数理解析研究所講究録
第327巻 1978年 47-51

RANK TESTS FOR THE k-SAMPLE PROBLEM

S. SHIRAHATA

Osaka University

1. Introduction. Let us consider k populations Π_i, $i\in\bar{k}$ and assume that each Π_i has a continuous distribution function F_i. Here by $\bar{k}$ we denote the set $\{1,\ldots,k\}$. Suppose X_{ij}, $j\in\bar{n}_i$ be a random sample drawn from Π_i and put $N=\sum_{i=1}^{k} n_i$. The purpose of this paper is to investigate the asymptotic properties of rank statistics as $N\to\infty$ in the problem of testing $H_0;F_1=\ldots=F_k=F$ where the functional form of F is unknown. Put $\lambda_i=n_i/N$ and suppose that $0<\underline{\lim}_{N\to\infty}\lambda_i\leq\overline{\lim}_{N\to\infty}\lambda_i<0$ for $i\in\bar{k}$.

There are two methods of ranking procedures. The first method is to combine k samples and then rank the observations. Define the rank R_{ij} of X_{ij} among $\{X_{rs};s\in\bar{n}_r,\ r\in\bar{k}\}$ as

$$R_{ij}=\sum_{r=1}^{k}\sum_{s=1}^{n_r}u(X_{ij}-X_{rs})$$

where $u(x)=1\ (0)$ according as $x\geq 0\ (<0)$. Let us call this ranking procedure as Type I ranking procedure. The rank statistics to be considered in Type I is

$$\underline{S}=(S_1,\ldots,S_k)'$$

where $S_i=\sum_{j=1}^{n_i}a_N(R_{ij})$ and $a_N(.)$ are score constants satisfying

$$\lim_{N\to\infty}\int_0^1(\phi(u)-a_N(1+[uN]))^2du=0$$

for some non-constant square integrable function ϕ with $\int_0^1\phi(u)du=1-\int_0^1\phi^2(u)du=0$.

The scores generating function ϕ is usually chosen to be increasing. For example, $\phi(u)=\sqrt{12}(u-\frac{1}{2})$, $\Phi^{-1}(u)$ etc.

The other method of ranking is to combine Π_i and Π_j, $(i<j)$ and then rank the observations, i.e., the rank of X_{js} among (Π_i, Π_j) is given by

$$R_{js}^{ij} = \sum_{t=1}^{n_j} u(X_{js}-X_{jt}) + \sum_{t=1}^{n_i} u(X_{js}-X_{it}).$$

Let us call this ranking procedure as Type II ranking procedure. The rank statistics to be considered in Type II is

$$\underline{S}^{II} = (S_{12}^{II}, S_{13}^{II}, \ldots, S_{1k}^{II}, S_{23}^{II}, \ldots, S_{2k}^{II}, \ldots, S_{k-1,k}^{II})'$$

where $S_{ij}^{II} = \sum_{t=1}^{n_j} a_{n_i+n_j}(R_{jt}^{ij})$, $i<j$. Here the scores constants are chosen as in Type I.

Type I ranking procedure is very convenient theoretically and is studied by many workers, see Hajek & Sidak(1967). However, the ranking is very tedious in practice when N is large. Type II ranking procedure has been studied in the problem of testing H_0 against ordered alternatives by Puri (1965), Jonkheere(1954) and Koziol & Reid(1977). The procedure is also practically inconvenient if we need every S_{ij}^{II} to get a good test. But if some of S_{ij}^{II}'s are enough, it is very useful in practice. We shall show that it is enough to consider adjacent part

$$\underline{S}_{ad}^{II} = (S_{12}^{II}, S_{23}^{II}, \ldots, S_{k-1,k}^{II})'.$$

Our results include Tryon & Hettmansperger(1973) as a special case. When the alternative is not ordered, the word "adjacent" seems to be meaningless. However, any arrangement of the populations will do in not ordered case.

2. Rank statistics. Now suppose that the density function of Π_i exists and is given by $f(x;c_i\theta)$. Then we have the folling theorem. The proof of the theorem is simple and is omitted.

Theorem 1. Suppose $\dot{f}(x;\theta)=(\partial/\partial\theta)f(x;\theta)$ exists in the neighbourhood of $\theta=0$ and satisfies

$$\lim_{\theta\to\infty}\int|\dot{f}(x;\theta)|dx=\int|\dot{f}(x;0)|dx.$$

Then the locally most powerful Type I rank tests for the one-sided alternative $\theta>0$ is based on

$$S(f)=\sum_{i=1}^{k}\sum_{j=1}^{n_i}c_i a_N(R_{ij};f)$$

where $a_N(r_{ij};f)=E(\dot{f}(X^{(r_{ij})};0)/f(X^{(r_{ij})};0))$ and where $X^{(r_{ij})}$ is the r_{ij}-th order statistic among N observations drawn from $f(x;0)$ and the expectation is calculated under $f(x;0)$.

This theorem implies that it is reasonable to consider tests based on $\underline{S}$. In fact the S(f) test gives an asymptotically most powerful test.

Now, let us consider the following classes of test statistics which include linear rank statistics.

$$T(\underline{S})=\{h(\underline{S});\ h \text{ is continuous}\},$$

$$T(\underline{S}^{II})=\{h^{II}(\underline{S}^{II});\ h^{II} \text{ is continuous}\}$$

and

$$T(\underline{S}^{II}_{ad})=\{h^{II}_{ad}(\underline{S}^{II}_{ad});\ h^{II}_{ad} \text{ is continuous}\}.$$

It is clear that $T(\underline{S}^{II}_{ad})\subset T(\underline{S}^{II})$. The statistic $\underline{S}^{II}_{ad}$ is more convenient in practice than both $\underline{S}$ and $\underline{S}^{II}$.

3. Asymptotic equivalence of $T(\underline{S})$, $T(\underline{S}^{II})$ and $T(\underline{S}^{II}_{ad})$. In this section we shall show that the three classes of rank statistics are asymptotically equivalent in the following sense. Let $\tilde{S}$ be an element of, say, $T(\underline{S})$, then there exist $\tilde{S}_1\in T(\underline{S}^{II})$ and $\tilde{S}_2\in T(\underline{S}^{II}_{ad})$ such that $\tilde{S}$, $\tilde{S}_1$ and $\tilde{S}_2$ are asymptotically equivalent in probability each other under H_0 and the contiguous alternatives.

The foundamental relations to establish the above assertion are

$$S_i \sim Y_i - \lambda_i \sum_{j=1}^{k} Y_j \tag{3.1}$$

and

$$S^{II}_{ij} \sim Y_j - \frac{\lambda_i}{\lambda_i+\lambda_j}(Y_i+Y_j) = \frac{\lambda_j Y_j - \lambda_i Y_i}{\lambda_i+\lambda_j} \tag{3.2}$$

where $Y_i = \sum_{t=1}^{n_i} \phi(F(X_{it}))$. These are given by Hajek & Sidak(1967). The sign $X_n \sim Y_n$ implies $\sqrt{n}(X_n - Y_x) \to 0$ in probability. The following theorem which was shown by Koziol & Reid(1977) is an easy consequence of (3.1) and (3.2).

Theorem 2. Under H_0 and the contiguous alternatives, $S^{II}_{ij} \sim (\lambda_i S_j - \lambda_j S_i)/(\lambda_i+\lambda_j)$.

Theorem 2 implies that $T(\underline{S}^{II}) \subset T(\underline{S})$. It is already stated that $T(\underline{S}^{II}_{ad}) \subset T(\underline{S}^{II})$ and hence the following theorem proves our assertion.

Theorem 3. Under H_0 and contiguous alternatives, $T(\underline{S}) \subset T(\underline{S}^{II}_{ad})$. More precisely, S_i is asymptotically equivalent to a linear function of $S^{II}_{i,i+1}$, $i \in \overline{k-1}$.

Proof. Since $S^{II}_{i,i+1} \sim (\lambda_i S_{i+1} - \lambda_{i+1} S_i)/(\lambda_i+\lambda_{i+1})$,

$$\begin{pmatrix} S^{II}_{12} \\ S^{II}_{23} \\ \vdots \\ S^{II}_{k-1,k} \\ 0 \end{pmatrix} \sim \begin{pmatrix} -a_1 & 1-a_1 & 0 & \cdots & 0 \\ 0 & -a_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & -a_{k-1} & 1-a_{k-1} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_{k-1} \\ S_k \end{pmatrix}$$

where $a_i = \lambda_{i+1}/(\lambda_i+\lambda_{i+1})$. The determinant of the matrix above is not zero. Thus, the theorem is proved.

From Theorems 2, 3 and the results of Hajek & Sidak(1967), the asymptotic normality of $\underline{S}$, $\underline{S}^{II}$ and $\underline{S}^{II}_{ad}$ can be shown.

It will be found that $ES_i = ES_{ij}^{II} = 0$ under H_0 and the asymptotic covariance matrices can be found.

Let us again consider the alternative

$$f_i(x) = f(x; c_i/N^{-1/2}). \tag{3.3}$$

From the following two theorems given in Hajek & Sidak(1967), it can be shown that any linear combinations $\underline{a}'\underline{S}$, $\underline{\bar{a}}'\underline{S}^{II}$ and $\underline{\bar{\bar{a}}}'\underline{S}_{ad}^{II}$ are asymptotically normal with the same asymptotic variances as in H_0 under the above alternative. The asymptotic means of them will be found too.

Theorem 4. Suppose $\int|f(x;\theta)|dx$ and $\int(\dot{f}(x;\theta))^2/f(x;\theta)\,dx$ are continuous at $\theta=0$. Then

$$L_N = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \dot{f}(X_{ij}; c_i/N^{-1/2})/f(X_{ij};0)$$

$$\sim -N^{-1/2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} c_i\dot{f}(X_{ij};0)/f(X_{ij};0) - \sigma^2/2$$

under H_0 where

$$\sigma^2 = \sum_{i=1}^{k} \lambda_i c_i^2 \int(\dot{f}(x;0))^2/f(x;0)\,dx .$$

Theorem 5. If a statistic S and L_N are asymptotically $N(\mu,\tilde{\sigma}^2)$ and $N(-\sigma^2/2,\sigma^2)$, respectively and if (S,L_N) is asymptotically bivariate normal with asymptotic covariance σ_{12}, then S is asymptotically $N(\mu+\sigma_{12},\tilde{\sigma}^2)$ under (3.3).

References

Hajek, J. & Sidak, Z. (1967). Theory of Rank Test. Academic Press.

Jonckheere, A.R. (1954). Biometrika, 41, 133-145.

Koziol, J.A. & Reid, N. (1977). Ann. Statist., 5, 1099-1106.

Puri, M.L. (1965). Comm. Pure Appl. Math., 18, 51-63.

Tryon, P.V. & Hettmansperger, T.P. (1973). Ann. Statist., 1, 1061-1070.