1 Motivation

- We want to model the birth-death process of topic evolution.
- We want to model the power-law phenomena appeared in most of natural datasets, e.g., text datasets.

2 Normalized Random Measures

**Poisson Processes**: A Poisson process on $S$ is a random subset $\{N(t), t \geq 0\}$ such that $N(t)$ is the number of points of $\{N(t), t \geq 0\}$, then $X(\cdot)$ is a Poisson random variable with mean $\mu(A)$, and $\{N(t), t \geq 0\}$ and $\{N(\cdot), \cdot \geq 0\}$ are independent if $A_1, \ldots, A_n$ are disjoint.

**Completely Random Measures (CRM)**: Let $S = \mathbb{R}^d \times \mathcal{F}$, a CRM is defined as a linear functional of the Poisson random measure $N(\cdot)$ (called $\nu$) the Lebesgue measure of $\mu$

$$\mu(B) = \int_B N(dx, d\gamma), \forall B \in \mathcal{B}(\mathbb{R}^d)$$

**Normalized Random Measures (NRM)**: An NRM is obtained by normalizing the CRM $\mu$ as

$$\tilde{\mu}(B) = \frac{\mu(B)}{\mu(S)}$$

**Normalized Generalized Gamma Process (NGG)**: An normalized generalized Gamma process (NGG) is an NRM with $\mu$-measure $\sum_{i=1}^n \delta_{\gamma_i}$, where $0 < \gamma_i < 1, n > 0$.

3 The three Dependency Operations

**Superposition of NRMs**:

Given $n$ independent NRMs $\mu_1, \ldots, \mu_n$ on $\mathbb{R}^d$, the superposition ($\mu$) is defined as

$$\mu_1 \oplus \mu_2 \oplus \cdots \oplus \mu_n = \sum_{i=1}^n \mu_i$$

**Subsampling of NRMs**:

Given a CRM $\mu = \sum_{i=1}^n \delta_{\gamma_i}$ on $\mathbb{R}^d$, and a Bernoulli parameter $\epsilon \in [0, 1]$, the subsampling of $\mu$ is defined as

$$S(\mu) = \sum_{i=1}^n \frac{\epsilon_i}{\sum_{j=1}^n \epsilon_j} \delta_{\gamma_i}$$

**Point transition of NRMs**:

Given a CRM $\mu = \sum_{i=1}^n \delta_{\gamma_i}$ on $\mathbb{R}^d$, the point transition of $\mu$ is to draw atoms $\epsilon_i$ from a transformed base measure to yield a new NRM as

$$T(\mu) = \sum_{i=1}^n \epsilon_i \delta_{\gamma_i}$$

4 Sampling

The statistics we are interested in are:

- $n_{i,j}$: the customer $i$ in the $j$th restaurant.
- $d_{i,j}$: the dish that $n_{i,j}$ is eating.
- $k_{i,j}$: the number of customers in $\mu_\theta$ eating dish $k$.
- $\theta_\mu = \sum_{i,j} k_{i,j}$.

At each time frame $m$, we do:

- Slice sample $x_{m,j}$ (ends up finite jumps).
- Subsample $x_{m,j}$ by inheriting from $x_{m,1:m}$ with Bernoulli trials.
- Construct $x_{m,j}$ by normalizing $x_{m,j}$.
- Sample $x_{m,j}$ using a generalized Blackwell-MacQueen sampling scheme for the hierarchical NRM.

5 Experiments

Evaluated on 9 datasets including news, blogs, academic and Twitter collections. See Figure 1, 2, 3 for demonstration and Table 1 for comparison.

![Figure 1: Left: Power-law phenomena in NGG: Right: topic evolution on JMLR. Shows a late developing topic on software, before during and after the start of MLOSS.org in 2008.](image1.png)

![Figure 2: Topic evolution on Twitter. Words in red have increased, and blue decreased.](image2.png)

![Figure 3: Training log-likelihoods influenced by the subsampling rate $\epsilon$. From top-down, left to right are the results on ICML, JMLR, TPMII, Person, Twitter, Twitter, Twitter, and BDT datasets, respectively.](image3.png)