Simulation results 3

Chanwoo Lee, September 2, 2021

1 Dealing with missing values

We consider the noisy observed tensor $\mathcal{Y}$ with possible missing entries. We use Ω to denote the set of observed indices in this section.

1.1 Borda count estimation

We modify Borda count estimation to handle the possible missingness. First, Define permuted set of observed indices given a permutation π,

$$\Omega_\pi = \{(\pi^{-1}(i_1), \ldots, \pi^{-1}(i_m) \colon (i_1 \ldots, i_m) \in \Omega\}.$$

We redefine the function τ which is empirical score function of g as

$$\tau(i|\pi) = \frac{1}{|\eta_i^\pi|} \sum_{(i_2,\ldots,i_m)\in\eta_i^\pi} \mathcal{Y}_{\pi^{-1}(i),i_2,\ldots,i_m}.$$

$\eta_i^\pi = \{(i_2, \ldots, i_m) \colon (\pi^{-1}(i), \pi^{-1}(i_2), \ldots, \pi^{-1}(i_m)) \in \Omega_\pi\}$.

Similar to complete observation, our estimation consists of two steps.

1. (Sorting stage): We find permutation $\hat{\pi}^{\mathrm{BC}}$ which makes $\tau(\cdot|\pi)$ monotonically increasing.

$$\tau(1|\hat{\pi}^{\mathrm{BC}}) \leq \cdots \leq \tau(d|\hat{\pi}^{\mathrm{BC}}).$$

   Then, we obtain the rearranged observation $\tilde{\mathcal{Y}}_{\Omega_{\hat{\pi}^{\mathrm{BC}}}}$,

$$\tilde{\mathcal{Y}}_{i_1,\ldots,i_m} = \mathcal{Y}_{(\hat{\pi}^{\mathrm{BC}})^{-1}(i_1),\ldots,(\hat{\pi}^{\mathrm{BC}})^{-1}(d)},$$

   for all $(i_1, \ldots, i_m) \in \Omega_{\hat{\pi}^{\mathrm{BC}}}$.

2. (Block-wise polynomial approximation stage): Now, we estimate the block-wise polynomial tensor based on the rearranged observation $\tilde{\mathcal{Y}}_{\Omega_{\hat{\pi}^{\mathrm{BC}}}}$.

$$\hat{\Theta}^{\mathrm{BC}} = \mathop{\arg\min}_{\mathcal{B}\in\mathscr{B}(k,\ell)} \|P_{\Omega_{\hat{\pi}^{\mathrm{BC}}}}(\tilde{\mathcal{Y}} - \Theta)\|_F^2.$$

1.2 Spectral estimation

We modify spectral estimation to handle the possible missingness here. Denote the unfolded set of indices Ω as $\Omega_1 \in [d] \times [d^2]$. Then, we solve

$$\hat{\boldsymbol{Z}} = \mathop{\arg\min}_{\boldsymbol{Z}\in\mathbb{R}^{d\times d^2}} \frac{1}{2}\|P_{\Omega_1}(\mathcal{M}(\mathcal{Y}) - \boldsymbol{Z})\|_F^2 + \lambda\|\boldsymbol{Z}\|_*$$

From $\hat{\boldsymbol{Z}}$, we take the singular vectors whose singular values are greater than $\sqrt{d}$ similar to [3]. Finally, we fold back the tensor obtaining $\hat{\Theta}^{\mathrm{SP}}$.

1.3 Other alternatives

There are several other possible alternative methods that we can compare. For example, Sort and smoothing algorithm (SAS) from [1] or biclustering algorithm in [2]. We can unfold tensor into matrix and apply their methods.

However, SAS does not have good performance because this algorithm is designed for $d \times d$ matrix not $d \times d^2$. Either sorting procedure on only row or both row and column does not give us good result. Above all, unfolding procedure destroys the monotonic degree assumption so we cannot expect the great performance.

Biclustering algorithm also does not show good simulation results. This algorithm has the same local minimum issue to find the clustering function z. [2] suggests crietrion to update the each clustering function z_1 and z_2 for row and column respectively as

$$z_1(i) = \arg\min_{a \in [k_1]} \sum_{j=1}^{d_2} (S_{az_2(j)} - Y_{ij})^2$$

$$z_2(i) = \arg\min_{a \in [k_2]} \sum_{i=1}^{d_1} (S_{z_1(i)a} - Y_{ij})^2,$$

where $S \in \mathbb{R}^{k_1 \times k_2}$ is a block mean tensor. First, I try to apply this method on unfolded matrix with $(k_1, k_2) = (k, k^2)$ and $(d_1, d_2) = (d, d^2)$ and fold back to tensor. However, this approach does not give meaningful results.

Second, I try to use symmetric tensor version of this criteria to estimate clustering function z as

$$z(i) = \arg\min_{a \in [k]} \sum_{i_2, i_3 = 1}^{d} (S_{a, z(i_2), z(i_3)} - Y_{i_1, i_2 i, _3})^2$$

The performance of this algorithm turns out to be similar to LSE with finidng z by spectral k-means or HSC, which are much worse than Borda count estimation and spectral estimation.

2 Observational fraction versus MSE

By the reasons in Section 1.3, I only compare our method (Borda count estimation) with spectral method. I fix the $d = 40$ and gradually increase the observation fraction $|\Omega|/d^3$ from 0.4 to 1. First simulation is based on "Simulation for the binary-valued observations" in previous meeting note ("Simulation result 2"). Figure 1 shows that our method achieves the smallest error for all scenerios. Secondly I also checked the performance when we consider the continuous valued observation. In this case, I observe that there are ups and downs between our method and spectral method as one can see in Figure 2. I will check whether this patterns will change when I increase the tensor dimension $d > 40$.

References

[1] Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pages 208–216, 2014.

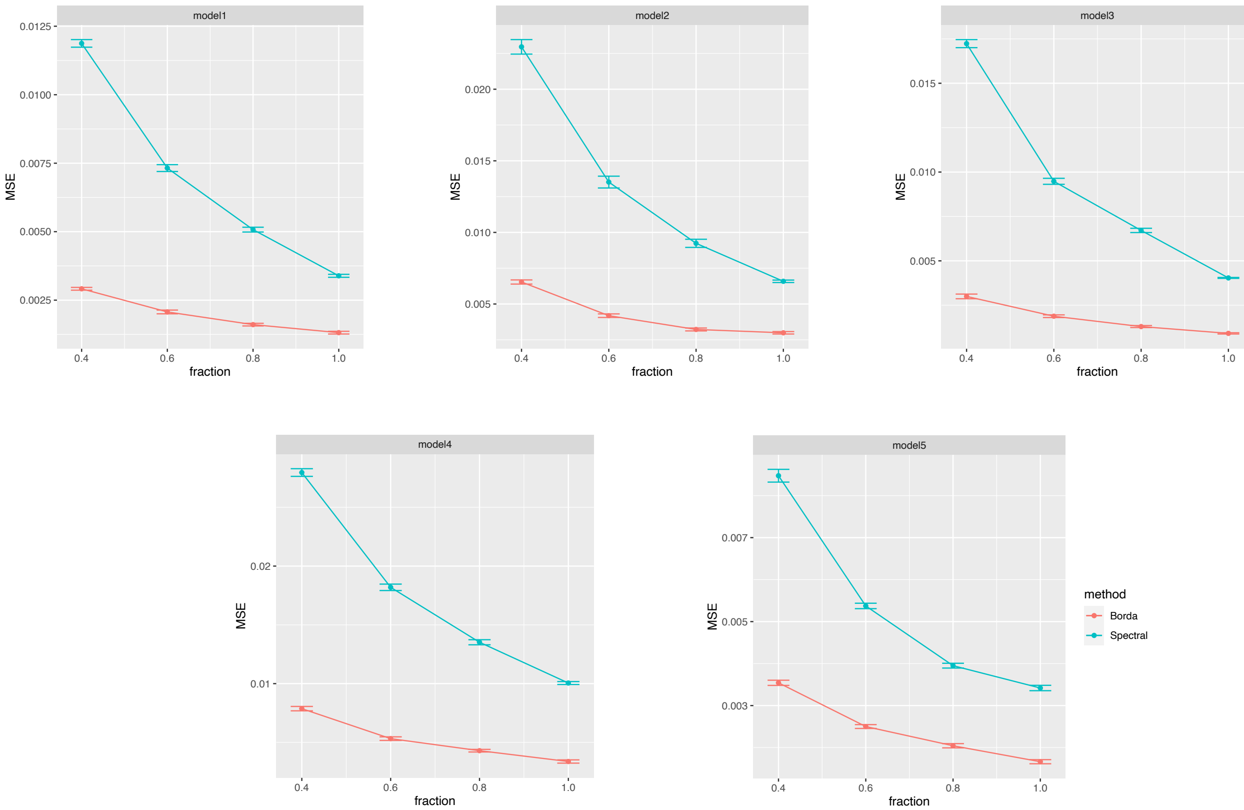

Figure 1: Estimation error versus observation fraction according to 5 different models and 2 different estimation methods. The observed tensor is generated by Bernoulli trial from generated graphons.

[2] Chao Gao, Yu Lu, Zongming Ma, and Harrison H Zhou. Optimal estimation and completion of matrices with biclustering structures. *Journal of Machine Learning Research*, 17(1):5602–5630, 2016.

[3] Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pages 5433–5442, 2018.

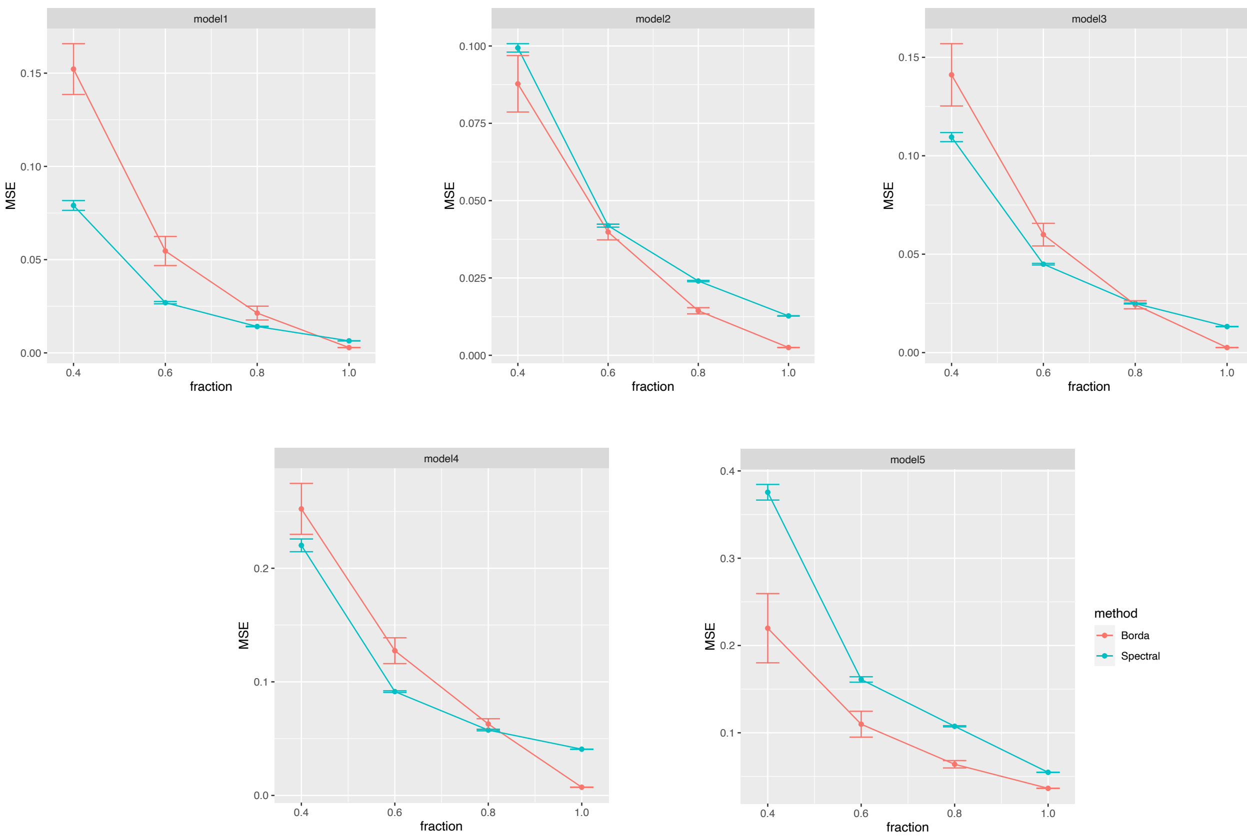

Figure 2: Estimation error versus observation fraction according to 5 different models and 2 different estimation methods. The observed tensor is generated by signal plus noise model where signal is generated by symmetric functions and noise by symmetirc normal tensors.