L. CIESIELSKA and Teresa LEDWINA (Wroclaw)

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 $X_1, \ldots, X_n$. Suppose that a random sample of $N$ $n$-dimensional vectors $(X_{1i}, X_{2i}, \ldots, X_{Ni})$, $i = 1, \ldots, N$, was drawn. We can present data in the form of a two-way layout of $N$ rows and $n$ columns. Let the vectors $X_1, X_2, \ldots, X_n$ 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 $(R_{j1}, R_{j2}, \ldots, R_{jN})$.

Consider the following model of dependence of $X_1, \ldots, X_n$. 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

$$X_j = X_j^* + \Delta Z, \quad j \neq i, k,$$
where $X_1^*, \ldots, X_n^*, Z, \{Z_j\}_{j=1}^n, j \neq i, k$, are mutually independent. Moreover, it is assumed that the random variables $X_1^*, \ldots, X_n^*$ have densities $f_i, \ldots, f_n$ 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 $X_1, \ldots, X_n$, say $q_d$, is of the form

\[
q_d(x_1, \ldots, x_n) = \sum_{i=1}^n \sum_{k=i+1}^n c_{ik} \left[ \int_R f_i(x_i - \Delta x) f_k(x_k - \Delta x) dM(x) \right] \prod_{j \neq i, k} \int_R f_j(x_j - \Delta x) dM(x),
\]

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 $K : q_d, \Delta > 0$, in the class of rank tests, i.e., tests based on $R_1, \ldots, R_n$. For the sake of completeness let us recall that a test is called \textit{locally most powerful} for $H$ against $K$ at some level $\alpha$ if it is uniformly most powerful at level $\alpha$ for $H$ against $K_\varepsilon = \{q_d, 0 < \Delta < \varepsilon\}$ for some $\varepsilon > 0$.

Before formulating the main result we state the following

\textbf{Theorem 1.} The most powerful $\alpha$-level test of $H$ against some simple alternative $q_d$ is given by

\[
\psi(r_1, \ldots, r_n) = \begin{cases} 
1 & \text{if } Q_d(R_1 = r_1, \ldots, R_n = r_n) > C, \\
\gamma & \text{if } Q_d(R_1 = r_1, \ldots, R_n = r_n) = C, \\
0 & \text{if } Q_d(R_1 = r_1, \ldots, R_n = r_n) < C,
\end{cases}
\]

where

\[
Q_d(R_1 = r_1, \ldots, R_n = r_n) = \int \cdots \int \prod_{i=1}^N q_d(x_{i1}, \ldots, x_{ni}) dx_{i1} \cdots dx_{ni},
\]

\[
B_{r_1, \ldots, r_n} = \{R_1 = r_1, \ldots, R_n = r_n\},
\]

while $C$ and $\gamma$ are determined by $E_H \psi(R_1, \ldots, R_n) = \alpha$.

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):

\textbf{Lemma 1.} Assume that $f_j(x)$ exist, are continuous almost everywhere, and

\[
\int_R |f_j(x)| dx < \infty, \quad j = 1, \ldots, n.
\]
Then for arbitrary \( i, k \) \((1 \leq i < k \leq n)\) we have

\[
limit_{\Delta \to 0} \Delta^{-2} \left[ \int_{\mathcal{R}} f_i(x_i - \Delta z) f_k(x_k - \Delta z) \, dM(z) - \int_{\mathcal{R}} f_i(x_i - \Delta z) \, dM(z) \int_{\mathcal{R}} f_k(x_k - \Delta z) \, dM(z) \right] = (\text{Var}Z) f'_i(x_i) f'_k(x_k).
\]

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

\[
\sum_{i=1}^{N} \sum_{t=1}^{n} \sum_{k=t+1}^{n} c_{ik} a_N(R_{it}, f_i) a_N(R_{kt}, f_k),
\]

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

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

**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

\[
\lim_{\Delta \to 0} \Delta^{-2} [Q_\Delta(R_1 = r_1, \ldots, R_n = r_n) - (N!)^n] = \frac{\text{Var}Z}{(N!)^n} \sum_{i=1}^{N} \sum_{t=1}^{n} \sum_{k=t+1}^{n} c_{ik} a_N(r_{it}, f_i) a_N(r_{kt}, f_k)
\]

which guarantees that for sufficiently small \( \Delta \) 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

\[
\int R_{x_1} \cdots \int R_{x_n} f_{jd}(x_{jt}) \, dx_1 \ldots dx_n = (N!)^{-n},
\]

where

\[
f_{jd}(x_{jt}) = \int_{\mathcal{R}} f_j(x_{jt} - \Delta z) \, dM(z).
\]

Using this fact and applying the formula

\[
\prod_{i=1}^{N} a_i - \prod_{i=1}^{N} b_i = \sum_{i=1}^{N} (a_i - b_i) \prod_{s=i+1}^{N} b_s \prod_{s=1}^{i-1} a_s
\]
for \( a_i = q_d(x_{i1}, \ldots, x_{ni}) \) and \( b_i = \prod_{j=1}^{n} f_{jd}(x_{ji}) \), we obtain

\[
L = \lim_{\Delta \to 0} \Delta^{-2} \int_{B_{r_1}, \ldots, v_{n}} \sum_{i=1}^{N} \left\{ \sum_{l=1}^{n} \left[ q_d(x_{il}, \ldots, x_{nl}) - \prod_{j=1}^{n} f_{jd}(x_{ji}) \right] \times \prod_{s=1}^{n} \prod_{p=1}^{l-1} f_{psd}(x_{ps}) \prod_{s=1}^{l-1} q_d(x_{is}, \ldots, x_{ns}) \right\} \, dx_1 \cdots dx_n.
\]

Replacing now \( q_d \) by (1) we get

(4)

\[
L = \lim_{\Delta \to 0} \sum_{l=1}^{N} \int_{B_{r_1}, \ldots, v_{n}} \left\{ \sum_{i=1}^{n} \sum_{k=l+1}^{n} c_{ik} \Delta^{-2} \times \right.
\]

\[
\times \left[ \int_{R} f_i(x_{il} - \Delta z) f_k(x_{kl} - \Delta z) \, dM(z) - f_{id}(x_{il}) f_{kd}(x_{kl}) \right] \times \prod_{j=1}^{n} \prod_{s=1}^{n} \prod_{p=1}^{l-1} f_{psd}(x_{ps}) \prod_{s=1}^{l-1} q_d(x_{is}, \ldots, x_{ns}) \right\} \, dx_1 \cdots dx_n.
\]

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

\[
L = (\text{Var}Z) \sum_{l=1}^{N} \sum_{i=1}^{n} \sum_{k=l+1}^{n} c_{ik} \times \prod_{j=1}^{n} \prod_{s=1}^{n} f_{jd}(x_{js}) \, dx_1 \cdots dx_n
\]

\[
= (\text{Var}Z) \sum_{l=1}^{N} \sum_{i=1}^{n} \sum_{k=l+1}^{n} c_{ik} \left\{ \int_{(R_i=r_j)} \frac{f_i(x_{il})}{f_i(x_{il})} \prod_{s=1}^{n} f_{j}(x_{js}) \, dx_i \right\} \times \prod_{(R_k=r_j)} \frac{f_k(x_{kl})}{f_k(x_{kl})} \prod_{j=1}^{n} \prod_{j \neq i, k} \int_{(R_k=r_j)} f_{j}(x_{js}) \, dx_j.
\]

Since

\[
\int_{(R_a=r_a)} \frac{f_i'(x_{al})}{f_i'(x_{al})} \prod_{s=1}^{n} f_{a}(x_{as}) \, dx_a
\]

\[
= (N!)^{-1} \mathbb{E} \{ [f_i'(x_{al})/f_i'(x_{al})] \mid R_a = r_a \}
\]

\[
= -(N!)^{-1} a_N(r_a, f_a), \quad a = 1, \ldots, n
\]

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

\[
L = (N!)^{-n}(\text{Var}Z) \sum_{l=1}^{N} \sum_{i=1}^{n} \sum_{k=l+1}^{n} c_{ik} a_N(r_{il}, f_i) a_N(r_{kl}, f_k).
\]
4. Remarks. 1. The following model of dependence is a special case of the model considered by Shirahata [4]:

$$X_k = X^*_k + \Delta Y, \quad k = 1, \ldots, n,$$

where $X^*_1, \ldots, X^*_n, Y$ are mutually independent. Moreover, it is assumed that $X^*_1, \ldots, X^*_n$ have densities $f_1, \ldots, f_n$ with respect to the Lebesgue measure, while $0 < \text{Var } Y < \infty$. Then the density $p_d$ of $X_1, \ldots, X_n$ is given by

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

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

If $\mathbb{E}Y = 0$, then the first and the second derivatives with respect to $\Delta$ can be taken under the integral sign in $p_d$ and

$$\int \int_\Delta \left. \frac{\partial^i}{\partial \Delta^i} p_d(x_1, \ldots, x_n) \right|_{\Delta=0} dx_1 \ldots dx_n$$

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 $K : q_d, \Delta > 0$, 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 $q_d$ satisfies the same regularity conditions as those given above for $p_d$.

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 $a_N(R_{st}, f_d) = 2R_{st}(N+1)-1$. 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 $X_1, \ldots, X_n$ 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


INSTITUTE OF MATHEMATICS
WROCŁAW TECHNICAL UNIVERSITY
50-370 WROCŁAW

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