Bulletin of Shimane University (Natural Science), No. 13, pp. 123—127.
December, 1963.

On certain c-sample problem against ordered alternatives

By

Ryoji TAMURA

1. Introduction. Many investigations have shown considerable advantage of nonparametric rank procedure over the one based on the normal theory for slippage alternatives in the two sample problem. For example, Hodges-Lehmann [4] and [5] gave the values of lower limit 0.864 and 1 of the asymptotic efficiency of Wilcoxon and the Normal scores tests with regard to the classical t-test, respectively. Any tests have not unfortunately ever been found for scale alternatives with such properties as the tests above. On the other hand, construction of optimum rank tests or modification of well-known rank tests have been successfully tried by Capon [2], Dwass [3] and Tamura [8], [9].

We now feel keenly that nonparametric two sample procedure shoud be developed following the line along which the normal theory has been done, that is to say the c-sample problem, multivariate analysis, successive process and etc.

We shall in this paper deal with tests against some alternatives considered in the c-sample problem.

2. Jonckheere's statistic. Let $\{X_{ij}\}$ $i = 1, 2, \cdots\cdots, c$; $j = 1, 2, \cdots\cdots, n_i$ be c samples of size n_i from the populations with continuous distributions $F_i(x)$ and densities $f_i(x)$. Jonckheere [7] has proposed a test statistic J_N to test the hypothesis

$$(1) \quad F_1(X) = F_2(X) = \cdots\cdots = F_c(X)$$

against the ordered alternative

$$(2) \quad F_1(X) > F_2(X) > \cdots\cdots > F_c(X) .$$

Let U_{ij} be Wilcoxon statistics for the i-th and j-th samples

$$(3) \quad U_{ij} = \sum_{ki=1}^{ni} \sum_{kj=1}^{nj} \delta(X_{iki}, X_{jkj})$$

,where $\delta(y_i, y_j)$ is 1 or 0 if $y_i < y_j$ or $y_i > y_j$ holds.
Then J_N is defined as follows

$$(4) \quad J_N = 2\sum_{i<j} U_{ij} - \sum_{i<j} n_i n_j \quad , \quad N = \sum_{i=1}^{c} n_i \quad .$$

The theorical result was only proof of the asymptotic normality of J_N by the methods of moments under the hypothesis. We here add to his result *the fact that its asymptotic normality holds also under the alternative*. The proof may be given by using the same technique as Andrews [1] and it is omitted (see Section 3).

It may be in general too difficult to construct optimum rank tests against the alternative (2). Hence we shall consider for our purpose some narrower alternativer in the following sections.

3. One-parameter alternatives. In this section, we are concerned the slippage and contam-

ination alternatives with only one parameter. We here construct locally most powerful rank tests which are respectively a generalization of the Normal scores and Wilcoxon tests.

(i) Slippage alternative. Assume the following alternative instead of (2)

(5) $F_j(X) = F(X - \delta_j(\theta)) \qquad j = 1, 2, \cdots\cdots, c$

,where the functions $\delta_j(\theta)$ are known and

$$\delta_j(0) = 0 \qquad j = 1, 2, \cdots\cdots, c$$

$$\delta_1(\theta) < \delta_2(\theta) < \cdots\cdots < \delta_c(\theta), \quad \text{for } \theta > 0 .$$

Then our problem is to test the hypothesis $\theta = 0$ agaist the alternative $\theta > 0$. I regret that we don't know whether such alternative is serviciable in practical applications or not.

Now let $R_j^{(i)}$ be the rank of X_{ij} in the combined sample of size N and $R = (R_i^{(1)}, \cdots\cdots, R_{n_c}^{(c)})$. Then the random rank vector R takes on for its values each of the Ni Permutations of $(1, \cdots\cdots, N)$. Let S_k be the set in N-space where R equals a fixed permutation. Denote the probability that the sample point is contained in S_k by $P_\theta(S_k)$. Then it is expressed as

(6) $$P_\theta(S_k) = \int_{S_k}\cdots\int \prod_{i=1}^{c}\prod_{j=1}^{ni} dF_i(x_{ij}) .$$

Or we expand $P_\theta(S_k)$ in power series of θ as

(7) $P_\theta(S_k) = P_0(S_k) + \theta P_0'(S_k) + o(\theta)$

Then under some regularity conditions, we may get

(8) $$\begin{aligned} P_0'(S_k) &= \int_{S_k}\cdots\int \sum_i \sum_j \left(\frac{\partial \log f(x_{ij} - \delta_i(\theta))}{\partial \theta}\right)_{\theta=0} dF(x_{ij}) \\ &= \frac{1}{N!}\sum_i \sum_j \delta_j'(0) E_0\left[-f'(Z_{R_i^{(j)}})/f(Z_{R_i^{(j)}})\right] \end{aligned}$$

,where $Z_1 \leqq Z_2 \leqq \cdots\cdots \leqq Z_N$ is the ordered statistics and E_o expresses the expectation under the hypothesis.

Especially, assuming the following

(9) $f(x - \delta) = N(\delta, \sigma^2)$,

our locally most powerful test procedure is to reject the hypothesis when

(10) $$\sum_i \sum_j \delta_j'(0) E_0(Z_{R_i^{(j)}}) \geqq c_N$$

, where Z_i are the ordered value of N independent normal N(0,1) random variables and the constant c_N is determined by the level of test. The results express apparently a generalization of the Normal scores test in the two sample problem.

(ii) Contamination alternative. We assume the following contamination alternative for $F_j(x)$

(11) $F_j(x) = \theta F_{j-1}^2(x) + (1-\theta) F_{j-1}(x) \qquad j = 1, 2, \cdots\cdots, c$

$$F_0^2(x) = F_0(x) = F(x)$$

and consider the testing hypothesis $\theta=0$ against the alternative $\theta>0$. We may obtain the following locally most powerful rank test by the similar technique as the case of the slippage alternative from (6) and (7)

$$(12)\qquad \sum_{i=1}^{c}\sum_{j=1}^{ni} i\, E_0\left[2F(Z_{R_j}(i))-1\right]$$

or the equivalent statistic

$$(13)\qquad t_N=\frac{1}{(N+1)^2}\sum_{i=1}^{c}\sum_{j=1}^{ni} iR_j^{(i)}.$$

It is also a generalization of Wilcoxon statistic in the two sample problem.

Secondly we shall prove the asymptotic normality of t_N. Now assume that $n_i=ns_i$ with integers n and s_i. Define the statistic

$$(14)\qquad h^{(\alpha)}(X_{1j_1}, X_{2j_2}, \cdots\cdots, X_{cj_c})=\sum_{\beta=1}^{c}\delta(X_{\beta j_\beta}, X_{\alpha j_\alpha})\, s_\alpha s_\beta$$

, then we may get the identity

$$\frac{1}{\prod n_i}\sum_{j_1=1}^{n_1}\cdots\cdots\sum_{j_c=1}^{n_c} h^{(\alpha)}(X_{1j_1},\cdots\cdots,X_{cj_c})=\frac{1}{n^2}\sum_{j=1}^{n_2} R_j^{(\alpha)}-\frac{1}{2}s_\alpha(s_\alpha+1).$$

We now also define U'

$$(15)\qquad U'=\sum_{\alpha=1}^{c}\frac{\alpha}{\prod_{i=1}^{c} n_i}\sum_{j_1=1}^{n_1}\cdots\cdots\sum_{j_c=1}^{n_c} h^{(\alpha)}(X_{1j_1},\cdots\cdots,X_{cj_c})$$

, then it is evident that the statistic t_N is equivalent to U' above.

Construct random variables X_k for $k=1,2,\cdots\cdots,n$ by

$$(X_{1(k-1)s_1+1},\cdots\cdots,X_{1\,ks_1},\cdots\cdots,X_{c(k-1)s_c+1},\cdots\cdots X_{c\,ks_c}).$$

That is, the components of X_k are constructed by combination of groups containing each s_i $(i=1,2,\cdots,c)$ variables from each sample.

After the following procedures

(i) taking out c vectors X_k from the total n vectors above

(ii) taking out each group without belonging to sames ample from each vector in (i) and constructing the statistic $h^{(\alpha)}(X_{1j_1},\cdots\cdots,X_{cj_c})$ such that its argument belongs to different group respectively,

define the following statistic

$$(16)\qquad \varphi(X_{i_1},\cdots\cdots,X_{i_c})=\sum_{\alpha=1}^{c}\frac{\alpha}{c!\prod_{1}^{c}s_i}\sum{}' h^{(\alpha)}(X_{i_1 j_{i_1}},\cdots\cdots,X_{i_c j_{i_c}})$$

where Σ' is the summation over all indeces (j_{i_1}, j_{i_c}) and

$$(17)\qquad U(X_1,\cdots\cdots,X_n)=\frac{1}{\binom{n}{c}}\sum\varphi(X_{i_1},\cdots\cdots,X_{i_c})$$

,where the summation extends over all indeces $1\leqq i_1<\cdots\cdots<i_c\leqq n$.

Then it is easy to show that the stattisic $U(X_1, \cdots\cdots, X_n)$ is what is called *U-statistic* following Hoeffding [6]. In order to prove the asymptotic normality of U', we shall now show by using Andrews's methos that the statistic U and U' have the same limiting distribution as $n \to \infty$.

In fact, define the statistic T by the identity

$$\left(\prod_1^c n_i\right) U' = c!\binom{n}{c}\left(\prod_1^c s_i\right) U + T$$

and T contains the statistic $h(X_{1j_1}, \cdots\cdots, X_{cj_c})$ by $\left(\prod_1^c n_i - c!\binom{n}{c}\prod_1^c s_i\right)$ in number. By taking the expectation, we get

$$\text{(18)} \qquad E\left[n(U' - U)^2\right] \leqq 2n\left(\frac{c!\binom{n}{c}}{n^c} - 1\right) EU^2 + \frac{1}{\left(\prod_1^c n_i\right)^2} ET^2 .$$

On the other hand, it follows from (14)

$$E\left[h^{(\alpha)}(X_{1j_1}, \cdots\cdots, X_{cj_c})^2\right] \leqq \left(\sum_{\beta=1}^c s_\alpha s_\beta\right)^2$$

and from (16) and the inequality above

$$E\varphi^2(X_1, \cdots\cdots, X_c) \leqq c\sum_{\alpha=1}^c \alpha^2\left(\sum_{\beta=1}^c s_\alpha s_\beta\right)^2 .$$

Thus

$$\text{(19)} \qquad EU^2 \leqq c\left(\sum_1^c s_\beta\right)^2\left(\sum_1^c \alpha s_\alpha\right)^2$$

$$\text{(20)} \qquad ET^2 \leqq \left(\prod_1^c n_i - c!\binom{n}{c}\prod_1^c s_i\right)^2\left(\sum_{\beta=1}^c s_\alpha s_\beta\right)^2 .$$

Introducing these inequalities into (18), we establish that

$$\text{(21)} \qquad E\left[n(U' - U)^2\right] \leqq 4n\left(1 - \frac{c!\binom{n}{c}}{n^c}\right)^2 c\left(\sum_1^c s_\beta\right)^2\left(\sum_1^c \alpha s_\alpha\right)^2$$

$$\longrightarrow 0 \quad (\text{as } n \longrightarrow \infty) .$$

Thus we have proved the asymptotic normality of the statistic $\sqrt{n}\,(U' - EU')$.

4. c-parameters alternative. Lastly, we assume $F_j(x)$ to be the form

(22) $F_j(x) = F(x - \phi_j),\ \phi_1 = \theta.$

We consider the testing hypothesis $\phi_j = \theta$ against the alternative $\phi_j > \theta$ for $j \geqq 2$. That is, the alternative is a special form of

$$F_1(x) > F_j(x) \quad , \quad j \geqq 2 .$$

Then it holds under the alternative

$$\text{(23)} \qquad P(S_k) = P_0(S_k) + \sum_{j=1}^c (\phi_j - \theta)\left(\frac{\partial P(S_k)}{\partial \phi_j}\right)_{\phi_j = \theta} + o(|\phi_j - \theta|) .$$

Hence we may obtain, for the values of ϕ_j in the neighbourhood of θ, a most powerful rank test by adopting the following test statistic

(24) $$S_N = \sum_{j=1}^{c} (\phi_j - \theta)\left(\frac{\partial P(S_k)}{\partial \phi_j}\right)\phi_i = \theta \quad .$$

However the statistic S_N depends on the value of ϕ_j, so that we may not construct a uniformly-most powerful test. From the expression (6) and (24), S_N may be written as follows,

(25) $$S_N = \sum_{j=1}^{c} \sum_{i=1}^{j} (\phi_j - \theta) E_0\left[\left(\frac{\partial \log f(Z_{R_i}^{(j)} - \phi_j)}{\partial \phi_j}\right)_{\phi_j = \theta}\right] .$$

We may get in detail, from (25), the generalizations of Capon's results in the two sample problem by assuming various forms of F(x).

REFERENCES

〔1〕 Andrews, F.C.(1954). Asymptotic behaviour of some rank tests for analysis of variance. Ann. Math. Statist. Vol. 25 724-736.

〔2〕 Capon, J. (1961). Asymptotic efficiency of certain locally most powerful rank tests. Ann. Math. Statist. Vol. 32. 88-100.

〔3〕 Dwass, M. (1956). The large-sample power of rank order tests in the two sample problem. Ann. Math. Statist. Vol. 27 352-374.

〔4〕 Hodges, J.L.Jr. and Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t-test. Ann. Math. Statist. Vol. 27 324-335.

〔5〕 Hodges, J.L.Jr. and Lehmann, E.L. (1961). Comparison of the Normal scores and Wilcoxon tests. Proc. Fourth Berkeley Sympos. Math. Stat. and Prob. 307-318.

〔6〕 Hoeffding, W. (1948). A class of statistics with asymptotic normal distribution. Ann. Math. Statist. Vol. 19. 293-325.

〔7〕 Jonckheere, A.R. (1954). A distribution-free k-sample test against ordered alternative. Biometrika Vol. 41 133-143.

〔8〕 Tamura, R. (1963) On a modification of certain rank tests. Ann. Math. Statist. Vol. 34 (1101~1104)

〔9〕 Tamura, R. On a construction of some optimum rank tests. (submitted to Bull. Math. Statist.)