Beta-Negative Binomial Process and Poisson Factor Analysis

Mingyuan Zhou, Lauren Hannah, David Dunson and Lawrence Carin

Duke University, Durham, NC 27708, USA

Introduction

A beta-negative binomial (BNB) process is proposed, leading to a beta-gamma-Poisson process, which may be viewed as a “multi-scoop” generalization of the beta-Bernoulli process.

- The BNB process is augmented into a beta-gamma-gamma-Poisson hierarchical structure, and applied as a nonparametric Bayesian prior for an infinite Poisson factor analysis model.
- A finite approximation for the beta process Levy random measure is constructed for convenient implementation.
- Efficient MCMC computations are performed with data augmentation and marginalization techniques.
- Encouraging results are shown on document count matrix factorization.

Beta-Negative Binomial Process

Negative binomial distribution

$$f_X(k) = \int_0^\infty \text{Pois}(k;\lambda)\text{Gamma}\left(\lambda; r, \frac{p}{1-p}\right) d\lambda = \frac{\Gamma(r+k)}{k!\Gamma(r)}(1-p)^r p^k$$

Beta process

$$\nu_{\text{BP}}(dpd\omega) = cp^{-1}(1-p)^{c-1}dpB_0(d\omega) \qquad B = \sum_{k=1}^\infty p_k \delta_{\omega_k}$$

Marked beta process

$$\nu^*_{\text{BP}}(dpdrd\omega) = cp^{-1}(1-p)^{c-1}dpR_0(dr)B_0(d\omega) \quad B^* = \sum_{k=1}^\infty p_k \delta_{(r_k,\omega_k)}$$

Negative binomial process $X_i \sim \text{NBP}(B^*)$

$$X_i = \sum_{k=1}^\infty \kappa_{ki}\delta_{(r_k,\omega_k)}, \quad \kappa_{ki} \sim \text{NB}(r_k, p_k)$$

Gamma Poisson process $X_i \sim \mathcal{P}(\mathcal{T}(B^*))$

$$\kappa_{ki} \sim \text{Pois}(\theta_{ki}), \quad \theta_{ki} \sim \text{Gamma}(r_k, p_k/(1-p_k))$$

Posterior

$$B^*|\{X_i\}_{1,n} \sim \text{BP}\left(c_n, \frac{c}{c_n}R_0B_0 + \frac{1}{c_n}\sum_k m_{nk}\delta_{(r_k,\omega_k)}\right)$$

$$c_n = \begin{cases} c + m_{nk} + nr_k, & \text{if } (r,\omega) = (r_k,\omega_k) \in \mathcal{D} \\ c + nr, & \text{if } (r,\omega) \in (\mathbb{R}^+ \times \Omega)\backslash\mathcal{D} \end{cases}$$

Poisson Factor Analysis

Finite approximation

$$\nu^*_{\epsilon\text{BP}}(dpdrd\omega) = cp^{c\epsilon-1}(1-p)^{c(1-\epsilon)-1}dpR_0(dr)B_0(d\omega)$$

$$\nu^+_{\epsilon\text{BP}} = \nu^*_{\epsilon\text{BP}}([0,1]\times\mathbb{R}^+\times\Omega) = c\gamma\alpha\text{B}(c\epsilon, c(1-\epsilon))$$

$$K \sim \text{Pois}(\nu^+_{\epsilon\text{BP}}) = \text{Pois}(c\gamma\alpha\text{B}(c\epsilon, c(1-\epsilon)))$$

Count matrix factorization

$\mathbf{X} = \text{Pois}(\mathbf{\Phi\Theta}) \quad \mathbf{\Phi} \in \mathbb{R}^{P\times K} \quad \mathbf{\Theta} \in \mathbb{R}^{K\times N}$

Augmentation 1 (addition)

$$x_{pi} = \sum_{k=1}^K x_{pik}, \quad x_{pik} \sim \text{Pois}(\phi_{pk}\theta_{ki})$$

Augmentation 2 (thinning)

$$x_{pi} \sim \text{Pois}\left(\sum_{k=1}^K \phi_{pk}\theta_{ki}\right), \quad \zeta_{pik} = \frac{\phi_{pk}\theta_{ki}}{\sum_{k=1}^K \phi_{pk}\theta_{ki}}$$

$$[x_{pi1},\cdots,x_{piK}] \sim \text{Mult}(x_{pi}, \zeta_{pi1},\cdots,\zeta_{piK})$$

BNB process Poisson factor analysis

$$x_{pi} = \sum_{k=1}^K x_{pik}, \quad x_{pik} \sim \text{Pois}(\phi_{pk}\theta_{ki})$$

$$\phi_k \sim \text{Dir}(a_\phi,\cdots,a_\phi)$$

$$\theta_{ki} \sim \text{Gamma}\left(r_k, \frac{p_k}{1-p_k}\right)$$

$$r_k \sim \text{Gamma}(c_0r_0, 1/c_0)$$

$$p_k \sim \text{Beta}(c\epsilon, c(1-\epsilon)).$$

MCMC Inference

$$x_{\cdot ik} = \sum_{p=1}^P x_{pik}, \; x_{p\cdot k} = \sum_{i=1}^N x_{pik}, \; x_{\cdot\cdot k} = \sum_{p=1}^P\sum_{i=1}^N x_{pik} \text{ and } x_{\cdot i\cdot} = \sum_{p=1}^P\sum_{k=1}^K x_{pik}.$$

$$[x_{pi1},\cdots,x_{piK}] \sim \text{Mult}(x_{pi}, \zeta_{pi1},\cdots,\zeta_{piK}) \quad \zeta_{pik} = \frac{\phi_{pk}\theta_{ki}}{\sum_{k=1}^K \phi_{pk}\theta_{ki}}$$

$$p(\phi_k|-) \sim \text{Dir}(a_\phi + x_{1\cdot k},\cdots,a_\phi + x_{P\cdot k})$$

$$p(p_k|-) \sim \text{Beta}(c\epsilon + x_{\cdot\cdot k}, c(1-\epsilon) + Nr_k)$$

$$p(\theta_{ki}|-) \sim \text{Gamma}(r_k + x_{\cdot ik}, p_k)$$

$$p(r_k|-) \propto \text{Gamma}(r_k; c_0r_0, 1/c_0)\prod_{i=1}^N \text{NB}(x_{\cdot ik}; r_k, p_k)$$

Example Results

Table 1: Algorithms related to PFA under various prior settings. The gamma scale and shape parameters and the sparsity of the factor scores are controlled by p_k, r_k and z_{ki}, respectively.

PFA priors	Infer p_k	Infer r_k	Infer z_{ki}	Infer K	Related algorithms
Γ	×	×	×	×	NMF
Dir	×	×	×	×	LDA
$\beta\Gamma$	✓	×	×	✓	GaP
S$\gamma\Gamma$	×	✓	✓	✓	FTM
$\beta\gamma\Gamma$	✓	✓	×	✓	

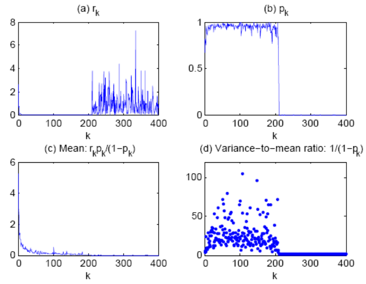

Figure 1: Inferred r_k, p_k, mean $r_kp_k/(1-p_k)$ and variance-to-mean ratio $1/(1-p_k)$ for each factor by $\beta\gamma\Gamma$-PFA with $a_\phi = 0.05$. The factors are shown in decreasing order based on the total number of word counts assigned to them. Of the $K_{\max} = 400$ possible factors, there are 209 active factors assigned nonzero counts. The results are obtained on the training count matrix of the PsyRev corpus based on the last MCMC iteration.

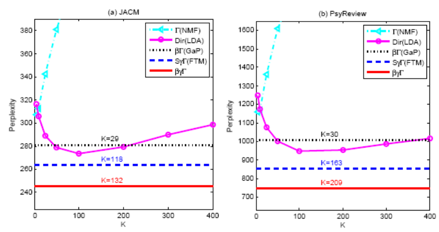

Figure 2: Per-word perplexities on the test count matrix for (a) JACM and (b) PsyRev with $a_\phi = 0.05$. The results of Γ- and Dir-PFAs are a function of K. $\beta\Gamma$-, S$\gamma\Gamma$- and $\beta\gamma\Gamma$-PFAs all automatically infer the number of active factors, with the number at the the last MCMC iteration shown on top of the corresponding lines.

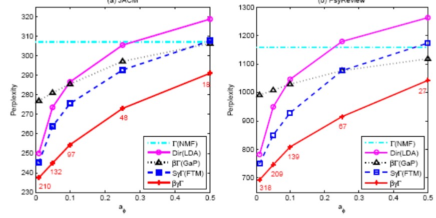

Figure 3: Per-word perplexities on the test count matrix of (a) JACM and (b) PsyRev as a function of the factor loading (topic) Dirichlet prior $a_\phi \in \{0.01, 0.05, 0.1, 0.25, 0.5\}$. The results of Γ- and Dir-PFAs are shown with the best settings of K under each a_ϕ. The number of active factors under each a_ϕ are all automatically inferred by $\beta\Gamma$-, S$\gamma\Gamma$- and $\beta\gamma\Gamma$-PFAs. The number of active factors inferred by $\beta\gamma\Gamma$-PFA at the last MCMC interaction are shown under the corresponding points.