41500: High-Dimensional Statistics	**Winter 2020**

Precision Matrix Inference

Lecturer: Chao Gao *Scribe: Xinze Li*

Suppose that $X \sim \mathcal{N}\left(0, \Omega^{-1}\right)$ and that $\Omega = \Sigma^{-1}$ is the precision matrix. And thus

$$X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}^{-1}\right) \tag{1}$$

And so we have the conditional probability distribution as follows

$$x_1 | x_2 \sim \mathcal{N}\left(-\Omega_{11}^{-1}\Omega_{12}x_2, \Omega_{11}^{-1}\right) \tag{2}$$

Suppose that $A = 1, 2$, and that $A^{\complement} = \{3, \cdots, p\}$. So

$$X = \begin{pmatrix} x_1 \\ | \\ x_p \end{pmatrix} = \begin{pmatrix} X_A \\ X_{A^{\complement}} \end{pmatrix} \tag{3}$$

Thus

$$X_A | X_{A^{\complement}} \sim \mathcal{N}\left(-\Omega_{AA}^{-1}\Omega_{AA^{\complement}}X_{A^{\complement}}, \Omega_{AA}^{-1}\right) \tag{4}$$

Let

$$B^{\top} = -\Omega_{AA}^{-1}\Omega_{AA^{\complement}}, \quad B \in \mathbb{R}^{(p-2)\times 2} \tag{5}$$

Now

$$\Omega_{AA} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \tag{6}$$

We suppose that $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Omega^{-1})$ and that

$$\max_j \sum_k \mathbb{1}\left(\Omega_{jk} \neq 0\right) \leq s \tag{7}$$

Also suppose

$$M^{-1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq M \tag{8}$$

and that $\widehat{B}$ is the lasso estimate

$$\widehat{B} = \arg\min_{B \in \mathbb{R}^{(p-2)\times 2}} \sum_i \left\| X_{iA} - B^{\top} X_{iA^{\complement}} \right\|^2 + \lambda \left\| B \right\|_1 \tag{9}$$

By HW6 P2, we know that w.h.p. the following holds

$$\begin{gathered} \frac{1}{n}\sum_{i=1}^{n} \left\| \left(\widehat{B} - B\right)^T X_{iA^{\complement}} \right\|^2 \lesssim \frac{s \log p}{n} \\ \left\| \widehat{B} - B \right\|_1 \lesssim s\sqrt{\frac{\log p}{n}} \end{gathered} \tag{10}$$

B is defined in 5. We now define the estimator

$$\widehat{\Omega}_{AA}^{-1} = \frac{1}{n}\sum_i \left(X_{iA} - \widehat{B}^\top X_{iA^\complement}\right)\cdot\left(X_{iA} - \widehat{B}^\top X_{iA^\complement}\right)^\top \tag{11}$$

Now note that from 4, we could write

$$X_{iA} = B^\top X_{iA^\complement} + \Omega_{AA}^{-\frac{1}{2}} w_i, \quad w_i \sim \mathcal{N}(0, I_2) \tag{12}$$

Combining eqs. (11) and (12), we have

$$\begin{aligned}
\widehat{\Omega}_{AA}^{-1} &= \frac{1}{n}\sum_i \left(B^\top X_{iA^\complement} + \Omega_{AA}^{-\frac{1}{2}} w_i - \widehat{B}^\top X_{iA^\complement}\right)\cdot\left(B^\top X_{iA^\complement} + \Omega_{AA}^{-\frac{1}{2}} w_i - \widehat{B}^\top X_{iA^\complement}\right)^\top \\
&= \underbrace{\frac{1}{n}\sum_i \Omega_{AA}^{-\frac{1}{2}} w_i w_i^\top \Omega_{AA}^{-\frac{1}{2}}}_{\text{(I)}} + \underbrace{\frac{1}{n}\sum_i \left(\left(B - \widehat{B}\right)^\top X_{iA^\complement}\right)\cdot\left(\left(B - \widehat{B}\right)^\top X_{iA^\complement}\right)^\top}_{\text{(II)}} \\
&\quad + \underbrace{\frac{1}{n}\sum_i \left(\Omega_{AA}^{-\frac{1}{2}} w_i\right)\cdot X_{iA^\complement}^\top \left(B - \widehat{B}\right)}_{\text{(III)}} + \underbrace{\frac{1}{n}\sum_i \left(B - \widehat{B}\right)^\top X_{iA^\complement}\cdot\left(\Omega_{AA}^{-\frac{1}{2}} w_i\right)^\top}_{\text{(IV)}}
\end{aligned} \tag{13}$$

Note that

$$\text{(I)} - \Omega_{AA}^{-1} = \frac{1}{n}\Omega_{AA}^{-\frac{1}{2}}\left(\sum_i w_i w_i^\top - I_2\right)\Omega_{AA}^{-\frac{1}{2}} \tag{14}$$

which is asymptotically normal. For (II), from 10, we have

$$\|\text{(II)}\|_{op} \leq \frac{1}{n}\sum_{i=1}^n \left\|\left(\widehat{B} - B\right)^T X_{iA^\complement}\right\|^2 \lesssim \frac{s\log p}{n} \tag{15}$$

The magnitude of (III) and (IV) is the same, so we simply analyze $\|\text{(III)}\|_F$

$$\begin{aligned}
\|\text{(III)}\|_F &= \left\|\left(B - \widehat{B}\right)^\top \cdot \left(\frac{1}{n}\sum_i X_{iA^\complement}\left(\Omega_{AA}^{-\frac{1}{2}} w_i\right)^\top\right)\right\|_F \\
&\leq \left\|B - \widehat{B}\right\|_1 \cdot \left\|\frac{1}{n}\sum_i X_{iA^\complement}\left(\Omega_{AA}^{-\frac{1}{2}} w_i\right)^\top\right\|_\infty \\
&\lesssim s\sqrt{\frac{\log p}{n}} \cdot \max_{j\in[2], k\in[p-2]} \frac{1}{n}\sum_i \left(\Omega_{AA}^{-\frac{1}{2}} w_i\right)_j \cdot (X_{iA^\complement})_k \\
&\lesssim s\sqrt{\frac{\log p}{n}} \cdot \sqrt{\frac{\log p}{n}} = s\frac{\log p}{n}
\end{aligned} \tag{16}$$

Thus

$$\sqrt{n}\left(\widehat{\Omega}_{AA}^{-1}-\Omega_{AA}^{-1}\right)=\frac{1}{\sqrt{n}}\Omega_{AA}^{-\frac{1}{2}}\left(\sum_{i} w_i w_i^{\top}-I_2\right)\Omega_{AA}^{-\frac{1}{2}}+O_p\left(\frac{s\log p}{\sqrt{n}}\right) \tag{17}$$

We conclude with the following theorem (See HW6 2(c) for details)

Theorem 0.1 ([Ren et al., 2015])**.** *If* $\frac{s\log p}{\sqrt{n}}\to 0$*, then*

$$\sqrt{n}\left(\widehat{\Omega}_{12}-\Omega_{12}\right)\rightsquigarrow\mathcal{N}\left(0,\Omega_{11}\Omega_{22}+\Omega_{12}^2\right) \tag{18}$$

From [Ren et al., 2015], define parameter space

$$\mathcal{G}_0\left(M,k_{n,p}\right)=\left\{\begin{array}{r}\Omega=\left(\omega_{ij}\right)_{1\le i,j\le p}:\max\limits_{1\le j\le p}\sum\limits_{i=1}^{p}1\left\{\omega_{ij}\ne 0\right\}\le k_{n,p}\\ \text{and } 1/M\le\lambda_{\min}(\Omega)\le\lambda_{\max}(\Omega)\le M\end{array}\right\} \tag{19}$$

we also know the minimax rate

$$\inf_{\hat{\omega}_{ij}}\sup_{\mathcal{G}_0(M,k_{n,p})}\mathbb{E}\left|\hat{\omega}_{ij}-\omega_{ij}\right|\asymp\max\left\{n^{-1}k_{n,p}\log p,n^{-1/2}\right\} \tag{20}$$

Thus we could observe that ther condition $\frac{s\log p}{\sqrt{n}}$ is necessary.

References

[Ren et al., 2015] Ren, Z., Sun, T., Zhang, C.-H., and Zhou, H. H. (2015). Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):9911026.