Marginalized Samplers for Normalized Random Measure Mixture Models

Yee Whye Teh and Stefano Favaro

University of Oxford and University of Torino

2013

Outline

Outline

Dirichlet Process

- Random probability measure $\mu \sim \mathrm{DP}(\alpha, H)$.
- For each partition $(A_1, \ldots, A_m)$,

$$(\mu(A_1), \ldots, \mu(A_m)) \sim \mathrm{Dirichlet}(\alpha H(A_1), \ldots, \alpha H(A_m))$$

- Draws from Dirichlet processes are discrete probability measures,

$$\mu = \sum_{k=1}^{\infty} w_k \delta_{\phi_k^*}$$

  where w_k, ϕ_k^* are random.
- Large support over space of probability measures.
- Analytically tractable posterior distribution.

[Ferguson 1973]

Dirichlet Process Mixture Models

- A hierarchical model, with sample $x_{1:n} = (x_1, \dots, x_n)$:

$$\begin{aligned} \mu &\sim \mathrm{DP}(\alpha, H) \\ \phi_i|\mu &\sim \mu \\ x_i|\phi_i &\sim F(\phi_i) \end{aligned}$$

- Discrete nature of μ induces repeated values among $\phi_{1:n}$.
  - Induces a partition π of observations $x_{1:n}$.
  - Each cluster $c \in \pi$ corresponds to a distinct value ϕ^*_c.
  - Leads to a clustering model with a varying/infinite number of clusters.
- Properties of model for cluster analysis depends on the properties of the induced random partition π (a CRP).
- Generalisations of DPs allow for more flexible prior specifications.

[Antoniak 1974, Lo 1984]

MCMC Inference in DP Mixtures

- Conditional samplers [Ishwaran and James 2001, Walker 2007, Kalli and Walker 2006, Papaspiliopoulos and Roberts 2008]
  - Simulate from the joint posterior of μ and $\phi_{1:n}$.
  - Difficulty is in the infinite nature of μ, requiring truncations and retrospective sampling techniques.
  - Easier to understand, parallelizable.
- Marginal samplers [Escobar and West 1995, Bush and MacEachern 1996, Neal 2000]
  - Marginalize out μ, and simulate from posterior for π and $\{\phi_c^*\}$.

$$
\begin{aligned}
\pi &\sim \mathrm{CRP}(\alpha) & \\
\phi_c^* &\sim H & \text{for } c \in \pi \\
x_i|\pi, \{\phi_c^*\} &\sim F(\phi_c^*) & \text{for } c \in \pi \text{ with } i \in c
\end{aligned}
$$

  - Generally better mixing.
  - Only developed for models with analytically known EPPFs so far.

Outline

Dirichlet Process Mixture Models

Normalized Random Measures

MCMC Samplers for NRM Mixture Models

Numerical Illustrations

Completely Random Measures

- A completely random measure (CRM) ν is a random measure such that

$$\nu(A) \perp\!\!\!\perp \nu(B)$$

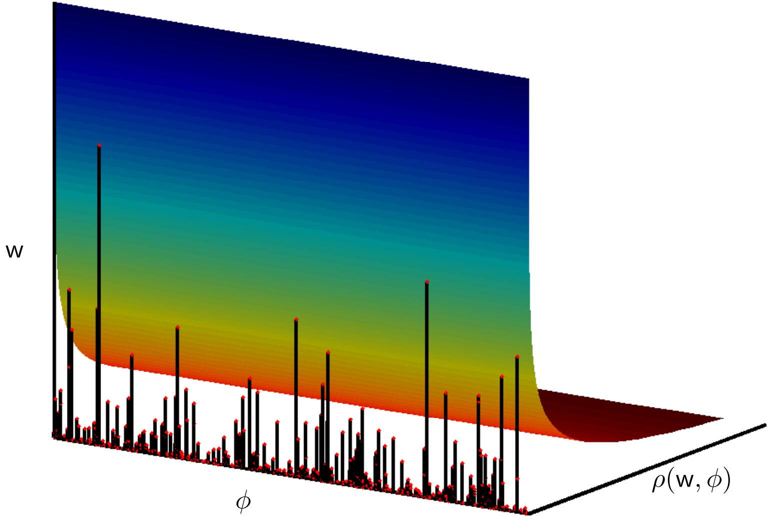
  whenever A and B are disjoint sets.
- The CRM can always be decomposed into 3 components:

$$\nu = \underbrace{\nu_0}_{\text{fixed measure}} + \underbrace{\sum_{j=1}^{\infty} v_j \delta_{\eta_j^*}}_{\text{fixed atoms}} + \underbrace{\sum_{k=1}^{\infty} w_k \delta_{\phi_k^*}}_{\text{random atoms}}$$

  - $\nu_0, \{\eta_j^*\}$ are not random,
  - $\{v_j\}$ are mutually independent and independent of $\{w_k, \phi_k^*\}$,
  - $\{(w_k, \phi_k^*)\}$ is drawn from a Poisson process over $\mathbb{R}_+ \times \Phi$ with rate measure $\rho(w, \phi) dw d\phi$ (the Lévy measure).
- In most modelling applications, only require third component.

[Kingman 1967]

Completely Random Measures



The Lévy Measure

$$\nu = \sum_{k=1}^{\infty} w_k \delta_{\phi_k^*} \qquad \{(w_k, \phi_k^*)\} \sim \text{Poisson}(\rho)$$

- Homogeneous CRMs have $\rho(w, \phi) = \rho(w)h(\phi)$, implies:

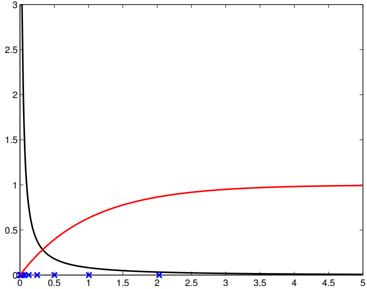
$$\{w_k\} \sim \text{Poisson}(\rho) \qquad \phi_k^* \sim H$$

- Want ν to have infinitely many atoms:

$$\Rightarrow \int_0^{\infty} \rho(w) dw = \infty$$

- Want ν to have finite total mass:

$$\Rightarrow \int_0^{\infty} (1 - e^{-w}) \rho(w) dw < \infty$$



Normalized Random Measures

- Normalizing a CRM gives a normalized random measure (NRM):

$$\mu = \frac{\nu}{\nu(\Phi)}$$

- μ is a random discrete probability measure.
- To study random partition structure, it suffices to assume
  - No fixed measure ($\nu_0 = 0$),
  - No fixed atoms ($v_i = 0$),
  - Homogeneous NRMs $\rho(w, \phi) = \rho(w)h(\phi)$.
- Examples: DP (normalized gamma process), normalized stable process, normalized inverse Gaussian process.

[Regazzini et al. 2003, Nieto-Barajas et al. 2004, James et al. 2009]

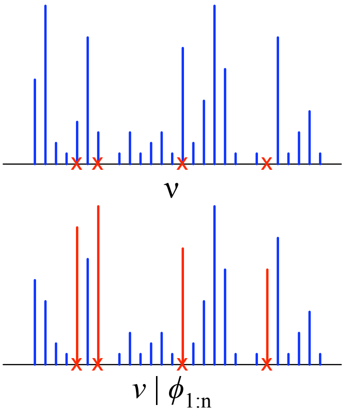
Normalized Generalized Gamma Process

- A large family is obtained by normalizing the generalized gamma process:
$$\rho_{\sigma,\alpha,\tau}(w) = \frac{\alpha}{\Gamma(1-\sigma)} w^{-\sigma-1} e^{-\tau w}$$
- It also has power-law properties (like the Pitman-Yor).
- Specializes to DP when $\sigma = 0$.
- Specializes to normalised stable when $\tau = 0, \alpha = \sigma$.
- Specializes to normalized inverse Gaussian process when $\sigma = \frac{1}{2}$.

[Brix 1999, James 2002]

NRM Hierarchical Model

$$\nu \sim \text{CRM}(\rho, H)$$
$$\mu = \nu / \nu(\Phi)$$
$$\phi_i | \mu \sim \mu \quad \text{for } i = 1 \ldots n$$



- Let $\phi_1^*, \ldots, \phi_K^*$ be the K unique values among $\phi_{1:n}$, with ϕ_k^* occurring n_k times.
- Intuitively,

$$\nu | \phi_{1:n} = \nu^* + \sum_{k=1}^{K} v_k \delta_{\phi_k^*}$$

$$p(\phi_{1:n} | \nu) = \frac{\prod_{k=1}^{K} \nu(\{\phi_k^*\})^{n_k} h(\phi_k^*)}{\nu(\Phi)^n}$$

Data Augmentation

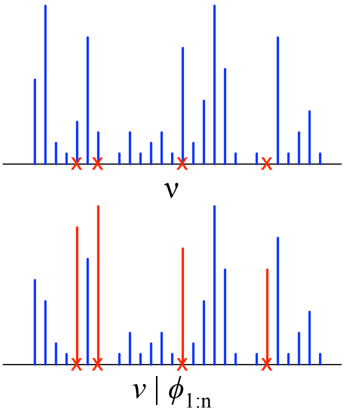
$$p(\phi_{1:n}|\nu) = \frac{\prod_{k=1}^{K} \nu(\{\phi_k^*\})^{n_k} h(\phi_k^*)}{\nu(\Phi)^n}$$

$$= \prod_{k=1}^{K} \nu(\{\phi_k^*\})^{n_k} h(\phi_k^*) \int_0^\infty \frac{u^{n-1} e^{-u\nu(\Phi)}}{\Gamma(n)} du$$

- Introducing an auxiliary variable u:

$$p(\phi_{1:n}, u|\nu)$$

$$= \frac{u^{n-1}}{\Gamma(n)} e^{-u(\nu^*(\Phi) + \sum_{k=1}^{K} v_k)} \prod_{k=1}^{K} v_k^{n_k} h(\phi_k^*)$$



$$\nu|\phi_{1:n} = \sum_{k=1}^{K} v_k \delta_{\phi_k^*} + \nu^*$$

[James 2002, James et al. 2009]

NRM Marginal Characterisation

$$p(\phi_{1:n}, u|\nu) = \frac{u^{n-1}}{\Gamma(n)} e^{-u\nu^*(\Phi)} \prod_{k=1}^{K} v_k^{n_k} e^{-uv_k} h(\phi_k^*)$$

$$p(\phi_{1:n}, u) = \mathbb{E}[p(\phi_{1:n}, u|\nu)] = \frac{u^{n-1}}{\Gamma(n)} e^{-\psi(u)} \prod_{k=1}^{K} \kappa(u, n_k) h(\phi_k^*)$$

The Laplace transform of ρ is

$$\psi(u) = -\log \mathbb{E}[e^{-u\nu(\Phi)}] = \int_0^\infty (1 - e^{-uw})\rho(w)dw$$

The mth moment of the u-exponentially tilted Lévy measure:

$$\kappa(u, m) = \int_0^\infty w^m e^{-uw} \rho(w)dw$$

[James 2002, James et al. 2009]

Outline

NRM Mixture Model

$$\nu \sim \mathrm{CRM}(\rho, H)$$

$$\mu = \nu/\nu(\Phi)$$

For $i = 1, \ldots, n$:

$$\phi_i | \mu \sim \mu$$

$$x_i | \phi_i \sim F(\phi_i)$$

- The distinct values among $\phi_{1:n}$ induces a partition π.
- Let ϕ_c^* be the the distinct value associated with cluster $c \in \pi$.
- Conditional and Pólya urn based samplers have been developed in the literature [James et al. 2009, Nieto-Barajas and Prüenster 2009, Griffin and Walker 2011, Favaro and Walker 2011, Barrios et al. 2013].
- Marginalized samplers operate on the joint distribution of π, $\{\phi_c^*\}$ and $x_{1:n}$.

NRM Marginal Sampler

Conjugate case

$$p(\pi, u, x_{1:n}) = \frac{u^{n-1} e^{-\psi(u)}}{\Gamma(n)} \prod_{c \in \pi} \left(\kappa(u, |c|) \int_{\Phi} h(\phi_c^*) \prod_{i \in c} f(x_i|\phi_c^*) d\phi_c^* \right)$$

- Gibbs sampling:
  - Marginalize out cluster parameters $\{\phi_c^*\}$.
  - Update u given π using slice sampling or MH.
  - Gibbs sample cluster assignment of each item x_i in turn:

$$p(i \in c|\pi_{\setminus i}, x_{1:n})$$

$$\propto \begin{cases} \frac{\kappa(u,|c|+1)}{\kappa(u,|c|)} \int f(x_i|\phi_c^*) h(\phi_c^*|(x_j)_{j \in c}) d\phi_c^* & \text{for } c \in \pi_{\setminus i}, \\ \kappa(u, 1) \int f(x_i|\phi_c^*) h(\phi_c^*) d\phi_c^* & \text{for } c = \text{new cluster.} \end{cases}$$

NRM Marginal Sampler

Conjugate case

- Normalised generalised gamma processes:

$$p(i \in c | \pi_{\setminus i}, x_{1:n})$$
$$\propto \begin{cases} (|c| - \sigma) \int f(x_i|\phi_c^*) h(\phi_c^*|(x_j)_{j \in c}) d\phi_c^* & \text{for } c \in \pi_{\setminus i}, \\ \alpha(u+\tau)^\sigma \int f(x_i|\phi_c^*) h(\phi_c^*) d\phi_c^* & \text{for } c = \text{new cluster.} \end{cases}$$

NRM Marginal Sampler

Non-conjugate case

- Cannot marginalize out cluster parameters $\{\phi_c^*\}$.

$$p(\pi, u, \{\phi_c^*\}, x_{1:n}) \propto \frac{u^{n-1} e^{-\psi(u)}}{\Gamma(n)} \prod_{c \in \pi} \left(\kappa(u, |c|) h(\phi_c^*) \prod_{i \in c} f(x_i | \phi_c^*) \right)$$

- Gibbs sample cluster assignment of item x_i:

$$p(i \in c | \pi_{\setminus i}, x_{1:n}) \propto \begin{cases} \frac{\kappa(u, |c|+1)}{\kappa(u, |c|)} f(x_i | \phi_c^*) & \text{for } c \in \pi_{\setminus i}, \\ \kappa(u, 1) \int f(x_i | \phi_c^*) h(\phi_c^*) d\phi_c^* & \text{for } c = \text{new cluster.} \end{cases}$$

- Integral for new clusters expensive to evaluate.
- When singleton cluster is emptied, parameter is discarded.

NRM Marginal Sampler

Neal's Algorithm 8

- Framed as a data augmentation scheme:
  - Introduce M "new" clusters, with parameters ψ_k for $k = 1, \ldots, M$.

$$\psi_k \sim H$$

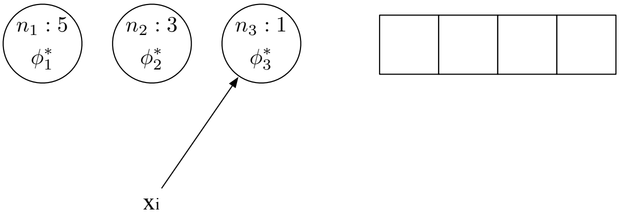
  - Drawn before each Gibbs update.
  - Exists only during the Gibbs update, and discarded afterwards.
  - The parameter of an emptied cluster is used as the parameter for one of the new cluster.

$$p(i \in c | \pi_{\setminus i}, x_{1:n}) \propto \begin{cases} \frac{\kappa(u, |c|+1)}{\kappa(u, |c|)} f(x_i | \phi^*_c) & \text{for } c \in \pi_{\setminus i}, \\ \frac{\kappa(u, 1)}{M} f(x_i | \psi_k) & \text{for } c = k \in \{1, \ldots, M\}. \end{cases}$$

[Neal 2000]

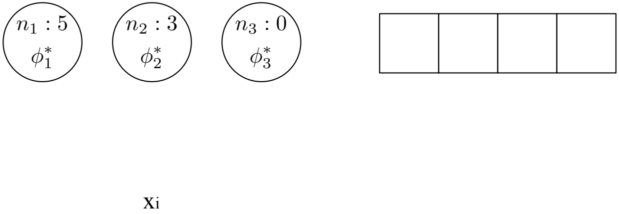
NRM Marginal Sampler

Neal's Algorithm 8



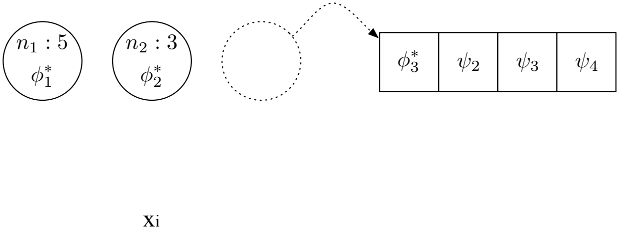
NRM Marginal Sampler

Neal's Algorithm 8



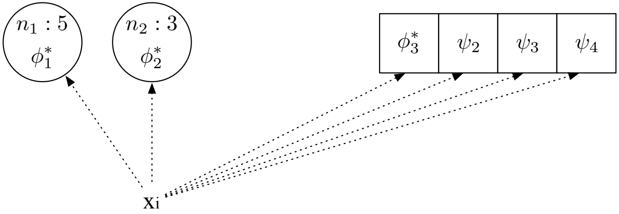
NRM Marginal Sampler

Neal's Algorithm 8



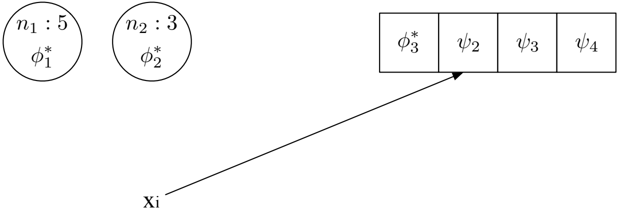
NRM Marginal Sampler

Neal's Algorithm 8



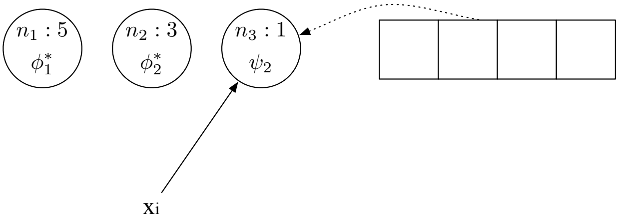
NRM Marginal Sampler

Neal's Algorithm 8



NRM Marginal Sampler

Neal's Algorithm 8



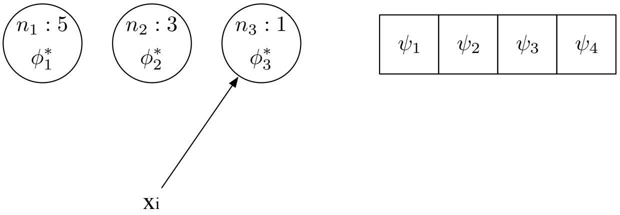
NRM Marginal Sampler

Reuse Algorithm

- Computationally expensive to generate many parameters from base distribution h.
- Would like to somehow reuse unused parameters.
- A transdimensional algorithm:
    - Augment state space permanently with M new clusters.
    - Reversible jump Metropolis-Hastings updates [Green 1995].

NRM Marginal Sampler

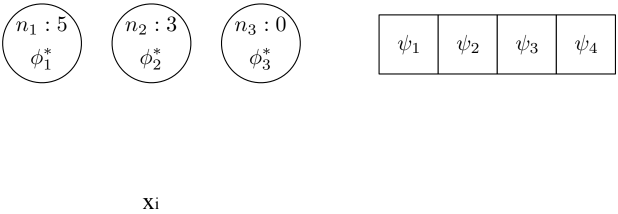
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

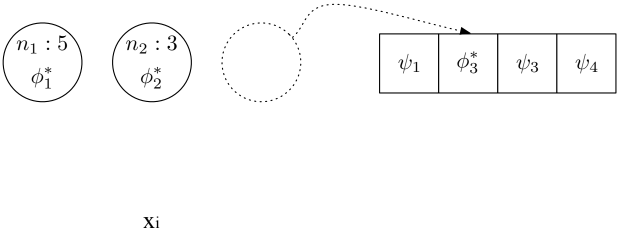
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

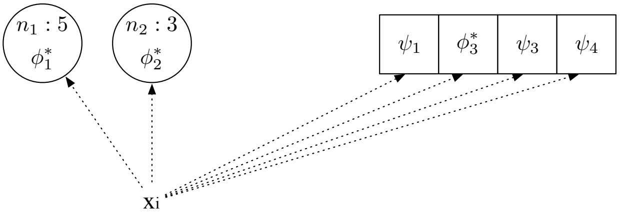
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

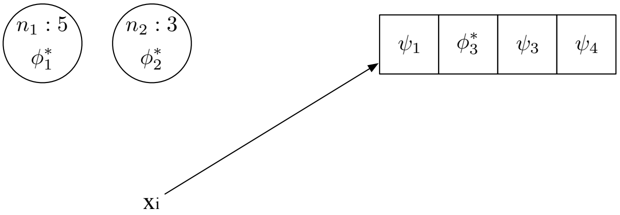
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

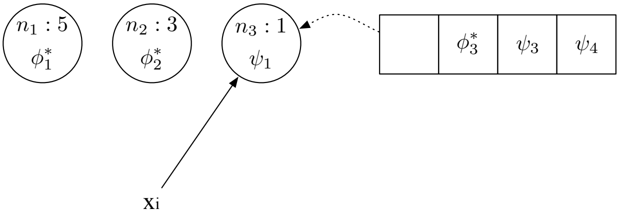
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

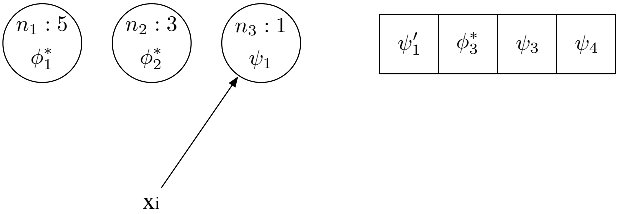
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

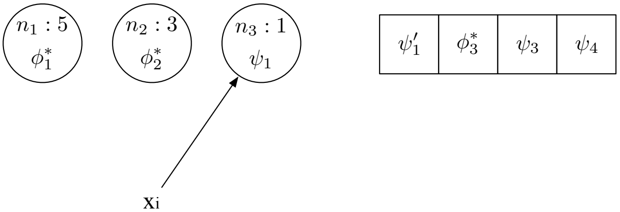
Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

NRM Marginal Sampler

Reuse Algorithm



- Augment state space with M new clusters.
- Unassign x_i; if current cluster is a singleton,
  - Replace the parameter of a randomly chosen new cluster with its parameter.
- Reassign cluster assignment of x_i.
- If x_i is assigned to a new cluster,
  - Create a cluster with the parameter,
  - Generate a new parameter from base distribution.
- Acceptance probability always one.

Outline

NRM Mixture of Normals

- One-dimensional examples:
  - Galaxy ($n = 82$)
  - Acidity ($n = 155$)
- Multi-dimensional examples:
  - Old Faithful, ($p = 2, n = 272$)
  - Neural spike sorting, ($p = 6, n = 1000, 2000$)

- Non-conjugate prior over mean and covariance of normals:

$$m \sim \mathcal{N}(m_0, S_0) \qquad \Sigma \sim \mathcal{IW}(\alpha_0, \Sigma_0)$$

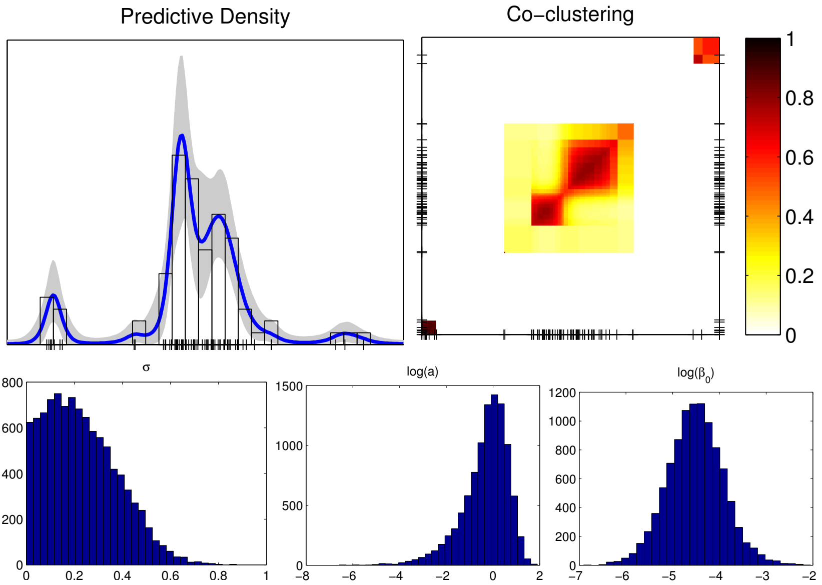
- Hierarchical prior for $\Sigma_0 \sim \mathcal{IW}(\beta_0, \gamma_0 S_0)$.
- Weakly informative, using prior knowledge of data range.
- In 1D case reduces to prior used in [Richardson and Green 1997].

[Richardson and Green 1997]

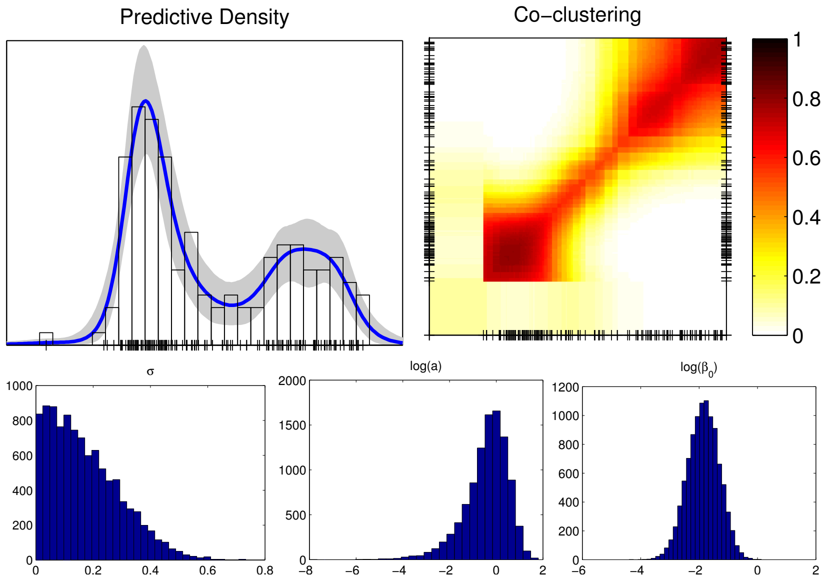
Efficiency Evaluation

- 10000 iterations burn-in, 10000 samples collected from 200000 iterations.
- Effective sample size of number of clusters K using Coda.
- Reports mean ESS and standard error over 10 repeats.
- Compared:
  - Conjugate marginalized sampler
  - Neal's Algorithm 8 marginalized sampler
  - Reuse Algorithm marginalized sampler
  - Slice sampler based on posterior representation
    - Variation on [Griffin and Walker 2011]
    - Truncation required.

Galaxy Dataset



Acidity Dataset



Comparative Results

Galaxy and Acidity datasets (conjugate model)

Sampler	Galaxy		Acidity	
	Runtime (s)	ESS	Runtime (s)	ESS
Cond Slice	239.1 ± 4.2	2004 ± 178	196.5 ± 1.0	910 ± 142
Marg ($C = 1$)	215.7 ± 1.4	7809 ± 87	395.5 ± 1.7	5236 ± 181
Cond Slice	133.0 ± 3.2	1594 ± 117	77.4 ± 0.7	1099 ± 49
Marg Neal 8 (C=1)	74.4 ± 0.6	5815 ± 145	133.3 ± 1.8	4175 ± 85
Marg Neal 8 (C=2)	87.9 ± 0.6	6292 ± 94	163.8 ± 1.5	4052 ± 158
Marg Neal 8 (C=3)	101.9 ± 0.7	6320 ± 137	188.2 ± 1.1	4241 ± 99
Marg Neal 8 (C=4)	115.9 ± 0.6	6283 ± 86	216.6 ± 1.7	4266 ± 122
Marg Neal 8 (C=5)	130.0 ± 0.6	6491 ± 203	243.8 ± 2.0	4453 ± 123
Marg Reuse (C=1)	64.3 ± 0.3	4451 ± 79	114.6 ± 2.0	3751 ± 65
Marg Reuse (C=2)	67.6 ± 0.5	5554 ± 112	123.1 ± 1.9	4475 ± 110
Marg Reuse (C=3)	71.3 ± 0.5	5922 ± 157	128.2 ± 2.2	4439 ± 158
Marg Reuse (C=4)	74.9 ± 0.5	6001 ± 101	140.1 ± 1.6	4543 ± 108
Marg Reuse (C=5)	78.7 ± 0.6	6131 ± 124	147.7 ± 1.5	4585 ± 116

Comparative Results

Galaxy and Acidity datasets (non-conjugate model)

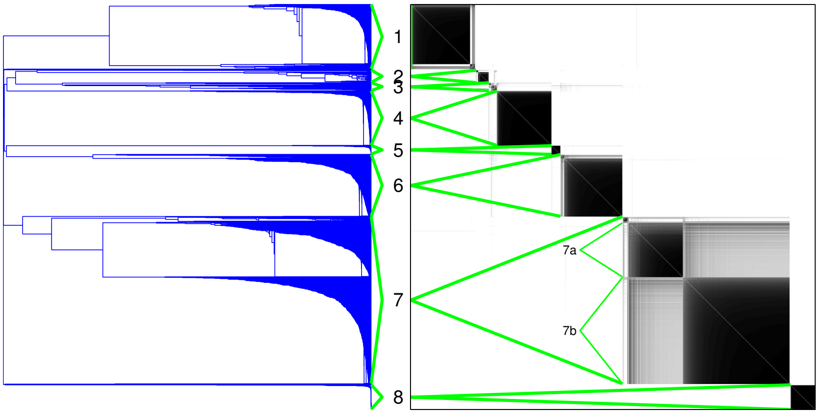
Sampler	Galaxy		Acidity	
	Runtime (s)	ESS	Runtime (s)	ESS
Cond Slice	75.5 ± 1.2	939 ± 92	50.9 ± 0.5	949 ± 70
Marg Neal 8 (C=1)	65.0 ± 0.5	4313 ± 172	110.9 ± 0.8	4144 ± 64
Marg Neal 8 (C=2)	78.6 ± 0.4	4831 ± 168	139.2 ± 1.8	4290 ± 125
Marg Neal 8 (C=3)	92.5 ± 0.5	4785 ± 97	162.7 ± 0.9	4368 ± 72
Marg Neal 8 (C=4)	106.3 ± 0.5	4849 ± 120	187.6 ± 1.1	4234 ± 142
Marg Neal 8 (C=5)	119.7 ± 0.6	5029 ± 89	215.4 ± 1.3	4144 ± 213
Marg Reuse (C=1)	55.2 ± 0.5	3830 ± 103	91.3 ± 0.9	4007 ± 122
Marg Reuse (C=2)	58.7 ± 0.5	4286 ± 101	98.1 ± 0.9	4192 ± 138
Marg Reuse (C=3)	62.4 ± 0.6	4478 ± 124	105.1 ± 0.9	4260 ± 136
Marg Reuse (C=4)	66.1 ± 0.5	4825 ± 63	112.3 ± 1.0	4191 ± 139
Marg Reuse (C=5)	69.8 ± 0.6	4755 ± 141	121.0 ± 1.8	4186 ± 121

Comparative Results

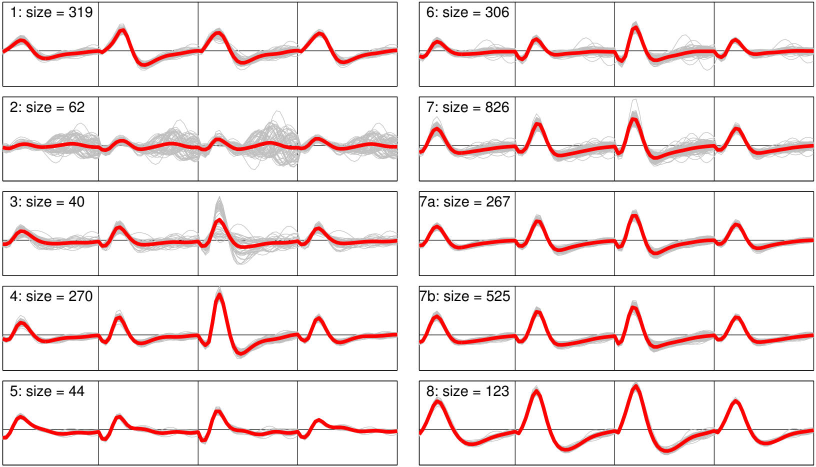
Old Faithful and spike sorting datasets (non-conjugate model)

Sampler	Old Faithful		Spike Sorting	
	Runtime (s)	ESS	Runtime (s)	ESS
Cond Slice	142.6 ± 1.1	574 ± 36	732.6 ± 8.1	17.1 ± 2.3
Marg Reuse (C=1)	208.0 ± 1.3	2770 ± 209	1120.3 ± 8.8	35.7 ± 2.4
Marg Reuse (C=2)	225.3 ± 1.4	3236 ± 73	1164.5 ± 5.4	46.9 ± 2.9
Marg Reuse (C=3)	241.5 ± 1.3	3148 ± 71	1204.1 ± 7.3	57.0 ± 3.9
Marg Reuse (C=4)	257.7 ± 1.7	3291 ± 145	1238.5 ± 7.8	61.4 ± 3.3
Marg Reuse (C=5)	274.8 ± 1.7	3144 ± 70	1291.8 ± 7.9	69.8 ± 4.9
Marg Reuse (C=10)	356.3 ± 2.5	3080 ± 135	1513.8 ± 11.9	90.8 ± 5.6
Marg Reuse (C=15)	446.6 ± 4.9	3312 ± 154	1746.3 ± 10.7	95.9 ± 4.2
Marg Reuse (C=20)	550.4 ± 3.5	3336 ± 109	1944.0 ± 14.7	114.5 ± 8.4

Spike Sorting Dataset



Spike Sorting Dataset



Conclusion

- Marginalised samplers for NRMs more efficient than conditional slice samplers.
- Simple algorithms, introducing an additional auxiliary variable u.

- Pitman-Yor processes are not normalised random measures.
- Marginalised samplers for all σ-stable Poisson-Kingman mixture models (including Pitman-Yor) (Lomeli et al).
- Motivation for normalised random measures?
  - Power-law properties
  - Dependent normalised random measures

Normalized Gamma Process

- A gamma process is a CRM with Lévy measure

$$\rho_{\alpha,\tau}(w, \phi) = \alpha w^{-1} e^{-\tau w} h(\phi)$$

- The gamma process has gamma marginals:

$$\nu(A) \sim \text{Gamma}\left(\alpha \int_A h(\phi) d\phi, \tau\right)$$

- Normalizing a gamma process gives a Dirichlet process, with mass parameter α and base distribution H with density h.

Normalized Stable Process

- A stable process is a CRM with Lévy measure

$$\rho_\sigma(w) = \frac{\sigma}{\Gamma(1-\sigma)} w^{-\sigma-1}$$

- It has positive stable marginals with index σ.
- Normalizing a stable process, and deriving the induced exchangeable partition process, leads to the following random partition:
  - Customer 1 sits at first table.
  - Subsequent customer $n+1$:
    - sits at table c with probability $\frac{|c|-\sigma}{n}$,
    - sits at new table with probability $\frac{|\pi|\sigma}{n}$.
- Related to the two-parameter Poisson-Dirichlet process (aka Pitman-Yor process).

[Perman et al. 1992, Pitman and Yor 1997]

NRM Posterior Characterisation

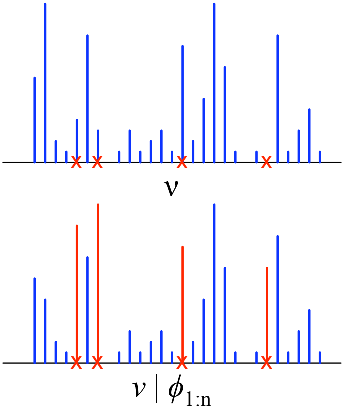
$$\nu|\phi_{1:n}, u = \nu^* + \sum_{k=1}^{K} v_k \delta_{\phi_k^*}$$

$$p(\phi_{1:n}, u|\nu) = \frac{u^{n-1} e^{-u\nu^*(\Phi)}}{\Gamma(n)} \prod_{k=1}^{K} v_k^{n_k} e^{-u v_k}$$

$$\nu^*|\phi_{1:n} \sim \mathrm{CRM}(\rho^*, H)$$

$$\rho^*(w, \phi) = e^{-uw} \rho(w)$$

$$p(v_k|\phi_{1:n}, u) \propto v_k^{n_k} e^{-u v_k} \rho(v_k)$$



[James et al. 2009]

References I

- Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2(6):1152–1174.
- Barrios, E., Lijoi, A., Nieto-Barajas, L. E., and Prüenster, I. (2013). Modeling with normalized random measure mixture models. Unpublished manuscript.
- Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, 31:929–953.
- Bush, C. A. and MacEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika*, 83:275–285.
- Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588.
- Favaro, S. and Walker, S. G. (2011). Slice sampling σ–stable Poisson–Kingman mixture models.
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1(2):209–230.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.
- Griffin, J. E. and Walker, S. G. (2011). Posterior simulation of normalized random measure mixtures. *Journal of Computational and Graphical Statistics*, 20(1):241–259.
- Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173.
- James, L. F. (2002). Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics.
- James, L. F., Lijoi, A., and Prüenster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, 36:76–97.
- Kalli, M. and Walker, S. G. (2006). Slice sampling for the Dirichlet process mixture model. Poster presented at the Eighth Valencia International Meeting on Bayesian Statistics.
- Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.

References II

- Lo, A. (1984). On a class of bayesian nonparametric estimates: I. density estimates. *Annals of Statistics*, 12(1):351–357.
- Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265.
- Nieto-Barajas, L. E. and Prüenster, I. (2009). A sensitivity analysis for Bayesian nonparametric density estimators. *Statistica Sinica*, 19:685–705.
- Nieto-Barajas, L. E., Pruenster, I., and Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *Annals of Statistics*, 32(6):2343–2360.
- Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, 95(1):169–186.
- Perman, M., Pitman, J., and Yor, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields*, 92(1):21–39.
- Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability*, 25:855–900.
- Regazzini, E., Lijoi, A., and Prüenster, I. (2003). Distributional results for means of random measures with independent increments. *Annals of Statistics*, 31:560–585.
- Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B*, 59(4):731–792.
- Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, 36:45.