Commentationes Mathematicae Universitatis Carolinae

12,1 (1971)

ASYMPTOTIC DISTRIBUTION OF RANK STATISTICS USED FOR MULTIVARIATE TESTING SYMMETRY

(Preliminary communication)

Marie HUŠKOVÁ, Praha

This preliminary communication contains assertions on asymptotic distributions of statistics used for the nonparametric multivariate testing symmetry. The results are proved under the hypothesis of symmetry, a near alternative and a general alternative. The proofs are based on the corresponding theorems for univariate case and the theorem on convergence in distribution for vectors (see Theorem V.2.1 in [5]).

Let $\mathbf{X}_j = (X_{j1}, \dots, X_{jp})$, $1 \leq j \leq N$, be independent p-dimensional random variables and let R^+_{ji} be the rank of $|X_{ji}|$ in the sequence of absolute values $|X_{1i}|, \dots, |X_{Ni}|$. Put

$$\mathbf{S}_c = (S_{1c}, \dots, S_{pc})' ,$$

$$S_{ic} = \sum_{j=1}^{N} c_{ji}\, a_{Ni}(R^+_{ji})\, \operatorname{sgn} X_{ji} , \quad 1 \leq i \leq p ,$$

with c_{ji} being regression constants, $a_{Ni}(j)$ scores and

$$\operatorname{sgn} x = \begin{cases} 1 & \text{if } x \geq 0 , \\ 1 & \text{if } x < 0 . \end{cases}$$

AMS, Primary 62E20, 62H10, 62G99 Ref.Ž. 9.77 , 9.74
Secondary -

By Σ_{pc} we denote the conditional matrix of $\mathbf{S}_c$ given $|X_{ji}| \frac{sgn\, X_{ji}}{sgn\, X_{j1}}$, $1 \leq j \leq N$, $1 \leq i \leq p$, under (1) given below and by Σ^-_{pc} the generalized inverse of Σ_{pc} (see [4]).

We are interested in an investigation of the asymptotic distribution of the statistics

$$Q_c = \mathbf{S}'_c \, \Sigma^-_{pc} \, \mathbf{S}_c$$

under various systems of conditions.

The problem was solved for example in the papers of Puri and Sen [3], Patel [2] and Adichie [1]. The attention has been devoted to the case $c_{ji} = 1$ or $X_{j1} = X_{j2} = \ldots = X_{jp}$.

At first let us consider the following system of conditions for the distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_N$:

(1)

a) $\mathbf{X}_1, \ldots, \mathbf{X}_N$ are independent;

b) $F_{1ik} = \ldots = F_{Nik}$, $i \neq k$, $1 \leq i, k \leq p$;

c) $F_i(x) = 1 - F_i(-x)$, $1 \leq i \leq p$;

d) F_i are continuous;

e) $P(sgn\, X_{j1} = v_1, \ldots, sgn\, X_{jp} = v_p) =$
$= P(sgn\, X_{j1} = -v_1, \ldots, sgn\, X_{jp} = -v_p)$,
$1 \leq j \leq N$;

where F_{jik} and F_i, $1 \leq j \leq N$, $1 \leq i \leq p$, are the distribution functions of (X_{ji}, X_{jk}) and X_{ji}, respectively.

These conditions are fulfilled when $\mathbf{X}_1, \ldots, \mathbf{X}_N$ satisfy the multivariate hypothesis of symmetry (definition see [3]).

Let us denote by $\mathbb{D}_c = (d_{ik})_{i,k=1,\dots,p}$ the diagonal matrix with

$$d_{ii} = \left(\sum_{j=1}^{N} c_{ji}^2 \int_0^1 \varphi_i^2(u)\,du\right)^{-1/2} .$$

Further we shall suppose the covariance matrix Σ_c of S_c under (1) satisfies:

(2) If $\{\mathbb{D}_{c_\nu} \Sigma_{c_\nu} \mathbb{D}_{c_\nu}\}_{\nu=1}^{\infty}$ has a limit Σ for (4) given below with $c_{ji} = c_{ji\nu}$ then Σ is regular.

On the asymptotic distribution of Q_c under (1) we can state:

Theorem 1. Let (1),(2) and

$$(3) \qquad \int_0^1 (a_{Ni}(1 + [uN]) - \varphi_i(u))^2 du \longrightarrow 0 , \quad 1 \leqq i \leqq p ,$$

where φ_i is squared integrable and $[uN]$ is the largest integer not exceeding uN . Then the statistics Q_c are for

$$(4) \qquad \frac{\max\limits_{1 \leqq j \leqq N} c_{ji}^2}{\sum\limits_{j=1}^{N} c_{ji}^2} \longrightarrow 0 , \quad 1 \leqq i \leqq p ,$$

asymptotically χ^2-distributed with p-degrees of freedom.

Now we turn to another case. Under (6) given below the following conditions ensure that $X_1, \dots, X_N$ "nearly" satisfy the hypothesis of symmetry:

(5)
- a) $X_1, \dots, X_N$ are independent;
- b) X_{ji} has a density $f_i(x, \theta_{ji})$ where θ_{ji} is an unknown parameter;
- c) $f_i(x, \theta)$ is absolutely continuous at

θ for almost all x, $1 \leq i \leq p$;

d) $\lim_{\theta \to 0} \int \frac{\dot{f}_i(x,\theta))^2}{f_i(x,\theta)} dx = \int \frac{(\dot{f}_i(x,0))^2}{f_i(x,0)} dx = I(f_i)$,

where $\dot{f}_i(x,\theta) = \frac{\partial f_i(x,\theta)}{\partial \theta}$, $1 \leq i \leq p$;

e) $\lim_{\theta \to 0} \frac{1}{\theta}(f_i(x,\theta) - f_i(x,0)) = \dot{f}_i(x,0)$ for almost all x,

f) $f_i(x,0)$ are symmetric about 0, $1 \leq i \leq p$;

g) $F_{ik}(x, y, \theta_{ji} \; \theta_{jk})$ is continuous at $\theta_{ji} = \theta_{jk} = 0$ for all x, y, $1 \leq i, k \leq p$, with $F_{ik}(x, y, \theta_{ji} \; \theta_{jk})$ being the distribution function of (X_{ji}, X_{jk}) respectively.

Under (5) it can be stated about Q_c ($F_i(x, \theta_{ji})$ denotes the distribution function of X_{ji}):

Theorem 2. Let (5), (3) with φ_i, $1 \leq i \leq p$ being squared integrable, (2) with Σ_c being the covariance matrix of S_c under (1) with $\theta_{ji} = 0$, $1 \leq i \leq p$, $1 \leq j \leq N$, and

(6) $\max_{1 \leq j \leq N} \theta_{ji}^2 \longrightarrow 0$, $\sum_{j=1}^{N} \theta_{ji}^2 I(f_i) \leq b^2$, $0 < b^2 < +\infty$, $1 \leq i \leq p$,

hold. Then for (4) it holds

$$\sup_x |P(Q_c < x) - G_p(x; \mu_{\theta c}' \Sigma_c \mu_{\theta c})| \to 0 ,$$

where the components of $\mu_{\theta c} = (\mu_{\theta c1}, \dots, \mu_{\theta cp})'$ are given by

$$\mu_{\theta ci} = \sum_{j=1}^{N} \theta_{ji} c_{ji} \int sgn\, x \; \varphi_i (F_i(|x|, 0) - F_i(-|x|, 0)) \dot{f}_i(x,0) dx$$

and where $G_p(x, \delta)$ is the distribution function of noncentral χ^2-distribution with p degrees of freedom and noncentrality parameter δ.

At the end the case of a general alternative will be considered. We shall suppose that $X_1, \dots, X_N$ satisfy only the following:

(7) a) $X_1, \dots, X_N$ are independent;

b) the distribution function of X_{ji} is continuous.

Let us denote by Σ_c or $\Sigma_c^o = (G^o_{ikc})_{i,k=1,\dots,p}$ the covariance matrix under (7) or the expectation of Σ_{pc} under (7) respectively. Here we shall need also the following notation

$$D_c^o = (d^o_{ik})_{i,k=1,\dots,p} \quad ,$$

$$d^o_{ik} = \begin{cases} d_{ii} & \text{if } \varphi_i \text{ satisfies (12), } i = k \text{ ,} \\ var^o\, S_{ic}) & \text{if } \varphi_i \text{ satisfies (13) but not (12),} \\ 0 & \text{if } i \neq k \text{ ,} \end{cases}$$

where var^o denotes var under (7).

Further we shall suppose that Σ_c^o satisfies:

(8) If there exists a matrix $\Sigma = (G_{ik})_{i,k=1,\dots,p}$ with the property, for every $\varepsilon > 0$ and $\eta > 0$ there exist an $N_{\varepsilon\eta}$ and $\delta_\varepsilon > 0$ such that the conditions

(9) $N > N_{\varepsilon\eta}$, $var\, S_{ic} > N\eta \max\limits_{1\leq j\leq N} c^2_{ji}$ if φ_i satisfies (13) but not (12) ,

$$\left.\begin{array}{l} \left\{ \begin{array}{l} \operatorname{var}^{\circ} S_{ic} > \delta_{\varepsilon}^{-1} \max_{1 \leq j \leq N} c_{ji}^{2} \quad \text{if } \varphi_i \\ \text{satisfies (12)} \end{array} \right. \\ \text{entail} \\ \qquad |d_{ii}^{\circ} d_{kk}^{\circ} \sigma_{ikc}^{\circ} - \sigma_{ik}| < \varepsilon , \\ \text{then } \Sigma \text{ is regular.} \end{array}\right.$$

The condition (7) is weaker than (1) and (5). On the other side we restrict ourselves to scores of the form either

(10) $a_{Ni}(j) = E\,\varphi_i(U_N^{(i)}),\ 1 \leq j \leq N,\ 1 \leq i \leq p$,

or

(11) $a_{Ni}(j) = \varphi_i\left(\frac{i}{N+1}\right),\ 1 \leq j \leq N,\ 1 \leq i \leq p$,

with $U_N^{(i)}$ denoting the i-th order statistics in a sample of size N from the uniform distribution on and with φ_i defined on $(0, 1)$ that either

(12) has a bounded second derivative on $(0, 1)$

or

(13) has a form $\varphi_i = \varphi_{1i} - \varphi_{2i}$, where φ_{ki} is nondecreasing square integrable and absolutely continuous inside $(0, 1)$.

<u>Theorem 3</u>. Let (7) and (8) be satisfied, let the scores be given by (10) or (11) and φ_i, $1 \leq i \leq p$, defined on $(0, 1)$, satisfy the condition (12) or (13). Then for every $\varepsilon > 0$ and $\eta > 0$ there exist an $N_{\varepsilon\eta}$ and a $\delta_\varepsilon > 0$ such that (9) entails

$$\sup_x |P(Q_c < x) - P(U_c' \Sigma_c^- U_c < x)| < \varepsilon ,$$

where $U_c = (U_{1c}, \dots, U_{pc})'$ has the normal distribution $(E\, S_c, \Sigma_c^{\circ})$.

R e f e r e n c e s

[1] J.N. ADICHIE: Asymptotic efficiency of a class of nonparametric tests for regression parameters, Ann.Math.Stat.37(1967),884-893.

[2] K.M. PATEL: Randomisation tests for multivariate two-sample and paired-sample problems and the problem of independence, thesis (unpublished).

[3] M.L. PURI and P.K. SEN: On the theory of rank order tests for location in the multivariate one sample problem, Ann.Math.Stat.38(1967), 1216-1228.

[4] R.C. RAO: Linear statistical inference and its application, New York 1965.

[5] J. HÁJEK and Z. ŠIDÁK: Theory of rank tests, Prague 1967.

Matematicko-fyzikální fakulta
Karlova universita
Sokolovská 83, Praha 8
Československo

(Oblatum 27.11.1970)