Supplementary Material: Bayesian Nonparametric Poisson Factorization for Recommendation Systems:

Prem Gopalan Princeton University

Francisco J. R. Ruiz University Carlos III in Madrid

Rajesh Ranganath Princeton University

David M. Blei Princeton University

1 Expected inner product

We show here that the expected value of the inner product $\boldsymbol{\theta}_u^\top \boldsymbol{\beta}_i$ is finite:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}_u,\boldsymbol{\beta}_i}\left[\boldsymbol{\theta}_u^\top \boldsymbol{\beta}_i\right] &= \mathbb{E}_{\boldsymbol{\theta}_u,\boldsymbol{\beta}_i}\left[\sum_{k=1}^{\infty} \beta_{ik}\theta_{uk}\right] \\ &= \mathbb{E}_{\boldsymbol{\theta}_u}\left[\sum_{k=1}^{\infty} \theta_{uk}\mathbb{E}_{\boldsymbol{\beta}_i}\left[\beta_{ik}\right]\right] \\ &= \mathbb{E}_{\boldsymbol{\theta}_u}\left[\sum_{k=1}^{\infty} \theta_{uk}\frac{a}{b}\right] \\ &= \frac{a}{b}\mathbb{E}_{\boldsymbol{\theta}_u}\left[\sum_{k=1}^{\infty} \theta_{uk}\right],\end{aligned} \tag{1}$$

since each β_{ik} is independent and identically distributed as a Gamma with shape a and rate b. The subscript of each expectation indicates the variables over which the expectation applies.

In order to compute the expectation $\mathbb{E}_{\boldsymbol{\theta}_u}[\sum_{k=1}^{\infty}\theta_{uk}]$, we apply Campbell's theorem (Kingman, 1993), and take into account that the Poisson process intensity of the Gamma process $\mathrm{GP}(c, H)$ is given by $H \times \nu(d\theta)$, where

$$\nu(d\theta) = \theta^{-1}\exp(-c\theta)d\theta. \tag{2}$$

Therefore, we have:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}_u,\boldsymbol{\beta}_i}\left[\boldsymbol{\theta}_u^\top \boldsymbol{\beta}_i\right] &= \frac{a}{b}\int_{\theta,\Omega}\theta\nu(d\theta)H(d\omega) \\ &= \frac{a}{b}\int_0^{\infty}\theta\nu(d\theta)\int_{\Omega}H(d\omega) \\ &= \frac{a\alpha}{bc}\end{aligned} \tag{3}$$

2 Sparsity of the observation matrix

We compute the log-probability $p(y_{ui}=0)$ as

$$\begin{aligned}&\log p(y_{ui}=0) \\ &= \log\left(\mathbb{E}_{\boldsymbol{\theta}_u,\boldsymbol{\beta}_i}\left[\prod_{k=1}^{\infty} p(y_{uik}=0|\theta_{uk},\beta_{ik})\right]\right) \\ &= \log\left(\mathbb{E}_{\boldsymbol{\theta}_u,\boldsymbol{\beta}_i}\left[\prod_{k=1}^{\infty} e^{-\theta_{uk}\beta_{ik}}\right]\right) \\ &\geq \mathbb{E}_{\boldsymbol{\theta}_u,\boldsymbol{\beta}_i}\left[\log\left(\prod_{k=1}^{\infty} e^{-\theta_{uk}\beta_{ik}}\right)\right],\end{aligned} \tag{4}$$

where the last step follows from Jensen's inequality. Then, we can write:

$$\begin{aligned}\log p(y_{ui}=0) &\geq \mathbb{E}_{\boldsymbol{\theta}_u}\left[\sum_{k=1}^{\infty}\mathbb{E}_{\beta_{ik}}\left[-\theta_{uk}\beta_{ik}\right]\right] \\ &= -\mathbb{E}_{\boldsymbol{\theta}_u}\left[\sum_{k=1}^{\infty}\frac{a}{b}\theta_{uk}\right].\end{aligned} \tag{5}$$

Using the results in the previous section, this quantity is equal to $-a\alpha/bc$. Therefore, the expected number of zeros in the user behavior matrix is lower bounded by

$$\mathbb{E}\left[\sum_{u=1}^{N}\sum_{i=1}^{M}\mathrm{I}\left\{y_{ui}=0\right\}\right] \geq NM\exp\left(-\frac{a\alpha}{bc}\right). \tag{6}$$

($\mathrm{I}\{\cdot\}$ is the indicator function).

3 Inference: Updates for τ_{uk}

As detailed in the main text, the updates for the stick proportions τ_{uk} are given by a quadratic equation of the form $A_{uk}\tau_{uk}^2 + B_{uk}\tau_{uk} + C_{uk} = 0$. Here we provide the specific equations for the coefficients A_{uk}, B_{uk} and

C_{uk}. First,

$$\begin{aligned}A_{uk} =& \mathbb{E}\left[s_u\right] \sum_{\ell=k+1}^{T} \tau_{u\ell} \left(\prod_{j=1, j\neq k}^{\ell-1} (1-\tau_{uj}) \right) \\ & \times \left(\sum_{i=1}^{M} \mathbb{E}\left[\beta_{i\ell}\right] \right) \\ & - \mathbb{E}\left[s_u\right] \left(\prod_{j=1}^{k-1} (1-\tau_{uj}) \right) \left(\sum_{i=1}^{M} \mathbb{E}\left[\beta_{ik}\right] \right) \\ & + \mathbb{E}\left[s_u\right] M \frac{a}{b} \prod_{j=1, j\neq k}^{T} (1-\tau_{uj}).\end{aligned}$$

For B_{uk}, we have

$$B_{uk} = \alpha - 1 - C_{uk} - A_{uk} + \sum_{i=1}^{M} y_{ui} \left(1 - \sum_{j=1}^{k} \phi_{ui,j} \right),$$

being

$$C_{uk} = -\sum_{i=1}^{M} y_{ui} \phi_{ui,k}.$$

4 Hyperparameter selection for the MovieLens1M dataset

See Figure 1 for the results obtained on this database. The best performance is obtained with unit scale and any shape. We fix unit scale and shape for the rest of our experiments.

References

Kingman, J. F. C. (1993). *Poisson processes* (Vol. 3). New York: The Clarendon Press Oxford University Press. (Oxford Science Publications)

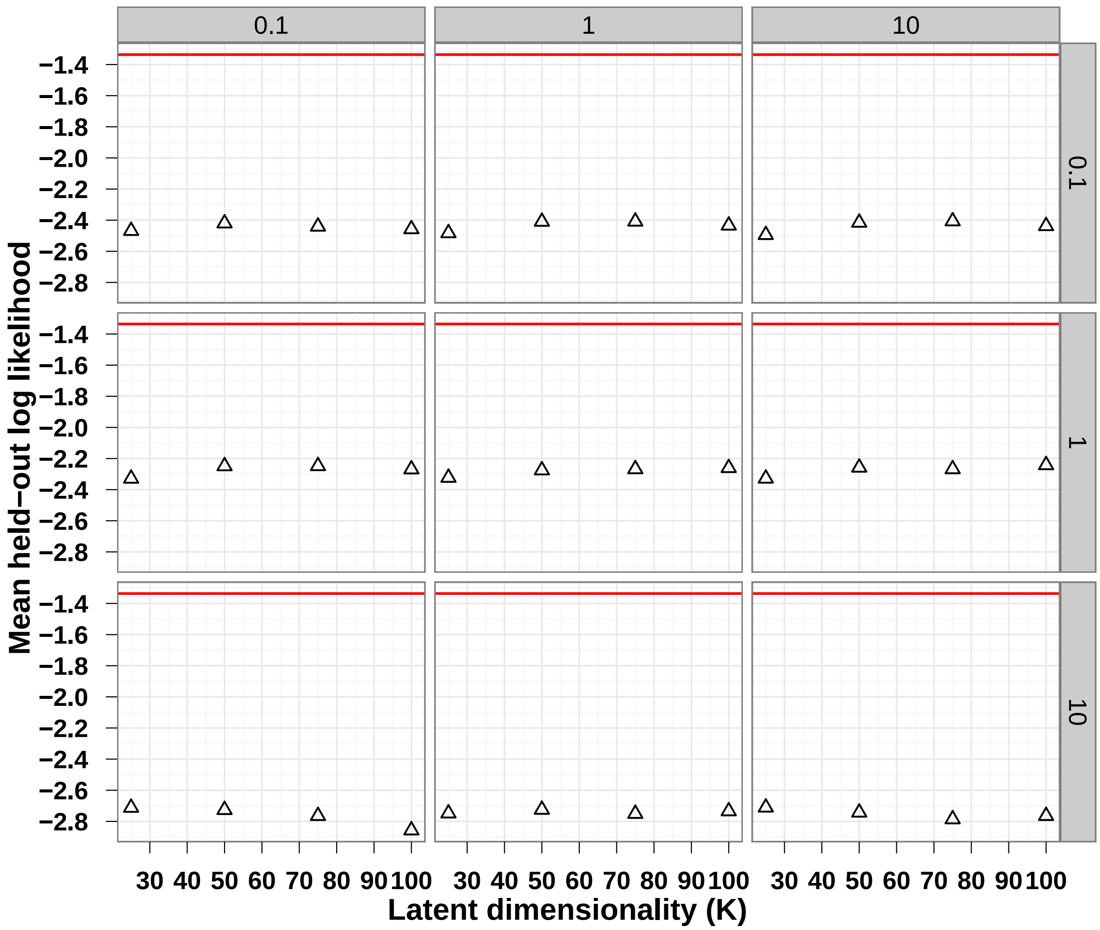

Figure 1: Mean held-out log-likelihood on the test set for the MovieLens1M dataset. The triangles show the performance of the finite model for varying values of K, while the horizontal lines show the performance of the BNPPF model. Each subfigure contains the results for a particular combination of the shape (top strip) and rate (right strip) Gamma hyperparameters of the user weights in the finite model.