Marie Hušková

Simultaneous rank test procedures

Aplikace matematiky, Vol. 25 (1980), No. 1, 33–38

Persistent URL: http://dml.cz/dmlcz/103835

SIMULTANEOUS RANK TEST PROCEDURES

MARIE HUŠKOVÁ

(Received February 15, 1978)

1. INTRODUCTION

Let $\boldsymbol{X}_j = (X_{1j}, \ldots, X_{pj})'$, $j = 1, \ldots, N$, be independent p-dimensional random variables with continuous distribution functions. Consider the hypotheses of randomness associated with some marginal distributions:

$$H_\nu : F_j^\nu(\boldsymbol{x}^\nu) = F^\nu(\boldsymbol{x}^\nu)\,, \quad j = 1, \ldots, N\,, \quad \nu = 1, \ldots, r\,,$$

where $F_j^\nu(\boldsymbol{x}^\nu)$ is the marginal distribution of the subvector $\boldsymbol{X}^\nu$, $\nu = 1, \ldots, r$, $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^r$ is a partition of the vector $\boldsymbol{x}$, i.e., $\boldsymbol{x} = (\boldsymbol{x}^{1\prime}, \ldots, \boldsymbol{x}^{r\prime})'$. We are interested in testing hypotheses $H_1, \ldots, H_r$ and $H_0 = \bigcap_{\nu=1}^{r} H_\nu$ against alternatives $A_1, \ldots, A_r$ and $A_0 = \bigcup_{\nu=1}^{r} A_\nu$, resp., where $A_\nu : F_j^\nu(\boldsymbol{x}^\nu) = F^\nu(\boldsymbol{x}^\nu; \theta_j^\nu)$, $j = 1, \ldots, N$, $\nu = 1, \ldots, r$, with $\theta_j = (\theta_j^{1\prime}, \ldots, \theta_j^{r\prime})'$ being a vector of unknown parameters.

Krishnaiah and some others (see [5]–[8]) developed several simultaneous test procedures for the classical multivariate normal theory. As for simultaneous rank test procedures, Krishnaiah and Sen [9] dealt with this problem for some MANOCOVA models, Jensen [3] for multivariate random blocks, Hušková [2] suggested a method for the problem considered in the present paper (see method I below).

Here we give three test procedures analogous to those proposed by Krishnaiah in [5–6] and based on the asymptotic distributions of quadratic rank statistics (for definition see (3) below).

Put

$$\boldsymbol{S}_c = (S_{c1}, \ldots, S_{cp})'\,, \tag{1}$$

$$S_{ci} = \sum_{j=1}^{N} (c_{ij} - \bar{c}_i)\, a_{Ni}(R_{ij})\,, \quad i = 1, \ldots, p\,, \tag{2}$$

with R_{ij} being the rank of X_{ij} in the sequence $X_{i1}, \ldots, X_{iN}$, c_{ij} regression constants, $a_{Ni}(j)$ scores and $\bar{c}_i = N^{-1} \sum_{j=1}^{N} c_{ij}$. Denote by $\boldsymbol{S}_c^\nu$ the subvector of $\boldsymbol{S}_c$ corresponding to $\boldsymbol{X}^\nu$, $\nu = 1, \ldots, r$. Define

$$Q_c = \boldsymbol{S}_c'(\operatorname{var}_p \boldsymbol{S}_c)^{-1}\, \boldsymbol{S}_c\,, \tag{3}$$

$$Q_c^\nu = \boldsymbol{S}_c^{\nu\prime}(\operatorname{var}_p \boldsymbol{S}_c^\nu)^{-1}\, \boldsymbol{S}_c^\nu\,, \quad \nu = 1, \ldots, r\,, \tag{4}$$

where the matrix $\text{var}_p\, \mathbf{S}_c$ is regular with elements

$$(N-1)^{-1}\sum_{j=1}^{N}(c_{ij}-\bar{c}_i)(c_{tj}-\bar{c}_t)\sum_{m=1}^{N}(a_{Ni}(R_{im})-\bar{a}_{Ni})(a_{Nt}(R_{tm})-\bar{a}_{Nt})$$

if

$$i, t \in \mathbf{I}_k, \quad k = 1, \ldots, r,$$

and

$$\sum_{j=1}^{N}(c_{ij}-\bar{c}_i)(c_{tj}-\bar{c}_t)(a_{Ni}(R_{ij})-\bar{a}_{Ni})(a_{Nt}(R_{tj})-\bar{a}_{Nt})$$

if

$$i \in \mathbf{I}_k, \quad t \notin \mathbf{I}_k, \quad k = 1, \ldots, r,$$

where $\mathbf{I}_1, \ldots, \mathbf{I}_r$ is the partition of the set $\mathbf{I} = \{1, \ldots, p\}$ considered in hypotheses H_v and $\text{var}_p\, \mathbf{S}_c^v$ is the submatrix of $\text{var}_p\, \mathbf{S}_c$ corresponding to $\mathbf{S}_c^v$ and $\bar{a}_{Ni} = N^{-1}\sum_{j=1}^{N} a_{Ni}(j)$. Denote by m_v the number of components of $\mathbf{x}^v$, $v = 1, \ldots, r$.

We shall impose usual conditions on scores, regression constants and the matrix $\text{var}_p\, \mathbf{S}_c$:

a. The scores $a_{Ni}(j)$ are generated by a nonconstant square integrable functions φ_i, $i = 1, \ldots, p$, i.e.,

$$\int_0^1 (\varphi_i(u) - a_{Ni}([uN]+1))^2\, \mathrm{d}u \to 0 \quad \text{for} \quad N \to \infty,\ i = 1, \ldots, p\,.$$

b. The regression constants fulfil:

$$\max_{1 \leq j \leq N}(c_{ij}-\bar{c}_i)^2 \left(\sum_{j=1}^{N}(c_{ij}-\bar{c}_i)^2\right)^{-1} \to 0\,, \quad i = 1, \ldots, p\,. \tag{5}$$

c. The matrices $\text{var}_p\, \mathbf{S}_c$ are regular and any accumulation point of the set $\{E\, \text{var}_p\, \mathbf{S}_c;\ c_{ij}\text{'s satisfy (5)}\}$ is a regular matrix.

In the sequel we shall often use the following results:

A. Under hypothesis H_0 and assumptions a, b, c the asymptotic distribution of $\mathbf{S}_c$ is multivariate normal $\mathfrak{N}(\mathbf{0}, \text{var}\, \mathbf{S}_c)$, where $\text{var}\, \mathbf{S}_c$ is the variance matrix of $\mathbf{S}_c$ under hypothesis H_0 (see [2]).

B. Under hypothesis H_0 and assumptions a, b, c the asymptotic distributions of Q_c and $Q_c^1, \ldots, Q_c^r$ are χ^2 with p and $m_1, \ldots, m_r$ degrees of freedom, resp. (see [2]).

C. Under hypothesis H_0 and assumptions a, b, c the matrix $\mathbf{S}_c\mathbf{S}_c'$ has asymptotically central Wishart distribution with 1 degree of freedom and positive definite matrix $\text{var}\, \mathbf{S}_c$ (it follows from A).

D. Under hypothesis H_0 and assumptions a, b, c the joint asymptotic distribution of $Q_c^1, \ldots, Q_c^r$ is the generalized multivariate χ^2-distribution defined by Jensen in [4], where the corresponding density is derived (it follows from C and [4]).

E. For an arbitrary subvector $\mathbf{S}_c^*$ of $\mathbf{S}_c$ the relation

$$\mathbf{S}_c^{*\prime}(\text{var}_p\, \mathbf{S}_c^*)^{-1}\, \mathbf{S}_c^* = \max_{u \neq 0} \frac{(\mathbf{u}'\mathbf{S}_c^*)^2}{\mathbf{u}'\, \text{var}_p\, \mathbf{S}_c^*\, \mathbf{u}}\,,$$

where $\boldsymbol{u}$ are nonzero real vectors, holds and thus

$$\mathbf{S}_c^{*\prime}(\operatorname{var}_p \mathbf{S}_c^*)^{-1} \mathbf{S}_c^{*\prime} \leqq Q_c$$

(as follows by Schwarz inequality).

F. Bonferroni inequality: For arbitrary events $A_1, \ldots, A_r$ the inequality

$$\mathrm{P}(\bigcap_{i=1}^{r} A_i) \geqq 1 - \sum_{i=1}^{r} (1 - \mathrm{P}(A_i))$$

is true.

G. Let a random p-vector $\mathbf{Y} = (Y_1, \ldots, Y_p)' = (\mathbf{Y}^{1\prime}, \ldots, \mathbf{Y}^{r\prime})'$ have the normal distribution $\mathfrak{N}(\mathbf{0}, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \Sigma_{11}, \ldots, \Sigma_{1r} \\ \vdots \qquad \vdots \\ \Sigma_{r1}, \ldots, \Sigma_{rr} \end{pmatrix}.$$

Assume that there exist vectors $\boldsymbol{b}_i$ with m_i components, $i = 1, \ldots, r$, $\sum_{l=1}^{r} m_i = p$, such that

$$\Sigma_{ij} = \boldsymbol{b}_i \boldsymbol{b}_j' , \quad i \neq j,\ i, j = 1, \ldots, r , \tag{6}$$

$$\Sigma_{ii} - \boldsymbol{b}_i \boldsymbol{b}_i' \geqq \mathbf{0} , \quad i = 1, \ldots, r , \tag{7}$$

then for arbitrary convex sets $C_1, \ldots, C_r$ symmetric about origin, $C_i \subset R_{m_i}$, the inequality

$$\mathrm{P}(\mathbf{Y}^i \in C_i,\ i = 1, \ldots, r) \geqq \prod_{i=1}^{r} \mathrm{P}(\mathbf{Y}^i \in C_i)$$

holds (see [1]).

The inequality always holds for $m_i = 1$, $i = 1, \ldots, r$ (see [10]).

2. TEST PROCEDURES

Procedure I. The author [2] proposed the test procedure with critical regions

$$Q_c > \chi^2_\alpha(p) , \tag{8}$$

where $\chi^2_\alpha(p)$ is $100\alpha\%$ critical value of the central χ^2-distribution with p degrees of freedom. This test can be used for a class of hypotheses that contain H_0 as a subhypothesis, e.g. for hypothesis that all $\mathbf{X}_j$, $j = 1, \ldots, N$, have the same distributions.

Procedure II. We base the test procedure on the statistics $Q_c^1, \ldots, Q_c^r$ given by (4). We reject the hypothesis H_v if

$$Q_c^v > d_v ,$$

where the d_v's are chosen so that

$$\lim_c \mathrm{P}(Q_c^v < d_v,\ v = 1, \ldots, r) = 1 - \alpha .$$

The total hypothesis H_0 is rejected if at least one of the hypotheses $H_1, \ldots, H_r$ is rejected. The optimal choice of the d_ν's is not known. Consistently with the classical normal case the values $d_1, \ldots, d_r$ are chosen either to be equal (i.e. $d_1 = \ldots = d_r = d$) or the individual critical regions are of equal sizes (denote them by $d_1^*, \ldots, d_r^*$). When $m_\nu = m$, $\nu = 1, \ldots, r$ then $d_\nu^* = d_\nu$, $\nu = 1, \ldots, r$. To find $d, d_1^*, \ldots, d_r^*$ with the requested properties is also very difficult for the asymptotic distribution of $(Q_c^1, \ldots, Q_c^r)$ includes numerous parameters. This problem was discussed by Jensen in [4] where some approximations are suggested.

We shall suggest here three approximations of $d, d_1^*, \ldots, d_r^*$. First consider the approximation of d. Using Bonferroni inequality we get an approximative value $\chi^2_{\alpha/r}(\max_{1 \leqq i \leqq r} m_i)$ and the critical region for testing H_ν against A_ν

$$Q_c^\nu > \chi^2_{\alpha/r}(\max_{1 \leqq i \leqq r} m_i) . \tag{9}$$

When the assumptions in G are satisfied then the critical region is

$$Q_c^\nu > \chi^2_{1-(1-\alpha)^{1/r}}(\max_{1 \leqq i \leqq r} m_i) . \tag{10}$$

Utilizing assertion E we get the third possible approximation of d. Then we reject the hypothesis H_ν if

$$Q_c^\nu > \chi^2_\alpha(p) . \tag{11}$$

Similarly we obtain the approximations of $d_1^*, \ldots, d_r^*$. By Bonferroni inequality and by G (if possible) we have the critical regions for testing H_ν against A_ν

$$Q_c^\nu > \chi^2_{\alpha/r}(m_\nu) \tag{12}$$

and

$$Q_c^\nu > \chi^2_{1-(1-\alpha)^{1/r}}(m_\nu) , \tag{13}$$

respectively.

If $m_i = 1$, $i = 1, \ldots, p$, the test procedure can be based on the statistics $S_{c1}, \ldots, S_{cp}$. Similarly, as in the general case we get critical regions

$$|S_{ci}| > \left(\sum_{j=1}^{N} (c_{ij} - \bar{c}_i)^2 (N-1)^{-1} \sum_{v=1}^{N} (a_{Ni}(v) - \bar{a}_{Ni})^2\right)^{1/2} u\left(1 - \frac{\alpha}{2p}\right) , \tag{14}$$

$$|S_{ci}| > \left(\sum_{j=1}^{N} (c_{ij} - \bar{c}_i)^2 (N-1)^{-1} \sum_{v=1}^{N} (a_{Ni}(v) - \bar{a}_{Ni})\right)^{1/2} u\left(\tfrac{1}{2} + \tfrac{1}{2}(1-\alpha)^{1/p}\right) , \tag{15}$$

$$|S_{ci}| > \left(\sum_{j=1}^{N} (c_{ji} - \bar{c}_i)^2 (N-1)^{-1} \sum_{v=1}^{N} (a_{Ni}(v) - \bar{a}_{Ni})^2\right)^{1/2} (\chi^2_\alpha(p))^{1/2} , \tag{16}$$

where $u(\cdot)$ is the $100\alpha\%$ quantile of the normal distribution (0, 1).

As for the comparison of the critical regions (9–10), (12–13), we can easily get the following relations among the approximations of $d_1, \ldots, d_r$

$$\chi^2_{1-(1-\alpha)^{1/r}}(\max_{1 \leqq i \leqq r} m_i) \geqq \chi^2_{1-(1-\alpha)^{1/r}}(m_\nu) ,$$

$$\chi^2_{\alpha/r}(\max_{1 \leqq i \leqq r} m_i) \geqq \chi^2_{\alpha/r}(m_\nu) \geqq \chi^2_{1-(1-\alpha)^{1/r}}(m_\nu) , \quad \nu = 1, \ldots, r .$$

Thus the critical region (13) is larger then (9), (10) and (12).The comparison of (11) with the other critical regions is more complicated, e.g.

$$\text{if} \quad \alpha \leqq 0{\cdot}05\,, \quad p - \max_{1 \leqq i \leqq r} m_i \geqq 5 \quad \text{then} \quad \chi^2_\alpha(p) > \chi^2_{\alpha/r}(\max_{1 \leqq i \leqq r} m_i)\,,$$

$$\text{if} \quad \alpha = 0{\cdot}05\,, \quad p = 22\,, \quad \max_{1 \leqq i \leqq r} m_i \leqq p - 2 \quad \text{then} \quad \chi^2_{0,05}(p) < \chi^2_{1-(0,95)^{1/r}}(\max_{1 \leqq i \leqq r} m_i)\,.$$

When $m_i = 1$ then the largest critical region is (15).

Procedure III. Define

$$Q^*_{cv} = \mathbf{S}^{*\prime}_{cv}(\operatorname{var}_p \mathbf{S}^*_{cv})^{-1}\, \mathbf{S}^*_{cv}\,, \quad v = 1, \ldots, r\,,$$

where

$$\mathbf{S}^*_{c1} = \mathbf{S}^1_c\,,$$

$$\mathbf{S}^*_{cv+1} = \mathbf{S}^{v+1}_c - \operatorname{cov}_p(\mathbf{S}^{v+1}_c; \mathbf{S}^1_c, \ldots, \mathbf{S}^v_c)\,(\operatorname{var}_p(\mathbf{S}^{1\prime}_c, \ldots, \mathbf{S}^{v\prime}_c)')^{-1}\,.$$
$$.\,(\mathbf{S}^{1\prime}_c, \ldots, \mathbf{S}^{v\prime}_c)'\,, \quad v = 1, \ldots, r-1\,,$$

$$\operatorname{cov}(\mathbf{S}^{v+1}_c; \mathbf{S}^1_c, \ldots, \mathbf{S}^v_c) = (\operatorname{cov}_p(\mathbf{S}^{v+1}_c, \mathbf{S}^1_c), \ldots, \operatorname{cov}_p(\mathbf{S}^{v+1}_c; \mathbf{S}^v_c))\,.$$

$$.\ \operatorname{var}_p \mathbf{S}^*_{cv+1} = \operatorname{var}_p \mathbf{S}^{v+1}_c - \operatorname{cov}_p(\mathbf{S}^{v+1}_c; \mathbf{S}^1_c, \ldots, \mathbf{S}^v_c;)\,(\operatorname{var}_p(\mathbf{S}^{1\prime}_c, \ldots, \mathbf{S}^{v\prime}_c)')^{-1}\,.$$
$$.\,(\operatorname{cov}_p(\mathbf{S}^{v+1}_c; \mathbf{S}^1_c, \ldots, \mathbf{S}^v_c))'\,,$$

with $\operatorname{var}_p(\ldots)$ and $\operatorname{cov}_p(\ldots)$ denoting the corresponding submatrices of $\operatorname{var}_p \mathbf{S}_c$.

The assertion A implies that the asymptotic distribution of $\mathbf{S}_c$ (under hypothesis H and assumptions a, b, c) is multivariate normal with mean $\mathbf{0}$ and the variance matrix

$$\operatorname{var} \mathbf{S}^*_{cv+1} = \operatorname{var} \mathbf{S}^{v+1}_c - \operatorname{cov}(\mathbf{S}^{v+1}_c; \mathbf{S}^1_c, \ldots, \mathbf{S}^v_c)\,(\operatorname{var}(\mathbf{S}^{1\prime}_c, \ldots, \mathbf{S}^{v\prime}_c)')^{-1}\,.$$
$$.\,(\operatorname{cov}(\mathbf{S}^{v+1}_c; \mathbf{S}^1_c, \ldots, \mathbf{S}^v_c))'$$

and Q^*_{cv} has asymptotically χ^2-distribution with m_i degrees of freedom. By direct computations we get that $\mathbf{S}^*_{c1}, \ldots, \mathbf{S}^*_{cr}$ are asymptotically independent and thus so are $Q^*_{c1}, \ldots, Q^*_{cr}$.

Using these arguments one can assert that

$$\lim_c \mathrm{P}(Q^*_{cr} < \chi^2_{1-(1-\alpha)^{1/r}}(\max_{1 \leqq i \leqq r} m_i)\,, \quad v = 1, \ldots, r) \geqq$$
$$\geqq \lim_c \mathrm{P}(Q^*_{cv} < \chi^2_{1-(1-\alpha)^{1/r}}(m_v)\,, \quad v = 1, \ldots, r) = 1 - \alpha\,.$$

Thus the critical region for testing the hypothesis H_v against A_v can be chosen in either of the following ways:

$$Q^*_c > \chi^2_{1-(1-\alpha)^{1/r}}(\max_{1 \leqq i \leqq r} m_i)\,, \tag{17}$$

$$Q^*_c > \chi^2_{1-(1-\alpha)^{1/r}}(m_v)\,. \tag{18}$$

Obviously, the critical region (18) contains (17).

We reject the hypothesis H_0 if we reject at least one of $H_1, \ldots, H_r$.

If $m_i = 1$, $i = 1, \ldots, p$ the test procedure can be based on the statistics S^*_{cv}, $v = 1, \ldots, p$. We reject the hypothesis H_v if

$$|S^*_{cv}| > (\text{var } S^*_{cv})^{1/2}\, u\left(\tfrac{1}{2} + \tfrac{1}{2}(1 - \alpha)^{1/p}\right).$$

References

[1] *Gupta, D. S.:* On a probability inequality for multivariate normal distribution, Aplikace Matematiky 21 (1976), 1—4.

[2] *Hušková, M.:* Multivariate rank statistics for testing randomness concerning some marginal distributions, J. Multivariate Anal. 5 (1975), 487—496.

[3] *Jensen, D. R.:* The joint distribution of Friedman's χ_r^2-statistics, Ann. Statist. 2 (1974), 311—323.

[4] *Jensen, D. R.:* The joint distribution of traces of Wishart matrices and some applications, Ann. Math. Statist. 41 (1970), 133—145.

[5] *Krishnaiah, P. R.:* On the simultaneous ANOVA and MANOVA tests, Ann. Inst. Statist. Math. 17 (1965), 35—53.

[6] *Krishnaiah, P. R.:* Simultaneous test procedures under general MANOVA models. In Multivariate Analysis — II (P. R. Krishnaiah, Ed.) pp. 121—143, Academic Press, New York (1969).

[7] *Roy, J.:* Step-down procedure in multivariate analysis, Ann. Math. Statist. 29 (1958), 1177—1187.

[8] *Roy, S. N.,* and *Gnanadesikan, R.:* Further contributions to multivariate confidence bounds, Biometrika 45 (1957), 581.

[9] *Sen, P. K.* and *Krishnaiah, P. K.:* On a class of simultaneous rank order tests in MANOCOVA, Ann. Inst. Statist. Math. 26 (1974), 135—145.

[10] *Šidák, Z.:* Rectangular confidence regions for the means of multivariate normal distributions, J. Amer. Stat. Assoc. 62 (1967), 626—633.

Souhrn

MARIE HUŠKOVÁ

SIMULTÁNNÍ PROCEDURY POŘADOVÝCH TESTŮ

Nechť X_j, $j = 1, \ldots, N$ jsou nezávislé p-rozměrné náhodné vektory se spojitou distribuční funkcí F_j. V článku jsou navržena tři testová kritéria založená na pořadích pro test nezávislosti marginálních rozdělení X_j na indexu j. Výchozím bodem pro konstrukci testových kritérii byl článek P. R. Krishnaiaha (Ann. Inst. Statist. Math. 17, 35—53, 1965).

Author's address: RNDr. *Marie Hušková*, CSc. Matematicko-fyzikální fakulta Karlovy univerzity, Sokolovská 83, 186 00 Praha 8.