The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1455–1460

ON A c-SAMPLE TEST BASED ON TRIMMED SAMPLES

BY RYOJI TAMURA[1]

Kyushu Institute of Design

Let $X_{i1} < \cdots < X_{in_i} (i = 1, \cdots, c)$ be the order statistics from an absolutely continuous cdf $F_i(x) = F(x-\theta_i)$ where $F(x)$ has symmetric density. The problem of testing the hypothesis $H_0: \theta_1 = \cdots = \theta_c$, which has been discussed by many authors, will be considered in this paper. We are concerned with the tests based on only the middle $n_i - 2k_i$ random variables $X_{ik_i+1} < \cdots < X_{in_i-k_i}$ where $k_i = [n_i\alpha]$ is the largest integer not exceeding $n_i\alpha$ for any α, $0 < \alpha < \frac{1}{2}$.

A test of Bhapkar's type [Bhapkar, V. P. (1961). A nonparametric test for the problem of several samples. *Ann. Math. Statist.* **32** 1108–1117] is proposed for this problem and it is shown that, for some distributions with heavy tails, the asymptotic relative efficiency of the proposed test relative to Bhapkar's test, which is based on the complete samples, is larger than one. The work presented in this paper is an attempt toward generalizing Hettmansperger's results [Hettmansperger, T. P. (1968). On the trimmed Mann–Whitney statistics. *Ann. Math. Statist.* **39** 1610–1614] to the c-sample problem.

1. Introduction. Hettmansperger [4] has recently shown that we can increase the asymptotic relative efficiency in Pitman's sense of the Mann–Whitney test for some distributions with heavy tails by using the trimmed samples instead of the complete samples. In this paper, we have an attempt toward generalizing his results to Bhapkar's test [3] for the c-sample problem. We here consider the following c-sample problem. Let $X_{i1} < \cdots < X_{in_i}$ be the order statistics from absolutely continuous distributions $F_i(x) = F(x-\theta_i)$, $i = 1, \cdots, c$, where $F(x)$ has symmetric density $f(x)$ of unknown functional form. We further assume for $0 < \alpha < \frac{1}{2}$ that (x) is continuously differentiable in some neighborhood of the unique population quantiles b_α and $b_{1-\alpha}$ of order α and $1-\alpha$, respectively. The hypothesis H_0, to be tested, is specified by $\theta_1 = \cdots = \theta_c$ against the alternative that not all θ's are equal.

For this problem, the test statistic V_α, which includes Bhapkar's as a special case, will be proposed on the basis of only the middle $n_i - 2k_i$ random variables $X_{ik_i+1} < \cdots < X_{in_i-k_i}$, $i = 1, \cdots, c$, where $k_i = [n_i\alpha]$ denotes the largest integer not exceeding $n_i\alpha$. We again emphasize that we are concerned with only the tests based on trimmed samples. Some definitions are given in Section 2. In Section 3 we derive the asymptotic distributions and obtain the asymptotic relative efficiency of the proposed test V_α with respect to Bhapkar's test. Some examples for distributions with heavy tails such as the logistic, double exponential and Cauchy are given in Section 4.

2. Some definitions. Let us define for $i = 1, \ldots, c$

$$(2.1) \qquad U_\alpha^{(i)} = \prod_{j=1}^{c} (n_j - 2k_j)^{-1} \sum_{\beta_1=k_1+1}^{n_1-k_1} \cdots \sum_{\beta_c=k_c+1}^{n_c-k_c} \varphi^{(i)}(X_{1\beta_1}, \cdots, X_{c\beta_c});$$

Received August 11, 1970.

[1]Now at Kumamoto University.

$$\text{(2.2)}\qquad \varphi^{(i)}(x_1,\cdots,x_c) = 1 \quad \text{if } x_i > x_j \text{ for all } j = 1,\cdots,c \text{ except } i,$$
$$= 0 \quad \text{otherwise.}$$

Further we define for $i = 1, \cdots, c$

$$\text{(2.3)}\qquad Y_{i1} = n_i^{\frac{1}{2}}(X_{ik_i+1} - b_\alpha - \theta_i), \qquad Y_{i2} = n_i^{\frac{1}{2}}(X_{in_i-k_i} - b_{1-\alpha} - \theta_i)$$
$$\mathbf{Y} = (Y_{11}, Y_{12}, \cdots, Y_{c1}, Y_{c2}).$$

We here notice that the statistic $U_\alpha^{(j)}$, given $\mathbf{Y}$, is a generalized U-statistic (see [5]) based on sample size $n_j - 2k_j$ from distribution with density $g_j(x)$, $j = 1, \cdots, c$,

$$\text{(2.4)}\qquad g_j(x) = f(x-\theta_j)/[F(b_{1-\alpha} + Y_{j2}/n_j^{\frac{1}{2}}) - F(b_\alpha + Y_{j1}/n_j^{\frac{1}{2}})]$$
$$\text{for} \quad b_\alpha + \theta_j + Y_{j1}/n_j^{\frac{1}{2}} < x < b_{1-\alpha} + \theta_j + Y_{j2}/n_j^{\frac{1}{2}}$$
$$= 0 \quad \text{otherwise.}$$

Finally, we define for $i = 1, \ldots, c$

$$\text{(2.5)}\qquad R_\alpha^{(i)} = (N-2k)^{\frac{1}{2}}[U_\alpha^{(i)} - E(U_\alpha^{(i)} \mid \mathbf{Y})], \qquad \mathbf{R}_\alpha = R_\alpha^{(1)}, \cdots, R_\alpha^{(c)}),$$

$$\text{(2.6)}\qquad W_\alpha^{(i)} = (N-2k)^{\frac{1}{2}}(U_\alpha^{(i)} - 1/c), \qquad \mathbf{W}_\alpha = (W_\alpha^{(1)}, \cdots, W_\alpha^{(c)}),$$

where $E(* \mid \mathbf{Y})$ is the expected value of the statistic $*$, given $\mathbf{Y}$, and $N = \sum_{j=1}^c n_j$, $k = \sum_{j=1}^c k_j$.

We assume throughout this paper that the sample sizes n_j, $j = 1, \cdots, c$ increase in such a way that $\lim_{N\to\infty} n_j/N = \lambda_j$, $0 < \lambda_j < 1$.

3. Asymptotic distributions. Now we shall consider the asymptotic distributions of the proposed statistics under the hypothesis H_0 and the following sequence of alternatives

$$\text{(3.1)}\qquad H_N: F_i(x) = F(x - v_i/N^{\frac{1}{2}}), \qquad i = 1, \cdots, c$$

where not all v's are equal.

LEMMA 3.1. *Under the sequence of alternatives H_N, the conditional mean vector and covariance matrix of $\mathbf{W}_\alpha$, given $\mathbf{Y}$, are asymptotically given by $\boldsymbol{\mu}(\mathbf{Y}) = (\mu_1(\mathbf{Y}), \cdots, \mu_c(\mathbf{Y}))$ and $\Omega = \|\omega^{(i,j)}\|$, respectively, if there exists a function $g(x)$ such that for any x and any sufficiently small h,*

$$\text{(3.2)}\qquad |[f(x+h) - f(x)]/h| \leqq g(x), \qquad \int_{-\infty}^{\infty} g(x)\, dF(x) < \infty$$

where

$$\text{(3.3)}\qquad \mu_i(\mathbf{Y}) = \mu_i + f(b_\alpha)[Z_i - (c-1)^{-1} \textstyle\sum_{j\neq i} Z_j]/c(1-2\alpha)^{\frac{1}{2}},$$

$$\text{(3.4)}\qquad \mu_i = (1-2\alpha)^{\frac{1}{2}-c} \int_{b_\alpha}^{b_{1-\alpha}} [F(x) - \alpha]^{c-2} f(x)\, dF(x) \textstyle\sum_{j=1}^c (v_i - v_j),$$
$$Z_i = \lambda_i^{-\frac{1}{2}} Y_{i1} + (c-1) Y_{i2},$$

$$\text{(3.5)}\qquad \omega^{(i,j)} = (\textstyle\sum_{k=1}^c \lambda_k^{-1} + c^2 \delta_{ij} \lambda_i^{-1} - c\lambda_i^{-1} - c\lambda_j^{-1})/c^2(2c-1),$$
$$\delta_{ij} = 1 \text{ or } 0 \text{ for } i = j \text{ or } i \neq j.$$

PROOF. First we get

$$E[\varphi^{(i)}(X_1, \cdots, X_c) \mid \mathbf{Y}] = \prod_{j=1}^{c} [F(b_{1-\alpha}+Y_{j2}/n_j^{\frac{1}{2}})-F(b_\alpha+Y_{j1}/n_j^{\frac{1}{2}})]^{-1} \times \int_{b_\alpha+Y_{i1}/n_i^{1/2}}^{b_{1-\alpha}+Y_{i2}/n_i^{1/2}} \prod_{j\neq i}[F(t+\theta_i-\theta_j)-F(b_\alpha+Y_{j1}/n_j^{\frac{1}{2}})]\, dF(t).$$

By expanding in a Taylor series and using the assumption (3.2), it follows that

$$(3.6)\quad E(\varphi^{(i)} \mid \mathbf{Y}) = c^{-1}+\int_{b_\alpha}^{b_{1-\alpha}}[F(x)-\alpha]^{c-2}f(x)\,dF(x)\textstyle\sum_{j\neq i}(v_i-v_j)/N^{\frac{1}{2}}(1-2\alpha)^c + f(b_\alpha)[Z_i-(c-1)^{-1}\sum_{j\neq i}Z_j]/cN^{\frac{1}{2}}(1-2\alpha)+O(N^{-1}).$$

Thus we get (3.3) by noticing the definitions (2.1) and (2.6). Secondly, we calculate the conditional covariance $\omega^{(ij)}$, given $\mathbf{Y}$. In this case, we get from the theory of the generalised U-statistics,

$$(3.7)\qquad \omega^{(ij)} = \lambda_1^{-1}\rho_{10\cdots 0}^{(ij)}+\cdots+\lambda_c^{-1}\rho_{0\cdots 01}^{(ij)}$$

where $\rho_{0\cdots 010\cdots 0}^{(ij)}$ (1 lies at the kth place) is the covariance of $\varphi^{(i)}(X_1, \cdots, X_k, \cdots, X_c)$ and $\varphi^{(j)}(X_1', \cdots, X_{k-1}', X_k, X_{k+1}', \cdots, X_c')$ and X_j and X_j' are independent and identically distributed as (2.4) for each j. After some calculations, we get

$$(3.8)\quad \rho_{0\cdots 010\cdots 0}^{(ij)} \text{ (1 lies at the } k\text{th place)} = (1-c\delta_{ik}-c\delta_{jk}+c^2\delta_{ij}\delta_{jk}\delta_{ki})/c^2(2c-1)+O(N^{-\frac{1}{2}}).$$

The identity (3.5) follows from (3.7) and (3.8). It is also concluded from (3.5) that $\omega^{(ij)}$ is asymptotically independent of $\mathbf{Y}$.

THEOREM 3.1. *The random vector* $\mathbf{W}_\alpha$ *has the joint asymptotic normal distribution* $N(\mathbf{0}, \mathbf{\Sigma}_\alpha)$ *or* $N(\boldsymbol{\mu}, \mathbf{\Sigma}_\alpha)$ *under the hypothesis* H_0 *or the alternatives* H_N *if the assumption* (3.2) *holds, where* $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_c)$ *and* $\mathbf{\Sigma}_\alpha = \|\sigma_\alpha^{(ij)}\|$

$$(3.9)\quad \sigma_\alpha^{(ij)} = \beta_c\omega^{(ij)}/(c-1)^2(1-2\alpha),$$

$$\beta_c = (c-1)^2(1-2\alpha)+(2c-1)(c^2-2c+2)\alpha(1-\alpha)+2(c-1)(2c-1)\alpha^2.$$

PROOF. We easily have $E(W_\alpha^{(i)}) = \mu_i+O(N^{-\frac{1}{2}})$ from (3.3) using the well-known fact that $\mathbf{Y}$ is asymptotically distributed as $N(\mathbf{0}, \mathbf{\Pi})$ where

$$(3.10)\qquad \mathbf{\Pi} = \begin{Vmatrix} \mathbf{\Pi}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{\Pi}_1 \end{Vmatrix}, \qquad \mathbf{\Pi}_1 = f(b_\alpha)^{-2}\begin{Vmatrix} \alpha(1-\alpha) & \alpha^2 \\ \alpha^2 & \alpha(1-\alpha) \end{Vmatrix}.$$

Now the identities

$$W_\alpha^{(i)} = R_\alpha^{(i)}+(N-2k)^{\frac{1}{2}}[E(U_\alpha^{(i)} \mid \mathbf{Y})-1/c], \qquad i = 1, \cdots, c$$

lead to the following expression

$$(3.11)\quad \sigma_\alpha^{(ij)} = \operatorname{Cov}(R_\alpha^{(i)}, R_\alpha^{(j)})+(N-2k)\operatorname{Cov}[E(U_\alpha^{(i)} \mid \mathbf{Y}), E(U_\alpha^{(j)} \mid \mathbf{Y})].$$

The second term of the right-hand side may be calculated from (3.3) and (3.10) as

$$(c^2-2c+2)\alpha(1-\alpha)+2\alpha^2(c-1)(\textstyle\sum_{k=1}^{c}\lambda_k^{-1}+c^2\delta_{ij}\lambda_i^{-1}-c\lambda_i^{-1}-c\lambda_j^{-1}) \div c^2(c-1)(1-2\alpha).$$

Thus we have (3.9) from (3.5), (3.10) and (3.11). It remains to show the asymptotic normality of $\mathbf{W}_\alpha$. First the asymptotic normality of $(\mathbf{Y}, \mathbf{R}_\alpha)$ follows by Theorem 2 of Sethuraman [6]. Secondly, for any constant vector $\mathbf{a}' = (a_1, \cdots, a_c)$, the expression

$$\mathbf{a}'\mathbf{W}_\alpha = \mathbf{a}'\mathbf{R}_\alpha + (N-2k)^{\frac{1}{2}} \sum_{i=1}^{c} a_i[E(U_\alpha^{(i)} \mid \mathbf{Y}) - 1/c] \tag{3.12}$$

is a function of $\mathbf{Y}$ and $\mathbf{R}_\alpha$. The asymptotic normality of $\mathbf{a}'\mathbf{W}_\alpha$ may be derived by an application of the theorem of Anderson [1], page 76, for (3.12) and that of $\mathbf{W}_\alpha$ may be established from this fact.

We here define the test statistic V_α,

$$V_\alpha = \mathbf{W}_\alpha' \mathbf{\Lambda} \mathbf{W}_\alpha \tag{3.13}$$

where

$$\mathbf{\Lambda} = (c-1)^2(2c-1)(1-2\alpha)\beta_c^{-1} \|\delta_{ij}\lambda_i - \lambda_i\lambda_j\|, \qquad i,j = 1, \cdots, c. \tag{3.14}$$

THEOREM 3.2. *The test statistic V_α is asymptotically distributed as χ^2_{c-1} with $c-1$ degree of freedom under H_0 and as noncentral $\chi^2_{c-1}(\delta)$ with $c-1$ degree of freedom and the noncentrality parameter δ if the assumption* (3.2) *holds, where*

$$\delta = c^2(c-1)^2(2c-1)(1-2\alpha)^{2-2c}\beta_c^{-1}\left(\sum_{i=1}^{c} \lambda_i(v_i - \bar{v})^2\right) \times \left[\int_{b_\alpha}^{b_{1-\alpha}} [F(x)-\alpha]^{c-2} f(x)\, dF(x)\right]^2, \qquad \bar{v} = \sum_{i=1}^{c} \lambda_i v_i. \tag{3.15}$$

The proof is essentially the same as that of Bhapkar [3] or Sugiura [7] and is omitted. We here notice that Bhapkar's test may be denoted by V_0.

It has been shown by Andrews [2] that the Pitman efficiency is given by the ratio of the noncentrality parameter of the V_α to that of the V_0 in their asymptotic distributions. From (3.15), we have the asymptotic relative efficiency of the test V_α with respect to the test V_0 as

$$e_c(\alpha) = (c-1)^2(1-2\alpha)^{2-2c}\beta_c^{-1}\left[\int_{b_\alpha}^{b_{1-\alpha}} [F(x)-\alpha]^{c-2} f(x)\, dF(x)\right]^2 \div \left[\int_{-\infty}^{\infty} F(x)^{c-2} f(x)\, dF(x)\right]^2. \tag{3.16}$$

4. Examples. We first express $e_c(\alpha)$ as a function of c for some distributions with heavier tails than the normal distribution and we compute the numerical values of $e_c(\alpha)$ for $c = 2, 3, 4$ and 5. After some calculations, we get the following expressions.

(a) Logistic distribution, $f(x) = \exp(-x)/(1+\exp(-x))^2$, $-\infty < x < \infty$.

$$e_c(\alpha) = \beta_c^{-1}[(c-)1(1-2\alpha)^2 + c(c+1)\alpha(1-\alpha)]^2.$$

(b) Double exponential distribution, $f(x) = \exp(-|x|)/2$, $-\infty < x < \infty$.

$$e_c(\alpha) = (c-1)^2(2^{c-1}-1)^{-2}\beta_c^{-1}[2^{c-1}(1-2\alpha+c\alpha) - (1-2\alpha)]^2.$$

(c) Cauchy distribution, $f(x) = 1/\pi(1+x^2)$, $-\infty < x < \infty$.

$$e_c(\alpha) = (c-1)^2(1-2\alpha)^{2-2c}\beta_c^{-1}[\{(1-2\alpha)^{c-1} + (c-1)J_c(\alpha)\} \div \{1+(c-1)J_c(0)\}]^2$$

where

$$J_c(\alpha) = (2\pi)^{-1}(1-2\alpha)^{c-2}\sin(\gamma)+(2\pi)^{-2}(c-2)(1-2\alpha)^{c-3}\cos(\gamma)$$
$$-(2\pi)^{-2}(c-2)(c-3)J_{c-2}(\alpha), \qquad c \geqq 4,$$
$$J_2(\alpha) = (\pi)^{-1}\sin(\gamma)$$
$$J_3(\alpha) = (2\pi)^{-1}(1-2\alpha)\sin(\gamma), \qquad \gamma = \pi(1-2\alpha_{,},$$

and $J_c(0)$ is the value of $J_c(\alpha)$ when $\alpha = 0$. The numerical values of $e_c(\alpha)$ for $c = 2, 3, 4$ and 5 are shown in the following tables.

TABLE 1
Logistic distribution

α / c	.1	.2	.25	.3	.35	.4	.45
2	.995	.968	.945	.917	.882	.842	.798
3	.986	.968	.953	.932	.807	.876	.840
4	.994	1.002	1	.992	.977	.956	.928
5	1.014	1.054	1.066	1.070	1.066	1.053	1.031

TABLE 2
Double exponential distribution

α / c	.1	.2	.25	.3	.35	.4	.45
2	1.029	1.089	1.125	1.164	1.204	1.246	1.289
3	1.019	1.089	1.134	1.184	1.238	1.296	1.357
4	1.022	1.129	1.200	1.279	1.367	1.463	1.567
5	1.042	1.203	1.306	1.422	1.551	1.693	1.850

TABLE 3
Cauchy distribution

α / c	.1	.2	.25	.3	.35	.4	.45
2	1.089	1.258	1.339	1.403	1.435	1.441	1.405
3	1.078	1.258	1.350	1.427	1.475	1.498	1.480
4	1.094	1.368	1.519	1.659	1.731	1.827	1.869
5	1.144	1.562	1.802	2.029	2.225	2.343	2.393

Acknowledgment. The author is grateful to the referee for his useful comments on this paper.

REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] Andrews, F. C. (1954). Asymptotic behavior of some rank tests for analysis of variance. *Ann. Math. Statist.* **25** 724–736.

[3] Bhapkar, V. P. (1961). A nonparametric test for the problem of several samples. *Ann. Math. Statist.* **32** 1108–1117.

[4] Hettmansperger, T. P. (1968). On the trimmed Mann–Whitney statistic. *Ann. Math. Statist.* **39** 1610–1614.

[5] Lehmann, E. L. (1951). Consistency and unbiasedness of certain nonparametric tests. *Ann. Math. Statist.* **22** 165–179.

[6] Sethuraman, J. (1961). Some limit theorems for joint distributions. *Sankhyā Ser. A* **23** 379–386.

[7] Sugiura, N. (1965). Multisample and multivariate nonparametric tests based on *U* statistics and their asymptotic efficiencies. *Osaka J. Math.* **2** 385–426.