УДК 519.233.32

Оптимальный тест проверки гипотезы об условной независимости в многомерном нормальном распределении

П. А. Колданов*, **А. П. Колданов***, **В. А. Калягин***, **П.М. Пардалос*†**

* *Лаборатория алгоритмов и технологий анализа сетевых структур, НИУ Высшая школа экономики, Нижегородский филиал ул. Б.-Печерская, д.25, Нижний Новгород, Россия*

† *Center for Applied Optimization, University of Florida, USA*

Аннотация. Для проверки гипотезы об условной независимости в многомерном нормальном распределении широко используется тест, основанный на выборочном частном коэффициенте корреляции. В настоящей работе построен равномерно наиболее мощный несмещенный (РНМН) тест структуры Неймана проверки этой гипотезы. Показано, что построенный РНМН тест эквивалентен тесту основанному на выборочном частном коэффициенте корреляции и, тем самым, тест частных корреляций является РНМН тестом проверки условной независимости компонент нормального вектора.

Ключевые слова: условная независимость, экспоненциальные семейства, многомерное нормальное распределение, тесты структуры Немана, равномерно наиболее мощный несмещенный тест.

1. Введение.

Пусть $X = (X_1, X_2, \dots, X_N)$ — случайный вектор с многомерным нормальным распределением. Частный коэффициент корреляции $\rho_{i,j\bullet N(i,j)}$ между X_i, X_j при заданных X_k, $k \in N(i,j) = \{1, 2, \dots, N\} \setminus \{i, j\}$, определяется как корреляция между X_i, X_j в двумерном условном распределении X_i, X_j при заданных X_k, $k \in N(i,j)$. Известно (см. [1], стр. 35), что это распределение является нормальным. Поэтому условная независимость X_i, X_j при заданных X_k, $k \in N(i,j)$, эквивалентна равенству $\rho_{i,j\bullet N(i,j)} = 0$.

Для проверки гипотезы $\rho_{i,j\bullet N(i,j)} = 0$ используется стандартный тест, основанный на выборочном частном коэффициенте корреляции (см. [1], стр. 143). В то же время, насколько известно авторам настоящей работы, неизвестно, является ли стандартный тест равномерно наиболее мощным несмещенным (РНМН). В настоящей работе построен РНМН тест структуры Неймана проверки гипотезы об условной независимости. Доказано, что РНМН тест эквивалентен стандартному. Таким образом, стандартный тест проверки условной независимости является РНМН.

2. Обозначения и постановка задачи.

Пусть случайный вектор $X = (X_1, X_2, \dots, X_N)$ имеет многомерное нормальное распределение $N(\mu, \Sigma)$, где $\mu = (\mu_1, \mu_2, \dots, \mu_N)$ — вектор математических ожиданий, $\Sigma = (\sigma_{i,j})$ — ковариационная матрица $i, j = 1, 2, \dots, N$. Пусть $x(t)$, $t = 1, 2, \dots, n$, — выборка размера n из распределения вектора X и

$$s_{i,j} = \frac{1}{n} \sum_{t=1}^{n} (x_i(t) - \overline{x_i})(x_j(t) - \overline{x_j}),$$

— выборочная ковариация между X_i, X_j, где $\overline{x_i} = (1/n) \sum_{t=1}^{n} x_i(t)$. Обозначим $S = (s_{i,j})$ — выборочная ковариационная матрица, $\Sigma^{-1} = (\sigma^{i,j})$ — обратная матрица к Σ, $\rho_{i,j \bullet N(i,j)} = \rho^{i,j}$. Задача проверки попарной условной независимости X_i, X_j заключается в проверке гипотезы:

$$h_{i,j} : \rho^{i,j} = 0 \text{ против альтернативы } k_{i,j} : \rho^{i,j} \neq 0. \tag{1}$$

Как показано в ([2], стр. 129) частный коэффициент корреляции между X_i, X_j при заданных X_k, $k \in N(i,j) = \{1, 2, \dots, N\} \setminus \{i, j\}$, можно представить через элементы обратной ковариакционной матрицы в виде

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

Поэтому задача проверки попарной условной независимости (1) может быть сформулирована как задача проверки гипотезы

$$h_{i,j} : \sigma^{i,j} = 0, \quad \text{против} \quad k_{i,j} : \sigma^{i,j} \neq 0. \tag{2}$$

3. Тесты структуры Неймана.

Для проверки гипотезы (2) будем использовать тесты структуры Неймана (см. [3], стр. 115). Пусть $f(x; \theta)$ — плотность экспоненциального семейства:

$$f(x; \theta) = c(\theta) \exp\left(\sum_{j=1}^{M} \theta_j T_j(x) \right) m(x).$$

РНМН тест проверки гипотезы

$$h_j : \theta_j = \theta_j^0 \text{ против альтернативы } k_j : \theta_j \neq \theta_j^0,$$

имеет вид:

$$
\begin{aligned}
&\varphi_j = 1 - \mathbf{I}(A_j),\\
&A_j = \left\{c'_j(t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_M) < t_j < c''_j(t_1, \ldots, t_{j-1}, t_{j+1}, ..., t_M)\right\}.
\end{aligned} \tag{3}
$$

где $t_i = T_i(x), i = 1, \ldots, M$. Константы c'_j, c''_j определяются из уравнений (см. [3], стр. 120)

$$
\int_{c'_j}^{c''_j} f(t_j; \theta_j^0 | T_i = t_i, i = 1, \ldots, M; i \neq j) dt_j = 1 - \alpha \tag{4}
$$

$$
\begin{aligned}
&\textstyle\int_{-\infty}^{c'_j} t_j f(t_j; \theta_j^0 | T_i = t_i, i = 1, \ldots, M; i \neq j) dt_j +\\
&+ \textstyle\int_{c''_j}^{+\infty} t_j f(t_j; \theta_j^0 | T_i = t_i, i = 1, \ldots, M; i \neq j) dt_j =\\
&= \alpha \textstyle\int_{-\infty}^{+\infty} t_j f(t_j; \theta_j^0 | T_i = t_i, i = 1, \ldots, M; i \neq j) dt_j
\end{aligned} \tag{5}
$$

где $f(t_j; \theta_j^0 | T_i = t_i, i = 1, \ldots, M; i \neq j)$ — плотность условного распределения T_j, α — уровень значимости. При этом если $\varphi_j = 1$, то гипотеза h_j отвергается.

4. Равномерно наиболее мощный несмещенный тест для проверки гипотезы об условной независимости.

Как известно (см. [1], стр. 252), совместным распределением $s_{k,l}$, $k, l = 1, 2, \ldots, N$, $n > N$, является распределение Уишарта с плотностью:

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

если матрица $S = (s_{k,l})$ положительно определена, и $f(\{s_{k,l}\}) = 0$ иначе. Функция плотности распределения Уишарта может быть записана следующим образом:

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

где

$$
C(\{\sigma^{k,l}\}) = c_1^{-1} [\det(\sigma^{k,l})]^{n/2},
$$

$$
c_1 = 2^{(Nn/2)} \times \pi^{N(N-1)/4} \times \Gamma(n/2)\Gamma((n-1)/2) \cdots \Gamma((n-N+1)/2),
$$

$$m(\{s_{k,l}\}) = [\det(s_{k,l})]^{(n-N-2)/2}.$$

В соответствии с (3), (4), (5) РНМН тест проверки гипотезы (2) имеет вид:

$$\varphi_{i,j}(\{s_{k,l}\}) =$$

$$= \begin{cases} 0, & \text{если } c'_{i,j}(\{s_{k,l}\}) < s_{i,j} < c''_{i,j}(\{s_{k,l}\}),\ (k,l) \neq (i,j) \\ 1, & \text{если } s_{i,j} \leq c'_{i,j}(\{s_{k,l}\}) \text{ или } s_{i,j} \geq c''_{i,j}(\{s_{k,l}\}),\ (k,l) \neq (i,j), \end{cases}$$

$$\frac{\int_{I\cap[c'_{i,j};c''_{i,j}]}[\det(s_{k,l})]^{(n-N-2)/2}ds_{i,j}}{\int_I[\det(s_{k,l})]^{(n-N-2)/2}ds_{i,j}} = 1-\alpha,$$

$$\int_{I\cap(-\infty;c'_{i,j}]} s_{i,j}[\det(s_{k,l})]^{(n-N-2)/2}ds_{i,j} +$$

$$+\int_{I\cap[c''_{i,j};+\infty)} s_{i,j}[\det(s_{k,l})]^{(n-N-2)/2}ds_{i,j} =$$

$$= \alpha\int_I s_{i,j}[\det(s_{k,l})]^{(n-N-2)/2}ds_{i,j},$$

где I — интервал значений $s_{i,j}$, такой что матрица $S = (s_{k,l})$ положительно определена и α — заданный уровень значимости теста. Если $\varphi_{i,j}(\{s_{k,l}\}) = 1$, то гипотеза $h_{i,j} : \sigma^{i,j} = 0$ отвергается. Подчеркнем условный характер теста $\varphi_{i,j}(\{s_{k,l}\})$, который заключается в том, что пороги $c'_{i,j}, c''_{i,j}$ зависят от значений $s_{k,l}, (k,l) \neq (i,j)$.

В [4] показано, что РНМН тест проверки гипотезы об условной независимости X_i, X_j можно представить в виде:

$$\varphi_{i,j} = \begin{cases} 0, & 2q-1 < \dfrac{as_{i,j}-\frac{b}{2}}{\sqrt{\frac{b^2}{4}+ac}} < 1-2q \\ 1, & \text{иначе}, \end{cases} \tag{6}$$

где a, b, c определяются, как коэффициенты в представлении $\det(s_{k,l})$ полиномом второй степени относительно $s_{i,j}$, q есть $\frac{\alpha}{2}$-квантиль бета-распределения $Be(\frac{n-N}{2}, \frac{n-N}{2})$.

5. Стандартный тест.

Хорошо известный тест, основанный на выборочных частных корреляциях $r^{i,j}$, проверки гипотезы $\rho^{i,j} = 0$ имеет вид:

$$\varphi_{i,j} = \begin{cases} 0, & |r^{i,j}| \leq c_{i,j} \\ 1, & |r^{i,j}| > c_{i,j}, \end{cases} \tag{7}$$

где $c_{i,j}$ есть $(1-\alpha/2)$-квантиль распределения с плотностью

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

Теорема [4]: *Тест* (7) *эквивалентен РНМН тесту* (6) *проверки гипотезы* $\rho^{i,j} = 0$ *против* $\rho^{i,j} \neq 0$.

Следовательно, РНМН тест проверки гипотезы об условной независимости X_i и X_j может быть записан в следующем виде:

$$\varphi_{i,j} = \begin{cases} 0, & 2q-1 < r^{i,j} < 1-2q \\ 1, & \text{иначе} \end{cases}$$

где q есть $\frac{\alpha}{2}$-квантиль бета-распределения $Be(\frac{n-N}{2}, \frac{n-N}{2})$.

Литература

1. *Anderson T.* An introduction to multivariate statistical analysis.3-d edition. — Wiley-Interscience, New York, 2003.
2. *Lauritzen S.L.* Graphical models. — Oxford University Press, 1996.
3. *Lehmann E.L., Romano J.P.* Testing statistical hypotheses. 3-d edition. — Springer, New York, 2005.
4. *Koldanov P., Koldanov A., Kalyagin V., Pardalos P.* Uniformly most powerful unbiased test for conditional independence in Gaussian graphical model // Statistics and Probability Letters. — 2017. — Vol.122. — P. 90-95.

UDC 519.233.32

Optimal test of conditional independence testing in multivariate normal distribution

P. A. Koldanov*, A. P. Koldanov*, V. A. Kalyagin*, P. M. Pardalos*†

* *Laboratory of algorithms and technologies for network analysis,*
NRU Hughre School of Economics
B.-Pecherskay str, 25, Nijni Novgorod, Russia
† *Center for Applied Optimization,*
University of Florida, USA

In practical applications partial correlation tests are widely used. In the paper uniformly most powerful unbiased (UMPU) test of Neymann structure is obtained. It turns out, that this test can be reduced to usual partial correlation test. It is shown that constructed UMPU test is equivalent to partial correlation test. Then partial correlation test is UMPU for conditional independence testing.

Keywords: conditional independence, exponential families, multivariate normal distribution, sample partial correlation test, tests of Neyman structure, uniformly most powerful unbiased tests.