Optimal test of conditional independence testing in multivariate normal distribution

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

Moscow, Russia, October 24, 2017

Introduction and problem statement

- Let $X = (X_1, \ldots, X_N)$ be random vector with multivariate normal distribution

$$\begin{pmatrix} X_1 \\ X_2 \\ \ldots \\ X_N \end{pmatrix} = N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \ldots \\ \mu_N \end{pmatrix} \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \ldots & \sigma_{1,N} \\ \sigma_{2,1} & \sigma_{2,2} & \ldots & \sigma_{2,N} \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{N,1} & \sigma_{N,2} & \ldots & \sigma_{N,N} \end{pmatrix}\right)$$

- Let $\rho^{i,j} = \rho_{i,j \cdot 1,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,N}$ be the partial correlation between X_i and X_j.
-

$$h_{i,j} : \rho^{i,j} = 0$$

versus

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

Partial correlation as correlation between residuals

For simplicity of notations let $i = 1, j = 2$. Define linear regression

$$X_1 = \beta_{1,3}X_3 + \ldots + \beta_{1,N}X_N + \epsilon_1$$

$$X_2 = \beta_{2,3}X_3 + \ldots + \beta_{2,N}X_N + \epsilon_2$$

Then residuals are

$$X_{1\cdot 3,\ldots,N} = X_1 - \beta_{1,3}X_3 - \ldots - \beta_{1,N}X_N$$

$$X_{2\cdot 3,\ldots,N} = X_2 - \beta_{2,3}X_3 - \ldots - \beta_{2,N}X_N$$

Partial correlation is

$$\rho^{1,2} = \rho_{1,2\cdot 3,\ldots,N} = \rho(X_{1\cdot 3,\ldots,N}, X_{2\cdot 3,\ldots,N}) = \frac{E(X_{1\cdot 3,\ldots,N}, X_{2\cdot 3,\ldots,N})}{\sqrt{E(X^2_{1}\cdot_{3,\ldots,N}), E(X^2_{2}\cdot_{3,\ldots,N})}}$$

If $N = 3$ then $\rho^{1,2} = \rho_{1,2\cdot 3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho^2_{13})(1-\rho^2_{23})}}$

Partial correlation as correlation in conditional distribution

Let $\Sigma_{1,2} = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{pmatrix}$; $\Sigma_{3,N} = \begin{pmatrix} \sigma_{3,3} & \dots & \sigma_{3,N} \\ \sigma_{4,3} & \dots & \sigma_{4,N} \\ \dots & \dots & \dots \\ \sigma_{N,3} & \dots & \sigma_{N,N} \end{pmatrix}$

$\Sigma^{1,2} = \begin{pmatrix} \sigma_{1,3} & \dots & \sigma_{1,N} \\ \sigma_{2,3} & \dots & \sigma_{2,N} \end{pmatrix}$; $\Sigma^{2,1} = \begin{pmatrix} \sigma_{3,1} & \sigma_{3,2} \\ \dots & \dots \\ \sigma_{N,1} & \sigma_{N,2} \end{pmatrix}$;

Conditional distribution $F_{X_1,X_2/X_3,\dots,X_N}$ is normal $N(\nu, \Sigma')$ where

$$\Sigma' = \Sigma_{1,2} - \Sigma^{1,2} \left(\Sigma_{3,N}\right)^{-1} \Sigma^{2,1}$$

Partial correlation is

$$\rho^{1,2} = \rho_{1,2\cdot 3,\dots,N} = \frac{\sigma'_{1,2}}{\sqrt{\sigma'_{1,1}\sigma'_{2,2}}}$$

Existing statistical procedures. Exact test[1]

Exact sample partial correlation test for testing hypothesis

$$h_{i,j} : \rho^{i,j} = 0$$

versus

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

is:

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

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

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

[1] Anderson (2003) An Introduction to Multivariate Statistical Analysis. New York, Springer.

Existing statistical procedures. Asymptotic test[2]

Asymptotic test of hypothesis $h_{i,j} : \rho^{i,j} = 0$ vs $k_{i,j} : \rho^{i,j} \neq 0$ has the form

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

where $z^{ij} = \frac{1}{2} \ln \left(\frac{1+r^{ij}}{1-r^{ij}} \right)$, r^{ij}-sample partial correlation.

$$z^{ij} \xrightarrow{d} N(0,1) \text{ if } n \to \infty$$

Then c_{ij} is $(1-\alpha/2)$-quantile of $N(0,1)$ distribution.

[2] Anderson (2003) An Introduction to Multivariate Statistical Analysis. New York, Springer.

Test of hypothesis

Hypothesis

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

According to Lauritzen S.L.[3]

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

Then

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

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

UMPU test for individual hypotheses

Theorem 1 Optimal in the class of unbiased statistical level α test for hypothesis $h_{ij} : \rho^{i,j} = 0$ against $k_{ij} : \rho^{i,j} \neq 0$ is:

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

where $\det(s_{kl}) = -as_{ij}^2 + bs_{ij} + c$, c_{α}^{beta} is the α-quantile of Beta distribution. $(a = a(\{s_{kl}\}), b = b(\{s_{kl}\}), c = c(\{s_{kl})\})$. [4]

[4] 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.

Wishart distribution

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

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

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 $i < j$:

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

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

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

where constants are defined from

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

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

UMPU test.

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

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

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

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

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 exact partial correlation test

Let us consider exact 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

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

Theorem 2 Exact 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 exact partial correlation and UMPU tests.

Since

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

it is sufficient to prove that

$$\frac{S^{i,j}}{\sqrt{S^{i,i}S^{j,j}}} = \frac{as_{i,j} - \frac{b}{2}}{\sqrt{\frac{b^2}{4} + ac}} \tag{10}$$

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 exact partial correlation and UMPU tests.

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

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 exact partial correlation and UMPU tests.

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

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

Conclusion

1. The UMPU test for testing hypothesis $h_{i,j} : \rho^{i,j} = 0$ versus $k_{i,j} : \rho^{i,j} \neq 0$ in multivariate normal distribution is constructed.
2. It is shown that UMPU test is equivalent to exact test based on partial correlation. Then the exact test based on partial correlation is UMPU one.

THANK YOU FOR YOUR ATTENTION!