Optimal statistical decision for Gaussian graphical model selection

Petr Koldanov

National Research University Higher School of Economics,
Laboratoty of Algorithms and Technologies for Network Analysis (LATNA)
Nizhny Novgorod, Russia
Statistical network analysis team
pkoldanov@hse.ru

Riverside, California, USA, June 23, 2017

Random variable network

Random variable network is a pair (X, γ):

- $X = (X_1, \dots, X_N) -$random vector
- $\gamma -$measure of pairwise dependence.

Gaussian graphical network - particular case of random variable network

- $X = (X_1, \dots, X_N) -$random vector with multivariate normal distribution $N(\mu, \Sigma)$.
- $\gamma = \gamma_{i,j}^{part} = \rho^{i,j} = \frac{-\sigma^{i,j}}{\sqrt{\sigma^{i,i}\sigma^{j,j}}}$

Network model

Any random variable network generate network model.

- Complete weighted graph $G = (V, E, \gamma)$.
- Nodes of the network model has been identified with random variables $X_i, i = 1, \dots, N$.
- Weights of edges in the network model are given by measure $\gamma_{i,j} = \gamma(X_i, X_j)$.
- Gaussian graphical network generate gaussian graphical model.

Network structures

Network structures - subgraphs of the network model.
$G' = (V', E') : V' \subseteq V, E' \subseteq E$

- Concentration graph - popular network structures for gaussian graphical model.
- Concentration graph - edge (i, j) is included in the concentration graph if random variables X_i and X_j are conditionally dependent.

References

Gene-expression network.

- Lauritzen(1996), Drton & Perlman(2007)
- Thousands of publications.
- Numerical algorithms. Only FWER under control.
- No results for finite sample size.

Problem statement. Multiple decision approach

- (X, γ) - Gaussian graphical network.
- $G = (V, E, \gamma)$ - Gaussian graphical model.
- $G' = (V, E') : E' \subseteq E$ - concentration graph.
- Let $S = (s_{i,j})$ - adjacency matrix of concentartion graph, $S \in \mathcal{G}$ - set of all adjacency matrices.
- Let $H_S : \theta \in \Omega_S$-hypothesis that concentration graph has adjacency matrix $S, S \in \mathcal{G}$.
- There are observation $x(t) = (x_1(t), \ldots, x_N(t)), t = 1, \ldots, n$

Problem: construct optimal statistical procedure $\delta(x)$ to identify concentration graph from observations i.e. to select one from disjoint hypotheses H_S.

Statistical procedures. Risk function.

- Statistical procedure $\delta(x) = \{ \quad d_Q, \quad x \in D_Q \; ; \bigcup_{Q \in \mathcal{G}} D_Q = \mathcal{X}$
- $\delta(x) = d_Q$ - decision, that concentration graph has adjacency matrix $Q, Q \in \mathcal{G}$.
- $w(H_S; d_Q) = w(S, Q)$ - loss from the decision d_Q when the hypothesis H_S is true, $w(S, S) = 0, S \in \mathcal{G}$.
- Risk function of statistical procedure $\delta(x)$ is defined by

$$Risk(S, \theta; \delta) = \sum_{Q \in \mathcal{G}} w(S, Q) P_\theta(\delta(x) = d_Q), \quad \theta \in \Omega_S, S \in \mathcal{G}$$

$P_\theta(\delta(x) = d_Q)$ - the probability that decision d_Q is taken while the true decision is d_S.

Optimal statistical procedure.

- **Definition:** Statistical procedure δ is optimal in class $\mathcal{D}$ if $R(S,\theta,\delta) \leq R(S,\theta,\delta'), \forall S, \forall \theta \in \Omega_S, \forall \delta' \in \mathcal{D}$.
- Restrict attention to W-unbiased statistical procedures

$$E_\theta w(\theta,\delta) \leq E_\theta w(\theta',\delta), \forall \theta, \theta' \in \Omega$$

$$R(S,\theta,\delta) \leq R(S',\theta,\delta), \forall S, S', \theta \in \Omega_S$$

Multiple hypotheses testing approach. Individual hypotheses

Individual hypotheses

$$h_{i,j} : \rho^{i,j} = 0 \text{ vs } k_{i,j} : \rho^{i,j} \neq 0$$

According to Lauritzen S.L.[1]

$$\rho^{i,j} = \frac{-\sigma^{i,j}}{\sqrt{\sigma^{i,i}\sigma^{j,j}}}$$

Then

$$h_{i,j} : \sigma^{i,j} = 0 \text{ vs } k_{i,j} : \sigma^{i,j} \neq 0$$

[1] Lauritzen S.L.(1996) Graphical model. Oxford university press.

Multiple statistical procedure

Let $\varphi_{i,j}(x)$ tests of individual hypotheses.
Define

$$\Phi(x) = \begin{pmatrix} 0, & \varphi_{1,2} & \dots & \varphi_{1,N} \\ \varphi_{1,2} & 0, & \dots & \varphi_{2,N} \\ \dots & \dots & \dots & \dots \\ \varphi_{1,N} & \varphi_{2,N} & \dots & 0 \end{pmatrix}$$

Define $\delta(x) = d_G$ if $\Phi(x) = G$

Existing statistical procedures. Single step procedure.[3]

Test of individual edge inclusion is

$$\varphi_{ij}^{St}(x) = \begin{cases} 1, & |z^{ij}| > c_{ij} \\ 0, & |z^{ij}| \leq c_{ij} \end{cases}$$

where $z^{ij} = \frac{1}{2}\ln\left(\frac{1+r^{ij}}{1-r^{ij}}\right)$, $r^{ij} = \frac{-s^{ij}}{\sqrt{s^{ii}s^{jj}}}$-sample partial correlation, s^{ij}-elements of matrix S^{-1}. c_{ij} from [2] $P_{\rho^{ij}=0}(|z^{ij}| > c_{ij}) = \alpha$
Then construct matrix $\Phi^{St}(x) = (\varphi_{ij}^{St}(x))$
Properties of the associated multiple decision **standard** statistical procedure $\delta^{St} = d_G$ if $\Phi^{St}(x) = G$ were not investigated.

---

[2]Anderson T.W.(2003) An introduction to multivariate statistical analysis.3-d edition. Wiley-Interscience, New York

[3]Edwards, D.M.(2000) Introduction to Graphical Modeling. New York, Springer.

Stepdown procedure.[4]

Let p_k are p-values of tests $\varphi_{ij}^{St}(x)$ $k = 1, \dots, \frac{N(N-1)}{2}$. Order $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(M-1)} \leq p_{(M)}$ and let $h_{(1)}, h_{(2)}, \dots, h_{(M)}$ be the corresponding hypotheses.

- Step 1: If $p_{(1)} \geq \frac{\alpha}{M}$ then the decision is : accept all hypotheses $h_{(i)}$, $i = 1, 2, \dots, M$ and stop, else reject hypothesis $h_{(1)}$ and go to the step 2.
- Step 2: If $p_{(2)} \geq \frac{\alpha}{M-1}$ then the decision is : accept all hypotheses $h_{(i)}$, $i = 2, \dots, M$ and stop, else reject hypothesis $h_{(2)}$ and go to the step 3.
- ...
- Step M: If $|p_M| \geq \alpha$ then the decision is: accept hypothesis $h_{(M)}$ else reject all hypotheses.

Properties - control of FWER.

Type II error are not under control.

[4] M. Drton, M.D. Perlman.(2007) Multiple testing and error control in Gaussian graphical model selection. Statistical Science, 22,3, 430-449.

Our approach. Additive loss function

$l'_{i,j}$-loss from false inclusion of edge (i,j)
$l''_{i,j}$-loss from false non inclusion of the edge (i,j) $i,j = 1,2,\ldots,N;\ i \neq j$.
Loss function $w(S,Q)$ is *additive*[5] if:

$$w(S,Q) = \sum_{(i,j):s_{i,j}=0,q_{i,j}=1} l'_{i,j} + \sum_{(i,j):s_{i,j}=1,q_{i,j}=0} l''_{i,j} \tag{1}$$

Theorem 1[6] Let the loss function w be additive and $l'_{i,j} = l'$, $l''_{i,j} = l''$, $i \neq j$, $i,j = 1,2,\ldots,N$. Then

$$Risk(S,\theta;\delta) = \sum_{i,j} r(s_{i,j},\varphi_{i,j}) = l' E_\theta[Y_I(S,\delta)] + l'' E_\theta[Y_{II}(S,\delta)], \quad \theta \in \Omega_S$$

where $Y_I(S,\delta)$, $Y_{II}(S,\delta)$ are the numbers of Type I and Type II errors by δ when the true decision is d_S.

[5] E.L.Lehmann (1957) A theory of some multiple decision problems.I Ann.Math.Stat.,28,1-25, 547-572.

[6] V.A. Kalyagin, A.P. Koldanov,P.A. Koldanov, P.M. Pardalos. Optimal statistical decision for Gaussian graphical model selection.arXiv:1701.02071v1

UMPU test for individual hypotheses

Theorem 2[7] Optimal in the class of unbiased statistical level α test for individual hypothesis $h_{ij} : \rho^{i,j} = 0$ against $k_{ij} : \rho^{i,j} \neq 0$ is:

$$\varphi_{ij}^{opt} = \begin{cases} 0, & \frac{|as_{ij} - \frac{b}{2}|}{\sqrt{\frac{b^2}{4} + ac}} < 1 - 2c_{\alpha}^{beta} \\ 1, & \frac{|as_{ij} - \frac{b}{2}|}{\sqrt{\frac{b^2}{4} + ac}} > 1 - 2c_{\alpha}^{beta} \end{cases} \quad (2)$$

where $\det(s_{kl}) = -as_{ij}^2 + bs_{ij} + c$, c_{α}^{beta} is the α-quantile of Beta distribution. $(a = a(\{s_{kl}\}), b = b(\{s_{kl}\}), c = c(\{s_{kl})\})$.

[7] Koldanov P., Koldanov A. P., Kalyagin V. A., Pardalos P. M. Uniformly most powerful unbiased test for conditional independence in Gaussian graphical model // Statistics & Probability Letters, 2017, Vol. 122, P. 90-95.

Sketch of proof. Wishart distribution

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1N} \\ s_{21} & s_{22} & \dots & s_{2N} \\ \dots & \dots & \dots & \dots \\ s_{N1} & s_{N2} & \dots & s_{NN} \end{pmatrix} \quad (3)$$

$$f(\{s_{k,l}\}) = \frac{[\det(\sigma^{kl})]^{n/2} \times [\det(s_{kl})]^{(n-N-2)/2} \times \exp[-(1/2)\sum_k \sum_l s_{k,l}\sigma^{kl}]}{2^{(Nn/2)} \times \pi^{N(N-1)/4} \times \Gamma(n/2)\Gamma((n-1)/2) \cdots \Gamma((n-N+1)/2)}$$

if the matrix (s_{kl}) is positive definite, and $f(\{s_{kl}\}) = 0$ otherwise. Let I be the interval of positive definiteness of the matrix. One has for a fixed $i < j$:

$$f(\{s_{kl}\}) = C(\{\sigma^{kl}\}) \times \exp[-\sigma^{ij}s_{ij} - \frac{1}{2} \sum_{(k,l)\neq(i,j);(k,l)\neq(j,i)} s_{kl}\sigma^{kl}] \times h(\{s_{kl}\})$$

Sketch of proof. UMPU test

UMPU test for testing hypothesis

$$h_{ij} : \rho^{i,j} = 0 \text{ vs } k_{ij} : \rho^{i,j} \neq 0$$

has the Neyman structure and can be written as

$$\delta_{i,j}(\{s_{kl}\}) = \begin{cases} \partial_{i,j}, & if\ c_1(\{s_{kl}\}) \leq s_{ij} \leq c_2(\{s_{kl}\}),\ (k,l) \neq (i,j) \\ \partial_{i,j}^{-1}, & if\ s_{ij} < c_1(\{s_{kl}\})\ \ s_{ij} > c_2(\{s_{kl}\}),\ (k,l) \neq (i,j) \end{cases} \tag{4}$$

where constants are defined from

$$\frac{\int_{I \cap [c_1;c_2]} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij}}{\int_I \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij}} = 1 - \alpha_{i,j}, \tag{5}$$

$$\begin{aligned} &\int_{I \cap [-\infty;c_1]} s_{ij} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij} + \\ &+ \int_{I \cap [c_2;+\infty]} s_{ij} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij} = \\ &= \alpha_{i,j} \int_I s_{ij} \exp[-\sigma_0^{ij} s_{ij}][\det(s_{kl})]^{(n-N-2)/2} ds_{ij}, \end{aligned} \tag{6}$$

Sketch of proof. UMPU test.

Under $\sigma_0^{i,j} = 0$ equation (5) is

$$\frac{\int_{I\cap[c_1;c_2]}[\det(s_{kl})]^{(n-N-2)/2}ds_{ij}}{\int_I[\det(s_{kl})]^{(n-N-2)/2}ds_{ij}} = 1-\alpha_{i,j} \tag{7}$$

Let $K = \frac{n-N-2}{2}, x = s_{ij}$. Then

$\int_f^d (ax^2 - bx - c)^K dx = (-1)^K a^K (x_2 - x_1)^{2K+1} \int_{\frac{f-x_1}{x_2-x_1}}^{\frac{d-x_1}{x_2-x_1}} u^K(1-u)^K du$

Equation (7) can be written as

$$\int_{\frac{c_1-x_1}{x_2-x_1}}^{\frac{c_2-x_1}{x_2-x_1}} u^K(1-u)^K du = (1-\alpha)\int_0^1 u^K(1-u)^K du = (1-\alpha)\frac{\Gamma(K+1)\Gamma(K+1)}{\Gamma(2K+2)} \tag{8}$$

Acceptance region is: $c_\alpha^{beta} \leq \frac{s_{i,j}-x_1}{x_2-x_1} \leq 1 - c_\alpha^{beta}$ or

$2c_\alpha^{beta} - 1 \leq \frac{as_{i,j}-b/2}{\sqrt{b^2/4+ac}} \leq 1 - 2c_\alpha^{beta}$

UMPU test is equivalent to partial correlation test

Sample partial correlation test for testing hypothesis $\rho^{i,j} = 0$:

$$\varphi_{i,j} = \begin{cases} 0, & |r^{i,j}| \leq c_{i,j} \\ 1, & |r^{i,j}| > c_{i,j} \end{cases} \tag{9}$$

where $c_{i,j}$ is $(1 - \alpha/2)$-quantile of the distribution with density function

$$f(x) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(n - N + 1)/2)}{\Gamma((n - N)/2)} (1 - x^2)^{(n-N-2)/2}, \quad -1 \leq x \leq 1 \tag{10}$$

Theorem 3 Sample partial correlation test (9) is equivalent to UMPU test (2) for testing hypothesis $\rho^{i,j} = 0$ vs $\rho^{i,j} \neq 0$.

Equivalence of partial correlation and UMPU tests. Sketch of proof.

It is sufficient to prove that

$$\frac{S^{i,j}}{\sqrt{S^{i,i}S^{j,j}}} = \frac{as_{i,j} - \frac{b}{2}}{\sqrt{\frac{b^2}{4} + ac}} \tag{11}$$

Let $A = (a_{k,l})$ be an $(N \times N)$ symmetric matrix. Fix $i < j$, $i, j = 1, 2, \ldots, N$. Denote by $A(x)$ the matrix obtained from A by replacing the elements $a_{i,j}$ and $a_{j,i}$ by x. Denote by $A^{i,j}(x)$ the cofactor of the element (i, j) in the matrix $A(x)$. Then the following statement is true
Lemma 1 One has $[\det A(x)]' = -2A^{i,j}(x)$.

Equivalence of partial correlation and UMPU tests. Sketch of proof.

$$\det(S(x)) = -ax^2 + bx + c \to [\det S(x)]' = -2ax + b = -2S^{i,j}(x)$$

i.e. $S^{i,j}(x) = ax - b/2$.

$$x = s_{i,j} \to as_{i,j} - \frac{b}{2} = S^{i,j}$$

It is sufficient to prove that $\sqrt{S^{i,i}S^{j,j}} = \sqrt{\frac{b^2}{4} + ac}$.

Let $x_2 = \frac{b+\sqrt{b^2+4ac}}{2a}$ be the maximum root of equation $ax^2 - bx - c = 0$.

Then $ax_2 - \frac{b}{2} = \sqrt{\frac{b^2}{4} + ac}$.

Equivalence of partial correlation and UMPU tests. Sketch of proof.

Consider

$$r^{i,j}(x) = \frac{-S^{i,j}(x)}{\sqrt{S^{i,i}S^{j,j}}}$$

According to Silvester determinant identity:

$$S^{\{i,j\},\{i,j\}} \det S(x) = S^{i,i}S^{j,j} - [S^{i,j}(x)]^2$$

Therefore for $x = x_1$ and $x = x_2$ one has

$$S^{i,i}S^{j,j} - [S^{i,j}(x)]^2 = 0$$

For $x = x_1$ and $x = x_2$ one has $r^{i,j}(x) = \pm 1$. The equation $S^{i,j}(x) = ax - \frac{b}{2}$ implies that when x is increasing from x_1 to x_2 then $r^{i,j}(x)$ is decreasing from 1 to -1. That is $r^{i,j}(x_2) = -1$, i.e. $ax_2 - \frac{b}{2} = \sqrt{S^{i,i}S^{j,j}}$. Therefore

$$\sqrt{S^{i,i}S^{j,j}} = \sqrt{\frac{b^2}{4} + ac}$$

Multiple decision statistical procedure

$$\Phi^{opt}(x) = \begin{pmatrix} 0, & \varphi^{opt}_{12}(x), & \ldots, & \varphi^{opt}_{1N}(x) \\ \varphi^{opt}_{21}(x), & 0, & \ldots, & \varphi^{opt}_{2N}(x) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi^{opt}_{N1}(x), & \varphi^{opt}_{N2}(x), & \ldots, & 0 \end{pmatrix}. \tag{12}$$

Define multiple statistical procedure for concentration graph identification

$$\delta^{opt}(x) = d_G, \text{ iff } \Phi^{opt}(x) = G \tag{13}$$

Multiple decision statistical procedure

Theorem 4[8] Let the loss function w be additive and

$$\alpha_{i,j} = \frac{l''_{i,j}}{l'_{i,j} + l''_{i,j}}, \ i \neq j, \ i,j = 1, 2, \ldots, p. \tag{14}$$

Then the procedure δ^{opt} is optimal multiple decision statistical procedure for Gaussian graphical model selection in the class of w-unbiased procedures.

Note: statistical procedure $\delta(x) = d_G$ iff $\Phi(x) = G$ and $\varphi(x)$ is defined by (9) with constant from (10) is optimal multiple decision statistical procedure for Gaussian graphical model selection in the class of w-unbiased procedures for additive loss function.

[8] V.A. Kalyagin, A.P. Koldanov,P.A. Koldanov, P.M. Pardalos. Optimal statistical decision for Gaussian graphical model selection.arXiv:1701.02071v1

Experimental result

Consider two multiple decision Bonferroni type statistical procedures for concentration graph identification. First procedure - standard procedure. Second procedure based on the UMPU individual tests.

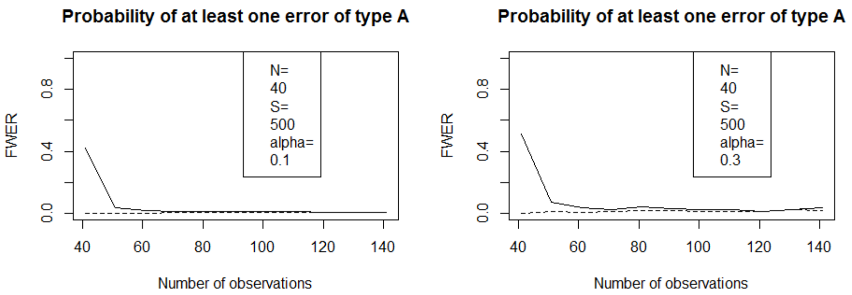

Figure: FWER as function from n. N=40. Left: significance level $\alpha = 0.1$. Right: significance level $\alpha = 0.3$ Solid line - procedure δ^{St} (standard). Dashed line - procedure δ^{opt} (optimal). Horizontal line - number of observations n.

Conclusions

- UMPU Neyman structure test for testing hypothesis $h^{i,j} : \rho^{i,j} = 0$ vs $k_{i,j} : \rho^{i,j} \neq 0$ is constructed.
- Standard test based on sample partial correlation with threshold from exact distribution is UMPU.
- Multiple decision statistical procedure based on UMPU individual tests is optimal unbiased procedure under additive loss function.

THANK YOU FOR YOUR ATTENTION!