Constructing dependent random probability measures from completely random measures

Changyou Chen[1], Vinayak Rao[2], Wray Buntine[1], Yee Whye Teh[3]

presented by **Sinead Williamson**[4]

[1]NICTA, [2]Duke University, [3]University of Oxford, [4]UT Austin

ICML 2013

Introduction

Two strands of research in NPBayes modelling of random probability measures (RPMs):

Introduction

Two strands of research in NPBayes modelling of random probability measures (RPMs):

priors that are more expressive than the Dirichlet Process

Introduction

Two strands of research in NPBayes modelling of random probability measures (RPMs):

priors that are more expressive than the Dirichlet Process

- e.g. power-law behaviour or more uncertainty on number of clusters:
  Normalized Random Measures [James et al., 2005, Kingman, 1975]
  (e.g. normalized generalized Gamma process)
  Poisson-Kingman processes [Pitman, 2003]
  (e.g. Pitman-Yor process [Pitman and Yor, 1997])

Introduction

Two strands of research in NPBayes modelling of random probability measures (RPMs):

priors that are more expressive than the Dirichlet Process

priors that model more structured data

Introduction

Two strands of research in NPBayes modelling of random probability measures (RPMs):

priors that are more expressive than the Dirichlet Process

priors that model more structured data

- for data violating the assumption of exchangeability:
  Time-series, spatial data, conditional density modelling
- Research traces back to work of [MacEachern, 1999] on *dependent RPMs*

Introduction

Two strands of research in NPBayes modelling of random probability measures (RPMs):

priors that are more expressive than the Dirichlet Process

priors that model more structured data

This talk:

Flexible constructions for dependent RPMs with flexible marginals

Relevant work

There is a rich literature on dependent RPMs

- the seminal work of [MacEachern, 1999] on dependent DPs
- Existing work that is directly relevant [Rao and Teh, 2009, Nipoti, 2010, Lijoi et al., 2012, Foti et al., 2012, Lin and Fisher, 2012, Griffin et al., 2013]

Relevant work

There is a rich literature on dependent RPMs

- the seminal work of [MacEachern, 1999] on dependent DPs
- Existing work that is directly relevant [Rao and Teh, 2009, Nipoti, 2010, Lijoi et al., 2012, Foti et al., 2012, Lin and Fisher, 2012, Griffin et al., 2013]
- C. Chen, V. Rao, W. Buntine and Y.W. Teh (2013)
  Dependent Normalized Random Measures

Completely random measures (CRMs)

- A random measure μ on some space $(\mathbb{X}, \Sigma_X)$ such that

$$\mu(A) \perp\!\!\!\perp \mu(B) \text{ if } A \text{ and } B \text{ are disjoint}$$

Completely random measures (CRMs)

- A random measure μ on some space $(\mathbb{X}, \Sigma_X)$ such that

$$\mu(A) \perp\!\!\!\perp \mu(B) \text{ if } A \text{ and } B \text{ are disjoint}$$

- The measure μ is atomic:

$$\mu = \sum_i w_i \delta_{x_i}$$

Completely random measures (CRMs)

- A random measure μ on some space $(\mathbb{X}, \Sigma_X)$ such that

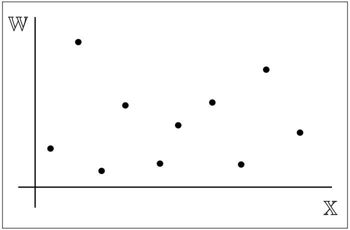
$$\mu(A) \perp\!\!\!\perp \mu(B) \text{ if } A \text{ and } B \text{ are disjoint}$$

- The measure μ is atomic:

$$\mu = \sum_i w_i \delta_{x_i}$$

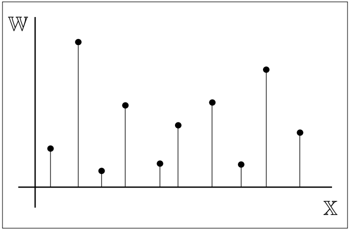
- (x_i, w_i) : events of a Poisson process on the space $\mathbb{X} \times \mathbb{W}$, where $\mathbb{W} = [0, \infty)$.
- The Poisson process has intensity $\nu(w, x) = \rho(w)h(x)$, where
  $\rho(w)$ is the Lévy intensity of the CRM, and
  $h(x)$ is the base probability density.

Normalized random measures



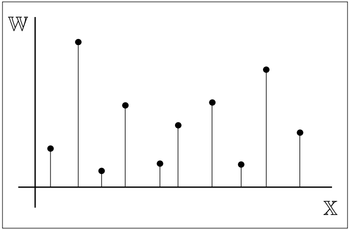
- Poisson process $\{w_i, x_i\}$

Normalized random measures



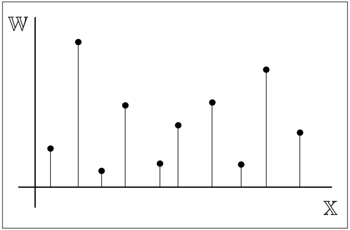
- Poisson process $\{w_i, x_i\}$
- CRM $\mu \equiv \{w_i, x_i\}$

Normalized random measures



- Poisson process $\{w_i, x_i\}$
- CRM $\mu \equiv \{w_i, x_i\}$
- Normalize to construct a random probability measure G: $G(\cdot) = \frac{\mu(\cdot)}{\mu(\mathbb{X})}$

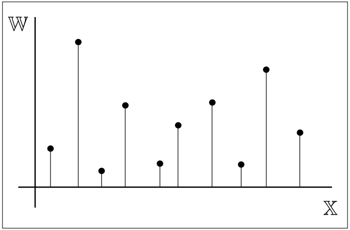
Normalized random measures



- Poisson process $\{w_i, x_i\}$
- CRM $\mu \equiv \{w_i, x_i\}$
- Normalize to construct a random probability measure G: $G(\cdot) = \frac{\mu(\cdot)}{\mu(\mathbb{X})}$

In the following, we set $\rho(w) = \alpha w^{-\sigma-1} \exp(-\tau w)$ corresponding to the generalized Gamma process.

Normalized random measures



- Poisson process $\{w_i, x_i\}$
- CRM $\mu \equiv \{w_i, x_i\}$
- Normalize to construct a random probability measure G: $G(\cdot) = \frac{\mu(\cdot)}{\mu(\mathbb{X})}$

In the following, we set $\rho(w) = \alpha w^{-\sigma-1} \exp(-\tau w)$ corresponding to the generalized Gamma process.

We want: Dependent normalized random measures, G_t

Dependent normalized random measures

- Define a common latent CRM/Poisson process.
- Define dependent measures via transformations of this process.

Dependent normalized random measures

- Define a common latent CRM/Poisson process.
- Define dependent measures via transformations of this process.
  - Superposition [Rao and Teh, 2009, Griffin et al., 2013]
  - Rescaling
  - Thinning [Lin et al., 2010, Lin and Fisher, 2012]

Normalize these dependent CRMs to produce dependent NRMs.

Superposition theorem

- The superposition of two independent Poisson processes with intensity $\nu_i(\cdot),\ i = 1, 2$ is a Poisson process with intensity $\nu_1(\cdot) + \nu_2(\cdot)$
- The resulting CRM has Lévy measure $\rho = \rho_1 + \rho_2$.

Superposition theorem

- The superposition of two independent Poisson processes with intensity $\nu_i(\cdot),\ i = 1, 2$ is a Poisson process with intensity $\nu_1(\cdot) + \nu_2(\cdot)$
- The resulting CRM has Lévy measure $\rho = \rho_1 + \rho_2$.
- The projection of a Poisson process from $\mathbb{X} \times \mathbb{W} \times \mathbb{A}$ to $\mathbb{X} \times \mathbb{W}$ is a Poisson process with intensity $\int_{\mathbb{A}} \nu(\mathrm{d}x, \mathrm{d}w, \mathrm{d}a)$

Superposition theorem

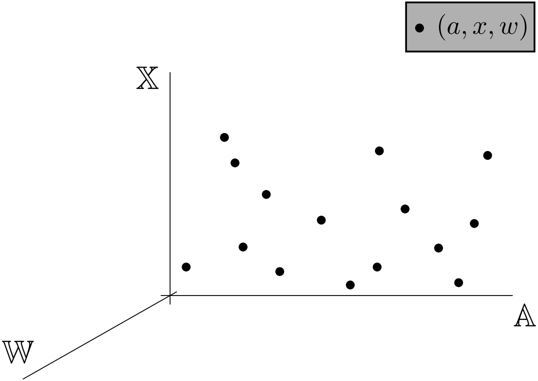
- The superposition of two independent Poisson processes with intensity $\nu_i(\cdot),\ i = 1, 2$ is a Poisson process with intensity $\nu_1(\cdot) + \nu_2(\cdot)$
- The resulting CRM has Lévy measure $\rho = \rho_1 + \rho_2$.
- The projection of a Poisson process from $\mathbb{X} \times \mathbb{W} \times \mathbb{A}$ to $\mathbb{X} \times \mathbb{W}$ is a Poisson process with intensity $\int_{\mathbb{A}} \nu(\mathrm{d}x, \mathrm{d}w, \mathrm{d}a)$
- If $\nu(\cdot)$ factors as $\rho(w)h(x)\nu_a(a)$, then the resulting CRM has Lévy intensity $\left(\int_{\mathbb{A}} \nu_a(a)\mathrm{d}a\right) \rho(w)h(x)$.

Spatial Normalized Gamma processes [Rao and Teh, 2009]

- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space

Spatial Normalized Gamma processes [Rao and Teh, 2009]

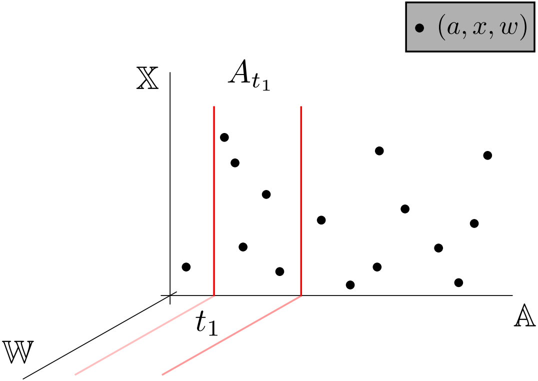
- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space



- Instantiate a Poisson process on some augmented space

Spatial Normalized Gamma processes [Rao and Teh, 2009]

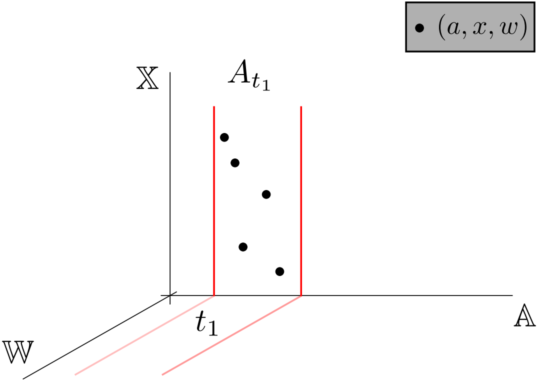
- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space



- Instantiate a Poisson process on some augmented space
- Associate each t with a subset $\mathbb{X} \times \mathbb{W} \times A_t$

Spatial Normalized Gamma processes [Rao and Teh, 2009]

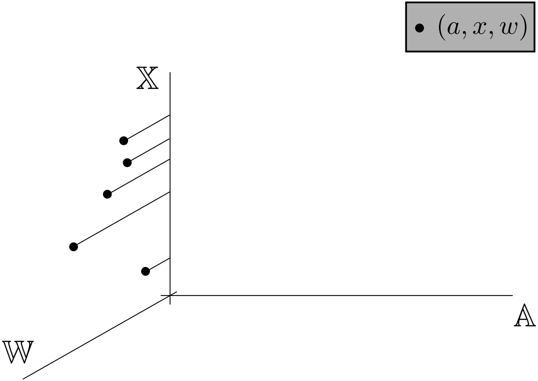
- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space



- Instantiate a Poisson process on some augmented space
- Associate each t with a subset $\mathbb{X} \times \mathbb{W} \times A_t$
- Restrict to A_t, and project onto the original space, defining an NRM

Spatial Normalized Gamma processes [Rao and Teh, 2009]

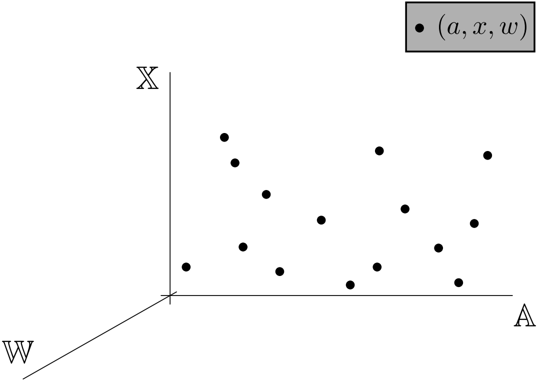
- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space



- Instantiate a Poisson process on some augmented space
- Associate each t with a subset $\mathbb{X} \times \mathbb{W} \times A_t$
- Restrict to A_t, and project onto the original space, defining an NRM

Spatial Normalized Gamma processes [Rao and Teh, 2009]

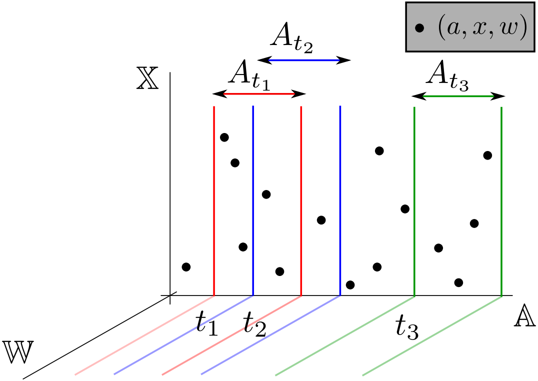
- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space



- Instantiate a Poisson process on some augmented space
- Associate each t with a subset $\mathbb{X} \times \mathbb{W} \times A_t$
- Restrict to A_t, and project onto the original space, defining an NRM

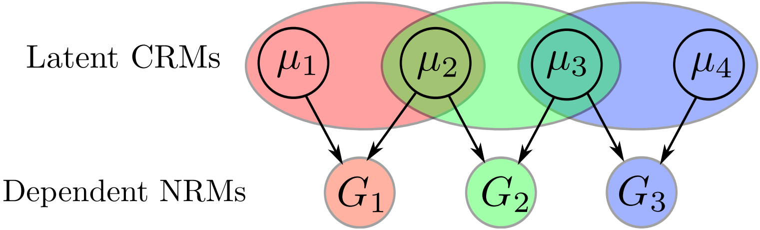
Spatial Normalized Gamma processes [Rao and Teh, 2009]

- A measure-valued stochastic process G_t, $t \in \mathbb{T}$ is an arbitrary space

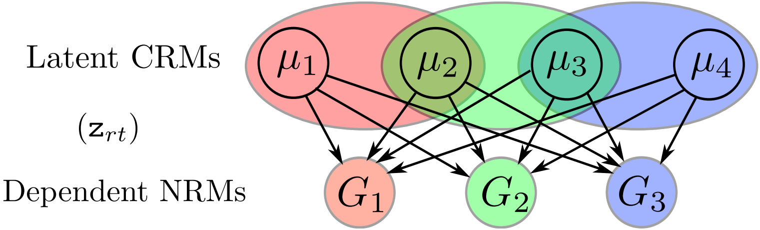


- Instantiate a Poisson process on some augmented space
- Associate each t with a subset $\mathbb{X} \times \mathbb{W} \times A_t$
- Restrict to A_t, and project onto the original space, defining an NRM
- Dependency across NRMs is controlled by amount of overlap of A_t's

Dependent normalized random measures ([Griffin et al., 2013])

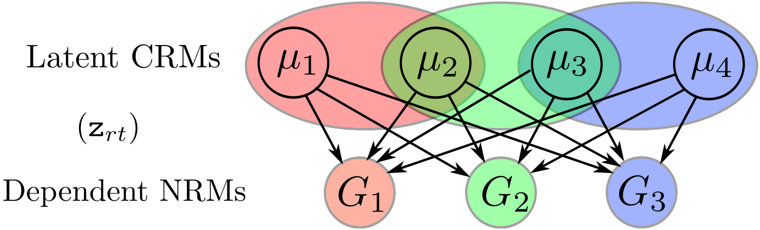


Dependent normalized random measures ([Griffin et al., 2013])



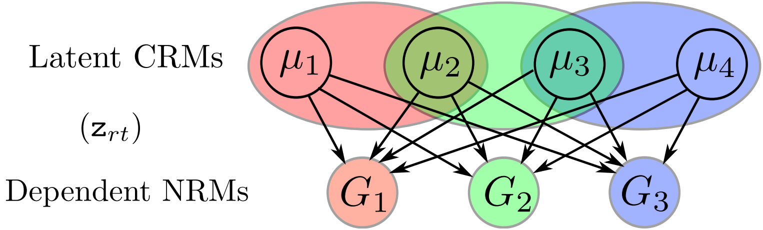
$$G_t \propto \sum_{r=1}^{R} z_{tr}\mu_r, \quad z_{tr} \in \{0, 1\} \tag{1}$$

Mixed normalized random measures



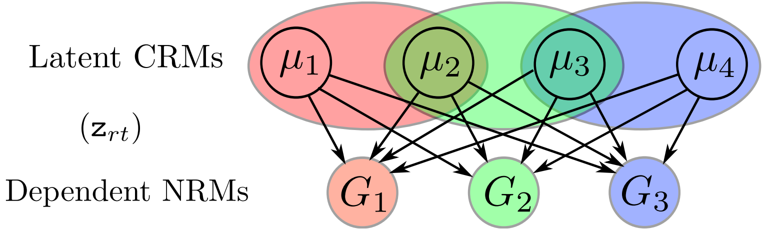
- A simple generalization: allow $\mathbf{z}_{tr} \in \mathbb{R}^+$

Mixed normalized random measures



- A simple generalization: allow $\mathbf{z}_{tr} \in \mathbb{R}^+$
- $\mathbf{z}\mu(\cdot)$ belongs to the same class of CRMs as $\mu(\cdot)$.
- Poisson mapping theorem:
  - If $\{w_i\}$ is a sample from a Poisson process with intensity $\nu(w)$, then $\{\mathbf{z}w_i\}$ is a Poisson process with intensity $\mathbf{z}^{-1}\nu(w/\mathbf{z})$.

Mixed normalized random measures



- A simple generalization: allow $\mathbf{z}_{tr} \in \mathbb{R}^+$
- $\mathbf{z}\mu(\cdot)$ belongs to the same class of CRMs as $\mu(\cdot)$.
- Poisson mapping theorem:
  - If $\{w_i\}$ is a sample from a Poisson process with intensity $\nu(w)$, then $\{\mathbf{z}w_i\}$ is a Poisson process with intensity $\mathbf{z}^{-1}\nu(w/\mathbf{z})$.
- $\mathbf{z}_{tr}$ governs how strongly atoms of CRM r contribute to covariate t.
- Given a set $\{\mathbf{z}_{tr}\}$, for each t, G_t is an NRM.

Thinning theorem

- If $\{w_i\}$ is a sample from a Poisson process with intensity $\nu(w)$, then $\{z_i w_i\}$, where $z_i \stackrel{i.i.d.}{\sim}$ Bernoulli(p) is Poisson with intensity $p\nu(w)$.
- Suggests independently thinning atoms of a CRM to form a new CRM ([Lin et al., 2010]).
- Corresponds to SNRM with an exponential number of CRM.

Thinned Normalized Random Measures

- Spatial NRM characterized by a set $\{\mathrm{z}_{tr} \in \{0,1\} \quad \forall t \in \mathcal{T}, r \in \mathcal{R}\}$
- z_{tr} specifies whether or not *all* atoms of CRM r are present at covariate t.

Thinned Normalized Random Measures

- Spatial NRM characterized by a set $\{z_{tr} \in \{0, 1\} \quad \forall t \in \mathcal{T}, r \in \mathcal{R}\}$
- z_{tr} specifies whether or not *all* atoms of CRM r are present at covariate t.
- Thinned NRM: introduce indicator variables $z_{trw} \in \{0, 1\}$ for each atom.

$$z_{rtk} \sim \text{Bernoulli}(q_{rt}) \quad k = 1, 2, \cdots$$

Then, the probability measure at covariate t is given by

$$\mu_t(\mathrm{d}\theta) = \frac{1}{\hat{\mu}_t(\Theta)} \hat{\mu}_t(\mathrm{d}\theta), \text{ where } \hat{\mu}_t(\mathrm{d}\theta) = \sum_{r=1}^{R} \sum_{k=1}^{\infty} z_{rtk} w_{rk} \tag{2}$$

Thinned Normalized Random Measures

- Spatial NRM characterized by a set $\{z_{tr} \in \{0,1\} \quad \forall t \in \mathcal{T}, r \in \mathcal{R}\}$
- z_{tr} specifies whether or not *all* atoms of CRM r are present at covariate t.
- Thinned NRM: introduce indicator variables $z_{trw} \in \{0,1\}$ for each atom.

$$z_{rtk} \sim \text{Bernoulli}(q_{rt}) \quad k = 1, 2, \cdots$$

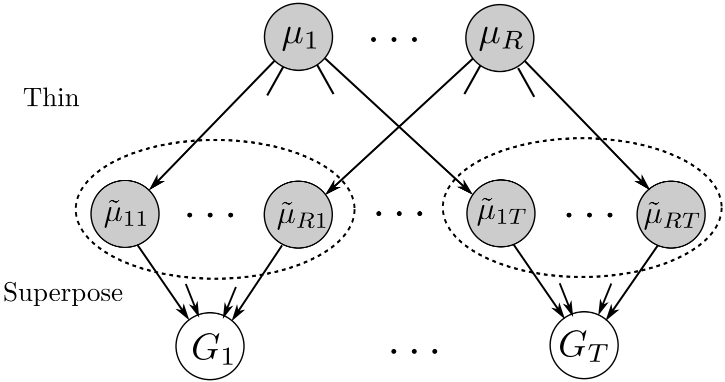
Then, the probability measure at covariate t is given by

$$\mu_t(\mathrm{d}\theta) = \frac{1}{\hat{\mu}_t(\Theta)} \hat{\mu}_t(\mathrm{d}\theta), \text{ where } \hat{\mu}_t(\mathrm{d}\theta) = \sum_{r=1}^{R} \sum_{k=1}^{\infty} z_{rtk} w_{rk} \tag{2}$$

Proposition

Conditioned on the set of q_{rt}'s, each random probability measure μ_t defined in (2) is marginally distributed as a normalized random measure with Lévy measure $\sum_r z_{rt} \nu_r(\mathrm{d}w, \mathrm{d}\theta)$.

Dependent normalized random measures



Inference

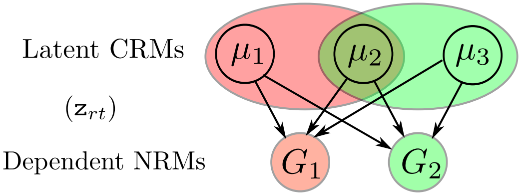
- [Lin et al., 2010] have proposed a similar model at NIPS this year.
- They provide a marginal sampler for posterior inference.

Inference

- [Lin et al., 2010] have proposed a similar model at NIPS this year.
- They provide a marginal sampler for posterior inference.
- Unfortunately, this sampler is incorrect.
- A similar error exists in [Lin and Fisher, 2012]

Inference

- [Lin et al., 2010] have proposed a similar model at NIPS this year.
- They provide a marginal sampler for posterior inference.
- Unfortunately, this sampler is incorrect.
- A similar error exists in [Lin and Fisher, 2012]
- At a high level:



  - One can superimpose 3 CRMs to construct 2 dependent RPMs, each marginally an NRM.
  - However, given observations from one , the other is *no longer* an NRM.
  - It becomes a *mixture* of NRMs.

Inference (a marginal sampling approach)

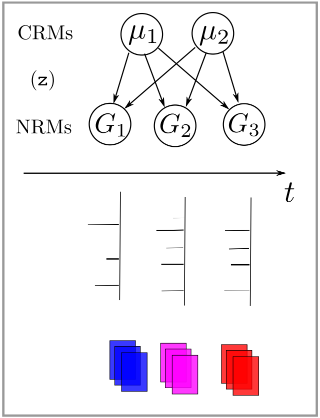
- Following [James et al., 2005], introduce auxiliary variables u_t $\forall t \in \mathcal{T}$
- Conditionally marginalize the CRMs μ_r to obtain a generalized CRP
- Alternately resample partition given $\{u_t\}$, and then $\{u_t\}$ given partition.
- Works for SNRM and MNRM, but impractical for TNRM.

Inference (a slice sampling approach)

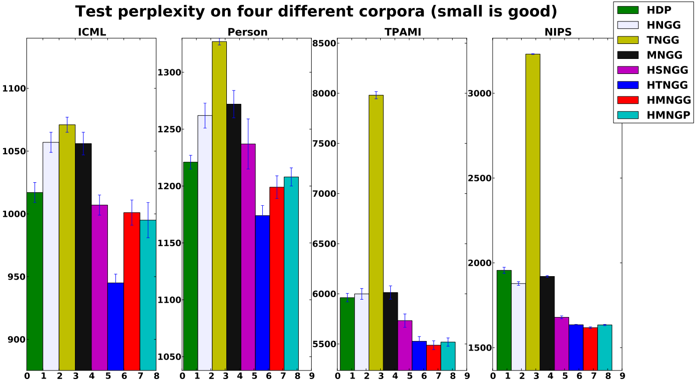
- Following [Walker, 2007], introduce auxiliary variables $s_r \; \forall r \in \mathcal{R}$
- Instantiate atoms of the CRM μ_r larger than s_r
- Conditionally sample partition of observations, and associated $z's$
- Alternately resample μ_r given $\{s_r\}$, and then $\{s_r\}$ given μ_r.

Application: Document modelling

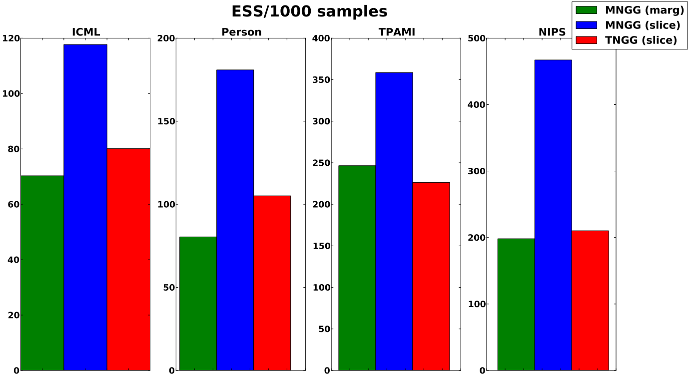
- Four corpora of documents (ICML, Person, TPAMI, NIPS)
- Documents organized by year.
- Largest corpus: NIPS
  - 17 years, 2483 documents, 3.28M words and a vocabulary of 14K
- Use a nonparametric topic model.



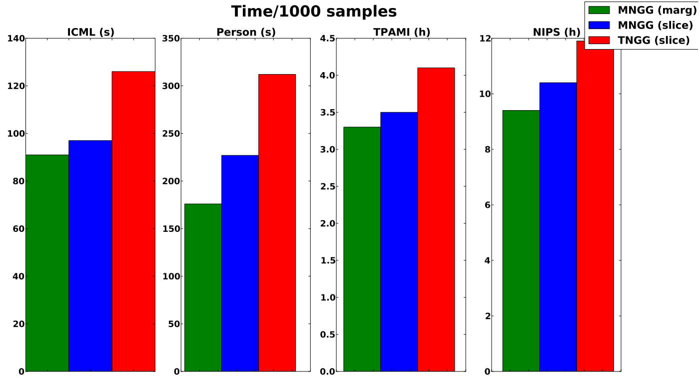
Application to document modelling



Application to document modelling



Application to document modelling



Thank you!

Bibliography I

Foti, N. J., Futoma, J., Rockmore, D., and Williamson, S. A. (2012).
A unifying representation for a class of dependent random measures.
Technical Report arXiv:1211.4753, Dartmouth College and CMU, USA.

Griffin, J. E., Kolossiatis, M., and Steel, M. F. J. (2013).
Comparing distributions by using dependent normalized random-measure mixtures.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), pages n/a–n/a.

James, L. F., Lijoi, A., and Pruenster, I. (2005).
Bayesian inference via classes of normalized random measures.
ICER Working Papers - Applied Mathematics Series 5-2005, ICER - International Centre for Economic Research.

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

Lijoi, A., Nipoti, B., and Pruenster, I. (2012).
Bayesian inference with dependent normalized completely random measures.
Technical report.

Lin, D., Grimson, E., and Fisher, J. (2010).
Construction of dependent Dirichlet processes based on Poisson processes.
In *NIPS*.

Lin, D. H. and Fisher, J. (2012).
Coupling nonparametric mixtures via latent Dirichlet processes.
In *NIPS*.

MacEachern, S. (1999).
Dependent nonparametric processes.
In *Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association.

Bibliography II

Nipoti, B. (2010).
Transformations of dependent completely random measures.
Technical report.

Pitman, J. (2003).
Poisson-kingman partitions.
In *of Lecture Notes-Monograph Series*, pages 1–34.

Pitman, J. and Yor, M. (1997).
The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator.
Annals of Probability, 25:855–900.

Rao, V. and Teh, Y. W. (2009).
Spatial normalized gamma processes.
In *Advances in Neural Information Processing Systems*.

Walker, S. G. (2007).
Sampling the Dirichlet mixture model with slices.
Communications in Statistics - Simulation and Computation, 36:45.