Supplementary Material for Incremental Tree-Based Inference with Dependent Normalized Random Measures

1 Review of Dirichlet Process Mixture Models

In this section, we briefly review Dirichlet process mixture model (DPM) and derive its marginal likelihood. Let $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$ be a dataset to model. The generative process of $\boldsymbol{X}$ under DPM is written as:

$$\mu|\alpha, H \sim \mathrm{DP}(\alpha, H) \tag{1}$$
$$\theta_i|\mu \sim \mu \tag{2}$$
$$\boldsymbol{x}_i|\theta_i \sim L(\cdot|\theta_i). \tag{3}$$

Here, μ can be marginalized out to yield Chinese restaurant process (CRP), which is defined as a conditional distribution of θ_i given $\theta_1, \ldots, \theta_{i=1}$.

$$p(\theta_i|\theta_1, \ldots, \theta_{i-1}) = \frac{\alpha H + \sum_{j=1}^{i-1} \delta_{\theta_j}}{\alpha + i - 1}. \tag{4}$$

Without loss of generality, assume that $\{\theta_i\}_{i=1}^n$ are clustered into K distinct values $\{\theta_k^*\}_{k=1}^K$ as follows:

$$\theta_1 = \theta_2 = \cdots = \theta_{n_1} = \theta_1^* \tag{5}$$
$$\theta_{n_1+1} = \theta_{n_1+2} = \theta_{n_1+n_2} = \theta_2^* \tag{6}$$
$$\vdots \qquad \vdots \tag{7}$$
$$\theta_{n_1+\cdots+n_{K-1}+1} = \cdots = \theta_{n_1+\cdots+n_K} = \theta_K^*, \tag{8}$$

where $n_1 + \cdots + n_K = n$. By (4), the joint probability of $\{\theta_i\}_{i=1}^n$ is computed as

$$\begin{aligned}
p(\{\theta_i\}_{i=1}^n) &= \prod_{i=1}^n p(\theta_i|\theta_1, \ldots, \theta_{i-1}) \\
&= H(\theta_1^*) \times \frac{1}{\alpha+1} \times \cdots \times \frac{n_1 - 1}{\alpha + n_1 - 1} \\
&\quad \times \frac{\alpha H(\theta_2^*)}{\alpha + n_1} \times \frac{1}{\alpha + n_1 + 1} \times \cdots \times \frac{n_2 - 1}{\alpha + n_1 + n_2 - 1} \\
&\quad \vdots \\
&\quad \frac{\alpha H(\theta_K^*)}{\alpha + n_1 + \cdots + n_{K-1}} \times \frac{1}{\alpha + n_1 + \cdots + n_{K-1} + 1} \times \cdots \times \frac{n_1 + \cdots + n_K - 1}{\alpha + n_1 + \cdots + n_K - 1} \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \prod_{k=1}^K \alpha \Gamma(n_k) H(\theta_k^*).
\end{aligned} \tag{9}$$

From (9), one can see that the joint probability of $\{\theta_i\}_{i=1}^n$ is invariant to the ordering of items or cluster labels. Hence, we can write (9) as a proability of a partition π of I_n and unique parameters $\{\theta_c^*\}_{c\in\pi}$ assigned to each cluster of the partition:

$$p(\{\theta_i\}_{i=1}^n) = p(\pi, \{\theta_c^*\}_{c\in\pi}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \prod_{c\in\pi} \alpha\Gamma(|c|)H(\theta_c^*). \tag{10}$$

The same result can be derived using **Theorem 2** of in our paper. By **Theorem 2**, the joint probability of $\{\theta_i\}$ and auxiliary variable ξ is

$$p(\pi, \{\theta_c^*\}, \xi) = \frac{\xi^{n-1}e^{-\psi_\rho(\xi)}}{\Gamma(n)} \prod_{c\in\pi} \kappa_\rho(|c|, \xi)H(\theta_c^*). \tag{11}$$

In case of DP, $\psi_\rho(\xi)$ and $\kappa_\rho(|c|, \xi)$ are computed as

$$\psi_\rho(\xi) = \int_{\mathbb{R}^+} (1 - e^{-\xi w})\alpha w^{-1} e^{-w} dw = \alpha \log(1+\xi) \tag{12}$$

$$\kappa_\rho(|c|, \xi) = \int_{\mathbb{R}^+} w^{|c|} e^{-\xi w} \alpha w^{-1} e^{-w} dw = \frac{\alpha\Gamma(|c|)}{(1+\xi)^{|c|}}. \tag{13}$$

Hence, we have

$$\begin{aligned} p(\pi, \{\theta_c^*\}, \xi) &= \frac{\xi^{n-1}}{\Gamma(n)(1+\xi)^{n+\alpha}} \prod_{c\in\pi} \alpha\Gamma(|c|)H(\theta_c^*) \\ &= \left(\frac{\xi}{1+\xi}\right)^{n-1} \left(\frac{1}{1+\xi}\right)^{\alpha+1} \frac{1}{\Gamma(n)} \prod_{c\in\pi} \alpha\Gamma(|c|)H(\theta_c^*). \end{aligned} \tag{14}$$

Let $\xi/(1+\xi) = v$. Then

$$\begin{aligned} p(\pi, \{\theta_c^*\}) &= \int p(\pi, \{\theta_c^*\}, \xi) d\xi \\ &= \int v^{n-1}(1-v)^{\alpha+1} \frac{1}{\Gamma(n)} \prod_{c\in\pi} \alpha\Gamma(|c|)H(\theta_c^*) \frac{1}{(1-v)^2} dv \\ &= \frac{1}{\Gamma(n)} \prod_{c\in\pi} \alpha\Gamma(|c|)H(\theta_c^*) \int v^{n-1}(1-v)^{\alpha-1} dv \\ &= \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \prod_{c\in\pi} \alpha\Gamma(|c|)H(\theta_c^*). \end{aligned} \tag{15}$$

Combined with the likelihood L, the joint probability of DPM is then written as:

$$p(\boldsymbol{X}, \pi, \{\theta_c^*\}_{c\in\pi}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \prod_{c\in\pi} \alpha\Gamma(|c|)L(\boldsymbol{X}^{(c)}|\theta_c^*)H(\theta_c^*), \tag{16}$$

and the marginal likelihood of $\boldsymbol{X}$ is

$$p(\boldsymbol{X}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \sum_\pi \prod_{c\in\pi} \alpha\Gamma(|c|)p(\boldsymbol{X}^{(c)}|\mathcal{H}_c). \tag{17}$$

2 Completely Random Measure

In this section, we provide more specific backgrounds on completely random measures (CRMs). Let us consider a measure space (Θ, Ω), where Ω is the σ-algebra of Θ. A random measure μ on (Θ, Ω) is a random variable whose values are measures on (Θ, Ω).

Definition 1. *A random measure μ on (Θ, Ω) is* **completly random** *if for any disjoint $A_1, \ldots, A_n \subset \Theta$, the random variables $\mu(A_1), \ldots, \mu(A_n)$ are independent.*

It has been shown [1] that a CRM is decomposed into the sum of three independent components: (1) a nonrandom measure; (2) a countable collection of nonnegative randomm masses at fixed locations, and (3) a countable collection of nonnegative random masses at random locations.

Theorem 1. *([1]) A CRM μ is decomposed into the sum of three independent components:*

$$\begin{aligned} \mu &= \mu_0 + \sum_j v_j \delta_{\theta_j^*} + \int_{\mathbb{R}^+} w \Pi(dw, \theta) \\ &= \mu_0 + \sum_j v_j \delta_{\theta_j^*} + \sum_k w_k \delta_{\theta_k}, \end{aligned} \tag{18}$$

where μ_0 is a nonrandom measure, $\{v_j\}$ are mutually independent random variables on $\mathbb{R}^+$ corresponding to random weights of fixed atoms $\{\theta_j^*\}$ of Θ, and $\Pi = \sum_k \delta_{(w_k, \theta_k)}$ is a Poisson process defined on a product space $\mathbb{R}^+ \times \Theta$ with Lévy intensity $\lambda(dw, d\theta)$.

Definition 2. *A CRM μ is said to be* **homogeneous** *if the Lévy intensity λ of the underlying Poisson process Π decomposes into a product of two intensities:*

$$\lambda(dw, d\theta) = \rho(dw) H(d\theta). \tag{19}$$

Throughout this article, we consider only homogeneous CRMs without nonrandom measures and fixed atoms, and we write

$$\mu = \int_{\mathbb{R}^+} \Pi(dw, \theta) = \sum_k w_k \delta_{\theta_k} \sim \mathrm{CRM}(\rho, H). \tag{20}$$

Without loss of generality, we assume that $\int_\Theta H(d\theta) = 1$. We also assume that ρ satisfies the following two conditions

$$\int_{\mathbb{R}^+} \rho(dw) = \infty, \quad \int_{\mathbb{R}^+} (1 - e^{-w}) \rho(dw) < \infty, \tag{21}$$

where the first condition ensures that μ has infinite atoms ($\mu = \sum_{k=1}^\infty w_k \delta_{\theta_k}$), and the second condition ensures that the total mass $\mu(\Theta) = \sum_{k=1}^\infty w_k$ is finite [2].

3 Posterior Analysis of Normalized Random Measure Mixture Models

In this section, we provide the proof of **Theorem 2** in our paper. The original proof can be found in [2, 3].

Theorem 2. *([2],* ***Theorem 2*** *in paper) Let $\mu \sim \mathrm{CRM}(\rho, H)$ on (Θ, Ω) and let $(\pi, \{\theta_c^*\}) \sim \widetilde{\mu}$. Introducing an auxiliary variable $\xi \sim \mathrm{Gamma}(n, \mu(\Theta))$, the posterior of μ is given by*

$$\mu | \xi, \pi, \{\theta_c^*\} = \bar{\mu} + \sum_{c \in \pi_n} w_c \delta_{\theta_c^*}, \tag{22}$$

where

$$\bar{\mu} \sim \mathrm{CRM}(\bar{\rho}, H), \quad \bar{\rho}(dw) \stackrel{\text{def}}{=} e^{-\xi w}\rho(dw),$$

and the posterior density over random weights of fixed atoms is given by

$$\begin{aligned} p(w_c|\xi, \pi, \{\theta_c^*\}) &= \frac{w_c^{|c|} e^{-\xi w_c} \rho(w_c)}{\kappa_\rho(|c|, \xi)}, \\ \kappa_\rho(m, \xi) &\stackrel{\text{def}}{=} \int_{\mathbb{R}^+} w^m e^{-\xi w} \rho(dw). \end{aligned}$$

The marginal distribution is given by

$$\begin{aligned} p(\pi, \{\theta_c^*\}, \xi) &= \frac{\xi^{n-1} e^{-\psi_\rho(\xi)}}{\Gamma(n)} \prod_{c\in\pi} \kappa_\rho(|c|, \xi) H(\theta_c^*), \\ \psi_\rho(\xi) &\stackrel{\text{def}}{=} \int_{\mathbb{R}^+} (1 - e^{-\xi w}) \rho(dw). \end{aligned}$$

The predictive distribution is given by

$$\theta|\xi, \pi, \{\theta_c^*\} \propto \kappa_\rho(1, \xi) H + \sum_{c\in\pi} \frac{\kappa_\rho(|c|+1, \xi)}{\kappa_\rho(|c|, \xi)} \delta_{\theta_c^*}.$$

Proof. The likelihood of $(\pi, \{\theta_c^*\}_{c\in\pi})$ given μ is

$$p(\pi, \{\theta_c^*\}_{c\in\pi}|\mu) = \mu(\Theta)^{-n} \prod_{c\in\pi} w_c^{|c|}, \tag{23}$$

where w_c is a mass of θ_c^*. We introduce an auxiliary variable ξ as follows:

$$\begin{aligned} p(\xi|\mu) &= \mathrm{Gamma}(\xi|n, \mu(\Theta)) \\ p(\pi, \{\theta_c^*\}_{c\in\pi}, \xi|\mu) &= \frac{\xi^{n-1}\exp(-\mu(\Theta)\xi)}{\Gamma(n)} \prod_{c\in\pi} w_c^{|c|}. \end{aligned} \tag{24}$$

We will compute the characteristic functional of the posterior process $\mu|\pi, \{\theta_c^*\}_{c\in\pi}, \xi$, which is given as:

$$\mathbb{E}\left[e^{-\mu(f)}|\pi, \{\theta_c^*\}_{c\in\pi}, \xi\right] = \frac{\mathbb{E}[p(\pi, \{\theta_c^*\}, \xi|\mu)e^{-\mu(f)}]}{\mathbb{E}[p(\pi, \{\theta_c^*\}, \xi|\mu)]}, \tag{25}$$

where f is an arbitrary bounded function on Θ and

$$e^{-\mu(f)} \stackrel{\text{def}}{=} \exp\left\{-\int f(\theta)\mu(d\theta)\right\}. \tag{26}$$

The denominator is a special case of the numerator when $f = 0$, so we first compute the numerator. Let Π be the underlying Poisson process of μ.

$$\begin{aligned} &\mathbb{E}\left[p(\pi, \{\theta_c^*\}, \xi|\mu)e^{-\mu(f)}\right] \\ &= \frac{\xi^{n-1}}{\Gamma(n)} \mathbb{E}\left[\exp\left\{-\int w(f(\theta)+\xi)\Pi(dw, d\theta)\right\} \prod_{c\in\pi} w_c^{|c|}\right] \end{aligned} \tag{27}$$

By applying the Palm formula or by **Lemma 2** in [4], we get

$$\propto \quad \mathbb{E}\Big[\exp\Big\{-\int w(f(\theta)+\xi)\Pi(dw,d\theta)\Big\}\Big]\prod_{c\in\pi}H(\theta_c^*)\int w_c^{|c|}\exp\{-w_c(f(\theta_c^*)+\xi)\}\rho(w_c) \tag{28}$$

By the Campbell's formula [5], this is computed as

$$= \quad \exp\Big\{\int(e^{-w(f(\theta)+\xi)}-1)\rho(dw)H(d\theta)\Big\}\prod_{c\in\pi}H(\theta_c^*)\int w_c^{|c|}\exp\{-w_c(f(\theta_c^*)+\xi)\}\rho(w_c) \tag{29}$$

If we set $f=0$, we get the denominator:

$$\begin{aligned} p(\pi,\{\theta_c^*\},\xi)=\mathbb{E}[p(\pi,\{\theta_c^*\},\xi|\mu)] \quad &\propto \quad \exp\Big\{\int(e^{-w\xi}-1)\rho(dw)H(d\theta)\Big\}\prod_{c\in\pi}H(\theta_c^*)\int w_c^{|c|}e^{-w_c\xi}\rho(w_c) \\ &= \quad \frac{\xi^{n-1}e^{-\psi_\rho(\xi)}}{\Gamma(n)}\prod_{c\in\pi}\kappa_\rho(|c|,\xi)H(\theta_c^*). \end{aligned} \tag{30}$$

Now dividing the numerator with the denominator gives

$$\begin{aligned} &\mathbb{E}\Big[e^{-\mu(f)}|\pi,\{\theta_c^*\}_{c\in\pi},\xi\Big] \\ &= \quad \exp\Big[\int(e^{-w(f(\theta)+\xi)}-e^{-w\xi})\rho(dw)H(d\theta)\Big]\prod_{c\in\pi}\int\frac{w_c^{|c|}\exp\{-w_c(f(\theta_c^*)+\xi)\}\rho(dw_c)}{\kappa_\rho(|c|,\xi)} \\ &= \quad \exp\Big[\int(e^{-wf(\theta)}-1)e^{-w\xi}\rho(dw)H(d\theta)\Big]\prod_{c\in\pi}\int\frac{w_c^{|c|}\exp\{-w_c(f(\theta_c^*)+\xi)\}\rho(dw_c)}{\kappa_\rho(|c|,\xi)} \end{aligned} \tag{31}$$

The first term inside the exponential is the characteristic functional of CRM $\bar{\mu}$ with Lévy measure $\bar{\rho}=e^{-\xi w}\rho(dw)$. The second terms inside the product is the characteristic functional of fixed atoms θ_c^* with random masses w_c with distribution

$$p(w_c)=\frac{w_c^{|c|}e^{-\xi w_c}\rho(w_c)}{\kappa_\rho(|c|,\xi)}. \tag{32}$$

□

4 Posterior analysis for Mixed Normalized Random Measure Mixtures

In this section, we describe the posterior analysis for mixed normalized random measure (MNRM) mixtures [6]. All the derivations and proofs can be found in the supplementary material of the original paper [4], and here we just rewrite them using our own notations.

4.1 Model

Let $\{\nu_r\}_{r=1}^R$ be mutually independent homogeneous CRMs on Θ. As in our paper, we will call these CRMs as *basis CRMs.*

$$\nu_r=\sum_{k=1}^\infty w_{rk}\delta_{\theta_{rk}}\sim\mathrm{CRM}(\rho_r,H_r)\text{ for }r=1,\dots,R. \tag{33}$$

Suppose that we are to model a dataset $\boldsymbol{X}_{1:T} = \bigcup_{t=1}^{T} \boldsymbol{X}_t$. The CRM μ_t to generate $\boldsymbol{X}_t$ is then represented as a mixture of the basis measures as follows:

$$\mu_t = \sum_{r=1}^{R} q_{t,r}\nu_r \text{ for } t = 1,\ldots,T. \tag{34}$$

Now for each data point $\boldsymbol{x}_{ti}$, a parameter θ_{ti} is drawn from its corresponding normalized random measure.

$$\{\theta_{ti}\}_{i=1}^{n_t}|\mu_t \overset{\text{iid}}{\sim} \widetilde{\mu}_t = \mu_t/\mu_t(\Theta). \tag{35}$$

Since all the θ_{ti} came from one of the (discrete) basis measures, $\{\{\theta_{ti}\}_{i=1}^{n_t}\}_{t=1}^{n_t}$ induces a partition π of the set of indices $I_{1:T}$. Moreover, each cluster c in π has its own indicator variable z_c which indicates that the parameter allocated to the cluster c had come from the basis measure ν_r. We write θ_c^* as a representation of the parameter allocated to the cluster c. Also, let w_c be the weight corresponding to θ_c^*.

$$\begin{aligned} \{\{\theta_{ti}\}_{i=1}^{n_t}\}_{t=1}^{T} &= (\pi, \{z_c, \theta_c^*\}_{c\in\pi}) && (36)\\ w_c &= \nu_{z_c}(\theta_c^*). && (37) \end{aligned}$$

4.2 Posterior processes

The likelihood of $(\pi, \{z_c, \theta_c^*\})$ given $\{\nu_r\}$ is written as

$$p(\pi, \{z_c, \theta_c^*\}|\{\nu_r\}) = \prod_{t=1}^{T} \frac{\prod_{c\in\pi}(q_{tz_c}w_c)^{|c\cap I_{t,n_t}|}}{(\sum_r q_{tr}\nu_r(\Theta))^{n_t}}. \tag{38}$$

Now we introduce a set of auxiliary Gamma variables $\{\xi_t\}_{t=1}^{T}$ to obtain closed-form posterior processes.

$$\begin{aligned} \xi_t|\{\nu_r\} &\sim \text{Gamma}\bigg(n_t, \sum_{r=1}^{R} q_{t,r}\nu_r(\Theta)\bigg) && (39)\\ p(\pi, \{z_c, \theta_c^*\}, \{\xi_t\}|\{\nu_r\}) &= \prod_{t=1}^{T} \frac{\xi_t^{n_t-1}}{\Gamma(n_t)} \prod_{r=1}^{R} \exp\bigg(- \nu_r(\Theta) \sum_{t=1}^{T} q_{t,r}\xi_t\bigg) \prod_{c\in\pi} \bar{q}_{c,z_c} w_c^{|c|}, && (40) \end{aligned}$$

where $\bar{q}_{cz_c} \overset{\text{def}}{=} \prod_t q_{tz_c}^{|c\cap I_{t,n_t}|}$.

Theorem 3. *The posterior process* $\nu_r|\pi, \{z_c, \theta_c^*\}, \{\xi_t\}$ *is written as*

$$\nu_r|\pi, \{z_c, \theta_c^*\}, \{u_t\} = \bar{\nu}_r + \sum_{\substack{c\in\pi \\ z_c=r}} \mathbb{I}[z_c = r] w_c \delta_{\theta_c^*}, \tag{41}$$

where

$$\begin{aligned} \bar{\rho}_r(dw) &\overset{\text{def}}{=} \exp\bigg\{ - w \sum_{t=1}^{T} q_{t,r}\xi_t\bigg\}\rho_r(dw) && (42)\\ \bar{\nu}_r &\sim \text{CRM}(\bar{\rho}_r, H_r) && (43)\\ p(w_c|\ldots) &= \frac{w_c^{|c|}\exp(-w_c\sum_t q_{t,r}\xi_t)\rho_r(w_c)}{\kappa_{\rho_r}(|c|, \sum_t q_{t,r}\xi_t)}. && (44) \end{aligned}$$

Moreover, the marginal distribution is written as

$$p(\pi, \{z_c, \theta_c^*\}, \{\xi_t\}) = \prod_{t=1}^{T} \frac{\xi_t^{n_t-1}}{\Gamma(n_t)} \prod_{r=1}^{R} \exp\bigg\{ - \psi_{\rho_r}\bigg(\sum_{t=1}^{T} q_{t,r}\xi_t\bigg)\bigg\} \prod_{c\in\pi} \bar{q}_{c,z_c}\kappa_{\rho_r}\bigg(|c|, \sum_{t=1}^{T} q_{t,z_c}\xi_t\bigg) H_{z_c}(\theta_c^*). \tag{45}$$

Proof. We will compute the characteristic functional of the process $\nu_r|\pi, \{z_c, \theta_c^*\}, \{\xi_t\}$ for an arbitrary bounded measurable function f on Θ.

$$\mathbb{E}\Big[e^{-\nu_r(f)}|\pi, \{z_c, \theta_c^*\}, \{u_t\}\Big] = \mathbb{E}\Big[\exp\Big\{-\int f(\theta)\nu_r(d\theta)\Big\}\Big|\pi, \{z_c, \theta_c^*\}, \{\xi_t\}\Big]. \tag{46}$$

One can easily see that this characteristic functional is written as

$$\mathbb{E}\Big[e^{-\nu_r(f)}|\pi, \{z_c, \theta_c^*\}, \{\xi_t\}\Big] = \frac{\mathbb{E}[p(\pi, \{z_c, \theta_c^*\}, \{\xi_t\}|\nu_r)e^{-\nu_r(f)}]}{\mathbb{E}[p(\pi, \{z_c, \theta_c^*\}, \{\xi_t\}|\nu_r)]}, \tag{47}$$

and the denominator is a special case of the numerator when $f = 0$ for all $\theta \in \Theta$. Hence we focus on computing the numerator. Let Π be the underlying Poisson process of ν_r.

$$\begin{aligned}
&\mathbb{E}\Big[p(\pi, \{z_c, \theta_c^*\}, \{\xi_t\}|\nu_r)e^{-\nu_r(f)}\Big] \\
&= \prod_{t=1}^{T} \frac{\xi_t^{n_t-1}}{\Gamma(n_t)} \prod_{r' \neq r} \exp\Big\{-\nu_{r'}(\Theta)\Big(\sum_{t=1}^{T} q_{t,r'}\xi_t\Big)\Big\} \prod_{c\in\pi} \bar{q}_{c,z_c} \prod_{\substack{c\in\pi \\ z_c\neq r}} w_c^{|c|} \\
&\quad \times\mathbb{E}\Big[\exp\Big\{-\int w\Big(f(\theta) + \sum_{t=1}^{T} q_{t,r}\xi_t\Big)\Pi(dw, d\theta)\Big\} \prod_{\substack{c\in\pi \\ z_c=r}} w_c^{|c|}\Big].
\end{aligned} \tag{48}$$

By applying the Palm formula or simply by **Lemma 2** in [4], we get

$$\begin{aligned}
\propto\ & \mathbb{E}\Big[\exp\Big\{-\int w\Big(f(\theta) + \sum_{t=1}^{T} q_{t,r}\xi_t\Big)\Pi(dw, d\theta)\Big\}\Big] \\
& \times \prod_{\substack{c\in\pi \\ z_c=r}} H_r(\theta_c^*) \int w_c^{|c|} \exp\Big\{-w_c\Big(f(\theta_c^*) + \sum_{t=1}^{T} q_{t,r}\xi_t\Big)\Big\}\rho_r(w_c).
\end{aligned} \tag{49}$$

By the Campbell's formula [5], the expectation for Poisson process is evaluated as

$$\begin{aligned}
=\ & \exp\Big[\int \Big[\exp\Big\{-w\Big(f(\theta) + \sum_{t=1}^{T} q_{t,r}\xi_t\Big)\Big\} - 1\Big]\rho_r(dw)H_r(d\theta)\Big] \\
& \times \prod_{\substack{c\in\pi \\ z_c=r}} H_r(\theta_c^*) \int w_c^{|c|} \exp\Big\{-w_c\Big(f(\theta_c^*) + \sum_{t=1}^{T} q_{t,r}\xi_t\Big)\Big\}\rho_r(w_c).
\end{aligned} \tag{50}$$

Now we set $f = 0$ to get the denominator.

$$\begin{aligned}
\mathbb{E}[p(\pi, \{z_c, \theta_c^*\}, \{u_t\}|\nu_r)] \ \propto\ & \exp\Big[\int \Big\{\exp\Big(-w\sum_{t=1}^{T} q_{t,r}\xi_t\Big) - 1\Big\}\rho_r(dw)H_r(d\theta)\Big] \\
& \times \prod_{\substack{c\in\pi \\ z_c=r}} H_r(\theta_c^*)\kappa_{\rho_r}\Big(|c|, \sum_{t=1}^{T} q_{t,r}\xi_t\Big).
\end{aligned} \tag{51}$$

Then dividing the numerator with the denominator yields

$$\mathbb{E}\Big[e^{-\nu_r(f)}|\pi,\{z_c,\theta_c^*\},\{u_t\}\Big] = \exp\left[\int(e^{-wf(\theta)}-1)\exp\Big(-w\sum_{t=1}^{T}q_{t,r}\xi_t\Big)\rho_r(dw)H_r(d\theta)\right] \times\prod_{\substack{c\in\pi\\ z_c=r}}\int\frac{w_c^{|c|}\exp\{-w_c(f(\theta_c^*)+\sum_t q_{t,r}\xi_t)\}}{\kappa_{\rho_r}(|c|,\sum_t q_{t,r}\xi_t)}\rho_r(dw_c). \quad (52)$$

This is the characteristic functional of a CRM with exponentially tilted Lévy measure $\bar{\rho}_r$ and fixed atoms as defined in (42). The marginal likelihood can easily be obtained by similar computations since

$$p(\pi,\{z_c,\theta_c^*\},\{\xi_t\}) = \mathbb{E}\Big[p(\pi,\{z_c,\theta_c^*\},\{\xi_t\})|\{\nu_r\})\Big]. \quad (53)$$

□

4.3 Posterior inference via marginal Gibbs sampling

We describe ths marginal Gibbs sampling algorithm we used in our paper. Again, the same stuff can be found in [4]. Combined with the likelihood function L, the joint likelihood of mixture model is written as:

$$p(\boldsymbol{X}_{1:T},\pi,\{z_c,\theta_c^*\}) = \prod_{t=1}^{T}\frac{\xi_t^{n_t-1}}{\Gamma(n_t)}\prod_{r=1}^{R}\exp\left\{-\psi_{\rho_r}\Big(\sum_{t=1}^{T}q_{t,r}\xi_t\Big)\right\} \times\prod_{c\in\pi}\bar{q}_{cz_c}\kappa_{\rho_r}\Big(|c|,\sum_{t=1}^{T}q_{t,z_c}\xi_t\Big)L(\boldsymbol{X}^{(c)}|\theta_c^*)H_{z_c}(\theta_c^*). \quad (54)$$

Suppose that we use NGGP with the same base measure H as Lévy measures for ν_r:

$$H_r = H \quad (55)$$

$$\rho_r(w) = \frac{\alpha_r}{\Gamma(1-\sigma)}w^{-1-\sigma}e^{-\tau w} \quad (56)$$

$$\kappa_{\rho_r}(m,u) = \frac{\alpha_r\Gamma(m-\sigma)}{(u+\tau)^{m-\sigma}\Gamma(1-\sigma)} \quad (57)$$

$$\psi_{\rho_r}(u) = \frac{\alpha_r}{\sigma}\{(\tau+u)^{\sigma}-\tau^{\sigma}\}. \quad (58)$$

Here, for simplicity, we only varied the hyperparemeter α_r and fixed other hyperparameters σ and τ. We also assume that H is a conjugate prior for L. Then we get

$$p(\boldsymbol{X}_{1:T},\pi,\{z_c\}) = \prod_{t=1}^{T}\frac{\xi_t^{n_t-1}}{\Gamma(n_t)}\exp\left[-\sum_{r=1}^{R}\frac{\alpha_r}{\sigma}\left\{\Big(\sum_{t=1}^{T}q_{t,r}\xi_t+\tau\Big)^{\sigma}-\tau^{\sigma}\right\}\right]\prod_{c\in\pi}\frac{\bar{q}_{c,z_c}\alpha_{z_c}\Gamma(|c|-\sigma)p(\boldsymbol{X}^{(c)}|\mathcal{H}_c)}{(\sum_t q_{t,z_c}\xi_t+\tau)^{|c|-\sigma}\Gamma(1-\sigma)}. \quad (59)$$

Sampling c and z_c

The cluster membership of each index (t,i) is sampled at each iteration. (t,i) may be allocated to a existing cluster $c\in\pi_{\backslash ti}$ where $\pi_{\backslash ti}$ is a partition of $I_{1:T}$ except (t,i). Also, (t,i) may create a new cluster which may come from one of the basis measures $\{\nu_r\}_{r=1}^{R}$.

$$p((t,i)\in c|\dots) \propto \frac{q_{t,z_c}(|c|-\sigma)p(\boldsymbol{X}^{(c\cup(t,i))}|\mathcal{H}_{c\cup(t,i)})}{(\sum_t q_{t,z_c}\xi_t+\tau)p(\boldsymbol{X}^{(c)}|\mathcal{H}_c)} \quad (60)$$

$$p((t,i)\in c\notin\pi_{\backslash ti},z_c=r|\dots) \propto \frac{q_{t,r}\alpha_r p(\boldsymbol{X}^{(c\cup(t,i))}|\mathcal{H}_{c\cup(t,i)})}{(\sum_t q_{t,r}\xi_t+\tau)^{1-\sigma}}. \quad (61)$$

Sampling ξ_t

The posterior distribution of ξ_t is written as

$$p(\xi_t|\dots) \propto \xi_t^{n_t-1} \exp\left\{ -\sum_{r=1}^{R} \frac{\alpha_r}{\sigma}\left(q_{t,r}\xi_t + \sum_{t'\neq t} q_{t'r}\xi_{t'} + \tau\right)^{\sigma} \right\} \prod_{c\in\pi} \left(q_{z_c,t}\xi_t + \sum_{t'\neq t} q_{z_c,t'}\xi_t + \tau\right)^{\sigma-|c|}. \tag{62}$$

As in [7, 4], we sample ξ_t vial slice-sampling [8] in log domain. Let $v_t = \log(\xi_t)$. Then

$$\begin{aligned}\log p(v_t|\dots) &= n_t v_t - \sum_{r=1}^{R} \frac{\alpha_r}{\sigma}\left(q_{t,r}e^{v_t} + \sum_{t'\neq t} q_{t',r}\xi_{t'} + \tau\right)^{\sigma} \\ &\quad - \sum_{c\in\pi}(|c|-\sigma)\log\left(q_{t,z_c}e^{v_t} + \sum_{t'\neq t} q_{t',z_c}\xi_t + \tau\right) + \text{const.}\end{aligned} \tag{63}$$

Sampling α

We place a Gamma prior on $\alpha_r \sim \text{Gamma}(a_\alpha, b_\alpha)$. Then

$$\alpha_r|\dots \sim \text{Gamma}\left(a_\alpha + \sum_{c\in\pi}\mathbb{I}[z_c = r], b_\alpha + \sum_{r=1}^{R} \frac{(\sum_t q_{t,r}\xi_t + \tau)^{\sigma} - \tau^{\sigma}}{\sigma}\right). \tag{64}$$

Hence, we can easily sample α_r from Gamma distribution.

Sampling τ

We place a Gamma prior on $\tau \sim \text{Gamma}(a_\tau, b_\tau)$. Then

$$p(\tau|\dots) \propto \tau^{a_\tau-1}e^{-b\tau} \exp\left[-\sum_{r=1}^{R} \frac{\alpha_r}{\sigma}\left\{ \left(\sum_{t=1}^{T} q_{t,r}\xi_t + \tau\right)^{\sigma} - \tau^{\sigma} \right\}\right] \prod_{c\in\pi}\left(\sum_{t=1}^{T} q_{t,z_c}\xi_t + \tau\right)^{\sigma-|c|}. \tag{65}$$

Again, as in [7, 4], τ can easily be sampled via slice sampling in log domian. Let $\xi = \log(\tau)$.

$$\begin{aligned}\log p(\xi|\dots) &= a_\tau\xi - b_\tau e^{\xi} - \sum_{r=1}^{R} \frac{\alpha_r}{\sigma}\left\{ \left(\sum_{t=1}^{T} q_{t,r}\xi_t + e^{\xi}\right)^{\sigma} - e^{\xi\sigma} \right\} \\ &\quad - \sum_{c\in\pi}(|c|-\sigma)\log\left(\sum_{t=1}^{T} q_{t,z_c}\xi_t + e^{\xi}\right) + \text{const.}\end{aligned} \tag{66}$$

In practice, following [7], we just fixed $\tau = 10^{-3}$.

Sampling σ

We place a Beta prior on $\sigma \sim \text{Beta}(a_\sigma, b_\sigma)$.

$$p(\sigma|\dots) \propto \sigma^{a_\sigma-1}(1-\sigma)^{b_\sigma-1} \exp\left[-\sum_{r=1}^{R} \frac{\alpha_r}{\sigma}\left\{ \left(\sum_{t=1}^{T} q_{t,r}\xi_t + \tau\right)^{\sigma} - \tau^{\sigma} \right\}\right] \prod_{c\in\pi} \frac{\Gamma(|c|-\sigma)}{(\sum_t q_{t,z_c}\xi_t + \tau)^{|c|-\sigma}\Gamma(1-\sigma)}, \tag{67}$$

which can easily be sampled using slice sampling, as in [7].

References

[1] J. F. C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

[2] L. F. James, A. Lijoi, and I. Prünster. Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics*, 36(1):76–97, 2009.

[3] Y. W. Teh. Bayesian nonparametrics rough notes. 2012.

[4] C. Chen, V. Rao, W. Buntine, and Y. W. Teh. Supplementary material for dependent normalized random measures. *Proceedings of the International Conference on Machine Learning (ICML)*, 2013.

[5] J. F. C. Kingman. *Poisson Processes*. Oxford Science Publications, 1993.

[6] C. Chen, V. Rao, W. Buntine, and Y. W. Teh. Dependent normalized random measures. In *Proceedings of the International Conference on Machine Learning (ICML)*, Atlanta, Georgia, USA, 2013.

[7] S. Favaro and Y. W. Teh. MCMC for normalized random measure mixture models. *Statistical Science*, 28(3):335–359, 2013.

[8] R. M. Neal. Slice sampling. *The Annals of Statistics*, 31(3):705–767, 2003.