Estimation Error for Intercept Case

Jiaxin Hu

June 28, 2021

1 Hard Constraint

Consider the optimization problem

$$\begin{aligned}
\min_{U,\Theta_r} \quad & \mathcal{L}(U,\Theta_r) = \sum_{k=1}^{K} \langle S_k, \Omega_k \rangle - \log\det(\Omega_k) \\
s.t. \quad & \Omega_k = \Theta_0 + \sum_{r=1}^{R} u_{kr}\Theta_r, \\
& \|\Theta_0\|_0 \leq s_0, \quad \|\Theta_r\|_0 \leq s_r \\
& \|u_{\cdot r}\|_F = 1, \quad \sum_{k=1}^{K} u_{kr} = 0, \quad \text{for all } r \in [R].
\end{aligned}$$

Notations.

1. Let U^*, Θ_r^*, I_r^* denote the true parameters and membership.
2. Let $I_r = \{k \in [K] : u_{kr} \neq 0\}$ collects the categories that belong to group r with given membership U, and $I_{ar} = \{k \in [K] : u_{kr}, u_{ka}^* \neq 0\}$ collects the categories that belong to group r and true group a with given membership U and the true membership U^*.
3. Let $\Sigma_k = (\Theta_0^* + u_{kr}^*\Theta_r^*)^{-1}$ be the true precision matrix for $k \in I_r^*$.
4. Let $0 < \min_{k\in[K]} \varphi_{\min}(\Sigma_k) \leq \max_{k\in[K]} \varphi_{\max}(\Sigma_k) < \tau^{-1}$.
5. Let $\Delta_0 = \Theta_0 - \Theta_0^*$, $\Delta_{ar} = \Theta_r - \Theta_a^*$, and $\Delta_{k,ar} = \Delta_0 + u_{kr}\Theta_r - u_{ka}^*\Theta_a^*$.

Lemma 1. *There exists a local minimizer for the optimization problem 1 satisfies the following inequalities simultaneously with high probability.*

$$\|\Delta_0\|_F \leq M_0\sqrt{\frac{s_0 \log p}{nK}}, \quad \|\Delta_{ar}\|_F \leq M_{ar}\sqrt{\frac{(s_r + s_a)\log p}{n|I_{ar}|}}, \quad |u_{kr} - u_{ka}^*| \leq M_k\sqrt{\frac{s_r \log p}{n}},$$

for $k \in I_{ar}, a, r \in [R]$ and some large positive constants M_0, M_{ar}, M_k.

Remark 1. The above lemma approximately agrees with the heuristic that

$$\left\|\hat{\theta}-\theta^*\right\|_F^2 = \frac{\text{degree of freedom}}{\text{sample size}}.$$

Note that there are nK samples include the intercept matrix, $|I_{rr}n|$ samples contribute to the estimation of Θ_r, and only n samples contributes to the estimation of u_{kr}. Thus, the inequality of u_{kr} may be further sharpened.

Proof. Consider the estimate (U, Θ_r) and the true parameters (U^*, Θ_r^*). Define the function

$$G(U, \Theta_r) = \mathcal{L}(U, \Theta_r) - \mathcal{L}(U^*, \Theta_r^*).$$

Note that $G(U^*, \Theta_r^*) = 0$. Therefore, our goal is to find a set $\mathcal{A}$ such that when $(U, \Theta_r) \in \partial\mathcal{A}$ we have $G(U, \Theta_r) > 0$. Thus, there exists a local minimizer inside the set $\mathcal{A}$. For simplicity, we does not consider the group only with intercept, and we assume $I_{ar} > 0$ for all $a, r \in [R]$. In next step, we may consider the group only with intercept and the case with $I_{ar} = 0$.

Rewrite the function G, we have

$$\begin{aligned} G(U, \Theta_r) &= \sum_{r=1}^{R}\sum_{a=1}^{R}\left[\sum_{k\in I_{ar}} \langle S_k, \Delta_{k,ar}\rangle - \log\det(\Theta_0 + u_{kr}\Theta_r) + \log\det(\Theta_0^* + u_{ka}^*\Theta_a^*)\right] \\ &\geq I_1 + I_2, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \sum_{r=1}^{R}\sum_{a=1}^{R}\sum_{k\in I_{ar}} \langle S_k - \Sigma_k, \Delta_{k,ar}\rangle \\ I_2 &= \frac{1}{4\tau^2}\sum_{r=1}^{R}\sum_{a=1}^{R}\sum_{k\in I_{ar}} \|\Delta_{k,ar}\|_F^2. \end{aligned}$$

For the first term, we have

$$I_1 = \langle \sum_{k\in[K]} (S_k - \Sigma_k), \Delta_0\rangle + \sum_{r=1}^{R}\sum_{a=1}^{R}\langle \sum_{k\in I_{ar}} u_{ka}^*(S_k - \Sigma_k), \Delta_{ar}\rangle + \sum_{r=1}^{R}\sum_{a=1}^{R}\sum_{k\in I_{ar}} (u_{kr} - u_{ka}^*)\langle S_k - \Sigma_k, \Theta_r\rangle.$$

By Lemma 2, with high probability, we have

$$\begin{aligned} \left\|\sum_{k\in[K]} (S_k - \Sigma_k)\right\|_{\max} &\leq C_0\sqrt{\frac{\log pK}{n}}, \\ \left\|\sum_{k\in I_{ar}} u_{ka}^*(S_k - \Sigma_k)\right\|_{\max} &\leq C_{ar}\sqrt{\frac{\log p|I_{ar}|}{n}}, \\ \|(S_k - \Sigma_k)\|_{\max} &\leq C_k\sqrt{\frac{\log p}{n}}, \end{aligned}$$

for positive constants $C_0, C_{ar}, C_k, a, r \in [R], k \in [K]$. By the inequality $|\langle A, B\rangle| \le \|A\|_{\max} \|B\|_1$ and the fact that $\|\Delta\|_1 \le \sqrt{\|\Delta\|_0} \|\Delta\|_F$, we obtain the lower bound for I_1,

$$\begin{aligned} I_1 \ge & -C_0\sqrt{\frac{2s_0 \log pK}{n}} \|\Delta_0\|_F - \sum_{r=1}^R \sum_{a=1}^R C_{ar}\sqrt{\frac{(s_r+s_a)\log p|I_{ar}|}{n}} \|\Delta_{ar}\|_F \\ & - \sum_{r=1}^R \sum_{a=1}^R \sum_{k\in I_{ar}} |u_{kr} - u_{ka}^*| C_k \sqrt{\frac{s_r \log p}{n}} \|\Theta_r\|_F \end{aligned}$$

For the second term, we have

$$\|\Delta_{k,ar}\|_F^2 = \|\Delta_0\|_F^2 + \|u_{ka}^*\Delta_{ar} + (u_{kr} - u_{ka}^*)\Theta_r\|_F^2 + 2\langle \Delta_0, u_{kr}\Theta_r - u_{ka}^*\Theta_a^*\rangle.$$

Note that

$$\sum_{r=1}^R \sum_{a=1}^R \sum_{k\in I_{ar}} \langle \Delta_0, u_{kr}\Theta_r - u_{ka}^*\Theta_a^*\rangle = \sum_{r=1}^R \sum_{k\in I_r} u_{kr}\langle \Delta_0, \Theta_r\rangle - \sum_{a=1}^R \sum_{k\in I_a^*} u_{ka}^*\langle \Delta_0, \Theta_a^*\rangle = 0.$$

Then, we have

$$\begin{aligned} I_2 &= \frac{1}{4\tau^2} \sum_{r=1}^R \sum_{a=1}^R \sum_{k\in I_{ar}} \|\Delta_0\|_F^2 + \|u_{ka}^*\Delta_{ar} + (u_{kr} - u_{ka}^*)\Theta_r\|_F^2 \\ &= \frac{1}{4\tau^2}\left\{ K\|\Delta_0\|_F^2 + \sum_{r=1}^R \sum_{a=1}^R \sum_{k\in I_{ar}} \left[(u_{ka}^*)^2 \|\Delta_{ar}\|_F^2 + (u_{kr} - u_{ka}^*)^2 \|\Theta_r\|_F^2 + 2\langle u_{ka}^*\Delta_{ar}, (u_{kr} - u_{ka}^*)\Theta_r\rangle \right] \right\}, \end{aligned}$$

where the last term satisfies

$$\begin{aligned} 2\langle u_{ka}^*\Delta_{ar}, (u_{kr} - u_{ka}^*)\Theta_r\rangle &\ge -2|u_{ka}^*||(u_{kr} - u_{ka}^*)| \|\Delta_{ar}\|_F \|\Theta_r\|_F \\ &\ge -2|u_{ka}^*||(u_{kr} - u_{ka}^*)| \left[\|\Delta_{ar}\|_F^2 + \right] \end{aligned}$$

Now consider the set

$$\begin{aligned} \mathcal{A} = \Bigg\{ (U, \Theta_r) : \|\Delta_0\|_F \le M_0\sqrt{\frac{s_0 \log p}{nK}}, \|\Delta_{ar}\|_F \le M_{ar}\sqrt{\frac{(s_r+s_a)\log p}{n|I_{ar}|}}, \\ |u_{kr} - u_{ka}^*| \le M_k\sqrt{\frac{s_r \log p}{n}}, k \in I_{ar}, a, r \in [R] \Bigg\}, \end{aligned}$$

for some large constants M_0, M_{ar}, M_k. For $(U, \Theta_r) \in \partial\mathcal{A}$, we have

$$\begin{aligned} G(U, \Theta_r) = & \frac{M_0 s_0 \log p}{n}\left[\frac{M_0}{4\tau^2} - C_0\sqrt{2}\right] + \sum_{r=1}^R \sum_{a=1}^R \frac{M_{ar}(s_r+s_a)\log p}{n}\left[\frac{\sum_{k\in I_{ar}} (u_{ka}^*)^2 M_{ar}}{|I_{ar}|} - C_{ar}\right] \\ & + \sum_{r=1}^R \sum_{a=1}^R \sum_{k\in I_{ar}} \frac{M_k\sqrt{s_r}\log p}{n} \|\Theta_r\|_F \left[M_k\sqrt{s_r}\|\Theta_r\|_F - \frac{2M_{ar}|u_{ka}^*|\sqrt{s_r+s_a}}{\sqrt{|I_{ar}|}} - C_k \right]. \end{aligned}$$

Choosing proper M_0, M_{ar}, M_k, we have $G(U, \Theta_r) > 0$, which implies these is a local minimizer lies inside $\mathcal{A}$.

□

Lemma 2. *Let* $Z_i \sim \mathcal{N}_p(\mathbf{0}, \Sigma_i)$ *i.i.d. with* $\Sigma_i = [\![\Sigma_{i,jk}]\!]$ *for* $i \in [n]$ *and* $\max_{i\in[n]} \lambda_{\max}(\Sigma_i) \leq \epsilon_0 < \infty$. *Then, we have*

$$P\left(|\frac{1}{n}\sum_{i=1}^{n}(Z_{i,j}Z_{i,k} - \Sigma_{i,jk})| \geq t\right) \leq c_1 \exp\left(-c_2 n t^2\right), \quad \textit{for} \quad t \leq |b|,$$

where c_1, c_2, b *depend on* ϵ_0.

Proof. The result follows by the equation (2.20) in (**?**). □

References