MATRIX ESTIMATION BASED ON SIMULTANEOUS CONFIDENCE INTERVALS

Cun-Hui Zhang

Department of Statistics and Biostatistics, Rutgers University

czhang@stat.rutgers.edu

BIRS, January 20, 2011

Thanks for the invitation!

We derive methods of learning population covariance and correlation matrices and their inverses based on simultaneous confidence intervals. The performance of these methods will be compared with the Lasso in selection consistency and estimation under the spectrum norm.

- Co-authors: Fei Ye and Tingni Sun

- Data: $\mathbf{x}_i$ iid from $N(0, \mathbf{\Sigma}) \in \mathbb{R}^p$, $i \leq n$
- Problem: estimation of the precision matrix $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$
- Difficulty: For $p > n$, $\widehat{\mathbf{\Sigma}} = \mathbf{X}'\mathbf{X}/n$ is singular
- Sparsity: the sparsity of $\mathbf{\Theta} = (\theta_{jk})_{p \times p}$;

$$S = \{(j,k) : j \neq k, \theta_{jk} \neq 0\} \text{ is not too large}$$

- Scale: $\|\mathbf{\Sigma}\|_{2,2} + \|\mathbf{\Theta}\|_{2,2} = O(1)$
- Maximum note degree: $d = \max_j \#\{k : \theta_{jk} \neq 0\} \ll n/\log p$
- Estimation error measure: spectrum norm $\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_{2,2}$
- Graphical model selection: $\widehat{S} = S$

- G-Lasso: With $D = \{(j,k) : j = k\}$,

$$\widehat{\mathbf{\Theta}} = \arg\min_{\mathbf{\Theta}} \left\{ \text{trace}(\mathbf{\Theta}\widehat{\mathbf{\Sigma}}) - \log\det(\mathbf{\Theta}) + \lambda \|\mathbf{\Theta}_{D^c}\|_1^{(vec)} \right\}$$

- ℓ_1 penalized MLE
- Yuan and Lin (2007), Friedman, Hastie and Tibshirani (2008, GLasso), Rothman-Bickel-Levina-Zhu (2008, SPICE), Ravikumar et al (2008a,b), ... Lam-Fan (2007)

Rothman *et al* (2008), Ravikumar et al (2008b):

- Let $\lambda = C\sqrt{(\log p)/n}$
- Error bound in the Frobenius norm

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_2^{(vec)} = O_P(1)\sqrt{(|S| + p)(\log p)/n}$$

- A modification gives an error bound in the ℓ_2 operator norm

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{2,2} = O_P(1)\sqrt{|S|(\log p)/n}$$

- The result for $\| \cdot \|_2^{(vec)}$ is of sharp order, but not $\| \cdot \|_{2,2}$
- $\lambda = C\sqrt{(\log p)/n}$; $C = ?$

Let $(X_1, \ldots, X_p) \sim N(0, \mathbf{\Sigma})$ and $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$.

- Partial correlation:

$$\mathrm{Corr}(X_j, X_k | X_\ell, \ell \neq j, \ell \neq k) = -\theta_{jk}/\{\theta_{jj}\theta_{kk}\}^{1/2}$$

- Linear model:

$$X_j | (X_k, k \neq j) \sim N\Big(\sum_{k \neq j} \beta_{jk} X_k \, , \sigma_j^2\Big),$$
$$\beta_{jk} = -\theta_{jk}/\theta_{jj}, \;\; \sigma_j^2 = 1/\theta_{jj}$$

- One may piece together estimates in the p linear regression models to produce an estimate of $\mathbf{\Theta}$
- Need to estimate both σ_j and $\boldsymbol{\beta}_{-j} = (\beta_{jk}, k \neq j)'$

- Städler, Bühlmann and van de Geer (2010): PMLE

$$\{\widehat{\boldsymbol{\beta}}_{-j}, \widehat{\sigma}_j\} = \arg\min_{\boldsymbol{\beta},\sigma} \Big\{ \frac{\|\mathbf{x}_j - \mathbf{X}_{-j}\boldsymbol{\beta}_{-j}\|^2}{2\sigma_j^2 n} - \log \sigma_j + \lambda_0 \frac{\|\boldsymbol{\beta}_{-j}\|_1}{\sigma_j} \Big\}.$$

This is a convex minimization problem in $\{\boldsymbol{\beta}_{-j}/\sigma_j, 1/\sigma_j\}$

- Since $\theta_{jk} = -\beta_{jk}/\sigma_j^2$ and $\theta_{jj} = 1/\sigma_j^2$,

$$(\widehat{\theta}_{jk}, k \neq j)' = -\widehat{\boldsymbol{\beta}}_{-j}/\widehat{\sigma}_j^2,\ \widehat{\theta}_{jj} = 1/\widehat{\sigma}_j^2$$

- $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{2,2} \leq \|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{\infty,\infty} = \|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{1,1} = \max_j \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_1$

Column-by-column p-linear regressions:

- Analysis: For $\lambda_0 = 2\sqrt{(\log p)/n}$,

$$\|\widehat{\boldsymbol{\beta}}_{-j}/\widehat{\sigma}_j - \boldsymbol{\beta}_{-j}/\sigma_j\|_1 \lesssim \|\boldsymbol{\beta}_{-j}\|_0\lambda_0,\ |\widehat{\sigma}_j/\sigma_j - 1| \asymp \lambda_0\|\boldsymbol{\beta}_{-j}/\sigma_j\|_1$$

- Since $\widehat{\theta}_{jk} = -\widehat{\beta}_{jk}/\widehat{\sigma}_j^2$ and $\widehat{\theta}_{jj} = 1/\widehat{\sigma}_j^2$,

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{2,2} \le \max_j \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_1 \lesssim \lambda_0(d + \|\boldsymbol{\Theta}\|_{\infty,\infty}^2)$$

- Is $d\lambda_0$ optimal under the assumption $\|\boldsymbol{\Theta}\|_{\infty,\infty} = O(1)$?
- Is $\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{2,2} \lesssim \sqrt{d(\log p)/n}$ possible?

• Alternative analysis: $\|\widehat{\boldsymbol{\beta}}_{-j} - \boldsymbol{\beta}_{-j}\|_q \le (1 + 1/\sqrt{2})\widehat{\sigma}_j \lambda_0 / CIF_q$,

$$CIF_q = \inf_{cone} \frac{\|\mathbf{X}_{-j}'\mathbf{X}_{-j}\mathbf{u}\|_\infty |S_{-j}|^{1/q}}{n\|\mathbf{u}\|_q}$$

with $S_{-j} = \{k : k \ne j, \theta_{jk} \ne 0\}$
• $\max_j(1/CIF_\infty) \approx \|\boldsymbol{\Theta}\|_{\infty,\infty} = O(1)$
• $\|\widehat{\boldsymbol{\beta}}_{-j} - \boldsymbol{\beta}_{-j}\|_\infty \lesssim \sqrt{(\log p)/n}$
• Compatibility/restricted eigenvalues: van der Geer (2007), Bickel et al (2009),...

Cai, Liu and Luo (2010, CLIME):

$$\min \|\boldsymbol{\theta}\|_1 \ \text{ subj to } \ \|\widehat{\boldsymbol{\Sigma}}\boldsymbol{\theta}_j - \mathbf{e}_j\|_\infty \le \lambda$$

If $\max_j \|(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{\theta}_j\|_\infty \le \lambda$, then

$$\max_j \|\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j\|_1 = O(\|\boldsymbol{\Theta}\|_{\infty,\infty})\lambda d$$

- $\lambda = C\sqrt{(\log p)/n}$; $C \propto \|\boldsymbol{\Theta}\|_{\infty,\infty}$

GMACS (graphical maximum absolute control selector)

- Given symmetric data matrix $\widehat{\mathbf{M}}$,

$$\min \|\mathbf{\Theta}\|_{\infty,\infty} \text{ subject to } \|\mathbf{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_\infty^{(vec)} \le \lambda(\mathbf{\Theta})$$

- Convex minimization for fixed λ; General $\lambda(\mathbf{\Theta})$?
- Simple uniform bound:

$$\widehat{\mathbf{\Theta}} - \mathbf{\Theta} = \widehat{\mathbf{\Theta}}(\mathbf{I} - \widehat{\mathbf{M}}\mathbf{\Theta}) + (\widehat{\mathbf{\Theta}}\widehat{\mathbf{M}} - \mathbf{I})\mathbf{\Theta}$$

In the event $\|\mathbf{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_\infty^{(vec)} \le \lambda(\mathbf{\Theta})$ (feasibility of $\mathbf{\Theta}$),

$$\begin{aligned} &\|\widehat{\mathbf{\Theta}}\|_{\infty,\infty} \le \|\mathbf{\Theta}\|_{\infty,\infty} \\ &\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_\infty^{(vec)} \le \lambda(\mathbf{\Theta})\|\widehat{\mathbf{\Theta}}\|_{\infty,\infty} + \lambda(\widehat{\mathbf{\Theta}})\|\mathbf{\Theta}\|_{\infty,\infty} \end{aligned}$$

- $\lambda(\widehat{\mathbf{\Theta}}) \le \lambda(\mathbf{\Theta})$?

- Let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$, $\widehat{\mathbf{M}} = \widehat{\boldsymbol{\Sigma}}$ and κ_α satisfy $\sqrt{2}\kappa_\alpha + \log(1 + \sqrt{2}\kappa_\alpha) = (n/2)\log(p(p+1)/\alpha)$ and $\lambda_\alpha(\boldsymbol{\Theta}) = \kappa_\alpha\sqrt{2 + 2\max_i \Sigma_{ii} \max_j \theta_{jj}}$. Then,

$$\|\boldsymbol{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_\infty^{(vec)} \leq \lambda_\alpha(\boldsymbol{\Theta})$$

with at least probability $1 - \alpha$. In the same event

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_\infty^{(vec)} \leq \lambda(\boldsymbol{\Theta})\|\widehat{\boldsymbol{\Theta}}\|_{\infty,\infty} + \lambda(\widehat{\boldsymbol{\Theta}})\|\boldsymbol{\Theta}\|_{\infty,\infty} = \tau^*$$

- If we threshold $\widehat{\boldsymbol{\Theta}}$ are level τ^*, then for "most" norms $\|\cdot\|$

$$\|\widehat{\widehat{\boldsymbol{\Theta}}} - \boldsymbol{\Theta}\| \leq 2\tau^*\|\mathbf{1}_{D \cup S}\|$$

- $\tau^* \leq C\sqrt{(\log p)/n}$? $C = ?$

- If $\tau^* \leq C\sqrt{(\log p)/n}$ with a known C, then

$$\|\widehat{\widehat{\mathbf{\Theta}}} - \mathbf{\Theta}\|_{2,2} \lesssim \|\mathbf{1}_{D\cup S}\|_{2,2}\sqrt{(\log p)/n}$$

- This will give a better rate

$$\sqrt{d} \leq \|\mathbf{1}_{D\cup S}\|_{2,2} = \max_{\|\mathbf{u}\|=1} \sum_{(j,k)\in S\cup D} u_j u_k \leq d$$

• Let h_λ be the hard threshold function. If λ is sufficiently large,

$$\|h_\lambda(\widehat{\mathbf{\Sigma}}) - \mathbf{\Sigma}\|_{2,2} \leq \lambda \max_{\|\mathbf{u}\|=1} \sum_{\sigma_{jk}\neq 0} u_j u_k$$

where σ_{jk} are elements of $\mathbf{\Sigma}$

• Bickel-Levina (2008b); number of closed walks of length m on the graph, El Karoui (2008), ...

GMACS: $\lambda(\mathbf{\Theta}) = \xi \|\mathbf{\Theta}\|_{\infty,\infty}$

- Given symmetric data matrix $\widehat{\mathbf{M}}$,

$$\min \|\mathbf{\Theta}\|_{\infty,\infty} \text{ subject to } \|\mathbf{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_{\infty}^{(vec)} \leq \lambda = \xi \|\mathbf{\Theta}\|_{\infty,\infty}$$

- Linear program:

$$\min \left\{ t : \|\mathbf{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_{\infty}^{(vec)} / \xi = t, \|\mathbf{\Theta}\|_{\infty,\infty} = t \right\}$$

- Simple uniform bound: In the event $\|\widehat{\mathbf{M}} - \mathbf{M}\|_{\infty}^{(vec)} \leq \xi$,

$$\|\mathbf{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_{\infty}^{(vec)} = \|\mathbf{\Theta}(\widehat{\mathbf{M}} - \mathbf{M})\|_{\infty}^{(vec)} \leq \xi \|\mathbf{\Theta}\|_{\infty,\infty}$$
$$\|\widehat{\mathbf{\Theta}}\|_{\infty,\infty} \leq \|\mathbf{\Theta}\|_{\infty,\infty}$$
$$\|\widehat{\mathbf{\Theta}} - \mathbf{\Theta}\|_{\infty}^{(vec)} \leq \xi \|\widehat{\mathbf{\Theta}}\|_{\infty,\infty} \|\mathbf{\Theta}\|_{\infty,\infty}$$

- $\|\widehat{\mathbf{M}} - \mathbf{M}\|_{\infty}^{(vec)} \leq \xi$; $\xi = ?$

SCI for covarainces σ_{jk}

- Let $\widehat{\sigma}_{jk}$ be elements of $\widehat{\boldsymbol{\Sigma}}$
- Let κ_α satisfy (Bonferroni adjustment)

$$2\binom{p}{2}\exp\Big(-\tfrac{n}{2}\big\{\sqrt{2}\kappa_\alpha - \log(1+\sqrt{2}\kappa_\alpha)\big\}\Big) \\ +p\,P\Big\{\Big|\chi^2_n/n-1\Big| > \kappa_\alpha\Big\} \le \alpha,$$

- With $\widehat{V}_{ij} = (1-\kappa_\alpha)^{-2}\min\big\{2\widehat{\sigma}_{jj}\widehat{\sigma}_{kk}, \widehat{\sigma}_{jj}\widehat{\sigma}_{kk}/(1-\kappa_\alpha)^2 + \widehat{\sigma}^2_{jk}\big\}$,

$$P\big\{|\widehat{\sigma}_{jk}-\sigma_{jk}| \le \kappa_\alpha \widehat{V}^{1/2}_{jk}\,\forall jk\big\} \ge 1-\alpha$$

- $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_\infty^{(vec)} \le \xi = \kappa_\alpha \max_{jk} V_{jk}^{1/2} \asymp \sqrt{n^{-1}\log(p/\alpha)}$, w.p. $1-\alpha$

SCI for correlations $\rho_{jk} = \sigma_{jk}/\sqrt{\sigma_{jj}\sigma_{kk}}$ and

- Let $\widehat{\rho}_{jk} = \widehat{\sigma}_{jk}/\sqrt{\widehat{\sigma}_{jj}\widehat{\sigma}_{kk}}$ and κ_α satisfy

$$p(p-1)P\Big\{|N(0,1/n)| > \kappa_\alpha\Big\} + p\,P\Big\{\chi_n^2/n \le 1-\kappa_\alpha\Big\} \le \alpha$$

- SCI:

$$P\Big\{|\widehat{\rho}_{ij} - \rho_{ij}| \le \kappa_\alpha \, \forall ij\Big\} \ge 1-\alpha$$

- $\|\widehat{\boldsymbol{\rho}} - \boldsymbol{\rho}\|_\infty^{(vec)} \le \xi = \kappa_\alpha \approx \sqrt{(4/n)\log(p/\alpha)}$ w.p. $1-\alpha$

- GMACS based on SCI

$$\min \|\boldsymbol{\Theta}\|_{\infty,\infty} \quad \text{subject to} \quad \|\boldsymbol{\Theta}\widehat{\mathbf{M}} - \mathbf{I}\|_{\infty}^{(vec)} \leq \xi \|\boldsymbol{\Theta}\|_{\infty,\infty}$$

- For $\widehat{\mathbf{M}} = \widehat{\boldsymbol{\Sigma}}$ or $\widehat{\mathbf{M}} = \widehat{\boldsymbol{\rho}}$,

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{\infty}^{(vec)} \leq 2\xi \|\widehat{\boldsymbol{\Theta}}\|_{\infty,\infty} \|\boldsymbol{\Theta}\|_{\infty,\infty} = 2\widehat{\lambda}_{\alpha} \|\boldsymbol{\Theta}\|_{\infty,\infty}$$
$$\|\widehat{\boldsymbol{\Theta}}\|_{\infty,\infty} \leq \|\boldsymbol{\Theta}\|_{\infty,\infty}$$
$$\widehat{\lambda}_{\alpha} \asymp \sqrt{n^{-1} \log(p/\alpha)}$$

- Since $\|\boldsymbol{\Theta}\|_{\infty,\infty} \leq d \|\boldsymbol{\Theta}\|_{\infty}^{(vec)}$,

$$\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{\infty}^{(vec)} \leq 2\widehat{\lambda}_{\alpha} \|\widehat{\boldsymbol{\Theta}}\|_{\infty,\infty} / (1 - 2\widehat{\lambda}_{\alpha} d)$$

- Thresholding at level $4\widehat{\lambda}_\alpha \|\widehat{\mathbf{\Theta}}\|_{\infty,\infty}$:

$$\widehat{\widehat{\mathbf{\Theta}}} = \Big(\widehat{\theta}_{jk} I\{|\widehat{\theta}_{jk}| \geq 4\lambda_\alpha \|\widehat{\mathbf{\Theta}}\|_{\infty,\infty}\}\Big)_{p\times p}$$

- Suppose $\|\Sigma\|_{2,2} + \|\mathbf{\Theta}\|_{\infty,\infty} = O(1)$ and $d\sqrt{\log(p/\alpha)}/\sqrt{n} \leq a_0$ for a certain fixed $a_0 \in (0,1)$. Then,

$$\|\widehat{\widehat{\mathbf{\Theta}}} - \mathbf{\Theta}\|_{2,2} \lesssim \|\mathbf{1}_{D\cup S}\|_{2,2}\sqrt{\log(p/\alpha)}/\sqrt{n}$$

- Removing the condition $\|\mathbf{\Theta}\|_{\infty,\infty} = \max_j \sum_k |\theta_{jk}| = O(1)$
- Removing the condition $d\sqrt{\log(p/\alpha)}/\sqrt{n} \leq a_0$

- Multiple testing with FWER $\leq \alpha$:

$$\text{Rej } H_{0k} \text{ iff } \inf_{\boldsymbol{\Theta} \in H_{0k}} \frac{\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\|_{\infty}^{(vec)}}{2\widehat{\lambda}_{\alpha}\|\boldsymbol{\Theta}\|_{\infty,\infty}} > 1$$

- Testing a single model:

$$\text{Rej model } S \text{ iff}$$

$$\max_{(i,j) \notin S \cup D} |\widehat{\theta}_{ij}| > \max_{i} \sum_{j} \frac{2\widehat{\lambda}_{\alpha}|\widehat{\theta}_{ij}| I_{\{(i,j) \in S \cup D\}}}{1 - 2\widehat{\lambda}_{\alpha}(|S_i| + 1)}$$

where $S_j = \#\{i : (i,j) \in S\}$

- Candidate models:

$$\widehat{S}_\tau = \left\{(i,j) \notin D : |\widehat{\theta}_{ij}| > 2\widehat{\lambda}_\alpha \tau\right\}$$

- Statistical selection:

$$\widehat{S}_{\widehat{\tau}},\ \widehat{\tau} = \max\left\{\tau : \widehat{S}_\tau \text{ not rejected }\right\}$$

- All proper submodel of $\widehat{S}_{\widehat{\tau}}$ are rejected with FWER $\leq \alpha$
- If $\min_{(i,j)\in S} |\theta_{ij}| \geq 4\lambda^*_\alpha \|\mathbf{\Theta}\|_{\infty,\infty}/(1 - 2\lambda^*_\alpha d)$, then

$$P\{\widehat{S}_{\widehat{\tau}} = S\} \geq 1 - \alpha - P\{\widehat{\lambda}_\alpha > \lambda^*_\alpha\}$$

- ℓ_∞ incoherence condition is not needed
- Ravikumar et al (2008a) on G-Lasso

THANKS!