Graphical models: dependence structure for a set of variables

Gaussian graphical models

$X = (X_1, \dots, X_p) \sim N_p(\mu, \Sigma)$ positive definite, undirected graph $G = (V, E)$, $V = \{1, \dots, p\}$ is the vertex set & E is the edge set:

$$(i,j) \notin E \text{ iff } X_i \perp\!\!\!\perp X_j \mid X_{\{1,\dots,p\}\setminus\{i,j\}}$$ — conditional independence

Precision/concentration/inverse covariance matrix $\Theta = \Sigma^{-1}$

Prop. $X_i \perp\!\!\!\perp X_j \mid X_{\{1,\dots,p\}\setminus\{i,j\}}$ iff $\theta_{ij} = 0$.

Pf. By properties of the multivariate normal, the conditional distribution of $X_{(1)} = (X_i, X_j)$ given $X_{(2)} = X_{\{1,\dots,p\}\setminus\{i,j\}}$ is $N_2(\mu_{1|2}, \Sigma_{1|2})$, where

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Partition Σ as
$$\begin{array}{cc} (i,j) & \setminus\{i,j\} \\ \end{array}$$
$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Thus, $X_i \perp\!\!\!\perp X_j \mid X_{\{1,\dots,p\}\setminus\{i,j\}}$ iff $\sigma_{1|2,ij} = 0$.

On the other hand, by partitioning $\Theta\Sigma = I$,

$$\Theta_{11}\Sigma_{11} + \Theta_{12}\Sigma_{21} = I,$$
$$\Theta_{11}\Sigma_{12} + \Theta_{12}\Sigma_{22} = 0.$$

Then

$$\Theta_{11}(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) = \Theta_{11}\Sigma_{11} - \Theta_{11}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$
$$= I - \Theta_{12}\Sigma_{21} + \Theta_{12}\Sigma_{21} = I,$$

So that

$$\Sigma_{1|2} = \Theta_{11}^{-1} = \frac{1}{\det(\Theta_{11})}\begin{pmatrix} \theta_{jj} & -\theta_{ij} \\ -\theta_{ji} & \theta_{ii} \end{pmatrix}.$$

This implies that

$$\sigma_{1|2,ij} = 0 \text{ iff } \theta_{ij} = 0. \quad \boxtimes$$

Precision matrix estimation

Precision matrix estimation

Method 1. Neighborhood-based, Meinshausen & Bühlmann (2006, AOS)

From $X \sim N_p(\mu, \Theta^{-1})$ we have $X_A | X_{A^c} \sim N\left(-\Theta_{AA}^{-1}\Theta_{AA^c}X_{A^c}, \Theta_{AA}^{-1}\right)$,

suggesting the linear model

$$X_A = B_A^T X_{A^c} + \eta_A, \quad \text{where } B_A = -\Theta_{AA^c}\Theta_{AA}^{-1}.$$

When $A = \{i\}$, this reduces to

$$X_i = \beta_i^T X_{\setminus\{i\}} + \eta_i, \quad \text{where } \beta_{ij} = -\frac{\theta_{ij}}{\theta_{ii}}.$$

Thus, $\mathrm{supp}(B) = \mathrm{supp}(\Theta)$. "Nodewise regression"

Pros. use techniques for linear regression, fairly stable

Cons. not trivial to estimate the magnitude of θ_{ij}, not symmetric or positive definite

Method 2. Penalized likelihood (Graphical Lasso), Yuan & Lin (2007, Biometrika)

The Gaussian likelihood

$$\ell(\mu, \Theta) = \frac{n}{2}\log\det(\Theta) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^T\Theta(X_i-\mu).$$

Substituting the MLE $\bar{X}$ for μ,

$$\mathcal{L}(\Theta) = \frac{n}{2}\log\det(\Theta) - \frac{n}{2}\mathrm{Tr}(\Theta\hat{\Sigma}).$$

sample covariance matrix $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})^T$

Assume Θ is sparse,

$$\underset{\Theta \succ 0}{\text{minimize}} \; -\log\det(\Theta) + \mathrm{tr}(\Theta\hat{\Sigma}) + \lambda\|\Theta\|_1$$

entrywise L_1-norm

Solved by e.g. ADMM algorithms

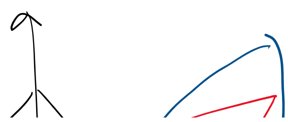
Method 3. CLIME (Constrained L_1-minimization), Cai, Liu & Luo (2011, JASA)

"Dantzig-selector version" of Graphical Lasso:

$$\text{minimize } \|\Theta\|_1$$
$$\text{subject to } \|\Theta^{-1} - \hat{\Sigma}\|_\infty \le \lambda \quad \xrightarrow{\times\Theta} \quad \|\hat{\Sigma}\Theta - I\|_\infty \le \lambda$$

too complex in θ_{ij}

entrywise L_∞-norm

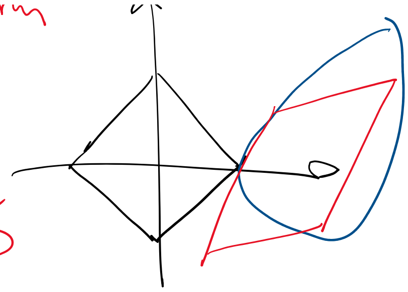


too complex in θ_{ij} ... ℓ_∞ norm

Equivalent to p linear programming problems:

$$\text{minimize } \|\theta_i\|_1$$
$$\text{subject to } \|\hat{\Sigma}\theta_i - e_i\|_\infty \le \lambda$$

solution denoted by $\tilde{\Omega}$



Symmetrization:

$$\hat{\Omega} = (\hat{\theta}_{ij}) \text{ w/ } \hat{\theta}_{ij} = \hat{\theta}_{ji} = \underbrace{\tilde{\theta}_{ij} I(|\tilde{\theta}_{ij}| \le |\tilde{\theta}_{ji}|) + \tilde{\theta}_{ji} I(|\tilde{\theta}_{ij}| > |\tilde{\theta}_{ji}|)}_{\text{take the smaller}}$$

Nonasymptotic error bounds for CLIME

Sparsity class $\mathcal{U}_q(M, s_0(p)) = \left\{\Omega : \Omega \succ 0, \|\Omega\|_{L_1} \le M, \max_i \sum_{j=1}^p |\theta_{ij}|^q \le s_0(p)\right\}$ for $0 \le q < 1$.

matrix L_1-norm

Lemma. Assume $\Omega_0 \in \mathcal{U}_q(M, s_0(p))$. If $\lambda \ge \|\Omega_0\|_{L_1} \|\hat{\Sigma} - \Sigma_0\|_\infty$, then

true precision matrix ... true

$$\|\hat{\Omega} - \Omega_0\|_\infty \le 4\|\Omega_0\|_{L_1}\lambda,$$
$$\|\hat{\Omega} - \Omega_0\|_{L_1} \le C s_0(p) \lambda^{1-q}.$$

Pf. Since
$$\|\hat{\Sigma}\Omega_0 - I\|_\infty \le \|(\hat{\Sigma} - \Sigma_0)\Omega_0\|_\infty \le \|\Omega_0\|_{L_1} \|\hat{\Sigma} - \Sigma_0\|_\infty \le \lambda,$$
by assumption

Ω_0 is a feasible solution. By the optimality of $\tilde{\Omega}$,
$$\|\tilde{\Omega}\|_{L_1} \le \|\Omega_0\|_{L_1}. \qquad (*)$$

Write
$$\|\tilde{\Omega} - \Omega_0\|_\infty = \|\Omega_0 \Sigma_0 (\tilde{\Omega} - \Omega_0)\|_\infty \le \|\Omega_0\|_{L_1} \|\Sigma_0(\tilde{\Omega} - \Omega_0)\|_\infty$$
$$\le \|\Omega_0\|_{L_1} \{\underbrace{\|\hat{\Sigma}(\tilde{\Omega} - \Omega_0)\|_\infty}_{T_1} + \underbrace{\|(\hat{\Sigma} - \Sigma_0)(\tilde{\Omega} - \Omega_0)\|_\infty}_{T_2}\}.$$

Note that
$$T_1 \le \|\hat{\Sigma}\tilde{\Omega} - I\|_\infty + \|\hat{\Sigma}\Omega_0 - I\|_\infty \le 2\lambda,$$
both feasible

$$T_2 \le \|\tilde{\Omega} - \Omega_0\|_{L_1} \|\hat{\Sigma} - \Sigma_0\|_\infty \le 2\|\Omega_0\|_{L_1} \|\hat{\Sigma} - \Sigma_0\|_\infty \le 2\lambda.$$

$$T_2 \le \|\tilde{\Omega}-\Omega_0\|_{L_1}\|\hat{\Sigma}-\Sigma_0\|_\infty \underset{\text{by }(*)}{\le} 2\|\Omega_0\|_{L_1}\|\hat{\Sigma}-\Sigma_0\|_\infty \underset{\text{by assumption}}{\le} 2\lambda.$$

Thus,

$$\|\tilde{\Omega}-\Omega_0\|_\infty \le 4\|\Omega_0\|_{L_1}\lambda,$$ which $\underset{\text{since } \hat{\theta}_{ij}=\tilde{\theta}_{ij} \text{ or } \tilde{\theta}_{ji}}{\text{implies}}$ $\|\hat{\Omega}-\Omega_0\|_\infty \le$ ——.

Let $t_n = \|\hat{\Omega}-\Omega_0\|_\infty$, $\delta_i = \hat{\theta}_i - \theta_i^0 \triangleq \delta_i^{(1)}+\delta_i^{(2)}$, where

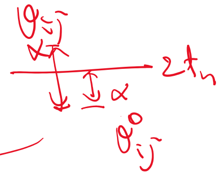
$$\delta_{ij}^{(1)} = \hat{\theta}_{ij} I(|\hat{\theta}_{ij}| \ge 2t_n) - \theta_{ij}^0, \quad \delta_{ij}^{(2)} = \hat{\theta}_{ij} I(|\hat{\theta}_{ij}| < 2t_n).$$

Then

$$\cancel{\|\theta_i^0\|_1} \ge \|\hat{\theta}_i\|_1 = \|\theta_i^0+\delta_i^{(1)}\|_1 + \|\delta_i^{(2)}\|_1 \ge \cancel{\|\theta_i^0\|_1} - \|\delta_i^{(1)}\|_1 + \|\delta_i^{(2)}\|_1,$$

so that $\|\delta_i^{(2)}\|_1 \le \|\delta_i^{(1)}\|_1$ & $\|\delta_i\|_1 \le 2\|\delta_i^{(1)}\|_1$. By the sparsity assumption,

$$\begin{aligned}
\|\delta_i^{(1)}\|_1 &= \sum_{j=1}^p \left|\hat{\theta}_{ij} I(|\hat{\theta}_{ij}| \ge 2t_n) - \theta_{ij}^0\right| \\
&\le \sum_{j=1}^p |\theta_{ij}^0| I(|\theta_{ij}^0| < 2t_n) + \sum_{j=1}^p \left|\hat{\theta}_{ij} I(|\hat{\theta}_{ij}| \ge 2t_n) - \theta_{ij}^0 I(|\theta_{ij}^0| \ge 2t_n)\right| \\
&\le (2t_n)^{1-q} \underbrace{\sum_{j=1}^p |\theta_{ij}^0|^q}_{\le s_0(p)} + \sum_{j=1}^p \underbrace{|\hat{\theta}_{ij}-\theta_{ij}^0|}_{\le t_n} I(|\hat{\theta}_{ij}| \ge 2t_n) \\
&\quad + \sum_{j=1}^p |\theta_{ij}^0| \underbrace{\left|I(|\hat{\theta}_{ij}| \ge 2t_n) - I(|\theta_{ij}^0| \ge 2t_n)\right|}_{=1 \text{ iff}} \\
&\le (2t_n)^{1-q} s_0(p) + t_n \sum_{j=1}^p I(|\theta_{ij}^0| \ge t_n) \\
&\quad + \sum_{j=1}^p |\theta_{ij}^0| I(|\theta_{ij}^0 - 2t_n| \le |\hat{\theta}_{ij}-\theta_{ij}^0|) \\
&\le (2t_n)^{1-q} s_0(p) + t_n^{1-q} \underbrace{\sum_{j=1}^p |\theta_{ij}^0|^q}_{\le s_0(p)} + \sum_{j=1}^p |\theta_{ij}^0| I(|\theta_{ij}^0| \le 3t_n) \\
&\le (2t_n)^{1-q} s_0(p) + t_n^{1-q} s_0(p) + (3t_n)^{1-q} s_0(p) \\
&= (1+2^{1-q}+3^{1-q})\, t_n^{1-q} s_0(p).
\end{aligned}$$



Combining the above to conclude

$= (1 + 2$ [illegible] $s_0(p)$.

Combining the above to conclude

$$\|\hat{\Omega} - \Omega_0\|_{L_1} = \max_i \|\delta_i\|_1 \le 2(1 + 2^{1-q} + 3^{1-q})(4\underbrace{\|\Omega_0\|_{L_1}}_{\le M}\lambda)^{1-q} s_0(p)$$

$$= C\, s_0(p)\, \lambda^{1-q}. \quad \#$$

Thm. $\|\hat{\Omega} - \Omega_0\|_2 \le C_1 M^{1-q} s_0(p) \left(\frac{\log p}{n}\right)^{(1-q)/2}$ w.h.p.

Pf. We can show that $\|\hat{\Sigma} - \Sigma_0\|_\infty \le C_2 \sqrt{\frac{\log p}{n}}$ w.h.p. Take $\lambda = C_2 M \sqrt{\frac{\log p}{n}}$.

Use the fact that $\|\hat{\Omega} - \Omega_0\|_2 \le \|\hat{\Omega} - \Omega_0\|_{L_1}$. #