Lemma 1 *If we draw N i.i.d. samples $x_1, x_2 \dots x_N$ through the generative process in Equation 1 (main paper) corresponding to N users, and the vectors probability mass function of the items y estimated from these N samples are $\hat{p}(y)$ whereas the true p.m.f is $p(y)$ with $y \in \{y_1, y_2 \dots y_D\}$, then with probability at least $1-\delta$ with $\delta \in (0,1)$,*

$$||\hat{p}(y) - p(y)||_F \leq \frac{2}{\tilde{d}_{1s}\sqrt{N}} \left(1 + \sqrt{\frac{\log(1/\delta)}{2}}\right) \tag{1}$$

$$||\hat{p}(y,y) - p(y,y)||_F \leq \frac{2}{\tilde{d}_{2s}\sqrt{N}} \left(1 + \sqrt{\frac{\log(1/\delta)}{2}}\right) \tag{2}$$

$$||\hat{p}(y,y,y) - p(y,y,y)||_F \leq \frac{2}{\tilde{d}_{3s}\sqrt{N}} \left(1 + \sqrt{\frac{\log(1/\delta)}{2}}\right) \tag{3}$$

where, $\tilde{d}_{1s} = \frac{1}{N}\sum_{i=1}^{N} nnz(x_i)$, $\tilde{d}_{2s} = \frac{1}{N}\sum_{i=1}^{N} nnz(x_i)^2$, $\tilde{d}_{3s} = \frac{1}{N}\sum_{i=1}^{N} nnz(x_i)^3$, and $nnz(x_i)$ is the non-zero entries in row x_i of the data X as described in section 3.

Proof The generative process in Equation 1 (main paper) results in samples $x_{1:N}$ that are vectors of count data, with $\sum_y [x_u]_d = n_u$, where x_u is the sample corresponding to the user u, and n_u is the sum of the counts of all the items for u. The operation $\sum_y$ denotes the sum across the dimensions. From here, we can show that $||x_u|| = \sqrt{\sum_y [x_u]_d^2} \leq \sum_y [x_u]_d = n_u$, since $[x_u]_d \geq 0, \forall d \in 1, 2 \dots D$. Therefore, the samples have bounded norm.

Without loss of generality, if we assume $||x|| \leq 1 \ \forall x \in X$, then from Lemma 7 of supplementary material of Wang and Zhu (2014), with probability at least $1-\delta$ with $\delta \in (0,1)$,

$$\left|\left|\hat{\mathbb{E}}[x] - \mathbb{E}[x]\right|\right|_F \leq \frac{2}{\sqrt{N}} \left(1 + \sqrt{\frac{\log(1/\delta)}{2}}\right) \tag{4}$$

$$\left|\left|\hat{\mathbb{E}}[x \otimes x] - \mathbb{E}[x \otimes x]\right|\right|_F \leq \frac{2}{\sqrt{N}} \left(1 + \sqrt{\frac{\log(1/\delta)}{2}}\right) \tag{5}$$

$$\left|\left|\hat{\mathbb{E}}[x \otimes x \otimes x] - \mathbb{E}[x \otimes x \otimes x]\right|\right|_F \leq \frac{2}{\sqrt{N}} \left(1 + \sqrt{\frac{\log(1/\delta)}{2}}\right) \tag{6}$$

where $\mathbb{E}$ stands for true expectation, and $\hat{\mathbb{E}}$ stands for the expectation estimated from the N samples, i.e.,

$$\hat{\mathbb{E}}[x] = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N} X^\top \mathbf{1}$$

$$\hat{\mathbb{E}}[x \otimes x] = \frac{1}{N}\sum_{i=1}^{N} x_i \otimes x_i = \frac{1}{N} X^\top X$$

$$\hat{\mathbb{E}}[x \otimes x \otimes x] = \frac{1}{N}\sum_{i=1}^{N} x_i \otimes x_i \otimes x_i = \frac{1}{N} X \otimes X \otimes X$$

Now, since each of our samples $x_{1:N}$ contains binary data, probability of the items can be estimated from the training data as $\hat{p}(y) = \frac{\hat{\mathbb{E}}[x]}{\sum_y \hat{\mathbb{E}}[x]}$, where $\sum_y \hat{\mathbb{E}}[x]$ is the sum of $\hat{\mathbb{E}}[x]$ across the dimensions, i.e., all the items. Also, it can be shown that $\sum_y \hat{\mathbb{E}}[x] = \tilde{d}_{1s}$. Therefore $\hat{p}(y) = \frac{\hat{\mathbb{E}}[x]}{\tilde{d}_{1s}}$. Please note that $\sum_y \mathbb{E}[x] \approx \sum_y \hat{\mathbb{E}}[x] = \tilde{d}_{1s}$, and therefore, $\hat{p}(y) - p(y) = \frac{1}{\tilde{d}_{1s}}(\hat{\mathbb{E}}[x] - \mathbb{E}[x])$, and using this in Equation 4, we get the first inequality of the Lemma (Equation 1).

Since $\tilde{d}_{2s} = \sum_y \sum_y \hat{\mathbb{E}}[x \otimes x]$ and $\tilde{d}_{3s} = \sum_y \sum_y \sum_y \hat{\mathbb{E}}[x \otimes x \otimes x]$, the pairwise and triple-wise probability matrices can be estimated as,

$$\hat{p}(y, y) = \frac{\hat{\mathbb{E}}[x \otimes x]}{\sum_y \sum_y \hat{\mathbb{E}}[x \otimes x]} = \frac{\hat{\mathbb{E}}[x \otimes x]}{\tilde{d}_{2s}}$$
$$\hat{p}(y, y, y) = \frac{\hat{\mathbb{E}}[x \otimes x]}{\sum_y \sum_y \sum_y \hat{\mathbb{E}}[x \otimes x \otimes x]} = \frac{\hat{\mathbb{E}}[x \otimes x \otimes x]}{\tilde{d}_{3s}}$$

Since $\sum_y \sum_y \mathbb{E}[x \otimes x] \approx \sum_y \sum_y \hat{\mathbb{E}}[x \otimes x] = \tilde{d}_{2s}$, and $\sum_y \sum_y \sum_y \mathbb{E}[x \otimes x \otimes x] \approx \sum_y \sum_y \sum_y \hat{\mathbb{E}}[x \otimes x \otimes x] = \tilde{d}_{3s}$, we can establish the following equations,

$$\hat{p}(y, y) - p(y, y) = \frac{1}{\tilde{d}_{2s}} \left(\hat{\mathbb{E}}[x \otimes x] - \mathbb{E}[x \otimes x] \right)$$
$$\hat{p}(y, y, y) - p(y, y, y) = \frac{1}{\tilde{d}_{3s}} \left(\hat{\mathbb{E}}[x \otimes x \otimes x] - \mathbb{E}[x \otimes x \otimes x] \right)$$

Substituting these equations in Equation 5 and 6, we complete the proof. ■

References

Yining Wang and Jun Zhu. Spectral methods for supervised topic models. In *Advances in Neural Information Processing Systems*, pages 1511–1519, 2014. URL https://papers.nips.cc/paper/5517-spectral-methods-for-supervised-topic-models.