ASYMPTOTIC NORMALITY AND EFFICIENCY OF CERTAIN NONPARAMETRIC TEST STATISTICS[1]

By Herman Chernoff and I. Richard Savage

Stanford University and *University of Minnesota*

1. Summary. Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be ordered observations from the absolutely continuous cumulative distribution functions $F(x)$ and $G(x)$ respectively. If $z_{Ni} = 1$ when the ith smallest of $N = m + n$ observations is an X and $z_{Ni} = 0$ otherwise, then many nonparametric test statistics are of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni}.$$

Theorems of Wald and Wolfowitz, Noether, Hoeffding, Lehmann, Madow, and Dwass have given sufficient conditions for the asymptotic normality of T_N. In this paper we extend some of these results to cover more situations with $F \neq G$. In particular it is shown for all alternative hypotheses that the Fisher-Yates-Terry-Hoeffding c_1-statistic is asymptotically normal and the test for translation based on it is at least as efficient as the t-test.

2. Introduction. Finding the distributions of nonparametric test statistics and establishing optimum properties of these tests for small samples has progressed slower than the corresponding large sample theory. Even so, it is not possible to state that the basic framework of the large sample theory has been completed. Dwass [3] has recently presented a general theorem on the asymptotic normality of certain nonparametric test statistics under alternative hypotheses. His results, however, do not apply to such important and interesting procedures as the c_1-test [11]. Many papers have appeared giving the asymptotic efficiency of particular tests. Hodges and Lehmann [7] have discussed the asymptotic efficiency of the Wilcoxon test with respect to all translation alternatives. In the same paper they have conjectured that the c_1-test is as efficient as the t-test for normal alternatives and at least as efficient as the t-test for all other alternatives.

The beginning of our work came from a desire to verify the Hodges and Lehmann conjecture. Related to the conjecture is the hypothesis that the c_1-statistic is asymptotically normally distributed. Thus our work has two parts: developing a new theorem for asymptotic normality of nonparametric test statistics and the establishing of the variational argument required for determining the minimum efficiency of test procedures.

Received November 26, 1957; revised May 2, 1958.

[1] This work was done with the partial support of the Office of Naval Research while I. R. Savage was at Stanford University and the Center for Advanced Study in the Behavioral Sciences.

Our basic result on the asymptotic normality of statistics of the form T_N is Theorem 1 of Section 4. This theorem is a partial generalization of results of Dwass [3] summarized in our Theorem 4. Theorem 1 is not given in the most general form possible. Our choice of the level of generality was to facilitate our writing and your reading.

Section 3 contains our basic notation and assumptions. Section 4 contains statements of the theorem on asymptotic normality as well as the basic portion of the proof. Details regarding the negligibility of the remainder terms are given in Section 7. The variational arguments are presented in Section 5 and Section 6 relates our Theorem 1 to Dwass's results. Applications of Theorem 1 to several nonparametric tests are given in Section 6.

3. Assumptions and notation. Let $X_1, X_2, \cdots, X_m$ be the ordered observations of a random sample from a population with continuous cumulative distribution function $F(x)$. Let $Y_1, Y_2, \cdots, Y_n$ be the ordered observations of a random sample from a population with continuous cumulative distribution function $G(x)$. Let $N = m + n$ and $\lambda_N = m/N$ and assume that for all N the inequalities $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$ hold for some fixed $\lambda_0 \leqq \frac{1}{2}$.

Let $F_m(x) =$ (number of $X_i \leqq x)/m$ and $G_n(x) =$ (number of $Y_i \leqq x)/n$. Thus $F_m(x)$ and $G_n(x)$ are the sample cumulative distribution functions of the X's and Y's respectively. Define $H_N(x) = \lambda_N F_m(x) + (1 - \lambda_N)G_n(x)$. Thus $H_N(x)$ is the combined sample cumulative distribution function. The combined population cumulative distribution function is $H(x) = \lambda_N F(x) + (1 - \lambda_N)G(x)$. Even though $H(x)$ depends on N (or rather m and n) through λ_N our notation suppresses this fact for convenience. In fact $F(x)$ and $G(x)$ may actually depend on N although this will not be stated explicitly. In Corollary 1 the distributions do depend on N. The point for suppressing this fact is that our limit theorems are "uniform" and hold, whether the distributions are constant, tend to a limit, or vary rather arbitrarily with the sample size N.

If the ith smallest in the combined sample is an X let $z_{Ni} = 1$ and otherwise let $z_{Ni} = 0$. Then our concern is with statistics of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni}\, z_{Ni}, \tag{3.1}$$

where the E_{Ni} are given numbers. (The special case where $E_{Ni} = E(i/N)$ is particularly easily handled by our methods. For the Wilcoxon test this condition is met with $E_{Ni} = i/N$, and Freund and Ansari [6] have considered $E_{Ni} = E(i/N) = |\frac{1}{2} - i/N|$ in testing for the equality of dispersion of two populations.) The definition (3.1) of T_N is the one conventionally used. We shall, however, use the following representation:

$$T_N = \int_{-\infty}^{\infty} J_N[H_N(x)]\, dF_m(x). \tag{3.2}$$

The definitions (3.1) and (3.2) are equivalent when $E_{Ni} = J_N(i/N)$. A representation like (3.2) was used by Blum and Weiss [1, page 243, Eq. 2.4] and R. v. Mises considered $\int \varphi(x)\, dF_m(x)$ in detail [9].

Throughout our proofs K will be used as a generic constant which may depend on J_N but it will not depend on $F(x)$, $G(x)$, m, n, N. Statements involving o_p or O_p will always be uniform in $F(x)$, $G(x)$, and $H(x)$, and λ_N in the interval $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$.

While J_N need be defined only at $1/N, 2/N, \cdots, N/N$, we shall find it convenient to extend its domain of definition to (0,1] by some convention such as letting J_N be constant on $(i/N, (i + 1)/N]$.

Let I_N be the interval in which $0 < H_N(x) < 1$. Then I_N is closed on the left at the smallest observation and open on the right at the largest observation. The interval, I_N, has a random location.

4. Asymptotic normality.

THEOREM 1. *If*

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

(2) $\int_{I_N} [J_N(H_N) - J(H_N)]\, dF_m(x) = o_p(N^{-1/2})$,

(3) $J_N(1) = o(\sqrt{N})$,

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

for $i = 0, 1, 2,$ *and for some* $\delta > 0$,

then, for fixed F, G *and* λ_N,

$$\lim_{N\to\infty} P\left(\frac{T_N - \mu_N}{\sigma_N} \leqq t\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \tag{4.1}$$

where

$$\mu_N = \int_{-\infty}^{\infty} J[H(x)]\, dF(x) \tag{4.2}$$

and

$$\begin{aligned} N\sigma_N^2 = 2(1 - \lambda_N)\Bigg\{ &\iint_{-\infty<x<y<\infty} G(x)[1 - G(y)]J'[H(x)]J'[H(y)]\, dF(x)\, dF(y) \\ &+ \frac{(1 - \lambda_N)}{\lambda_N} \iint_{-\infty<x<y<\infty} F(x)[1 - F(y)]J'[H(x)]J'[H(y)]\, dG(x)\, dG(y)\Bigg\}, \end{aligned} \tag{4.3}$$

providing $\sigma_N \neq 0$.

In Eqs. 4.1 and 4.3 we put subscripts on μ and σ to recall that these depend on F, G and λ_N and are meaningful in the more general case where F, G, and λ_N are not fixed. Corollary 1 will extend Theorem 1 to obtain convergence to normality uniformly with respect to F, G, and λ_N for a broad range of F, G, and λ_N.

To facilitate the proof of Corollary 1, we will regard F, G, and λ_N as variable throughout the proof of Theorem 1 except where it is specified otherwise.

Assumption 1 is likely to be filled whenever one speaks of a sequence of tests. In the special case $E_{Ni} = E(i/N)$ of course $J_N = E = J$ and Assumption 2 will automatically be satisfied. Theorem 2 shows that Assumptions 1, 2 and 3 are often satisfied when the E_{Ni} are the mean values of order statistics. Assumption 4 is the basic condition. The assumption has two functions: it limits the growth of the coefficients E_{Ni} and it supplies certain smoothness properties. Both conditions are essential to our argument. We believe that the theorem is true without the smoothness condition.

Proof. To begin the proof we rewrite T_N as

$$T_N = \int_{-\infty}^{\infty} J_N(H_N)\,dF_m(x) = \int_{I_N} [J_N(H_N) - J(H_N)]\,dF_m(x) + \int_{I_N} J(H_N)\,dF_m(x) + \int_{H_N=1} J_N(H_N)\,dF_m.$$

In the second integral we write $dF_m = d(F_m - F + F)$, $J(H_N) = J(H) + (H_N - H)J'(H) + [(H_N - H)^2/2]J''[\varphi H_N + (1-\varphi)H]$, where $0 < \varphi < 1$, and $H = \lambda_N F + (1 - \lambda_N)G$. After multiplying out the expression becomes

$$T_N = A + B_{1N} + B_{2N} + \sum_{i=1}^{6} C_{iN},$$

where

$$A = \int_{0<H<1} J(H)\,dF(x), \tag{4.4}$$

$$B_{1N} = \int_{0<H<1} J(H)\,d[F_m(x) - F(x)], \tag{4.5}$$

$$B_{2N} = \int_{0<H<1} (H_N - H)J'(H)\,dF(x), \tag{4.6}$$

$$C_{1N} = \lambda_N \int_{0<H<1} (F_m - F)J'(H)\,d[F_m(x) - F(x)] \tag{4.7}$$

$$C_{2N} = (1 - \lambda_N) \int_{0<H<1} (G_n - G)J'(H)\,d[F_m(x) - F(x)], \tag{4.8}$$

$$C_{3N} = \int_{I_N} \frac{(H_N - H)^2}{2} J''[\varphi H_N + (1-\varphi)H]\,dF_m(x), \tag{4.9}$$

$$C_{4N} = \int_{H_N=1} [-J(H) - (H_N - H)J'(H)]\,dF_m(x), \tag{4.10}$$

$$C_{5N} = \int_{I_N} [J_N(H_N) - J(H_N)]\,dF_m(x), \tag{4.11}$$

$$C_{6N} = \int_{H_N=1} J_N(H_N)\,dF_m(x). \tag{4.12}$$

The A, B, C terms represent the "constant," "first order random," and "higher order random" portions respectively of T_N. In this section a detailed study of the A and B terms is made and in Section 7 it is shown that the C terms are of higher order.

The "constant" term, $A = \int_{0<H<1} J(H)\, dF(x)$, is finite as a result of Assumption 4 of Theorem 1; see Section 7.A.10. Since A depends on λ_N as well as $F(x)$ and $G(x)$ it need not converge as $N \to \infty$, but it does remain bounded.

Integrating B_{2N} by parts and using the fact that

$$\int_{-\infty}^{\infty} d[F_m(x) - F(x)] = 0,$$

we obtain

$$B_{1N} + B_{2N} = [1 - \lambda_N]\left\{\int_{-\infty}^{\infty} B(x)\, d[F_m(x) - F(x)] - \int_{-\infty}^{\infty} B^*(x)\, d[G_n(x) - G(x)]\right\}, \tag{4.13}$$

where

$$B(x) = \int_{x_0}^{x} J'[H(y)]\, dG(y) \tag{4.14}$$

$$B^*(x) = \int_{x_0}^{x} J'[H(y)]\, dF(y) \tag{4.15}$$

and

$$\lambda_N B^*(x) + (1 - \lambda_N)B(x) = J[H(x)] - J[H(x_0)]$$

with x_0 determined somewhat arbitrarily, say by $H(x_0) = 1/2$.

Thus,

$$B_{1N} + B_{2N} = [1 - \lambda_N]\left\{\frac{1}{m}\sum_{i=1}^{m} [(BX_i) - \mathcal{E}B(X)] - \frac{1}{n}\sum_{i=1}^{n} [B^*(Y_i) - \mathcal{E}B^*(Y)]\right\}, \tag{4.16}$$

where $\mathcal{E}$ represents expectation and X and Y have the F and G distributions respectively.

The two summations involve independent samples of identically distributed random variables. Therefore, if F, G, and λ_N are fixed, $B(X)$ and $B^*(Y)$ are specified random variables and we may apply the central limit theorem to show that $B_{1N} + B_{2N}$ when properly normalized has a Gaussian distribution in the limit. The central limit theorem applies if the variances of $B(X)$ and $B^*(Y)$ are finite and at least one is positive.

First, we shall find a bound on the moments of $B(X)$ and $B^*(Y)$:

$$|B(x)| = \left|\int_{x_0}^{x} J'[H(y)]\, dG(y)\right| \leq K[H(x)[1 - H(x)]]^{-\frac{1}{2}+\delta}.$$

Thus for $\delta' > 0$ such that $(2 + \delta')(-\frac{1}{2} + \delta) > -1$,

$$\mathcal{E}\{| B(X) |\}^{2+\delta'} \leqq K \int_{-\infty}^{\infty} [H(x)[1 - H(x)]]^{(-\frac{1}{2}+\delta)(2+\delta')} \, dF(x)$$

$$\leqq K \int_0^1 [H(1 - H)]^{(-\frac{1}{2}+\delta)(2+\delta')} \, dH \leqq K,$$

having made use of $dG \leqq (1/\lambda_0) \, dH$. (See Section 7.A.8.)

Similarly, we may bound the $2 + \delta'$ absolute moments of $B^*(Y)$. The asymptotic normality of $B_{1N} + B_{2N}$ follows providing $B(X)$ and $B^*(Y)$ do not both have zero variance.

We compute the variances of $B(X)$ and $B^*(Y)$. These can be expressed in terms of $\int B(x) \, dF(x)$, $\int B^2(x) \, dF(x)$, etc., but we shall use a slightly different approach.

$$B(X) - \mathcal{E}B(X) = \int_{-\infty}^{\infty} B(x) \, d[F_1(x) - F(x)]$$

$$= -\int_{-\infty}^{\infty} [F_1(x) - F(x)] J'[H(x)] \, dG(x)$$

has variance

$$\sigma^2_{B(X)} = \mathcal{E}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F_1(x) - F(x)][F_1(y) - F(y)] J'[H(x)] J'[H(y)] \, dG(x) \, dG(y)\right\},$$

and

$$\sigma^2_{B(X)} = 2 \iint_{-\infty<x<y<\infty} F(x)[1 - F(y)] J'[H(x)] J'[H(y)] \, dG(x) \, dG(y), \tag{4.17}$$

if it is permitted to interchange expectation and integral. That this may be done follows from Fubini's theorem when it is seen that for $x < y$,

$$\mathcal{E}\{| F_1(x) - F(x) | \, | F_1(y) - F(y) |\} \leqq KF(x)[1 - F(y)]$$

and that the last integral above is finite. (In fact this integral is bounded in the argument dealing with (C_{23N}) in Section 7.B.)

Similarly, the variance of $B^*(Y)$ is given by

$$\sigma^2_{B^*(Y)} = 2 \iint_{-\infty<x<y<\infty} G(x)[1 - G(y)] J'[H(x)] J'[H(y)] \, dF(x) \, dF(y). \tag{4.18}$$

These two variances when combined give the variance result stated in (4.3). We review the status of our proof. In Section 7, the C terms are shown to be "higher order uniformly." The A term is non-random and finite. Finally

$$B_{1N} + B_{2N}$$

is the sum of two independent terms each of which is the average of random variables with mean 0 and finite second moments. Theorem 1 follows.

The proof given can be extended to the case where F, G and λ_N are not fixed. To obtain uniform convergence to normality, we apply a theorem of Esseen

([4], p. 43) which is a generalization of the so-called Berry-Esseen theorem ([8], p. 288)[2]. Since the C terms are uniformly $o_p(1/\sqrt{N})$ it suffices to obtain uniform convergence for $B_{1N} + B_{2N}$. For this it suffices to bound $\rho_{2+\delta'}$ for $B(X)$ and $B^*(Y)$. Since we bounded the absolute $2 + \delta'$ moments, all that is required is to bound the variances of $B(X)$ and $B^*(Y)$ away from 0 and to have m and $n \to \infty$. Thus we have

COROLLARY 1. *If the conditions* 1 *to* 4 *of Theorem* 1 *are satisfied, and* F, G, *and* $\lambda_N(0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1)$ *are restricted to a set for which* $B(X)$ *and* $B^*(Y)$ *have variances bounded away from* 0, *then Eq.* 4.1 (*asymptotic normality*) *holds uniformly with respect to* F, G, *and* λ_N.

COROLLARY 2. *If conditions* 1 *to* 4 *of Theorem* 1 *are satisfied,* $0 < \lambda_0 \leqq \lambda_N \leqq 1 - \lambda_0 < 1$,

$$F(x) = \Psi(x - \theta_N),$$
$$G(x) = \Psi(x - \varphi_N),$$

where Ψ *has a density* ψ, *then Eq.* 4.1 *holds uniformly with respect to* λ_N, θ_N *and* φ_N *for* $\varphi_N - \theta_N$ *in some neighborhood of* 0. *If* $\varphi_N - \theta_N \to 0$,

$$\lim_{N\to\infty} \frac{\lambda_N N \sigma_N^2}{(1 - \lambda_N)} = 2 \iint_{0<x<y<1} x(1 - y)J'(x)J'(y)\, dx\, dy = \int_0^1 J^2(x)\, dx - \left[\int_0^1 J(x)\, dx\right]^2. \tag{4.19}$$

PROOF. It suffices to show that $B(X)$ and $B^*(Y)$ have variances bounded away from zero and to establish Eq. 4.19. Since J is not constant and has a second derivative, there is an interval of u in which $J'(u)$ is bounded away from 0 and in which $J'(u) > 0$ or in which $J'(u) < 0$. There is a corresponding interval of x for which $\Psi(x)$ lies in the u interval and its density $\psi(x)$ is almost everywhere bounded away from 0. For $\varphi_N - \theta_N$ small enough, there is an x interval whose length is bounded away from 0 where the densities $f(x) = \psi(x - \theta_N)$ and $g = \psi(x - \varphi_N)$ are almost everywhere bounded away from 0 and $J'[H(x)]$ is bounded away from zero. It follows that $B(X)$ and $B^*(Y)$ have variances bounded away from zero.

All that remains is to establish Eq. 4.19. The first equality follows directly from Theorem 1 by letting $F(x) = x^*$ and $G(x) \to x^*$. The second equality can be obtained by interpreting the double integral as

$$\iiiint_{0<u<x<y<v<1} J'(x)J'(y)\, du\, dx\, dy\, dv$$

[2] Esseen's theorem states that if $X_1, X_2, \cdots, X_n$ are independent observations from a population with mean 0, variance σ^2, and finite absolute $2 + \delta'$ moment $\beta_{2+\delta'}$, $0 < \delta' \leqq 1$, then $|F^* - \Phi^*| < C(\delta')\left[\frac{\rho_{2+\delta'}}{n^{\delta'/2}} + \frac{\rho_{2+\delta'}^{1/\delta'}}{n^{1/2}}\right]$ where F^* is the cdf. of $\bar{X}$, Φ^* is the approximating normal cdf, C depends only on δ' and $\rho_{2+\delta'} = \beta_{2+\delta'}/\sigma^{2+\delta'}$.

and integrating with respect to y first and x second. It can also be obtained by considering a standard derivation [13] of the asymptotic distribution of T_N when $F = G$ where T_N is regarded as the average of a sample of m from the population of N numbers $J_N(1/N), J_N(2/N), \cdots, J_N(N/N)$.

We remark that normalizing J so that $\int_0^1 J(x)\,dx = 0$ and $\int_0^1 J^2(x)\,dx = 1$ will not affect the efficiency of the test. Furthermore, if J is the inverse of a cdf, the right-hand side of (4.19) is the variance of that distribution.

In applying Theorem 1 the verification of condition 2 may cause some difficulty. The following Theorem 2 gives a simple sufficient condition under which conditions 1, 2, and 3 hold. In particular with the use of Theorem 2 it is simple to verify that the distribution of the c_1-statistic does approach a Gaussion distribution for alternative hypotheses.

THEOREM 2. *If $J_N(i/N)$ is the expectation of the ith order statistic of a sample of size N from a population whose cumulative distribution function is the inverse function of J and*

$$| J^{(i)}(u) | \leqq K[u(1-u)]^{-i-\frac{1}{2}+\delta}, \qquad i = 0, 1, 2,$$

then

$$\lim_{N\to\infty} J_N(H) = J(H), \qquad 0 < H < 1,$$

$$J_N(1) = o(N^{1/2}),$$

and

$$\int_{I_N} [J_N(H_N) - J(H_N)]\,dF_m(x) = o(N^{-1/2}).$$

(We write o instead of o_p because the random sequence is bounded by a nonrandom sequence which is $o(N^{-1/2})$. In fact $| \int [J_N(H_N) - J(H_N)]\,dF_m(x) | \leqq (1/\lambda) \int | J_N(H_N) - J(H_N) |\,dH_N(x)$ and our proof essentially shows that this latter integral which is non-random and independent of F and G, is $o(N^{-1/2})$.)

PROOF. It is well known that $J_N(H) \to J(H)$. A proof of the other two results is given in Section 7.C.

5. Variational argument. We have now established that the limiting distribution of the c_1-statistic is Gaussian. Thus we may proceed with the study of the efficiency of this test procedure. We will examine translation alternatives only. Since the power of the c_1-test approaches one when the distributions F and G are held fixed as N approaches infinity we restrict our consideration to the following situation.

There is a distribution function $\Psi(x)$ which does not depend on N and $F(x) = \Psi(x - \theta)$ and $G(x) = \Psi(x - \varphi)$. We test the hypothesis that $\Delta = \theta - \varphi = 0$ vs. "near" alternatives of the form $\Delta = \Delta_N = cN^{-1/2}$. We will also assume that

$$0 < \lim_{N\to\infty} \lambda_N = \lambda < 1.$$

With this framework we are able to use the Pitman criterion (the one considered by Hodges and Lehmann) for finding efficiencies of test procedures. The following conditions have been established for the c_1-statistic if Ψ has a density and clearly hold for the t-statistic if Ψ has finite second moments. There are functions $a_N(\Delta)$ and $b_N(\Delta)$ such that for Δ in some neighborhood of 0,

$$\mathfrak{L}\left(\frac{T_N - a_N(\Delta)}{b_N(\Delta)}\right) \Rightarrow N(0, 1), \tag{5.1}$$

$$\operatorname*{limit}_{N\to\infty} \frac{b_N(\Delta_N)}{b_N(0)} = 1, \tag{5.2}$$

$$E_T = \operatorname*{limit}_{N\to\infty} \left[\frac{[a_N(\Delta_N) - a_N(0)]}{\Delta_N N^{1/2} b_N(0)}\right]^2 \tag{5.3}$$

exists and is independent of c.

The quantity E_T is called the efficacy of the procedure based on the sequence of statistics T_N. Of course E_T depends on Ψ. In comparing two sequences of tests, say T_N and T_N^*, for the same pair of near alternatives the two tests will have the same power only when the corresponding sample sizes, N and N^*, satisfy the following relationship

$$\operatorname*{limit}_{N\to\infty} \frac{N^*}{N} = \frac{E_T}{E_{T^*}} = E_{T,T^*} \tag{5.4}$$

if $E_{T^*} \neq 0$. *E_{T,T^*} is called the asymptotic relative efficiency of T_N with respect to T_N^*.*

Let $E_{c_1,t}(\Psi)$ denote the asymptotic efficiency relative to the t-test of the c_1-test against translation alternatives. Then we have $J = J_0$ the inverse of the normal $N(0, 1)$ cdf Φ and applying Corollary 1 and using derivatives in the expression for E_T, we have

$$E_{c_1,t}(\Psi) = I_{1\Psi}^2/\sigma^2, \tag{5.5}$$

where

$$I_{1\Psi} = \int J_0'[\Psi(x)]\psi^2(x)\,dx \tag{5.6}$$

and σ^2 is the variance of the distribution with cdf Ψ (and density ψ).[3] Normalizing Ψ to have mean 0 and variance 1 does not affect $E_{c_1,t}(\Psi)$ which then becomes equal to $I_{1\Psi}^2$. In this section we shall prove

THEOREM 3. *If Ψ is a cdf with a density and finite second moment, then $E_{c_1,t}(\Psi) \geqq 1$, and $E_{c_1,t}(\Psi) = 1$ only if Ψ is normal.*

PROOF. It suffices to show that the minimum of $I_{1\Psi}$ subject to the restrictions

$$I_{2\Psi} = \int x\psi(x)\,dx = 0$$

[3] If Ψ does not have finite variance σ^2, $E_{c_1,t}$ is not defined but it makes sense to regard it as ∞.

and

$$I_{3\Psi} = \int x^2\psi(x)\,dx = 1$$

is attained only for $\Psi = \Phi$ and that $I_{1\Phi} = 1$.

A density $\psi(x)$ assigns to each x a value of Ψ and a corresponding value of $J_0[\Psi(x)]$. If $\psi(x) = 0$ a.e. on an interval, this interval corresponds to a fixed value of $J_0[\Psi(x)]$. If x is then regarded as a function of J_0, it is multivalued at that value of J_0. Otherwise x is continuous and it is increasing in J_0. Conversely any monotone non-decreasing function x of J_0 determines a corresponding cdf Ψ. We have

$$u = \Phi[J_0(u)],$$

$$J_0'(u) = \frac{1}{\varphi[J_0(u)]},$$

and

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}.$$

Furthermore

(5.7)
$$\int_{-\infty}^{x}\psi(t)\,dt = \Psi(x) = \int_{-\infty}^{J_0}\varphi(t)\,dt$$

and

$$\psi(x)\,dx = d\Psi(x) = \varphi(J_0)\,dJ_0.$$

Consequently our problem consists of finding a monotone function $x(J_0)$ which minimizes

(5.8)
$$I_{1\Psi} = \int \frac{1}{\varphi(J_0)}\,\frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)}\,\varphi(J_0)\,dJ_0 = \int \frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)}\,dJ_0$$

subject to the restrictions and

(5.9)
$$I_{2\Psi} = \int x\psi(x)\,dx = \int x\varphi(J_0)\,dJ_0 = 0,$$

(5.10)
$$I_{3\Psi} = \int x^2\psi(x)\,dx = \int x^2\varphi(J_0)\,dJ_0 = 1.$$

In the above form it is immediately obvious that if $\Psi = \Phi$, $x = J_0$ and hence $I_{1\Phi} = 1$. This form is also more suitable for our variational approach.

Suppose now that x is replaced by $x^* = cx$. Then I_1, I_2 and I_3 are replaced by $I_1^* = I_1/c$, $I_2^* = cI_2$, and $I_3^* = c^2I_3$. Thus if $I_2 = 0$ and $I_3 < 1$, we can obtain $I_2^* = 0$ and $I_3^* = 1$ with $I_1^* < I_1$. This discussion is relevant to the proof of the following lemma.

LEMMA 1. *The solution of the minimization problem is unique if it exists.*

PROOF. Suppose x_1 and x_2 are distinct functions with non-negative derivatives. Then let $x = (1 - w)x_1 + wx_2$, where $0 \leqq w \leqq 1$. Then, by convexity

$$I_1(w) = \int \frac{\varphi(J_0)}{\left(\frac{dx}{dJ_0}\right)} dJ_0 < (1 - w) \int \frac{\varphi(J_0)}{\left(\frac{dx_1}{dJ_0}\right)} dJ_0 + w \int \frac{\varphi(J_0)}{\left(\frac{dx_2}{dJ_0}\right)} dJ_0,$$

$$I_2(w) = \int x\varphi(J_0)\, dJ_0 = (1 - w) \int x_1\varphi(J_0)\, dJ_0 + w \int x_2\varphi(J_0)\, dJ_0,$$

and

$$I_3(w) = \int x^2\varphi(J_0)\, dJ_0 < (1 - w) \int x_1^2\varphi(J_0)\, dJ_0 + w \int x_2^2\varphi(J_0)\, dJ_0.$$

Hence x_1 and x_2 cannot both be solutions of the minimization problem since otherwise a multiple of $(x_1 + x_2)/2$ would then satisfy the side conditions and yield a smaller I_1.

With this lemma, all that remains is to show that $x = J_0$ is a solution of the problem. To this end we establish a sufficient condition for the solution of the problem as follows. Suppose that x_1 and x_2 are monotone functions satisfying the restrictions where x_2 gives a lower value for I_1 than does x_1. Then using the convexity again, we have

$$I_1'(0) = -\int \frac{\frac{d(x_2 - x_1)}{dJ_0}}{\left(\frac{dx_1}{dJ_0}\right)^2} \varphi(J_0)\, dJ_0 < 0,$$

$$I_2'(0) = \int (x_2 - x_1)\varphi(J_0)\, dJ_0 = 0,$$

and

$$I_3'(0) = 2 \int x_1(x_2 - x_1)\varphi(J_0)\, dJ_0 < 0.$$

Consequently we have

LEMMA 2. *If x_1 satisfies the restrictions and if for each x_2 which does so also there is a $\xi \geqq 0$ such that*

$$I_1'(0) + \xi I_3'(0) \geqq 0,$$

then x_1 is the unique solution of the minimization problem.[4]

[4] This sufficient condition is essentially the usual Euler equation except that with the convexity at our disposal and the monotonicity restriction, it plays the role of a sufficient instead of a necessary condition.

Now

$$I_1'(0) = \frac{-(x_2 - x_1)}{\left(\frac{dx_1}{dJ_0}\right)^2} \varphi(J_0) \Big|_{-\infty}^{\infty} + \int (x_2 - x_1) \left[\frac{\varphi'(J_0)}{\left(\frac{dx_1}{dJ_0}\right)^2} - \frac{\frac{2d^2x_1}{dJ_0^2}\varphi(J_0)}{\left(\frac{dx_1}{dJ_0}\right)^3} \right] dJ_0 .$$

Now let $x_1(J_0) = J_0$. Then

$$I_1'(0) + \xi I_3'(0) = \int (x_2 - x_1)[\varphi'(J_0) + 2\xi J_0 \varphi(J_0)] \, dJ_0 ,$$

which vanishes for $\xi = 1/2$. Applying Lemma 2 establishes our theorem.

If we regarded the c_1-test as one tailor made to compete against the best parametric test for translation when F and G are normal, we may inquire about nonparametric tests designed to compete against the best parametric tests when F and G have some other form.

Suppose F and G are known to be of the form $F_0(x - \theta)$ and $F_0(x - \varphi)$ respectively where F_0 has a twice differentiable density f_0. Then an efficient[5] test statistic for $\Delta = \theta - \varphi = 0$ would be the maximum-likelihood estimate

$$\hat{\Delta} = \hat{\theta} - \hat{\varphi}$$

for which the asymptotic distribution is normal with mean Δ and variance $[N\lambda(1 - \lambda)(\text{inf}_{F_0})]^{-1}$, where

$$\text{inf}_{F_0} = \int \frac{[f_0'(x)]^2}{f_0(x)} \, dx, \tag{5.11}$$

providing the above integral exists. The relative efficiency of our nonparametric test based on the test statistic T with a specified normalized[6] J to the $\hat{\Delta}$ test is

$$E_{T,\hat{\Delta}}(F_0) = \frac{I_{1F_0}^2}{\text{inf}_{F_0}}, \tag{5.5a}$$

where

$$I_{1F_0} = \int J'(F_0) f_0^2(x) \, dx. \tag{5.6a}$$

It can be shown that the best J in the sense that it maximizes $E_{T,\hat{\Delta}}(F_0)$ is given by

$$J(u) = \frac{-f_0'(x)}{f_0(x)} (\text{inf}_{F_0})^{-1/2} \tag{5.12}$$

---

[5] There seems to be no clear-cut statement in the literature which would establish the test based on $\hat{\Delta}$ as an efficient test invariant under the same translation of the X_i and Y_j. The authors wish to thank the referee who pointed out the following elegant proof. The efficacy of the $\hat{\theta} - \hat{\varphi}$ test is $\lambda(1 - \lambda) \text{ inf}_{F_0}$, where inf_{F_0} is the information of F_0. No invariant test of $\Delta = \Delta_N$ vs. $\Delta = 0$ can have greater efficacy than the likelihood ratio test for testing $\Delta = \Delta_N$ vs. $\Delta = 0$ when the densities of X and Y are $f_0(x + (1 - \lambda)\Delta)$ and $f_0(x - \lambda\Delta)$. A standard calculation gives this test efficacy $\lambda(1 - \lambda) \text{ inf}_{F0}$. Thus our $\hat{\theta} - \hat{\varphi}$ test is efficient.

[6] Let J be normalized so that $\int J(u) \, du = 0$ and $\int J^2(u) \, du = 1$.

where $u = F_0(x)$. In fact for this J, we have

$$I_{1F_0} = -(\inf_{F_0})^{-1/2} \int \left[\frac{f_0''(x)}{f_0(x)} - \frac{[f_0'(x)]^2}{f_0^2(x)}\right] \frac{1}{f_0(x)} f_0'(x)\, dx = (\inf_{F_0})^{1/2}$$

and

$$E_{T,\hat{\Delta}} = 1.$$

As it is to be expected, if $F_0 = \Phi(N(0, 1))$, the corresponding $J = J_0$, the inverse of Φ. The problem of comparing the nonparametric with the parametric procedures designed for F_0 when F and G are translates of $\Psi \neq F_0$ is hindered by our ignorance of the behavior of the *parametric* procedure when $\Psi \neq F_0$.

6. Orientation and applications.

6.A. *Orientation.* In Fraser's book [5] it is shown that the c_1-test has a limiting normal distribution for normal alternatives. We have now shown this to be the case for all alternatives (if we include the cases where $N\sigma_N^2 = 0$ or $N\sigma_N^2 \to 0$ as degenerate cases). Hoeffding's U-statistics include many nonparametric test statistics and he, Lehmann, and Dwass have shown that U-statistics are asymptotically normal under the alternative hypothesis. The U-statistics do not include all statistics of the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni}. \tag{3.1}$$

In particular c_1 is not a U-statistic. Dwass's results [3], summarized in Theorem 4, appear to be the only useful results for statistics of the form (3.1) under general alternative hypotheses.

THEOREM 4. *Suppose*

(1) *The conditions of the first paragraph of our Section* 3 *hold* (*Dwass has written to us indicating that it is sufficient to have m and n approach* ∞);

(2) *The polynomial*

$$P(t) = \sum_{k=1}^{h} b_k t^k$$

is non-degenerate, i.e.,

$$\max(|b_1|, \cdots, |b_h|) > 0;$$

(3) $(X_1, \cdots, X_m, Y_1, \cdots, Y_n) = (U_1, \cdots, U_N)$ *and* R_i *is the number of* U*'s less than or equal to* U_i,

$$a_{Ni} = \begin{cases} a_1 = (n/mN)^{1/2}, & i = 1, \cdots, m, \\ a_2 = -(m/nN)^{1/2}, & i = m+1, \cdots, N; \end{cases} \tag{4}$$

$$t_N = \sum_{i=1}^{N} a_{Ni} P(R_i/N); \tag{5}$$

then

$$\lim_{N\to\infty} P\left(\frac{t_N - E(t_N)}{\sigma_{t_N}} < s\right) = \int_{-\infty}^{s} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx.$$

First note

$$\begin{aligned} t_N &= \sum_{i=1}^{N} P\left(\frac{i}{N}\right)[a_1 z_{Ni} + a_2(1 - z_{Ni})] \\ &= \sum_{i=1}^{N} P\left(\frac{i}{N}\right) z_{Ni}\left(\frac{1}{\sqrt{N}}\left[\left(\frac{1-\lambda_N}{\lambda_N}\right)^{1/2} + \left(\frac{\lambda_N}{1-\lambda_N}\right)^{1/2}\right]\right) + a_2 \sum_{i=1}^{N} P\left(\frac{i}{N}\right) \\ &= \sqrt{N} T_N\left(\left(\frac{1-\lambda_N}{\lambda_N}\right)^{1/2} + \left(\frac{\lambda_N}{1-\lambda_N}\right)^{1/2}\right) + K, \end{aligned}$$

where in T_N we have $E_{Ni} = P(i/N)$. Thus there is a non-stochastic linear relationship between t_N and T_N. Hence, from the statistical viewpoint t_N is equivalent to T_N, a statistic of the form (3.1). Now let us compare Dwass's conditions with ours.

(1) Requiring λ_N to be bounded away from 0 and 1 seems to be essential in our Theorem 1.

(2) The condition $E_{Ni} = J_N(i/N) = P(i/N) = \sum_{k=1}^{h} b_k(i/N)^k$ is much stronger than our condition 4 in Theorem 1 in two respects: We only require that $J_N(x)$ have a limit and the limit need not be a polynomial in x. Of particular importance we do not require $J(x)$ to be bounded on $0 < x < 1$. The requirement $\max(|b_1|, \cdots, |b_h|) > 0$ is to insure that $E_{Ni} \equiv 0$, a trivial case which causes no difficulty.

6.B. *Applications.*

Example 1: Let $E_{Ni} = \sum_{j=i}^{N} j^{-1}$. Then Savage has proved [10] that T_N has a limiting Gaussian distribution under the hypothesis and is the test statistic for the locally most powerful rank test of $\theta_1 = \theta_2$ against the alternative $\theta_1 \neq \theta_2$ where $F(x) = e^{\theta_1 x}$ and $G(x) = e^{\theta_2 x}$, $-\infty < x \leqq 0$ and $F(x) = G(x) = 1$, $x > 0$. In order to verify that T_N has a limiting Gaussian distribution under the alternative hypothesis let us check the conditions of Theorem 1. To do so we note that $J_N(i/N)$ is the expected value of the ith smallest observation of a sample from the exponential distribution and that Theorem 2 is applicable. Hence T_N is asymptotically normal in all cases.

Example 2: Van der Waerden [12] has developed the theory of the test statistic

$$T_N = \int_{-\infty}^{\infty} J\left(\frac{NH_N(x)}{N+1}\right) dF_m(x),$$

where J is the inverse of the normal $N(0, 1)$ cumulative distribution. It can be shown that

$$\int_{-\infty}^{\infty} \left| J\left(\frac{NH_N(x)}{N+1}\right) - J(H_N(x)) \right| dH_N(x) = o\left(\frac{1}{\sqrt{N}}\right).$$

Then conditions 2 and 3 of Theorem 1 are established and the asymptotic normality and efficiency properties for this statistic are verified to be the same as those of the c_1-statistic.

7. Higher order terms. In proving that the C terms of Theorem 1 are uniformly of higher order the following elementary results are used repeatedly.

7.A. *Elementary results.*

1. $H \geqq \lambda_N F \geqq \lambda_0 F$.
2. $H \geqq (1 - \lambda_N)G \geqq \lambda_0 G$.
3. $1 - F \leqq \dfrac{1 - H}{\lambda_N} \leqq \dfrac{1 - H}{\lambda_0}$.
4. $1 - G \leqq \dfrac{1 - H}{1 - \lambda_N} \leqq \dfrac{1 - H}{\lambda_0}$.
5. $F(1 - F) \leqq \dfrac{H(1 - H)}{\lambda_N^2} \leqq \dfrac{H(1 - H)}{\lambda_0^2}$.
6. $G(1 - G) \leqq \dfrac{H(1 - H)}{\lambda_0^2}$.
7. $dH \geqq \lambda_N \, dF \geqq \lambda_0 \, dF$.
8. $dH \geqq (1 - \lambda_N) \, dG \geqq \lambda_0 \, dG$.
9. Let (a_N, b_N) be the interval $S_{N\epsilon}$, where

$$S_{N\epsilon} = \left\{x : H(1 - H) > \frac{\eta_\epsilon \lambda_0}{N}\right\}. \tag{7.1}$$

Then η_ϵ can be chosen independently of F, G and λ_N so that

$$P\{X_i \in S_{N\epsilon}, Y_j \in S_{N\epsilon}, i = 1, 2, \cdots, m, j = 1, 2, \cdots, n\} \geqq 1 - \epsilon. \tag{7.2}$$

10. $\int_{-\infty}^{\infty} J(H(x)) \, dF(x)$ is finite.

PROOF. Using assumption 4 of Theorem 1 and A.7

$$\begin{aligned} \left| \int_{-\infty}^{\infty} J(H(x)) \, dF(x) \right| &\leqq K \int_0^1 [H(1 - H)]^{-\frac{1}{2}+\delta} \, dH \\ &\leqq K \int_0^1 \frac{dH}{[H(1 - H)]^{1/2}} \leqq K. \end{aligned} \tag{7.3}$$

7.B. *Detailed consideration of the second order terms of Theorem* 1. We are now ready to show that the C terms are uniformly of higher order. We begin with C_{1N} and prove the following identity:

$$\begin{aligned} C_{1N} &= \lambda_N \int_{-\infty}^{\infty} (F_m - F)J'(H) \, d(F_m(x) - F(x)) \\ &= \frac{\lambda_N}{2}\left[\int J'(H) \, d(F_m - F)^2 + \frac{1}{m} \int J'(H) \, dF_m\right]. \end{aligned} \tag{7.4}$$

Let R be the set of points of increase of F_m. Then the right-hand side of the identity becomes

$$\frac{\lambda_N}{2}\left[\int_{\bar{R}} J'(H)\, d(F_m - F)^2 + \int_R J'(H)\, d(F_m - F)^2 + \frac{1}{m}\sum_{i=1}^m J'(H(X_i))\frac{1}{m}\right]$$

$$= \frac{\lambda_N}{2}\left[2\int_{\bar{R}} J'(H)(F_m - F)\, d(F_m - F) + \sum_{i=1}^m J'(H(X_i)) \cdot \left[\left(\frac{i}{m} - F(X_i)\right)^2 - \left(\frac{i-1}{m} - F(X_i)\right)^2\right] + \frac{1}{m}\sum_{i=1}^m J'(H(X_i))\frac{1}{m}\right]$$

$$= \frac{\lambda_N}{2}\left[2\int_{\bar{R}} J'(H)(F_m - F)\, d(F_m - F) + \sum_{i=1}^m J'(H(X_i)) \cdot \left[\frac{2}{m}\left[\frac{i}{m} - F(X_i)\right] - \frac{1}{m^2}\right] + \sum_{i=1}^m J'(H(X_i))\frac{1}{m^2}\right]$$

$$= \lambda_N \int (F_m - F) J'(H)\, d(F_m - F).$$

Using this identity we integrate by parts and obtain

$$C_{1N} = -\frac{\lambda_N}{2}(C_{11N} + C_{12N} - C_{13N}), \tag{7.5}$$

where

$$C_{11N} = \int_{S_{N\epsilon}} (F_m - F)^2 J''(H)\, dH,$$

$$C_{12N} = \int_{\bar{S}_{N\epsilon}} (F_m - F)^2 J''(H)\, dH,$$

$$C_{13N} = \frac{1}{m}\int J'(H(x))\, dF_m(x) = \frac{1}{m^2}\sum_{i=1}^m J'(H(X_i)),$$

where $S_{N\epsilon}$ was defined in 7.A.9.

Now let us consider the random variable C_{11N}. We find

$$\mathcal{E} \mid C_{11N} \mid = \mathcal{E}\left\{\int_{S_{N\epsilon}} (F_m - F)^2 \mid J''(H) \mid dH\right\} = \int_{S_{N\epsilon}} \frac{F(1-F)}{N\lambda_N} \mid J''(H) \mid dH.$$

Now using assumption 4 of Theorem 1 and 7.A.5 we obtain

$$\mathcal{E} \mid C_{11N} \mid \leqq \frac{K}{N}\int_{S_{N\epsilon}} \frac{H(1-H)\, dH}{[H(1-H)]^{\frac{5}{2}-\delta}} \leqq \frac{K}{N}\int_{\frac{1}{KN}}^1 \frac{1}{H^{\frac{3}{2}-\delta}}\, dH \leqq \frac{K}{N^{\frac{1}{2}+\delta}}.$$

Now using the Markoff inequality ([2], p. 182),

$$\Pr\left(|C_{11N}| > aN^{-1/2}\right) \leqq \frac{K}{N^{\frac{1}{2}+\delta}} \frac{N^{1/2}}{a} = \frac{K}{aN^{\delta}},$$

where K may depend on ϵ. Now consider C_{12N}.

Let $H_1 = H(a_N)$, $H_2 = H(b_N)$ as in 7.A.9. Then $H_1 = 1 - H_2 < K/N$. With probability greater than $1 - \epsilon$ we have

$$C_{12N} = \int_{S_{N\epsilon}} (F_m - F)^2 J''(H)\, dH = \int_0^{H_1} F^2 J''(H)\, dH + \int_{H_2}^1 (1 - F)^2 J''(H)\, dH$$

$$|C_{12N}| \leqq K\left[\int_0^{H_1} \frac{H^2\, dH}{(H(1 - H))^{\frac{5}{2}-\delta}} + \int_{H_2}^1 \frac{(1 - H)^2\, dH}{(H(1 - H))^{\frac{5}{2}-\delta}}\right]$$

$$\leqq K \int_0^{H_1} H^{-\frac{1}{2}+\delta}\, dH \leqq KN^{-\frac{1}{2}-\delta}.$$

Hence $C_{11N} + C_{12N}$ which does not involve ϵ is $o_p(N^{-\frac{1}{2}})$. Now to complete the study of C_{1N} we investigate C_{13N}:

$$|C_{13N}| = \frac{1}{m^2}\left|\sum_{i=1}^m J'[H(X_i)]\right| \leqq \frac{K}{m^2}\sum_{i=1}^m [H(X_i)(1 - H(X_i))]^{-\frac{3}{2}+\delta}.$$

We may assume $\delta < \frac{3}{2}$ or $\delta < \frac{1}{2}$ without loss of generality. Then using 7.A.5

$$|C_{13N}| \leqq \frac{K}{N}\frac{1}{m}\sum_{i=1}^m [F(X_i)[1 - F(X_i)]]^{-\frac{3}{2}+\delta},$$

which is distribution free. By a theorem of Marcinkiewicz ([8], pp. 242–243) if a random variable Y has rth order moment finite $(0 < r < 1)$, then the sum of N independent observations on Y is $o_p(N^{1/r})$. If X has cdf F,

$$[F(X)[1 - F(X)]]^{-\frac{3}{2}+\delta}$$

has a finite moment of order $2/(3 - \delta)$ and hence

$$C_{13N} = o_p\left[\frac{1}{m^2} N^{\frac{3}{2}-\frac{\delta}{2}}\right] = o_p[N^{-\frac{1}{2}}].$$

Consequently $C_{1N} = o_p(N^{-\frac{1}{2}})$.

Next consider

$$C_{2N} = (1 - \lambda_N)\int_{-\infty}^{\infty} (G_n - G)J'(H)\, d[F_m(x) - F(x)]. \tag{7.6}$$

We have

$$C_{2N} = (1 - \lambda_N)(C_{21N} + C_{22N})$$

where

$$C_{21N} = \int_{S_{N\epsilon}} (G_n - G)J'(H)\, d[F_m(x) - F(x)],$$

$$C_{22N} = \int_{S_{N\epsilon}} (G_n - G)J'(H)\, d[F_m(x) - F(x)].$$

With probability greater than $1 - \epsilon$, there are no observations in $\bar{S}_{N\epsilon}$ and

$$|C_{21N}| \leqq K \int_{\bar{S}_{N\epsilon}} H(1 - H)[H(1 - H)]^{-\frac{3}{2}+\delta} \, dH(x) \leqq K \left(\frac{\eta_\epsilon}{N}\right)^{\frac{1}{2}+\delta}.$$

Since the two samples are independent and $\mathcal{E}(G_n - G) = 0$, we have

$$\mathcal{E}(C_{22N}) = \mathcal{E}\{\mathcal{E}C_{22N} \mid X_1, X_2, \cdots, X_m\} = 0,$$

$$\mathcal{E}(C_{22N}^2 \mid X_1, X_2, \cdots, X_m) = C_{23N} + C_{24N},$$

$$C_{23N} = \frac{2}{n} \iint_{\substack{x,y \in S_{N\epsilon} \\ x<y}} G(x)[1 - G(y)]J'[H(x)]J'[H(y)] \cdot d[F_m(x) - F(x)] \, d[F_m(y) - F(y)],$$

$$C_{24N} = \frac{1}{nm} \int_{S_{N\epsilon}} G(x)[1 - G(x)]\{J'[H(x)]\}^2 \, dF_m(x),$$

$$\mathcal{E}(C_{23N}) = \frac{-2}{nm} \iint_{\substack{x,y \in S_{N\epsilon} \\ x<y}} G(x)[1 - G(y)]J'[H(x)]J'[H(y)] \, dF(x) \, dF(y)^{7}$$

$$\leqq \frac{K}{N^2} \iint_{x<y} H(x)[1 - H(y)] \mid J'[H(x)]J'[H(y)] \mid dH(x) \, dH(y)$$

$$\leqq \frac{K}{N^2} \iint_{0<x<y<1} x^{-\frac{1}{2}+\delta}(1 - x)^{-\frac{3}{2}+\delta} y^{-\frac{3}{2}+\delta}(1 - y)^{-\frac{1}{2}+\delta} \, dx \, dy \leqq \frac{K}{N^2},$$

$$\mathcal{E}(C_{24N}) = \frac{1}{nm} \int_{S_{N\epsilon}} G(1 - G)(J'[H])^2 \, dF(x) \leqq \frac{K}{N^2} \int_{S_{N\epsilon}} [H(1 - H)]^{-2+2\delta} \, dH(x)$$

$$\leqq \frac{K\eta_\epsilon^{-1+2\delta}}{N^{1+2\delta}} = o(N^{-1}).$$

Hence

$$\mathcal{E}(C_{22N}^2 \mid X_1, X_2, \cdots, X_m) \leqq K o_p(N^{-1}),$$

where K may depend on ϵ and

$$|C_{22N}| \leqq K o_p(N^{-1/2})$$

since

$$P(C_{22N}^2 > a\mathcal{E}(C_{2N}^2 \mid X_1, \cdots, X_m)) < 1/a.$$

Consequently $C_{2N} = (1 - \lambda_N)(C_{21N} + C_{22N})$ which does not involve ϵ, satisfies

$$C_{2N} = o_p(N^{-1/2}).$$

---

[7] This integrand has already appeared as part of the variance in Eq. (4.3).

Now consider

$$C_{3N} = \int_{0<H_N(x)<1} [H_N(x) - H(x)]^2 J''[\varphi H_N(x) + (1-\varphi)H(x)]\, dF_m(x), \tag{7.7}$$
$$0 < \varphi < 1.$$

With probability greater than $1 - \epsilon$, the range of integration $0 < H_N(x) < 1$ can be replaced by $S_{N,\epsilon}$ without changing C_{3N}. Since

$$\sup_{H_N>0} \left| \frac{H(x)}{H_N(x)} \right| = O_p(1), \tag{7.8a}$$

and

$$\sup_{H_N<1} \left| \frac{1 - H(x)}{1 - H_N(x)} \right| = O_p(1), \tag{7.8b}$$

for each $\epsilon > 0$, there is an $\eta_\epsilon^* > 0$ such that with probability greater than $1 - \epsilon$, we have for $0 < H_N(x) < 1$,

$$[\varphi H_N + (1-\varphi)H][1 - (\varphi H_N + (1-\varphi)H)] > \eta_\epsilon^* H(x)[1 - H(x)]. \tag{7.9}$$

Then

$$| C_{3N} | \leqq \int_{S_{N,\epsilon}} [H_N(x) - H(x)]^2 (\eta_\epsilon^*)^{-\frac{5}{2}+\delta} \{H[1-H]\}^{-\frac{5}{2}+\delta}\, dF_m(x) = (\eta_\epsilon^*)^{-\frac{5}{2}+\delta} C_{31N},$$

$$\begin{aligned}\mathcal{E}(| C_{31N} |) \leqq \frac{1}{N} \int_{S_{N,\epsilon}} \Bigg[\lambda_N F(1-F) + \frac{(1-F)(1-2F)}{N} \\ + (1-\lambda_N)G(1-G)\Bigg][H(1-H)]^{-\frac{5}{2}+\delta}\, dF(x)\end{aligned}$$

$$\leqq \frac{K}{N} \int_{S_{N,\epsilon}} [H(1-H)]^{-\frac{3}{2}+\delta}\, dH + \frac{K}{N^2} \int_{S_{N,\epsilon}} [H(1-H)]^{-\frac{5}{2}+\delta}\, dF$$

$$\leqq \frac{K\eta_\epsilon^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}} + \frac{K\eta_\epsilon^{-\frac{3}{2}+\delta}}{N^{\frac{1}{2}+\delta}}.$$

Consequently

$$C_{3N} = o_p(N^{-1/2}).$$

The C_{4N} term vanishes unless the greatest of the $N = m + n$ observations is an X. In that case

$$C_{4N} = \frac{1}{m}\{ -J[H(X_m)] - [1 - H(X_m)]J'[H(X_m)]\}. \tag{7.10}$$

Using 7.A.9, however,

$$\frac{1}{m} | J[H(X_m)] | \leqq \frac{[H(X_m)[1 - H(X_m)]]^{-\frac{1}{2}+\delta}}{m} \leqq \frac{(\eta_\epsilon^*)^{-\frac{1}{2}+\delta}}{N^{\frac{1}{2}+\delta}}$$

with probability at least $1 - \epsilon$. Hence

$$\frac{1}{m} J[H(X_m)] = o_p(N^{-1/2}).$$

Similarly

$$\left| \frac{[1 - H(X_m)]J'[H(X_m)]}{m} \right| \leqq \frac{[1 - H(X_m)]}{m} \{H(X_m)[1 - H(X_m)]\}^{-\frac{3}{2}+\delta}$$

$$\leqq \frac{\{[H(X_m)][1 - H(X_m)]\}^{-\frac{1}{2}+\delta}}{mH(X_m)} = o_p(N^{-1/2})O_p(1) = o_p(N^{-1/2}).$$

Hence

$$C_{4N} = o_p(N^{-1/2}).$$

The neglibility of C_{5N} and C_{6N} follows immediately from Assumptions 2 and 3 of Theorem 1.

7.C. *Proof of Theorem* 2. First we note that

$$J_N\left(\frac{i}{N}\right) = E_{N,i} = \int_0^1 J(u) g_{i,N}(u)\, du, \tag{7.11}$$

where

$$g_{i,N}(u) = \frac{N!}{(i-1)!\,(N-i)!} u^{i-1}(1-u)^{(N-i)} \tag{7.12}$$

is the density of the ith order statistic from the uniform distribution on [0, 1] and incidentally has mean $i/(N + 1)$ and variance $i(N - i + 1)/[(N + 1)^2 (N + 2)]$. Then we have

$$\begin{aligned} | E_{N,1} | &\leqq KN \int_0^1 [u(1-u)]^{-\frac{3}{2}+\delta}(1-u)^{N-1}\, du \\ &= \frac{KN\Gamma(N - \frac{1}{2} + \delta)\Gamma(\frac{1}{2} + \delta)}{\Gamma(N + 2\delta)} \leqq KN^{\frac{1}{2}-\delta}. \end{aligned} \tag{7.13}$$

By a symmetric argument the desired result $J_N(1) = o(N^{1/2})$ follows. Furthermore we have

$$\left| J_N\left(\frac{1}{N}\right) - J\left(\frac{1}{N}\right) \right| \leqq KN^{\frac{1}{2}-\delta} + K\left[\frac{1}{N}\left(1 - \frac{1}{N}\right)\right]^{-\frac{1}{2}+\delta} \leqq KN^{\frac{1}{2}-\delta}. \tag{7.14}$$

Before proceeding to bound $J_N(i/N) - J(i/N)$ for $1 < i \leqq N/2$ we apply the Stirling formula

$$\begin{aligned} \log x! &= \log \Gamma(x + 1) \\ &= \frac{1}{2} \log 2\pi - x + \left(x + \frac{1}{2}\right) \log x + \frac{\theta}{12x}, \qquad 0 < \theta < 1, \end{aligned} \tag{7.15}$$

with a rather standard argument to obtain for $1 < i \leqq N/2, 0 < u \leqq (i-1)/(N-1)$,

$$g_{i,N}(u) \leqq \sqrt{\frac{(N-1)^3}{2\pi(i-1)(N-i)}}\, e^{-\frac{v^2}{2}\left[\frac{(N-1)}{(i-1)(N-i)}\right]} \left[1 + \frac{K}{N}\right], \tag{7.16}$$

(7.16)[8]

where

$$v = (N-1)u - (i-1). \tag{7.17}$$

For $1 < i \leqq N/2$,

$$\begin{aligned} J_N\left(\frac{i}{N}\right) - J\left(\frac{i}{N}\right) &= \int_0^1 \left[J(u) - J\left(\frac{i}{N}\right)\right] g_{i,N}(u)\, du \\ &= D_{11} + D_{12} + D_{21} + D_{22} + D_3 + D_4, \end{aligned} \tag{7.18}$$

where

$$D_{11} = \int_0^{u_1} J(u) g_{i,N}(u)\, du, \qquad D_{12} = \int_{1-u_1}^1 J(u) g_{i,N}(u)\, du,$$

$$D_{21} = -\int_0^{u_1} J\left(\frac{i}{N}\right) g_{i,N}(u)\, du, \qquad D_{22} = -\int_{1-u_1}^1 J\left(\frac{i}{N}\right) g_{i,N}(u)\, du,$$

$$D_3 = \int_{u_1}^{1-u_1} \left(u - \frac{i}{N}\right) J'\left(\frac{i}{N}\right) g_{i,N}(u)\, du$$

$$D_4 = \frac{1}{2}\int_{u_1}^{1-u_1} \left(u - \frac{i}{N}\right)^2 J''(u^*) g_{i,N}(u)\, du,$$

u^* between u and i/N, and $u_1 = (i-1)/[2(N-1)]$.

$$\begin{aligned} g_{i,N}(u) &= u^\alpha \frac{u^{i-1-\alpha}(1-u)^{N-i}(N-\alpha)!}{(i-1-\alpha)!\,(N-i)!} \frac{N!}{(N-\alpha)!} \frac{(i-1-\alpha)!}{(i-1)!} \\ &\leqq K u^\alpha N^\alpha g_{i-\alpha,N-\alpha}(u), \end{aligned} \tag{7.19}$$

where $\alpha = \frac{1}{2} - \delta$ and we assume $\delta < \frac{1}{2}$ and thus $\alpha > 0$ without loss of generality. Let Φ be the normal cdf. Then

$$|D_{11}| \leqq \int_0^{u_1} K[u(1-u)]^{-\alpha} K u^\alpha N^\alpha g_{i-\alpha,N-\alpha}(u)\, du \leqq KN^\alpha \cdot \Phi\left[\frac{\left(u_1 - \frac{i-1-\alpha}{N-1-\alpha}\right)(N-1-\alpha)^{3/2}}{\sqrt{(i-1-\alpha)(N-i)}}\right],$$

$$|D_{11}| \leqq KN^\alpha \Phi\left(\frac{-\sqrt{i}}{K}\right). \tag{7.20}$$

---

[8] K represents a generic constant independent of i, N, λ_N, F, and G. This equation is related to the asymptotic normality of order statistics and is derived by an operation similar to the direct proof of the asymptotic normality of the binomial distribution.

Since $g_{i,N}(u) \geqq g_{i,N}(1 - u)$ for $1 < i \leqq N/2$ and $0 \leqq u \leqq 1/2$, $| D_{12} |$ has the same bound as $| D_{11} |$. Similarly

$$(7.21) \qquad | D_{21} | \leqq K \left(\frac{i}{N}\right)^{-\alpha} \Phi \left[\frac{\left(u_1 - \frac{i-1}{N-1}\right)(N-1)^{3/2}}{\sqrt{(i-1)(N-i)}} \right] \leqq KN^{\alpha} \Phi \left(\frac{-\sqrt{i}}{K}\right)$$

and $| D_{22} |$ has the same bound too. Since the expectation of the ith order statistic from the uniform distribution is $i/(N + 1)$,

$$D_3 = -J'\left(\frac{i}{N}\right)\left\{\int_0^{u_1} \left(u - \frac{i}{N}\right) g_{i,N}(u)\, du + \int_{1-u_1}^{1} \left(u - \frac{i}{N}\right) g_{i,N}(u)\, du + \frac{i}{N(N+1)}\right\}$$

Now

$$h(u) = \left| u - \frac{i}{N} \right| g_{i,N}(u) \leqq Kh(1 - u) \qquad \text{for } u < u_1.$$

Hence

$$(7.22) \qquad \begin{aligned} | D_3 | &\leqq K \left(\frac{i}{N}\right)^{-\alpha-1} \left[K \frac{i}{N} \Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{i}{N(N+1)} \right] \\ &\leqq KN^{\alpha}\Phi\left(\frac{-\sqrt{i}}{K}\right) + KN^{\alpha-1} \end{aligned}$$

Finally

$$| D_4 | \leqq Ku_1^{-\frac{5}{2}+\delta} \int_0^1 \left(u - \frac{i}{N}\right)^2 g_{i,N}(u)\, du,$$

$$| D_4 | \leqq Ku_1^{-\frac{5}{2}+\delta} \left[\frac{i(N - i + 1)}{(N+1)^2(N+2)} + \left(\frac{i}{N+1} - \frac{i}{N}\right)^2 \right],$$

$$(7.23) \qquad | D_4 | \leqq Ku_1^{-\frac{5}{2}+\delta} \left[K \frac{u_1}{N} + K \frac{u_1^2}{N^2} \right] \leqq \frac{Ku_1^{-\frac{3}{2}+\delta}}{N} \leqq \frac{KN^{\alpha}}{i^{1+\alpha}}.$$

Thus, for $1 < i \leqq N/2$,

$$(7.24) \qquad J_N\left[\frac{i}{N}\right] - J\left[\frac{i}{N}\right] \leqq KN^{\alpha}\left[\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{1}{N} + \frac{1}{i^{1+\alpha}}\right]$$

and

$$(7.25) \qquad \begin{aligned} &\left| \int_{1 \leqq NF_m \leqq N/2} [J_N(H_N) - J(H_N)]\, dF_m \right| \\ &\quad \leqq \frac{1}{m}\left\{ KN^{\frac{1}{2}-\delta} + \sum_{i=2}^{N/2} KN^{\alpha}\left[\Phi\left(\frac{-\sqrt{i}}{K}\right) + \frac{1}{N} + \frac{1}{i^{1+\alpha}}\right]\right\} \\ &\quad \leqq KN^{-\frac{1}{2}-\delta} \end{aligned}$$

since $\sum_{i=1}^{\infty} \Phi(-\sqrt{i}/K)$ and $\sum_{i=1}^{\infty} i^{-(1+\alpha)}$ converge. By a symmetric argument we can cover the range $N/2 < NF_m \leqq N$ and our theorem follows.

BIBLIOGRAPHY

[1] J. R. Blum and Lionel Weiss, "Consistency of certain two-sample tests," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 242–246.

[2] Harald Cramér, *Mathematical Methods of Statistics*, Princeton, 1946.

[3] Meyer Dwass, "The large sample powers of rank order tests in the two sample problem," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 352–374.

[4] C. G. Esseen, "Fourier analysis of distribution functions," *Acta Math.*, Vol. 77 (1945), pp. 1–125.

[5] D. A. S. Fraser, *Nonparametric Methods of Statistics*, John Wiley and Sons, Inc., 1956.

[6] J. E. Freund and A. R. Ansari, "Two-way rank sum test for variances," Virginia Polytechnic Institute report supported by the Office of Ordnance Research and National Science Foundation, 1957.

[7] J. L. Hodges, Jr., and Erich L. Lehmann, "The efficiency of some nonparametric competitors of the t-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.

[8] Michel Loève, *Probability Theory*, D. Van Nostrand Co., Inc., 1955.

[9] R. v. Mises, "On the asymptotic distribution of differentiable statistical functions," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 309–348.

[10] I. R. Savage, "Contributions to the theory of rank order statistics: Two sample case," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 590–616.

[11] M. E. Terry, "Some rank order tests which are most powerful against specific parametric alternatives," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 346–366.

[12] B. L. Van der Waerden, "The Computation of the X-distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, University of California Press, 1956, pp. 207–208.

[13] Abraham Wald and Jacob Wolfowitz, "Statistical tests based on permutations of the observations," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 358–372.