Stochastic Tensor Block Model
– Statistical Limits

Jiaxin Hu

1 Model

- **Model.** Consider an order-K binary tensor $\mathcal{Y} = [\![\mathcal{Y}_{i_1,\ldots,i_K}]\!] \in \{0,1\}^{p_1\times\cdots\times p_K}$. Suppose there are r_k communities for the p_k nodes on mode k for all $k\in[K]$. We assume that the entries in $\mathcal{Y}$ follow independent Bernoulli distribution such that

$$\mathbb{P}(\mathcal{Y}_{i_1,\ldots,i_K}=1)=\mathcal{S}_{z_1(i_1),\ldots,z_K(i_K)}, \tag{1}$$

where tensor $\mathcal{S}\in[0,1]^{r_1\times\cdots\times r_K}$ collects community means, and $z_k:[p_k]\mapsto[r_k], k\in[K]$ are community assignment functions. Equivalently, we are able to write the model (1) in a tensor form:

$$\mathbb{E}[\mathcal{Y}]=\mathcal{S}\times_1 \boldsymbol{M}_1\times_2\cdots\times_K \boldsymbol{M}_K, \tag{2}$$

where $\boldsymbol{M}_k\in\{0,1\}^{p_k\times r_k}$ are membership matrices corresponding to the assignment z_k for all $k\in[K]$. We call the model (2) *stochastic tensor block model* (STBM).

- **Notation.**
  - $\mathrm{Mat}_k(\mathcal{S})\in\mathbb{R}^{p_k\times\prod_{l\neq k}p_l}$: the k-th mode matricization of $\mathcal{S}$; we use $\boldsymbol{S}_k=\mathrm{Mat}_k(\mathcal{S})$ to denote the matricization of the core tensor throughout the note;
  - $\boldsymbol{S}_{a:}\in\mathbb{R}^n$: the a-th row of matrix $\boldsymbol{S}\in\mathbb{R}^{m\times n}$;

- **Identifiability.**

Theorem 1 (Identifiability (Proposition 1 in Wang and Zeng (2019)))**.** *The parameterization $(\mathcal{S},\boldsymbol{M}_k)$ in model (2) is identifiable up to label permutation if $\mathcal{S}$ is an irreducible core; i.e., $\boldsymbol{S}_{k,a:}\neq\boldsymbol{S}_{k,b:}$ for all $a\neq b\in[r_k], k\in[K]$.*

2 Statistical limit

- **Parameter space.** We consider the fundamental limit in the following parameter space for $(\mathcal{S},z_k)$.

$$\mathcal{P}=\{(\mathcal{S},z_k):\mathcal{S}\in[0,1]^{r_1\times\cdots\times r_K}, c_1p_k/r_k\leq|z_k^{-1}(a)|\leq c_2p_k/r_k, a\in[r_k],k\in[K]\},$$

where $c_1<c_2$ are two positive constants. The space $\mathcal{P}$ requires balanced cluster sizes on every mode. For simplicity, we discuss the statistical limit under the special case

$$p_k=p,\quad r_k=r.$$

- **Misclassification error.** Let $\hat{z}$ and z^* denote the estimated and true clustering assignment. We define following misclassification error to evaluate the performance of $\hat{z}$

$$\ell(\hat{z}, z^*) = \frac{1}{p} \min_{\pi \in \Pi_p} \sum_{i \in [p]} \mathbb{1}\{\hat{z}(i) \neq \pi \circ z^*(i)\}.$$

- **Signal.** Define the signal of $\mathcal{S}$ as

$$\Delta^2_{\min} = \min_{k \in [K]} \Delta^2_k, \quad \Delta^2_k = \min_{a \neq b \in [r_k]} \| \boldsymbol{S}^{1/2}_{k,a:} - \boldsymbol{S}^{1/2}_{k,b:} \|^2.$$

  Under the matrix "assortative" setting in Gao et al. (2018), i.e., $\min_{i \in [p]} \boldsymbol{S}_{ii} = s_1 > s_2 = \max_{i \neq j \in [p]} \boldsymbol{S}_{ij}$, the signal is equal to $\Delta^2_{\min} \geq 2(\sqrt{s_1} - \sqrt{s_2})^2$, which serves as the signal term in the error bounds.

  We define the parameter space with a particular signal level:

$$\mathcal{P}(\Delta^2_{\min}) = \mathcal{P} \cap \{\text{parameter } \mathcal{S} \text{ has signal } \Delta^2_{\min}\}.$$

Theorem 2 (Statistical lower bound of STBM)**.** *Consider the STBM (2) with parameter $(\mathcal{S}, z_k) \in \mathcal{P}$. We have minimax lower bound*

$$\lim_{p \to \infty} \inf_{\hat{z}_k} \sup_{(\mathcal{S}, z_k) \in \mathcal{P}(\Delta^2_{\min})} \mathbb{E}[\ell(\hat{z}_k, z_k)] \geq \exp\left(-Cp^{K-1} \Delta^2_{\min}\right)$$

Proof sketch of Theorem 2. It suffices to show that for a particular $(\mathcal{S}^*, z^*_k) \in \mathcal{P}(\Delta^2_{\min})$

$$\inf_{\hat{z}_1} \mathbb{E}[\ell(\hat{z}_1, z^*_1) | (\mathcal{S}^*, z^*_k)] \gtrsim \exp\left(-Cp^{K-1} \Delta^2_{\min}\right).$$

We construct a $\mathcal{S}^*$ such that $\Delta^2_{\min} = \| \boldsymbol{S}^{1/2}_{1,1:} - \boldsymbol{S}^{1/2}_{1,2:} \|^2$. With the construction of z^*_k in Gao et al. (2018); Han et al. (2020), we have

$$\inf_{\hat{z}_1} \mathbb{E}[\ell(\hat{z}_1, z^*_1) | (\mathcal{S}^*, z^*_k)] \gtrsim \inf_{\hat{z}(1)} \mathbb{P}(\hat{z}(1) = 1 | z^*_1(1) = 2) + \mathbb{P}(\hat{z}(1) = 2 | z^*_1(1) = 1).$$

It is equivalent to consider the hypothesis test

$$H_0 : \boldsymbol{y} \sim \text{Ber}(\boldsymbol{x}^*_1) \quad \leftrightarrow \quad H_1 : \boldsymbol{y} \sim \text{Ber}(\boldsymbol{x}^*_2), \tag{3}$$

where $\boldsymbol{y} = \text{Mat}_1(\mathcal{Y})_{1:} \in \{0,1\}^{p^{K-1}}$ is the Bernoulli observation, $\boldsymbol{x}^*_a = \text{Mat}_1(\mathcal{S}^* \times_2 \boldsymbol{M}^*_2 \times \cdots \times_K \boldsymbol{M}^*_K)_{a:} \in [0,1]^{p^{K-1}}, a \in [r]$ are the true mean vectors of the random Bernoulli observation.

Lemma 1 is the key lemma that describes the sum of Type I + II error of the optimal test for the hypothesis testing (3). The optimal error is obtained by the Neyman-Pearson Lemma and the Carmer-Chernoff Theorem.

Therefore, we have Theorem 2 by the construction of $(\mathcal{S}^*, z^*_k)$ and Lemma 1.

$\square$

Lemma 1 (Optimal Type I + II error)**.** *Consider the hypothesis testing (3). For $p \to \infty$, we have*

$$\inf_{\phi} \mathbb{P}_{H_0}(\phi) + \mathbb{P}_{H_1}(\phi) \geq \exp(-Cp^{K-1} \Delta^2_{\min}).$$

Proof of Lemma 1. For simplicity, we use $\mathcal{S}, z_k$ for true parameters. By Neyman-Pearson Lemma, likelihood ratio test is the uniform most powerful test of the testing (3). The likelihood function of $\boldsymbol{y}$ with mean $\boldsymbol{x}_1$ is

$$\mathcal{L}(\boldsymbol{x}_1, \boldsymbol{y}) = \prod_{a_2\in[r_2],\ldots,a_K\in[r_K]} \prod_{i_2\in z_2^{-1}(a_2),\ldots,i_K\in z_K{}^{-1}(a_K)} \mathcal{S}_{1,a_2,\ldots,a_K}^{\mathcal{Y}_{1,i_2,\ldots,i_K}} (1-\mathcal{S}_{1,a_2,\ldots,a_K})^{1-\mathcal{Y}_{1,i_2,\ldots,i_K}}.$$

For simplicity, we use $a \in [R]$ with $R = \prod_{k=2}^K r_k$ to represent the index $(a_2, \ldots, a_K)$ and use $i \in [p_a]$ with $p_a = \prod_{k=2}^K |z_k^{-1}(a_k)|$ to represent the index $(i_2, \ldots, i_K)$. Let $\boldsymbol{S} = \mathrm{Mat}_1(\mathcal{S})$. Then, the log-likelihood is

$$\ell(\boldsymbol{x}_1, \boldsymbol{y}) = \sum_{a\in[R]} \sum_{i\in[p_a]} \left(\boldsymbol{y}_i \log \frac{\boldsymbol{S}_{1a}}{1-\boldsymbol{S}_{1a}} + \log(1-\boldsymbol{S}_{1a}) \right).$$

Hence, we have

$$\begin{aligned}\mathbb{P}(\ell(\boldsymbol{x}_1, \boldsymbol{y}) < \ell(\boldsymbol{x}_2, \boldsymbol{y}) &= \mathbb{P}\left(\sum_{a\in[R]} \sum_{i\in[p_a]} \left(\boldsymbol{y}_i \log \frac{\boldsymbol{S}_{2a}(1-\boldsymbol{S}_{1a})}{\boldsymbol{S}_{1a}(1-\boldsymbol{S}_{2a})} + \log \frac{(1-\boldsymbol{S}_{2a})}{(1-\boldsymbol{S}_{1a})} \right) > 0 \right) \\ &= \mathbb{P}\left(\sum_{a\in[R]} \sum_{i\in[p_a]} W_i^{(a)} > 0 \right), \qquad (4)\end{aligned}$$

where $W_i^{(a)}$ has distribution

$$\mathbb{P}\left(W_i^{(a)} = t \log \frac{\boldsymbol{S}_{2a}}{\boldsymbol{S}_{1a}} \right) = \boldsymbol{S}_{1a}, \quad \mathbb{P}\left(W_i^{(a)} = t \log \frac{1-\boldsymbol{S}_{2a}}{1-\boldsymbol{S}_{1a}} \right) = 1 - \boldsymbol{S}_{1a}$$

for some $t \in (0, 1]$. To lower bound the inequality (4), we are not able to directly apply Cramer-Chernoff Theorem Van der Vaart (2000, Proposition 14.23) due to the non-iidness of W_i. But we are able to adopt the same proof idea in Cramer-Chernoff Theorem.

Consider

$$M_W(a, t) = \mathbb{E}[\exp(W_i^{(a)})] = \boldsymbol{S}_{2a}^t \boldsymbol{S}_{1a}^{1-t} + (1-\boldsymbol{S}_{2a})^t (1-\boldsymbol{S}_{1a})^{1-t}.$$

The term $M_W(a, t)$ can be considered as the moment generating function of the random variable $W_i^{(a)}$ with $t = 1$ and with respect to the probability $\mathbb{P}$.

Note that

$$\begin{aligned}\mathbb{P}\left(\sum_{a\in[R]} \sum_{i\in[p_a]} W_i^{(a)} > 0 \right) &\geq \sum_{0<\sum_{a,i} w_i^{(a)} \leq L} \prod_{a,i} \mathbb{P}(W_i^{(a)} = w_i^{(a)}) \\ &\geq \left(\prod_a M_W^{p_a}(a, t) \right) e^{-L} \sum_{0<\sum_{a,i} w_i^{(a)} \leq L} \prod_{a,i} \frac{\mathbb{P}(W_i^{(a)} = w_i^{(a)}) e^{w_i^{(a)}}}{M_W(a, t)} \\ &= \left(\prod_a M_W^{p_a}(a, t) \right) e^{-L} \sum_{0<\sum_{a,i} w_i^{(a)} \leq L} \prod_{a,i} \mathbb{Q}(W_i^{(a)} = w_i^{(a)}) \\ &= \left(\prod_a M_W^{p_a}(a, t) \right) e^{-L} \mathbb{Q}\left(0 < \sum_{a,i} w_i^{(a)} \leq L \right),\end{aligned}$$

where

$$\mathbb{Q}\left(W_i^{(a)} = t\log\frac{\boldsymbol{S}_{2a}}{\boldsymbol{S}_{1a}}\right) = \frac{\mathbb{P}(W_i^{(a)} = t\log\frac{\boldsymbol{S}_{2a}}{\boldsymbol{S}_{1a}})e^{t\log\frac{\boldsymbol{S}_{2a}}{\boldsymbol{S}_{1a}}}}{M_W(a,t)} = \frac{\boldsymbol{S}_{2a}^t\boldsymbol{S}_{1a}^{1-t}}{M_W(a,t)}$$

$$\mathbb{Q}\left(W_i^{(a)} = t\log\frac{1-\boldsymbol{S}_{2a}}{1-\boldsymbol{S}_{1a}}\right) = \frac{\mathbb{P}(W_i^{(a)} = t\log\frac{1-\boldsymbol{S}_{2a}}{1-\boldsymbol{S}_{1a}})e^{t\log\frac{1-\boldsymbol{S}_{2a}}{1-\boldsymbol{S}_{1a}}}}{M_W(a,t)} = \frac{(1-\boldsymbol{S}_{2a})^t(1-\boldsymbol{S}_{1a})^{1-t}}{M_W(a,t)}.$$

The measurement $\mathbb{Q}$ is a valid distribution since $\mathbb{Q}\left(W_i^{(a)} = t\log\frac{\boldsymbol{S}_{2a}}{\boldsymbol{S}_{1a}}\right)+\mathbb{Q}\left(W_i^{(a)} = t\log\frac{1-\boldsymbol{S}_{2a}}{1-\boldsymbol{S}_{1a}}\right) = 1.$

First, we lower bound the term $e^{-L}\mathbb{Q}\left(0 < \sum_{a,i} w_i^{(a)} \le L\right)$ with some particular t and L.

□

References

Gao, C., Ma, Z., Zhang, A. Y., and Zhou, H. H. (2018). Community detection in degree-corrected block models. *The Annals of Statistics*, 46(5):2153–2185.

Han, R., Luo, Y., Wang, M., and Zhang, A. R. (2020). Exact clustering in tensor block model: Statistical optimality and computational limit. *arXiv preprint arXiv:2012.09996*.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Wang, M. and Zeng, Y. (2019). Multiway clustering via tensor block models. *Advances in neural information processing systems*, 32.