Appendix: On the Model Shrinkage Effect of Gamma Process Edge Partition Models

Iku Ohama[⋆‡] **Issei Sato**[†] **Takuya Kida**[‡] **Hiroki Arimura**[‡]
[⋆]Panasonic Corp., Japan [†]The Univ. of Tokyo, Japan [‡]Hokkaido Univ., Japan
ohama.iku@jp.panasonic.com sato@k.u-tokyo.ac.jp {kida,arim}@ist.hokudai.ac.jp

A Gibbs Samplers for the EPM

A.1 Model Description

The full description of the generative model for the EPM [1] is described as follows:

$$x_{i,j} = \mathbb{I}(m_{i,j,\cdot} \geq 1), \quad m_{i,j,\cdot} \,|\, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\lambda} \sim \text{Poisson}\left(\sum_{k=1}^{K} U_{i,k} V_{j,k} \lambda_k\right),$$
$$U_{i,k} \sim \text{Gamma}(a_1, b_1), \quad V_{j,k} \sim \text{Gamma}(a_2, b_2), \quad \lambda_k \sim \text{Gamma}(\gamma_0/T, c_0). \tag{9}$$

A.2 Closed-form Gibbs Samplers

Posterior inference for all parameters and hyperparameters of the EPM can be performed using Gibbs sampler.

Sampling $\boldsymbol{m}$: From Eq. (9), as $m_{i,j,\cdot} = 0$ if and only if $x_{i,j} = 0$, posterior sampling of $\boldsymbol{m}$ is required only for non-zero entries ($x_{i,j} = 1$), and can be performed using zero-truncated Poisson (ZTP) distribution [2] as follows:

$$m_{i,j,\cdot} \,|\, \boldsymbol{U}, \boldsymbol{\lambda}, \boldsymbol{V} \sim \begin{cases} \delta(0) & \text{if } x_{i,j} = 0, \\ \text{ZTP}(\sum_{k=1}^{T} U_{i,k} \lambda_k V_{j,k}) & \text{if } x_{i,j} = 1. \end{cases} \tag{10}$$

Then, latent count $m_{i,j,k}$ related to the k-th atom can be obtained by partitioning $m_{i,j,\cdot}$ into T atoms as

$$\{m_{i,j,k}\}_{k=1}^{T} \,|\, m_{i,j,\cdot}, \boldsymbol{U}, \boldsymbol{\lambda}, \boldsymbol{V} \sim \text{Multinomial}\left(m_{i,j,\cdot}; \left\{\frac{U_{i,k} \lambda_k V_{j,k}}{\sum_{k'=1}^{T} U_{i,k'} \lambda_{k'} V_{j,k'}}\right\}_{k=1}^{T}\right). \tag{11}$$

Sampling $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\lambda}$: As the generative model for $m_{i,j,k}$ can be given as $m_{i,j,k} \,|\, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\lambda} \sim \text{Poisson}(U_{i,k} V_{j,k} \lambda_k)$, according to the additive property of the Poisson distributions, generative models for aggregated counts also can be expressed as follows:

$$m_{i,\cdot,k} = (\textstyle\sum_j m_{i,j,k}) \,|\, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\lambda} \sim \text{Poisson}(U_{i,k} (\sum_j V_{j,k}) \lambda_k), \tag{12}$$
$$m_{\cdot,j,k} = (\textstyle\sum_i m_{i,j,k}) \,|\, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\lambda} \sim \text{Poisson}((\sum_i U_{i,k}) V_{j,k} \lambda_k), \tag{13}$$
$$m_{\cdot,\cdot,k} = (\textstyle\sum_i \sum_j m_{i,j,k}) \,|\, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\lambda} \sim \text{Poisson}((\sum_i U_{i,k})(\sum_j V_{j,k}) \lambda_k). \tag{14}$$

Therefore, thanks to the conjugacy between Poisson and gamma distributions, posterior samplers for $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{\lambda}$ are straightforwardly derived as follows:

$$U_{i,k} \,|\, - \sim \text{Gamma}(a_1 + m_{i,\cdot,k}, b_1 + (\textstyle\sum_j V_{j,k}) \lambda_k), \tag{15}$$
$$V_{j,k} \,|\, - \sim \text{Gamma}(a_2 + m_{\cdot,j,k}, b_2 + (\textstyle\sum_i U_{i,k}) \lambda_k), \tag{16}$$
$$\lambda_k \,|\, - \sim \text{Gamma}(\gamma_0/T + m_{\cdot,\cdot,k}, c_0 + (\textstyle\sum_i U_{i,k})(\sum_j V_{j,k})). \tag{17}$$

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

A.3 Sampling Hyperparameters

Sampling b_1, b_2, c_0: Thanks to the conjugacy between gamma distributions, posterior samplers for b_1, b_2, and c_0 are straightforwardly performed as follows:

$$b_1 \mid - \sim \text{Gamma}(e_0 + ITa_1, f_0 + \textstyle\sum_i \sum_k \phi_{i,k}), \tag{18}$$

$$b_2 \mid - \sim \text{Gamma}(e_0 + JTa_2, f_0 + \textstyle\sum_j \sum_k \psi_{j,k}), \tag{19}$$

$$c_0 \mid - \sim \text{Gamma}(e_0 + \gamma_0, f_0 + \textstyle\sum_k \lambda_k). \tag{20}$$

For the remaining hyperparameters (i.e., a_1, a_2, and γ_0), we can construct closed-form Gibbs samplers using data augmentation techniques [3, 1, 4, 5], that consider an expanded probability over target and some auxiliary variables. The key strategy is the use of the following expansions:

$$\frac{\Gamma(u)}{\Gamma(u+n)} = \frac{B(u,n)}{\Gamma(n)} = \Gamma(n)^{-1} \int_0^1 v^{u-1}(1-v)^{n-1} \mathrm{d}v, \tag{21}$$

$$\frac{\Gamma(u+n)}{\Gamma(u)} = \sum_{w=0}^{n} S(n,w) u^w, \tag{22}$$

where $B(\cdot,\cdot)$ is the beta function and $S(\cdot,\cdot)$ is the Stirling number of the first kind.

Sampling a_1, a_2: For shape parameter a_1, marginalizing $\boldsymbol{U}$ from Eq. (12), we have a partially marginalized likelihood related to target variable a_1 as:

$$P(\{m_{i,\cdot,k}\}_{i,k} \mid \boldsymbol{V}, \boldsymbol{\lambda}) \propto \prod_{k=1}^{T} \left\{ \left(\frac{b_1}{b_1 + (\sum_j V_{j,k})\lambda_k} \right)^{Ia_1} \prod_{i=1}^{I} \frac{\Gamma(a_1 + m_{i,\cdot,k})}{\Gamma(a_1)} \right\}. \tag{23}$$

Therefore, expanding Eq. (23) using Eq. (22) and assuming gamma prior as $a_1 \sim \text{Gamma}(e_0, f_0)$, posterior sampling for a_1 can be performed as follows:

$$w_{i,k} \mid - \sim \text{Antoniak}(m_{i,\cdot,k}, a_1), \tag{24}$$

$$a_1 \mid - \sim \text{Gamma}\left(e_0 + \textstyle\sum_i \sum_k w_{i,k}, f_0 - I \times \sum_k \ln \dfrac{b_1}{b_1 + (\sum_j V_{j,k})} \right), \tag{25}$$

where $\text{Antoniak}(m_{i,\cdot,k}, a_1)$ is an Antoniak distribution [6]. This is the distribution of the number of occupied tables if $m_{i,\cdot,k}$ customers are assigned to one of an infinite number of tables using the Chinese restaurant process (CRP) [7, 8] with concentration parameter a_1, and is sampled as $w_{i,k} = \sum_{p=1}^{m_{i,\cdot,k}} w_{i,k,p}, w_{i,k,p} \sim \text{Bernoulli}\left(\frac{a_1}{a_1+p-1}\right)$. Similarly, posterior sampler for a_2 can be derived from Eqs. (13) and (22) (omitted for brevity).

Sampling γ_0: Similar to the samplers for a_1 and a_2, according to Eqs. (14) and (22), γ_0 can be updated as follows:

$$w_k \mid - \sim \text{Antoniak}(m_{\cdot,\cdot,k}, \gamma_0/T), \tag{26}$$

$$\gamma_0 \mid - \sim \text{Gamma}\left(e_0 + \textstyle\sum_k w_k, f_0 - \dfrac{1}{T} \sum_k \ln \dfrac{c_0}{c_0 + (\sum_i U_{i,k})(\sum_j V_{j,k})} \right). \tag{27}$$

B Gibbs Samplers for the CEPM

Posterior inference for the CEPM can be performed using Gibbs sampler as same as that for the EPM. However, only a_1 and a_2 do not have closed-form sampler because of introduced constraints $b_1 = C_1 \times a_1$ and $b_2 = C_2 \times a_2$. Therefore, instead of sampling from true posterior, we use the grid Gibbs sampler [9] to sample from a discrete probability distribution

$$P(a_1 \mid -) \propto \text{Eq (23)} \times P(a_1) \tag{28}$$

over a grid of points $\frac{1}{1+a_1} = 0.01, 0.02, \ldots, 0.99$. Note that a_2 can be sampled in a same way as a_1 (omitted for brevity).

C Gibbs Samplers for the DEPM

C.1 Closed-form Gibbs Samplers

Sampling $\boldsymbol{\phi}, \boldsymbol{\psi}$: Given $m_{\cdot,\cdot,k} = \sum_i \sum_j m_{i,j,k}$, generative process for latent count $m_{i,\cdot,k}$ can be expressed as

$$\{m_{i,\cdot,k}\}_{i=1}^{I} \,|\, m_{\cdot,\cdot,k}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda} \sim \text{Multinomial}\left(m_{\cdot,\cdot,k}; \{\phi_{i,k}\}_{i=1}^{I}\right). \tag{29}$$

Thanks to conjugacy between Eq. (29) and Dirichlet prior in Eq. (4), posterior sampling for $\boldsymbol{\phi}$ can be performed as

$$\{\phi_{i,k}\}_{i=1}^{I} \,|\, - \sim \text{Dirichlet}(\{\alpha_1 + m_{i,\cdot,k}\}_{i=1}^{I}). \tag{30}$$

Similarly, $\boldsymbol{\psi}$ can be updated as

$$\{\psi_{j,k}\}_{j=1}^{J} \,|\, - \sim \text{Dirichlet}(\{\alpha_2 + m_{\cdot,j,k}\}_{j=1}^{J}). \tag{31}$$

Sampling $\boldsymbol{m}, \boldsymbol{\lambda}$: Posterior samplers for remaining latent variables $\boldsymbol{m}$ and $\boldsymbol{\lambda}$ are straightforwardly given from Eqs. (10), (11), and (17) by replacing $\boldsymbol{U}$ and $\boldsymbol{V}$ with $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, respectively.

C.2 Sampling Hyperparameters

Sampling α_1, α_2: Similar to Appendix A.3, marginalizing $\boldsymbol{\phi}$ out from Eq. (4) and expanding the marginal likelihood using Eqs. (21) and (22), posterior sampling for α_1 can be derived as follows:

$$v_{1,k} \,|\, - \sim \text{Beta}(I\alpha_1, m_{\cdot,\cdot,k}), \tag{32}$$

$$w_{1,i,k} \,|\, - \sim \text{Antoniak}(m_{i,\cdot,k}, \alpha_1), \tag{33}$$

$$\alpha_1 \,|\, - \sim \text{Gamma}(e_0 + \textstyle\sum_i \sum_k w_{1,i,k}, f_0 - I \times \sum_k \ln v_{1,k}). \tag{34}$$

Note that the posterior sampler for α_2 can be derived in same way (omitted for brevity).

Sampling γ_0, c_0: The remaining hyperparameters (i.e., γ_0 and c_0) can be updated as same as in the EPM. Similar to the sampler for the EPM, c_0 can be updated using Eq. (20). Finally, posterior sampler for γ_0 can be derived as

$$w_k \,|\, - \sim \text{Antoniak}(m_{\cdot,\cdot,k}, \gamma_0/T), \tag{35}$$

$$\gamma_0 \,|\, - \sim \text{Gamma}\left(e_0 + \textstyle\sum_k w_k, f_0 - \ln \dfrac{c_0}{c_0+1}\right). \tag{36}$$

D Proof of Theorem 4

Considering a joint distribution for $m_{i,j,\cdot}$ customers and their assignments $\boldsymbol{z}_{i,j} = \{z_{i,j,s}\}_{s=1}^{m_{i,j,\cdot}} \in \{1, \cdots, T\}^{m_{i,j,\cdot}}$ to T tables, we have following lemma for the truncated DEPM:

Lemma 1. *The joint distribution over $\boldsymbol{m}$ and $\boldsymbol{z}$ for the DEPM is expressed by a fully factorized form as*

$$P(\boldsymbol{m}, \boldsymbol{z} \,|\, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) = \prod_{i=1}^{I}\prod_{j=1}^{J} \frac{1}{m_{i,j,\cdot}!} \times \prod_{i=1}^{I}\prod_{k=1}^{T} \phi_{i,k}^{m_{i,\cdot,k}} \times \prod_{j=1}^{J}\prod_{k=1}^{T} \psi_{j,k}^{m_{\cdot,j,k}} \times \prod_{k=1}^{T} \lambda_k^{m_{\cdot,\cdot,k}} e^{-\lambda_k}. \tag{37}$$

Proof. As the likelihood functions $P(m_{i,j,\cdot} \,|\, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda})$ and $P(z_{i,j,s} \,|\, m_{i,j,\cdot}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda})$ are given as

$$P(m_{i,j,\cdot} \,|\, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) = \frac{1}{m_{i,j,\cdot}!} \left(\sum_{k=1}^{T} \phi_{i,k}\psi_{j,k}\lambda_k\right)^{m_{i,j,\cdot}} e^{-\sum_{k=1}^{T} \phi_{i,k}\psi_{j,k}\lambda_k}, \tag{38}$$

$$P(z_{i,j,s} = k^* \,|\, m_{i,j,\cdot}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) = \frac{\phi_{i,k^*}\psi_{j,k^*}\lambda_{k^*}}{\sum_{k'=1}^{T} \phi_{i,k'}\psi_{j,k'}\lambda_{k'}}, \tag{39}$$

respectively, we obtain the joint likelihood function for $\boldsymbol{m}$ and $\boldsymbol{z}$ as follows:

$$
\begin{aligned}
&P(\boldsymbol{m}, \boldsymbol{z} \,|\, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) \\
&= \prod_{i=1}^{I} \prod_{j=1}^{J} \left\{ P(m_{i,j,\cdot} \,|\, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) \prod_{s=1}^{m_{i,j,\cdot}} P(z_{i,j,s} \,|\, m_{i,j,\cdot}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\lambda}) \right\} \\
&= \prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{m_{i,j,\cdot}!} \times \prod_{i=1}^{I} \prod_{k=1}^{T} \phi_{i,k}^{m_{i,\cdot,k}} \times \prod_{j=1}^{J} \prod_{k=1}^{T} \psi_{j,k}^{m_{\cdot,j,k}} \times \prod_{k=1}^{T} \lambda_k^{m_{\cdot,\cdot,k}} e^{-\lambda_k (\sum_i \phi_{i,k})(\sum_j \psi_{j,k})}.
\end{aligned}
\tag{40}
$$

Thanks to the l_1-constraints for $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ we introduced in Eq. (3), substituting $\sum_i \phi_{i,k} = \sum_j \psi_{j,k} = 1$ for Eq. (40), we obtain Eq. (37) in Lemma 1. □

Thanks to the conjugacy between Eq. (37) in Lemma 1 and prior construction in Eq. (4), marginalizing $\boldsymbol{\phi}$, $\boldsymbol{\psi}$, and $\boldsymbol{\lambda}$ out, we obtain the following marginal likelihood for the DEPM:

$$
\begin{aligned}
P(\boldsymbol{m}, \boldsymbol{z}) = &\prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{m_{i,j,\cdot}!} \times \prod_{k=1}^{T} \frac{\Gamma(I\alpha_1)}{\Gamma(I\alpha_1 + m_{\cdot,\cdot,k})} \prod_{i=1}^{I} \frac{\Gamma(\alpha_1 + m_{i,\cdot,k})}{\Gamma(\alpha_1)} \\
&\times \prod_{k=1}^{T} \frac{\Gamma(J\alpha_2)}{\Gamma(J\alpha_2 + m_{\cdot,\cdot,k})} \prod_{j=1}^{J} \frac{\Gamma(\alpha_2 + m_{\cdot,j,k})}{\Gamma(\alpha_2)} \times \prod_{k=1}^{T} \frac{\Gamma\left(\frac{\gamma_0}{T} + m_{\cdot,\cdot,k}\right) c_0^{\frac{\gamma_0}{T}}}{\Gamma\left(\frac{\gamma_0}{T}\right) (c_0+1)^{\frac{\gamma_0}{T} + m_{\cdot,\cdot,k}}}.
\end{aligned}
\tag{41}
$$

Considering a partition $[\boldsymbol{z}]$ instead of the assignments $\boldsymbol{z}$ as same as in [10], the marginal likelihood function $P(\boldsymbol{m}, [\boldsymbol{z}])$ for a partition of the truncated DEPM can be expressed as

$$
\begin{aligned}
P(\boldsymbol{m}, [\boldsymbol{z}]) = &\frac{T!}{(T-K_+)!} P(\boldsymbol{m}, \boldsymbol{z}) \\
= &\prod_{i=1}^{I} \prod_{j=1}^{J} \frac{1}{m_{i,j,\cdot}!} \times \prod_{k=1}^{K_+} \frac{\Gamma(I\alpha_1)}{\Gamma(I\alpha_1 + m_{\cdot,\cdot,k})} \prod_{i=1}^{I} \frac{\Gamma(\alpha_1 + m_{i,\cdot,k})}{\Gamma(\alpha_1)} \\
&\times \prod_{k=1}^{K_+} \frac{\Gamma(J\alpha_2)}{\Gamma(J\alpha_2 + m_{\cdot,\cdot,k})} \prod_{j=1}^{J} \frac{\Gamma(\alpha_2 + m_{\cdot,j,k})}{\Gamma(\alpha_2)} \\
&\times \frac{T!}{(T-K_+)! T^{K_+}} \times \gamma_0^{K_+} \left(\frac{c_0}{c_0+1} \right)^{\gamma_0} \prod_{k=1}^{K_+} \frac{\prod_{l=1}^{m_{\cdot,\cdot,k}-1} (l + \gamma_0/T)}{(c_0+1)^{m_{\cdot,\cdot,k}}}.
\end{aligned}
\tag{42}
$$

Therefore, taking $T \to \infty$ in Eq. (42), we obtain the marginal likelihood function for the truly infinite DEPM (i.e., IDEPM) as in Eq. (5) of Theorem 4.

E Sampling Hyperparameters for the IDEPM

Sampling α_1, α_2: Posterior samplers for α_1 and α_2 of the IDEPM are equivalent to those of the truncated DEPM as in Appendix C.2.

Sampling γ_0: From Eq. (5), we straightforwardly obtain the posterior sampler for γ_0 as

$$
\gamma_0 \,|\, - \sim \text{Gamma}\left(e_0 + K_+, f_0 - \ln \frac{c_0}{c_0+1} \right). \tag{43}
$$

Note that γ_0 in Eq. (5) can be marginalized out assuming gamma prior. However, we explicitly sample γ_0 for simplicity in this paper.

Sampling c_0: As derived in Sec. 4.3 of main article, c_0 is updated as

$$
\lambda_k \,|\, - \sim \text{Gamma}(m_{\cdot,\cdot,k}, c_0 + 1) \quad k \in \{1, \ldots, K_+\}, \tag{44}
$$

$$
\lambda_{\gamma_0} \,|\, - \sim \text{Gamma}(\gamma_0, c_0 + 1), \tag{45}
$$

$$
c_0 \,|\, - \sim \text{Gamma}(e_0 + \gamma_0, f_0 + \lambda_{\gamma_0} + \textstyle\sum_{k=1}^{K_+} \lambda_k). \tag{46}
$$

References

[1] Mingyuan Zhou. "Infinite Edge Partition Models for Overlapping Community Detection and Link Prediction". In: *Proc. AISTATS*. Vol. 38. 2015, pp. 1135–1143.

[2] Charles J. Geyer. *Lower-Truncated Poisson and Negative Binomial Distributions*. Tech. rep. Working Paper Written for the Software R. University of Minnesota, MN (available: http://cran.r-project.org/web/packages/aster/vignettes/trunc.pdf), 2007.

[3] Michael D. Escobar and Mike West. "Bayesian Density Estimation and Inference Using Mixtures". In: *J. Am. Stat. Assoc.* 90 (1994), pp. 577–588.

[4] Yee Whye Teh, Michael I. Jordan, Matthew J. Beal, and David M. Blei. "Hierarchical Dirichlet Processes". In: *J. Am. Stat. Assoc.* 101.476 (2006), pp. 1566–1581.

[5] David Newman, Arthur U. Asuncion, Padhraic Smyth, and Max Welling. "Distributed Algorithms for Topic Models". In: *J. Mach. Learn. Res.* 10 (2009), pp. 1801–1828.

[6] Charles E. Antoniak. "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems". In: *The Annals of Statistics* 2.6 (1974), pp. 1152–1174.

[7] David Blackwell and James B. MacQueen. "Ferguson distributions via Polya urn schemes". In: *The Annals of Statistics* 1 (1973), pp. 353–355.

[8] David Aldous. "Exchangeability and related topics". In: *Ecole d'Ete de Probabilities de Saint-Flour XIII*. 1985, pp. 1–198.

[9] Mingyuan Zhou. "Beta-Negative Binomial Process and Exchangeable Random Partitions for Mixed-Membership Modeling". In: *Proc. NIPS*. 2014, pp. 3455–3463.

[10] Thomas L. Griffiths and Zoubin Ghahramani. "The Indian Buffet Process: An Introduction and Review". In: *J. Mach. Learn. Res.* 12 (2011), pp. 1185–1224.