Kumamoto J. Sci. (Math.)
Vol. 14, 34~41 March (1980)

A TEST OF WILCOXON TYPE FOR HOMOGENEITY OF MARGINAL DISTRIBUTIONS AGAINST ORDERED ALTERNATIVES

Ryoji TAMURA

(Received September 10, 1979)

1. Introduction.

Let $\mathbf{X}=(X_1, \cdots, X_p)$ be a p-dimensional random vector with an absolutely continuous distribution function $G(x_1, \cdots, x_p)$ and the marginal distribution function of the i-th component variable X_i be $F_i(x)$ for $i=1, \cdots, p$. Let $\mathbf{X}_\alpha\equiv(X_{1\alpha}, \cdots, X_{p\alpha})$, $\alpha=1, \cdots n$ be a random sample of size n from the p-variate population with the distribution function G where the forms of G and F_i are unknown. This paper is concerned with the problem of testing the hypothesis

$$H_0:\ F_1(x)=\cdots=F_p(x) \text{ for all } x$$

against the ordered alternatives

$$H_1:\ F_1(x)\leqq\cdots\leqq F_p(x)$$

where at least one of the inequalities is strict.

Many workers have investigated the nonparametric p-sample problem where the alternatives are ordered. But, most of them have dealt with this problem under the assumption that the components $X_1, \cdots, X_p$ are mutually independent, that is,

$$G(x_1, \cdots, x_p)=F_1(x_1)\cdots F_p(x_p) \text{ for all } \mathbf{x}=(x_1, \cdots, x_p).$$

For example, see Jonckheere [3], Tamura [6], Chacho [1] and Puri [4]. The investigations for the case where the independence of $X_1, \cdots, X_p$ is violated, can be only found, for $p=2$, in Hollander, Pledger and Lin [2] and Raviv [5]. In this paper, we shall propose a test of Wilcoxon type for this problem where the independence of the component variables is not assumed and $p\geqq 2$. And its asymptotic properties will be investigated and we shall find that the proposed test is not strictly distribution-free, but is asymptotially distribution-free.

In what follows, the summation $\sum_{i<j}$ (or $\sum_{k<l}$) extends over all possible $1\leqq i<$

$j \leqq p$(or $1 \leqq k < l \leqq p$) and $\sum_{\alpha \neq \beta}$ (or $\sum_{\gamma \neq \delta}$) is over all possible $1 \leqq \alpha$, $\beta \leqq n$, $\alpha \neq \beta$ (or $1 \leqq \gamma$, $\delta \leqq n$, $\gamma \neq \delta$).

2. Test statistic and asymptotic distributions.

We first define for $i, j = 1, \cdots, p$ and $\alpha, \beta = 1, \cdots, n$

(2. 1) $$\varphi(X_{i\alpha}, X_{j\beta}) = \begin{cases} 1 & \text{if } X_{i\alpha} > X_{j\beta} \\ 0 & \text{otherwise,} \end{cases}$$

(2. 2) $$W_{ij} = \sum_{\alpha \neq \beta} \varphi(X_{i\alpha}, X_{j\beta}) / n(n-1)$$

and

(2. 3) $$W = \sqrt{n} \sum_{i<j} (W_{ij} - p_{ij}), \text{ where } p_{ij} = \mathrm{p}(X_{i\alpha} > X_{j\beta}) \text{ for } \alpha \neq \beta.$$

Then $\mathrm{E}(W_{ij}) = p_{ij} = \int_{-\infty}^{\infty} F_j(x)\, dF_i(x)$ is easily seen. To derive the variance of W_{ij}, we use the following (2.4) shown by Hollander and others [2] or Raviv [5],

(2. 4) $$\lim_{n\to\infty} \mathrm{E}\Big[\sqrt{n}\,(W_{ij} - p_{ij}) - \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} \{F_j(X_{i\alpha}) + 1 - F_i(X_{j\alpha}) - 2p_{ij}\}\Big]^2 = 0.$$

Then we get (2.5) from both (2.4) and the fact that $F_j(X_{i\alpha}) - F_i(X_{j\alpha})$, $\alpha = 1, \cdots, n$ are independently and identically distributed random variables,

(2. 5) $$\lim_{n\to\infty} \mathrm{Var}\{\sqrt{n}\,(W_{ij} - p_{ij})\} = \mathrm{Var}\{F_j(X_i) - F_i(X_j)\}.$$

We shall now derive the asymptotic distribution of the statistic W.

LEMMA 2.1. *W is asymptotically normally distributed with mean* 0 *and variance A^2 where*

(2. 6) $$A^2 = \mathrm{Var}\,(\sum_{i<j} \{F_j(X_i) - F_i(X_j)\})$$

PROOF. We first get from (2.4) that

$$\sqrt{n}\,(W_{ij} - p_{ij}) - \frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} \{F_j(X_{i\alpha}) + 1 - F_i(X_{j\alpha}) - 2p_{ij}\} \xrightarrow{(P)} 0 \text{ as } n \longrightarrow \infty.$$

Then the following is easily obtained,

(2. 7) $$W - \frac{1}{\sqrt{n}} \sum_{i<j} \sum_{\alpha=1}^{n} \{F_j(X_{i\alpha}) + 1 - F_i(X_{j\alpha}) - 2p_{ij}\} \xrightarrow{(P)} 0 \text{ as } n \longrightarrow \infty.$$

Since $Y_\alpha = \sum_{i<j} \{F_j(X_{i\alpha})+1-F_i(X_{j\alpha})-2p_{ij}\}$, $\alpha=1, \cdots, n$ are independently and identically distributed random variables with mean 0 and variance A^2, the central limit theorem leads the asymptotic normality of the statistic

$$\frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} Y_\alpha = \frac{1}{\sqrt{n}} \sum_{i<j} \sum_{\alpha=1}^{n} \{F_j(X_{i\alpha})+1-F_i(X_{j\alpha})-2p_{ij}\},$$

where the mean and variance are respectively 0 and A^2.
The fact that W has the same asymptotic distribution as $\frac{1}{\sqrt{n}} \sum_{\alpha=1}^{n} Y_\alpha$, proves Lemma 2.1.

Now we shall derive a consistent estimator for A^2 by the similar idea to Raviv [5]. Define for each α,

$$S_\alpha = \sum_{i<j} \sum_{\beta,\beta\neq\alpha} \{\varphi(X_{i\alpha},X_{j\beta})+\varphi(X_{i\beta},X_{j\alpha})\}/(n-1) \tag{2.8}$$

where the summation $\sum_{\beta,\beta\neq\alpha}$ is over all possible $1\leqq\beta\leqq n$, $\beta\neq\alpha$,

$$\hat{A}_n^2 = \sum_{\alpha=1}^{n} (S_\alpha - \bar{S})^2/n, \text{ where } \bar{S} = \sum_{\alpha=1}^{n} S_\alpha/n. \tag{2.9}$$

LEMMA 2.2. *$\hat{A}_n^2$ is a consistent estimator for A^2.*

PROOF. (i) We first show $E(\hat{A}_n^2) \longrightarrow A^2$ as $n \longrightarrow \infty$. From (2.8),

$$E(S_\alpha^2)| = \frac{1}{(n-1)^2} \sum_{i<j} \sum_{k<l} \sum_{\beta,\beta\neq\alpha} \sum_{\gamma,\gamma\neq\alpha} E[\{\varphi(X_{i\alpha}, X_{j\beta})+\varphi(X_{i\beta}, X_{j\alpha})\} \times \{\varphi(X_{k\alpha}, X_{l\gamma})+\varphi(X_{k\gamma}, X_{l\alpha})\}]$$

By using the identity $E\{\varphi(X_{i\alpha}, X_{j\beta})\ \varphi(X_{k\alpha}, X_{l\gamma})\} = E\{F_j(X_i)\ F_l(X_k)\}$ for $\alpha\neq\beta\neq\gamma$, we can get the expression

$$E(S_\alpha^2) = \sum_{i<j} \sum_{k<l} E[\{F_j(X_i)+1-F_i(X_j)\}\{F_l(X_k)+1-F_k(X_l)\}]+O(1/n). \tag{2.10}$$

We may also write as $\bar{S}=2 \sum_{i<j} \sum_{\alpha\neq\beta} \varphi(X_{i\alpha}, X_{j\beta})/n(n-1)$ and we get

$$E(\bar{S}^2) = 4 \sum_{i<j} \sum_{k<l} \sum_{\alpha\neq\beta} \sum_{\gamma\neq\delta} E\{\varphi(X_{i\alpha}, X_{j\beta})\ \varphi(X_{k\gamma}, X_{l\delta})\}/n^2(n-1)^2.$$

Using the identities $E\{\varphi(X_{i\alpha}, X_{j\beta})\ \varphi(X_{k\gamma}, X_{l\delta})\} = p_{ij}p_{kl}$ for $\alpha\neq\beta\neq\gamma\neq\delta$, $i<j$, $k<l$, we can obtain

$$E(\bar{S}^2) = 4 (\sum_{i<j} p_{ij})^2 + O(1/n). \tag{2.11}$$

From (2.10) and (2.11), it holds

$$\lim_{n\to\infty} \mathrm{E}(\hat{A}_n^2) = \sum_{i<j} \sum_{k<l} \mathrm{E}[\{F_j(X_i)+1-F_i(X_j)\}\{F_l(X_k)+1-F_k(X_l)\}] - 4(\sum_{i<j} p_{ij})^2 = A^2.$$

(ii) We next show $\mathrm{Var}(\hat{A}_n^2) \longrightarrow 0$ as $n \longrightarrow \infty$. First we write $\hat{A}_n^2$ as

(2.12) $$\hat{A}_n^2 = \sum_{\alpha=1}^{n} (\sum_{\beta,\beta\neq\alpha} u_{\alpha\beta})^2/n(n-1)^2 - (\sum_{\alpha\neq\beta} u_{\alpha\beta})^2/n^2(n-1)^2,$$

where

$$u_{\alpha\beta} = \sum_{i<j} \{\varphi(X_{i\alpha}, X_{j\beta}) + \varphi(X_{i\beta}, X_{j\alpha})\} \text{ for } \alpha \neq \beta.$$

Then Var $(\hat{A}_n^2)$ is also written as

(2.13) $$\mathrm{Var}(\hat{A}_n^2) = \mathrm{Var}(\sum_{\alpha\neq\beta\neq\gamma} u_{\alpha\beta}u_{\alpha\gamma})/n^2(n-1)^4 + \mathrm{Var}(\sum_{\alpha\neq\beta\neq\gamma\neq\delta} u_{\alpha\beta}u_{\gamma\delta})/n^4(n-1)^4 - 2\mathrm{Cov}(\sum_{\alpha\neq\beta\neq\gamma} u_{\alpha\beta}u_{\gamma\alpha}, \sum_{\alpha\neq\beta\neq\gamma\neq\varrho} u_{\alpha\beta}u_{\gamma\delta})/n^3(n-1)^4 + O(1/n).$$

Now, Var $(\sum_{\alpha\neq\beta\neq\gamma} u_{\alpha\beta}u_{\alpha\gamma})$ is the sum of $n^2(n-1)^2(n-2)^2$ terms and among of them, $6!\binom{n}{6}$ terms such as $\mathrm{E}[\{u_{\alpha\beta}u_{\alpha\gamma} - \mathrm{E}(u_{\alpha\beta}u_{\alpha\gamma})\}\{u_{\alpha'\beta'}u_{\alpha'\gamma'} - \mathrm{E}(u_{\alpha'\beta'}u_{\alpha'\beta'})\}]$ with $\alpha\neq\beta\neq\gamma\neq\alpha'\neq\beta'\neq\gamma'$ become zero. Therefore the number of non-zero terms in Var $(\sum_{\alpha\neq\beta\neq\gamma} u_{\alpha\beta}u_{\alpha\gamma})$ is at most of order n^5. Thus the first term in (2.13) tends to zero as $n \longrightarrow \infty$. Similarly, the other terms tend to zero as $n \longrightarrow \infty$.
The Lemma may be proved from (i) and (ii).

The following theorem is straight from Lemma 2.1 and Lemma 2.2.

THEOREM 2.1. *The asymptotic distribution of* $\sqrt{n}\sum_{i<j}(W_{ij}-p_{ij})/\hat{A}_n$ *is normal with mean* 0 *and variance* 1.

We here propose the test W:

$$\text{If } \sqrt{n}\sum_{i<j}(W_{ij}-\frac{1}{2})/\hat{A}_n \geq z_\alpha, \text{ reject } H_0,$$

where $1-\Phi(z_\alpha)=\alpha$ and $\Phi(x)$ is the standard normal distribution function.
Since $p_{ij}=\frac{1}{2}$ under H_0, it is seen from Theorem 2.1 that the test W has asymptotically level α of significance.

3. Asymptotic properties of the test W.

In this section, we shall restrict ourselves to the translation model as

$$F_i(x)=F(x-\triangle_i), \qquad i=1,\cdots,p. \tag{3. 1}$$

Then the hypothesis and the alternatives are respectively expressed as follows,

$$H_0:\ \triangle_1=\cdots=\triangle_p, \qquad H_1:\ \triangle_1\geqq\cdots\geqq\triangle_p.$$

In addition, we adopt the alternatives H_1^* converging to the hypothesis, in order to discuss the asymptotic power and Pitman relative efficiency,

$$H_1^*:\ \triangle_i=\triangle+\theta_i/\sqrt{n}, \qquad \theta_1=0\geqq\theta_2\geqq\cdots\geqq\theta_p \tag{3. 2}$$

THEOREM 3.1. *Under the model* (3.1) *and the alternatives* (3.2), *the asymptotic power of the test* W *is given by*

$$1-\Phi(z_\alpha-(\int_{-\infty}^{\infty}f^2(x)dx)\sum_{i<j}(\theta_i-\theta_j)/A_0), \tag{3. 3}$$

where $f(x)$ *is the density of* $F(x)$ *and*

$$A_0^2=p(p^2-1)/36+2\sum_{i<j}c_{ij}\,\mathrm{Cov}_0(F(X_i),\ F(X_j)) \tag{3. 4}$$

$$c_{ij}=(p-2i+1)(p-2j+1)$$

and Cov_0 *is the covariance under the hypothesis* H_0.

PROOF. The asymptotic power of the test W is written as

$$\lim_{n\to\infty}P[\sqrt{n}\sum_{i<j}(W_{ij}-\frac{1}{2})/\hat{A}_n\geqq z_\alpha|H_1^*]$$

$$=\lim_{n\to\infty}P[\sqrt{n}\sum_{i<j}(W_{ij}-p_{ij})/A\geqq z_\alpha-\sqrt{n}\sum_{i<j}(p_{ij}-\frac{1}{2})/A|H_1^*], \tag{3. 5}$$

$$=1-\Phi(\lambda_\alpha)$$

where

$$\lambda_\alpha=z_\alpha-\lim_{n\leftarrow\infty}\sqrt{n}\sum_{i<j}(p_{ij}-\frac{1}{2})/A.$$

Since it holds under H_1^*

$$p_{ij}=\frac{1}{2}+(\int_{-\infty}^{\infty}f^2(x)dx)\,(\theta_i-\theta_j)/\sqrt{n}+O(1/n),$$

we get

$$\lim_{n\to\infty} \sqrt{n} \sum_{i<j} (p_{ij} - \frac{1}{2}) = (\int_{-\infty}^{\infty} f^2(x)dx) \sum_{i<j} (\theta_i - \theta_j). \tag{3. 6}$$

To derive $\lim_{n\to\infty} A^2$, we first note that

$$\text{Cov}\ (F_j(X_i),\ F_l(X_k)) = \text{Cov}_0(F(X_i), F(X_k)) + O(1/\sqrt{n})$$

and therefore,

$$\lim_{n\to\infty} A^2 = \text{Var}_0(\sum_{i<j}\{F(X_i) - F(X_j)\}),$$

where Var_0 is the variance under H_0.

In $\text{Var}_0\ (\sum_{i<j} \{F(X_i) - F(X_j)\})$, we can show that the coefficent of $\text{Cov}_0\ (F(X_i),\ F(X_j))$ becomes $(p-2i+1)(p-2j+1)$ for $1 \leqq i \leqq j \leqq p$ after elementary but some tedious calculations Thus we can obtain the following by using $\text{Var}_0\ (F(X_i)) = 1/12$ and $\sum_{i=1}^{p} (p-2i+1)^2 = p(p^2-1)/3$,

$$\begin{aligned} &\text{Var}_0\ (\sum_{i<j} \{F(X_i) - F(X_j)\}) \\ &= p(p^2-1)/36 + 2 \sum_{i<j} c_{ij}\ \text{Cov}_0\ (FX_i),\ F(X_j)) \\ &= A_0^2 \end{aligned} \tag{3. 7}$$

The relations (3.4), (3.5) and (3.7) prove the Theorem 3.1.

We are now in a position to make large sample comparision between the test W and the normal theory competitor based on Student statistic,

$$T = \sqrt{n} \sum_{i<j} (\bar{X}_i - \bar{X}_j),\ \bar{X}_i = \sum_{\alpha=1}^{n} X_{i\alpha}/n \text{ for } i = 1,\ \cdots,\ p. \tag{3. 8}$$

Then the procedure of the test T is as follows,

if $T/\sqrt{\text{Var}_0(T)} \geqq z_\alpha$, reject H_0.

THEOREM 3.2. *The Pitman's asymptotic relative efficiency of the test W relative to the test T is the following,*

$$e_{W,T} = B_0^2 (\int_{-\infty}^{\infty} f^2(x)dx)^2 / A_0^2, \tag{3. 9}$$

$$B_0^2 = p(p^2-1)\sigma^2/3 + 2 \sum_{i<j} c_{ij}\ \text{Cov}_0\ (X_i, X_j), \tag{3.10}$$

where σ^2 *is the variance of* X_i *and* c_{ij} *is given in* (3.4).

PROOF. *We first show the asymptotic normality of* T. *We represent* T as $T=\sum_{\alpha=1}^{n}(\sum_{i<j}(X_{i\alpha}-X_{j\alpha}))/\sqrt{n}$. Since $Z_\alpha=\sum_{i<j}(X_{i\alpha}-X_{j\alpha})$, $\alpha=1,\cdots,n$ are independently and identically distributed random variables, we get the asymptotic normality of $T=\sum_{\alpha=1}^{n} Z_\alpha/\sqrt{n}$ by the central limit theorem. It is evident that $\mathrm{E}(T)=\sum_{i<j}(\theta_i-\theta_j)$ or 0 under H_1^* or H_0 and $\mathrm{Cov}_0\ (\sqrt{n}\,\bar{X}_i, \sqrt{n}\bar{X}_j)=\mathrm{Cov}_0\ (X_i,X_j)$. By the similar discussions to A_0^2 in Theorem 3.1, we obtain

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Var}(T)&=\mathrm{Var}_0\ (T)\\ &=p(p^2-1)\sigma^2/3+2\sum_{i<j} c_{ij}\ \mathrm{Cov}_0\ (X_i,X_j)\end{aligned}$$

Thus the asymptotic power of the test T is given by

$$1-\Phi(z_\alpha-\sum_{i<j}(\theta_i-\theta_j)/B_0). \tag{3.11}$$

From Theorem 3.1 and (3.11), we can get (3.9) as the asymptotic efficiency of the test W relative to the test T.

Finally, we consider the model with more strong restriction than (3.1), that is, we assume in addition to (3.1) that

$$\mathrm{P}[X_i\leqq x,\ X_j\leqq y]=K(x-\triangle_i,\ y-\triangle_j) \text{ for } i,j=1,\ \cdots,\ p. \tag{3.12}$$

We here note that the multivariate normal distrbution with covariance matrix $\|\sigma_{ij}\|$ where $\sigma_{ij}=\sigma^2$ or $\rho\sigma^2$ for $i=j$ or $i\neq j$ is a special case of the model above. Then we get

COROLLARY 3.1. *Under the model* (3.1), (3.2) *and* (3.12), *the Pitman relative efficiency* $e_{W,T}^*$ *of the test* W *relative to the test* T *is given by*

$$e_{W,T}^*=12\sigma^2\ (\int_{-\infty}^{\infty} f^2(x)dx)^2\ (1-\rho_0\,(X_i,X_j))/(1-\rho_0(F(X_i),\ F(X_j))), \tag{3.13}$$

where ρ_0 *is the correlation coefficent under* H_0.

PROOF. Under the model above, $\mathrm{Cov}_0\ (F(X_i),\ F(X_j))$ and $\mathrm{Cov}_0\ (X_i,X_j)$ have respectivly the same values fos all $1\leqq i<j\leqq p$ and moreover we get $2\sum_{i<j} c_{ij}=p(1-p^2)/3$. Hence, we get

$$A_0^2=p(p^2-1)\,(1-\rho_0(F(X_i),\ F(X_j)))/36 \text{ and } B_0^2=p(p^2-1)\sigma^2(1-\rho_0\,(X_i,\ X_j))/3.$$

These show (3.13).

It should be noted that the form (3.13) of $e^*_{W,T}$ consists with that in Raviv [5] where the problem has been dealt for $p=2$.

References

[1] Chacko, V. J.: Testing homogeneity against ordered aternatives. Ann. Math Statist. **34** (1963) 945-956.

[2] Hollander, M., Pledger, G. and Lin, P. E.: Robustness of the Wilcoxon test to a certain dependency between samples. Ann. Statist. **2** (1974) 177-181.

[3] Jonckheere, A. R.: A distribution-free k-sample test against ordered alternatives. Biometrika **41** (1954) 135-145.

[4] Puri, M. L.: Some distribution-free k-sample rank tests of homogeneity against ordered alternatives. Commu. Pure and Applied Math. **18** (1965) 51-63.

[5] Raviv, A.: A non-parametric test for comparing two non-independent distributions. J. Roy. Statist. Soc. B **40** (1978) 253-261.

[9] Tamura, R.: On certain c-sample problem against ordered alternatives. Bull. Shimane Univ. (Natural Science) **13** (1964) 123-127.

Department of Mathematics
Faculty of Science
Kumamoto University