Dependent Normalized Random Measures

Changyou Chen, Vinayak Rao, Wray Buntine, Yee Whye Teh

changyou.chen@nicta.com.au vrao@gatsby.ucl.ac.uk
wray.buntine@nicta.com.au y.w.teh@stats.ox.ac.uk

ANU, NICTA, Duke, Oxford

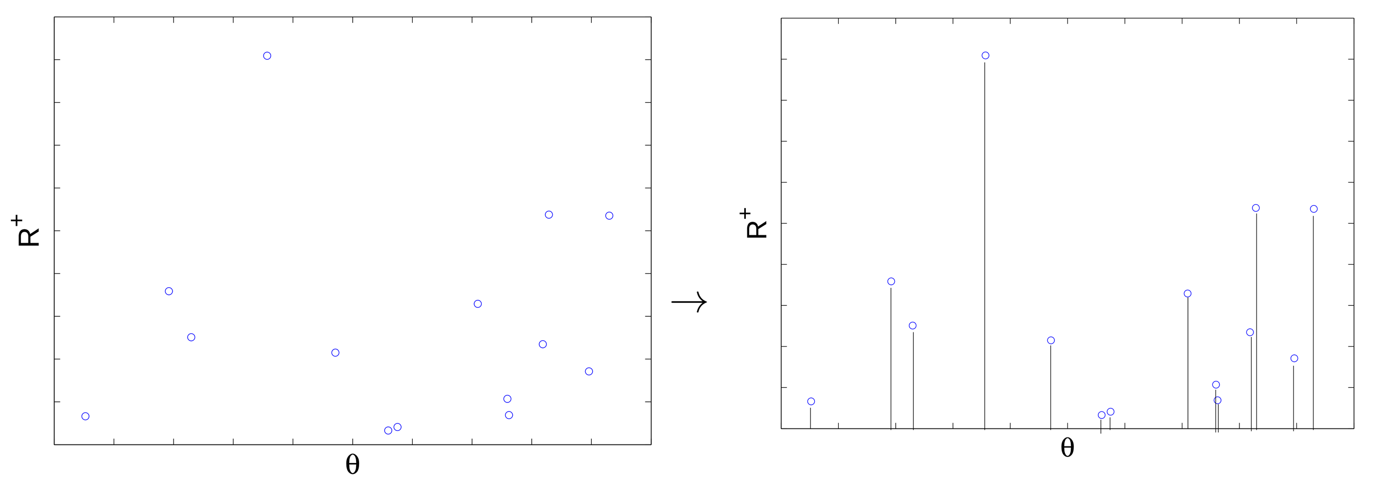
Contribution

- Propose two constructions of Dependent Normalized Random Measures related to [RaoTeh09, LinFisher12]:
  - Mixed Normalized Random Measures
  - Thinned Normalized Random Measures
- Analyze their distributional properties
- Analyze their distributional properties and posterior structures
- Provide alternatives to dependent DP and IBP
- Application to time series dynamic topic modeling

Normalized Random Measures

Completely Random Measure (CRM): Let $\mathbb{S} = R^+ \times \Theta$, a CRM $\tilde{\mu}$ is defined as a linear functional of the Poisson random measure $N(\cdot)$ (the intensity of the Poisson process $\nu(\cdot)$ is the Lévy intensity of $\tilde{\mu}$)

$$\tilde{\mu}(B) = \int_{\mathbb{R}^+ \times B} t N(\mathrm{d}t, \mathrm{d}\theta), \forall B \in \mathcal{B}(\Theta).$$



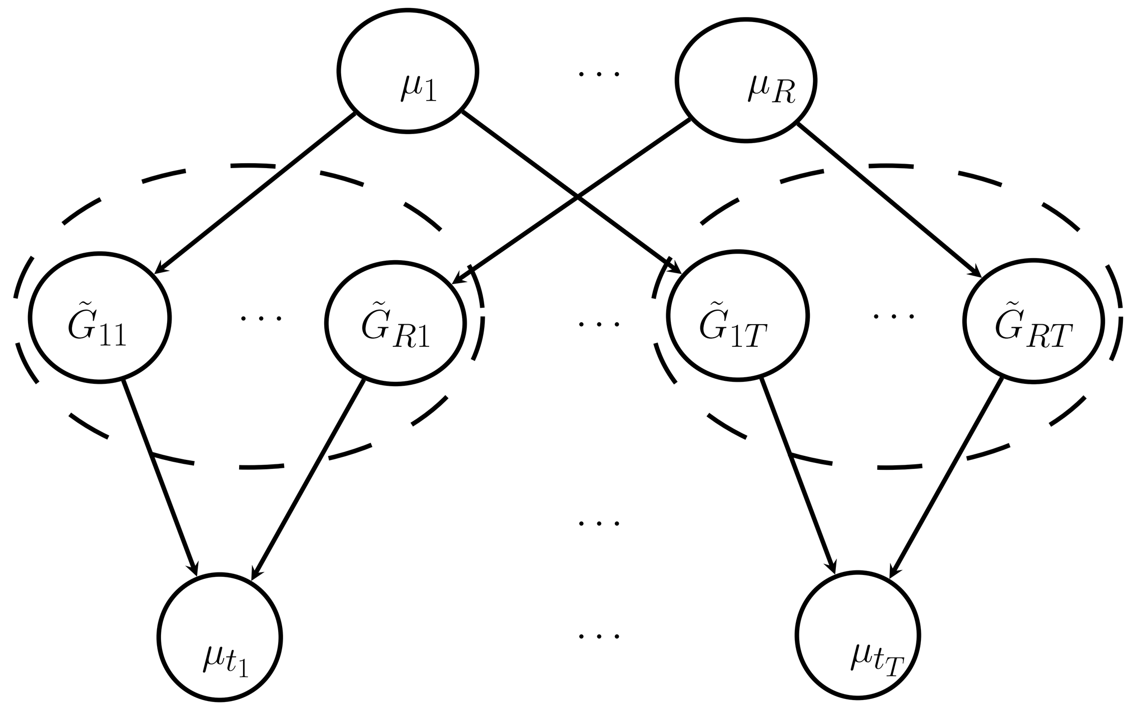
Poisson processes: $N(A) = \sum_{(w,\theta)\in A} \delta_{(w,\theta)}$

Completely random measures: $\tilde{\mu}(A) = \sum_{(w,\theta)\in A} w\delta_\theta$

Normalized Random Measure (NRM): An NRM is obtained by normalizing the CRM $\tilde{\mu}$ as: $\mu = \frac{\tilde{\mu}}{\tilde{\mu}(\Theta)}$. A normalized generalized Gamma process (NGG) is an NRM with Lévy measure being $\frac{e^{-bt}}{t^{1+a}} H(\mathrm{d}\theta), b > 0, 0 < a < 1$.

Graphical Construction of dNRMs

- $\mu_1, \cdots, \mu_R$: R independent NRMs, each for a *Region* r.
- $\mu_{t_1}, \cdots, \mu_{t_T}$: the constructed dependent NRMs, each for a *Time* t_i.



dNRM-1: Mixed Normalized Random Measures

- Construction by weighting:

$$\tilde{\mu}_r(\mathrm{d}\theta) = \int_{\mathbb{R}^+ \times \tilde{R}_r} w\mathcal{N}(\mathrm{d}w, \mathrm{d}\theta, \mathrm{d}a), \quad \text{for each region } r$$

$$\tilde{\mu}_t(\mathrm{d}\theta) = \sum_{r=1}^{\#\mathcal{R}} q_{rt}\tilde{\mu}_r(\mathrm{d}\theta) \quad \text{for each time } t$$

$$\mu_t(\mathrm{d}\theta) = \frac{1}{Z_t}\tilde{\mu}_t(\mathrm{d}\theta), \text{ where } Z_t = \tilde{\mu}_t(\Theta) \quad \text{for each time } t$$

dNRM-2: Thinned Normalized Random Measures

- Construction by thinning:

$$\tilde{\mu}_r(\mathrm{d}\theta) = \int_{\mathbb{R}^+ \times \tilde{R}_r} w\mathcal{N}(\mathrm{d}w, \mathrm{d}\theta, \mathrm{d}a), \quad \text{for each region } r$$

$$z_{rtk} \sim \text{Bernoulli}(q_{rt}), \quad \text{for each atom } k$$

$$\hat{\mu}_t(\mathrm{d}\theta) = \sum_{k=1}^{\infty} z_{rtk} w_{rk} \delta_{\theta_{rk}}, \quad \text{for each time } t$$

$$\mu_t(\mathrm{d}\theta) = \frac{1}{Z_t}\hat{\mu}(\mathrm{d}\theta), \text{ where } Z_t = \tilde{\mu}_t(\Theta) \quad \text{for each time } t$$

Distributional Properties

μ_t's in both MNRM and TNRM are marginally Normalized Random Measures, with Lévy intensities having the following forms:

MNRM

$$\nu_t(w, \theta) = \sum_{r=1}^{R} \nu_r(w/q_{rt}, \theta)/q_{rt}$$

TNRM

$$\nu_t(w, \theta) = \sum_{r=1}^{R} q_{rt}\nu_r(w, \theta)$$

Marginal VS Slice Sampler with Effective Sample Sizes

	ICML$_2$	Person$_2$	NIPS$_2$
Models	ESS (Ave/Med/Min) | Time	ESS (Ave/Med/Min) | Time	ESS (Ave/Med/Min) | Time
HMNGG	57.4/52.5/7.3 | 66s	119.4/102.0/3.1 | 1.0h	111.1/73.8/3.3 | 1.5h
HMNGG$_s$	125.4/112.5/15.0 | 69s	212.9/212.0/5.9 | 1.1h	205.2/203.0/5.5 | 1.9h
HTNGG	50.3/46.9/3.0 | 71s	144.8/170.6/4.2 | 1.3h	119.1/130.0/2.8 | 2.3h
HTNGG$_s$	94.9/90.9/4.0 | 76s	153.2/113.5/2.7 | 1.1h	176.1/151.0/3.3 | 1.9h

Topic Modeling: Perplexities

Datasets	ICML		Person	
Models	train perplexity	test perplexity	train perplexity	test perplexity
HDP	580 ± 6	1017 ± 8	4541 ± 33	5962 ± 43
HNGG	575 ± 5	1057 ± 8	4565 ± 60	5999 ± 54
TNGG	681 ± 23	1071 ± 6	5815 ± 122	7981 ± 36
MNGG	569 ± 6	1056 ± 9	4560 ± 63	6013 ± 66
HSNGG	550 ± 5	1007 ± 8	4324 ± 77	5733 ± 66
HTNGG	572 ± 7	$\mathbf{945 \pm 7}$	4196 ± 29	5527 ± 47
HMNGG	$\mathbf{535 \pm 6}$	1001 ± 10	$\mathbf{4083 \pm 36}$	$\mathbf{5488 \pm 44}$
HMNGP	561 ± 10	995 ± 14	4118 ± 45	5519 ± 41

Conditional Posterior of MNRM

- MNRM has a nice marginal posterior.
- Conditioned on some auxiliary variables u_t's[a], the posterior of MNRM is a generalization of a CRP via the following prediction rules:

$$p(s_{tl} = k, g_{tl} = r | \text{others}\}) \propto \begin{cases} \frac{q_{rt}(n_{\cdot rk}^{\setminus tl} - \sigma)}{1 + \sum_{t'} q_{rt'} u_{t'}} F_{rk}^{\setminus tl}(x_{tl}), & \text{if } k \text{ already exists,} \\ \sigma \left(\sum_{r'} \frac{M_{r'}}{(1 + \sum_{t'} q_{r't'} u_{t'})^{1-\sigma}} \right) \int_\Theta F(x_{tl}|\theta) H(\theta) \mathrm{d}\theta \,, & \end{cases}$$

where $F_{rk}^{\setminus tl}(x_{tl})$ is the conditional density of the observations.

- This allows a marginal sampler as well as a slice sampled to be developed.

[a] see the paper for details.

Conditional Posterior Lévy Measure of TNRM

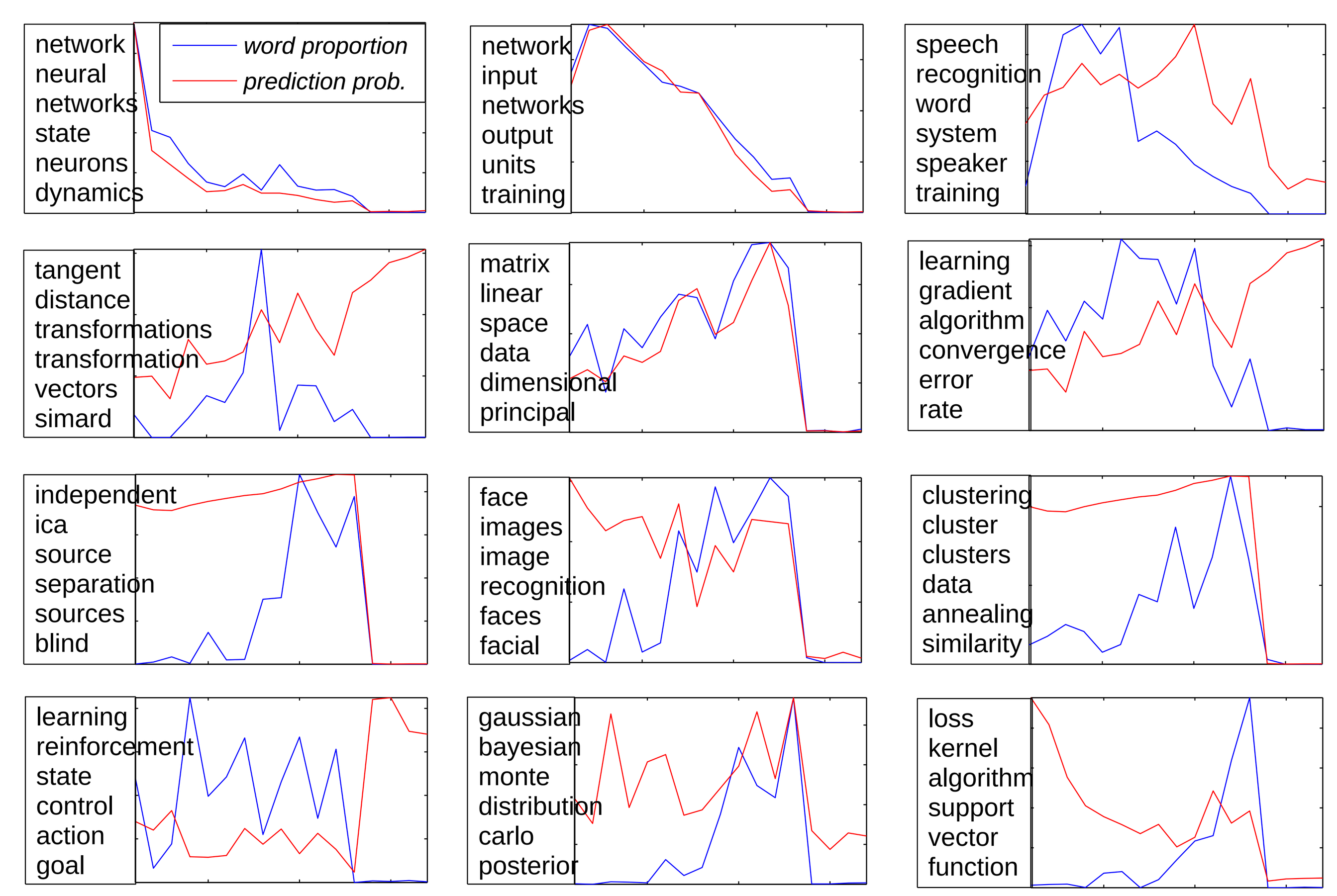
- Posterior structure of TNRM is complex, *i.e.*, it is equivalent to a NRM mixture of 2^T independent NRMs, so the complexity increases exponentially fast with #times T.
- Marginal sampler for TNRM is infeasible, thus a slice sampler is needed relying on the following conditional posterior of TNRM (built on Poisson process partition calculus [James05]):

Theorem 1 *Given observations, some auxiliary variables u_t for each ν_t, the points in ν_r without observations are distributed as a CRM with Lévy measure*

$$\nu_r'(\mathrm{d}w, \mathrm{d}\theta) = \prod_t \left(1 - q_{rt} + q_{rt}e^{-u_t w}\right) \nu_r(\mathrm{d}w, \mathrm{d}\theta) \,.$$

Remark Posterior inference for dependent DPs via thinning can not be performed via the standard Chinese restaurant processes prediction rules.

Topic Evolution on NIPS with HMNRM



References

[RaoTeh09] Rao, V., Teh, Y. W.: Spatial normalized Gamma processes. NIPS (2009)

[LinFisher12] Lin, D., Fisher, J.: Coupling Nonparametric Mixtures via Latent Dirichlet Processes. NIPS (2012)

[James05] James, L. F.: Poisson process partition calculus with an application to Bayesian Lévy moving averages. Anna. Stats. (2005)