164)

5104.713(200.949) 4848.560(312.820) NA 3587.733(569.984) 4646.987(370.955) 4343.555(173.113) 1000.214(70.148) 4443.905(367.297)

-

2520.6(121.044) NA NA

3068.174(540.111) NA NA

Table 1: Running times in seconds and ESS averaged over 10 chains, 30,000 iterations, 10,000 burn in. Table 1 shows that different choices of ? result in differences in the algorithm?s running times and ESS. The reason for this is that in the ? ? 0.5 case there are readily available random number generators which do not increase the computational cost. In contrast, in the ? ? 0.3 case, a rejection sampler method is needed every time a new size-biased weight is sampled which increases the computational cost, see Favaro et al. [8] for details. Even so, in most cases, we outperform both marginal and conditional MCMC schemes in terms of running times and in all cases, in terms of ESS. In the Hybrid-MH case, even thought the ESS and running times are competitive, we found that the acceptance rate is not optimal, we are currently exploring other choices of proposals. Finally, in Example b), our approach is the only one

available and it has good running times and ESS. This qualitative comparison confirms our previous statements about our novel approach.

5

Discussion

Our main contribution is our Hybrid MCMC sampler as a general purpose tool for inference with a very large class of infinite mixture models. We argue in favour of an approach in which a generic algorithm can be applied to a very large class of models, so that the modeller has a lot of flexibility in choosing specific models suitable for his/her problem of interest. Our method is a hybrid approach since it combines the perks of the conditional and marginal schemes. Indeed, our experiments confirm that our hybrid sampler is more efficient since it outperforms both marginal and conditional samplers in running times in most cases and in ESS in all cases. We introduced a new compact way of representing the infinite dimensional component such that it is feasible to perform inference and how to deal with the corresponding intractabilities. However, there are still various challenges that remain when dealing with these type of models. For instance, there are some values for ? which we are unable to perform inference with our novel sampler. Secondly, when a Metropolis-Hastings step is used, there could be other ways to improve the mixing in terms of better proposals. Finally, all BNP MCMC methods can be affected by the dimensionality and size of the dataset when dealing with an infinite mixture model. Indeed, all methods rely on the same way of dealing with the likelihood term. When adding a new cluster, all methods sample its 7

corresponding parameter from the prior distribution. In a high dimensional scenario, it could be very difficult to sample parameter values close to the existing data points. We consider these points to be an interesting avenue of future research. Acknowledgments We thank Konstantina Palla for her insightful comments. Mar??a Lomel?? is funded by the Gatsby Charitable Foundation, Stefano Favaro is supported by the European Research Council through StG N-BNP 306406 and 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.

2 References

[1] De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Pr?uenster, 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. [2] Devroye, L. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. [3] Devroye, L. 2009. Random variate generation for exponentially and polynomially tilted Stable distributions. ACM Transactions on Modelling and Computer Simulation, 19, 1?20. [4] Escobar, M. D. 1994. Estimating normal means with a Dirichlet process prior. Journal of the American Statistical Association, 89, 268?277. [5] Escobar, M. D., & West, M. 1995. Bayesian density estimation and inference using mixtures. Journal of the Amer-

ican Statistical Association, 90, 577?588. [6] Favaro, S., & Teh, Y. W. 2013. MCMC for Normalized Random Measure Mixture Models. Statistical Science, 28(3), 335?359. [7] Favaro, S., & Walker, S. G. 2012. Slice sampling ?-Stable Poisson-Kingman mixture models. Journal of Computational and Graphical Statistics, 22, 830?847. [8] Favaro, S., Lomeli, M., Nipoti, B., & Teh, Y. W. 2014. On the Stick-Breaking representation of ?-Stable Poisson-Kingman models. Electronic Journal of Statistics, 8, 1063?1085. [9] Ghahramani, Z. 2015. Probabilistic Machine Learning and Artificial Inteligence. Nature, 521, 452459. [10] Gnedin, A., & Pitman, J. 2006. Exchangeable Gibbs partitions and Stirling triangles. Journal of Mathematical Sciences, 138, 5674?5684. [11] Hofert, M. 2011. Efficiently sampling nested Archimedean copulas. Comput. Statist. Data Anal., 55, 5770. [12] Ishwaran, H., & James, L. F. 2001. Gibbs Sampling Methods for Stick-Breaking Priors. Journal of the American Statistical Association, 96(453), 161?173. [13] James, L. F. 2002. Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics. ArXiv:math/0205093. [14] Kanter, M. 1975. Stable densities under change of scale and total variation inequalities. Annals of Probability, 3, 697?707. [15] Kingman, J. F. C. 1967. Completely Random Measures. Pacific Journal of Mathematics, 21, 59?78. [16] Kingman, J. F. C. 1978. The representation of partition structures. Journal of the London Mathematical Society, 18, 374?380. [17] Lomeli, M., Favaro, S., & Teh, Y. W. 2015. A marginal sampler for ?-stable Poisson-Kingman mixture models. Journal of Computational and Graphical Statistics (To appear). [18] Neal, R. M. 1998. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. Tech. rept. 9815. Department of Statistics, University of Toronto. [19] Neal, R. M. 2003. Slice sampling. Annals of Statistics, 31, 705?767. [20] Papaspiliopoulos, O., & Roberts, G. O. 2008. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. Biometrika, 95, 169?186. [21] Perman, M., Pitman, J., & Yor, M. 1992. Size-biased sampling of Poisson point processes and excursions. Probability Theory and Related Fields, 92, 21?39. [22] Pitman, J. 1996. Random discrete distributions invariant under size-biased permutation. Advances in Applied Probability, 28, 525?539.

8

[23] 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. [24] Pitman, J. 2006. Combinatorial Stochastic Processes. Lecture Notes in Mathematics. Springer-Verlag, Berlin. [25] Regazzini, E., Lijoi, A., & Pr?uenster, I. 2003. Distributional results for means of normalized random measures with independent increments. Annals of Statistics, 31, 560?585. [26] 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. [27] von Renesse, M., Yor, M., & Zambotti, L. 2008. Quasi-invariance properties of a class of subordinators. Stochastic Processes and their Applications, 118, 2038?2057. [28] Walker, Stephen G. 2007. Sampling the Dirichlet Mixture Model with Slices. Communications in Statistics - Simulation and Computation, 36, 45.

9