Beta-Negative Binomial Process and Exchangeable Random Partitions for Mixed-Membership Modeling

Mingyuan Zhou

Department of Information, Risk, and Operations Management, McCombs School of Business

The University of Texas at Austin, Austin, TX, USA

Introduction

The integer-valued beta-negative binomial process (BNBP) is employed to partition a count vector into a latent column-exchangeable random count matrix. The paper makes the following contributions:

- ➢An exchangeable partition probability function (EPPF) for mixed-membership modeling is proposed.
- ➢The prediction rule to simulate the exchangeable random partitions for grouped data governed by the BNBP is derived.
- ➢A fully collapsed Gibbs sampler, with closed-form update equations for model parameters, is constructed.
- ➢The BNBP topic model is shown to converge fast, mix well, and provide state-of-the-art prediction performance when a compact representation of the corpus is desired.

Model and Inference

Beta Negative Binomial Process

$$X_j|B \sim \mathrm{NBP}(r_j, B),\ B \sim \mathrm{BP}(c, B_0)$$

Augmented representation:

$$X_j|\Theta_j \sim \mathrm{PP}(\Theta_j),\ \Theta_j|r_j, B \sim \Gamma\mathrm{P}[r_j, B/(1-B)],\ B \sim \mathrm{BP}(c, B_0)$$

Group size dependent mixture model

$$z_{ji} \sim \sum_{k=1}^{\infty} \frac{\theta_{jk}}{\Theta_j(\Omega)}\delta_k,\ m_j \sim \mathrm{Pois}(\Theta_j(\Omega)),$$
$$\Theta_j \sim \Gamma\mathrm{P}[r_j, B/(1-B)],\ B \sim \mathrm{BP}(c, B_0)$$

Conditional Likelihood:

$$f(\boldsymbol{z}_j, m_j|r_j, B) = \frac{1}{m_j!}\prod_{k=1}^{\infty}\frac{\Gamma(n_{jk}+r_j)}{\Gamma(r_j)}p_k^{n_{jk}}(1-p_k)^{r_j}$$

Exchangeable Cluster Probability Function:

$$f(\boldsymbol{z}, \boldsymbol{m}|\boldsymbol{r}, \gamma_0, c) = \frac{\gamma_0^{K_J} e^{-\gamma_0[\psi(c+r_\cdot)-\psi(c)]}}{\prod_{j=1}^J m_j!}\prod_{k=1}^{K_J}\left[\frac{\Gamma(n_{\cdot k})\Gamma(c+r_\cdot)}{\Gamma(c+n_{\cdot k}+r_\cdot)}\prod_{j=1}^J\frac{\Gamma(n_{jk}+r_j)}{\Gamma(r_j)}\right]$$

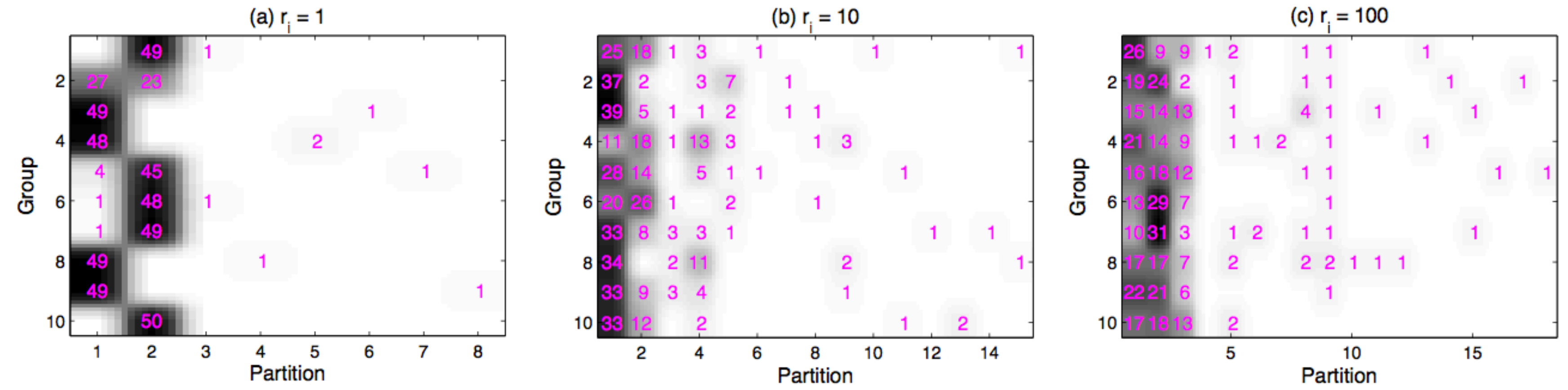
Exchangeable Partition Probability Function:

$$f(\boldsymbol{z}|\boldsymbol{m}, \boldsymbol{r}, \gamma_0, c) = \frac{\frac{\gamma_0^{K_J}}{\prod_{j=1}^J m_j!}\prod_{k=1}^{K_J}\left[\frac{\Gamma(n_{\cdot k})\Gamma(c+r_\cdot)}{\Gamma(c+n_{\cdot k}+r_\cdot)}\prod_{j=1}^J\frac{\Gamma(n_{jk}+r_j)}{\Gamma(r_j)}\right]}{\sum_{K'=1}^{m_\cdot}\frac{\gamma_0^{K'}}{K'!}\sum_{\sum_{k'=1}^{K'}\boldsymbol{n}_{:k'}=\boldsymbol{m}}\prod_{k'=1}^{K'}\frac{\Gamma(n_{\cdot k'})\Gamma(c+r_\cdot)}{\Gamma(c+n_{\cdot k'}+r_\cdot)}\prod_{j=1}^J\frac{\Gamma(n_{jk'}+r_j)}{n_{jk'}!\Gamma(r_j)}}$$

Prediction Rule

$$P(z_{ji}=k|\boldsymbol{z}^{-ji}, \boldsymbol{m}, \boldsymbol{r}, \gamma_0, c) \propto \begin{cases}\frac{n_{\cdot k}^{-ji}}{c+n_{\cdot k}^{-ji}+r_\cdot}(n_{jk}^{-ji}+r_j), & \text{for } k=1,\dots,K_J^{-ji};\\ \frac{\gamma_0}{c+r_\cdot}r_j, & \text{if } k=K_J^{-ji}+1.\end{cases}$$

Simulated Exchangeable Random Partitions



Mixed-Membership Modeling

BNBP Topic Model

$$x_{ji} \sim \mathrm{Mult}(\boldsymbol{\phi}_{z_{ji}}),\ \boldsymbol{\phi}_k \sim \mathrm{Dir}(\eta,\dots,\eta),\ z_{ji} \sim \sum_{k=1}^{\infty}\frac{\theta_{jk}}{\Theta_j(\Omega)}\delta_k,\ m_j \sim \mathrm{Pois}(\Theta_j(\Omega)),$$
$$\Theta_j \sim \Gamma\mathrm{P}(r_j, \tfrac{B}{1-B}),\ r_j \sim \mathrm{Gamma}(a_0, 1/b_0),\ B \sim \mathrm{BP}(c, B_0),\ \gamma_0 \sim \mathrm{Gamma}(e_0, 1/f_0)$$

Parameter Inference

$$(\gamma_0|-) \sim \mathrm{Gamma}\left(e_0 + K_J, \frac{1}{f_0 + \psi(c+r_\cdot) - \psi(c)}\right),$$
$$(p_k|-) \sim \mathrm{Beta}(n_{\cdot k}, c+r_\cdot),\ \ (Q(\Omega\backslash\mathcal{D}_J)|-) \sim \mathrm{logBeta}(\gamma_0, c+r_\cdot),$$
$$(l_{jk}|-) = \sum_{t=1}^{n_{jk}} u_t,\ u_t \sim \mathrm{Bernoulli}\left(\frac{r_j}{r_j+t-1}\right),$$
$$(r_j|-) \sim \mathrm{Gamma}\left(a_0 + \sum_{k=1}^{K_J} l_{jk}, \frac{1}{b_0 + Q(\Omega\backslash\mathcal{D}_J) - \sum_{k=1}^{K_J}\ln(1-p_k)}\right)$$

HDP-LDA Collapsed Gibbs Sampler

$$P(z_{ji}=k|\boldsymbol{x}, \boldsymbol{z}^{-ji}, \boldsymbol{m}, \alpha, \tilde{\boldsymbol{r}}) \propto \begin{cases}\frac{\eta+n_{v_{ji}\cdot k}^{-ji}}{V\eta+n_{\cdot k}^{-ji}}\cdot(n_{jk}^{-ji}+\alpha\tilde{r}_k), & \text{for } k=1,\dots,K_J^{-ji};\\ \frac{1}{V}\cdot(\alpha\tilde{r}_*), & \text{if } k=K_J^{-ji}+1;\end{cases}$$

BNBP Collapsed Gibbs Sampler

$$P(z_{ji}=k|\boldsymbol{x}, \boldsymbol{z}^{-ji}, \gamma_0, \boldsymbol{m}, c, \boldsymbol{r}) \propto \begin{cases}\frac{\eta+n_{v_{ji}\cdot k}^{-ji}}{V\eta+n_{\cdot k}^{-ji}}\cdot\frac{n_{\cdot k}^{-ji}}{c+n_{\cdot k}^{-ji}+r_\cdot}\cdot(n_{jk}^{-ji}+r_j), & \text{for } k=1,\dots,K_J^{-ji};\\ \frac{1}{V}\cdot\frac{\gamma_0}{c+r_\cdot}\cdot r_j, & \text{if } k=K_J^{-ji}+1.\end{cases}$$