Bayesian non-parametric approaches for dependent processes

Kaspar Märtens Xenia Miscouridou Paul Vanetti

11th March 2016

Random Measures

Completely Random Measures

- Distributions over random measures defined on a measurable space Θ.
- For disjoint subsets $A_1, ..., A_r \in \Theta$ and a realization $G \sim CRM$, the rv's $G(A_1), ..., G(A_r)$ are independent.
- Poisson Process representation
- e.g. Gamma Process, Beta Process

Normalized Random Measures

- obtained via normalization of a CRM
- G is a NRM if $G = \frac{G'}{G'(\Theta)}$, for $G' \sim CRM$
- e.g. Dirichlet Process is a Normalized Gamma Process

Dirichlet Process

- DP$(\alpha; G_0)$, on Θ
  concentration parameter $\alpha > 0$,
  base measure G_0 on Θ
- A realization $G \sim DP$ can be written as

  $$G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i} \qquad \text{(i.e. a.s. atomic)}$$

  where $\theta_k \sim G_0$ and π_k obtained via stick breaking:

  $$\pi_k = V_k \prod_{l<k} (1 - V_l),\ V_k \sim \text{Beta}(1, \alpha)$$

- For measurable disjoint sets $\{A_1, ..., A_r\} \in \Theta$,
  $(G(A_1), ..., G(A_r)) \sim \text{Dir}(\alpha G_0(A_1), ..., \alpha G_0(A_r))$

Exchangeable Priors for Bayesian nonparametric models

- Dirichlet Process
- Chinese Restaurant Process
- Pitman - Yor Process
- Indian Buffet Process

EXCHANGEABILITY ASSUMPTION

For a sequence $\{x_1, x_2 ...\}$ to be exchangeable:

$$\mathbb{P}\{(x_1, ... x_n)\} = \mathbb{P}\{(x_{\sigma(1)}, ..., x_{\sigma(n)})\},$$

for any σ permutation of $\{1, .., n\}$

Dependent non-parametric random processes

- Model data containing **spatial or time dependencies**.
- Extent nonparametric processes from distributions over random measures to distributions over **collections** of random measures.
- Families of random measures indexed by some **covariate**. $\{G^{(x)} : x \in \mathbb{X}\}$ e.g. $\mathbb{X} = \mathbb{R}^+$ for time or $\mathbb{X} = \mathbb{R}^2$ for space

$$G = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i} \rightarrow G^{(x)} = \sum_{i=1}^{\infty} \pi_i^{(x)} \delta_{\theta_i^{(x)}}$$

- The closer two covariates are in covariate space the greater amount of overlap among the corresponding processes .

Construction

- Different forms of dependencies
  1. dependence on atom location; the weights are shared
     $\forall x \ \pi_k^{(x)} = \pi_k$
  2. dependence on atom weight; the locations are shared
     $\forall x \ \theta_k^{(x)} = \theta_k$

- Different constructions; many are based on extensions of the Dirichlet process and we focus on two of those namely
  1. **spatial normalized gamma process**
  2. **probit stick-breaking process**

Spatial Normalized Gamma Process [1]

- Properties of gamma process:

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process
  - Marginalized gamma process $\rightarrow$ gamma process

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process
  - Marginalized gamma process $\rightarrow$ gamma process
  - Normalized gamma process $\rightarrow$ Dirichlet process

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process
  - Marginalized gamma process $\rightarrow$ gamma process
  - Normalized gamma process $\rightarrow$ Dirichlet process
- Define gamma process on extended space $\Theta \times \mathcal{Y}$

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process
  - Marginalized gamma process $\rightarrow$ gamma process
  - Normalized gamma process $\rightarrow$ Dirichlet process
- Define gamma process on extended space $\Theta \times \mathcal{Y}$
- For each observation:

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process
  - Marginalized gamma process $\rightarrow$ gamma process
  - Normalized gamma process $\rightarrow$ Dirichlet process
- Define gamma process on extended space $\Theta \times \mathcal{Y}$
- For each observation:
  - Choose subset of $\mathcal{Y}$ based on some covariate (time, space)

Spatial Normalized Gamma Process [1]

- Properties of gamma process:
  - Subset of gamma process $\rightarrow$ gamma process
  - Marginalized gamma process $\rightarrow$ gamma process
  - Normalized gamma process $\rightarrow$ Dirichlet process
- Define gamma process on extended space $\Theta \times \mathcal{Y}$
- For each observation:
  - Choose subset of $\mathcal{Y}$ based on some covariate (time, space)
  - Marginalize and normalize to yield DP

Spatial Normalized Gamma Process

- Example: time series.

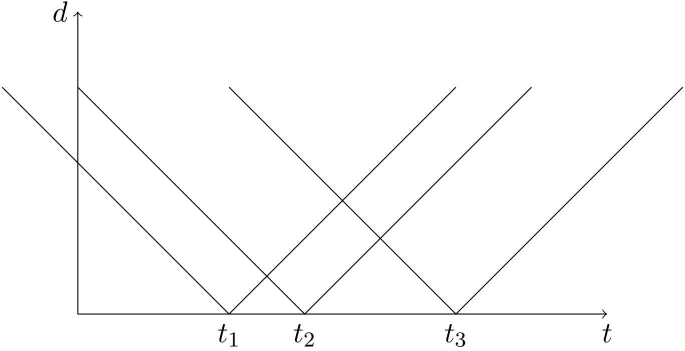
Spatial Normalized Gamma Process

- Example: time series.
  - Set extended space over time and duration (t, d)

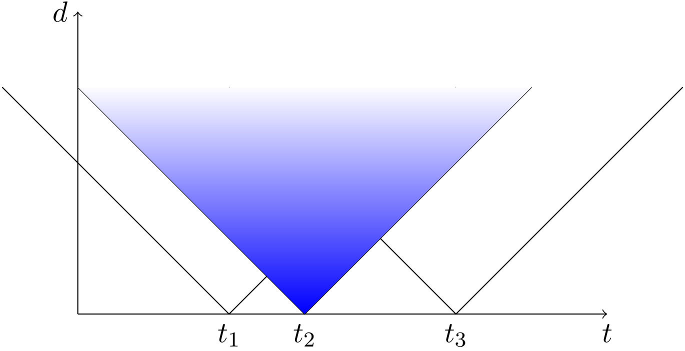
Spatial Normalized Gamma Process

- Example: time series.
  - Set extended space over time and duration (t, d)
  - Region for time t_i: $\{(t, d) : t - d < t_i < t + d\}$

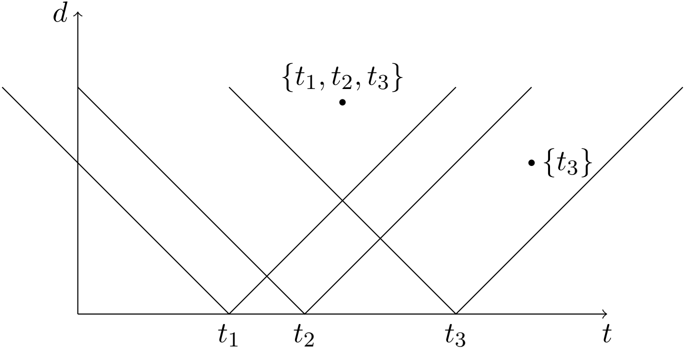
Spatial Normalized Gamma Process



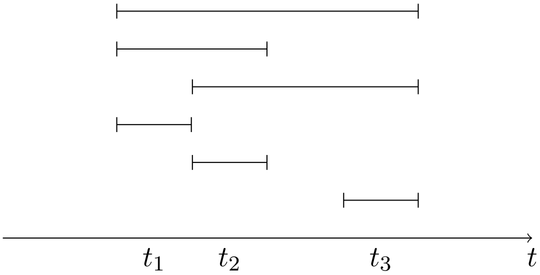
Spatial Normalized Gamma Process



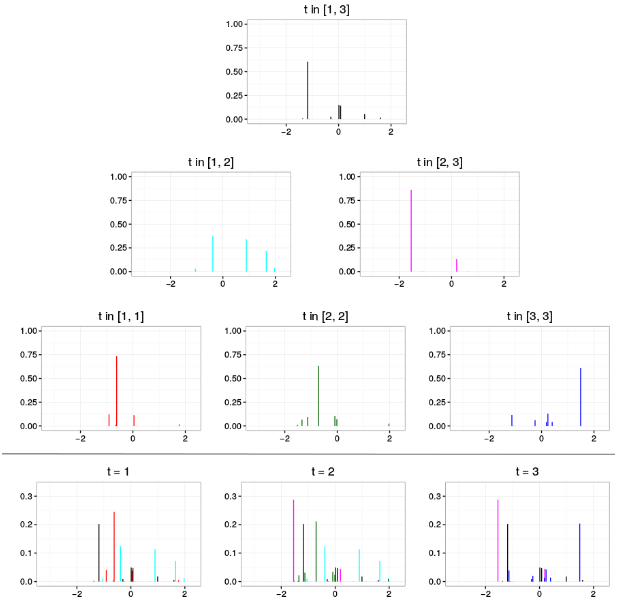
Spatial Normalized Gamma Process



Spatial Normalized Gamma Process



Spatial Normalized Gamma Process



Dependence via stick-breaking

Recall the stick-breaking construction

$$\pi_k = V_k \prod_{i<k} (1 - V_i)$$

Idea: introduce dependence on x, by replacing V_k with $V_k^{(x)}$

Dependence via stick-breaking

Recall the stick-breaking construction

$$\pi_k = V_k \prod_{i<k} (1 - V_i)$$

Idea: introduce dependence on x, by replacing V_k with $V_k^{(x)}$

Several ways to model $V_k^{(x)}$, e.g.

- Kernel stick-breaking process
- Probit stick-breaking process

Kernel stick-breaking process

Idea: associate each stick to a covariate location μ_k, and specify

$$V_k^{(x)} = U_k \, K(x, \mu_k)$$

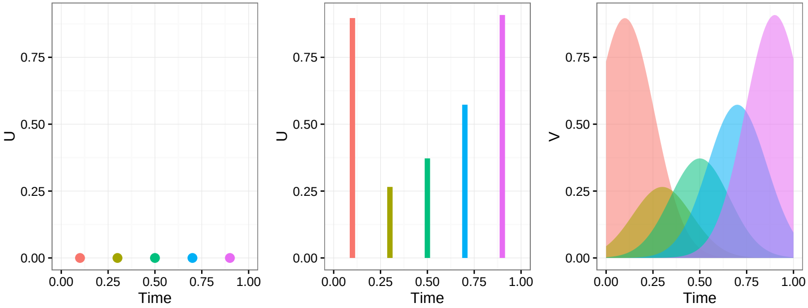
for some kernel function $K(\cdot, \cdot)$

Kernel stick-breaking process

Idea: associate each stick to a covariate location μ_k, and specify

$$V_k^{(x)} = U_k \, K(x, \mu_k)$$

for some kernel function $K(\cdot, \cdot)$



Probit stick-breaking process

Idea: specify

$$V_k^{(x)} = \Phi(\alpha_k^{(x)})$$

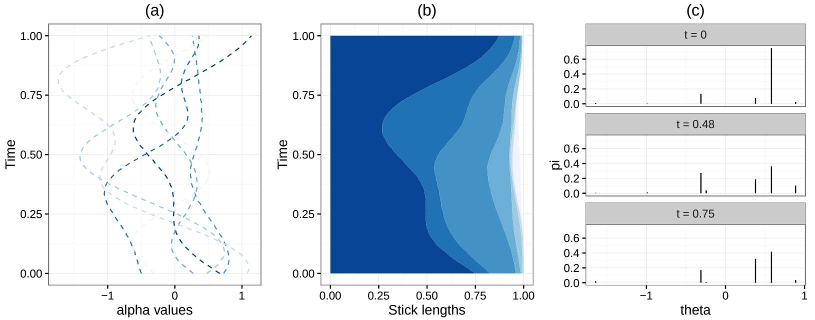
where $\alpha_k^{(x)}$ is a latent Gaussian process.

Probit stick-breaking process

Idea: specify

$$V_k^{(x)} = \Phi(\alpha_k^{(x)})$$

where $\alpha_k^{(x)}$ is a latent Gaussian process.



References

Vinayak Rao and Yee W Teh.
Spatial normalized gamma processes.
In *Advances in neural information processing systems*, pages 1554–1562, 2009.