ON WEIGHTED RANK-SUM TESTS FOR DISPERSION

By P. K. SEN

(Received Feb. 11, 1963)

Summary

Here is proposed a class of two sample non-parametric scale tests (for both the situations, viz., with and without assuming the identity of the locations) based on a family of congruent inter-quantile numbers, and their various properties have been studied. A class of appropriate weight-functions has also been proposed here, and the corresponding weighted rank-sum tests appear to be asymptotically more (power-) efficient, in many cases, than most of the other ones available in the literature.

1. Introduction

Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be two independent samples of m and n independent observations from two populations with continuous cumulative distribution functions (cdf's) $F_1(x)$ and $F_2(x)$ respectively, where

$$F_i(x)=F([x-\mu_i]/\delta_i) \qquad \text{for } i=1,\ 2\,; \tag{1.1}$$

μ_i ane δ_i being respectively the location and the scale parameters of the i-th population, for $i=1$, 2. We are interested in testing the null hypothesis $H_0: \delta_1=\delta_2$ against the set of alternatives that they are not equal (alternatives being one sided or both sided). And, in testing this hypothesis, we are faced with two different situations, according as we assume the homogeneity of the locations μ_1 and μ_2, or we want to test for the identity of δ_1 and δ_2 without any regard to the possible heterogeneity of μ_1 and μ_2. In the first case, various tests have been proposed by Lehmann [10], Wilks-Rosenbaum [13], Mood [12], Kamat [8], Barton and David [2], Sukhatme ([18], [20]), Tamura [21], Ansari and Bradley [1], Siegel and Tukey [17], Capon [4], Klotz [9], among others. Location-free scale-tests has mainly been studied by Sukhatme [19], while Tamura, and Ansari and Bradley have also considered such tests. A discussion of the relative properties of some of these tests, along with some other ones, has also been made by the author [15], elsewhere. Here the following type of tests has been proposed, their properties studied, and contrasted with those of the other ones, referred to above.

Let $X_{(1)}<\cdots<X_{(m)}$ be the first sample ordered variables. We then define a system of (first sample) congruent inter-quantile cells as

(1.2) $I_j:\ X_{(j)} \leqq x \leqq X_{(m-j+1)}$ for $j=0, 1, \cdots, [m/2]$;

where conventionally $X_{(0)}=-\infty$ and $X_{(m+1)}=\infty$, and where $[s]$ stands for the largest integer contained in s. Let us also define the corresponding ring-cells as

(1.3) $L_j:\ X_{(j)} \leqq x < X_{(j+1)}$; $\quad X_{(m-j)} < x \leqq X_{(m-j+1)}$

i.e., $L_j \equiv I_j - I_{j+1}$, for $j=0, 1, \cdots, [m/2]-1$,

and

$$L_{[m/2]} = I_{[m/2]} .$$

Let now in the second sample, r_j observations belong to the ring-cell L_j, and R_j observations to the cell I_j, so that $R_j=\sum_{i=j}^{[m/2]} r_{ji}$ for $j=0, \cdots, [m/2]$ and $R_0=n$. Let us also select a sequence $\{a(j, m)\}$ or real and finite elements, satisfying the following conditions:

(c.1) $$|a(j, m)| \leqq K\{(j+1)(m-j)/m^2\}^{\delta-1/2} ,$$

for some $\delta>0$, and for all $j=0, \cdots, [m/2]$. Let then

(1.4)
$$\bar{a}(m)=\sum_{j=0}^{[m/2]} a(j, m)/\{[m/2]+1\} , \quad \text{and}$$
$$\sigma^2_{a(m)}=\frac{1}{[m/2]+1}\sum_{j=0}^{[m/2]} a^2(j, m)-[\bar{a}(m)]^2 .$$

It then follows that

$$\frac{1}{[m/2]+1}\sum_{j=0}^{[m/2]} a^2(j, m) \leqq K^2 \sum_{j=0}^{[m/2]} \left(\frac{j+1}{m}\right)^{2\delta-1} /\{[m/2]+1\} ,$$

and as

$$\lim_{n=\infty} \sum_{j=0}^{n} (j/n)^{2\delta-1}/n = \int_0^1 x^{2\delta-1}dx = \frac{1}{2\delta} < \infty ,$$

it follows that $\bar{a}(m)$ and $\sigma^2_{a(m)}$ have finite limits as $m\to\infty$. We thus write

(1.5) $$\bar{a} = \lim_{m=\infty} \bar{a}(m) \quad \text{and} \quad \sigma^2_a = \lim_{m=\infty} \sigma^2_{a(m)} ,$$

and assume further that

(1.6) $$\sigma^2_{a(m)} > 0 \ \text{ for all } m, \text{ and } \ \sigma^2_a > 0 .$$

Naturally, (1.6) implies that $a(j, m)$ is not a constant.

(c.2) If we write $F=j/m$, then $a(j, m)(=a(F, m))$ is an explicit function of F and m, and for any given m, $a(F, m)$ is a strictly monotonic function of F. This condition, as would be shown later on, is required only for the scale tests of this type.

(c.3) $a(F, m)$ has a continuous first order derivative (with respect to F) $G(F, m)$, which is finite for all F: $0<F<1$, and let

$$G(F)=\lim_{m=\infty} G(F, m) . \tag{1.7}$$

We then assume that $G(F)$ is a continuous and integrable function of F.

Then our proposed test is based on

$$S_N=n^{-1}\sum_{j=0}^{[m/2]} a(j, m)r_j , \tag{1.8}$$

where $N=m+n$. Also, if we write $b(j, m)=a(j, m)-a(j-1, m)$ for $j=1, \cdots, [m/2]$, and $b(0, m)=a(0, m)$, we can then equivalently write

$$S_N=n^{-1}\sum_{j=0}^{[m/2]} b(j, m)R_j . \tag{1.9}$$

In this paper, we have considered first, various properties of the S_N-test for arbitrary sequences $\{a(j, m)\}$ of weight-functions, satisfying the conditions (c.1), (c.2) and (c.3). Subsequently, we have proposed a class of such weight-functions, and have shown that the corresponding S_N-test, in many cases, is asymptotinally more (power-) efficient than most of the other ones, referred to earlier. Now, the procedure sketched so far, is obviously based upon the assumed identity of the locations μ_1 and μ_2. In order to make the tests location-free, we, following Sukhatme [19], centre the sample observations at the respective sample medians, and investigate the regularity conditions under which the S_N-test based on these centred variables would be asymptotically distribution-free. These conditions are essentially the same as those of Sukhatme [19].

2. Null distribution and moments of S_N

Let us first consider the joint distribution of $r_0, \cdots, r_{[m/2]-1}$ under the null hypothesis H_0: $F_1(x)\equiv F_2(x)$ (i.e., $\mu_1=\mu_2$ and $\delta_1=\delta_2$ in (1.1)). It then follows that if *m be even*, this joint distribution, after some simplification, reduces to

$$\binom{m+n}{m}^{-1}\prod_{j=0}^{(m-2)/2}(r_j+1) , \tag{2.1}$$

where $r_j \geqq 0$ for all j, and $\sum_{j=0}^{m/2} r_j = n$. If on the other hand, *m be odd*, the joint distribution of $r_0, \cdots, r_{(m-3)/2}$ comes out as

$$\binom{m+n}{m}^{-1} \prod_{j=0}^{(m-1)/2} (r_j+1) , \tag{2.2}$$

where $r_j \geqq 0$ for all j, and $\sum_{j=0}^{(m-1)/2} r_j = n$.

Now, we can always enumerate the set of points $(r) = ((r_0, \cdots, r_{[m/2]}))$ for which $S_N = n^{-1} \sum_{j=0}^{[m/2]} a(j, m) r_j \leqq S_0$ and the probability of each element r in this set can be evaluated by using (2.1) or (2.2), according as m is even or odd. Thus we get the values of $P\{S_N \leqq S_0 | H_0\}$ for any S_0. This procedure yields the null distribution of S_N. However, for large samples (even for moderately large ones) the labour in this computation increases so much, that one is naturally inclined to adopt some simple limiting form of distribution, for the purpose, and the same has been accomplished in the following section. But before that, we consider here the first two moments of S_N under H_0.

It follows by a straightforward computation that for *an odd m*,

$$E\{ \prod_{i=1}^{k} r_j^{[s_i]} | H_0\} = \{n^{[s]}/(m+s)^{[s]}\} \prod_{i=1}^{k} (s_i+1)! \tag{2.3}$$

where $p^{[q]} = p(p-1)\cdots(p-q+1)$, $s_1, \cdots, s_k$ are non-negative integers, $k \geqq 1$, and $s = \sum_{i=1}^{k} s_i$. Whence, we get

$$\begin{aligned} & E\{r_j | H_0\} = 2n/(m+1) , \\ & V\{r_j | H_0\} = 2n(m-1)(n+m+1)/(m+1)^2(m+2) , \quad \text{and} \\ & Cov_{j \neq j'} \{r_j, r_{j'} | H_0\} = -4n(n+m+1)/(m+1)^2(m+2) . \end{aligned} \tag{2.4}$$

Thus,

$$\begin{aligned} & E(S_N | H_0) = \frac{2}{m+1} \sum_{j=0}^{(m-1)/2} a(j, m) = \bar{a}(m) , \quad \text{and} \\ & V(S_N | H_0) = \frac{(n+m+1)}{n(m+2)} \sigma^2_{a(m)} . \end{aligned} \tag{2.5}$$

If on the other hand, *m is even*, $r_0, \cdots, r_{(m-2)/2}$ will have the same moments as in (2.3) and (2.4), while $r_{m/2}$ will have mean $n/(m+1)$, variance $nm(n+m+1)/(m+1)^2(m+2)$ and covariance with any other r_j equal to $-2n(n+m+1)/(m+1)^2(m+2)$. Thus, the mean and variance of S_N would be changed to

$$E(S_N|H_0)=\frac{2}{m+1}\sum_{j=0}^{(m-2)/2} a(j,m)+\frac{1}{m+1}a(m/2, m) ,$$

$$V(S_N|H_0)=\frac{(n+m+1)}{n(m+2)}\left\{\frac{2}{m+1}\sum_{j=0}^{(m-2)/2} a^2(j, m)-\left(\frac{2}{m+1}\sum_{j=0}^{(m-2)/2} a(j, m)\right)^2 + \frac{m}{(m+1)^2}a^2(m/2, m)-\frac{4a(m/2, m)}{(m+1)^2}\sum_{j=0}^{(m-2)/2} a(j, m)\right\} . \tag{2.6}$$

Hence, from (2.5) and (2.6), we get that the difference of $E(S_N|H_0)$ between the cases with m being odd and even is of the order m^{-1} (as $a(m/2, m)<\infty$, by (c.1)) and that of $(m+1)V(S_N|H_0)$ between those cases is also of the order m^{-1}. Hence, (2.6) and (2.5) are asymptotically equal.

3. Asymptotic normality of S_N

The asymptotic normality of the distribution of S_N (standardised by appropriate location and scale parameters) under the null as well as a family of alternative hypotheses, has been established here through the use of an elegant theorem by Blum, Chernoff, Rosenblatt and Teicher ([3], theorem 3 ; corollary) on interchangeable processes. For our purpose, let us first state the theorem.

For each positive integer n, let $\{X_{n_i}, i=1, 2, \cdots\}$ be an interchangeable process with mean zero and variance unity. If then

$$\lim_{n=\infty} E_n(X_{n_1}^2 X_{n_2}^2)=1 , \tag{3.1}$$

$$E_n\{|X_{n_1}|^3\}=o(n^{1/2}) , \quad \text{and} \tag{3.2}$$

$$\lim_{n=\infty} n^{k/2}E_n\{X_{n_1}\cdots X_{n_k}\}=a_k , \quad \text{for} \quad k=1, 2, \cdots , \tag{3.3}$$

where a_k is the k-th moment of a normal distribution with mean a_1 and variance b, then $n^{-1/2}\sum_{i=1}^{n} X_{n_i}$ has asymptotically a normal distribution with mean a_1 and variance $(1+b)$.

Let now $N=m+n$, $m/N=\lambda_N$, and we consider values of m and n such that for some $\lambda_0\leqq 1/2$, $0<\lambda_0\leqq\lambda_N\leqq 1-\lambda_0<1$. Let then X_m stand for the order statistics $X_{(1)}<\cdots<X_{(m)}$, and we adopt the definition of the ring-cells L_j : $j=0, 1, \cdots, [m/2]$ given in (1.3). We then define the counter functions $\phi(Y_j|X_m)$ for $j=1, \cdots, n$, by

$$\phi(Y_j|X_m)=a(i, m) , \quad \text{if} \quad Y_j\in L_i \quad \text{for} \quad i=0, \cdots, [m/2] . \tag{3.4}$$

It then follows readily from (1.8) and (3.4) that

$$S_N=n^{-1}\sum_{j=1}^{n}\phi(Y_j|X_m) . \tag{3.5}$$

Since $Y_1, \cdots, Y_n$ are all independently and identically distributed, it follows that whatever be the two cdf's $F_1(x)$ and $F_2(x)$, $\phi(Y_j|X_m)$ has the same marginal distribution for all $j=1, \cdots, n$, and further the joint distribution of any subset of these variables is a symmetric function of them. Thus for any two arbitrary cdf's $F_1(x)$ and $F_2(x)$, the sequence of variables $\{\phi(Y_j|X_m): j=1, 2, \cdots\}$ forms an interchangeable process (cf. Chernoff and Teicher [6], case b). Thus, we require to prove only that for this sequence of interchangeable random variables, the theorem stated earlier holds true. We propose to prove the theorem first in the null case (i.e., when $F_1(x) \equiv F_2(x)$) and later append a lemma by which the theorem can be readily extended to the more general case of a family of alternative hypotheses. Further, we would consider here only the case of m being *an odd integer*. The case with an even m would follow precisely on the same line.

We now adopt the notations and definitions in (1.4), (1.5) and (1.6). We then write $\Psi_m(Y_j) = \phi(Y_j| X_m) - \bar{a}(m)$, for $j=1, \cdots, n$, and $a'(j, m) = a(j, m) - \bar{a}(m)$ for $j=0, 1, \cdots, [m/2]$, and propose to prove that under $H_0: F_1(x) \equiv F_2(x)$,

$$\text{(3.6)} \qquad \text{(i)} \quad E_{N\cdot\lambda_N}\{|\Psi_m(Y_j)|^3\} = O(N^{1/2-\delta}) = o(N^{1/2}) ,$$

$$\text{(3.7)} \qquad \text{(ii)} \quad \lim_{N=\infty} E_{N\cdot\lambda_N}\{\Psi_m^2(Y_j) \underset{j\neq j'}{\Psi_m^2}(Y_{j'})\} = \sigma_a^4 , \quad \text{and}$$

$$\text{(3.8)} \qquad \text{(iii)} \quad \lim_{N=\infty} E_{N\cdot\lambda_N}\{n^{k/2}\Psi_m(Y_{j_1}) \underset{j_1\neq\cdots\neq j_k}{\cdots} \Psi_m(Y_{j_k})\} = \left(\frac{1-\lambda_N}{\lambda_N}\right)^k \sigma_a^k a_k$$

where

$$a_k = 0, \qquad \text{if } k \text{ is odd}$$
$$= \frac{k!}{2^{k/2}(k/2)!}, \qquad \text{if } k \text{ is even,}$$

and $E_N\lambda_N$ denotes the expectation for a given value of N and $\lambda_N = m/N$, satisfying $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$ for some $\lambda_0 \leqq 1/2$.

The proof of (3.6) follows from (c.1) in section 1. We have

$$E_{N\cdot\lambda_N}\{|\Psi_m(Y_j)|^3\} \leqq 2^3 E_{N\cdot\lambda_N}\{|\phi(Y_j|X_m)|^3\} \leqq 4km^{1/2-\delta} E_{N\cdot\lambda_N}\{|\phi(Y_j|X_m)|^2\}$$
$$= O(N^{1/2-\delta}) = o(N^{1/2}) ,$$

as by (c.1)

$$E_{N\cdot\lambda_N}\{|\phi(Y_j| X_m|^2\} < \infty .$$

While, to prove (3.7) and (3.8), let us first prove the following lemma.

LEMMA 3.1. *Let* $P\{\overset{k}{\underset{j=1}{\Omega}}[\phi(Y_j|X_m) = a(i_j, m)] | H_0\} = P^0_{i_1\cdots i_k}$. *Let now* s_u

of the i_j*'s be equal to* i_u *for* $u=1, 2, \cdots l$, *so that* $\sum_{u=1}^{l} s_u = k$. *Then*

$$p^0_{i_1^{s_1}\cdots i_l^{s_l}} = \prod_{u=1}^{l} (s_u+1)!/(m+k)^{[k]} ,$$

where

$$p^{[q]} = p(p-1)\cdots(p-q+1) ,$$

and l *may range from* 1 *to* k.

The lemma can readily be proved by straightforward computation, and hence the proof is left to the reader.

Now

$$E_{N\cdot\lambda_N}\{\Psi^2_m(Y_{j_1})\Psi^2_m(Y_{j_2})\}_{j_1\neq j_2} = \sum_{i_1=0}^{(m-1)/2}\sum_{i_2=0}^{(m-1)/2} [a'(i_1, m)]^2[a'(i_2, m)]^2 p^0_{i_1 i_2} ,$$

and hence by lemma 3.1, the above reduces to

$$(3.9)\qquad \frac{1}{m+2}\left\{\frac{2}{m+1}\sum_{i=0}^{(m-1)/2}[a'(i, m)]^4\right\} + \frac{m+1}{m+2}\left\{\frac{2}{m+1}\sum_{i=0}^{(m-1)/2}[a'(i, m)]^2\right\}^2$$
$$= \sigma^4_{a(m)} + O(N^{-2\delta}) ,$$

as

$$\frac{2}{m+1}\sum_{i=0}^{(m-1)/2}[a'(i, m)]^4 \leqq \sigma^2_{a(m)} k^2 m^{1-2\delta} ,$$

by condition (c.1). Hence (3.7) is proved.

Let us next prove (3.8) for an even k. Thus, we have for $2k$ variables,

$$(3.10)\qquad \begin{aligned} &E_{N\cdot\lambda_N}\{n^k\Psi_m(Y_{j_1})\cdots\Psi_m(Y_{j_{2k}})\} \\ &= n^k \sum_{i_1=0}^{(m-2)/2}\cdots\sum_{i_{2k}=0}^{(m-1)/2}\{\prod_{j=1}^{2k} a'(i_j, m)\}\, p^0_{i_1\cdots i_{2k}} \\ &= n^k\{\sum_{l=1}^{2k}[\sum_{S_l}(\prod_{j=1}^{l}[a'(i_j, m)]^{\alpha_j}) p^0_{i_1^{\alpha_1}\cdots i_l^{\alpha_l}} ,\end{aligned}$$

where $\sum_{j=1}^{l}\alpha_j = 2k$, and where the summation S_l extends over all possible $i_1 \neq \cdots \neq i_l = 0, 1, \cdots, [m/2]$, and $(\alpha_1, \cdots, \alpha_l)$ with $\alpha_i \geqq 1$ for $i=1, \cdots, l$. Let us next prove the following lemma.

LEMMA 3.2. *Let* $A_l(\alpha_1 \cdots, \alpha_l) = n^k \sum_{S^0_l}(\prod_{j=1}^{l}(\alpha'(i_j, m)]^{\alpha_j}) p^0_{i_1^{\alpha_1}\cdots i_l^{\alpha_l}}$ *where the summation* S^0_l *extends over all possible* $i_1 \neq \cdots \neq i_l = 0, \cdots (m-1)/2$. *Then*

$$A_l(\alpha_1, \cdots, \alpha_l) = O(N^{-2\delta}) ,\ \textit{at most, if } \sup.(\alpha_1, \cdots \alpha_l) > 2 .$$

PROOF. Let us first consider the case $l \geqq k$. Let $l=k+s$, and thus s may range from 0 to $k-2$, as $\sup.(\alpha_1, \cdots, \alpha_l)>2$. It then follows by simple arguments that if $l=k+s$, and if $\sup.(\alpha_1, \cdots, \alpha_l)>2$ then at least $2s+1$ of the α_j's are equal to one. It may be further noted that $\sum_{i=0}^{(m-1)/2} a'(i, m)=0$. Hence, we have for $l=k+s$

$$\begin{aligned} &A_l(\alpha_1 \cdots, \alpha_l) \\ (3.11) \quad &= \frac{n^k \prod_{j=1}^{l} (\alpha_j+1)!}{(m+2k)^{[2k]}} C_l(\alpha_1, \cdots, \alpha_l) \sum_{i_1 \neq \cdots \neq i_l=0}^{(m-1)/2} \prod_{j=1}^{2s+1} a'(i_j, m) \prod_{j=2s+2}^{k+s} [a'(i_j, m)]^{\alpha_j} \\ &= (-1) \sum_{j=2}^{l} \frac{C_l(\alpha_1, \cdots, \alpha_l) A_{l-1}(\alpha_2, \cdots, \alpha_{j-1}, \alpha_j+1, \alpha_{j+1}, \cdots, \alpha_l)}{C_{l-1}(\alpha_2, \cdots, \alpha_{j-1}, \alpha_j+1, \alpha_{j+1}, \cdots, \alpha_l)}, \end{aligned}$$

where

$$C_l(\alpha_1, \cdots, \alpha_l) = \frac{(2k)!}{\alpha_1! \cdots \alpha_l! \; r_1! \cdots r_{l'}!},$$

where r_i of the α_j's are equal to α_i for $i=1, \cdots l'$, and thus it is a finite quantity, independent of (m, n). Using this relation repeatedly, we ultimately get that $A_l(\alpha_1, \cdots, \alpha_l)$ is a linear function of a finite number of $A_p(\beta_1, \cdots, \beta_p)$'s where $p \leqq k-1$, $\sum_{i=1}^{p} \beta_i = 2k$, and none of $\beta_1, \cdots, \beta_p$ are equal to 1, and the compounding coefficients are all finite. In the case of $l<k$, we also similarly arrive at that $A_l(\alpha_1, \cdots \alpha_l)$ is a linear function of $A_p(\beta_1, \cdots, \beta_p)$'s with $p \leqq l$, $\sum_{i=1}^{p} \beta_i = 2k$ and none of $\beta_1, \cdots, \beta_p$ are equal to 1. Thus, we require only to show that

$$(3.12) \qquad A_p(\beta_1, \cdots, \beta_p) = O(N^{-2\delta}),$$

at most, for all $p \leqq k-1$, $\sum_{i=1}^{p} \beta_i = 2k$ and none of $\beta_1, \cdots, \beta_p$ being equal to 1. Let now $p=k-s$, where s may range from 1 to $k-1$. Then we have

$$\begin{aligned} A_p(\beta_1, \cdots, \beta_p) &= n^k \sum_{S_p^0} \Big(\prod_{j=1}^{p} [a'(i_j, m)]^{\beta_p} \Big) p^0_{i_1^{\beta_1} \cdots i_p^{\beta_p}} \\ &= \frac{\prod_{j=1}^{p} (\beta_j+1)! n^k}{(m+2k)^{[2k]}} \sum_{S_p^0} \Big(\prod_{j=1}^{p} [a'(i_j, m)]^{\beta_p} \Big) \\ (3.13) \quad &\leqq \frac{C_p(\beta_1, \cdots, \beta_p) n^k \prod_{j=1}^{p} (\beta_j+1)!}{(m+2k)^{[2k]}} \prod_{j=1}^{p} \Big(\sum_{i=0}^{(m-1)/2} |a'(i_j, m)|^{\beta_j} \Big) \\ &\leqq \frac{C_p(\beta_1, \cdots, \beta_p) n^k \prod_{j=1}^{p} (\beta_j+1)! (m+1)^p}{(m+2k)^{[2k]} 2^p} k^{2s} m^{2s(1/2-\delta)} \sigma^{2p}_{a(m)} \\ &= O(N^{-2s\delta}), \end{aligned}$$

for any given λ_N: $0<\lambda_0\leqq\lambda_N\leqq 1-\lambda_0<1$.

Hence, the lemma follows through (3.12).

In view of this, in (3.10), we need consider only terms $A_l(\alpha_1, \cdots, \alpha_l)$ with $\alpha_i\leqq 2$ for all $i=1, \cdots, l$. Consequently, here $l\geqq k$. Let now $l=k+s$. Then among these α_j's $(k-s)$ are equal to 2, while the remaining $2s$ α_j's are equal to 1. Since, the summation S_l^0 extends over all possible $\alpha_1\neq \cdots \neq\alpha_{k+s}=0, \cdots, (m-1)/2$, the possible number of such sets of α_j's will be equal to

$$\binom{2k}{2s}\frac{\{2(k-s)\}\,!}{2^{k-s}(k-s)\,!}=\frac{(2k)\,!}{2^{k-s}(2s)\,!\,(k-s)\,!}\,.$$

Hence

$$\begin{aligned}
&A_l(\alpha_1, \cdots \alpha_l)\\
&=\frac{(2k)\,!\,n^k6^{k-s}2^{2s}}{2^{k-s}(2s)\,!\,(k-s)\,!\,(m+2k)^{[2k]}}\sum_{i_1\neq\cdots\neq i_l=0}^{(m-1)/2}\prod_{j=1}^{2s}a'(i_j, m)\prod_{j=2s+1}^{k+s}[a'(i_j, m)]^2\\
&=\frac{(-1)(2s-1)(2k)\,!\,n^k6^{k-s}2^{2s}}{2^{k-s}(2s)\,!\,(k-s)\,!\,(m+2k)^{[2k]}}\sum_{i_2\neq\cdots\neq i_l=0}^{(m-1)/2}\prod_{j=2}^{2s-1}a'(i_j, m)\prod_{j=2s}^{k+s}[a'(i_j, m)]^2\\
&\qquad +O(N^{-2\delta})\qquad \text{(by lemma (3.2))}\\
&=\cdots\cdots
\end{aligned}$$

$$\begin{aligned}
(3.14)\quad &=\frac{(-1)^s(2s-1)(2s-3)\cdots 3\cdot 1(2k)!\,n^k\,6^{k-s}2^{2s}}{2^{k-s}(2s)\,!\,(k-s)\,!\,(m+2k)^{[2k]}}\sum_{i_{s+1}\neq\cdots\neq i_{k+s}=0}^{(m-1)/2}\prod_{j=s+1}^{k+s}[a'(i_j, m)]^2\\
&\qquad +O(N^{-2\delta})\\
&=\frac{(-1)^s(2s)\,!\,(2k)\,!\,n^k6^{k-s}2^{2s}}{2^s\,s\,!\,2^{k-s}(2s)\,!\,(k-s)\,!\,(m+2k)^{[2k]}}\prod_{j=1}^{k}\Big(\sum_{i=0}^{(m-1)/2}[a'(i_j, m)]^2+O(N^{-2\delta})\\
&=\frac{(-1)^sn^k}{(m+2k)^{[2k]}}\frac{(2k)\,!}{2^k\,s!\,(k-s)!}6^{k-s}2^{2s}\frac{\sigma^{2k}_{a(m)}(m+1)^k}{2k}+O(N^{-2\delta})\,.
\end{aligned}$$

Thus summing over s from 0 to k, we get from (3.10) through (3.14) that

$$\begin{aligned}
&E_{N\cdot\lambda_N}\{n^k\Psi_m(Y_{j_1})\cdots\Psi_m(Y_{j_{2k}})\}\\
(3.15)\quad &=\left(\frac{n}{m}\right)^k\frac{(2k)\,!}{2^{2k}k\,!}\sigma^{2k}_{a(m)}\sum_{s=0}^{k}(-1)^s\binom{k}{s}2^{2s}6^{k-s}+O(N^{-2\delta})\\
&=\left\{\frac{1-\lambda_N}{\lambda_N}\right\}^k\cdot\frac{(2k)\,!}{2^kk\,!}\sigma^{2k}_{a(m)}+O(N^{-2\delta})\,,
\end{aligned}$$

whence, (3.8) is proved for an even k. The case with an odd k follows precisely on the same line. Thus, under the null hypothesis, all the regularity conditions (3.1), (3.2) and (3.3) for the applicability of the Blum-Chernoff-Rosenblatt-Teicher theorem hold true. Hence, under H_0,

$$\left(\frac{nm}{n+m}\right)^{1/2}\{S_N-\bar{a}(m)\}/\sigma_{a(m)}$$

has asymptotically a normal distribution with zero mean and unit variance.

Let us now consider the family of alternative hypotheses, as specified by

$$(3.16) \qquad \boldsymbol{H}:\ F_i(x)=F([x-\mu]/\delta_{i\cdot N})\ , \qquad \text{for } i=1,2$$

where $\{\delta_{i\cdot N}\}$ is a sequence of real and positive elements, converging to unity as $N\to\infty$. For our purpose, we would assume that the cdf F possesses a continuous density function f, everywhere. This restriction on the cdf F, will be required only for the asymptotic normality of S_N under $\boldsymbol{H}$, but will not be required for the consistency of the S_N-test (as would be shown later on). We then propose to prove the following lemma first.

LEMMA 3.3. *Let* $P\{\Omega_{j=1}^{k}[\phi(Y_j|X_m)=a(i_j, m)]|H\}=p^H_{i_1\cdots i_k}$. *Then for all* $H\in\boldsymbol{H}$, *and all possible* $(\alpha_1, \cdots, \alpha_k)$

$$\lim_{m=\infty} m^k|p^H_{i_1\cdots i_k}-p^0_{i_1\cdots i_k}|=O\ .$$

PROOF: Let us now write

$$(3.17) \qquad \begin{aligned} u_{i(1)}&=F_1(X_{(m-i+1)})-F_1(X_{(m-i)})+F_1(X_{(i+1)})-F_1(X_{(i)})\ , \quad \text{and} \\ u_{i(2)}&=F_2(X_{(m-i+1)})-F_2(X_{(m-i)})+F_2(X_{(i+1)})-F_2(X_{(i)})\ . \end{aligned}$$

It then follows that

$$(3.18) \qquad p^0_{i_1\cdots i_k}=E\{\prod_{j=1}^{k}u_{i_j(1)}|F_1\} \quad \text{and} \quad p^H_{i_1\cdots i_k}=E\{\prod_{j=1}^{k}u_{i_j(2)}|F_2\}\ ,$$

where $E\{\ \ |F_1\}$ indicates that the expectation extends over $X_{(1)}<\cdots<X_{(m)}$ and is operated by the parent cdf F_1. Since by definition $u_{j(i)}$'s are all finite and lie between zero and one, it follows that

$$m^k|p^H_{i_1\cdots i_k}-p^0_{i_1\cdots i_k}|\to 0\ , \quad \text{as} \quad m\to\infty\ ,$$

if we can show that

$$(3.19) \qquad m|u_{i(1)}-u_{i(2)}|\xrightarrow{P} 0\ ,$$

uniformly in $i=0, \cdots, [m/2]$, and in $\boldsymbol{H}$.

Now, using (3.16) and some simple but lengthy computations, it can be readily shown that

(3.20) $$mu_{i(2)}=\alpha_m\theta_N\{mu_{i(1)}\}$$

uniformly in $i=0, \cdots [m/2]$; where θ_N converges to unity as $N\to\infty$ for all alternative in $\boldsymbol{H}$, and α_m converges in probability to unity. Also, by the well-known results on the elementary sample coverages, $mu_{i(1)}$ has a limiting non-degenerate distribution with mean 2 and variance 6, for all $i=0, 1, \cdots, [m/2]$. Hence, (3.19) follows from (3.20), and the lemma is proved.

Combining now lemma 3.3 with lemmas 3.1 and 3.2, the proof of (3.6), (3.7) and (3.8) follows precisely on the same line as in (3.9) through (3.15). Hence, the asymptotic normality of S_N, under the family of alternative hypotheses in $\boldsymbol{H}$, sketched in (3.16). It may be noted in this connection that the mean and variance of S_N under such a hypothesis will be asymptotically equal to $\bar{a}_{(m)}$ and $\{nm/(n+m)\}^{-1}\sigma^2_{a(m)}$ respectively, but the distribution of

$$\left(\frac{nm}{n+m}\right)^{1/2}\{S_N-\bar{a}(m)\}/\sigma_{a(m)}$$

will then be normal with unit variance, and mean equal to

(3.21) $$\Delta_H=\left(\frac{nm}{n+m}\right)^{1/2}\sum_{i=0}^{(m-1)/2}a(i, m)\{p_i^H-p_i^0\}/\sigma_{a(m)}$$

and thus $|\Delta_H|$ may be finite, zero or indefinitely large, depending on the particular alternative specification.

REMARK ONE. There is some analogy of our statistic S_N with the one by Chernoff and Savage [5]. For our purpose, let us first introduce the following notations.

Let

$$N=m+n\ , \qquad m/N=\lambda_N\ , \qquad 0<\lambda_0\leqq\lambda_N\leqq1-\lambda_0<1$$

for some $\lambda_0\leqq1/2$. Let us also define the empirical cdf's as

(3.22)
$$\begin{aligned}&F_{1m}(x)=(\text{Number of } X_i\leqq x)/m\ ,\\&F_{2n}(x)=(\text{Number of } Y_i\leqq x)/n\ , \quad \text{and}\\&H_N(x)=\lambda_NF_{1m}(x)+(1-\lambda_N)F_{2n}(x)\ ; \qquad H(x)=\lambda_NF_1(x)+(1-\lambda_N)F_2(x)\ .\end{aligned}$$

A statistic T_N has then been defined by Chernoff and Savage [5] as

(3.23) $$T_N=\int_{-\infty}^{\infty}J_N[H_N(x)]\,dF_{1m}(x)\ ,$$

where $J_N=J_N(i/N)$ is an explicit function of (i, N) and need be defined only at $1/N, \cdots, N/N$, but may have its domain of definition extended to (0.1], by some convention. Then under the following regularity conditions:

(a) $J(H)=\lim_{N=\infty} J_N(H)$ exists for all $0<H<1$, and is not constant,

(b) $\int_{I_N} [J_N(H_N)-J(H_N)]dF_{1m}(x)=O_p(N^{-1/2})$ where I_N: $O<H_N(x)\leqq 1$,

(c) $J_N(1)=O(\sqrt{N})$, and

(d) $|J^{(i)}(H)|=\left|\frac{d^i J(H)}{dH^i}\right|\leqq k[H(1-H)]^{-i-1/2+\delta}$

for almost all H: $O<H<1$ and, for $i=0, 1, 2$ and for some $\delta>0$; $(T_N-\mu_N)/\sigma_N$ (where μ_N and σ_N has been defined by them in [5], theorem 1) has asymptotically a normal distribution with zero mean and unit variance, for any given F_1, F_2 and λ_N. It may be noted that the statistic T_N may also be written as

$$T_N=m^{-1}\sum_{j=1}^{m} J_N[H_N(x_{(j)})]=m^{-1}\sum_{i=1}^{N} E_{Ni}Z_{Ni} \tag{3.24}$$

where $Z_{Ni}=1$, or 0, according as the i-th smallest in the combined sample is a X or Y, $E_{Ni}=J_N(i/N)$ are pure numbers, and $X_{(1)}<\cdots<X_{(m)}$ are the first sample ordered variables. Adopting now the definition of $b(j, m)$ in section 1, we let

$$\begin{aligned} b(j, m)&=b(m-j+1, m)\\ &=m^{-1}G_1(j/m) \quad \text{for} \quad j=1, \cdots, [m/2], \end{aligned} \tag{3.25}$$

where the function G_1 satisfies the conditions (c.1), (c.2) and (c.3), and thus

$\lim_{m=\infty} G_1=G$ for all $0<F<1$, G being defined in (1.7). We then let

$$J_N[H_N(X_{(j)})]=G_1(j/m)|H_N(X_{(j)})-1/2| \quad \text{for} \quad j=1, \cdots, m. \tag{3.26}$$

It can then be shown with some lengthy but direct computations that

$$N^{1/2}\{T_N-[(1-\lambda_N)S_N+\lambda_N C_N]\} \xrightarrow{P} 0 \tag{3.27}$$

uniformly in $\boldsymbol{H}$, defined in (3.16), where

$$\begin{aligned} C_N&=\bar{a}(m)+1/m\{(m+1)a(1, m)+2a(o, m)-2\bar{a}(m)\}\\ &\to\bar{a}+a_1 \text{ as } m\to\infty; \ \bar{a}=\lim_{m=\infty}\bar{a}(m), \text{ and } a_1=\lim_{m=\infty}a(1, m). \end{aligned}$$

We thus arrive at a linear relationship (almost sure) of our statistic S_N and their T_N. But unfortunately, the function J_N, defined in (3.26) fails to satisfy the basic requirement of Chernoff-Savage theorem, since their J_N depends only on (i/N) but our J_N depends on (j/m) as well as (i/N). Thus we fail to arrive at the asymptotic normality of S_N, through that of T_N.

However, it is well-worth comparing the regularity conditions pertaining to the asymptotic normality of S_N and of T_N. It follows from (c.1), (c.2) and (c.3) that we do not require the regularity conditions (a), (b), (c) of the Chernoff-Savage theorem, though (a) is implicit in (c.1). Our condition (c.1) is in a sense, less restrictive than (d) of theirs. In (c.1), we do not require the existence of the second order derivative of our weight-function, as well as the restriction on the first order derivative of it, while the two weight-functions satisfy a similar condition. However, their theorem applies to any arbitrary F_1 and F_2, while ours only to a class of alternative hypotheses $\boldsymbol{H}$, considered in (3.16); though for large samples this class $\boldsymbol{H}$ covers the entire domain of interest.

4. Consistency and asymptotic efficiency of the S_N-test

The consistency of the S_N-test against any difference of the scale parameters (assuming, of course, the homogeneity of the locations) can be proved by Blum-Chernoff-Rosenblatt-Teicher theorem [3], provided we assume that the cdf F admits a continuous density function f everywhere. However, as for the consistency, we do not require the existence and finiteness of the density function, and the same may be established as follows.

It follows from (2.5) and (2.6) that under H_0: $F_1(x)\equiv F_2(x)$,

$$S_N \xrightarrow{P} \bar{a} \qquad \text{where} \qquad \bar{a}=\lim_{m=\infty}\bar{a}(m)\,. \tag{4.1}$$

It follows from the definition of S_N in (1.9) and the condition (c.3) that under condition which are even less stringent than (c.1), (c.2) and (c.3), and for arbitrary F_1 and F_2, $|S_N-E(S_N|F_1, F_2)| \xrightarrow{P} 0$. The proof is somewhat lengthy, and for the intended brevity of our discussion here, is omitted. Thus, we require to prove only that for any $\boldsymbol{H}$: $F_2(x)=F_1(\theta x)$ with $\theta\neq 1$, (conventionally, under the assumption of the homogeneity of locations, we take $\mu=0$), $|E(S_N|\theta)-\bar{a}(m)|$ tends to a non-zero limit, as $N\to\infty$. Again, with the definition of S_N in (1.9), conditions (c.2) and (c.3), and by an application of the Glivenko-Cantelli's theorem (cf. Loève [11], pp. 20–21), it can be shown readily that

$$\lim_{N=\infty} E(S_N|\theta)=\int_0^1 G(F)\,F(\theta x)\,dF(x)\,, \qquad |G(F)|>0 \tag{4.2}$$

where $G(F)$ has the same sign for all $0<F<1$, (by (c.2) and (c.3)). Consequently,

$$\lim_{N=\infty} |E(S_N|\theta)-\bar{a}(m)|=\int_0^1 |G(F)|\,|F(\theta x)-F(x)|\,dF(x)>0 \quad \text{for any } \theta\neq 1\,. \tag{4.3}$$

Hence, we have the consistency of the S_N-test.

As the S_N-test has been shown to be consistent for any $\theta\neq 1$, the power of the test will be asymptotically equal to unity. Hence, we relate θ with N, in such a way that the power asymptotically lies between 0 and 1. Thus, as in (3.16), we take

$$F_i(x)=F([x-\mu]/\delta_{i\cdot N}) \qquad \text{for} \qquad i=1,\ 2\,, \tag{4.4}$$

where $\theta=\delta_{2\cdot N}/\delta_{1\cdot N}=1+\gamma/\sqrt{N}$, with a real and finite γ.

It then follows from the definition of Pitman-efficiency (cf. Fraser [7], pp. 271–273) that the necessary regularity conditions are all satisfied here, provided we assume further that the cdf F has a continuous density function f, everywhere. Hence, the relative Pitman-efficiency of the S_N-test with respect to the variance ratio (F-) test is given as

$$E_{S\cdot F}=\frac{\beta_2-1}{4\sigma_a^2}\left[\int_{-\infty}^{\infty} |x\,G(F)|\,f^2(x)\,dx\right]^2, \tag{4.5}$$

where

$$\sigma_a^2=\lim_{m=\infty}\sigma_{a(m)}^2\,, \qquad F=F(x) \quad \text{and} \quad \beta_2=E\{X-E(X)\}^4/[E\{X-E(X)\}^2]^2$$

is the conventional measure of Kurtosis of the cdf F.

5. Location-free S_N-test

Our S_N-test is based on the postulated identity of the locations. In order to test for the difference in the scale parameters, without presuming the identity of the locations, we, following Sukhatme [19], express the sample observations as deviations from the respective sample medians, and thus base the test on these centred observations. We thus require to study the regularity conditions, under which, our S_N-test thus modified, will be distribution-free. Now, for small samples, the distribution of this modified test criterion will be highly involved and will depend appreciably on the parent cdf, through the sampling distribution of the sample medians. Hence, we will confine ourselves only to large samples, and study the regularity conditions, under which the modified S_N-test would be distribution-free.

Let $\tilde{X}$ and $\tilde{Y}$ denote respectively the first and the second sample medians. We then define $X_i^* = X_i - \tilde{X}$, $X_{(i)}^* = X_{(i)} - \tilde{X}$ for $i=1, \cdots, m$, $X_{(0)}^* = -\infty$, $X_{(m+1)}^* = \infty$, and $Y_i^* = Y_i - \tilde{Y}$ for $i=1, \cdots, n$. The definitions of the modified inter-quantile cells (I_j^*: $j=0, \cdots, [m/2]$) and the modified ring-cells (L_j^*: $j=0, \cdots, [m/2]$) will then go precisely on the same line as in (1.2) and (1.3) respectively, where only $X_{(i)}$ and x are to be replaced by $X_{(i)}^*$ and x^* for $i=0, \cdots, m+1$. Let then r_j^* of the Y^* values in the second sample belong to L_j^* and R_j^* to I_j^*, for $j=0, \cdots, [m/2]$. Then the modified test criterion S_N^* may be written as

$$(5.1) \qquad S_N^* = n^{-1} \sum_{j=0}^{[m/2]} a(j, m) r_j^* = n^{-1} \sum_{j=0}^{[m/2]} b(j, m) R_j^* \, .$$

The asymptotic distribution-freeness of S_N^* would be established, provided we can show that $N^{1/2}\{S_N - S_N^*\} \xrightarrow{P} 0$ for any given λ_N and under H_0 .

Now

$$(5.2) \qquad \sqrt{n}\,(S_N - S_N^*) = n^{-1/2} \sum_{j=0}^{[m/2]} b(j, m)(R_j - R_j^*)$$

and hence, it can be shown following some lengthy computations, that (5.2) converges in probability to zero, provided that (i) the cdf F has everywhere a continuous density function f, which is symmetrical about the population median, and (ii) the density function at the population median is non-zero. The details of the deduction are omitted here for the intended brevity of the discussion. The deduction, however, follows precisely on the same line as in the case with modified inter-quantile tests, considered by the author [16] elsewhere. It may be noted that Sukhatme [19] has put forward the same set of regularity conditions, while studying the same property of his test [18] based on a U-statistic, though the other tests by Mood [12] or by Lehmann [10] fail to be asymptotically distribution-free (when thus modified) even under such regularity conditions. The modified form of the test by Tamura [21] is also asymptotically distribution-free under the same set of regularity conditions, while Ansari and Bradley [1] guessed that under these conditions their test might also be distribution-free.

The consistency of the S_N^* test for any difference of the scale parameters, without any regard to the possible difference of the locations, follows precisely on the same line as in the preceding section, and further S_N^* and S_N have the same asymptotic normal distribution, and hence, the same Pitman-efficiency.

6. A proposed class of weight-functions

Here a class of weight-functions has been proposed, and the resulting Pitman-efficiency of the corresponding S_N-test has been studied and contrasted with those of the others. We now let

$$(6.1) \qquad a(j, m)=\left\{\left(\frac{m+1}{2}-j\right)\Big/(m+1)\right\}^k \qquad \text{for} \quad j=0, \cdots, [m/2] ;$$

where the possible values of k (>0) generate the class. It then follows that

(i) $\bar{a}=\lim_{m=\infty} \bar{a}(m)=1/2^k(k+1)$,

(ii) $\sigma_a^2 =\lim_{m=\infty} \sigma_{a(m)}^2= \{k/(k+1)\}^2/2^{2k}(2k+1)$, and

(iii) $G(F)=\lim_{m=\infty} G(F, m)=k|F-1/2|^{k-1}$.

Consequently, from (4.5) we get that the Pitman-efficiency of the corresponding S_N-test with respect to the variance ratio (F-) test is

$$(6.2) \qquad E_{S.F}^{(k)}=\frac{\beta_2-1}{4}\left[\int_{-\infty}^{\infty}|x|k|F(x)-1/2|^{k-1}f^2(x)\,dx\right]^2\frac{2^{2k}(2k+1)(k+1)^2}{k^2}$$
$$=2^{2(k-1)}(2k+1)(k+1)^2(\beta_2-1)\left[\int_{-\infty}^{\infty}|x|\,|F(x)-1/2|^{k-1}f^2(x)\,dx\right]^2 .$$

Now we have compared (6.2) with the corresponding values of the tests by Mood [12], Sukhatme ([18], [20]), Tamura [21], and Ansari and Bradley [1]. The test by Lehmann [10], unfortunately, fails to be distribution-free, even under H_0. The present author [14] has utilized a distribution-free and consistent estimate of the variance of U-statistic, to render the studentized form of Lehmann's statistic asymptotically distribution-free, and to have the same asymptotic power properties as that of the original test. Sukhatme's [18] first test and the test by Ansari and Bradley [1] are asymptotically power-equivalent to our S_N-test in the particular case $k=1$. Sukhatme's [20] second test and Lehmann's [10] test have more or less the same power efficiency, and they are more-power efficient than the former ones, while they are less efficient than the ones by Mood [12] and by Tamura [21], these latter two being again asymptotically power-equivalent, and also power-equivalent to our S_N-test in the particular case: $k=2$. The test by Siegel and Tukey [17] is again power-equivalent to Ansari and Bradley's test. The asymptotic power-efficiency of the tests by Kamat [8] and by Wilks-Rosenbaum [13] has been considered in detail by the author [15], elsewhere; where it has been shown that the asymptotic efficiency of such tests (with respect to either of the above tests) would be equal to zero, if the range of the parent cdf extends to infinity, in at least one extremity, while for distributions having finite extremities these two tests would be suitable only if the order of terminal contact be either zero or unity. In fact, for all distributions with non-zero density functions at both the extremities, the asymptotic efficiency of any one of the earlier tests with respect to either of these two tests would be equal to zero. Capon [4] and Klotz [9] have considered locally most powerful rank-sum tests based on the principle of likelihood ratio, and their tests are also asymptotically power-equivalent, and are most efficient for certain specified type of parent cdf (under the alternative specifications). But the main drawback of these two tests is that the weight-functions depend explicitly on the assumed parent density function, for which we intend to maximise locally the power function. Consequently, if such a density function can not fairly be guessed (so as to be really appropriate for the given data), application of such a test seems to be dubious from the point of view of the power of the test, as the locally most powerful test for a particular parent cdf, may lose its power efficiency considerably for other cdf's. Further, for the computation of the weight-functions in these two tests, we require a tabulation of the expected values of the order statistics for various sample sizes, and this may be quite tedious in computation for many parent cdf's (other than the normal one). From these considerations, it is clear that we should use some rank-sum tests, which are reasonably efficient for many parent cdf's (and more particularly in the vicinity of the normal distribution), whose weight functions do not depend on the particular cdf, (which may be assumed to fit adequately the given data), and which are not computationally cumbrous. The class of S_N-test appears to be very suitable from this standpoint. The following table illustrates this.

Thus, in many cases, our S_N-test with $k \geqq 3$ has the power-efficiency substantially higher than that of the others. For $k > 4$, the efficiency of the S_N-test decreases, for normal alternatives, with k increasing. In view of this, the S_N-test with $k=4$ appears to be very suitable for a broad class of parent cdf's, including the normal one.

7. A few concluding remarks

The test proposed here is based on a family of congruent inter-quantile cells of the first sample. We may, however, use similarly a family of such cells of the pooled sample. We will have then $[N/2]$

TABLE 1

The Pitman-efficiency of some two sample scale tests with respect to the variance ratio test for certain common type of cdf's.

Tests	Relative Pitman-efficiency with respect to the variance ratio test for the parent cdf.		
	Normal	Laplace	Uniform
Sukhatme's [18] first test, Ansari and Bradley's [1] test, S_N-test with $k=1$	$6/\pi^2=0.61$	0.94	0.60
Sukhatme's [20] second test	$=0.69$	1.03	0.80
Mood's [12] test, Tamura's [21] test, S_N-test with $k=2$	$15/2\pi^2=0.76$	1.08	1.00
S_N-test with $k=3$	$=0.84$	1.14	1.40
S_N-test with $k=4$	$=0.88$	1.16+	1.80
S_N-test with $k=5$	$=0.85$	1.16−	2.20

+ (or −) indicates that the actual value is slightly more (or less) than the tabulated one.

(where N is the pooled sample size) such cells, and the test may then be based on the number of observations of either of the samples, belonging to these cells. The asymptotic properties of such tests may then be studied by the application of the well-known theorem by Chernoff and Savage [5], following the lines of Ansari and Bradley [1]. If we use the weight function $a(j, N)$ (i.e., the same as $a(j, m)$ with m replaced by N), then it can be shown that our S_N-test and this one would be asymptotically power equivalent. As the ordering of the observations (in order of magnitude) in any single sample is always less tedious than that in the samples pooled, from computational aspect, our S_N-test appear to be comparatively more suitable. There is a further point in favour of our S_N-test. The null distribution of S_N, for small samples, as has been considered in section 2, also appears to be more simple than the one, for the later test, which may be found by the same technique as by Ansari and Bradley [1]. Thus the finding out of the critical values of S_N, for small samples, also appears to be comparatively simpler, in relation to the later class of tests.

The extension of the findings to the case of more than two samples is kept pending, as it requires first of all, a c-sample extension of the theorem by Blum et al [3] or by Chernoff and Savage [5].

Acknowledgements

The author remains indebted to Prof. H. K. Nandi, Reader in Statistics, Calcutta University, for his generous help in course of this investigation. Thanks are also due to Prof. H. Chernoff, for some useful references.

DEPT. OF STAT., CALCUTTA UNIVERSITY, INDIA

REFERENCES

[1] A. R. Ansari and R. A. Bradley, "Rank-sum tests for dispersions," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 1172–1189.

[2] D. E. Barton and Florence N. David, "A test for birth order-effects," *Ann. Human Genetics*, Vol. 22 (1958), pp. 250–257.

[3] J. R. Blum, H. Chernoff, M. Rosenblatt and H. Teicher, "Central limit theorems for interchangeable processes," *Can. Jour. Math.*, Vol. 10 (1958), pp. 222–229.

[4] Jack Capon, "Asymptotic efficiency of certain locally most powerful rank-tests," *Ann. Math. Stat.*, Vol. 32 (1961), pp. 88–100.

[5] Herman Chernoff and I. Richard Savage, "Asymptotic normality and efficiency of certain non-parametric test statistics," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 972–994.

[6] H. Chernoff and H. Teicher, "A central limit theorem for sums of interchangeable random variables," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 118–130.

[7] D. A. S. Fraser, *Non-Parametric Methods in Statistics*, New York, John Wiley & Sons, 1957.

[8] A. R. Kamat, "A two-sample distribution-free test," *Biometrika*, Vol. 43 (1956), pp. 377–385.

[9] Jerome Klotz, "Non-parametric tests for scale," *Ann. Math. Stat.*, Vol. 33 (1962), pp. 498–512.

[10] E, L. Lehmann, "Consistency and unbiasedness of certain non-parametric tests," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 165–179.

[11] M. Loève, *Probability Theory*, New York, D. Van Nostrand Co., 1955.

[12] A. M. Mood, "On the asymptotic efficiency of certain non-parametric two sample tests," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 514–522.

[13] S. Rosenbaum, "Tables for a non-parametric test of dispersion," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 663–668.

[14] P. K. Sen, "On some convergence properties of *U*-statistics," *Calcutta Stat. Assoc. Bull.*, Vol. 10 (1960), pp. 1–18.

[15] P. K. Sen, *Order Statistics and Their Role in Some Problems of Statistical Inference*, D. Phil. Thesis, Calcutta University, (1961).

[16] P. K. Sen, "On the role of a class of quantile tests in some non-parametric problems," *Calcutta Stat. Assoc. Bull.*, Vol. 11 (1962), pp. 125–143.

[17] Sidney Siegel and John W. Tukey, "A non-parametric sum of ranks procedure for relative spread in unpaired samples," *Jour. Amer. Stat. Assoc.*, Vol. 55 (1960), pp. 429–445.

[18] B. V. Sukhatme, "On certain two-sample non-parametric tests for variances," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 188–194.

[19] B. V. Sukhatme, "Testing the hypothesis that two populations differ only in location," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 60–78.

[20] B. V. Sukhatme, "A two-sample distribution-free test for variances," *Biometrika*, Vol. 45 (1958), pp. 544–548.

[21] Ryoji Tamura, "On the non-parmetric tests based on certain *U*-statistics," *Bull. Math. Stat.*, Vol. 9 (1960), pp. 61–68.