NONPARAMETRIC TESTS FOR INDEPENDENCE BASED ON INTRACLASS RANKS

Shirahata, Shingo
Department of Statistics, Faculty of General Education, Osaka University

https://doi.org/10.5109/13344

出版情報：Bulletin of informatics and cybernetics. 20 (3/4), pp.45-55, 1983-03. Research Association of Statistical Sciences

Bulletin of Informatics and Cybernetics Vol. 20, No. 3~4, 1983

NONPARAMETRIC TESTS FOR INDEPENDENCE BASED ON INTRACLASS RANKS

By

Shingo SHIRAHATA*

Abstract

Intraclass rank statistics are introduced to test for independence in a bivariate population when it has the same continuous marginal distributions. Locally most powerful intraclass rank tests (LMPIRT) are derived for a one-parameter family and asymptotic normality of a family of intraclass rank statistics including LMPIRT is shown under the hypothesis of independence and its contiguous alternatives. Furthermore, approximations of the null distributions of the statistics are discussed.

1. Introduction

Let (X_i, Y_i), $i=1, \cdots, n$ be a random sample from a population with continuous distribution function. In this paper intraclass rank tests of the null hypothesis H; X and Y are independent is considered under the constraint that X and Y have a common continuous marginal distribution function $F(x)$.

Let R_i and Q_i be the intraclass ranks of X_i and Y_i among overall observations, respectively, i. e.,

$$R_i = \sum_{j=1}^{n} \{u(X_i - X_j) + u(X_i - Y_j)\}$$

and

$$Q_i = \sum_{j=1}^{n} \{u(Y_i - X_j) + u(Y_i - Y_j)\}$$

where $u(x)=1$ or 0 according as $x \geqq 0$ or $x<0$. The random variables Y_i and Q_i are also denoted by X_{i+n} and R_{i+n}, respectively. In the second notations, the intraclass ranks are given by

$$R_i = \sum_{j=1}^{2n} u(X_i - X_j), \qquad i=1, \cdots, 2n .$$

We shall use both notations for simplicity. We believe no confusions will occur.

Since X and Y have the same distribution, it is preferable that the conclusion of a test does not altered if (X_i, Y_i) is replaced by (Y_i, X_i) for some i. In this sense, the tests based on the intraclass ranks are more sound than the tests based on the usual ranks.

* Department of statistics, Faculty of General Education, Osaka University, Toyonaka, Osaka 560, Japan.

In Section 2 the locally most powerful intraclass rank test (LMPIRT) for a one-parameter family is derived. As in the case of the usual ranks in Shirahata (1974), the LMPIRT has critical region of the form $\sum_{i=1}^{n} a_n(R_i, Q_i) \geqq c$ in many models. The asymptotic properties of intraclass rank statistics including LMPIRT are considered under H and its contiguous alternatives in Sections 3 and 4. Some special attentions are paid to statistics of the product form $\sum_{i=1}^{n} a_n(R_i)a_n(Q_i)$ in the same sections. Though these can not be the LMPIRT, the product form is easy to treat and is useful in the practical situations.

Usual linear rank statistics with symmetric scores constants are symmetric about the origin under H. However, our intraclass rank statistics are not symmetric exactly and symmetric asymptotically. Hence, the accuracy of the approximations by the asymptotic distribution will be not so good. In Section 5, the approximations using exact moments under H are discussed.

2. Locally Most Powerful Intraclass Rank Tests

Let us consider a one-parameter family of density functions $f(x, y; \theta)$ with the same marginal density function $f(x; \theta)$. Here θ represents the parameter such that $\theta=0$ implies that X and Y are independent. By denoting $f(x)=f(x; 0)$, $f(x, y; 0)=f(x)f(y)$. In this section, we derive the LMPIRT of the null hypothesis $\theta=0$ against the alternative $\theta>0$ or $\theta\neq 0$.

Put $\boldsymbol{R}=(R_1, \cdots, R_{2n})$ and $\boldsymbol{X}^*=(X_{(1)}, \cdots, X_{(2n)})$ where $X_{(i)}$ is the i-th smallest order statistic of $\boldsymbol{X}=(X_1, \cdots, X_{2n})$. Denote by $r(X_i)$ the intraclass rank of X_i for $i=1, \cdots, 2n$. Recalling the notations in Section 1, $r(X_{i+n})=r(Y_i)=R_{i+n}=Q_i$. We need the following assumptions.

ASSUMPTION 2.1. The derivative

$$f'(x, y; \theta)=(\partial/\partial\theta)f(x, y; \theta)$$

exists for almost everywhere (x, y) and the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f'(x, y; \theta)|dxdy$$

is finite and continuous in a neighbourhood of $\theta=0$.

ASSUMPTION 2.2. The derivative

$$f''(x, y; \theta)=(\partial^2/\partial\theta^2)f(x, y; \theta)$$

exists for almost everywhere (x, y) and the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f''(x, y; \theta)|dxdy$$

is finite and continuous in a neighbourhood of $\theta=0$.

Note that the continuity of the integrals of the above assumptions are weaker than the usual Lebesgue's assumptions that $|f'(x, y; \theta)|\leqq M_1(x, y)$ and $|f''(x, y; \theta)|\leqq M_2(x, y)$ for some integrable functions M_1 and M_2.

We can obtain the following theorem.

THEOREM 2.1. *Under Assumption* 2.1, *the locally most powerful intraclass rank tests of the null hypothesis* $\theta=0$ *against the alternative* $\theta>0$ *have the critical regions of the form*

$$S_L=\sum_{i=1}^{n} E_0\Big(\frac{f'(X_i, Y_i\,;0)}{f(X_i)f(Y_i)}\,|\,r(X_i)=R_i\,,\quad r(Y_i)=Q_i\Big)\geqq c \tag{2.1}$$

at the respective levels.

Here E_0 is calculated under $\theta=0$. From now on, E_0, var_0 and cov_0 imply that the calculations are performed under H or $\theta=0$.

PROOF. We have

$$\begin{aligned}(2.2)\qquad & P_\theta(\boldsymbol{R}=\boldsymbol{r})-P_0(\boldsymbol{R}=\boldsymbol{r})\\ &=\sum_{i=1}^{n}\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int(f(x_i,\ y_i\,;\theta)-f(x_i)f(y_i))\prod_{j=1}^{i-1}f(x_j,\ y_j\,;\theta)\prod_{j=i+1}^{n}(f(x_j)f(y_j))\boldsymbol{dx}\\ &=\sum_{i=1}^{n}\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int\int_0^\theta f'(x_i,\ y_i\,;t)dt\prod_{j=1}^{i-1}f(x_j,\ y_j\,;\theta)\prod_{j=i+1}^{n}(f(x_j)f(y_j))\boldsymbol{dx}\\ &\equiv\sum_{i=1}^{n}\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int\int_0^\theta A_i dt\boldsymbol{dx}\,,\quad \text{(say)},\end{aligned}$$

where

$$\boldsymbol{dx}=\prod_{i=1}^{n}(dx_i dy_i)\,.$$

Clearly

$$\lim_{\theta\to 0} A_i/\theta=\frac{f'(x_i,\ y_i\,;0)}{f(x_i)f(y_i)}\prod_{j=1}^{n}(f(x_j)f(y_j))\,.$$

Furthermore,

$$\int_{R^{2n}}\cdots\int\int_0^\theta|A_i/\theta|\,dt\boldsymbol{dx}\leqq\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta^{-1}\int_0^\theta|f'(x,\ y\,;t)|\,dtdxdy\,. \tag{2.3}$$

From Assumption 2. 1, the right hand side of (2.3) converges to $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|f'(x,\ y\,;0)|\,dxdy$ as $\theta\to 0$. Hence, from the convergence theorem II 4.2 of Hájek and Šidák (1967),

$$\lim_{\theta\to 0}\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int\theta^{-1}\int_0^\theta A_i dt\boldsymbol{dx}=\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int f'(x_i,\ y_i\,;0)\prod_{j(\neq i)}(f(x_j)f(y_j))\boldsymbol{dx}\,.$$

Thus, it holds that

$$\lim_{\theta\to 0}(P_\theta(\boldsymbol{R}=\boldsymbol{r})-P_0(\boldsymbol{R}=\boldsymbol{r}))/\theta=((2n)\,!)^{-1}\sum_{i=1}^{n}E_0\Big(\frac{f'(X_i,\ Y_i\,;0)}{f(X_i)f(Y_i)}\,|\,r(X_i)=R_i,\ r(Y_i)=Q_i\Big)\,.$$

From the Neyman-Pearson lemma, we can get the desired result.

If (X, Y) is normal or it satisfies a model of Farlie type (1960)

$$f(x,\ y\,;\theta)=f(x)f(y)(1+\theta A(x,\ y)+o(\theta)) \tag{2.4}$$

where $A(x,\ y)=A(y,\ x)$, then the LMPIRT can be given by Theorem 2.1. However, in some models the test statistic S_L vanishes. The Hájek model

$$(X,\ Y)=(X^*+\Delta Z,\ Y^*+\Delta Z) \tag{2.5}$$

is a typical one, since in this model $f'(x,\ y\,;0)$ is proportional to $f(x)f'(y)+f'(x)f(y)$.

This is because X and Y are always positively correlated for $\Delta \neq 0$. Hence, we must consider the two-sided alternative $\Delta \neq 0$ in this case.

THEOREM 2.2. *Suppose $S_L=0$ with probability one and Assumptions 2.1 and 2.2 are satisfied, then the locally most powerful intraclass rank tests of the null hypothesis $\theta=0$ against the alternative $\theta \neq 0$ have the critical regions of the form*

$$(2.6) \qquad \begin{aligned} T_L = &\frac{1}{2}\sum_{i=1}^{n} E_0\Big(\frac{f''(X_i, Y_i; 0)}{f(X_i)f(Y_i)} \,|\, r(X_i)=R_i, r(Y_i)=Q_i\Big) \\ &+\sum\sum_{i>j} E_0\Big(\frac{f'(X_i, Y_i; 0)f'(X_j, Y_j; 0)}{f(X_i)f(Y_i)f(X_j)f(Y_j)} \,|\, r(X_i)=R_i, r(Y_i)=Q_i, \\ &\qquad\qquad , r(X_j)=R_j, r(Y_j)=Q_j\Big) \\ \geqq c& \end{aligned}$$

at the respective levels.

PROOF. From (2.2), we have

$$(2.7) \qquad \begin{aligned} &P_\theta(\boldsymbol{R}=\boldsymbol{r})-P_0(\boldsymbol{R}=\boldsymbol{r}) \\ &= \sum_{i=1}^{n}\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int\int_0^\theta\int_0^t f''(x_i, y_i; s)\prod_{j=1}^{i-1} f(x_j, y_j; t)\prod_{j=i+1}^{n}(f(x_j)f(y_j))dsdt\boldsymbol{dx} \\ &\quad + \sum_{i=1}^{n}\theta\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int f'(x_i, y_i; 0)\Big(\prod_{j=1}^{i-1} f(x_j, y_j; \theta)-\prod_{j=1}^{i-1}(f(x_j)f(y_j))\Big) \\ &\quad \times\prod_{j=i+1}^{n}(f(x_j)f(y_j))\boldsymbol{dx}+\sum_{i=1}^{n}\theta\int_{\boldsymbol{R}=\boldsymbol{r}}\cdots\int f'(x_i, y_i; 0)\prod_{j(\neq i)}(f(x_j)f(y_j))\boldsymbol{dx}\,. \end{aligned}$$

From the assumptions, the last term of (2.7) is equal to zero. From the same convergence theorem used in the proof of Theorem 2.1, the second and the first terms each multiplied by $(2n)!/\theta^2$ converges to the first and the second terms of T_L, respectively. The Neyman-Pearson lemma ensures the conclusion of the theorem.

The test statistic T_L seems very complicated. However, it is not necessarily so. For example, in the Hájek model (2.5), $f''(x, y; 0)$ is proportional to $f''(x)f(y)+2f'(x)f'(y)+f(x)f''(y)$ and T_L is equivalent to

$$\sum_{i=1}^{n} E_0\Big(\frac{f'(X_i)f'(Y_i)}{f(X_i)f(Y_i)} \,|\, r(X_i)=R_i, r(Y_i)=Q_i\Big)$$

which is analogous to S_L in its form. Two statistics S_L and T_L are analogous to statistics in Shirahata (1974) in which usual ranks are adopted.

In many model $f'(x, y; 0)/(f(x)f(y))$ is of the form $A(x)A(y)$ and the statistic

$$(2.8) \qquad S_L^* = \sum_{i=1}^{n} E_0(A(X_i)\,|\,r(X_i)=R_i)E_0(A(Y_i)\,|\,r(Y_i)=Q_i)$$

is of interest although it is not equivalent to S_L or T_L.

In the following sections the asymptotic properties of intraclass rank statistics including S_L will be investigated and some special attentions are paid to a class including S_L^*.

3. Asymptotic Distributions of Intraclass Rank Statistics under the Hypothesis of Independence

Let us consider the asymptotic property of the statistic

$$S_{nG}=\sum_{i=1}^{n} a_n(R_i,\, Q_i) \tag{3.1}$$

under H where $a_n(i, j)$, $i, j=1, \cdots, 2n$ be some given constants such that $a_n(i, j)=a_n(j, i)$. The LMPIRT S_L in Theorem 2.1 is a special case of S_{nG} and many of T_L in Theorem 2.2 will be special cases of S_{nG}. In the usual rank tests, Shirahata (1974) considered the asymptotic normality of locally most powerful rank tests. In our situation, the results of Jogdeo (1968) which are extensions of Hájek (1961) are interesting. The following theorems 3.1, 3.2 and 3.3 are based on Jogdeo's results.

Define the function $\phi_n(u, v)$ on $(0, 1]\times(0, 1]$ as

$$\phi_n(u,\, v)=a_n(i,\, j) \qquad \frac{i-1}{2n}<u\leqq\frac{i}{2n},\ \frac{j-1}{2n}<v\leqq\frac{j}{2n} \tag{3.2}$$

and suppose that the following is satisfied.

ASSUMPTION 3.1. A collection of constants $\{a_n(i, j)\}$ is Δ-monotone in the sense that

$$\Delta_{ij}=a_n(i+1,\, j+1)-a_n(i+1,\, j)-a_n(i,\, j+1)+a_n(i,\, j)\geqq 0$$

for all (i, j) or $\Delta_{ij}\leqq 0$ for all (i, j).

Put

$$T_{nG}=\sum_{i=1}^{n}\phi_n(F(X_i),\, F(Y_i))-\frac{1}{2(2n-1)}\sum_{i\neq j=1}^{2n}\sum\phi_n(F(X_i),\, F(X_j)) \tag{3.3}$$
$$+\frac{1}{2(2n-1)}\sum_{i\neq j=1}^{2n}\sum a_n(i,\, j)$$

where $F(x)$ is the common marginal distribution function. Note that $Y_i=X_{i+n}$ in (3.3). Then from Jogdeo (1968. Theorem 4.1), we have

THEOREM 3.1. *Suppose that* $\{a_n(i, j)\}$ *satisfies Assumption* 3.1 *and*

$$\lim_{n\to\infty} n^{-1}\max_{(i,j)}\Big(a_n(i,\, j)-\frac{1}{4n^2}\sum_{k,m=1}^{2n}\sum a_n(k,\, m)\Big)^4=0\,.$$

Then, under H, it holds that

$$\lim_{n\to\infty} E_0\{(S_{nG}-T_{nG})^2/n\}=0\,. \tag{3.4}$$

From Theorem 3.1, S_{nG} is asymptotically equivalent to T_{nG}. Now, consider a piecewise monotone function $\phi(u, v)$ on $(0, 1)\times(0, 1)$ such that $\phi(u, v)=\phi(v, u)$. Let $\{a_n(i, j)\}$ be a set of constants constructed from one of

$$a_n(i,\, j)=\phi(i/(2n+1),\, j/(2n+1))\,, \tag{3.5}$$

$$a_n(i,\, j)=4n^2\int_{(i-1)/(2n)}^{i/(2n)}\int_{(j-1)/(2n)}^{j/(2n)}\phi(u,\, v)\,du\,dv \tag{3.6}$$

and

$$a_n(i,\, j)=E_0(\phi(F(X_k),\, F(X_m))\,|\,r(X_k)=i,\, r(X_m)=j)\,. \tag{3.7}$$

The statistic S_L is a special case of (3.7). Then from Jogdeo (1968, Theorem 4.2), we have

THEOREM 3.2. *Suppose that* $\{a_n(i, j)\}$ *is constructed from one of* (3.5), (3.6) *and* (3.7) *and it satisfies Assumption* 3.1. *Furthermore, suppose*

$$\int_0^1\int_0^1 \phi^8(u, v)dvdv<\infty . \tag{3.8}$$

Then, we have the convergence (3.4) *under* H.

In Jogdeo (1968), the case (3.7) is not treated. But, we can prove the theorem easily by showing that the function $\phi_n(u, v)$ is uniformly integrable in n. The proof of the uniform integrability is due to the martingale theory adopted in Shirahata (1974). From Theorem 3.2, many practical scores constants give the convergence (3.4). However, the random variable T_{nG} is complicated. Instead of T_{nG}, let us introduce

$$\begin{aligned} T^*_{nG}= \sum_{i=1}^{n} \phi(F(X_i), F(Y_i))-\frac{1}{2(2n-1)}\sum_{i\neq j=1}^{2n}\sum\phi(F(X_i), F(X_j)) \\ +\frac{1}{2(2n-1)}\sum_{j\neq j=1}^{2n}\sum a_n(i, j) . \end{aligned} \tag{3.9}$$

Then we have

THEOREM 3.3. *Under the same assumptions of Theorem* 3.2,

$$\lim_{n\to\infty} E_0\{(T_{nG}-T^*_{nG})^2/n\}=0 . \tag{3.10}$$

The proof of Theorem 3.3 is along a similar line of the proof of Theorem 3.2 (see Jogdeo (1968)) and hence it is omitted.

Now, without loss of generality, it is assumed that

$$\int_0^1\phi(u, v)du=\int_0^1\phi(u, v)dv=0 . \tag{3.11}$$

Then, it is easily seen that T^*_{nG} is asymptotically equivalent to

$$T^{**}_{nG}= \sum_{i=1}^{n} \phi(F(X_i), F(Y_i)) . \tag{3.12}$$

Thus, we can get the following theorem.

THEOREM 3.4. *Suppose that the assumptions of Theorem* 3.3 *and* (3.11) *are satisfied. Then* S_{nG} *is asymptotically normal with mean zero and the asymptotic variance* $n\int_0^1\int_0^1\phi^2(u, v)dudv$ *under* H.

The conditions to ensure the asymptotic normality of S_{nG} are strong and it is tedious to check them. However, if $a_n(i, j)=a_n(i)a_n(j)$ for some scores constants $\{a_n(i), i=1, \cdots, 2n\}$, then it is easier to investigate though R_i and Q_i are not independent even under H. Hence, let us consider

$$S_{nP}= \sum_{i=1}^{n} a_n(R_i)a_n(Q_i) \tag{3.13}$$

which is a generalization of Shirahata (1981) in which $a_n(i)=i$. We need the following assumptions

ASSUMPTION 3.2. The scores constants $a_n(i)$, $i=1, \cdots, 2n$ satisfy

(3.14) $$\lim_{n\to\infty}\int_0^1 (a_n(1+[2nu])-\phi(u))^4 du=0$$

for some non-constant function $\phi(u)$ on (0, 1) such that $\int_0^1 \phi^4(u)du<\infty$. Here $[2nu]$ is the integer part of $2nu$.

ASSUMPTION 3.3. It holds that

(3.15) $$\int_0^1 \phi(u)du=\sum_{i=1}^{2n} a_n(i)=0\,.$$

Assumption 3.3 is for the sake of simplicity of the calculations. We can get the following theorem.

THEOREM 3.5. *If Assumptions* 3.2 *and* 3.3 *hold, then* S_{nP} *is asymptotically normal with mean zero and the asymptotic variance* $n\left(\int_0^1 \phi^2(u)du\right)^2$ *under* H.

PROOF. It sufficies to show

(3.16) $$\lim_{n\to\infty} E_0\{(S_{nP}-T_{nP})^2/n\}=0$$

where

(3.17) $$T_{nP}=\sum_{i=1}^{n} \phi(F(X_i))\phi(F(Y_i))\,.$$

Now

$$E_0\{(S_{nP}-T_{nP})^2/n\}=E_0(D_1^2)+(n-1)E_0(D_1D_2)$$

where

$$D_i=a_n(R_i)a_n(Q_i)-\phi(F(X_i))\phi(F(Y_i))\,,\qquad i=1,\ 2.$$

Clearly

$$E_0(D_1^2)\leqq 2E_0\{a_n^2(R_1)(a_n(Q_1)-\phi(F(Y_1)))^2\}+2E_0\{\phi^2(F(Y_1))(a_n(R_1)-\phi(F(X_1)))^2\}\,.$$

From (3.14) and the Schwartz inequality, it is found that $E_0(D_1^2)$ converges to zero as $n\to\infty$.

On the other hand, from (3.15),

(3.18) $$\begin{aligned} E_0(D_1D_2)=&E_0(a_n(R_1)a_n(R_2)a_n(Q_1)a_n(Q_2))\\ &-2E_0(a_n(R_1)a_n(Q_1)\phi(F(X_2))\phi(F(Y_2)))\,. \end{aligned}$$

The first term of (3.18) is

$$3\left\{\left(\sum_{i=1}^{2n} a_n^2(i)\right)^2-2\sum_{i=1}^{2n} a_n^4(i)\right\}/\{2n(2n-1)(2n-2)(2n-3)\}=0(n^{-2})\,.$$

The second term is, up to the multiplicative constants,

$$\begin{aligned} &E_0\{E_0(a_n(R_1)a_n(Q_1)\phi(F(X_2))\phi(F(Y_2))\,|\,\boldsymbol{X}^*,\ r(X_2)=R_2,\ r(Y_2)=Q_2)\}\\ &\quad=E_0\left\{\phi(F(X_2))\phi(F(Y_2))\left(-\sum_{i=1}^{2n} a_n^2(i)+2a_n^2(R_2)+2a_n^2(Q_2)+2a_n(R_2)a_n(Q_2)\right)\right\}\\ &\qquad\qquad/\{(2n-2)(2n-3)\}\\ &\quad=0(n^{-2})\,. \end{aligned}$$

Thus, $E_0(D_1D_2)=0(n^{-2})$ and hence (3.16) is established.

4. Asymptotic Distributions of Intraclass Rank Statistics under Contiguous Alternatives

Let us consider the one-parameter family $f(x, y; \theta)$ in Section 2. The alternative hypothesis to be considered in this section is $H_n(\theta_0)$; the density function of (X_i, Y_i) is $f(x, y; n^{-1/2}\theta_0)$. We need the following assumptions.

ASSUMPTION 4.1. Let $s(x, y; t)=f^{1/2}(x, y; t)$ and $s'(x, y; t)=(\partial/\partial t)s(x, y; t)$. The integral $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(s'(x, y; t))^2 dxdy$ is continuous in a neighbourhood of $t=0$.

At first, let us show that $H_n(\theta_0)$ is contiguous to H. Introduce

$$\log(L_n)=\sum_{i=1}^{n}\log(f(X_i, Y_i; n^{-1/2}\theta_0)/(f(X_i)f(Y_i))),$$

$$W_n=2\sum_{i=1}^{n}\{(f(X_i, Y_i; n^{-1/2}\theta_0)/(f(X_i)f(Y_i)))^{1/2}-1\}$$

and

$$T_n=n^{-1/2}\theta_0\sum_{i=1}^{n} f'(X_i, Y_i; 0)/(f(X_i)f(Y_i)).$$

To establish the contiguity, we may show the following two lemmas (see Hájek and Šidák (1967)).

LEMMA 4.1. *If Assumption* 4.1 *is fulfilled, then*

(4.1) $$\lim_{n\to\infty} E_0(W_n)=-\sigma^2/4$$

under $H_n(0)$ *where*

$$\sigma^2=var_0(f'(X, Y; 0)/(f(X)f(Y)))$$
$$=4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(s'(x, y; 0))^2 dxdy.$$

PROOF. $$E_0(W_n)=-n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(s(x, y; n^{-1/2}\theta_0)-s(x, y; 0))^2 dxdy$$
$$=-\theta_0^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(n^{1/2}\theta_0^{-1}\int_0^{n^{-1/2}\theta_0} s'(x, y; t)dt\right)^2 dxdy$$
$$=-\theta_0^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A_n(x, y; \theta_0)dxdy, \qquad \text{say.}$$

Then obviously

$$\lim_{n\to\infty} A_n(x, y; \theta_0)=(s'(x, y; 0))^2.$$

Furthermore,

(4.2) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A_n(x, y; \theta_0)dxdy\leqq\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}n^{1/2}\theta_0^{-1}\int_0^{n^{-1/2}\theta_0}(s'(x, y; t))^2 dtdxdy.$$

From Assumption 4.1 the right hand side of (4.2) converges to $\sigma^2/4$. Hence, from the convergence theorem II 4.2 of Hájek and Šidák (1967), (4.1) holds.

LEMMA 4.2. *Suppose Assumption* 4.1 *is fulfilled, then*

$$\lim_{n\to\infty} var_0(W_n-T_n)=0$$

under $H_n(0)$.

PROOF. Since $W_n - T_n$ is n independent summands,

$$var_0(W_n - T_n) \leqq 4n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{s(x, y; n^{-1/2}\theta_0) - s(x, y; 0) \\ - n^{-1/2}\theta_0 s'(x, y; 0)\}^2 dxdy . \tag{4.3}$$

As in Lemma 4.1, the right hand side of (4.3) converges to zero from the convergence theorem V 1.3 of Hájek and Šidák (1967).

From the Le Cam's second lemma and the above lemmas, it holds that $\log(L_n)$ is asymptotically equivalent to $W_n - \sigma^2/4$ and consequently to $T_n - \sigma^2/2$. Thus, we can get the following theorem.

THEOREM 4.1. *Suppose that the convergence*

$$\lim_{n\to\infty} E_0\{(S_{nG} - T^*_{nG})^2/n\} = 0$$

holds under $H_n(0)$. *Also suppose that* (3.8) *and Assumption* 4.1 *are fulfilled. Then* S_{nG} *is, under* $H_n(\theta_0)$, *asymptotically normal with mean* $n^{1/2}\theta_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(x, y; 0)\phi(F(x), F(y))dxdy$ *and the variance given in Theorem* 3.4.

PROOF. From the Le Cam's third lemma, S_{nG} is, under $H_n(\theta_0)$, asymptotically normal with mean $cov_0(T^*_{nG}, T_n)$ and the same variance as in $H_n(0)$. The calculation of $cov_0(T^*_{nG}, T_n)$ leads us to the conclusion.

Note that when S_{nG} satisfies the assumptions of Theorem 3.4, the asymptotic normality of S_{nG} follows from Theorem 4.1.

In the model (2.5), $cov_0(T^*_{nG}, T)=0$ and the above theorem seems to be useless. However, the model is interpretted as

$$f(x, y; \Delta) = f(x)f(y)\left(1 + E(Z^2)\Delta^2\left(\frac{f'(x)f'(y)}{f(x)f(y)} + \frac{f''(x)}{2f(x)} + \frac{f''(y)}{2f(y)} + 0(\Delta^2)\right)\right) \tag{4.4}$$

provided $E(Z)=0$. The condition $E(Z)=0$ does not affect the distribution of S_{nG}. Hence, by putting $\theta_0 = \Delta^2$ and assuming some regularity conditions, the asymptotic consideration of S_{nG} for the model (2.5) is possible from Theorem 4.1 and the asymptotic mean is given by

$$n^{1/2}\theta_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(x)f'(y)\phi(F(x))\phi(F(y))dxdy .$$

From Theorem 3.5 and the contiguity, we can get further result when the scores constants are of the product form.

THEOREM 4.2. *Suppose the assumptions in Theorem* 3.5 *and Assumption* 4.1 *are satisfied, then* S_{nP} *given in Section* 3 *is asymptotically normal with mean*

$$n^{1/2}\theta_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'(x, y; 0)\phi(F(x))\phi(F(y))dxdy$$

and the variance given in Theorem 3.5 *under* $H_n(\theta_0)$.

Let us consider (4.4) again. Then we can derive the asymptotic distribution of S_{nP} under the model (2.5) and the asymptotic mean is $n^{1/2}\theta_0\left(\int_{-\infty}^{\infty} f'(x)\phi(F(x))dx\right)^2$. A similar result will hold for the usual rank statistics with possibly different marginal distributions.

5. Approximation of distribution of S_{nG} under H

In order to use the statistic S_{nG}, it is required to get the percentage points under H. Theorems in Section 3 can be utilized to determine the asymptotic critical points. However, the asymptotic distribution of S_{nG} is symmetric but the exact distributions are not so even if symmetric scores are adopted. Therefore, the approximation of the asymptotic distribution will be not so good and we need further considerations to apply S_{nG} for relatively small n.

Without loss of generality, assume that $\sum_{i=1}^{2n} a_n(i, j)=0$. Note that it is already assumed that $a_n(i, j)=a_n(j, i)$. A useful method is to standardize S_{nG} with its exact moments. The exact means of S_{nG} and S_{nG}^2 are given by

$$E_0(S_{nG})=-\sum_{i=1}^{2n} a_n(i, i)/\{2(2n-1)\}$$

and

$$E_0(S_{nG}^2)=\left\{\left(\sum_{i=1}^{2n} a_n(i, i)\right)^2-4n\sum_{i=1}^{2n} a_n^2(i, i)+4(n-1)\left(\sum\sum_{i,j=1}^{2n} a_n^2(i, j)\right)\right\} \Big/ \{4(2n-1)(2n-3)\}.$$

Hence

$$var_0(S_{nG})=\left\{2\left(\sum_{i=1}^{2n} a_n(i, i)\right)^2-4n(2n-1)\sum_{i=1}^{2n} a_n^2(i, i)+4(n-1)(2n-1)\sum\sum_{i,j=1}^{2n} a_n^2(i, j)\right\} \Big/ \{4(2n-1)^2(2n-3)\}.$$

Therefore, for the scores generating function $\phi(u, v)$ satisfying $\phi(u, v)=\phi(v, u)$ and $\int_0^1\phi(u, v)du=0$, the mean and the variance of S_{nG} under H are approximately

$$E_0(S_{nG})\sim-\frac{1}{2}\int_0^1\phi(u, u)du$$

and

$$var_0(S_{nG})\sim n\int_0^1\int_0^1\phi^2(u, v)dudv\,.$$

Thus, it will be more accurate to determine the critical points from approximations

$$S_{nG}\sim N(E_0(S_{nG}), var_0(S_{nG}))$$

or

$$S_{nG}\sim N\left(-\frac{1}{2}\int_0^1\phi(u, u)du,\ n\int_0^1\int_0^1\phi^2(u, v)dudv\right).$$

If we use formal Edgeworth approximation, we can obtain further approximation. Put

$$S_{nG}^*=(S_{nG}-E_0(S_{nG}))/var_0^{1/2}(S_{nG})\,.$$

Then the approximation is

$$P(S_{nG}^*\leqq x)\sim N(x)-n(x)\{(x^2-1)\mu_3/6+(x^3-3x)(\mu_4-3)/24\} \tag{5.1}$$

where $N(x)$ and $n(x)$ are the distribution function and the density function of the standard normal distribution, respectively and where μ_3 and μ_4 are third and fourth moments of S^*_{nG}, respectively. The calculations of the moments are tedious and are omitted. However, for S_{nP} given in Section 3, the calculations are easy when $\{a_n(i)\}$ is symmetric i. e., $a_n(i)=-a_n(2n+1-i)$. In Shirahata (1981), the approximations when $a_n(i)=i-n-\frac{1}{2}$ are given and are very satisfactory.

References

[1] FARLIE, D. J. G.: *The performance of some correlation coefficient for a general bivariate distribution*. Biometrika, **47** (1960), 307–323.

[2] HÁJEK, J.: *Some extensions of the Wald-Wolfowitz-Noether theorem*. Ann. Math. Statist., **32** (1961), 506–523.

[3] HÁJEK, J. and ŠIDÁK, Z.: Theory of Rank Tests. (1967). Academic Press, New York.

[4] JOGDEO, K.: *Asymptotic normality in nonparametric methods*. Ann. Math. Statist., **39** (1968), 905–922.

[5] SHIRAHATA, S.: *Locally most powerful rank tests for independence*. Bull. Math. Statist., **16** (1974), 11-21.

[6] SHIRAHATA, S.: *Intraclass rank tests for independence*. Biometrika, **68** (1981), 451-456.

Communicated by N. Furukawa
Received May 28, 1982