High Dimensional Non Paranormal Graphical Models

Reviewer: Abhirup Mallik

Gaussian Graphical Model

- $X = (X_1, \cdots, X_p)$ be a random vector from $N_p(\mu, \Sigma)$.
- $G := (V, E)$ is encoded by $\Theta = \Sigma^{-1}$.
- $X_i \perp\!\!\!\perp X_j \big| X_{\setminus\{i,j\}}$ holds iff, $\Theta_{ij} = 0$.
- $\Sigma_{ij} = 0$ implies marginal independence. $\implies X_i \perp\!\!\!\perp X_j$
- $S = \frac{1}{n}\sum_{i=1}^{n}(X^{(i)} - \bar{X})(X^{(i)} - \bar{X})^T$ is unbiased but very noisy estimate of Σ.

What is Non Paranormal?

- $X = (X_1, \cdots, X_p)$. $X \sim NPN(\mu, \Sigma, f)$ if $\{f_j\}_{j=1}^p$, such that, $Z \equiv f(X) \sim N(\mu, \Sigma)$.
- $f(X) = (f_1(X_1), \cdots, f_p(X_p))$, f_i are monotone and differentiable.
- $NPN(\mu, \Sigma, f)$ is a Gaussian copula if f_i's are monotone and differentiable.
- For identifiability, $\mu_j = E(X_j) = E(Z_j)$, $\sigma_j^2 \equiv \Sigma_{jj} = var(X_j) = var(Z_j)$.
- $Z_i \perp\!\!\!\perp Z_j \big| Z_{\setminus\{i,j\}} \iff X_i \perp\!\!\!\perp X_j \big| X_{\setminus\{i,j\}}$ holds iff, $\Omega_{ij} = 0$.

More about Non Paranormal

- PDF=$p_x(X) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(f(x)-\mu)^T\Sigma^{-1}(f(x)-\mu))\Pi_{j=1}^{p}|f_j'(x_j)|$
- jth Marginal = $F_j(x) = P(X \leq x) = P(Z_j \leq f_j(x)) = \Phi(\frac{f_j(x)-\mu_j}{\sigma_j})$
- inverse transform: $f_j(x) = \mu_j + \sigma_j \Phi^{-1}(F_j(x))$, where, $h_j(x) = \Phi^{-1}(F_j(x))$. Or, $Z = \Phi^{-1}(F(X))$
- $\Sigma_{jk} = cov(f_j(X_j), f_k(X_k)) = \sigma_j \sigma_k cov(h_j(X_j), h_k(X_k))$
- $\Sigma = D\Lambda D$, so, $\Sigma^{-1} = D^{-1}\Lambda^{-1}D^{-1}$

Free lunch!!

- "Extra modeling flexibility and robustness comes at almost no cost in terms of statistical efficiency. So, rank based methods can replace the Gaussian estimators even when data is true Gaussian".

Rank Correlations

- $\hat{\rho}_{jk} = \frac{\sum_{i=1}^{n}(r_j^i - \bar{r}_j)(r_k^i - \bar{r}_k)}{\sum_{i=1}^{n}(r_j^i - \bar{r}_j)^2 \sum_{i=1}^{n}(r_k^i - \bar{r}_k)^2}$
- $\hat{\tau}_{jk} = \frac{2}{n(n-1)} \sum_{1 \leq i \leq i' \leq n} sign((x_j^i - x_j^{i'})(x_k^i - x_k^{i'}))$
- $\hat{r}_{ij}^s = 2 \sin \left(\frac{\pi}{6} \hat{\rho}_{ij} \right)$
- $\hat{r}_{ij}^t = 2 \sin \left(\frac{\pi}{2} \hat{\tau}_{ij} \right)$
- $\hat{R}^s = (\hat{r}_{ij}^s)_{1 \leq i,j \leq p}$
- $\hat{R}^t = (\hat{r}_{ij}^t)_{1 \leq i,j \leq p}$

Methods

- Lasso: $\hat{\beta} = argmin_{\beta \in \mathbb{R}^p}\{\frac{1}{n}||Y - X\beta||_2^2 + \lambda||\beta||_1\}$
- Dantzig Selector: $\hat{\beta} = argmin||\beta||_1$, s.t. $||X^T(Y - X\beta)||_\infty \leq \lambda$
- Graphical Lasso: $argmin_{\Omega \in \mathbb{S}_{++}} - \log|\Omega| + tr(\Omega S) + \lambda \sum_{j \neq k} |\Omega_{jk}|$
- CLIME: $argmin_\Omega ||\Omega||_1$, s.t. $||S\Omega - I||_{\max} \leq \lambda$
- $\min ||\beta||_1$ s.t. $||S\beta - e_i||_\infty \leq \lambda_n$
- Can be relaxed into linear programming problem.

Methods

consider $\mu = 0$

- 'Neighborhood' Lasso:
  $Z_k|Z_{(k)} \sim N(Z_{(k)}^T \Sigma_{(k)}^{-1} \sigma_{(k)}, 1 - \sigma_{(k)}^T \Sigma_{(k)}^{-1} \sigma_{(k)})$
- $Z_k = Z_{(k)}^T \beta_k + \epsilon_k$, $\theta_{(k)} = -\beta_k / var(\epsilon_k)$
- $\min \beta^T \hat{\Sigma}_{(k)}^o \beta - 2\beta^T \hat{\sigma}_{(k)}^o + \lambda ||\beta||_{l1}$
- Graphical Dantzig Selector:
  $min||\beta||_1$ s.t. $||\hat{\Sigma}_{(k)}^o \beta - \hat{\sigma}_{(k)}^o||_\infty \leq \lambda$

Rank based sample correlation matrix

For $\epsilon \in (0,1)$, $n \geq 12\pi/\epsilon$, there exist some ansolute constant c_0 with:

$$P(|\hat{r}^s_{ij} - \sigma_{ij}| > \epsilon) \leq 2exp(-c_0 n\epsilon^2)$$
$$P(||\hat{R}^s - \Sigma||_{max} > \epsilon) \leq p^2 exp(-c_0 n\epsilon^2)$$
$$||\hat{R}^s - \Sigma||_{max} = O(\sqrt{\log p/n}) \text{ wp } 1 - 1/p^2$$
$$||\hat{R}^\tau - \Sigma||_{max} = O(\sqrt{\log p/n}) \text{ wp } 1 - 1/p$$

Proof

$$\hat{r}^s_{ij} = \frac{n-2}{n+1} u_{ij} + \frac{3}{n+1} d_{ij}$$

$$d_{ij} = \frac{1}{n(n-1)} \sum_{k \neq l} sign(x_{ki} - x_{li}).sign(x_{kj} - x_{li})$$

$$u_{ij} = \frac{3}{n(n-1)(n-2)} \sum_{k \neq l, k \neq m, l \neq m} sign(x_{ki} - x_{li}).sign(x_{kj} - x_{mj})$$

$$E(u_{ij}) = \frac{6}{\pi} sin^{-1}(\sigma_{ij}/2), \text{ by def, } r^s_{ij} = 2\sin(\pi/6\hat{r}_{ij})$$

As $2\sin(\frac{\pi}{6}x)$ is a lipschitz function with lipschitz constant $\pi/3$,

$$
\begin{aligned}
P(|\hat{r}_{ij}^{s} - \sigma_{ij}| > \epsilon) &\leq P(|\hat{r}_{ij} - E(u_{ij})| > \frac{3\epsilon}{\pi}) \\
&\leq P(|u_{ij} - E(u_{ij})| > 3\epsilon/2\pi) \text{ under assumptions} \\
&\leq exp(-c_0 n\epsilon^2) \text{ McDiarmid's}
\end{aligned}
$$

$$
\begin{aligned}
P(|\hat{r}_{ij}^{\tau} - \sigma_{ij}| > t) &= P(|\sin(\pi/2\hat{\tau}_{ij}) - \sin(\pi/2\tau_{ij})| > t) \\
&\leq P(|\hat{\tau}_{ij} - \tau_{ij}| > \frac{2}{\pi}t) \\
&\leq p^2 exp(\frac{-nt^2}{2\pi^2}) \text{ (Hoeffding)}
\end{aligned}
$$

Result:

- Rank based estimator $\hat{R}^s$ and $\hat{R}^\tau$ can be plugged in to parametric graphical lasso, or graphical Dantzig selector, or CLIME.
- Rank based methods achieve the same rate of convergence for both precision matrix estimation and graph recovery under the non paranormal model.

Rank based Graphical Lasso

- $\hat{\Theta}_g^s = \arg\min_{\Theta \in \mathbb{S}_{++}} - \log(det(\Theta)) + tr(\hat{R}^s\Theta) + \lambda \sum_{i \neq j} |\theta_{ij}|$
- High probability result.
- $||\hat{\Theta}_g^s - \Theta^*||_{\max} \leq 2K_{H^*}(1 + k/4)\lambda$
- Sign consistency.
- $||\hat{\Theta}_g^s - \Theta^*||_{\max} = O_P\left(\sqrt{\frac{\log p}{n}}\right)$
- $||\hat{\Theta}_g^s - \Theta^*||_{l1} = O_P\left(\sqrt{\frac{\min\{s+p,d^2\} \log p}{n}}\right)$
- for $n >> d^2$, for $\lambda = O(log(p/n))^{1/2}$, then sign consistent.

Rank based Neighborhood Dantzig Selector

- $\hat{\beta}_k^{s.nd} = \arg\min_{\beta \in \mathbb{R}^{p-1}} ||\beta||_{l1}$ s.t. $||\hat{R}^s_{(k)}\beta - \hat{r}^s_{(k)}||_{l\infty} \leq \lambda$
- $\hat{\theta}_{kk}^{s.nd} = ((\hat{\beta}_k^{s.nd})^T \hat{R}^s_{(k)} \beta_k^{s.nd} - 2(\hat{\beta}_k^{s.nd})^T \hat{r}^{s.nd}_{(k)} + 1)^{-1}$
- $\hat{\theta}_{(k)}^{s.nd} = -\hat{\theta}_{kk}^{s.nd} \hat{\beta}_k^{s.nd}$
- $\hat{\Theta}^s_{nd} = (\hat{\theta}^{s.nd}_{(k)})_{p \times p}$ symmetrize it:
  $\breve{\Theta}^s_{nd} = \arg\min_{\Theta} ||\Theta - \hat{\Theta}^s_{nd}||_1$
- $||\breve{\Theta}^s_{nd} - \Theta^*||_{l1} = O_P\left(d\sqrt{\frac{\log p}{n}}\right)$
- Extension to adaptive lasso and consistency.

Proof

$$||\breve{\Theta}^s_{nd} - \Theta^*||_1 \leq ||\breve{\Theta}^s_{nd} - \hat{\Theta}^s_{nd}|| + ||\hat{\Theta}^s_{nd} - \Theta^*||_1 \leq 2||\hat{\Theta}_{nd} - \Theta^*||_1$$

$$||\hat{R}^s - \Sigma^*||_{max} \leq \frac{b}{M}\lambda \text{ (under this probability event)}$$

$$|\hat{\Theta}^{s.nd}_{kk} - \Theta^*_{kk}| \leq C_1(\lambda)$$

$$||\hat{\theta}^{s.nd}_{(k)} - \theta^*_{(k)}||_1 \leq C_2(\lambda)$$

Rank Based CLIME

- $\hat{\Theta}_c^s = \arg\min ||\Theta||_1$ s.t. $||\hat{R}^s\Theta - I||_{max} \leq \lambda$
- $M = ||\Theta^*||_1$. For $n\lambda \geq 12\pi M$, with high probability $||\hat{\Theta}_c^s - \Theta^*||_{max} \leq 2M\lambda$.
- $n >> d^2 \log p$ and $\lambda = O(\sqrt{(\frac{logp}{n}})$, $||\hat{\Theta}_c^s - \Theta^*||_{max} = O_P(\sqrt{(\frac{logp}{n}})$
- Rates for other norms similar to original paper.
- Adaptive CLIME: $\hat{\Theta}_{ac}^s = \arg\min ||Wo\Theta||_1$, s.t. $||\hat{R}^s\Theta - I||_{max} \leq \lambda W$
- Sign consistency of adaptive version.

Rank based Neighborhood Lasso

- $\min_{\beta \in \mathbb{R}^{p-1}} \beta^T \hat{R}^s_{(k)} \beta - 2\beta^T \hat{r}^s_{(k)} + \lambda ||\beta||_{l1}$
- $\hat{R}_s$ could be negative definite, although $\hat{R} \in \mathbb{S}_{++}$
- $A = \begin{pmatrix} 1 & 0.7 & 0 \\ 0.7 & 1 & 0.7 \\ 0 & 0.7 & 1 \end{pmatrix}$ Eigenvalues: $\{1.99, 1.00, 0.01\}$.
  But $2\sin(\pi/6A)$ has Eigenvalues $\{2.01, 1.00, -0.01\}$.
- The objective function becomes ill defined.
- Other methods above are not affected by this.
- Possible to correct $\hat{R}_s$ but not implemented.

Transform function estimation

- Not needed, but just for theoretical interests.
- Winsorize: $T_{\delta_n}(x) = \begin{cases} \delta_n & \text{if } x < \delta_n, \\ x & \text{if } \delta_n \leq x \leq 1 - \delta_n, \\ 1 - \delta_n & \text{if } x > 1 - \delta_n \end{cases}$
- $\tilde{F}_j(t) = \frac{1}{n} \sum_{i=1}^{n} I(x_j^i \leq t)$
- $\tilde{f}_j(x) = \Phi^{-1}(T_{1/(2n)}(\tilde{F}_j(x)))$
- $I_n := [f^{-1}(-\sqrt{\frac{7}{4}(1-\gamma)\log n}), f^{-1}(\sqrt{\frac{7}{4}(1-\gamma)\log n})]$
- $\sup_{t \in I_n} |\tilde{f}_j(t) - f_j(t)| = o_P(1)$

Numerical Experiments

Summary:

- Non Gaussian data without outliers: npn-ns $\approx$ npn $\approx$ npn-sp $\approx$ npn-tau $>>$ normal
- Non Gaussian data with lower level of outliers: npn-tau $\approx$ npn-sp $>$ npn $>$ npn-ns $>>$ normal
- Non Gaussian data with higher level of outliers: npn-tau $>$ npn-sp $>>$ npn $>$ npn-ns $>>$ normal
- Gaussian data without outliers: normal $\approx$ npn-ns $\approx$ npn $>$ npn-sp $\approx$ npn-tau
- Gaussian data with low level of outliers: npn-tau $\approx$ npn-sp $>$ npn $>$ npn-ns $>>$ normal
- Gaussian data with higher level of outliers: npn-tau $>$ npn-sp $>$ npn $>$ npn-ns $>$ normal.

Connection of Rank Correlations

- $sign(\xi) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{it\xi}}{it} dt$
- $(x_1, y_1), \cdots, (x_n, y_n)$ from iid $N_2(\mu, \Sigma)$. cor(x,y)=ρ.

Let, $t = sign(u)sign(v)$, where $u = x_1 - x_2$, $v = y_1 - y_2$ then,

$$E(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} sign(u)sign(v)dF \text{ (as u,v are Gaussian)}$$
$$= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{dt_1}{it_1}\int_{-\infty}^{\infty}\frac{dt_2}{it_2}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iut_1+ivt_2}dF\}$$
$$= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dt_1}{it_1}\frac{dt_2}{it_2}exp(-\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2))$$
$$\frac{\delta E(t)}{\delta\rho} = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dt_1 dt_2 exp(-\frac{1}{2}(t_1^2 + 2\rho t_1 t_2 + t_2^2))$$

Cont'd

$$
\begin{aligned}
\frac{\delta E(t)}{\delta \rho} &= \frac{2}{\pi\sqrt{1-\rho^2}} \\
E(t) &= \frac{2}{\pi}\sin^{-1}\rho \\
E(r^s) &= \frac{3}{(n+1)}E(\tau_{jk}) + \frac{3(n-2)}{n+1}U_{jk} \\
&= \frac{3}{(n+1)}\frac{2}{\pi}\sin^{-1}\rho + \frac{3(n-2)}{n+1}\frac{2}{\pi}\sin^{-1}(\rho/2) \\
&= \frac{6}{\pi}\sin^{-1}(\frac{\rho}{2})
\end{aligned}
$$

Thank you

Reference:

- High-Dimensional Semi-parametric Gaussian Copula Graphical Models, Han Liu, Fang Han, Ming Yuan, John Lafferty, Larry Wasserman, Annals of Statistics, 2012, 14 (4).
- Regularized Rank Based Estimation of High Dimensional Non-Paranormal Graphical Models, Lingzhou Xue, Hui Zhou, Annals of Statistics, 2012, 40(5).