ASYMPTOTIC NORMALITY AND EFFICIENCY OF CERTAIN NONPARAMETRIC TEST STATISTICS

BY
HERMAN CHERNOFF
I. RICHARD SAVAGE

TECHNICAL REPORT NO. 15

PREPARED UNDER CONTRACT Nonr-225 (21)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

SEPTEMBER 13, 1957
ASYMPTOTIC NORMALITY AND EFFICIENCY OF CERTAIN
NONPARAMETRIC TEST STATISTICS

by

Herman Chernoff and I. Richard Savage

TECHNICAL REPORT NO. 15

PREPARED UNDER CONTRACT Nonr-225(21)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

REPRODUCTION IN WHOLE OR IN PART IS PERMITTED FOR
 ANY PURPOSE OF THE UNITED STATES GOVERNMENT

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
September 13, 1957
ASYMPTOTIC NORMALITY AND EFFICIENCY OF CERTAIN
NONPARAMETRIC TEST STATISTICS

by
Herman Chernoff and I. Richard Savage

1. Summary

Let \( X_1', \ldots, X_m \) and \( Y_1', \ldots, Y_n \) be ordered observations from the continuous cumulative distribution functions \( F(x) \) and \( G(x) \) respectively. If \( z_{N_1} = 1 \) when the \( i \)-th smallest of \( N = m+n \) observations is an \( X \) and \( z_{N_1} = 0 \) otherwise, then many nonparametric test statistics are of the form

\[
\frac{mT_N}{N} = \sum_{i=1}^{N} E_{N_1} z_{N_1}.
\]

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 [2] 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 \[8\]. Many papers have appeared giving the asymptotic efficiency of particular tests. Hodges and Lehmann \[3\] 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.

Our basic result on the asymptotic normality of statistics of the form $T_N$ is Theorem I of Section 4. This theorem is a generalization of Dwass's \[2\] Theorem 7.2. Theorem I 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 I to Dwass's Theorem 7.2. Applications of Theorem I to several nonparametric tests are given in Section 6.
3. **Assumptions and notation**

Assume that $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ are $N = m + n$ mutually independent random variables. The $X$'s have a common continuous cumulative distribution function $F(x)$ and the $Y$'s have a common continuous distribution function $G(x)$. Let $\lambda_N = m/N$ and assume that for all $N$ the inequalities

\[ 0 < \lambda_o \leq \lambda_N \leq 1 - \lambda_o < 1 \]

hold for some fixed $\lambda_o \leq \frac{1}{2}$.

Let $F_m(x) = (\text{number of } X_i \leq x)/m$ and $G_n(x) = (\text{number of } Y_i \leq y)/n$. 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 $N$ through $\lambda_N$ 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 I 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 arbitrarily with the sample size $N$.

If the $i$-th 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

\[(3.1) \quad mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni}\]

where the $E_{Ni}$ are given numbers. (The special case where $E_{Ni} = E(1/N)$ is particularly easily handled by our methods. For the Wilcoxon test this condition is met with $E_{Ni} = 1/N$.) The definition (3.1) of $T_N$ is the one con-
ventionally used. We shall, however, use the following representation:

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

The definitions (3.1) and (3.2) are seen to be equivalent when \( E_{Ni} = J_N(i/N) \).

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 \( \lambda_N \) in the interval \( 0 < \lambda_0 \leq \lambda_N \leq 1 - \lambda_0 < 1 \).

While \( J_N \) need be defined only at \( \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N} \), 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 \((\frac{1}{N}, \frac{i+1}{N}]\).

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 I.** If

1. \( J(H) = \lim_{N \to \infty} J_N(H) \) exists for all \( 0 < H < 1 \),

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. \( |J^{(1)}(H)| = \left| \frac{d^1 J}{dH} \right| \leq K[H(1-H)]^{-1-\frac{1}{2}+\delta} \) for \( i = 0, 1, 2 \) and for some \( \delta > 0 \).
then

\begin{equation}
\lim_{N \to \infty} P \left( \frac{T_N - \mu_N}{\sigma_N} \leq t \right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt
\end{equation}

uniformly with respect to F, G and \( \lambda_N \), where

\begin{equation}
\mu_N = \int_{-\infty}^{\infty} J(H(x)) \, dF(x)
\end{equation}

and

\begin{equation}
N\sigma_N^2 = 2(1-\lambda_N) \left\{ \int_{-\infty}^{\infty} G(x)(1-G(y))J'(H(x))J'(H(y)) \, dF(x) \, dF(y)
+ \frac{(1-\lambda_N)}{\lambda_N} \int_{-\infty}^{\infty} F(x)(1-F(y))J'(H(x))J'(H(y)) \, dG(x) \, dG(y) \right\}.
\end{equation}

Assumption 1 is likely to be filled whenever one speaks of a sequence of tests. In the special case \( E_{N_n} = E(1/N) \) of course \( J_N = E = J \) and assumption 2 will automatically be satisfied. Theorem II shows that assumptions 1, 2 and 3 are often satisfied when the \( E_{N_n} \) 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 \( E_{N_n} \) 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

\[
T_N = \int_{-\infty}^{\infty} J_N(H_N) \, dF_m(x) = \int_{0}^{1} \left[ J_N(H_N) - J(H_N) \right] \, dF_m(x)
+ \int_{0}^{1} J(H_N) \, dF_m(x)
\]
In the second integral we write \( dF_m = d(F_m - F) \), \( J(H_N) = J(H) + (H_N - H) J'(H) + \frac{(H_N - H)^2}{2} \frac{J''(\phi H_N + (1 - \phi) H)}{2} \) where \( 0 < \phi < 1 \), and \( H = \lambda^N_N F + (1 - \lambda^N_N) G \). After multiplying out the expression becomes

\[
T_N = A + B_{1N} + B_{2N} + \sum_{i=1}^{6} C_{1N}
\]

where

\[
A = \int_{0}^{1} J(H) dF(x)
\]

\[
B_{1N} = \int_{0}^{1} J(H) d(F_m(x) - F(x))
\]

\[
B_{2N} = \int_{0}^{1} (H_N - H) J'(H) dF(x)
\]

\[
C_{1N} = \lambda^N_N \int_{0}^{1} (F_m - F) J'(H) d(F_m(x) - F(x))
\]

\[
C_{2N} = (1 - \lambda^N_N) \int_{0}^{1} (G_n - G) J'(H) d(F_m(x) - F(x))
\]

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

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

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

\[
C_{6N} = \int_{H_N = 1} J_N(H_N) dF_m(x)
\]
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, $A = \int_0^1 J(H) dF(x)$, is finite as a result of assumption 4 of Theorem I, see Section 7.A.10. Since A depends on $\lambda_N$ as well as $F(x)$ and $G(x)$ it need not converge as $N \to \infty$, but it does remain bounded.

To begin with the first order term first integrate by parts

$$B_{2N} = \int_0^1 (H_N - H) J'(H) dF(x)$$

$$= - \int B(x) d[H_N(x) - H(x)]$$

where

$$B(x) = \int_{x_0}^x J'(H(y)) dF(y)$$

(4.13)

with $x_0$ determined rather arbitrarily by $H(x_0) = \frac{1}{2}$.

Now then we may write

$$B_{1N} + B_{2N} = \int_{-\infty}^{\infty} J(H) d[F_m(x) - F(x)] - \int_{-\infty}^{\infty} B(x) d[\lambda_N [F_m(x) - F(x)] + (1 - \lambda_N) [G_n(x) - G(x)]]$$

$$= \int_{-\infty}^{\infty} [J(H(x)) - \lambda_N B(x)] d(F_m(x) - F(x)) - \int_{-\infty}^{\infty} (1 - \lambda_N) B(x) d(G_n(x) - G(x)).$$

Hence
(4.15)  \[ B_{1N} + B_{2N} = \sum_{i=1}^{m} \frac{J(H(X_i)) - \lambda_N B(X_i)}{m} - \mathcal{E}[J(H(X)) - \lambda_N B(X)] - (1 - \lambda_N) \sum_{i=1}^{n} \frac{B(Y_i)}{n} - \mathcal{E}B(Y) \]

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, provided the first two moments of these random variables exist, we may apply the central limit theorem to show that when properly normalized each sum has a Gaussian distribution in the limit and that the sum of the two summations is Gaussian in the limit.

Now let us turn our attention to the moments. First

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

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

\[ \mathcal{E}\left\{ B(Y) \right\}^{2+\delta'} \leq K \int_{-\infty}^{\infty} \left[ H(x)(1-H(x)) \right]^{-\frac{1}{2} + \delta}(2+\delta') dG(x) , \]

\[ \leq K \int_{\lambda_0}^{1} \left[ H(1-H) \right]^{-\frac{1}{2} + \delta}(2+\delta') dH \leq K , \]

having made use of \( dG \leq \frac{1}{\lambda_0} dH \) (see Section 7.A.8).

Since

\[ |J(H(x)) - \lambda_N B(x)| \leq K \left[ H(x)(1-H(x)) \right]^{-\frac{1}{2} + \delta} \]

the existence of the \( 2+\delta' \) absolute moments of both terms in equation (4.15) follows. The variance may be computed directly in terms of \( \int B(x) dF(x) \), \( \int B^2(x) dF(x) \), etc., but we shall use a slightly different approach. First
considering the last part of 4.14,

\[- \int_{-\infty}^{\infty} (1-\lambda_N) B(x) d[G_n(x) - G(x)] = (1-\lambda_N) \int_{-\infty}^{\infty} [G_n(x) - G(x)] J'[H(x)] dF(x)\]

which evidently has mean 0 and variance

\[
\mathbb{E}\left\{ (1-\lambda_N) \int_{-\infty}^{\infty} [G_n(x) - G(x)] J'[H(x)] dF(x) \right\}^2 = 
\mathbb{E}\left\{ \frac{(1-\lambda_N)^2}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [G_1(x) - G(x)] [G_1(y) - G(y)] J'[H(x)] J'[H(y)] dF(x) dF(y) \right\}
\]

\[
= \frac{2(1-\lambda_N)^2}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(x) [1-G(y)] J'[H(x)] J'[H(y)] dF(x) dF(y),
\]

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 \( x < y, \)

\[
\mathbb{E}\left\{ |G_1(x) - G(x)| |G_1(y) - G(y)| \right\} \leq K G(x)[1-G(y)]
\]

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.)

By a similar argument, the variance of

\[
\int_{-\infty}^{\infty} [J(H(x)) - \lambda_N B(x)] d[F_m(x) - F(x)] = -(1-\lambda_N) \int_{-\infty}^{\infty} [F_m(x) - F(x)] J'[H(x)] dG(x)
\]

is given by

\[
\frac{2(1-\lambda_N)^2}{m} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) [1-F(y)] J'[H(x)] J'[H(y)] dG(x) dG(y).
\]
These two variances when added 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 absolute (2+$\delta'$) moments. If in fact they had finite absolute third moments, as they would if $\delta$ were greater than 1/6, we would be able to apply the Berry-Esseen Theorem [5,p.288] to get a uniform bound on the error of the normal approximation to the distribution of $B_{1N} + B_{2N}$, and our proof would be complete. If however $\delta$ does not exceed 1/6 we must apply Parzen's generalization of the Berry-Esseen Theorem to complete the proof.\footnote{Parzen [6] has proved but not yet published that if $X_1, X_2, \ldots, X_n$ are independent observations from a population with finite absolute 2+$\delta'$ moment where $\delta' > 0$, then $|F_n - F*| < f(n,\delta')$ where $F_n$ is the cdf of $\bar{X}$, $F*$ is the approximating normal distribution, $f(n,\delta') \to 0$ as $n \to \infty$ and finally $f(n,\delta')$ does not involve the distribution of $X$. The Berry-Esseen Theorem gives this result for $\delta' = 1$.}

In considering alternatives "near" the null hypothesis, as will be done in Section 5, the following corollary will be useful.

**COROLLARY I.** If

$$F(x) = \overline{v}(x - \theta_N)$$

$$G(x) = \overline{v}(x - \phi_N)$$

and

$$\lim_{N \to \infty} \Phi_N - \theta_N = 0$$

\footnote{Parzen [6] has proved but not yet published that if $X_1, X_2, \ldots, X_n$ are independent observations from a population with finite absolute 2+$\delta'$ moment where $\delta' > 0$, then $|F_n - F*| < f(n,\delta')$ where $F_n$ is the cdf of $\bar{X}$, $F*$ is the approximating normal distribution, $f(n,\delta') \to 0$ as $n \to \infty$ and finally $f(n,\delta')$ does not involve the distribution of $X$. The Berry-Esseen Theorem gives this result for $\delta' = 1$.}
then the limit of (4.3) is given by

\[
(4.16) \quad \lim_{N \to \infty} \left[ \frac{\lambda N}{(1-\lambda N)^2} \right]^2 c_N^2 = 2 \int \int x(1-y)J'(x)J'(y) \, dx \, dy .
\]

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

**THEOREM II.** If $J_{\frac{i}{N}}(\lambda)\,_{\tilde{N}}$ is the expectation of the $i$-th order statistic of a sample of size $N$ from a population whose cumulative distribution function is the inverse function of $J$ and

\[
|J(i)(u)| \leq K[u(1-u)]^{-i-\frac{1}{2} + \delta}
\]

then

\[
\lim_{N \to \infty} J_{\frac{i}{N}}(H) = J(H) , \quad 0 < H < 1 ,
\]

\[
J_{\frac{i}{N}}(1) = o(N^{1/2} )
\]

and

\[
\int \left[ J_{\frac{i}{N}}(H) - J(H) \right] \, dF_m(x) = o(N^{-1/2} ) .
\]

(We write $o$ instead of $o_p$ because the random sequence is bounded by a non-random sequence which is $o(N^{-1/2})$. In fact, $\int |J_{\frac{i}{N}}(H) - J(H)| \, dF_m(x) \leq \frac{1}{N} \int |J(H) - J_{\frac{i}{N}}(H)| \, dH_N(x)$ 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 \( \Delta = \Delta_N = c N^{-1/2} \). 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 and clearly hold for the \( t \)-statistic

\[
(5.1) \quad \mathcal{L} \left( \frac{\bar{T}_N - \bar{a}_N}{b_N} \right) \Rightarrow N(0,1)
\]

\[
(5.2) \quad \lim_{N \to \infty} \frac{b_N(\Delta_N)}{b_N(0)} = 1
\]

\[
(5.3) \quad E_T = \lim_{N \to \infty} \frac{\left[ (a_N(\Delta_N) - a_N(0)) \right]^2}{\Delta_N N^{1/2} b_N(0)}
\]
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 \( \Psi \). In comparing two sequences of tests, say \( T_N \) and \( T^*_N \), for near alternatives the two tests will have the same power only when the corresponding sample sizes, \( N \) and \( N^* \), satisfy the following relationship

\[
(5.4) \quad \lim_{N \to \infty} \frac{N^*}{N} = \frac{E_T}{E_{T^*}} = E_{T,T^*}
\]

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 cdf \( \Phi \) and applying Corollary I and using derivatives in the expression for \( E_T \), we have

\[
(5.5) \quad E_{c_1,t}(\Psi) = I_{1\Psi}/\sigma^2
\]

where

\[
(5.6) \quad I_{1\Psi} = \int J_0[\Psi(x)]\Psi^2(x)dx
\]

and \( \sigma^2_{\Psi} \) is the variance of the distribution with cdf \( \Psi \) (and density \( \Psi \)). Normalizing \( \Psi \) 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.** For all continuous cdf \( \Psi \), \( E_{c_1,t}(\Psi) \geq 1 \) and \( E_{c_1,t}(\Psi) = 1 \) only if \( \Psi \) is normal.
Proof: It suffices to show that the minimum of $I_1 \Psi$ subject to the restrictions

$$I_2 \Psi = \int x \psi(x) \, dx = 0$$

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 $\Psi$ 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 (having a jump) 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 $\Psi$. We have

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

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

and

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

Furthermore

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

and

$$\psi(x) \, dx = d\Psi(x) = \Phi(J_0) \, dJ_0.$$
Consequently our problem consists of finding a monotone function $x(J_0)$ which minimizes

$$I_1\Psi = \int \frac{1}{\varphi(J_0)} \frac{\varphi(J_0)}{\frac{dx}{dJ_0}} \varphi(J_0) dJ_0 = \int \frac{\varphi(J_0)}{\frac{d\varphi}{dJ_0}} dJ_0$$

subject to the restrictions and

$$I_2\Psi = \int x\Psi(x)dx = \int x\varphi(J_0)dJ_0 = 0$$

$$I_3\Psi = \int x^2\Psi(x)dx = \int x^2\varphi(J_0)dJ_0 = 1$$

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 $x = (1-\lambda)x_1 + \lambda x_2$. Then, by convexity

$$I_1(\lambda) = \int \frac{\varphi(J_0)}{\frac{d\varphi}{dJ_0}} dJ_0 < (1-\lambda) \int \frac{\varphi(J_0)}{\frac{d\varphi}{dJ_0}} dJ_0 + \lambda \int \frac{\varphi(J_0)}{\frac{d\varphi}{dJ_0}} dJ_0$$

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

\[ I_3(\lambda) = \int x_1^2 \varphi(J_0) dJ_0 < (1-\lambda) \int x_1^2 \varphi(J_0) dJ_0 + \lambda \int x_2^2 \varphi(J_0) dJ_0. \]

Hence \( x_1 \) and \( x_2 \) cannot both be solutions of the minimization problem since otherwise a multiple of \( \frac{x_1 + x_2}{2} \) 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 \( I_1 \) then does \( x_1 \). Then using the convexity again, we have

\[
I_1'(0) = -\int \frac{d(x_2-x_1)}{dJ_0} \frac{dJ_0}{(dx_2/dJ_0)^2} \varphi(J_0) dJ_0 < 0
\]

\[
I_2'(0) = \int (x_2-x_1) \varphi(J_0) dJ_0 = 0
\]

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 \( \xi \geq 0 \) such that

\[
I_1'(0) + \xi I_3'(0) \geq 0,
\]
then $x_1$ is the unique solution of the minimization problem. 2/

Now

$$I_1'(0) = \frac{-(x_2-x_1)}{(dx_1^2)_{x_1}} \phi(J_0) \left[ 1 \int_{-\infty}^{\infty} (x_2-x_1) \left( \frac{d^2 x_1}{dx_1^2} \right) - \frac{2dx_1}{dJ_0} \right] \phi(J_0) \left[ \frac{d^2 x_1}{dx_1^2} \right] \phi(J_0) \left[ \frac{d x_1}{dJ_0} \right] \phi(J_0) \right] dJ_0.

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

$$I_1'(0) + \xi I_3'(0) = f(x_2-x_1)[\phi(J_0) + \xi J_0 \phi(J_0)] dJ_0$$

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

If we regard 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 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 $F_0(x-\phi)$ respectively where $F_0$ has a twice differentiable density $f_0$. Then an efficient test statistic for $\Delta = \theta - \phi = 0$ would be

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

2/ 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.
for which the asymptotic distribution is normal with mean $\Delta$ and variance 
$[\mathbb{N}(1-\lambda)(\text{Inf}_{i_0}^o)]^{-1}$ where

\begin{equation}
\text{Inf}_{i_0}^o = \int \frac{[f_o'(x)]^2}{f_o(x)} \, dx
\end{equation}

providing the above integral exists. The relative efficiency of our non-parametric test $T$ with a specified $J$ to the efficient $\hat{\Delta}$ test is

\begin{equation}
E_{T,\hat{\Delta}} = \frac{I_{1F}^o}{(\text{Inf})_{i_0}^o}
\end{equation}

where

\begin{equation}
I_{1F}^o = \int J'(f_o) f_o^2(x) \, dx
\end{equation}

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

\begin{equation}
J(u) = \frac{f_o'(x)}{f_o(x)} (\text{Inf}_{i_0}^o)^{-1/2} \quad \text{where} \quad u = f_o(x)^3/2.
\end{equation}

In fact for this $J$, we have

\begin{align*}
I_{1F}^o &= (\text{Inf}_{i_0}^o)^{-1/2} \int \left[ \frac{f_o''(x)}{f_o(x)} - \frac{[f_o'(x)]^2}{f_o(x)} \right] \frac{1}{f_o(x)} f_o^2(x) \, dx = (\text{Inf}_{i_0}^o)^{-1/2}
\end{align*}

and

\begin{equation*}
E_{T,\hat{\Delta}} = 1.
\end{equation*}

\text{3/} \ J \ has \ been \ normalized \ to \ have \ \int J(u) \, du = 0 \ \text{and} \ \int J^2(u) \, du = 1.
As it is to be expected if \( F_0 = \hat{\Phi} \) (normal), the corresponding \( J = J_0 \), the inverse of \( \hat{\Phi} \). The problem of comparing the non-parametric 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.

Hoeffding [4] has 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. Hoeffding U-statistics include many non-parametric test statistics and he and Lehmann have shown that U-statistics are asymptotically normal under the alternative hypothesis. The U-statistics do not include all statistics of the form

\[
T_N = \sum_{i=1}^{N} \frac{E_{N_i}}{N_i} z_{N_i}.
\]

In particular \( c_1 \) is not a U-statistic. Dwass's [2] Theorem 7.2 (quoted below) appears to be the only result for statistics of the form (3.1) under general alternative hypotheses.

We first state Dwass's theorem and examine its relationship to our Theorem I.

**Theorem 7.2.** Suppose

1. \[ \max(|b_1|, \ldots, |b_h|) > 0 \]
2. \[ \lim_{N \to \infty} \lambda_N = \lambda \quad (0 < \lambda < 1) \]

Then
\[ \lim_{N \to \infty} \Pr \left( \frac{t_N - \mu_0 \theta}{\sigma_0 \theta} < s \right) = \int_{-\infty}^{s} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx \]

where
\[ t_N = \sum_{i=1}^{N} a_{Ni} P(R_i/N) \]

and
\[ P(t) = \sum_{k=1}^{h} b_k t^k , \]
\[ a_1 = (n/mN)^{1/2} \quad i = 1, \ldots, m \]
\[ a_{Ni} \]
\[ a_2 = -(m/nN)^{1/2} \quad i = m+1, \ldots, N \]

(4) \[ F(x) = M(x, \theta) , \quad G(x) = M_2(x, \theta) \quad \text{and} \quad M_1(x, 0) = M_2(x, 0) \]
\[ \theta = \theta(N) = \delta N^{-1/2} , \quad \delta > 0 \]

(5) \[ \lim \frac{\sigma_{\theta N}^2}{\theta} = b'DADB \]
\[ A_{1, j} = \left( \frac{1}{1+j+1} \right) \]
\[ D = \begin{pmatrix}
1/2 & 0 \\
0 & 2/3 \\
0 & h \\
h & h+1
\end{pmatrix} \]

and if \( (X_1, \ldots, X_m, Y_1, \ldots, Y_n) \) is written as \( (U_1, \ldots, U_N) \) then \( R_1 \) is
the number of \( U \)'s less or equal to \( U_1 \). First note

\[
t_N = \sum_{i=1}^{N} P\left(\frac{i}{N}\right) \left[ a_{1\cdot Ni} + a_2(1 - z_{Ni})\right]
\]

\[
= \sum_{i=1}^{N} P\left(\frac{i}{N}\right) z_{Ni} \left( \frac{1}{\sqrt{N}} \left( \frac{1 - \lambda_N}{\lambda_N} \right)^{1/2} - \left( \frac{\lambda_N}{1 - \lambda_N} \right)^{1/2} \right) + a_2 \sum_{i=1}^{N} P\left(\frac{i}{N}\right)
\]

\[
= \sqrt{N} T_N \left( \frac{1 - \lambda_N}{\lambda_N} \right)^{1/2} - \left( \frac{\lambda_N}{1 - \lambda_N} \right)^{1/2} \right) + K
\]

where in \( T_N \) we have \( E_{Ni} = P\left(\frac{i}{N}\right) \). Thus there is a non-stochastic linear relationship between \( t_N \) and \( T_N \). Hence, from the statistical viewpoint \( t_N \) and \( T_N \) are "equivalent". Now let us consider Dwass's conditions.

(1) The requirement \( \max(|b_1|, \ldots, |b_h|) > 0 \) is to insure that \( E_{Ni} \neq 0 \), a trivial case which causes no difficulty.

(2) Dwass requires \( \lambda_N \to \lambda \) and we only require \( 0 < \lambda_0 \leq \lambda_N \leq 1 - \lambda_0 < 1 \).

(3) The condition \( E_{Ni} = J_N(i/N) = P(i/N) = \sum_{k=1}^{h} b_k \left(\frac{i}{N}\right)^k \) is much stronger than our condition 4 in Theorem I 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 \( 0 < x < 1 \).

(4) Dwass's results only apply to alternatives \( (F, G) \) near each other. We have made no restrictions of this kind and our conclusions are "uniform".

(5) Using Corollary I Dwass's form of the variance is easily obtained.
6.B. Applications.

Example 1: Let $E_{N1} = \sum_{j=1}^{N} j^{-\frac{1}{2}}$. Then Savage has proved [7] 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 $G(x) = e^{-\theta_2 x}$ for $-\infty < x \leq 0$ 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 I. To do so we note that $J_N\left(\frac{i}{N}\right)$ is the expected value of the $i$-th 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 [9] 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 $(0,1)$ 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 I 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 I are uniformly of higher order the following elementary results are used repeatedly.

7.A. **Elementary results.**

1. $H > \lambda_N F > \lambda_0 F$

2. $H > (1 - \lambda_N)G > \lambda_0 G$

3. $1 - F \leq \frac{1-H}{\lambda_N} \leq \frac{1-H}{\lambda_0}$

4. $1 - G \leq \frac{1-H}{1-\lambda_N} \leq \frac{1-H}{\lambda_0}$

5. $F(1-F) \leq \frac{H(1-H)}{\lambda_N^2} \leq \frac{H(1-H)}{\lambda_0^2}$

6. $G(1-G) \leq \frac{H(1-H)}{\lambda_0^2}$

7. $dH > \lambda_N dF > \lambda_0 dF$

8. $dH > (1 - \lambda_N)dG > \lambda_0 dG$

9. Let $(a_N, b_N)$ be the interval $S_{N\epsilon}$ where

\begin{equation}
S_{N\epsilon} = \left\{x : H(1-H) > \frac{\eta\lambda_0}{N}\right\}.
\end{equation}
Then $\eta_\epsilon$ can be chosen independently of $F, G$ and $\lambda_N$ so that

\[(7.2)\quad P\{X_i \in S_{N\epsilon}, Y_j \in S_{N\epsilon}, i = 1,2,\ldots,m, j = 1,2,\ldots,n\} \geq 1 - \epsilon .\]

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

**Proof:** Using assumption 4 of Theorem I and A.7

\[(7.3)\quad \int_{-\infty}^{\infty} J(H(x))dF(x) \leq K \int_{0}^{1} \frac{1}{[H(1-H)]^{1/2}} dH \leq K .\]

\[\leq K \int_{0}^{1} \frac{dH}{[H(1-H)]^{1/2}} \leq K .\]

7.B. **Detailed consideration of the second order terms of Theorem I.**

We are now ready to show that the $C$ terms are uniformly of higher order.

We begin with $C_{IN}$ and prove the following identity.

\[(7.4)\quad C_{IN} = \lambda_N \int_{-\infty}^{\infty} (F_m - F)J(H)d(F_m(x) - F(x)) = \frac{\lambda_N}{2} \left[ \int J'(H)d(F_m - F)^2 + \frac{1}{m} \int J'(H)dF_m \right] .\]

Let $R$ be the set of points of increase of $F_m$. Then the right hand side of the identity becomes
\[ \frac{\lambda_N}{2} \left[ \int_R J'(H) d(F_m - F)^2 + \int_R J'(H) d(F_m - F)^2 + \frac{1}{m} \sum_{i=1}^{m} J'(H(X_i)) \frac{1}{m} \right] \]

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

\[ + \left. \frac{1}{m} \sum_{i=1}^{m} J'(H(X_i)) \frac{1}{m} \right] \]

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

\[ = \lambda_N \int (F_m - F) J'(H) d(F_m - F) \]

Using this identity we integrate by parts and obtain

(7.5) \[ C_{1N} = -\frac{\lambda_N}{2} \left( C_{11N} + C_{12N} - C_{13N} \right) \]

where

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

\[ C_{12N} = \int_{S_{N\epsilon}} (F_m - F)^2 J''(H) dH \]

\[ C_{13N} = \frac{1}{m} \int J'(H(x)) dF_m(x) \]

\[ = \frac{1}{m} \sum_{i=1}^{m} J'(H(X_i)) \]

where \[ S_{N\epsilon} \] was defined in A.9.
Now let us consider the random variable $C_{11N}$. We find

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

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

$$\mathbb{E}C_{11N} \leq \frac{K}{N} \int_{S_{N\epsilon}} \frac{H(1-H)}{[H(1-H)]^2} \frac{dH}{2 - 5} \leq \frac{K}{N} \int_{1}^{1} \frac{1}{2} \frac{dH}{H^2} \leq \frac{K}{N^{1/2} + 5}.$$  

Now using the Markoff inequality [1, p. 182]

$$\Pr(C_{11N} > aN^{-1/2}) \leq \frac{K}{N^{1/2} + 5} \frac{N^{1/2}}{a} = \frac{K}{aN^2}$$

where $K$ may depend on $\epsilon$. Now consider $C_{12N}$.

Let $H_1 = H(a_n)$, $H_2 = H(b_n)$ as in 7.A.9. Then $H_1 = 1 - H_2 < \frac{K}{N}$. With probability greater than $1 - \epsilon$ we have

$$C_{12N} = \int_{S_{N\epsilon}} (F_m-F)^2 J''(H) dH = \int_{H_1}^{H} F^2 J''(H) dH + \int_{H_2}^{1} (1-F)^2 J''(H) dH.$$
\[ |c_{12N}| \leq K \int_0^1 \frac{H^2 dH}{(H(1-H))^\frac{5}{2} - \delta} + \int_{H_2}^1 \frac{(1-H)^2 dH}{(H(1-H))^\frac{5}{2} - \delta} \]

\[ \leq K \int_0^K \frac{H - \frac{1}{2} + \delta}{H} dH \leq KN - \frac{1}{2} - \delta . \]

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

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

We may assume \( \delta < \frac{3}{2} \) or \( \delta < \frac{1}{2} \) without loss of generality. Then

\[ |c_{13N}| \leq \frac{K}{N} \frac{1}{m} \sum_{i=1}^m [F(X_i)[1-F(X_i)]]^{-\frac{3}{2} + \delta} . \]

which is distribution free. By a theorem of Marcinkiewicz [5, p. 242-3] if a random variable \( Y \) has \( r \)-th order moment finite \( (0 < r < 1) \) then the sum of \( N \) independent observations on \( Y \) is \( o_p(N^{-\frac{1}{r}}) \). If \( X \) has cdf \( F \),

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

has a finite moment of order \( \frac{2}{3-\delta} \) and hence

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

Consequently \( c_{1N} = o_p(N^{-\frac{1}{2}}) \).
Next consider

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

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

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 \(S_{N\epsilon}\) and

\[ C_{21N} \leq K \int_{S_{N\epsilon}} H(1-H) [H(1-H)]^{-\frac{3}{2}} \delta dH(x) \leq K \eta_\epsilon \frac{1}{2} + \delta - \frac{1}{2} - \delta \]

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

\[ \mathcal{E}(C_{22N}) = \mathcal{E} \{ \mathcal{E} C_{22N} \mid X_1, X_2, \ldots, X_m \} = 0 \]

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

\[ C_{23N} = \frac{2}{n} \int_{S_{N\epsilon}} \int_{x < y} g(x)(1-G(y)) J'[H(x)] J'[H(y)] d[F_m(x)-F(x)] d[F_m(y)-F(y)] \]
\[ c_{24N} = \frac{1}{nm} \int_{S_{N \epsilon}} G(x)[1-G(x)]\left\{ J'[H(x)] \right\}^2 dF_m(x) \]

\[ \mathcal{E}(c_{23N}) = \frac{2}{nm} \int_{x,y \in S_{N \epsilon}} G(x)[1-G(y)]J'[H(x)]J'[H(y)]dF(x)dF(y) \]

\[ \leq \frac{K}{N^2} \int_{x < y} H(x)[1-H(y)]J'[H(x)]J'[H(y)]dH(x)dH(y) \]

\[ \leq \frac{K}{N^2} \int_{0 < x < y < 1} x^{-\frac{1}{2} + \delta} (1-x)^{-\frac{3}{2} + \delta} y^{-\frac{3}{2} + \delta} (1-y)^{-\frac{1}{2} + \delta} \, dx \, dy \leq \frac{K}{N^2} \]

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

\[ \leq \frac{Kn_\epsilon^{-1+2\delta}}{N^{1+2\delta}} = o(N^{-1}) \]

Hence

\[ \mathcal{E}(c_{22N}^2 | X_1, X_2, \ldots, X_m) \leq K \, o_p(N^{-1}) \]

where \( K \) may depend on \( \epsilon \) and

\[ |c_{22N}| \leq K \, o_p(N^{-1/2}) \]

since

\[ P(c_{22N}^2 > a \, \mathcal{E}(c_{22N}^2 | X_1, \ldots, X_m)) < 1/a \]

Consequently \( c_{2N} = c_{1N} + c_{22N} \), which does not involve \( \epsilon \), satisfies

\[ c_{2N} = o_p(N^{-1/2}) \]

\[ ^{\frac{1}{2}} \text{ This integrand has already appeared as part of the variance in equation (4.3).} \]
Now consider

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

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

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

and

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

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

\[(7.9) \quad [\varphi_{H_N} + (1 - \varphi)H][1 - (\varphi_{H_N} + (1 - \varphi)H)] > \eta_{\varepsilon}^* H(x)[1 - H(x)].\]

Then

\[
|C_{3N}| \leq \frac{1}{N} \sup_{S_{N, \varepsilon}} \left[ \int [H_N(x) - H(x)]^2 (\eta_{\varepsilon}^*)^{-\frac{5}{2} + \delta} \left\{H[1 - H]\right\}^{-\frac{5}{2} + \delta} dF_{m}(x) = (\eta_{\varepsilon}^*)^{-\frac{5}{2} + \delta} C_{31N} \right.
\]

\[
E(|C_{31N}|) \leq \frac{1}{N} \left[ \int_{S_{N, \varepsilon}} \left[ \lambda_N P(1 - P) + \frac{(1 - P)(1 - 2P)}{N} + (1 - \lambda_N) G(1 - G) \right] [H(1 - H)]^{-\frac{5}{2} + \delta} dF(x) \right.
\]

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

\[
- \frac{1}{2} + \delta + \frac{3}{2} + \delta
\]

\[
\leq \frac{K \eta_{\varepsilon}^*}{N^2} + \frac{K \eta_{\varepsilon}^*}{N^2} + \delta
\]
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

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

\[
\frac{1}{m} \left| J[H(X_m)] \right| \leq \frac{[H(X_m)][1-H(X_m)]}{m} \leq \frac{(\eta^*_e) - \frac{1}{2} + \delta}{N^2 + \delta}
\]

with probability at least \( 1 - \epsilon \). Hence

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

Similarly

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

\[
= o_p(N^{-1/2})o_p(1) = o_p(N^{-1/2})
\]

Hence

\[ C_{4N} = o_p(N^{-1/2}) \]

The negligibility of \( C_{5N} \) and \( C_{6N} \) follows immediately from assumptions 2 and 3 of Theorem I.
7.C. Proof of Theorem II.

First we note that

\begin{align}
J_N(\frac{1}{N}) = E_{1,N} = \int_0^1 J(u)g_{i,N}(u)du
\end{align}

where

\begin{align}
\ell_{i,N}(u) = \frac{N!}{(i-1)! (N-i)!} u^{i-1} (1-u)^{(N-i)}
\end{align}

is the density of the i-th order statistic from the uniform distribution on [0,1] and incidentally has mean \( \frac{1}{N+1} \) and variance \( \frac{i(N+1)}{(N+1)^2(N+2)} \).

Then we have

\begin{align}
|E_{1,N}| \leq KN \int_0^1 [u(1-u)]^{-\frac{1}{2}\delta} (1-u)^{N-1} du
\end{align}

\begin{align*}
= \frac{KN \Gamma(N - \frac{1}{2} + \delta) \Gamma(\frac{1}{2} + \delta)}{\Gamma(N+2\delta)} \leq \frac{1}{N^2} - \delta
\end{align*}

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

\begin{align}
|J_N(\frac{1}{N}) - J(\frac{1}{N})| \leq \frac{1}{N^2} - \delta + K \left[ \frac{1}{N} \left( 1 - \frac{1}{N} \right) \right] - \frac{1}{2} + \delta \leq \frac{1}{N^2} - \delta
\end{align}

Before proceeding to bound \( J_N(\frac{1}{N}) - J(\frac{1}{N}) \) for \( 1 < i \leq \frac{N}{2} \) we apply the Stirling formula

\begin{align}
\log x! = \log \Gamma(x+1) = \frac{1}{2} \log 2\pi - x + (x + \frac{1}{2}) \log x + \frac{\theta}{12x} \quad 0 < \theta < 1
\end{align}
with a rather standard argument to obtain for \(1 < i \leq \frac{N}{2},\; 0 < u < \frac{i-1}{N-1}\),

\[(7.16)\quad \varepsilon_{i,N}(u) \leq \sqrt{\frac{(N-1)^3}{2\pi(i-1)(N-1)}} e^{-\frac{u^2}{2} \left[ \frac{(N-1)}{\lambda N} \right]} \left[ 1 + \frac{K}{N} \right]^{\frac{5}{2}}\]

where

\[(7.17)\quad v = (N-1)u - (i-1)\,.

For \(1 < i \leq N/2\)

\[(7.18)\quad J_{\frac{i}{N}}^N - J_{\frac{i}{N}} = \int_0^1 \left[ J(u) - J_{\frac{i}{N}} \right] \varepsilon_{i,N}(u) du

= D_{11} + D_{12} + D_{21} + D_{22} + D_3 + D_4\]

where

\[D_{11} = \int_0^{u_1} J(u) \varepsilon_{i,N}(u) du, \quad D_{12} = \int_{u_1}^1 J(u) \varepsilon_{i,N}(u) du\]

\[D_{21} = \int_0^{u_1} J_{\frac{i}{N}}^N \varepsilon_{i,N}(u) du, \quad D_{22} = \int_{u_1}^1 J_{\frac{i}{N}}^N \varepsilon_{i,N}(u) du\]

\[D_3 = \int_{u_1}^{1-u_1} (u - \frac{i}{N}) J_{\frac{i}{N}}^N \varepsilon_{i,N}(u) du\]

\[D_4 = \frac{1}{2} \int_{u_1}^{1-u_1} (u - \frac{i}{N})^2 J''(u^*) \varepsilon_{i,N}(u) du,\]

where \(u^*\) between \(u\) and \(\frac{i}{N}\), and \(u_1 = \frac{i-1}{2(N-1)}\).

\[\text{\(\frac{5}{2}\)}\quad K\] represents a generic constant independent of \(i, N, \lambda N, F\) 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.
(7.19) \[ g_{1,N}(u) = u^{\xi} \frac{u^{1-\alpha}(1-u)^{N-1}(N-\alpha)!}{(1-\alpha)!} \frac{N!}{(N-\alpha)!} \frac{(i-1-\alpha)!}{(i-1)!} \leq K u^{\alpha} N g_{1-\alpha,N-\alpha}(u) \]

where \( \alpha = \frac{1}{2} - \delta \). Let \( \Phi \) be the normal cdf. Then

\[ |D_{11}| \leq \int_{0}^{u_1} \Phi \left[ \frac{(u_1 - \frac{1-\alpha}{N-1-\alpha}(N-1-\alpha)^{3/2}}{\sqrt{(1-\alpha)(N-1)}} \right] \]

(7.20) \[ |D_{11}| \leq K N^{\alpha} \Phi \left( \frac{-\sqrt{1}}{K} \right) \]

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

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

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

\[ D_3 = - \frac{1}{N} \left[ \int_{0}^{u_1} (u - \frac{1}{N}) g_{1,N}(u) du + \int_{1-u_1}^{1} (u - \frac{1}{N}) g_{1,N}(u) du + \frac{i}{N(N+1)} \right] \]

Now

\[ h(u) = |u - \frac{1}{N}| g_{1,N}(u) \leq Kh(1-u) \quad \text{for} \quad u < u_1 \]
Hence

\[ |D_3| \leq K \left( \frac{1}{N} \right)^{\alpha-1} \left[ K \frac{1}{N} \Phi \left( \frac{-\sqrt{I}}{K} \right) + \frac{1}{N(N+1)} \right] \leq KN^\alpha \Phi \left( \frac{-\sqrt{I}}{K} \right) + KN^{\alpha-1}. \]

Finally

\[ |D_4| \leq Ku_1 \left( \frac{5}{2} + \delta \right) \int_0^1 (u - \frac{i}{N})^2 \varepsilon_{i,N}(u) du \]

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

\[ |D_4| \leq Ku_1 \left( \frac{5}{2} + \delta \right) \left[ \frac{u_1}{N} + \frac{u_2}{N^2} \right] \leq \frac{Ku_1}{N} \leq \frac{KN^\alpha}{1+\alpha}. \]

Thus, for \( 1 < i \leq N/2 \),

\[ J_N \left[ \frac{i}{N} \right] - J_N \left[ \frac{i-1}{N} \right] \leq KN^\alpha \left[ \Phi \left( \frac{-\sqrt{I}}{K} \right) + \frac{1}{N} + \frac{1}{1+\alpha} \right] \]

and

\[ \left| \int_{1 \leq m \leq N/2} [J_N(H_N) - J(H_N)] dF_m \right| \leq \frac{1}{m} \left\{ \frac{1}{2} + \delta + \frac{N/2}{1+\alpha} \left[ \Phi \left( \frac{-\sqrt{I}}{K} \right) + \frac{1}{N} + \frac{1}{1+\alpha} \right] \right\} \]

\[ \leq \frac{1}{m} - \delta \]

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


<table>
<thead>
<tr>
<th>Chief, Bureau of Ships</th>
<th>Scientific Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept. of the Navy</td>
<td>Office of Naval Research</td>
</tr>
<tr>
<td>Washington 25, D.C.</td>
<td>Dept. of the Navy</td>
</tr>
<tr>
<td>Attn: H. Weingarten</td>
<td>1000 Geary St.</td>
</tr>
<tr>
<td>Code 223</td>
<td>San Francisco 9, Calif.</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Commanding Officer</td>
<td></td>
</tr>
<tr>
<td>U. S. Naval Powder Factory</td>
<td>Statistical Engineering Lab.</td>
</tr>
<tr>
<td>Indianhead, Md.</td>
<td>National Bureau of Standards</td>
</tr>
<tr>
<td>Attn: F. Freshman</td>
<td>Washington 25, D.C.</td>
</tr>
<tr>
<td>R. and D.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Director, Naval Research Lab.</td>
<td>Dept. of Statistics</td>
</tr>
<tr>
<td>Washington 25, D.C.</td>
<td>University of California</td>
</tr>
<tr>
<td>Attn: Technical Information Officer</td>
<td>Berkeley 4, Calif.</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Document Service Center</td>
<td></td>
</tr>
<tr>
<td>A.S.T.I.A.</td>
<td>Statistical Laboratory</td>
</tr>
<tr>
<td>Knott Building</td>
<td>University of Washington</td>
</tr>
<tr>
<td>Dayton 2, Ohio</td>
<td>Seattle 5, Wash.</td>
</tr>
<tr>
<td>Attn: DSC-S</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Dept. of Mathematical Statistics</td>
<td>Professor O. P. Aggarwal</td>
</tr>
<tr>
<td>University of North Carolina</td>
<td>Statistical Lab.</td>
</tr>
<tr>
<td>Chapel Hill, N. C.</td>
<td>Purdue University</td>
</tr>
<tr>
<td></td>
<td>Lafayette, Indiana</td>
</tr>
<tr>
<td>Office of Naval Research</td>
<td></td>
</tr>
<tr>
<td>Dept. of the Navy</td>
<td>Dr. Stephen G. Allen</td>
</tr>
<tr>
<td>17th and Constitution Aves.</td>
<td>244 Laurel St.</td>
</tr>
<tr>
<td>Attn: Code 433</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Office of the Asst. Naval Attaché for Research</td>
<td>Professor Fred C. Andrews</td>
</tr>
<tr>
<td>Naval Attaché</td>
<td>Mathematics Dept.</td>
</tr>
<tr>
<td>American Embassy</td>
<td>University of Nebraska</td>
</tr>
<tr>
<td>Navy No. 100</td>
<td>Lincoln 8, Nebr.</td>
</tr>
<tr>
<td>Fleet Post Office</td>
<td>1</td>
</tr>
<tr>
<td>New York, N. Y.</td>
<td></td>
</tr>
<tr>
<td>Office of Technical Services</td>
<td>Professor T. W. Anderson</td>
</tr>
<tr>
<td>Dept. of Commerce</td>
<td>Dept. of Mathematical Statistics</td>
</tr>
<tr>
<td>Washington 25, D.C.</td>
<td>Columbia University</td>
</tr>
<tr>
<td></td>
<td>New York 27, N.Y.</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Professor David Blackwell</td>
</tr>
<tr>
<td></td>
<td>Dept. of Statistics</td>
</tr>
<tr>
<td></td>
<td>University of California</td>
</tr>
<tr>
<td></td>
<td>Berkeley 4, Calif.</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dr. Julius R. Blum</td>
</tr>
<tr>
<td></td>
<td>Dept. of Mathematics</td>
</tr>
<tr>
<td></td>
<td>Indiana University</td>
</tr>
<tr>
<td></td>
<td>Bloomington, Indiana</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Dr. Paul Blunk  
Box 532  
Fair Oaks, Calif.  

Dr. Charles Boll  
Hughes Aircraft Co.  
Bldg. 12, Room 2343  
Culver City, Calif.  

Professor W. G. Cochran  
Biostatistics Dept.  
The Johns Hopkins University  
Baltimore 5, Md.  

Professor Lee Cronbach  
Bureau of Education Research  
1007 S. Wright  
Champaign, Ill.  

Dr. Joseph Daly  
Bureau of the Census  
Washington 25, D.C.  

Dr. Francis Dresch  
Stanford Research Institute  
1915 University Ave.  
Palo Alto, Calif.  

Professor Meyer Dwass  
Dept. of Mathematics  
Northwestern University  
Evanston, Ill.  

Mr. Murray A. Geisler  
Logistics Section  
The RAND Corporation  
1700 Main St.  
Santa Monica, Calif.  

Mr. Geoffrey Gregory  
4 Osborne Grove  
Gatley, Cheadle  
Cheshire, England  
c/o Office of Naval Research  
London Branch  
Navy No. 100, Fleet Post Office  
New York, N. Y.  

Professor E. J. Gumbel  
Industrial Engr. Dept.  
409 Engineering Bldg.  
Columbia University  
New York 27, N. Y.  

Dr. Theodore E. Harris  
The RAND Corporation  
1700 Main St.  
Santa Monica, Calif.  

Professor Leonid Hurwicz  
School of Business Administration  
University of Minnesota  
Minneapolis 14, Minn.  

Professor Stanley Issacson  
4715 Pleasant Street  
Des Moines, Iowa  

Professor Leo Katz  
Statistics Dept.  
Michigan State University  
East Lansing, Mich.  

Professor Tosio Kitagawa  
Mathematical Institute  
Faculty of Science  
Kyuusy University  
Fukuoka, Japan  

Dr. Dennis V. Lindley  
Statistical Lab.  
University of Cambridge  
Cambridge, England  
c/o Office of Naval Research  
London Branch  
Navy No. 100, Fleet Post Office  
New York, N. Y.  

Dr. Eugene Lukacs  
Dept. of Mathematics  
Catholic University  
Washington 17, D. C.
Mr. Monroe Norden  
Research Division  
Engineering Statistics Group  
College of Engineering  
New York University  
401 W. 205 St.  
New York 54, N. Y.  

Dr. A. R. Roy  
Statistical Wing  
Indian Council of Agricultural Research  
Linlithgow Ave.  
New Delhi, India  

Dr. Jagdish Rustagi  
Dept. of Mathematics  
College of Engineering and Science  
Carnegie Inst. of Tech.  
Pittsburgh 13, Pa.  

Dr. Milton Sobel  
Bell Telephone Labs  
555 Union Blvd.  
Allentown, Pa.  

Dr. Dan Teichroew  
National Cash Register Co.  
Product Development Dept.  
1401 E. El Segundo Blvd.  
Hawthorne, California  

Professor Donald Truax  
Mathematics Dept.  
University of Kansas  
Lawrence, Kansas  

Mr. Cesareo Villegas  
Institute de Mathematica y Estadistica  
Av. J. Herrera y Reissig 565  
Montevideo, Uruguay  

Professor W. Allen Wallis  
Committee on Statistics  
University of Chicago  
Chicago 37, Ill.  

Dr. Oscar Wesler  
Dept. of Mathematics  
University of Michigan  
Ann Arbor, Mich.  

Dr. John D. Wilkes  
Office of Naval Research  
Code 200  
Washington 25, D. C.  

Professor J. Wolfowitz  
Mathematics Dept.  
Cornell University  
Ithaca, N. Y.  

Professor M. A. Woodbury  
Dept. of Mathematics  
New York University  
New York 53, N. Y.  

Additional copies for project leader and assistants and reserve for future requirements 50