ASYMPTOTIC NORMALITY AND ROBUSTNESS OF ONE SAMPLE CHERNOFF-SAVAGE STATISTICS FOR HETEROGENEOUS DISTRIBUTIONS

by

Pranab Kumar Sen

University of North Carolina

Institute of Statistics Mimeo Series No. 556

November 1967

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

DEPARTMENT OF BIOSTATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

ASYMPTOTIC NORMALITY AND ROBUSTNESS OF ONE SAMPLE CHERNOFF-SAVAGE STATISTICS FOR HETEROGENEOUS DISTRIBUTIONS*

By PRANAB KUMAR SEN

University of North Carolina at Chapel Hill

1. Introduction and Summary. Let $X_1,\dots,X_n$ be independent random variables having continuous cumulative distribution functions (cdf) $F_1(x),\dots,F_n(x)$, respectively. Consider the Chernoff-Savage (1958) statistic

$$T_n = (1/n)\sum_{i=1}^{n} E_{n,i} Z_{n,i}, \tag{1.1}$$

where $E_{n,i} = J_n(i/(n+1))$, $(i=1,\dots,n)$ are (explicitly known) rank scores (satisfying the conditions 1, 2 and 3 of section 2), and $Z_{n,i}$ is 1 or 0 according as the ith smallest observation among $|X_1|,\dots,|X_n|$ corresponds to a positive X or not $(i=1,\dots,n)$. For $F_1\equiv\dots\equiv F_n\equiv F$, the asymptotic normality of the standardized form of T_n has been obtained by Govindarajulu (1960) (See also Sen and Puri (1967) and Pyke and Shorack (1967)). The present paper is concerned with (i) the asymptotic normality of T_n for arbitrary continuous $F_1,\dots,F_n$ and (ii) the robust efficiency of T_n for shift alternatives when $F_1,\dots,F_n$ are not all identical.

2. Preliminary notions. Let $c(u)$ be 1 or 0 according as u is ≥ 0 or not. Define

$$F_n^*(x) = (1/n)\sum_{i=1}^{n} c(x-X_i),\quad \bar{F}_{(n)}(x) = (1/n)\sum_{i=1}^{n} F_i(x) \quad (-\infty<x<\infty); \tag{2.1}$$

$$H_i(x) = F_i(x)-F_i(-x-),\ i=1,\dots,n;\quad \bar{H}_{(n)}(x) = (1/n)\sum_{i=1}^{n} H_i(x) \quad (x\geq 0); \tag{2.2}$$

$$H_n^*(x) = (1/n)\sum_{i=1}^{n} c(x-|X_i|) = F_n^*(x)=F_n^*(-x-),\ (x\geq 0). \tag{2.3}$$

As in Chernoff and Savage (1958), we extend the domain of $J_n(u)$ to (0,1) by letting

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

it have constant values on $(\frac{i-1}{n+1} < u \leq \frac{i}{n+1})$, $i=1,\ldots,n$. Then, T_n in (1.1) may be written as

(2.4) $$T_n = \int_0^\infty J_n(\frac{n}{n+1} H_n^*(x))dF_n^*(x).$$

It is assumed that

(2.5) (1) $\lim_{n\to\infty} J_n(u) = J(u)$ exists for all $0<u<1$ and is not a constant;

(2.6) (2) $$\int_0^\infty [J_n(\frac{n}{n+1} H_n^*(x))-J(\frac{n}{n+1}H_n^*(x))]dF_n^*(x) = o_p(n^{-\frac{1}{2}});$$

(2.7) (3) $$|J^{(r)}(u)| = |(d^r/du^r)J(u)| \leq K[u(1-u)]^{-r-\frac{1}{2}+\delta},\ r=0,1,$$

for some $\delta>0$, where K is a finite positive constant. Let us also define

(2.8) $$\mu_n^* = \int_0^\infty J[\bar{H}_{(n)}(x)]d\bar{F}_{(n)}(x) \text{ (so that } |\mu_n^*| \leq \int_0^1 |J(u)|du < \infty),$$

(2.9) $$\gamma_{n,i}^2 = \int_0^\infty J^2[\bar{H}_{(n)}(x)]dF_i(x)-[\int_0^\infty J[\bar{H}_{(n)}(x)]dF_i(x)]^2 +$$
$$2[\iint_{0<x<y<\infty} H_i(x)[1-H_i(y)]J'[\bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(y)]d\bar{F}_{(n)}(x)d\bar{F}_{(n)}(y)$$
$$+ \iint_{0<x<y<\infty} J[\bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(y)]d\bar{F}_{(n)}(x)dF_i(y)$$
$$- \{\int_0^\infty J[\bar{H}_{(n)}(x)]dF_i(x)\}\{\int_0^\infty H_i(x)J'[\bar{H}_{(n)}(x)]d\bar{F}_{(n)}(x)\}],\ i=1,\ldots,n;$$

(2.10) $$\gamma_n^2 = (1/n)\sum_{i=1}^n \gamma_{n,i}^2.$$

The main theorem of the paper is the following.

<u>THEOREM 1</u>. <u>If assumptions 1, 2 and 3 hold and if</u> $\inf_n \gamma_n^2 > 0$,

$$\lim_{n\to\infty} P\{n^{\frac{1}{2}}[T_n-\mu_n^*]/\gamma_n \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\{-\tfrac{1}{2}t^2\}dt,$$

<u>uniformly in x and</u> $F_1,\ldots,F_n$.

The proof is postponed to Section 4.

3. Some fundamental lemmas. Define

$$\gamma^2(\bar F_{(n)}) = \int_0^\infty J^2[\bar H_{(n)}(x)]d\bar F_{(n)}(x) - (\mu_n^*)^2 +$$

(3.1)
$$2\Big[\iint_{0<x<y<\infty} \bar H_{(n)}(x)[1-\bar H_{(n)}(y)]J'[\bar H_{(n)}(x)]J'[\bar H_{(n)}(y)]d\bar F_{(n)}(x)d\bar F_{(n)}(y) + \iint_{0<x<y<\infty} J[\bar H_{(n)}(x)]J'[\bar H_{(n)}(y)]d\bar F_{(n)}(x)d\bar F_{(n)}(y) - \{\int_0^\infty J[\bar H_{(n)}(x)]d\bar F_{(n)}(x)\}\{\int_0^\infty \bar H_{(n)}(x)J'[\bar H_{(n)}(x)]d\bar F_{(n)}(x)\}\Big];$$

(3.2)
$$\alpha_{n,i} = \int_0^\infty J[\bar H_{(n)}(x)]dF_i(x) - \mu_n^*, \quad i=1,\dots,n;$$

(3.3)
$$\beta_{n,i} = \int_0^\infty [H_i(x)-\bar H_{(n)}(x)]J'[\bar H_{(n)}(x)]d\bar F_{(n)}(x), \quad i=1,\dots,n.$$

<u>LEMMA 3.1</u>. <u>For arbitrary</u> $F_1,\dots,F_n$,

$$\gamma_n^2 = \gamma^2(\bar F_{(n)}) - (1/n)\sum_{i=1}^n(\alpha_{n,i}+\beta_{n,i})^2 \le \gamma^2(\bar F_{(n)}) < \infty.$$

The proof follows by straightforward computations using (2.1), (2.2), (2.7), (2.8), (2.9), (2.10), (3.1), (3.2) and (3.3). Hence, for brevity, the details are omitted.

Define

(3.4)
$$B_{n1}(X_i) = J[\bar H_{(n)}(|X_i|)]c(X_i);$$

(3.5)
$$B_{n2}(X_i) = \int_0^\infty [c(x-|X_i|)-H_i(x)]J'[\bar H_{(n)}(x)]d\bar F_{(n)}(x),$$

(3.6)
$$B_n(X_i) = B_{n1}(X_i)+B_{n2}(X_i), \quad i=1,\dots,n.$$

<u>LEMMA 3.2</u>. <u>Under assumption 3</u>, $(1/n)\sum_{i=1}^n E\{|B_n(X_i)|^{2+\delta}\} < \infty$, <u>uniformly in</u> n <u>and</u> $F_1,\dots,F_n$.

<u>PROOF</u>. By virtue of the inequality that $|a+b|^{2+\delta} \le 2^{1+\delta}\{|a|^{2+\delta}+|b|^{2+\delta}\}$, it suffices to show that uniformly in $F_1,\dots,F_n$ and n,

(3.7) $$(1/n)\sum_{i=1}^{n} E\{|B_{nj}(X_i)|^{2+\delta}\} < \infty \quad \text{for } j=1,2.$$

Upon using assumption 3, it follows from (3.4) that

(3.8) $$(1/n)\sum_{i=1}^{n} E\{|B_{n1}(X_i)|^{2+\delta}\} = \int_0^\infty |J[\bar{H}_{(n)}(x)]|^{2+\delta} d\bar{F}_{(n)}(x)$$
$$\leq \int_0^1 |J(u)|^{2+\delta} du < \infty, \text{ as } d\bar{F}_{(n)} \leq d\bar{H}_{(n)} \text{ and } (2+\delta)(-\tfrac{1}{2}+\delta) > -1.$$

Let now Y_n be a random variable (independent of $X_1, \ldots, X_n$) following the cdf $\bar{F}_{(n)}(x)$. Define

(3.9) $$d_n(X_i, Y_n) = [c(Y_n - |X_i|) - H_i(Y_n)] J'[\bar{H}_{(n)}(Y_n)], \; i=1,\ldots,n.$$

It is easy to verify that

(3.10) $$B_{n2}(X_i) = E\{d_n(X_i, Y_n)|X_i\}, \; i=1,\ldots,n.$$

Consequently, by straightforward computations, we obtain that

(3.11) $$E\{|B_{n2}(X_i)|^{2+\delta}\} \leq E\{[E|d_n(X_i,Y_n)|^{1+\delta/2}|X_i]^2\}$$
$$= 2E\{\iint_{0<x<y<\infty} |\{c(x-|X_i|)-H_i(x)\}\{c(y-|X_i|)-H_i(y)\}J'[\bar{H}_{(n)}(x)] J'[\bar{H}_{(n)}(y)]|^{1+\delta/2} d\bar{F}_{(n)}(x) d\bar{F}_{(n)}(y)$$
$$\leq 6 \iint_{0<x<y<\infty} H_i(x)[1-H_i(y)]|J'[\bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(y)]|^{1+\delta/2} d\bar{F}_{(n)}(x) d\bar{F}_{(n)}(y),$$

(as $E[|\{c(x-|X_i|)-H_i(x)\}\{c(y)-|X_i|\}-H_i(y)\}|^{1+\delta/2}|x<y\} \leq 3H_i(x)[1-H_i(y)]$.)

Upon noting that $(1/n)\sum_{i=1}^{n} H_i(x)[1-H_i(y)] = \bar{H}_{(n)}(x)[1-\bar{H}_{(n)}(y)] - (1/n)\sum_{i=1}^{n}[H_i(x) - \bar{H}_{(n)}(x)][H_i(y)-\bar{H}_{(n)}(y)]$, we obtain from (3.11) that

(3.12) $$(1/n)\sum_{i=1}^{n} E\{|B_{n2}(X_i)|^{2+\delta}\}$$
$$\leq 6[\iint_{0<x<y<\infty} \bar{H}_{(n)}(x)[1-\bar{H}_{(n)}(y)]|J'[\bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(y)]|^{1+\delta/2} d\bar{F}_{(n)}(x) d\bar{F}_{(n)}(y) - $$
$$- \frac{1}{2n}\sum_{i=1}^{n}\{\int_0^\infty [H_i(x)-\bar{H}_{(n)}(x)]|J'[\bar{H}_{(n)}(x)]|^{1+\delta/2} d\bar{F}_{(n)}(x)\}^2]$$
$$\leq 6 \iint_{0<u<v<1} u(1-v)|J'(u)J'(v)|^{1+\delta/2} du dv < \infty, \text{ by (2.7)},$$

(as $d\bar{F}_{(n)} \leq d\bar{H}_{(n)}$). Therefore the proof follows from (3.8) and (3.12). Q.E.D.

4. The proof of theorem 1. Using (2.4), (3.4), and (3.6), one can write

$$(4.1) \qquad T_n = (1/n)\sum_{i=1}^{n} B_n(X_i) + \sum_{r=1}^{4} C_{r,n},$$

where $B_n(X_i)$ is defined by (3.6) and

$$(4.2) \qquad C_{1,n} = \int_0^\infty [J_n(\tfrac{n}{n+1} H_n^*(x)) - J(\tfrac{n}{n+1} H_n^*(x))]dF_n^*(x) = o_p(n^{-\frac{1}{2}}), \text{ by } (2.6);$$

$$(4.3) \qquad C_{2,n} = [-1/(n+1)] \int_0^\infty H_n^*(x)J'[\bar{H}_{(n)}(x)]dF_n^*(x);$$

$$(4.4) \qquad C_{3,n} = \int_0^\infty [H_n^*(x) - \bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(x)]d[F_n^*(x) - \bar{F}_{(n)}(x)];$$

$$(4.5) \qquad C_{4,n} = \int_0^\infty \{J[\tfrac{n}{n+1} H_n^*(x)] - J[\bar{H}_{(n)}(x)] - [\tfrac{n}{n+1} H_n^*(x) - \bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(x)]\}dF_n^*(x).$$

Straightforward computations using (2.8), (3.4), (3.5), (3.6), (2.9) and (2.10) yield that

$$(4.6) \qquad (1/n)\sum_{i=1}^{n} E\{B_n(X_i)\} = \mu_n^*, \ |\mu_n^*| < \infty;$$

$$(4.7) \qquad (1/n)\sum_{i=1}^{n} V\{B_n(X_i)\} = \gamma_n^2.$$

Further by lemma 3.1 and the assumption that $\inf_n \gamma_n > 0$,

$$(4.8) \qquad 0 < \gamma_n < \infty, \text{ uniformly in } F_1,\ldots,F_n \text{ and } n.$$

Finally, by lemma 3.2 and (4.6)

$$(4.9) \qquad (1/n)\sum_{i=1}^{n} E\{|B_n(X_i) - E[B_n(X_i)]|^{2+\delta}\} < \infty,$$

uniformly in $F_1,\ldots,F_n$ and n. Hence, the independent random variables $\{B_n(X_1),\ldots,B_n(X_n)\}$ satisfy Liapounoff's condition of the central limit theorem [cf. Gnedenko (1962, p. 322)]. Consequently,

$$(4.10) \qquad \mathcal{L}(n^{1/2}[(1/n)\sum_{i=1}^{n} B_n(X_i) - \mu_n^*]/\gamma_n) \to \mathcal{N}(0,1).$$

It remains only to prove that $C_{r,n} = o_p(n^{-\frac{1}{2}})$ for r=2,3,4 (uniformly in $F_1,\ldots,F_n$), and this is accomplished in the Appendix. Q.E.D.

5. General hypothesis of symmetry and robustness of T_n. Let $\mathcal{F}_o$ be the class of all continuous cdf's symmetric about 0. We want to test the general hypothesis

(5.1) $$H_o: \quad F_i \in \mathcal{F}_o \;\forall\; i=1,\ldots,n,$$

without bringing in the assumption that $F_1 \equiv \ldots \equiv F_n$. (5.1) is less restrictive than H_o^*: $F_1 \equiv \ldots \equiv F_n \equiv F \in \mathcal{F}_o$. It will be seen that T_n in (1.1) provides a robust test for H_o in (5.1), for all $F_i \in \mathcal{F}_o$, i=1,...,n. For this, it may be noted that

(5.2) $$F_i \in \mathcal{F}_o \;\forall\; i \Rightarrow \bar{F}_{(n)} \in \mathcal{F}_o \Rightarrow \bar{H}_{(n)}(x) = 2\bar{F}_{(n)}(x)-1 \text{ for } x \geq 0.$$

From (2.8) and (5.2), it follows that under H_o in (5.1),

(5.3) $$\mu_n^* = \tfrac{1}{2} \int_o^1 J(u)\,du = \tfrac{1}{2}\mu(\text{say}).$$

Upon using (5.2) and (3.1), straightforward computations yield that

(5.4) $$\gamma^2(\bar{F}_{(n)}) = \tfrac{1}{4} \int_o^1 J^2(u)\,du = \tfrac{1}{4}A^2(\text{say,}) \text{ if } \bar{F}_{(n)} \in \mathcal{F}_o.$$

Again, using (5.2) and integrating (3.3) by parts, it readily follows that

(5.5) $$\alpha_{n,i}+\beta_{n,i} = 0, \forall\; i, \text{ if } H_o \text{ in (5.1) holds.}$$

Consequently, from theorem 1, (5.3), (5.4), (5.5) and lemma 3.1, it follows that if H_o in (5.1) holds

(5.6) $$\mathcal{L}(2n^{\frac{1}{2}}[T_n-\tfrac{1}{2}\mu]/A) \to \mathcal{N}(0,1), \text{ uniformly in } F_1,\ldots,F_n.$$

This clearly indicates the robustness of T_n for arbitrarily symmetric $F_1,\ldots,F_n$. Thus, like the well-known sign-test (for location) we need not assume the identity of $F_1,\ldots,F_n$ (only identity of their medians is enough). However, unlike the sign-test, symmetry of each F_i (i=1,...,n) appears to be necessary.

REMARK 1. The sign-invariant permutation distribution theory of one-sample Chernoff-Savage statistics, developed in the more general multivariate case by Sen and Puri (1967), can be easily shown to be valid (for the univariate case) even when the sample observations are drawn from different (but symmetric) distributions. This permutation principle leaves scope for an exact test for H_o in (5.1) (based on T_n in (1.1)), when n is small but $F_1,\ldots,F_n$ are not necessarily all identical. Also, along the line of theorem 3.2 of Sen and Puri (1967), the asymptotic convergence of the sign-invariant permutation distribution of $2n^{\frac{1}{2}}[T_n-\frac{1}{2}\mu]/A$ to a standard normal distribution can be readily deduced. This patches up the link between small sample and large sample tests for H_o in (5.1) based on T_n.

REMARK 2. Not only T_n is used to test H_o in (5.1), but also it may be used to estimate the common median (μ, say) of $F_1,\ldots,F_n$. Thus the Hodges-Lehmann estimate of μ based on T_n is robust for any possible heterogeneity of (symmetric) $F_1,\ldots,F_n$.

REMARK 3. Recently, Puri (1967) has considered the problem of combining independent one sample tests of significance. His procedure encounters some difficulty when the number of sources is large but the number of observations in (all) the sources are not large. This difficulty can be readily avoided by considering an one-sample test based on all the samples pooled together. Here also, the symmetry of the different cdf's is enough, their identity is not necessary.

6. *Robust-efficiency of T_n*. Consider the sequence of shift alternatives $\{H_n\}$, where H_n specifies that $X_1,\ldots,X_n$ are independent random variables having absolutely continuous cdf's $F_{n1},\ldots,F_{nn}$, respectively, where $F_{ni}(x) = F_i(x-n^{\frac{1}{2}}c_i\theta)$, $i=1,\ldots,n$, (F_i's being all symmetric about 0), and $c_1,\ldots,c_n$ and θ are all real and finite. It is also assumed that F_i has a continuous (a.e.) density function f_i for all

$i=1,\ldots,n$. We define

(6.1) $$\bar{f}_{(n)}(x) = (1/n)\sum_{i=1}^{n} f_i(x) \text{ and } \bar{F}_{(n)}(x) = (1/n)\sum_{i=1}^{n} F_i(x).$$

Further, it is assumed that $\bar{f}_{(n)}(x)J[\bar{H}_{(n)}(x)]$ is bounded as $x\to\pm\infty$. Concerning the rank scores $\{E_{n,i}\}$, it is assumed that $E_{n,i}$ is the expected value of the ith order statistic of a sample of size n from a distribution function $\Psi(x)$ given by

(6.2) $$\Psi(x) = \Psi^*(x)-\Psi^*(-x-); \quad \Psi^*(-x-) = 1-\Psi^*(x),$$

where $\Psi^*(x)$ is a continuous cdf. This implies that

(6.3) $$J(u) = \Psi^{-1}(u) = \Psi^{*-1}(\tfrac{1+u}{2}) = J^*(\tfrac{1+u}{2}), \quad 0<u<1.$$

Define μ and A^2 as in (5.3) and (5.4), and let

(6.4) $$B(F) = \int_{-\infty}^{\infty} J^{*\prime}[F(x)]f^2(x)dx \text{ for all } F \in \mathfrak{F}_0.$$

Then by routine computations, it follows that under $\{H_n\}$

(6.5) $$\mu_n^* = \tfrac{1}{2}\mu+(\theta/2n^{\frac{1}{2}})(\frac{1}{n}\sum_{i=1}^{n} c_i \int_{-\infty}^{\infty} J^{*\prime}[\bar{F}_{(n)}(x)]\bar{f}_{(n)}(x)dF_i(x))+o(n^{-\frac{1}{2}});$$

(6.6) $$\gamma_n^2 = \tfrac{1}{4}A^2 + o(1).$$

Thus, it follows from theorem 1, (6.5) and (6.6) that

(6.7) $$\lim_{n\to\infty} P_{H_n}\{2n^{\frac{1}{2}}[T_n-\tfrac{1}{2}\mu]/A \le x+(\theta/A)\cdot$$

$$(\frac{1}{n}\sum_{i=1}^{n} c_i \int_{-\infty}^{\infty} J^{*\prime}[\bar{F}_{(n)}(x)]\bar{f}_{(n)}(x)dF_i(x))\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}dt,$$

for all real x. Let us now assume that the cdf F_i has the variance σ_i^2 for $i=1,\ldots,n$ and denote $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$, $\bar{\sigma}_n^2 = (1/n)\sum_{i=1}^{n}\sigma_i^2$. Then, by the well-known central limit theorem (for non-identically distributed independent random variables), we have for any real x

(6.8) $$\lim_{n\to\infty} P_{H_n}\{n^{\frac{1}{2}}\bar{X}_n|\bar{\sigma}_n \leq x+(\theta/\bar{\sigma}_n)\bar{c}_n\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}dt;\ \bar{c}_n = \frac{1}{n}\sum_{i=1}^{n} c_i.$$

Thus, if $\bar{c}_n$ is different from 0, the asymptotic relative efficiency (A.R.E.) of T_n with respect to $\bar{X}_n$ may be computed as

(6.9) $$e_n = e(T_n;\bar{X}_n) = [\bar{\sigma}_n^2/(A^2\bar{c}_n^2)][\frac{1}{n}\sum_{i=1}^{n} c_i \int_{-\infty}^{\infty} J^{*\prime}[\bar{F}_{(n)}(x)]\bar{f}_{(n)}(x)dF_i(x)]^2.$$

Thus, in general, the A.R.E. depends on $(c_1,\ldots,c_n)$ as well as on $F_1,\ldots,F_n$. Two special cases are of interest and yield some interesting results.

<u>Case I</u>. $c_1=\ldots=c_n=c\neq 0$, i.e., <u>equal shift but not necessarily identical cdfs</u>. It readily follows from (6.4) and (6.9), that $e_n (=e_n^{(1)})$ reduces to

(6.10) $$e_n^{(1)} = [B(\bar{F}_{(n)})\bar{\sigma}_n/A]^2,$$

which agrees with the expression for the Chernoff-Savage (1958) efficiency but for the cdf $\bar{F}_{(n)}$. As such, for normal scores statistic (i.e., when Ψ^* in (6.2) is a standard normal cdf), (6.10) will be at least as large as 1, where the equality sign holds only when $\bar{F}_{(n)}$ is itself normal. This clearly illustrates the robust-efficiency of the normal scores test. Incidentally, when $F_1,\ldots,F_n$ are all normal cdf's differing only in $\sigma_1^2,\ldots,\sigma_n^2$, $\bar{F}_{(n)}$ can not be normal, unless $\sigma_1=\ldots=\sigma_n$. Hence, for normal cdf's, the normal scores statistic will have an A.R.E. (relative to $\bar{X}_n$) ≥ 1, where the equality sign holds only when $\sigma_1=\ldots=\sigma_n$. For the Wilcoxon's signed rank statistic, similar results are already deduced in an earlier note [Sen (1968)] and hence the discussion is omitted.

<u>Case II</u>. $F_1=\ldots F_n=F$ <u>but</u> c_i's <u>are not all equal. i.e., homogeneous cdf's but heterogeneous shifts</u>. It follows from (6.4) and (6.9) that

(6.11) $$e_n = e_n^{(2)} = [B(F)\sigma/A]^2,$$

where σ^2 is the variance of the cdf F. This indicates that the A.R.E. is not affected by heterogeneity of the shifts, and the Chernoff-Savage (1958) bounds are equally applicable in this situation.

REMARK. In the two sample case, a distribution-free estimate of B(F) (defined by (6.4)), has been considered by Sen (1966) and the same procedure yields a similar estimate of B(F) in the one sample case when $F_1 \equiv \ldots \equiv F_n \equiv F$. It follows from (5.6), (6.7) and some routine computations along the line of Sen (1966) that this one-sample estimator estimates consistently (i) $B(\bar{F}_{(n)})$ in case I when $F_1, \ldots, F_n$ are not necessarily identical or (ii) B(F) in case II when $c_1, \ldots, c_n$ are not necessarily identical. Also, the interval estimation of the common median of $F_1, \ldots, F_n$ based on T_n remains valid even when $F_1, \ldots, F_n$ are not necessarily identical.

7. Appendix: higher order terms. Let (a_n, b_n) be the interval $S_{n,\varepsilon}$ such that

(7.1) $$S_{n,\varepsilon} = \{x: \bar{H}_{(n)}(x)[1-\bar{H}_{(n)}(x)] > n^{-1}\eta_\varepsilon\},$$

where ε is an arbitrary positive number and $\eta_\varepsilon(>o)$ depends on ε. Upon noting that

(7.2) $$P\{\max_i |X_i| \leq x\} = \prod_{i=1}^n H_i(x) \leq [\bar{H}_{(n)}(x)]^n,$$

(7.3) $$P\{\min_i |X_i| \geq x\} = \prod_{i=1}^n [1-H_i(x)] \leq [1-\bar{H}_{(n)}(x)]^n,$$

and proceeding as in Chernoff and Savage (1958, p. 986), it follows that

(7.4) $$P\{|X_i| \in S_{n,\varepsilon} \forall\ i\} \geq 1-\varepsilon, \text{ uniformly in } F_1, \ldots, F_n.$$

Thus, from (4.3) and (7.4), it follows that with probability $\geq 1-\varepsilon$

(7.5) $$|C_{2,n}| \leq \frac{1}{n+1} \int_{S_{n,\varepsilon}} |J'[\bar{H}_{(n)}(x)]| dF_n^*(x) \leq \frac{1}{n} \int_{S_{n,\varepsilon}} |J'[\bar{H}_{(n)}(x)]| dH_n^*(x).$$

From (2.7) and (7.1), it follows that for all $x \in S_{n,\varepsilon}$

(7.6) $$n^{-\frac{1}{2}(1+\delta)} |J'[\bar{H}_{(n)}(x)]| \leq K^*\{\bar{H}_{(n)}(x)[1-\bar{H}_{(n)}(x)]\}^{-1+\delta/2}, \ K^* < \infty .$$

Since, the extreme right hand side of (7.5) involves average over independent random variables, by Markov's law of large numbers

(7.7) $$\int_{S_{n,\varepsilon}} [\bar{H}_{(n)}(x)\{1-\bar{H}_{(n)}(x)\}]^{-1+\delta/2} dH_n^* \overset{p}{\to} \int_o^1 [u(1-u)]^{-1+\delta/2} du<\infty \ .$$

Therefore, (7.5), (7.6) and (7.7) yield that

(7.8) $$|C_{2,n}| = O_p(n^{-\frac{1}{2}(1+\delta)}) = o_p(n^{-\frac{1}{2}}), \text{ uniformly in } F_1,\ldots,F_n.$$

For $C_{3,n}$ and $C_{4,n}$, we require the following theorems.

<u>THEOREM 7.1</u>. <u>For any</u> $\varepsilon>o$ $\exists$ $c(\varepsilon)$ $(<\infty)$, <u>such that for</u> $\delta'>o$

$$P\{\sup_x \frac{n^{\frac{1}{2}}|H_n^*(x)-\bar{H}_{(n)}(x)|}{\{\bar{H}_{(n)}(x)[1-\bar{H}_{(n)}(x)]\}^{\frac{1}{2}-\delta'}} \geq c(\varepsilon)\} \leq \varepsilon,$$

<u>uniformly in</u> $F_1,\ldots,F_n$.

<u>PROOF</u>. Define the stochastic process $V_n(t)$ by

(7.9) $$V_n(t) = n[H_n^*(t)-(1-n^{-\frac{1}{2}})\bar{H}_{(n)}(t)]^2, \ t\geq 0.$$

Then, by direct computations

(7.10) $$\nu_n(t) = E\{V_n(t)\} = \bar{H}_{(n)}(t) - \frac{1}{n}\sum_{i=1}^n [H_i(t)-\bar{H}_{(n)}(t)]^2 \leq \bar{H}_{(n)}(t)<1,$$

for all $t\geq 0$. Also, on using (2.1), (2.2), (2.3) by straightforward manipulations, it follows that

(7.11) $$E\{V_n(t)|V_n(s),s<t\} = V_n(s)+[\bar{H}_{(n)}(t)-\bar{H}_{(n)}(s)]^2+2n^{\frac{1}{2}}[\bar{H}_{(n)}(t)-\bar{H}_{(n)}(s)]H_n^*(s)$$
$$+\frac{1}{n}\sum_{i=1}^n [H_i(t)-H_i(s)][1-H_i(t)+H_i(s)] \geq V_n(s)$$

Therefore $\{V_n(t), \ t\geq 0\}$ is a semimartingale, such that $0\leq d\nu_n(t)\leq(1+2\bar{H}_{(n)}(t) \cdot d\bar{H}_{(n)}(t)$.

Define $g_n(t) = K^2[\bar{H}_{(n)}(t)\{1-\bar{H}_{(n)}(t)\}]^{1-2\delta'}$ $(\delta>0)$. Then (i) $\lim_{t\to o} \nu_n(t)/g_n(t) = 0$ and (ii) $\int_o^\infty [1/g_n(t)]dH_n(t)<\infty$. Thus, on defining t_n^o by $\bar{H}_{(n)}(t_n^o) = \frac{1}{2}$ and using theorem 5.1 of Birnbaum and Marshall (1961), it follows that

$$P\{\sup_{t\varepsilon[o,t_n^o]} [V_n(t)/g_n(t)] \geq 1\} \leq \frac{2}{K^2} \int_o^{\frac{1}{2}} \{u(1-u)\}^{-1+2\delta'}du, \tag{7.12}$$

(as $1+2\bar{H}_n(t)\leq 2$ for all $t \leq t_n^o$). But, $V_n(t)/g_n(t)$ is equal to

$$\left[\frac{n^{\frac{1}{2}}[H_n^*(t)-\bar{H}_{(n)}(t)]}{K\{\bar{H}_{(n)}(t)[1-\bar{H}_{(n)}(t)]\}^{\frac{1}{2}-\delta'}} + \frac{\{\bar{H}_{(n)}(t)\}^{2\delta'}}{K}\left\{\frac{\bar{H}_{(n)}(t)}{1-\bar{H}_{(n)}(t)}\right\}^{\frac{1}{2}-\delta'}\right]^2 . \tag{7.13}$$

Hence, it follows from (7.12) and (7.13) that

$$P\{\sup_{t\varepsilon[o,t_n^o]} \frac{n^{\frac{1}{2}}|H_n^*(t)-\bar{H}_{(n)}(t)|}{\{\bar{H}_{(n)}(t)[1-\bar{H}_{(n)}(t)]\}^{\frac{1}{2}-\delta'}} \geq K+1\} \leq \frac{2}{K^2} \int_o^{\frac{1}{2}} \{u(1-u)\}^{-1+2\delta'}du. \tag{7.14}$$

In a similar manner, it can be shown that

$$P\{\sup_{t\varepsilon[t_n^o,\infty)} \frac{n^{\frac{1}{2}}|H_n^*(t)-\bar{H}_{(n)}(t)|}{\{\bar{H}_{(n)}(t)[1-\bar{H}_{(n)}(t)]\}^{\frac{1}{2}-\delta'}} \geq K+1\} \leq \frac{2}{K^2} \int_{\frac{1}{2}}^{1} \{u(1-u)\}^{-1+2\delta'}du. \tag{7.15}$$

Since $2\int_o^1\{u(1-u)\}^{-1+2\delta'}du = c_{\delta'}<\infty$, K can always be so selected that $c_\delta/K^2\leq\varepsilon$, $K+1=c(\varepsilon)$. With this choice of K (say, K_ε), the proof of the theorem readily follows from (7.14) and (7.15). Q.E.D.

REMARK. The theorem generalizes lemma 6 of [5] to non-identically distributed random variables without unnecessarily using the Poisson distribution in conjunction with a binomial distribution; the latter approach becomes quite involved in the case of non-identical cdf's considered above.

THEOREM 7.2. $\sup_{H_n^*>o} \bar{H}_{(n)}(t)/H_n^*(t) = O_p(1)$, uniformly in $F_1,\ldots,F_n$.

PROOF. For $\bar{H}_n(t) \geq cn^{-\frac{1}{2}}$, $c>o$, the proof readily follows from that of theorem 7.1 and some routine computations. For $\bar{H}_n(t) < cn^{-\frac{1}{2}}$, one can apply the well-known results on standard Poisson process to deduce the desired result. Q.E.D.

COROLLARY. $\sup_{H_n^*<1} \{[1-\bar{H}_{(n)}(t)]/[1-H_n^*(t)]\} = O_p(1)$, uniformly in $F_1,\ldots,F_n$.

Let now $\bar{S}_{n,\varepsilon}$ be the complementary interval to $S_{n,\varepsilon}$. Then,

$$C_{3,n} = \int_{S_{n,\varepsilon}} + \int_{\bar{S}_{n,\varepsilon}} [H_n^*(x)-\bar{H}_{(n)}(x)]J'[\bar{H}_{(n)}(x)]d[F_n^*(x)-\bar{F}_{(n)}(x)]. \tag{7.16}$$

Since, with probability $\geq 1-\varepsilon$, there is no observation in $\bar{S}_{n,\varepsilon}$, and as $d\bar{F}_{(n)} \leq d\bar{H}_{(n)}$, it is easily seen on using (2.7) that the integral over $\bar{S}_{n,\varepsilon}$ is $O_p(n^{-\frac{1}{2}-\delta})$ i.e., $o_p(n^{-\frac{1}{2}})$. Again, on making use of theorem 7.1 and (2.7) (with $0<\delta'<\delta$), straightforward computations yield that in (7.16), the integral over $S_{n,\varepsilon}$ is also $o_p(n^{-\frac{1}{2}})$. Thus, $C_{3,n} = o_p(n^{-\frac{1}{2}})$. Finally, with theorem 7.2 and its corollary, the proof of $C_{4,n} = o_p(n^{-\frac{1}{2}})$ follows by routine computations and is omitted.

REFERENCES

[1] BIRNBAUM, Z. W., and MARSHALL, A. W. (1961) Some multivariate Chebyshev inequalities with extensions to continuous parameter process. Ann. Math. Statist. 32, 687-703.

[2] CHERNOFF, H., and SAVAGE, I. R. (1958) Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. 29, 972-994.

[3] GNEDENKO, B. V. (1962) Theory of probability. (Translated by B. D. Seckler). New York: Chelsea Publishing Co.

[4] GOVINDARAJULU, Z. (1960) Central limit theorem and asymptotic efficiency of one sample nonparametric procedures. Tech. Rep. 11. Dept. Statist., Univ. Minnesota.

[5] GOVINDARAJULU, Z., LECAM, L., and RAGHAVACHRI, M. (1965) Generalizations of theorems of Chernoff and Savage on the asymptotic normality of test statistics. Proc. Fifth Berkeley Symp. Math. Statist. Prob. (Univ. Calif. Press) 1, 608-638.

[6] PURI, M. L. (1967) Combining independent one-sample tests of significance. Ann. Inst. Statist. Math. 19, 285-300.

[7] PYKE, R., and SHORACK, G. (1967) A Chernoff-Savage theorem for random sample sizes. (Abstract). Ann. Math. Statist. 38, 1313.

[8] SEN, P. K. (1966) On a distribution-free method of estimating asymptotic efficiency of a class of nonparametric tests. Ann. Math. Statist. 37, 1759-1770.

[9] SEN, P. K. (1968) On a further robustness property of the test and estimator based on Wilcoxon's signed rank statistic. Ann. Math. Statist. 39, No. 1 (in press).

[10] SEN, P. K. and PURI, M. L. (1967) On the theory of rank order tests for location in the multivariate one sample problem. Ann. Math. Statist. 38, 1216-1228.