Supplement to Hierarchical Topic Modeling for Analysis of Time-Evolving Personal Choices

XianXing Zhang
Duke University
xianxing.zhang@duke.edu

David B. Dunson
Duke University
dunson@stat.duke.edu

Lawrence Carin
Duke University
lcarin@ee.duke.edu

1 Retrospective sampling for change point $p_{i,t}$

Due to the non-conjugacy between the Poisson and multinomial distributions, the exact form of its posterior distribution is difficult to compute. Additionally, in order to sample $p_{i,t}$, we require imputation of an infinite-dimensional process, and implementation of the sampling algorithm often relies on finite approximations [1]. To address the above two problems, we design a retrospective sampler, which allows us to obtain samples from the exact posterior distribution of $p_{i,t}$, and no truncation approximation is required. The retrospective sampler was proposed originally in [2] as an inference engine for Dirichlet process hierarchical models, and in this paper we constitute another application by inferring the Poisson distributed change points $p_{i,t}$.

Denote $P_i = \max_t p_{i,t}$ as the inferred maximum value of the change point in the current sampling iteration; then, given the samples of all other latent variables, $p_{i,t}$ can be sampled from a Metropolis-Hastings (M-H) step, where the proposed $p^\star$ is generated from the following distribution:

$$q(p_{i,t} = p^\star | \boldsymbol{\theta}_{i,t}^{p^\star}, \lambda_{i,t}, \omega_{i,t}, \boldsymbol{l}_{i,t}) \propto \begin{cases} p(p_{i,t} = p^\star | \lambda_{i,t}) p(\boldsymbol{l}_{i,t} | \boldsymbol{\theta}_{i,t}^{p^\star}, \omega_{i,t}), & \text{for } p^\star \leq P_i \\ p(p_{i,t} = p^\star | \lambda_{i,t}) M(\boldsymbol{l}_{i,t}, P_i), & \text{for } p^\star > P_i \end{cases} \quad (1)$$

where $\boldsymbol{l}_{i,t} = \{l_{i,n} : t(i,n) = t\}$ are all layer allocations of choices made by individual i at time t. $p(p_{i,t} = p^\star | \lambda_{i,t}) = \text{Poi}(p_{i,t} = p^\star | \lambda_{i,t})$ follows a Poisson distribution, $p(\boldsymbol{l}_{i,t} | \boldsymbol{\theta}_{i,t}^{p^\star}, \omega_{i,t}) = \text{Mult}(\boldsymbol{l}_{i,t} | \{\omega_{i,t} \hat{\boldsymbol{\theta}}_{i,t}^{p^\star}, (1 - \omega_{i,t}) \tilde{\boldsymbol{\theta}}_{i,t}^{p^\star}\})$ is the multinomial density function over the layer allocations $\boldsymbol{l}_{i,t}$, and $M(\boldsymbol{l}_{i,t}, P_i)$ is chosen according to (2) to ensure that the tails of the proposal distribution (1) are heavier than the tails of the target probability which leads to faster mixing [2]:

$$M(\boldsymbol{l}_{i,t}, P_i) = \max_{p^\star \leq P_i} p(\boldsymbol{l}_{i,t} | \boldsymbol{\theta}_{i,t}^{p^\star}, \omega_{i,t}) \quad (2)$$

The normalization constant of (1) is

$$C(P_i) = \sum_{p=1}^{P_i} p(p_{i,t} = p | \lambda_{i,t}) p(\boldsymbol{l}_{i,t} | \boldsymbol{\theta}_{i,t}^{p}, \omega_{i,t}) + M(\boldsymbol{l}_{i,t}, P_i) \left(1 - \sum_{p=1}^{P_i} p(p_{i,t} = p | \lambda_{i,t}) \right)$$

The acceptance probability for the proposed $p_{i,t} = p^\star$ from the previous change point assignment $p_{i,t} = p$ from the last iteration is defined as

$$\alpha_{i,t}(p, p^\star) = \begin{cases} 1, & \text{if } p^\star \leq P_i \text{ and } P_i^\star = P_i \\ \min\{1, \frac{C(P_i) M(\boldsymbol{l}_{i,t}, P_i^\star)}{C(P_i^\star) p(\boldsymbol{l}_{i,t} | \boldsymbol{\theta}_{i,t}^{p}, \omega_{i,t})}\}, & \text{if } p^\star \leq P_i \text{ and } P_i^\star < P_i \\ \min\{1, \frac{C(P_i) p(\boldsymbol{l}_{i,t} | \boldsymbol{\theta}_{i,t}^{p^\star}, \omega_{i,t})}{C(P_i^\star) M(\boldsymbol{l}_{i,t}, P_i)}\}, & \text{if } p^\star > P_i \end{cases} \quad (3)$$

where $P_i^\star$ is the updated maximum value of change point after replace $p_{i,t} = p$ with $p_{i,t} = p^\star$. Through the M-H sampling step defined above, $p_{i,t}$ is updated from its exact posterior without resorting to any approximation. The detailed derivation and discussion of the retrospective sampler can be found in [2].

References

[1] H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.

[2] O. Papaspiliopoulos and G. O. Roberts. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchiacal models. *Biometrika*, 95(1):169–186, 2008.