ON A ROBUSTNESS PROPERTY OF A CLASS OF NONPARAMETRIC TESTS BASED ON U-STATISTICS

by

Pranab Kumar Sen

University of North Carolina Chapel Hill,
and University of Calcutta

Institute of Statistics Mimeo Series No. 513

February 1967

Work supported by the National Institute of Health,
Public Health Service, Grant GM-12868.

DEPARTMENT OF BIOSTATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

ON A ROBUSTNESS PROPERTY OF A CLASS OF NONPARAMETRIC TESTS BASED ON U-STATISTICS*

By

Pranab Kumar Sen

University of North Carolina Chapel Hill,
and University of Calcutta.

1. Summary and introduction. A class of nonparametric tests based on Hoeffding's (1948) U-statistics is shown to be robust for heterogeneity of the distributions of the sample observations; the theory is illustrated by means of some known tests.

Let $X_1,\ldots,X_n$ be independent (real or vector valued) random variables having cumulative distribution functions (cdf's) $F_1(x),\ldots,F_n(x)$, respectively. Let Ω be the set of all cdf's closed under the following (countable) set operation: if $F_i \in \Omega$ for $i=1,2,\ldots$, then $(1/n)\Sigma_{i=1}^{n} F_i \in \Omega$ for all $n \geq 1$. It is assumed that $\underset{\sim}{F}_n = (F_1,\ldots,F_n)$ is an element of the product set Ω^n, and the null hypothesis states that

$$(1.1) \qquad H_o: \underset{\sim}{F}_n \in \omega_n \subset \Omega^n.$$

Let now

$$(1.2) \qquad \mathcal{F}_n = \{\underset{\sim}{F}_n: F_1 \equiv \ldots \equiv F_n \in \Omega\}, \ \omega_n^* = \omega_n \cap \mathcal{F}_n \text{ and } \omega_n^{**} = \omega_n - \omega_n^*.$$

When ω_n^* is some hypothesis of invariance of $\underset{\sim}{F}_n$ (under certain finite group of transformations that maps the sample space onto itself), then granted certain conditions, tests based on Hoeffding's (1948) U-statistics are essentially distribution-free [e.g., the signed-rank test by Tukey (1949)(for the hypothesis

* Supported by the National Institute of Health, Public Health Service, Grant GM-12868.

of sign-invariance (i.e., symmetry)), the tests by Kendall (1938), Hotelling and Pabst (1936) and others [cf. Hoeffding (1948, pp. 316-321)] (for the hypothesis of matching-invariance (i.e., bivariate independence))]. However, if $\underset{\sim}{F}_n \in \omega_n^{**}$ (even when ω_n is some hypothesis of invariance), tests based on U-statistics are, in general, not distribution-free. A test $\psi(\underset{\sim}{E}_n)$ (where $\underset{\sim}{E}_n = (X_1, \ldots, X_n)$ and $0 \leq \psi \leq 1$) is termed robust for $\underset{\sim}{F}_n \in \omega_n$ if

$$(1.3) \qquad E\{\psi(\underset{\sim}{E}_n) | \underset{\sim}{F}_n \in \omega_n^*\} \leq \alpha \Rightarrow \sup_{\underset{\sim}{F}_n \in \omega_n^{**}} [E\{\psi(\underset{\sim}{E}_n) | \underset{\sim}{F}_n\}] \leq \alpha,$$

$\alpha (0 < \alpha < 1)$ being the preassigned level of significance of the test. If there is no confusion, $\psi(\underset{\sim}{E}_n)$ satisfying (1.3) only for large n will also be termed robust. The object of the present investigation is to show that under certain conditions on ω_n (to be stated in section 2), the usual nonparametric tests for $\underset{\sim}{F}_n \in \omega_n^*$ based on U-statistics are robust for $\underset{\sim}{F}_n \in \omega_n$. Some illustrative examples are also considered.

2. Statement of the main theorem. Corresponding to a symmetric kernel $\Phi(X_1, \ldots, X_m)$ of degree $m (\geq 1)$, Hoeffding's (1948) U-statistic is defined as

$$(2.1) \qquad U_n = \binom{n}{m}^{-1} \Sigma_S \Phi(X_{i_1}, \ldots, X_{i_m}); \quad S = \{1 \leq i_1 < \ldots < i_m \leq n\}.$$

Following Hoeffding, we define

$$(2.2) \qquad \theta_{i_1, \ldots, i_m} = \int \ldots \int \Phi(x_1, \ldots, x_m) \prod_{i}^{m} dF_{i_j}(x_j), \; 1 \leq i_1 < \ldots < i_m \leq n.$$

Then, U_n is an unbiased estimator of $\theta(\underset{\sim}{F}_n)$, defined by

$$(2.3) \qquad \theta(\underset{\sim}{F}_n) = \binom{n}{m}^{-1} \Sigma_S \theta_{i_1,\ldots,i_m}.$$

U_n is said to be <u>strictly distribution-free for</u> $\underset{\sim}{F}_n \epsilon \omega_n^*$, if its distribution is independent of $\underset{\sim}{F}_n (\epsilon \omega_n^*)$; this implies that

$$(2.4) \qquad \theta(\underset{\sim}{F}_n) = \theta_o \text{ (known) for all } \underset{\sim}{F}_n \epsilon \omega_n^*.$$

Since for $\underset{\sim}{F}_n \epsilon \omega_n^*$, $F_1 \equiv \ldots \equiv F_n \equiv F$, $\theta(\underset{\sim}{F}_n)$ reduces to

$$(2.5) \qquad \theta(F) = \int \ldots \int \phi(x_1,\ldots,x_m) \prod_1^m dF(x_j), \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n^*.$$

There are certain situations where U_n may not be properly distribution-free (for $\underset{\sim}{F}_n \epsilon \omega_n^*$), but (2.4) holds and further it is possible to find an estimator V_n of the variance of U_n such that $V_n^{-\frac{1}{2}}[U_n - \theta_o]$ converges in law to a standardized normal distribution for all $\underset{\sim}{F}_n \epsilon \omega_n^*$; U_n is then said to be <u>asymptotically distribution-free for</u> $\underset{\sim}{F}_n \epsilon \omega_n^*$. In either case, our contention is to show that the assumption $\underset{\sim}{F}_n \epsilon \mathcal{F}_n$ (implicit in ω_n^*) is redundant if n is not small and ω_n satisfies some conditions. To formulate the same, we define for any $(F_1, F_2) \epsilon \Omega^2$

$$(2.6) \qquad \phi_1(x;F_2) = \int \ldots \int \phi(x, x_2, \ldots, x_m) \prod_2^m dF_2(x_j), \quad \theta(F_1;F_2) = \int \phi_1(x;F_2) dF_1(x);$$

$$(2.7) \qquad \zeta_1(F_1;F_2) = \int \phi_1^2(x;F_2) dF_1(x) - \theta^2(F_1;F_2).$$

It may be noted that for $F_1 \equiv F_2 \equiv F$, we have $\theta(F_1;F_2) \equiv \theta(F)$ and

$$(2.8) \qquad \zeta_1(F;F) = \int \ldots \int \phi(x_1,\ldots,x_m) \phi(x_m,\ldots,x_{2m-1}) \prod_1^{2m-1} dF(x_j) - \theta^2(F) = \zeta_1(F),$$

where $\zeta_1(F)$ is defined by Hoeffding (1948, p. 298). Also, let

(2.9) $\omega_n^{(i)} = \{\underset{\sim}{F}_n \epsilon \omega_n: \ F_1 \equiv \ldots \equiv F_{i-1} \equiv F_{i+1} \equiv \ldots \equiv F_n\}$ for $i=1,\ldots,n$.

Then, we have

(2.10) $$\omega_n^* = \bigcap_i \omega_n^{(i)} \subset \bigcup_i \omega_n^{(i)} \subset \omega_n.$$

Now, concerning ω_n, we assume that (i) for some $\delta > 0$

(2.11) $$\int \ldots \int |\phi(x_1,\ldots,x_m)|^{2+\delta} \prod_1^m dF_{i_j}(x_j) < \infty ,$$

uniformly in $1 \leq i_1 < \ldots < i_m \leq n$ and $\underset{\sim}{F}_n \epsilon \omega_n$; and (ii)

(2.12) $$\zeta_1(F_i;F) > 0 \text{ uniformly in } \underset{\sim}{F}_n \epsilon \omega_n^{(i)}, \ i=1,\ldots,n.$$

In conjunction with (2.8), (2.12) implies that

(2.13) $$\zeta_1(F) > 0 \text{ uniformly in } \underset{\sim}{F}_n \epsilon \omega_n^*.$$

Now, according to Hoeffding (1948, p. 299), (2.13) implies that $\Theta(F)$ in (2.5) is stationary of order zero for all $\underset{\sim}{F}_n \epsilon \omega_n^*$. By analogy with this, we say that (2.12) means that $\Theta(\underset{\sim}{F}_n)$, defined by (2.4), <u>is stationary or order zero</u> for all $\underset{\sim}{F}_n \epsilon \bigcup_i \omega_n^{(i)}$. Further, if U_n is strictly distribution-free for $\underset{\sim}{F}_n \epsilon \omega_n^*$, we have from (2.13) that

(2.14) $$\zeta_1(F) = \zeta_{10} \text{ (known)} > 0, \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n^*.$$

Let us now define

(2.15) $$\bar{F}_{(n)}(x) = \frac{1}{n}\Sigma_{i=1}^{n}F_i(x); \text{ by definition } \bar{F}_{(n)}\epsilon\Omega.$$

We say that ω_n is a <u>resolving subset</u> of Ω^n if

(2.16) $$\underset{\sim}{F}_n\epsilon\omega_n \Rightarrow \bar{F}_{(n)}(1,\ldots,1)\epsilon\omega_n^*.$$

Further, if in addition to (2.16), we have on defining $\Theta(F_1;F_2)$ as in (2.6),

(2.17) $$\Theta(F_i,\bar{F}_{(n)}) = \Theta_o \text{ for all } \underset{\sim}{F}_n\epsilon\omega_n,$$

we say that ω_n is an <u>affine resolving subset</u> of Ω^n, the affinity being with respect to Θ. Finally, we shall consider the following types of tests:

(2.18) $$\psi(\underset{\sim}{E}_n) = \psi_1(\underset{\sim}{E}_n) = \begin{cases} 1, & U_n \geq c_n, \\ 0, & U_n < c_n; \end{cases}$$

(2.19) $$\psi(\underset{\sim}{E}_n) = \psi_2(\underset{\sim}{E}_n) = \begin{cases} 1, & |U_n-\Theta_o| \geq c_n, \\ 0, & |U_n-\Theta_o| < c_n; \end{cases}$$

where

(2.20) $$E\{\psi(\underset{\sim}{E}_n)|\underset{\sim}{F}_n\epsilon\omega_n^*\} \leq \alpha.$$

If we have a vector of U-statistics i.e., $\underset{\sim}{U}_n = (U_n^{(1)},\ldots,U_n^{(p)})$, $p \geq 1$, we define the corresponding vector of Θ_o's by $\underset{\sim}{\Theta}_o = (\Theta_o^{(1)},\ldots,\Theta_o^{(p)})$, the true covariance matrix of $\underset{\sim}{U}_n$ by $\underset{\sim}{\Sigma}_n(\underset{\sim}{F}_n)$ and the estimated covariance matrix by $\underset{\sim}{V}_n$. If $\underset{\sim}{U}_n$ is strictly distribution-free for $\underset{\sim}{F}_n\epsilon\omega_n^*$, we have, $\underset{\sim}{\Sigma}_n(\underset{\sim}{F}_n) = \underset{\sim}{\Sigma}_n^o$ (independent of $\underset{\sim}{F}_n \in \omega_n^*$), and by an assumption similar to (2.12), we have

(2.21) $n\underset{\sim}{\Sigma}_n^o$ is positive definite; $n\underset{\sim}{V}_n$ is positive definite, in probability.

We consider a test-statistic

$$S_n = [U_n - \theta_o]\underset{\sim}{A}_n^{-1}[\underset{\sim}{U}_n - \theta_o]' \tag{2.22}$$

where $\underset{\sim}{A}_n$ is $\underset{\sim}{\Sigma}_n^o$ or $\underset{\sim}{V}_n$ according as U_n is strictly or asymptotically distribution-free for $\underset{\sim}{F}_n \in \omega_n^*$. Then, the third type of test is

$$\psi(\underset{\sim}{E}_n) = \psi_3(\underset{\sim}{E}_n) = \begin{cases} 1, & S_n \geq c_n \\ 0, & S_n < c_n, \end{cases} \tag{2.23}$$

where c_n satisfies (2.20). The main theorem of the paper is the following.

<u>THEOREM 2.1. If (i) (2.11) holds, (ii) $\theta(\underset{\sim}{F}_n)$ is stationary of order zero for all $\underset{\sim}{F}_n \in \bigcup_i \omega_n^{(i)}$, (iii) ω_n is a resolving subset of Ω^n and (iv) U_n is strictly distribution-free for $\underset{\sim}{F}_n \in \omega_n^*$, then (a) for all $0 < \alpha < 1$, $\psi_2(\underset{\sim}{E}_n)$ and $\psi_3(\underset{\sim}{E}_n)$ are robust for $\underset{\sim}{F}_n \in \omega_n$, and (b) for $0 < \alpha \leq \frac{1}{2}$, $\psi_1(\underset{\sim}{E}_n)$ is robust for $\underset{\sim}{F}_n \in \omega_n$. If (i), (ii) and (iv) remain same, while (iii)' ω_n is an affine resolving subset of Ω^n then for all $0 < \alpha < 1$, $\psi_1(\underset{\sim}{E}_n)$ is robust for $\underset{\sim}{F}_n \in \omega_n$. If U_n is only asymptotically distribution-free for $\underset{\sim}{F}_n \in \omega_n^*$, we need to change (i) by (i)' that (2.11) holds for some</u> $\delta \geq 2$.

The proof of this theorem is postponed to section 4.

3. <u>Three useful lemmas.</u> Let us define $\theta(F)$ and $\zeta_1(F)$ as in (2.5) and (2.8), respectively, $\bar{F}_{(n)}$ as in (2.15), and let

$$(3.1)\qquad \theta(\bar{F}_{(n)}) = \theta(F)\big|_{F\equiv F_{(n)}},\quad \zeta_1(\bar{F}_{(n)}) = \zeta_1(F)\big|_{F\equiv \bar{F}_{(n)}}.$$

Also, we define $\theta(F_1;F_2)$ as in (2.6) and let

$$(3.2)\qquad \Delta_n^2 = \frac{1}{n}\sum_{i=1}^{n}[\theta(F_i;\bar{F}_{(n)}) - \theta(\bar{F}_{(n)})]^2 \geq 0.$$

LEMMA 3.1. *Under (2.11),* $|\theta(\underset{\sim}{F}_n) - \theta(\bar{F}_{(n)})| = 0(n^{-1})$, *where* $\theta(\underset{\sim}{F}_n)$ *is defined by (2.3).*

PROOF. From (2.1) and (2.3) we have after some simplifications

$$(3.3)\qquad \theta(\underset{\sim}{F}_n) = \frac{1}{n\ldots(n-m+1)} \sum_{i_1\neq\ldots\neq i_m=1}^{n} \int\ldots\int \phi(x_1,\ldots,x_m)\prod_{1}^{m} dF_{i_j}(x_j).$$

Also from (2.5) and (2.15), we have

$$(3.4)\qquad \theta(\bar{F}_{(n)}) = n^{-m}\sum_{i_1=1}^{n}\ldots\sum_{i_m=1}^{n}\int\ldots\int\phi(x_1,\ldots,x_m)\prod_{1}^{m} dF_{i_j}(x_j).$$

Comparing (3.3) and (3.4), it is easily seen that under (2.11), $|\theta(\underset{\sim}{F}_n)-\theta(\bar{F}_{(n)})| = 0(n^{-1})$. Hence, the lemma.

LEMMA 3.2. *Under* (2.11), $|nV\{U_n\} - [m^2(\zeta_1(\bar{F}_{(n)})-\Delta_n^2)]| = 0(n^{-1})$.

PROOF. We define $\theta_{i_1,\ldots,i_m}$ as in (2.2) and let

$$(3.5)\qquad \begin{aligned}&\psi_{c(i_1,\ldots,i_c),j_1,\ldots,j_{m-c}}(x_{i_1},\ldots,x_{i_c})\\ &= E\{\phi(x_{i_1},\ldots,x_{i_c},X_{j_1},\ldots,X_{j_{m-c}})\} - \theta_{i_1,\ldots,i_c,j_1,\ldots,j_{m-c}};\end{aligned}$$

$$\zeta_{c(i_1,\ldots,i_c)j_1,\ldots,j_{m-c};k_1,\ldots,k_{m-c}} =$$

(3.6)

$$E\{\psi_{c(i_1,\ldots,i_c)j_1,\ldots,j_{m-c}}(X_{i_1},\ldots,X_{i_c})\,\psi_{c(i_1,\ldots,i_c)k_1,\ldots,k_{m-c}}(X_{i_1},\ldots,X_{i_c})\}$$

for $c=1,\ldots,m$ and all possible distinct i's, j's and k's. Also, let

$$(3.7)\qquad \zeta_{c,n} = \{\binom{n}{2m-c}\binom{2m-c}{c}\binom{2m-2c}{m-c}\}^{-1} \Sigma_c^* \zeta_{c(i_1,\ldots,i_c)j_1,\ldots,j_{m-c};k_1,\ldots,k_{m-c}},$$

where the summation Σ_c^* extends over all possible (distinct) sets of i's, j's and k's. Then the variance of U_n is given by [cf. Hoeffding (1948)]

$$(3.8)\qquad \sigma^2(U_n) = \binom{n}{m}^{-1} \Sigma_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c} \zeta_{c,n} .$$

Thus, it suffices to prove that under (2.11)

$$(3.9)\qquad |\zeta_{1,n} - \{\zeta_1(\bar{F}_{(n)}) - \Delta_n^2\}| = o(1).$$

Now, using (3.6) and (3.7), $\zeta_{1,n}$ may also be written as

$$n^{-[2m-1]} \Sigma_1^{**}\{\int\ldots\int\phi(x_1,\ldots,x_m)\phi(x_m,\ldots,x_{2m-1})dF_{i_1}(x_m)\prod_1^{m-1} dF_{j_\ell}(x_\ell)dF_{k_\ell}(x_{m+\ell})$$

(3.10)

$$- \theta_{i_1,j_1,\ldots,j_{m-1}} \cdot \theta_{i_1,k_1,\ldots,k_{m-1}}\},$$

where $p^{[q]} = p\ldots(p-q+1)$ and the summation Σ_1^{**} extends over all possible $1 \leq i_1 \neq j_1 \neq \ldots \neq j_{m-1} \neq k_1 \neq \ldots \neq k_{m-1} \leq n$. Now, by the same method as in lemma 3.1, it can be shown that under (2.11)

(3.11) $$\left|n^{-[2m-1]}\Sigma_1^{**}\,\theta_{i_1,j_1,\ldots,j_{m-1}}\theta_{i_1,k_1,\ldots,k_{m-1}} - \frac{1}{n}\sum_{i=1}^{n}\theta^2(F_i;\bar{F}_{(n)})\right| = O(n^{-1}),$$

where $\theta(F_i;\bar{F}_{(n)})$ is defined by (2.6) and (2.15). Also, it is seen from (2.5), (2.6) and (2.15) that

(3.12) $$(1/n)\Sigma_{i=1}^{n}\,\theta(F_i;\bar{F}_{(n)}) = \theta(\bar{F}_{(n)};\bar{F}_{(n)}) = \theta(\bar{F}_{(n)}).$$

Finally, using (2.8) and (2.15), we obtain that

(3.13) $$\zeta_1(\bar{F}_{(n)}) + \theta^2(\bar{F}_{(n)}) = n^{-(2m-1)}\,\Sigma_1^{o}\int\ldots\int\phi(x_1,\ldots,x_m)\phi(x_m,\ldots,x_{2m-1})dF_{i_1}(x_m)\prod_{1}^{m-1}dF_{j_\ell}(x_\ell)dF_{k_\ell}(x_{m+\ell}),$$

where the summation Σ_1^{o} extends over all possible $1 \le i_1 \le n$, $1 \le j_\ell \le n$, $1 \le k_\ell \le n$ for $\ell=1,\ldots,m-1$. (3.9) readily follows from (2.11), (3.10, (3.11), (3.12) and (3.13). Hence, the lemma.

LEMMA 3.3. Under (2.11) and (2.12), $n^{\frac{1}{2}}[U_n-\theta(\bar{F}_{(n)})]/m\{\zeta_1(\bar{F}_{(n)}) - \Delta_n^2\}^{\frac{1}{2}}$ converges in law to a standardized normal distribution for all $\underset{\sim}{F}_n\epsilon\omega_n$.

PROOF. The proof of this lemma will follow readily from theorems 8.1 and 8.2 of Hoeffding (1948), and our lemmas 3.1 and 3.2, provided we can show that (in addition to (3.9)), $\zeta_1(\bar{F}_{(n)}) - \Delta_n^2 > 0$, uniformly in $\underset{\sim}{F}_n\epsilon\omega_n$. Now, by virtue of (2.7) and (2.12), we have

(3.14) $$\int\phi_1^2(x;\bar{F}_{(n)})dF_i(x) - \theta^2(\bar{F}_{(n)}) > \theta^2(F_i;\bar{F}_{(n)}) - \theta^2(\bar{F}_{(n)}),$$

for all i=1,...,n and uniformly in $\underset{\sim}{F}_n\epsilon\omega_n$. Hence, from (2.8), (3.2), (3.12) and (3.14), we obtain

(3.15) $$\zeta_1(\bar{F}_{(n)}) - \Delta_n^2 > 0 \text{ uniformly in } \underset{\sim}{F}_n \epsilon \omega_n.$$

Hence, the lemma.

[Incidentally, the condition (2.12) may be easier to verify than the parallel condition by Hoeffding (1948, p. 313; (8.14)). Also the third moment condition implicit in his theorems may be replaced by (2.11)].

4. Proof of theorem 2.1. As ω_n is a resolving subset of Ω^n, from (2.4) and (2.16), we have

(4.1) $$\theta(\bar{F}_{(n)}) = \theta_o \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n.$$

Hence, using the three lemmas in section 3, we obtain under the conditions (i) and (ii) of theorem 2.1 that

(4.2) $$\mathcal{L}(n^{\frac{1}{2}}[U_n - \theta_o]/m\{\zeta_1(\bar{F}_{(n)}) - \Delta_n^2\}^{\frac{1}{2}}) \to N(0,1) \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n.$$

Now, if U_n is strictly distribution-free for $\underset{\sim}{F}_n \epsilon \omega_n^*$, by (2.14) and (2.16) we have

(4.3) $$\zeta_1(\bar{F}_{(n)}) = \zeta_{10} > 0 \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n.$$

Hence, from (4.2) and (4.3)

(4.4) $$\mathcal{L}(n^{\frac{1}{2}}[U_n - \theta_o]/m\{\zeta_{10} - \Delta_n^2\}^{\frac{1}{2}}) \to N(0,1) \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n,$$

where Δ_n^2, defined by (3.2), is zero either when $\underset{\sim}{F}_n \epsilon \omega_n^*$ or when ω_n is an affine resolving subset of Ω^n. Let now t_α be the $100(1-\alpha)\%$ point of a standardized normal distribution. Then from (4.4), (2.18), (2.19) and (2.20), we obtain that

$$(4.5)\qquad |c_n - \theta_o - n^{-\frac{1}{2}}\, m t_\alpha\, \zeta_{10}^{\frac{1}{2}}| = o(n^{-\frac{1}{2}}) \text{ for } \psi_1(\underset{\sim}{E}_n);$$

$$(4.6)\qquad |c_n - n^{-\frac{1}{2}}\, m t_{\alpha/2}\, \zeta_{10}^{\frac{1}{2}}| = o(n^{-\frac{1}{2}}) \text{ for } \psi_2(\underset{\sim}{E}_n).$$

The robustness of $\psi_1(\underset{\sim}{E}_n)$ (for $0 < \alpha < \frac{1}{2}$) and $\psi_2(\underset{\sim}{E}_n)$ (for $0 < \alpha < 1$) then follows from (4.4), (4.5) and (4.6). Further, if ω_n is an affine resolving subset of Ω^n, $\Delta_n \equiv 0$ for all $\underset{\sim}{F}_n \epsilon \omega_n$ and hence from (4.4) and (4.5), the robustness of $\psi_1(\underset{\sim}{E}_n)$ (for all $0 < \alpha < 1$) follows.

Now, if U_n is asymptotically distribution-free for $\underset{\sim}{F}_n \epsilon \omega_n^*$, its estimated variance (V_n) is such that nV_n estimates $m^2\zeta_1(F)$ for all $\underset{\sim}{F}_n \epsilon \omega_n^*$. Now, by (2.8), $\zeta_1(F)$ is a regular functional of degree 2m, and hence, the corresponding U-statistic is an optimum estimator of it [cf. Fraser (1957, p. 142)]. If this estimator of $\zeta_1(F)$ is used then nV_n is an optimum unbiased estimator of $m^2\zeta_1(F)$ for all $\underset{\sim}{F}_n \epsilon \omega_n^*$. Now, if (2.11) holds for some $\delta \geq 2$, it follows from our lemmas 3.1 and 3.2 that under (2.12)

$$(4.7)\qquad nV_n/m^2\zeta_1(\bar{F}_{(n)}) \xrightarrow{p} 1 \text{ for all } F_n \epsilon \omega_n.$$

Now, for asymptotically distribution-free U_n, we need substitute $(nV_n)^{\frac{1}{2}}$ for $m\zeta_{10}^{\frac{1}{2}}$ in (4.5) and (4.6). By virtue of (4.7), the desired result will again follow from (2.18), (2.19), (4.4), (4.5) and (4.6).

It remains only to show that $\psi_3(\underset{\sim}{E}_n)$ in (2.23) is also robust. By a straightforward extension of lemma 3.3, we have

$$(4.8)\qquad \mathcal{L}(S_n) \rightarrow \chi_p^2 \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n^*,$$

where S_n is defined by (2.22) and χ_p^2 has a χ^2 distribution with p degrees of freedom (d.f.). Hence, from (2.23), we have

(4.9) $$|c_n - \chi^2_{p,\alpha}| = o(1), \text{ where } P\{\chi^2_p \geq \chi^2_{p,\alpha}\} = \alpha .$$

Let us furst consider the case when U_n is strictly distribution-free for $\underset{\sim}{F}_n \epsilon \omega_n^*$. On replacing the two Φ's in (2.8) by Φ_r and Φ_s (with m=m_r and m_s respectively,) we obtain $\zeta_{1(r,s)}(F)$ for all r, s=1,...,p. We consider then the following matrix

(4.10) $$n \underset{\sim}{\Sigma}_n(F) = ((m_r m_s \zeta_{1(r,s)}(F)))_{r,s=1,\ldots,p} .$$

Similarly, replacing the sum of squares in (3.2) by an analogous sum of products, we define $\Delta_{n(r,s)}$ for all r, s=1,...,p, and let

(4.11) $$n \underset{\sim}{\Delta}_n = ((m_r m_s \Delta_{n(r,s)})).$$

By assumption (2.21), $n \underset{\sim}{\Sigma}_n(F)$ is positive definite for all $\underset{\sim}{F}_n \epsilon \omega_n^*$, and $n \underset{\sim}{\Delta}_n$ is positive semi-definite for all $\underset{\sim}{F}_n \epsilon \omega_n$. As U_n is strictly distribution-free for $\underset{\sim}{F}_n \epsilon \omega_n^*$, by (2.21) and (2.22),

(4.12) $$n\, A_n = n \underset{\sim}{\Sigma}_n(F) = n \underset{\sim}{\Sigma}_n(\bar{F}_{(n)}) = \text{a known matrix, for all } \underset{\sim}{F}_n \epsilon \omega_n.$$

Let now for $\underset{\sim}{F}_n \epsilon \omega_n$,

(4.13) $$\underset{\sim}{B}_n = \underset{\sim}{\Sigma}_n(\bar{F}_{(n)}) - \underset{\sim}{\Delta}_n, \quad S_n^* = [\underset{\sim}{U}_n - \underset{\sim}{\theta}_o] \underset{\sim}{B}_n^{-1} [\underset{\sim}{U}_n - \underset{\sim}{\theta}_o]' .$$

Then, by a straightforward extension of (4.4), we have

(4.14) $$\mathcal{L}(S_n^*) \longrightarrow \chi^2 \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n .$$

Our desired result will follow then from (4.9), (4.13) and (4.14), provided we can show that

(4.15) $$S_n \geq S_n^* \text{ for all } \underset{\sim}{E}_n \text{ and all } \underset{\sim}{F}_n \epsilon \omega_n.$$

Upon noting that $\underset{\sim}{B}_n = \underset{\sim}{A}_n - \underset{\sim}{\Delta}_n$, $n\underset{\sim}{A}_n$ positive definite, (4.15) readily follows from (2.22), (4.13) and the following simple lemma (whose proof is omitted).

<u>LEMMA 4.1</u> <u>If</u> $\underset{\sim}{A}_1$ <u>(and</u> $\underset{\sim}{A}_2$<u>) are two positive definite (and semidefinite) symmetric matrices of the order</u> p x p <u>and</u> $\underset{\sim}{x}$ <u>is any p-vector with real elements,</u>
$\underset{\sim}{x}\, \underset{\sim}{A}_1^{-1}\, \underset{\sim}{x}' \geq \underset{\sim}{x}(\underset{\sim}{A}_1+\underset{\sim}{A}_2)^{-1}\, \underset{\sim}{x}'$.

This completes the proof for distribution-free U_n. When U_n is asymptotically distribution-free (for $\underset{\sim}{F}_n \epsilon \omega_n^*$), and $\underset{\sim}{V}_n$ is its estimated covariance matrix, as in (4.7), we have under (2.11) (for $\delta \geq 2$),

(4.16) $$|n\, \underset{\sim}{V}_n - n\, \underset{\sim}{\Sigma}_n(\bar{F}_{(n)})| \xrightarrow{p} \underset{\sim}{0}_{pxp} \text{ for all } \underset{\sim}{F}_n \epsilon \omega_n.$$

The rest of the proof will follow from (2.22), (4.13), (4.14) and (4.15). Hence the theorem.

5. Illustrations. Let us first consider the hypothesis of matching-invariance. Let $X_i = (X_{1i}, X_{2i})$ have the cdf $F_i(x_1, x_2)$, for i=1,...,n and let

(5.1) $$H_o: \; F_i(x_1,x_2) \equiv G_i(x_1)H_i(x_2);\; G_i(x)=F_i(x_1,\infty),\; H_i(x_2) = F_i(\infty,x_2),$$

for all i=1,...,n. If $F_1 \equiv \ldots \equiv F_n \equiv F$, (5.1) implies that the joint distribution of $(X_1,\ldots,X_n)$ remains invariant under arbitrary matching $\{(X_{1i}, X_{2R_i}),\ i=1,\ldots,n\}$, where $(R_1,\ldots,R_n)$ is any permutation of $(1,\ldots,n)$, and hence, (5.1) is termed the hypothesis of matching-invariance. When $F_1 \equiv \ldots \equiv F_n$, genuinely distribution-free tests for H_o in (5.1) based on U-statistics are due to Kendall (1938), Hotelling and Pabst (1936) and others [cf. Hoeffding (1948, pp. 316-321)]. We shall show that these tests are all robust for a wider class of distributions. We consider the following four cases:

I. $G_1 \equiv \ldots \equiv G_n \equiv G$, $H_1 \equiv \ldots \equiv H_n \equiv H$ (i.e., $F_1 \equiv \ldots \equiv F_n \equiv F$),

II. $G_1 \equiv \ldots \equiv G_n \equiv G$ but $H_1, \ldots, H_n$ are not all same,

III. $H_1 \equiv \ldots \equiv H_n \equiv H$ but $G_1, \ldots, G_n$ are not all same,

IV. Neither G_i's nor H_i's are all identical.

It is easy to verify that except in case IV, ω_n, defined by the set of equations in (5.1), is a resolving subset of Ω^n. Further, if in case II (or case III) $\int_0^1 H_i(x)dHj(x) = \frac{1}{2}$ (or $\int_0^1 G_i(x)dG_j(x) = \frac{1}{2}$) for all $i, j=1, \ldots, n$, ω_n is an affine resolving subset of Ω^n (when Kendall's tau or Spearman's rho is used). Also, for Kendall's tau or Spearman's rho, the kernels are bounded and hence (2.11) holds for any finite δ, and it may be verified that (2.12) will hold if the integrals $\int_0^1 H_i(x)dH_j(x)$ (in case III) or $\int_0^1 G_i(x)dG_j(x)$ (in case II) for $i, j=1, \ldots, n$ are bounded away from 0 and 1, i.e., the marginal cdf's are all mutually overlapping.

This clearly indicates that the nonparametric tests by Kendall (1938) and Hotelling and Pabst (1936) are robust for heterogeneity of the marginal cdf's of one of the two variates (i.e., case II or case III).

Next, let us consider the problem of sign-invariance. Let X_i have the cdf $F_i(x)$ defined on the real p-space, (for some $p \geq 1$), $i=1, \ldots, n$. The null hypothesis states that X_i and $(-1)X_i$ have the common cdf F_i for all $i=1, \ldots, n$, i.e., $F_1, \ldots, F_n$ are all diagonally symmetric about $\underset{\sim}{0}$. [For $p=1$, this simplifies to

$$(5.2) \qquad F_i(x) = 1 - F_i(-x-o) \text{ for all } i=1, \ldots, n.]$$

It is, again, easy to verify that if $F_1, \ldots, F_n$ are all diagonally symmetric, then $\bar{F}_{(n)}$, defined by (2.15), is also so. Thus, here also ω_n is a resolving subset of Ω^n. Moreover, ω_n can also be shown to be an affine resolving subset of Ω^n if the signed-rank statistic of Tukey (1949) is used. As the kernel of this U-statistic is bounded, (2.11) holds for all $\underset{\sim}{F}_n \in \omega_n$, if the marginal cdf's (of each variate) corresponding to $F_1, \ldots, F_n$ are mutually overlapping. Thus, the test based on the p-vector of signed-rank statistics corresponding to the p components of X_i $(i=1,\ldots,n)$ is robust for any heterogeneity of the cdf's $F_1, \ldots, F_n$. In the case of $p=1$, it has been shown by the author [cf. Sen (1967)] that when $F_1, \ldots, F_n$ are not all identical, the asymptotic relative efficiency of the signed rank test with respect to the t-test (for shift alternatives) is usually greater than that in the case when $F_1 \equiv \ldots \equiv F_n$. By virtue of the results derived in this paper, it can be shown by some routine analysis that the above results also hold for the general case of $p \geq 1$.

Other examples of robustness can also be cited.

REFERENCES

[1] FRASER, D. A. S. (1957). Nonparametric methods in statistics. New York: John Wiley.

[2] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distributions. Ann. Math. Statist. 19, 293-325.

[3] HOTELLING, H. and PABST, M. R. (1936). Rank correlation and tests of significance involving no assumptions of normality. Ann. Math. Statist. 7, 29-43.

[4] KENDALL, M. G. (1938). A new measure of rank correlation. Biometrika, 30, 81-93.

[5] SEN, P. K. (1967). On a further robustness property of the test and estimator based on Wilcoxon's signed rank statistic. Ann. Math. Statist.

[6] TUKEY, J. W. (1949). The simplest signed-rank tests. Mimeo report 17, Statistics Research Group, Princeton University.