

L. CIESIELSKA and Teresa LEDWINA (Wrocław)

ON LOCALLY MOST POWERFUL RANK TESTS OF INDEPENDENCE

1. Introduction. The result of Hájek and Šidák [2] on locally most powerful rank tests (LMPRT's) of independence of two random variables is extended to the case of n random variables. The precise statement of the problem is given in Section 2. Note that Shirahata [4] has considered a more general problem of testing and derived an LMPRT for it. However, it used slightly stronger regularity conditions than those required in this paper. We present these additional regularity conditions in the case of two models of dependence: one described in Section 2 and the other which is a straightforward generalization of the bivariate dependence model considered in [2]. It is interesting to observe that the same test is LMPRT when the alternatives are described by one of the above models. The known test of independence of n random variables, based on Friedman's statistics, is a special case of tests derived in this paper.

It should also be said that the authors did not know the paper of Shirahata [4] while working on this problem.

2. Preliminaries and formulation of the problem. Consider the problem of testing the independence of n random variables X_1, \dots, X_n . Suppose that a random sample of N n -dimensional vectors $(X_{1i}, X_{2i}, \dots, X_{ni})$, $i = 1, \dots, N$, was drawn. We can present data in the form of a two-way layout of N rows and n columns. Let the vectors X_1, X_2, \dots, X_n stand for the columns of the layout. Moreover, let R_{ij} stand for the rank of the i -th component of X_j and denote by R_j the vector $(R_{j1}, R_{j2}, \dots, R_{jN})$.

Consider the following model of dependence of X_1, \dots, X_n . Assume that there exist i, k ($1 \leq i < k \leq n$) such that

$$X_i = X_i^* + \Delta Z, \quad X_k = X_k^* + \Delta Z,$$

while

$$X_j = X_j^* + \Delta Z_j, \quad j \neq i, k,$$

where $X_1^*, \dots, X_n^*, Z, \{Z_j\}_{j=1}^n, j \neq i, k$, are mutually independent. Moreover, it is assumed that the random variables X_1^*, \dots, X_n^* have densities f_1, \dots, f_n with respect to the Lebesgue measure, while $Z, \{Z_j\}_{j=1}^n, j \neq i, k$, are arbitrary identically distributed random variables with distribution function $M(z)$, such that $0 < \text{Var } Z < \infty$. Therefore, the density of the joint distribution of X_1, \dots, X_n , say q_Δ , is of the form

$$(1) \quad q_\Delta(x_1, \dots, x_n) \\ = \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} \left[\int_R f_i(x_i - \Delta z) f_k(x_k - \Delta z) dM(z) \right] \prod_{\substack{j=1 \\ j \neq i, k}}^n \int_R f_j(x_j - \Delta z) dM(z),$$

where $0 \leq c_{ik} \leq 1$ are fixed and

$$\sum_{i=1}^n \sum_{k=i+1}^n c_{ik} = 1.$$

Denote by q_0 the density $\prod_{i=1}^n f_i(x_i)$. In the next section we obtain the locally most powerful test for testing $H: q_0$ against $K: q_\Delta, \Delta > 0$, in the class of rank tests, i.e., tests based on $\mathbf{R}_1, \dots, \mathbf{R}_n$. For the sake of completeness let us recall that a test is called *locally most powerful* for H against K at some level α if it is uniformly most powerful at level α for H against $K_\varepsilon = \{q_\Delta, 0 < \Delta < \varepsilon\}$ for some $\varepsilon > 0$.

Before formulating the main result we state the following

THEOREM 1. *The most powerful α -level test of H against some simple alternative q_Δ is given by*

$$(2) \quad \psi(\mathbf{r}_1, \dots, \mathbf{r}_n) = \begin{cases} 1 & \text{if } Q_\Delta(\mathbf{R}_1 = \mathbf{r}_1, \dots, \mathbf{R}_n = \mathbf{r}_n) > C, \\ \gamma & \text{if } Q_\Delta(\mathbf{R}_1 = \mathbf{r}_1, \dots, \mathbf{R}_n = \mathbf{r}_n) = C, \\ 0 & \text{if } Q_\Delta(\mathbf{R}_1 = \mathbf{r}_1, \dots, \mathbf{R}_n = \mathbf{r}_n) < C, \end{cases}$$

where

$$Q_\Delta(\mathbf{R}_1 = \mathbf{r}_1, \dots, \mathbf{R}_n = \mathbf{r}_n) = \int \dots \int \prod_{i=1}^N q_\Delta(x_{1i}, \dots, x_{ni}) dx_1 \dots dx_n, \\ B_{\mathbf{r}_1, \dots, \mathbf{r}_n} = \{\mathbf{R}_1 = \mathbf{r}_1, \dots, \mathbf{R}_n = \mathbf{r}_n\},$$

while C and γ are determined by $\mathbb{E}_H \psi(\mathbf{R}_1, \dots, \mathbf{R}_n) = \alpha$.

The proof of Theorem 1 can be easily obtained by repeating the reasoning of Hájek and Šidák (see [2], p. 52).

Moreover, the following lemma is used in the next section (see [2], p. 76):

LEMMA 1. *Assume that $f_j'(x)$ exist, are continuous almost everywhere, and*

$$\int_R |f_j'(x)| dx < \infty, \quad j = 1, \dots, n.$$

Then for arbitrary i, k ($1 \leq i < k \leq n$) we have

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \Delta^{-2} \left[\int_{\mathbb{R}} f_i(x_i - \Delta z) f_k(x_k - \Delta z) dM(z) - \right. \\ \left. - \int_{\mathbb{R}} f_i(x_i - \Delta z) dM(z) \int_{\mathbb{R}} f_k(x_k - \Delta z) dM(z) \right] \\ = (\text{Var} Z) f'_i(x_i) f'_k(x_k). \end{aligned}$$

3. Locally most powerful rank tests of independence. Denote by T_N the rank statistic

$$\sum_{l=1}^N \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} a_N(R_{il}, f_i) a_N(R_{kl}, f_k),$$

where the scores $a_N(r_{st}, f_s)$ are given by

$$a_N(r_{st}, f_s) = -\mathbb{E}_H \left\{ \frac{f'_s(X_{st})}{f_s(X_{st})} \mid R_{st} = r_{st} \right\}, \quad 1 \leq s, t \leq n.$$

THEOREM 2. *The test with critical region $\{T_N \geq t\}$ is the locally most powerful test for H against K at the appropriate level.*

Proof. The proof consists in showing the equality

$$\begin{aligned} (3) \quad \lim_{\Delta \rightarrow 0} \Delta^{-2} [Q_\Delta(\mathbf{R}_1 = \mathbf{r}_1, \dots, \mathbf{R}_n = \mathbf{r}_n) - (N!)^n] \\ = \frac{\text{Var} Z}{(N!)^n} \sum_{l=1}^N \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} a_N(r_{il}, f_i) a_N(r_{kl}, f_k) \end{aligned}$$

which guarantees that for sufficiently small Δ the rejection region $\{T_N \geq t\}$ coincides with the rejection region of (2) for some C . Denote by L the left-hand side of (3). To prove (3) let us observe first that

$$\int_{B_{\mathbf{r}_1, \dots, \mathbf{r}_n}} \dots \int \prod_{i=1}^n \prod_{j=1}^n f_{j\Delta}(x_{ji}) dx_1 \dots dx_n = (N!)^{-n},$$

where

$$f_{j\Delta}(x_{ji}) = \int_{\mathbb{R}} f_j(x_{ji} - \Delta z) dM(z).$$

Using this fact and applying the formula

$$\prod_{i=1}^N a_i - \prod_{i=1}^N b_i = \sum_{l=1}^N (a_l - b_l) \prod_{s=l+1}^N b_s \prod_{s=1}^{l-1} a_s$$

for $a_i = q_\Delta(x_{1i}, \dots, x_{ni})$ and $b_i = \prod_{j=1}^n f_{j\Delta}(x_{ji})$, we obtain

$$L = \lim_{\Delta \rightarrow 0} \Delta^{-2} \int \dots \int_{B_{\mathbf{r}_1, \dots, \mathbf{r}_n}} \sum_{l=1}^N \left\{ [q_\Delta(x_{1l}, \dots, x_{nl}) - \prod_{j=1}^n f_{j\Delta}(x_{ji})] \times \right. \\ \left. \times \prod_{s=l+1}^N \prod_{p=1}^n f_{p\Delta}(x_{ps}) \prod_{s=1}^{l-1} q_\Delta(x_{1s}, \dots, x_{ns}) \right\} d\mathbf{x}_1 \dots d\mathbf{x}_n.$$

Replacing now q_Δ by (1) we get

$$(4) \quad L = \lim_{\Delta \rightarrow 0} \sum_{l=1}^N \int \dots \int_{B_{\mathbf{r}_1, \dots, \mathbf{r}_n}} \left\{ \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} \Delta^{-2} \times \right. \\ \left. \times \left[\int_{\mathcal{R}} f_i(x_{i\ell} - \Delta z) f_k(x_{k\ell} - \Delta z) dM(z) - f_{i\Delta}(x_{i\ell}) f_{k\Delta}(x_{k\ell}) \right] \times \right. \\ \left. \times \prod_{\substack{j=1 \\ j \neq i, k}}^n f_{j\Delta}(x_{j\ell}) \prod_{s=l+1}^N \prod_{p=1}^n f_{p\Delta}(x_{ps}) \prod_{s=1}^{l-1} q_\Delta(x_{1s}, \dots, x_{ns}) \right\} d\mathbf{x}_1 \dots d\mathbf{x}_n.$$

Since the calculation of the limit under the integral sign can be justified by the same reasoning as in the case $n = 2$ (cf. [2], p. 77), from (4) and Lemma 1 we obtain

$$L = (\text{Var} Z) \sum_{l=1}^N \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} \times \\ \times \int \dots \int_{B_{\mathbf{r}_1, \dots, \mathbf{r}_n}} [f'_i(x_{i\ell}) f'_k(x_{k\ell}) / f_i(x_{i\ell}) f_k(x_{k\ell})] \prod_{j=1}^n \prod_{s=1}^N f_j(x_{js}) d\mathbf{x}_1 \dots d\mathbf{x}_n \\ = (\text{Var} Z) \sum_{l=1}^N \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} \left\{ \int_{\{\mathbf{R}_i = \mathbf{r}_i\}} [f'_i(x_{i\ell}) / f_i(x_{i\ell})] \prod_{s=1}^N f_i(x_{is}) d\mathbf{x}_i \right\} \times \\ \times \left\{ \int_{\{\mathbf{R}_k = \mathbf{r}_k\}} [f'_k(x_{k\ell}) / f_k(x_{k\ell})] \prod_{s=1}^N f_k(x_{ks}) d\mathbf{x}_k \right\} \prod_{\substack{j=1 \\ j \neq i, k}}^n \int_{\{\mathbf{R}_j = \mathbf{r}_j\}} \prod_{s=1}^N f_j(x_{js}) d\mathbf{x}_j.$$

Since

$$\int_{\{\mathbf{R}_a = \mathbf{r}_a\}} [f'_a(x_{a\ell}) / f_a(x_{a\ell})] \prod_{s=1}^N f_a(x_{as}) d\mathbf{x}_a \\ = (N!)^{-1} \mathbf{E} \{ [f'_a(x_{a\ell}) / f_a(x_{a\ell})] \mid \mathbf{R}_a = \mathbf{r}_a \} \\ = -(N!)^{-1} a_N(r_{a\ell}, f_a), \quad a = 1, \dots, n$$

(see [2], p. 38), we have

$$L = (N!)^{-n} (\text{Var} Z) \sum_{l=1}^N \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} a_N(r_{i\ell}, f_i) a_N(r_{k\ell}, f_k).$$

4. Remarks. 1. The following model of dependence is a special case of the model considered by Shirahata [4]:

$$X_k = X_k^* + \Delta Y, \quad k = 1, \dots, n,$$

where X_1^*, \dots, X_n^*, Y are mutually independent. Moreover, it is assumed that X_1^*, \dots, X_n^* have densities f_1, \dots, f_n with respect to the Lebesgue measure, while $0 < \text{Var } Y < \infty$. Then the density p_Δ of X_1, \dots, X_n is given by

$$p_\Delta(x_1, \dots, x_n) = \int \prod_{i=1}^n f_i(x_i - \Delta z) dM(z),$$

where $M(y)$ is the distribution function of Y .

If $EY = 0$, then the first and the second derivatives with respect to Δ can be taken under the integral sign in p_Δ and

$$\begin{aligned} \int \cdots \int \frac{\partial^i}{\partial \Delta^i} p_\Delta(x_1, \dots, x_n) dx_1 \cdots dx_n \\ \rightarrow \int \cdots \int \frac{\partial^i}{\partial \Delta^i} p_\Delta(x_1, \dots, x_n) |_{\Delta=0} dx_1 \cdots dx_n \end{aligned}$$

holds for $i = 1, 2$ and $\Delta \rightarrow 0$. Then, by Theorem 2 of [4], one can obtain (cf. [3], p. 590) the LMPRT which coincides with the test derived in this paper for $H : q_0$ against $K : q_\Delta$, $\Delta > 0$, where $c_{ik} = \binom{n}{2}^{-1}$. It is a quite rare case that the LMPRT's for two different alternatives are the same.

Moreover, note that the test derived in this paper can be obtained after long but elementary calculations from Theorem 2 of [4] when one assumes that q_Δ satisfies the same regularity conditions as those given above for p_Δ .

Observe also that for $n = 2$ both models of dependence presented here coincide with the bivariate dependence model proposed in [2], p. 75.

2. One can see that the test with the critical region $\{T_N \geq t\}$ is equivalent to the test of independence based on Friedman's statistics (see [1], Chapter 13) when $c_{ik} = \binom{n}{2}^{-1}$ and the observed random variables have the logistic distribution because then $a_N(R_{st}, f_s) = 2R_{st}(N+1) - 1$. Note also that Friedman's test is a straightforward generalization of the well-known Spearman rank correlation test.

3. Note that the test statistic T_N is a linear combination of test statistics of LMPRT's of independence of pairs of random variables (cf [2], p. 76). Therefore, in the case of independent and identically distributed X_1, \dots, X_n the moments and the asymptotic normality of T_N for

commonly used f_i 's can be immediately obtained from appropriate results for $n = 2$ (see [2], p. 112-114 or 167). The asymptotic normality of T_N under some alternatives can be investigated by using the results of Ruymgaart and van Zuijlen [3].

References

- [1] J. D. Gibbons, *Nonparametric statistical inference*, McGraw-Hill, New York 1971.
- [2] J. Hájek and Z. Šidák, *Theory of rank tests*, Academia, Prague 1967.
- [3] F. H. Ruymgaart and M. C. A. van Zuijlen, *Asymptotic normality of multivariate linear rank statistics in the non-i.i.d. case*, Ann. Statist. 6 (1978), p. 588-602.
- [4] S. Shirahata, *Locally most powerful rank tests for independence*, Bull. Math. Statist. 16 (1974), p. 11-21.

INSTITUTE OF MATHEMATICS
WROCLAW TECHNICAL UNIVERSITY
50-370 WROCLAW

Received on 7. 12. 1979;
revised version on 15. 5. 1981
