THE LARGE-SAMPLE POWER OF RANK ORDER TESTS IN THE TWO-SAMPLE PROBLEM[1]

By Meyer Dwass

Northwestern University

1. Summary. This paper studies the large-sample power of certain rank order tests against one-parameter alternatives in the two-sample problem.

The first m of N independent random variables are supposed identically distributed, each with a density function $f_1(x, \theta)$, the remaining $N - m$ with a density function $f_2(x, \theta)$. When $\theta = 0$ both density functions are the same. Let $a_{N1}, \cdots, a_{NN}$ be a set of constants defined by (3.2) below; let $b_{N1}, \cdots, b_{NN}$ be another set of constants; and let $R_1, \cdots, R_N$ be the ranks of the N random variables. A statistic of the type $\sum_{i=1}^{N} a_{Ni} b_{NR_i}$ is called an L statistic.

Part I of this paper characterizes the locally best rank order statistic for testing $H_0 : \theta = 0$ against the alternative that θ is positive and "close" to zero. This turns out to be any one of an equivalent class of L statistics. Under certain regularity conditions it is possible to determine the large-sample power of L statistics. Of particular interest is the large-sample power of the locally best L statistic.

For arbitrary $b_{N1}, \cdots, b_{NN}$ it is usually difficult to determine whether the regularity conditions hold. Hence, in Part II a special class of L statistics, the L_h statistics, are studied. For these, the regularity conditions are easier to verify and the large-sample power is determined. The best L statistic can, in a certain sense, be approximated by L_h statistics.

Part I

2. Large-sample power. Suppose it is known that for every positive integer N, the joint c.d.f. of the random vector $\underset{\sim}{X}_N = (X_{N1}, \cdots, X_{NN})$ is a member of the one-parameter class of c.d.f.'s, $\{F_{N\theta}, 0 \leqq \theta < \infty\}$. That is, the distribution of $\underset{\sim}{X}_N$ depends on a nonnegative parameter θ.

Consider

(a) the hypotheses, $H_0 : \theta = 0$, $H_1 : \theta > 0$,

(b) a statistic, $t_N = t_N(\underset{\sim}{X}_N)$, and

(c) a decision procedure:

$$\begin{aligned} &\text{Accept } H_0 \quad \text{if } t_N \leqq C_N , \\ &\text{Reject } H_0 \quad \text{if } t_N > C_N . \end{aligned} \tag{2.1}$$

Let $P_\theta(\ \ )$ be the probability of the event in parentheses when θ is the true parameter. Then the power function of (2.1) is $P_\theta\{t_N \geqq C_N\}$, $0 \leqq \theta \leqq \infty$. If

Received January 19, 1955.

[1] Much of the work of Part II of this paper was done under the sponsorship of the Office of Naval Research.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

t_N is asymptotically normal and certain regularity conditions hold, then it is easy to approximate this power for large N. This is done in Theorem 2.1, below.

The following notation is used: $E_\theta Y$, $\mathrm{Var}_\theta\, Y$ are the expectation and variance of a random variable Y when θ is the true parameter;

$$E'_{N,\theta} = \frac{\partial E_\theta t_N}{\partial \theta};$$

$\{\theta(N)\}$ is the sequence of parameter values,

$$\theta(N) = \delta N^{-1/2}, \qquad \delta > 0, N = 1, 2, \cdots,$$

$\Phi(x)$ is the normal (0, 1) c.d.f.; and λ_N is defined by

$$\lambda_N = (C_N - E_0(t_N))(\mathrm{Var}_0\, t_N)^{-1/2}.$$

THEOREM 2.1. (Pitman)

ASSUMPTIONS.

(a) $P_{\theta(N)}\{(t_N - E_{\theta(N)}t_N)(\mathrm{Var}_{\theta(N)}t_N)^{-1/2} < s\} \rightarrow \Phi(s)$ *as* $N \rightarrow \infty$ *for any* s *and any positive* δ.

(b) $E'_{N,\theta}$ *exists for all* θ *in a half-closed interval* $[0, a)$, *where* a *does not depend on* N.

(c) $E'_{N,\theta(N)}/E'_{N,0} \rightarrow 1$, $\mathrm{Var}_{\theta(N)}t_N/\mathrm{Var}_0 t_N \rightarrow 1$, *as* $N \rightarrow \infty$, *for any positive* δ.

(d) $E'_{N,0}(N\ \mathrm{Var}_0 t_N)^{-1/2} \rightarrow c$ *as* $N \rightarrow \infty$.

(e) $\lambda_N \rightarrow \lambda$, *as* $N \rightarrow \infty$, *where* $\Phi(\lambda) = 1 - \alpha$.

CONCLUSION. *Choose any positive* ϵ *and positive* δ. *Then there is an* $N' = N'(\epsilon, \delta)$ *such that*

$$(2.2) \qquad | P_\theta\{t_N \geqq C_N\} - (1 - \Phi(\lambda - \theta N^{1/2}c)) | < \epsilon.$$

for $\theta N^{1/2} = \delta$, *and all* $N \geqq N'$.

PROOF. See [13].

REMARKS ABOUT THEOREM 2.1:

(a) As justified by this theorem, $1 - \Phi(\lambda - \delta c)$ is called the *large-sample power* of the test described by (2.1).

(b) The following is a slight extension that is useful later: The statistics t_N and t'_N are called *asymptotically equivalent* if $(t_N - E_{\theta(N)}t_N) / ((\mathrm{Var}_{\theta(N)}t_N)^{1/2}) - (t'_N - E_{\theta(N)}t'_N) / ((\mathrm{Var}_{\theta(N)}t'_N)^{1/2})$ converges in probability to zero, as $N \rightarrow \infty$. Evidently the large-sample power is the same using either t_N or t'_N.

(c) Let t_N, $\bar{t}_{\bar{N}}$ be two competing statistics, the first based on sample size N, the second based on sample size $\bar{N}$. If

$$\frac{N}{\bar{N}} \rightarrow \frac{\bar{c}^2}{c^2} = e, \qquad (\bar{c} \text{ defined analogously to } c),$$

as $N, \bar{N} \rightarrow \infty$, then the two large-sample powers will be the same. The number e is called the asymptotic efficiency of $\bar{t}_{\bar{N}}$ relative to t_N.

3. Rank order tests. The remainder of this paper deals with rank order tests in the two-sample problem. What this means is now made specific.

ASSUMPTION A. $\underline{X}_N = (X_1, \cdots, X_m, X_{m+1}, \cdots, X_{m+n})$ consists of $m + n = N$ independent random variables. The first m are identically distributed, each with density function $f_1(x, \theta)$; the remaining n are identically distributed, each with density function $f_2(x, \theta)$. These are density functions with respect to Lebesgue measure on the real line, which satisfy the conditions

$$f_1(x, 0) = f_2(x, 0) = f(x), \qquad (-\infty < x < \infty). \tag{3.1}$$

Also $m/N \to K$, $(0 < K < 1)$, as $N \to \infty$.

Assumption A holds throughout this paper even when not explicitly mentioned. The hypotheses considered are always

$$H_0 : \theta = 0, \qquad H_1 : \theta > 0.$$

DEFINITION 3.1. Let $R_i =$ the number of $X_1, \cdots, X_N$ that are less than or equal to X_i; R_i is called the *rank* of X_i, $(i = 1, \cdots, N)$. Let $R = (R_1, \cdots, R_N)$.

Assumption A implies that when $\theta = 0$, the random vector R takes on for its values each of the $N!$ permutations of $(1, \cdots, N)$ with equal probability $1/N!$. (The assumption about density functions implies that $X_1, \cdots, X_N$ have continuous c.d.f.'s, and hence ties among them occur with probability zero.)

Let the $N!$ permutations of $(1, \cdots, N)$ be ordered in some fixed way. Denote these permutations by $p_1, p_2, \cdots, p_{N!}$. Let S_i be the set in N-space where the random vector R equals p_i.

Let S be any set of points in N-space which does not depend on θ. Let

$$\bar{f}(x, \theta) = \prod_{i=1}^{m} f_1(x_i, \theta) \prod_{j=m+1}^{N} f_2(x_j, \theta),$$

$$dx = \prod_{i=1}^{N} dx_i,$$

$$I_s(\theta) = \int_s \bar{f}(x, \theta)\, dx.$$

By the carrier of a density function $f(x)$ is meant the closure of the set of points on the real line where $f(x) > 0$. The following assumption is stated for later reference.

ASSUMPTION B. The carriers of $f_1(x, \theta)$, $f_2(x, \theta)$ do not depend on θ and the following differentiation under the integral sign is permissible:

$$I_s'(0) = \int_s \left. \frac{\partial \bar{f}(x, \theta)}{\partial \theta} \right|_{\theta=0} dx.$$

Also $I_s'(\theta)$ is a continuous function of θ in some half-closed interval $[0, a)$, $a > 0$.

DEFINITION 3.2. A rank order test is a set W in N-space which is a union of some of the sets $S_1, \cdots, S_{N!}$. H_0 is accepted if and only if the observed value of $\underline{X}_N$ is in W. (In other words, the acceptance of H_0 depends only on the ranks.) t_N is called a rank order statistic if it is constant on each set S_i. (In other words, t_N depends on $\underline{X}_N$ only through the ranks.)

Theorem 3.1, below, states some obvious but useful facts about rank order tests. The following notation is needed for Theorem 3.1: Let

$$P_0'(S_i) = \left.\frac{\partial P_\theta(S_i)}{\partial \theta}\right|_{\theta=0} \qquad (i = 1, \cdots, N).$$

Let θ' be positive and let the two sets of integers $(i_1, \cdots, i_r)$, $(j_1, \cdots, j_r)$ be determined by the requirements that

$$P_{\theta'}(S_{i_1}) \geqq \cdots \geqq P_{\theta'}(S_{i_r}) \geqq \cdots \geqq P_{\theta'}(S_{i_{N!}}).$$

and

$$P_0'(S_{j_1}) \geqq \cdots \geqq P_0'(S_{j_r}) \geqq \cdots \geqq P_0'(S_{j_{N!}}).$$

Let $\alpha = r/N!$.

THEOREM 3.1. *Assumptions* A *and* B *imply that*

(a) *A most powerful, size α, rank order test of H_0 against the specific alternative that $\theta = \theta'$ is given by*

$$W = S_{i_1} \cup \cdots \cup S_{i_r}.$$

(b) *There is a number $\theta'' = \theta''(N) > 0$ such that a uniformly most powerful size α, rank order test of H_0 against the alternative, $0 < \theta < \theta''$, is given by*

$$W' = S_{j_1} \cup \cdots \cup S_{j_r}.$$

The proof is an immediate consequence of the definition of the sets $(i_1, \cdots, i_r)$, $(j_1, \cdots, j_r)$.

DEFINITION 3.3. Two statistics t_N, t_N' are called equivalent $(t_N:t_N')$ if $t_N = at_N' + b$, where a, $(a > 0)$, and b are constants which may depend on N. (It is easy to verify that ":" is a bona fide equivalence relationship.)

DEFINITION 3.4. A test W is said to be derived from a statistic t_N if W is the set of points in N-space for which $t_N \geqq C$ for some constant C.

If W is derived from t_N and $t_N:t_N'$, then W is also derived from t_N'. (This follows immediately from the definitions.)

Theorem 3.2, below, gives the structure of a rank order test which is uniformly most powerful for θ close to 0. The following notation is introduced: Let $Z_{N1} \leqq Z_{N2} \leqq \cdots \leqq Z_{NN}$ be the ordered values of $X_1, \cdots, X_N$. Let

$$H_i(x) = \left.\frac{\partial \log f_i(x, \theta)}{\partial \theta}\right|_{\theta=0}, \qquad H(x) = H_1(x) - H_2(x).$$

Let

$$a_{Ni} = \begin{cases} (n/mN)^{1/2}, & i = 1, \cdots, m, \\ -(m/nN)^{1/2}, & i = m + 1, \cdots, N. \end{cases} \tag{3.2}$$

(The facts that $\Sigma a_{Ni} = 0$, $\Sigma a_{Ni}^2 = 1$ are used later.) (Σ, without display of indices, hereafter means $\sum_{i=1}^N$.)

THEOREM 3.2. *Assumptions* A *and* B *imply that the test* W' *of Theorem* 3.1 *is derived from*

$$t_N = \Sigma a_{Ni} E_0 H(Z_{R_i}). \tag{3.3}$$

PROOF. By differentiating and using Assumption B,

$$N!P_0'(S_j) = \sum_{i=1}^{m} E_0 H_1(Z_{Nl_i}) + \sum_{i=m+1}^{N} E_0 H_2(Z_{Nl_i}),$$

where $l_1, \cdots, l_N$ is the value of R on the set S_j.

Set $t_N'(X_N) = N!P_0'(S_j)$ for X_N in S_j, $(i = 1, \cdots, N!)$. Next notice that $\Sigma E_0 H_1(Z_{Ni}) = \Sigma E_0 H_2(Z_{Ni}) = 0$. This is because $\Sigma E_0 H_1(Z_{Ni}) = \Sigma E_0 H_1(X_1)$ and

$$E_0 H_1(X_i) = \int_{-\infty}^{\infty} H_1(t) f(t)\, dt = \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f_1(t, \theta)\, dt \bigg|_{\theta=0} = \frac{\partial(1)}{\partial \theta} = 0. \tag{3.4}$$

(Similarly for H_2.) Hence,

$$t_N' = \sum_{i=1}^{m} E_0 H(Z_{NR_i}) = - \sum_{i=m+1}^{N} E_0 H(Z_{NR_i}),$$

and

$$t_N = a_{N1} t_N' - a_{N,m+1} t_N' = (N/mn)^{1/2} t_N', \tag{3.5}$$

or $t_N : t_N'$. This completes the proof.

REMARKS ON THEOREM 3.2.

(a) Suppose $f_i(x, \theta)$ is the density function of a normal (m_i, σ) random variable $(i = 1, 2)$ and that $\theta = (m_1 - m_2)/\sigma$. Then the statistic t_N of Theorem 3.2 is equivalent to $\Sigma a_{Ni} E Z_{NRi}$, where the Z_{Ni} are the ordered values of N independent normal (0, 1) random variables. This was established differently by Hoeffding in [4].

(b) This is an example that will be used later. Let $f(x)$ be a density function that does not depend on θ. Let $F(x) = \int_{-\infty}^{x} f(t)\, dt$. Let

$$f_1(x, \theta) = f(x),$$

$$f_2(x, \theta) = 2\theta F(x) f(x) + (1 - \theta) f(x) \qquad (0 \leqq \theta \leqq 1)$$

($f_2 = d[\theta F^2 + (1 - \theta)F]$). Then for θ close to zero, Theorem 3.2 says the most powerful test is derived from $t_N = \Sigma a_{Ni} E Z_{NR_i}$ where $Z_{N1} \leqq \cdots \leqq Z_{NN}$ are the ordered values of N independent random variables, uniformly distributed on (0, 1). Since $EZ_{Ni} = i/N + 1$, t_N is equivalent to $\Sigma a_{Ni} R_i$, which is equivalent to the Wilcoxon-Mann-Whitney statistic. This result was found by Lehmann by examining the probabilities $P_\theta(S_i)$ (see [10]).

(c) It would be interesting to know when a test that is most powerful for θ close to zero is uniformly most powerful for all θ in a wider interval. Teichroew [16] has presented some empirical evidence that this may be so in the normal case discussed in Remark (a). This writer has computational evidence of this in some special cases. This is presumably an open problem.

DEFINITION 3.5. The statistic (3.3) is called the locally best rank order statistic and the derived test is called the locally best rank order test.

4. The large-sample power of L tests.

DEFINITION 4.1. Let $b_{N1}, \cdots, b_{NN}$ be a set of constants given for every N. Let $a_{N1}, \cdots, a_{NN}$ be defined by (3.2). The rank order statistic

$$t_N = \Sigma a_{Ni} b_{NR_i} \tag{4.1}$$

is called an L statistic. A test W derived from t_N is called an L test.

Theorem 3.2 states that the locally best rank order test is an L test.

LEMMA 4.1. *Let $b_{N1}, \cdots, b_{NN}$; $b'_{N1}, \cdots, b'_{NN}$ be two sets of constants given for every N. Let $\bar{b}_N$, $\bar{b}'_N$ be the averages of the two sets of numbers. Then*

$$E_0(\Sigma a_{Ni} b_{NR_i})(\Sigma a_{Ni} b'_{NR_i}) = \Sigma(b_{Ni} - \bar{b}_N)(b'_{Ni} - \bar{b}_N)/N - 1.$$

This is proved by an elementary computation.

The next theorem gives some information about the large-sample power of a rank order test derived from (4.1). It is assumed that $\Sigma b_{Ni} = 0$. This involves no loss of generality since it gives a t_N equivalent to (4.1). It should be recalled that $Z_{N1} \leqq \cdots \leqq Z_{NN}$ are the ordered values of $X_1, \cdots, X_N$, and that by Assumption A, $m/N \to K$ as $N \to \infty$.

THEOREM 4.1.

ASSUMPTIONS.

(a) *Assumptions* A *and* B *hold.*

(b) *Assumptions* (a), (b), (c) *and* (e) *of Theorem* 2.1 *hold.*

(c) $N^{-1} \Sigma b_{Ni} E_0 H(Z_{Ni}) \to c'$, $N^{-1} \Sigma b_{Ni}^2 \to (c'')^2$, *as* $N \to \infty$.

CONCLUSION. *The large-sample power of the rank order test derived from* (4.1) *is* $1 - \Phi(\lambda - \theta N^{1/2} c)$, *where* $c = K^{1/2}(1 - K)^{1/2} c'/c''$.

PROOF. The only thing that needs to be verified is Condition (d) of Theorem 2.1. Let $l_1, \cdots, l_N$ be the value of R on S_j. Then

$$\begin{aligned} E'_{N0} &= \left.\frac{\partial E_\theta t_N}{\partial \theta}\right|_{\theta=0} = \sum_{j=1}^{N!} (\Sigma a_{Ni} b_{Nl_i}) P'_0(S_j) \,/\, N! \\ &= (mn)^{1/2} N^{-1/2} (N - 1)^{-1} \Sigma b_{Ni} \, E_0 H(Z_{Ni}) \end{aligned}$$

by (3.5) and by Lemma 4.1. Also $\mathrm{Var}_0\, t_N = (N - 1)^{-1} \Sigma b_{Ni}^2$ by Lemma 4.1. Hence

$$(E'_{N0})(N \,\mathrm{Var}_0\, t_N)^{-1/2} \to c.$$

REMARKS ON THEOREM 4.1.

(a) If t_N is the locally best rank order statistic (according to Definition 3.5), then

$$c^2 = K(1 - K) \lim_{N\to\infty} N^{-1} \Sigma[E_0 H(Z_{Ni})]^2.$$

Hoeffding has studied limits of this sort in [6]. His results involve restrictions, some of which do not seem appropriate for the problem here. (The main restriction requires the convexity of $H(x)$.) A reasonable conjecture on the basis of Hoeffding's work in [6] is, however, that

$$N^{-1}\Sigma[E_0H(Z_{Ni})]^2 \rightarrow \int_{-\infty}^{\infty} H^2(x)f(x)\,dx, \qquad \text{as } N \rightarrow \infty$$

(f is defined in (3.1)). This conjecture tends to be borne out by the work of Part II of this paper.

(b) Let $(f_1(x, \theta), f_2(x, \theta))$, $(\bar{f}_1(x, \theta), \bar{f}_2(x, \theta))$ be two (possibly different) sets of alternatives. Let $\bar{f}(x) = \bar{f}_1(x, 0) = \bar{f}_2(x, 0)$ and let $\bar{E}_0$, $\bar{Z}_{Ni}$, $\bar{H}$ be defined in exact analogy to E_0, Z_{Ni}, H. If the actual alternative is (f_1, f_2) and the test is derived from $t_N = \Sigma a_{Ni}\bar{E}_0\bar{H}(\bar{Z}_{Ni})$, then

$$c = K^{1/2}(1 - K)^{1/2} \lim_{N\to\infty} N^{1/2} \frac{\Sigma\bar{E}_0\bar{H}(\bar{Z}_{Ni})E_0H(Z_{Ni})}{\{\Sigma[\bar{E}_0\bar{H}(\bar{Z}_{Ni})]^2\}^{1/2}}.$$

Let $F(x) = \int_{-\infty}^{x} f(t)\,dt$ and suppose $x = \rho(t)$, the inverse of $F(x) = t$, exists. Make analogous definitions of $\bar{F}$, $\bar{\rho}$. Since

$$\int_{-\infty}^{\infty} H^2(x)f(x)\,dx = \int_0^1 H^2(\rho(t))\,dt,$$

a possible extension of the conjecture in Remark (a) would be that

$$N^{-1}\Sigma E_0\bar{H}(\bar{Z}_{Ni})E_0H(Z_{Ni}) \rightarrow \int_0^1 \bar{H}(\bar{\rho}(t))H(\rho(t))\,dt$$

as $N \rightarrow \infty$, in which case

$$c = K^{1/2}(1 - K)^{1/2} \frac{\int_0^1 \bar{H}(\bar{\rho}(t))H(\rho(t))\,dt}{\left[\int_0^1 \bar{H}^2(\rho(t))\,dt\right]^{1/2}}.$$

This conjecture also tends to be borne out by the results of Part II of this paper.

(c) If

$$t_N = \Sigma a_{Ni}p(R_i/N), \tag{4.2}$$

where $p(x)$ is a polynomial $(0 \leqq x \leqq 1)$, then it will be shown below that the assumptions of Theorem 4.1 hold under easy conditions. The importance of this approach is the following: Let $r(x)$ be a continuous function on (0, 1) for which $r(i/N) = E_0h(Z_{Ni})$. As r can be approximated by a polynomial $p(x)$ of high degree, it is reasonable that the large-sample power of the test derived from (3.3) should be approximated by the large-sample power of the test derived from (4.2). In Section 5, a heuristic upper bound is given for the large-sample power of the locally best rank order test. In Part II it is shown that this upper bound can be approximated as closely as one pleases using rank order statistics of the form (4.2), where $p(x)$ is of sufficiently high degree.

5. A heuristic upper bound for the large-sample power of a rank order test. By the Neyman Pearson lemma, an optimum statistic for testing H_0 against the alternative of a specific θ is

$$t_N = \sum_{i=1}^{m} \log \frac{f_1(X_i, \theta)}{f_i(X_i, 0)} + \sum_{i=m+1}^{N} \log \frac{f_2(X_i, \theta)}{f_2(X_i, 0)}. \tag{5.1}$$

Of course, (5.1) is in general not a rank order statistic. As a sum of independent random variables, (5.1) will under quite general circumstances be asymptotically normal. Assuming that the conditions of Theorem 2.1 hold, a heuristic derivation of the large-sample power of the test based on (5.1) is now given. (Notice that here t_N depends on θ, which is not the case in Theorem 2.1.)

$$\begin{aligned}(E_\theta t_N - E_0 t_N)/N\theta^2 = \frac{m}{N} \int_{-\infty}^{\infty} \frac{\log f_1(x, \theta) - \log f_1(x, 0)}{\theta} \cdot \frac{f_1(x, \theta) - f_1(x, 0)}{\theta}\, dx \\ + \frac{n}{N} \int_{-\infty}^{\infty} \frac{\log f_2(x, \theta) - \log f_2(x, 0)}{\theta} \cdot \frac{f_2(x, \theta) - f_2(x, 0)}{\theta}\, dx.\end{aligned}$$

Set $\theta = \delta N^{-1/2}$ and assume that the limiting operations involved may be interchanged. Then

$$(E_\theta t_N - E_0 t_N)/N\theta^2 \to K \int_{-\infty}^{\infty} H_1^2(x) f(x)\, dx + (1 - K) \int_{-\infty}^{\infty} H_2^2(x) f(x)\, dx, \quad \text{as } N \to \infty. \tag{5.2}$$

In a similar way one finds that $\mathrm{Var}_0\, t_N/N\theta^2$ approaches the same limit. (The computations use (3.4).) Hence,

$$c^2 = \lim_{N \to \infty} \frac{(E_\theta t_N - E_0 t_N)^2}{N\theta^2 \, \mathrm{Var}_0\, t_N}$$

is equal to the right-hand side of (5.2). The Neyman-Pearson lemma implies that $1 - \Phi(\lambda - \delta c)$ is an upper bound for the large-sample power of the optimum rank order test.

The alternative considered above is that the distribution of $X_1, \cdots, X_N$ is determined by $(f_1(x, \theta), f_2(x, \theta))$. If one considers instead the alternative that the random variables are distributed as $T(X_1), \cdots, T(X_N)$ where T is an increasing function, then there is no difference between the ranks of the X_i and the ranks of the $T(X_i)$; hence, the power of one rank order test against either of these alternatives is the same. Thus, the upper bound found above may be tightened if in the alternative $(f_1(x, \theta), f_2(x, \theta))$, f_1 and f_2 are replaced by the density functions of suitably transformed variables. In particular, suppose that the N random variables are distributed as

$$KF_1(X_i, \theta) + (1 - K)F_2(X_i, \theta) \qquad (i = 1, \cdots, N),$$

where

$$F_j(x, \theta) = \int_{-\infty}^{x} f_j(t, \theta)\, dt, \qquad j = 1, 2.$$

Let $f_1^*(x, \theta)$, $f_2^*(x, \theta)$ be the density functions of the transformed random variables. It is easy to verify that

$$K f_1^*(x, \theta) + (1 - K) f_2^*(x, \theta) = 1. \tag{5.3}$$

Hence, for purposes of obtaining an upper bound for the large-sample power of a rank order test, there is no loss in supposing that (5.3) holds for f_1, f_2 in the earlier argument. An easy calculation shows that under condition (5.3) on f_1, f_2,

$$c^2 = K(1 - K) \int_{-\infty}^{\infty} H^2(x)\, f(x)\, dx.$$

Notice that this ties in with the conjecture of Remark (a), Theorem 4.1.

The development of this section is extremely heuristic and unrigorous. Suitable regularity conditions can no doubt be put down to make everything correct. Since the results of this section will not be used, the matter will be left as it is. In the case of the usual examples of particular density functions, f_1, f_2, the results can often be verified directly.

Part II

6. U statistics. In this section, U statistics are studied in order to obtain information about the related statistics (4.2).

Definition 6.1. Let $u(x_1, \cdots, x_p; x_{p+1}, \cdots, x_{p+q})$ be a function of $p + q$ variables which is symmetric in the first p variables (that is, invariant under all permutations of the labels $1, \cdots, p$) and which is symmetric in the last q variables. Such a u will be called a (p, q) symmetric function.

Let $\alpha_1, \cdots, \alpha_p; \beta_1, \cdots, \beta_q$ be $p + q$ integers subject to the restrictions

$$1 \leqq \alpha_1 < \cdots < \alpha_p \leqq m < m + 1 \leqq \beta_1 < \cdots < \beta_q \leqq N. \tag{6.1}$$

There are, of course, $\binom{m}{p}$ sets of α_i's and $\binom{n}{q}$ sets of β_i's satisfying (6.1).

Definition 6.2. Let

$$U = \binom{m}{p}^{-1} \binom{n}{q}^{-1} \Sigma' u(X_{\alpha_1}, \cdots, X_{\alpha_p}; X_{\beta_1}, \cdots, X_{\beta_q}), \tag{6.2}$$

where Σ' means summation over all $\binom{m}{p}\binom{n}{q}$ choices of the indices, subject to (6.1). Any statistic of the form (6.2) is called a U statistic. This generalizes the terminology of Hoeffding [3], who studied the case $q = 0$.

The first problem is to study the asymptotic normality of U statistics as $N \to \infty$ (p, q fixed). Under suitable regularity conditions, asymptotic normality was established by Hoeffding [3] for the case $q = 0$ and by Lehmann [9] for the case $p = q$. (Lehmann's proof was, for simplicity, done for $m = n$.) What follows admits the possibility that u may depend on N and that θ may depend on N. The methods are essentially those of Hoeffding in [3].

Let u, u_N be two (p, q) symmetric functions. Assumption α, below, will prescribe a certain sense in which $u_N \to u$ as $N \to \infty$. Let U_N denote the U statistic determined by u_N as in Definition 6.2. Let $\theta = \theta(N)$ be a function of N such that $\theta(N) \to \theta_0$ as $N \to \infty$. (This includes the possibility that $\theta(N) = \theta_0$ for all N.)

DEFINITION 6.3.

$$\begin{aligned} u_{N,r,s} &= u_{N,r,s}(x_{\alpha_1}, \cdots, x_{\alpha_r}; x_{\beta_1}, \cdots, x_{\beta_s}) \\ &= E_\theta u_N(x_{\alpha_1}, \cdots, x_{\alpha_r}, X_{\alpha_{r+1}}, \cdots, X_{\alpha_p}; x_{\beta_1}, \cdots, x_{\beta_s}, X_{\beta_{s+1}}, \cdots, X_{\beta_q}) \\ M_{N,\theta} &= E_\theta u_{N,r,s}(X_{\alpha_1}, \cdots, X_{\alpha_r}; X_{\beta_1}, \cdots, X_{\beta_s}) \quad \text{(this also equals } E_\theta u_N) \\ \Psi_{N,\theta} &= u_N - M_{N,\theta}, \\ \Psi_{N,\theta,r,s} &= u_{N,r,s} - M_{N,\theta}, \\ \rho_{N,\theta} &= E_\theta \Psi^2_{N,\theta}; \\ \rho_{N,\theta,r,s} &= E_\theta \Psi^2_{N,\theta,r,s}, \qquad r = 0, 1, \cdots, p; s = 0, 1, \cdots, q. \end{aligned}$$

By deleting the subscript N throughout, analogous definitions are made of $u_{r,s}$, M_θ, Ψ_θ, $\Psi_{\theta,r,s}$, ρ_θ, $\rho_{\theta,r,s}$.

Since $X_1, \cdots, X_m$ are identically distributed and $X_{m+1}, \cdots, X_N$ are identically distributed, the values of $M_{N,\theta}$, M_θ, $\rho_{N,\theta}$, ρ_θ, $\rho_{N,\theta,r,s}$, $\rho_{\theta,r,s}$ do not depend on the labels $\alpha_1, \cdots, \alpha_p; \beta_1, \cdots, B_q$. It is understood that $\Psi_{N,\theta,p,q} = \Psi_{N,\theta}$, $\Psi_{N,\theta,0,0} = 0$. (A similar convention holds for N removed.)

ASSUMPTION α.

$$M_{N,\theta} \to M_{\theta_0}, \qquad \rho_{N,\theta,r,s} \to \rho_{\theta_0,r,s}$$

as $N \to \infty$ for $r = 0, 1, \cdots, p$; $s = 0, 1, \cdots, q$. (Recall that $\theta = \theta(N)$.)

Let

$$Y_N = p(Km)^{-1/2} \sum_{i=1}^{m} \Psi_{N,\theta,1,0}(X_i) + q((1 - K)m)^{-1/2} \sum_{i=m+1}^{N} \Psi_{N,\theta,0,1}(X_i).$$

Evidently $\text{Var}_\theta\, Y_N \to (p^2/H)\rho_{\theta_0,0,1} + (q^2/1 - K)\rho_{\theta_0,0,1} = L$ as $N \to \infty$, by Assumption α.

ASSUMPTION β.

Y_N is asymptotically normal $(0, L^{1/2})$ as $N \to \infty$ and $m/N \to K$, $(0 < K < 1)$.

Since Y_N is a sum of independent random variables, Assumption β is not very restrictive. It will be satisfied, for instance, if u_N converges in probability to u as $N \to \infty$ and $\max(\rho_{\theta_0 0,1}; \rho_{\theta_0,1,0}) > 0$. (See [2], Theorem 3, p. 101.)

LEMMA 6.1. *Assumption α implies that N* $\text{Var}_\theta\, U_N \to L$, *$N$* $\text{Var}_\theta\, U \to L$, *as $N \to \infty$, $m/N \to K$.*

PROOF.

$$N \operatorname{Var}_\theta U_N = N \binom{m}{p}^{-2} \binom{n}{q}^{-2} \sum_{s=0}^{m} \sum_{r=0}^{m} \Sigma^{(r,s)} E_\theta[\Psi_{N,\theta}(X_{\alpha_1}, \cdots, X_{\alpha_p}; X_{\beta_1}, \cdots, X_{\beta_q})]$$
$$\cdot [\Psi_{N,\theta}(X_{\alpha'_1}, \cdots, X_{\alpha'_p}; X_{\beta'_1}, \cdots, X_{\beta'_q})],$$

where $\Sigma^{(r,s)}$ means summation over those subscripts where exactly r equations $\alpha_i = \alpha'_j$ are satisfied and exactly s equations $\beta_i = \beta'_j$ are satisfied. Each term in $\Sigma^{(r,s)}$ is equal to $\rho_{N,\theta,r,s}$ and the number of such terms is $\binom{p}{r}\binom{m-p}{p-r}\binom{m}{p}\binom{q}{s}$ $\binom{n-q}{q-s}\binom{n}{q}$. A similar expression holds for U. The required result follows on taking limits.

THEOREM 6.1. *Assumptions α and β imply that $N^{1/2}U_N$ is asymptotically normal, $(0, L^{1/2})$, as $N \to \infty$.*

PROOF. It is sufficient to show that $E_\theta(N^{1/2}U_N - Y_N)^2 \to 0$ as $N \to \infty$. Let $D_N = N \operatorname{Var}_\theta U_N + \operatorname{Var}_0 Y_N - 2N^{1/2}E_\theta U_N Y_N$. Consider the facts that

$$E_\theta \Psi_{N,\theta,1,0}(X_i)\Psi_{N,\theta}(X_{\alpha_1}, \cdots, X_{\alpha_p}; X_{\beta_1}, \cdots, X_{\beta_q}) = \begin{cases} \rho_{N,\theta,1,0} \text{ if } i \text{ is one of } \alpha_1, \cdots, \alpha_p, \\ 0 \qquad \text{otherwise;} \end{cases}$$

and

$$\Sigma_\theta \Psi_{N,\theta,0,1}(X_i)\Psi_{N,\theta}(X_{\alpha_1}, \cdots, X_{\alpha_p}; X_{\beta_1}, \cdots, X_{\beta_q}) = \begin{cases} \rho_{N,\theta,0,1} \text{ if } i \text{ is one of } \beta_1, \cdots, \beta_q, \\ 0 \qquad \text{otherwise.} \end{cases}$$

Hence,

$$N^{1/2}E_\theta U_N Y_N = p^2(N/mK)^{1/2}\rho_{N,\theta,1,0} + q^2(N/n(1-K))^{1/2}\rho_{N,\theta,0,1}$$
$$\to (p^2/K)\rho_{\theta_0,1,0} + (q^2/1-K)\rho_{\theta_0,0,1},$$

and $D_N \to 0$ as $N \to \infty$.

7. L_h statistics.

DEFINITION 7.1. Let $p(t) = b_1 t + \cdots + b_h t^h$ $(0 \leqq t \leqq 1)$ be a polynomial with real coefficients and let

$$t_N = \Sigma\, a_{Ni}p(R_i/N). \tag{7.1}$$

Such a t_N is called an L_h statistic. (Notice that including a constant term in $p(t)$ gives an equivalent t_N.) The test derived from t_N is called an L_h test.

The main purpose of this section is to show that an L_h statistic is a U statistic and is asymptotically normally distributed. Define

$$c(y) = \begin{cases} 1 & \text{if } y > 0, \\ 0 & \text{if } y \leqq 0. \end{cases}$$

Then with probability 1, $R_i = 1 + \sum_{j=1}^{N} c(X_i - X_j)$ and

$$\begin{aligned} R_i^p = \Sigma^{(p)} B_{pp} c(X_i - X_{j_1}) \cdots c(X_i - X_{j_p}) + \cdots \\ + \Sigma^{(t)} B_{pt} c(X_i - X_{j_1}) \cdots c(X_i - X_{j_t}) + \cdots \\ + \Sigma^{(1)} B_{p1} c(X_i - X_{j_1}) \\ + 1, \end{aligned}$$

where $\Sigma^{(t)}$ is summation over all subscripts satisfying $1 \leqq j_1 < j_2 < \cdots < j_t \leqq N$ $(t = 1, \cdots, p)$. B_{pt} depends on p and on t but *not on N and not on i*. Since $c(X_i - X_i) = 0$, it can be assumed that none of $j_1, \cdots, j_t$ is equal to i. The expansion of R_i^p makes use of the fact that $c^j(y) = c(y)$ $(j = 1, 2, \cdots)$.

Consider the $t + 1$ integers $i, j_1, \cdots, j_t$. Let $\alpha_1, \cdots, \alpha_s$ be those that are a subset of $1, \cdots, m$ and let $\beta_1, \cdots, \beta_{t+1-s}$ be those that are a subset of $m + 1, \cdots, N$.

DEFINITION 7.2. Define $u_{s,t+1-s}^{(p)}$ by

$$\begin{aligned} N^p (b_p B_{pt})^{-1} u_{s,t+1-s}^{(p)}(X_{\alpha_1}, \cdots, X_{\alpha_s}; X_{\beta_1}, \cdots, X_{\beta_{t+1-s}}) \\ = a_{Ni} c(X_i - X_{j_1}) c(X_i - X_{j_2}) \cdots c(X_i - X_{j_t}) \\ + a_{Nj_1} c(X_{j_1} - X_i) c(X_{j_1} - X_{j_2}) \cdots c(X_{j_1} - X_{j_t}) \\ \vdots \\ + a_{Nj_t} c(X_{j_t} - X_{j_1}) c(X_{j_t} - X_{j_2}) \cdots c(X_{j_t} - X_i), \end{aligned}$$

$$u_{0,1}^{(p)} = u_{0,1}^{(p)} = 0,$$

$$u_{0,0}^{(p)} = 1, \quad 0 \leqq s \leqq t + 1, \quad 0 \leqq t \leqq p.$$

Notice that $u_{s,t+1-s}^p$ is an $(s, t + 1 - s)$ symmetric function (see Definition 6.1).

Consider the $t + 1$ integers α_i, β_i satisfying

$$(7.2) \qquad 1 \leqq \alpha_1 < \cdots < \alpha_s \leqq m; \quad m + 1 \leqq \beta_1 < \cdots < \beta_{t+1-s} \leqq N,$$

and the $2(h + 1)$ integers α_i', β_i' satisfying

$$(7.3) \qquad 1 \leqq \alpha_1' < \cdots < \alpha_{h+1}' \leqq m; \quad m + 1 \leqq \beta_1' < \cdots < \beta_{h+1}' \leqq N.$$

The first set is called an associate of the second set if every α_i is some α_j' and every β_i is some β_j'. Thus, any fixed set satisfying (7.2) is an associate of

$\binom{m - s}{h + 1 - s}\binom{n - t - 1 + s}{h - t + s}$ different sets satisfying (7.3).

DEFINITION 7.3. Define u_N by

$$N^{1/2}\binom{m}{h+1}^{-1}\binom{n}{h+1}^{-1} u_N(X_{\alpha_1}, \cdots, X_{\alpha_{h+1}}; X_{\beta_1}, \cdots, X_{\beta_{h+1}}) = \sum_{p=1}^{h}\sum_{t=0}^{p}\sum_{s=0}^{t+1}\binom{m-s}{h+1-s}^{-1}\binom{n-t-1+s}{h-t+s}^{-1}\Sigma'' u_{s,t+1-s}^{(p)},$$

where Σ'' is summation over all the $u_{s,t+1-s}^{(p)}$ terms (for fixed s, t, p) whose indices are associates of $\alpha_1, \cdots, \alpha_{h+1}; \beta_1, \cdots, \beta_{h+1}$.

DEFINITION 7.4.

$$U_N = \binom{m}{h+1}^{-1}\binom{n}{h+1}^{-1}\Sigma' u_N,$$

where Σ' is summation over all sets of the $2(h+1)$ indices satisfying (7.3).

The following theorem now follows directly from the constructions made above.

THEOREM 7.1.

$$t_N = N^{1/2}U_N, \quad \textit{where } t_N \textit{ is given by } (7.1).$$

LEMMA 7.1. *u_N has the following properties:*

(a) *The $2(h+1)$-dimensional space over which u_N is defined is partitioned into a finite number (which depends on h but not on N) of disjoint sets on each of which u_N assumes a constant value.*

(b) *$\lim_{N\to\infty} u_N = u$ exists as $N \to \infty$, $m/n \to K$ $(0 < K < 1$*

(c) *u_N is a $(h+1, h+1)$ symmetric function.*

(d) *U_N is a U statistic.*

PROOF. (a), (b), (c), (d) all follow from the constructions made in definitions 7.1 and 7.2. In particular,

$$u = \lim_{N\to\infty} u_N = \lim_{N\to\infty}\sum_{p=1}^{h}\sum_{s=0}^{p+1}\binom{m}{h+1}\binom{n}{h+1}\binom{m-s}{h+1-s}^{-1}\binom{n-p-1+s}{h-p+s}^{-1} N^{-p-1/2}\,\Sigma''\, u_{s,p+1-s}^{(p)}.$$

LEMMA 7.2. *u_N, u satisfy Assumption α.*

This follows from (a) and (b) of Lemma 7.1.

As in Section 6, it is still supposed that $\theta = \theta(N) \to \theta_0$ as $N \to \infty$.

LEMMA 7.3.

$$\text{(a)}\quad \operatorname{Var}_0 t_N \to \sum_{i,j=1}^{h} b_i b_j ij(i+j+1)^{-1}(i+1)^{-1}(j+1)^{-1} = \sigma^2,$$

$$\text{(b)}\ \operatorname{Var}_\theta t_N \to \sigma^2 \qquad \textit{as } N \to \infty.$$

PROOF OF (a). By Lemma 4.1,

$$\operatorname{Var}_0 t_N = \Sigma[p(i/N)]^2/(N-1) - [\Sigma\, p(i/N)]^2 / N(N-1).$$

The lemma follows from the fact that

$$\sum_{i=1}^{N} i^r/N^{r+1} \to 1/r + 1 \text{ as } N \to \infty \qquad (r = 1, 2, \cdots).$$

PROOF OF (b). Because Assumption α is satisfied, $\mathrm{Var}_\theta\, N^{1/2} U_N$ and $\mathrm{Var}_0\, N^{1/2} U_N$ approach the same limit as $N \to \infty$. The proof follows from the fact that $t_N = N^{1/2} U_N$.

DEFINITION 7.5. Let $P_i(x) = p_{i0} + p_{i1}x + \cdots + p_{ii}x^i$ $(i = 0, 1, 2, \cdots)$ be the system of orthogonal polynomials associated with the weight function x^2 over $(0, 1)$; that is,

$$\int_0^1 P_i(x)P_j(x)x^2\,dx = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases} \tag{7.4}$$

In other words, $\{P_i(x)x\}$ is an orthonormal system. It is well known that this orthonormal system is complete with respect to the Lebesgue square integrable functions on (0, 1). For the very basic Hilbert space information needed in the remainder of this paper, see [15].

DEFINITION 7.6.

$$P = \begin{pmatrix} p_{00} & 0 & 0 & \cdots & 0 \\ p_{10} & p_{11} & 0 & \cdots & 0 \\ \vdots & & & & \\ p_{h-1,0} & p_{h-1,1} & 0 & \cdots & p_{h-1,h-1} \end{pmatrix},$$

$$A = \begin{pmatrix} \frac{1}{3} & \frac{1}{4} & \cdots & 1/(h+2) \\ \frac{1}{4} & \frac{1}{5} & \cdots & 1/(h+3) \\ \vdots & \vdots & & \\ 1/(h+2) & 1/(h+3) & \cdots & 1/(2h+1) \end{pmatrix},$$

$$D = \begin{pmatrix} \frac{1}{2} & & & 0 \\ & \frac{2}{3} & & \\ & & \ddots & \\ 0 & & & h/(h+1) \end{pmatrix},$$

$$s_h = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \vdots \\ 1/h + 1 \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_h \end{pmatrix}.$$

Any of the above matrix or vector symbols with a prime means a transpose.

LEMMA 7.4. *A is a positive definite matrix.*

PROOF. First notice that $1/i + j + 1$, the (i, j)th element of A is equal to $\int_0^1 x^{(i-1)+(j-1)}x^2\,dx$. Hence, the orthogonality conditions (7.4) mean that

$$PAP' = I \text{ (the unit matrix)}.$$

Hence, P is nonsingular and

$$A = P^{-1}P'^{-1},$$

which proves the positive definiteness of A. (The idea of this lemma was suggested by [1].)

THEOREM 7.2. *Suppose* $\max(|b_1|, \cdots, |b_h|) > 0$ *and* $m/N \to K$ $(0 < K < 1)$ *as* $N \to \infty$.

Then

$$P_\theta\{(t_N - E_\theta t_N)(\mathrm{Var}_\theta\, t_N)^{-1/2} < s\} \to \Phi(s)$$

as $N \to \infty$. (*Recall that* $\theta = \theta(N)$.)

PROOF. By Lemma 7.3, $\mathrm{Var}_\theta\, t_N \to \sigma^2 = b'DADb$. Hence, $\sigma^2 > 0$, since DAD is positive definite by Lemma 7.4. By (a) and (b) of Lemma 7.1, u_N converges in probability to u as $N \to \infty$; hence assumption β is satisfied. (See the remark following statement of Assumption β.) The required result follows from Theorem 6.1.

8. The large-sample power of L_h tests.

ASSUMPTION C.

$$0 < \int_{-\infty}^{\infty} H^2(x)\, f(x)\, dx < \infty \qquad \text{(Lebesgue integral).}$$

LEMMA 8.1. *Assumptions* A, B, C *imply that*

$$N^{-\frac{1}{2}}\, \Sigma a_{Ni} \frac{\partial E_\theta (R_i/N)^j}{\partial \theta}\bigg|_{\theta=0} \to l_j \qquad \text{as } N \to \infty, \tag{8.1}$$

where

$$l_j = K^{1/2}(1 - K)^{1/2} \int_{-\infty}^{\infty} H(x) F^j(x)\, f(x)\, dx, \quad \textit{and} \quad F(x) = \int_{-\infty}^{x} f(t)\, dt.$$

PROOF. Let P_{it} be the probability that the random variable X_i has rank t, $(i, t = 1, \cdots, N)$. Then

$$P_{it} = \Sigma' \int \prod_{i=1}^{m} f_1(x_i, \theta) \prod_{i=m+1}^{N} f_2(x_i, \theta) \prod_{i=1}^{N} dx_i,$$

where $\int$ is over the set where $x_{j_1} < \cdots < x_{j_{t-1}} < x_i < x_{j_{t+1}} < \cdots < x_{j_N}$ and Σ' is over all permutations $j_1, \cdots, j_{t-1}, j_{t+1}, \cdots, j_N$ of the $N - 1$ numbers $1, 2, \cdots, t - 1, t + 1, \cdots, N$.

An elementary but tedious computation shows that

$$\frac{\partial P_{it}}{\partial \theta}\bigg|_{\theta=0} = \frac{s_{it}}{N(N-1)} \binom{N}{t} \int_{-\infty}^{\infty} H(x) F(x)^{t-1} (1 - F(x))^{N-t} f(x)\, dx,$$

where

$$s_i = \begin{cases} n, & i = 1, \cdots, m \\ -m, & i = m + 1, \cdots, N. \end{cases}$$

Let $y^{(j)}$ denote the factorial $y(y-1)\cdots(y-j+1)$. Then

$$\left.\frac{\partial E_\theta \dfrac{(R_i-1)^{(j)}}{N^j}}{\partial\theta}\right|_{\theta=0} = \sum_{t=1}^{N} \frac{(t-1)^{(j)}}{N^j} \left.\frac{\partial P_{it}}{\partial\theta}\right|_{\theta=0}$$

$$= s_i(N-2)^{(j-1)}N^{-j}\int_{-\infty}^{\infty} H(x)F^j(x)\, f(x)\, dx$$

by another routine computation. The result of the theorem follows from the evaluation of the limit of

$$N^{-\frac{1}{2}}\sum_{i=1}^{N} a_{Ni} \left.\frac{\partial E_\theta(R_i-1)^{(j)}/N^j}{\partial\theta}\right|_{\theta=0}, \qquad \text{as } N\to\infty,$$

and from the fact that this limit must be the same as the limit of the left-hand side of (8.1). (Notice that by Assumption C and by the Schwartz inequality the integrals displayed above all exist.)

Definition 8.1. Let

$$l = \begin{pmatrix} l_1 \\ \vdots \\ l_h \end{pmatrix},$$

and define B by $B'B = DAD$. (This factorization makes sense by Lemma 7.4. Also, by the same lemma, B is nonsingular.)

The next theorem shows that the theorems on large-sample power can be applied to L_h tests.

Theorem 8.1.

Assumptions:

(a) $\max\{|b_1|, \cdots, |b_h|\} > 0$,

(b) $\lambda_N \to \lambda$, *as* $N\to\infty$, *where* $\Phi(\lambda) = 1-\alpha$,

(c) *Assumptions* A, B, *and* C *hold.*

Conclusion. *The large-sample power of the test derived from* (7.1) *is* $1-\Phi(\lambda-\delta c)$, *where*

$$c = (b'l)(b'B'Bb)^{-1/2} = (b'l)(b'DADb)^{-1/2}. \tag{8.2}$$

Proof. The proof will follow by verifying the assumptions of Theorem 2.1:

Condition (a) of Theorem 2.1 holds by Theorem 7.2.

Conditions (b) and (c) of Theorem 2.1 hold by Theorem 7.1 and Lemma 7.1.

Condition (d) of Theorem 2.1 together with the explicit value of c follow from Lemmas 7.3 and 8.1.

(Notice that it is no loss to suppose $c \geqq 0$, since otherwise the statistic $-t_N$ could be used.)

Corollary. *The maximum large-sample power obtainable with a* p_h*-rank order statistic is* $1-\Phi(\lambda-\delta\bar{c})$, *where*

$$\bar{c} = [l(DAD)^{-1}l]^{1/2}.$$

PROOF. $c = (b'B'B'^{-1}l)(b'B'Bb)^{-1/2}$. By Schwartz's inequality,

$$c^2 \leqq (b'B'Bb)(l'B^{-1}B'^{-1}l)(b'B'Bb)^{-1} = (l'B^{-1}B'^{-1}l).$$

This maximum is achieved by choosing the coefficients of the polynomial $p(t)$ as

$$b = (DAD)^{-1}l.$$

DEFINITION 8.2. The L_h statistic which maximizes the large-sample power (for fixed h) is called the locally *best* L_h statistic. The test derived from it is called the locally best L_h test.

ASSUMPTION D. $F(x) = \int_{-\infty}^{x} f(t)\, dt$ is an increasing function of x whenever $F(x)$ is not zero or one.

Let $F(x) = t$. Assumption D implies that this function has a continuous inverse, $x = \rho(t)$ $(0 < t < 1)$.

DEFINITION 8.3.

$$g(t) = K^{1/2}(1 - K)^{1/2} H(\rho(t)),$$

$$G(t) = \int_0^t g(x)\, dx \qquad (0 < t < 1).$$

Since

$$\int_0^1 H^2(\rho(t))\, dt = \int_{-\infty}^{\infty} H^2(x) f(x)\, dx.$$

Assumption C implies that $g(t)$ is Lebesgue square integrable on (0, 1). What follows also requires this to be true for $G(t) / t$ and Assumption E, below, is a convenient way of insuring this.

ASSUMPTION E. $G^2(t) / t \to 0$ as $t \to c$.

LEMMA 8.2. *Assumptions* C *and* E *imply that* $G(t) / t$ *is Lebesgue square integrable on* (0, 1).

PROOF. $G(t)$ is continuous and differentiable at any point of $(\epsilon, 1)$ $(0 < \epsilon < 1)$, hence $\int_\epsilon^1 G^2(t) / t^2\, dt$ exists. Differentiating by parts and using the Schwartz inequality implies that

$$\int_\epsilon^1 G^2(t) / t^2\, dt = -G^2(\epsilon) / \epsilon + 2 \int_\epsilon^1 G(t)t^{-1}g(t)\, dt$$

$$\leqq -G^2(\epsilon) / \epsilon + 2 \left[\int_\epsilon^1 G^2(t)t^{-2}\, dt\right]^{1/2} \left[\int_\epsilon^1 g^2(t)\, dt\right]^{1/2}.$$

The left-hand side of Assumption C implies that $[\int_\epsilon^1 G^2(t)t^{-2}\, dt]^{1/2} > 0$; hence dividing by this quantity gives the desired result.

LEMMA 8.3. *Let*

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_h \end{pmatrix}.$$

Let $s(t)$ be a function on (0, 1) *such that*

$$v_j = \int_0^1 t^j s(t)\, dt \qquad (i = 1, \cdots, h).$$

Then the vector

$$Pv = \begin{pmatrix} v_1 p_{00} \\ v_1 p_{10} + v_2 p_{11} \\ \vdots \\ v_1 p_{h-1,0} + v_2 p_{h-1,1} + \cdots + v_h p_{h-1,h-1} \end{pmatrix}$$

is equal to

$$\begin{bmatrix} \int_0^1 P_0(t) ts(t)\, dt \\ \vdots \\ \int_0^1 P_{h-1}(t) ts(t)\, dt \end{bmatrix}.$$

In other words, the jth element of Pv is the jth Fourier coefficient of $s(t)$ with respect to the orthonormal system $\{P_i(t)t\}$. The proof follows simply from the fact that

$$\int_0^1 P_j(t) ts(t)\, dt = \sum_{i=0}^{j} \int_0^1 p_{ij} t^{j+1}\, s(t)\, dt = \sum_{i=0}^{j} p_{ji} v_{i+1}.$$

Recall that the corollary to Theorem 8.1 says that the maximum large-sample power obtainable with a p_h-rank order statistic is

$$1 - \Phi(\lambda - \delta\bar{c}), \qquad \bar{c} = [l'(DAD)^{-1}l]^{1/2}.$$

The fact that $\bar{c}$ depends on h is now stressed by writing $\bar{c}_h$.

Lemma 8.4. *Assumptions* C, D, *and* E *imply that*

$$\bar{c}_h^2 \to \int_0^1 g^2(t)\, dt = K(1 - K) \int_{-\infty}^{\infty} H^2(x) f(x)\, dx, \qquad \text{as } h \to \infty.$$

Proof. Using integration by parts,

$$\int_0^1 [g(t) - G(t) / t] t^j\, dt = \int_0^1 g(t) t^j\, dt - \int_0^1 G(t) t^{j-1}\, dt$$

$$= j^{-1}(j + 1) \int_0^1 g(t) t^j\, dt = j^{-1}(j + 1) l_j.$$

By Lemma 8.3, the vector $PD^{-1}l$ is the vector of the first h Fourier coefficients of $g(t) = G(t) / t$. Hence, by Parseval's Theorem (see [15])

$$l'D^{-1}P'PD^{-1}l \to \int_0^1 [g(t) - G(t) / t]^2\, dt, \qquad \text{as } h \to \infty.$$

Integration by parts gives $\int_0^1 g^2(t)\,dt$ for this last integral, which proves the lemma.

It should be noticed that Lemma 8.4 implies that Assumption (a) of Theorem 8.1 is satisfied when $b = (DAD)^{-1}l$.

The results of Theorem 8.1 and Lemma 8.4 can be summarized in the following Theorem 8.2, which together with Theorem 8.3, might be considered the main results of Part II of this paper.

THEOREM 8.2. *Suppose Assumptions* A *through* E *hold and* $\lambda_N \to \lambda$ *as* $N \to \infty$.

Let $\varphi_{N,h}^{(\theta)}$ be the power function of the best L_h test. Choose any positive ϵ and a positive number δ. Then there is an $N' = N'(\epsilon, \delta)$ and an $h' = h'(\epsilon, \delta)$ such that

$$|\varphi_{Nh'}^{(\theta)}(\theta) - (1 - \Phi(\lambda - \theta N^{1/2}c)| < \epsilon$$

for $\theta N^{1/2} = \delta$, and all $N \geqq N'$.

$$c = \left(\int_0^1 g^2(t)\,dt\right)^{1/2}$$

REMARKS ON THEOREM 8.2.

(a) Roughly, this theorem says that the large-sample power of the best L_h test approaches $1 - \Phi(\lambda - \delta c)$ as h, the order of $p(t)$, is made large.

(b) This theorem lends credence to the conjecture made in remark (a) of Theorem 4.1. It is reasonable that the "polynomial approximation" to the best rank order test should behave almost like the best test itself, but this does not constitute a proof.

Let $(f_1(x, \theta), f_2(x, \theta))$, $(\bar{f}_1(x, \theta), \bar{f}_2(x, \theta))$ be two (possibly different) sets of alternatives. Let $\bar{g}(t)$, $\bar{G}(t)$, $\bar{H}(x)$, $\bar{l}$, $\bar{p}(t)$ be defined analogously to $g(t)$, $G(t)$, $H(x)$, l, $p(t)$.

THEOREM 8.3. *Suppose both sets of alternatives* (f_1, f_2), $(\bar{f}_1, \bar{f}_2)$ *satisfy Assumptions* A *through* E.

Let $\bar{l}_N$ *be the locally best* L_h *statistic against the alternative* $(\bar{f}_1, \bar{f}_2)$. *Then, if the true alternative is* (f_1, f_2),

(a) *the large-sample power of the test derived from* $\bar{l}_N$ *is* $1 - \Phi(\lambda - \delta c_h)$, *where*

$$c_h = (\bar{l}'(DAD)^{-1}l)(\bar{l}'(DAD)^{-1}\bar{l})^{-1/2}.$$

(b)

$$c_h \to \frac{\int_0^1 g(t)\bar{g}(t)\,dt}{\left[\int_0^1 \bar{g}^2(t)\,dt\right]^{1/2}} = c, \qquad \text{as } h \to \infty.$$

Hence (in the sense described in Theorem 8.2*), the large-sample power of the test derived from* $\bar{l}_N$ *approaches* $1 - \Phi(\lambda - \delta c)$ *as the order of* $p(t)$ *is made large.*

PROOF. Part (a) follows from Theorem 8.1 by setting the b vector in (8.2) equal to

$$\bar{b} = (DAD)^{-1}\bar{l}.$$

Part (b) follows from considerations exactly analogous to those used in proving Lemma 8.4. The vectors $PD^{-1}l$ and $PD^{-1}\bar{l}$ are the vectors of the first h Fourier coefficients of $g(t) - G(t) / t$ and $\bar{g}(t) - \bar{G}(t) / t$, respectively. Hence, by Parseval's theorem,

$$\bar{l}'D^{-1}P'PD^{-1}l \rightarrow \int_0^1 [g(t) - G(t) / t][\bar{g}(t) - \bar{G}(t) / t]\, dt$$

$$= \int_0^1 g(t)\bar{g}(t)\, dt, \qquad \text{as } h \rightarrow \infty.$$

COROLLARY. *If the locally best L statistic against* $(\bar{f}_1, \bar{f}_2)$ *is an* L_h *statistic for some* $h = h'$, *then*

$$c_h = c, \qquad \text{for all } h \geqq h'.$$

PROOF. This is because the locally best L_h statistic for all $h \geqq h'$ must be the locally best $L_{h'}$ statistic.

9. Applications and examples.

I. *Location-parameter alternatives.* Let $f_i(x, \theta) = f(x + m_i(\theta))$ $(i = 1, 2)$, where $f(x)$ is a density function not depending on θ; $m_1(0) = m_2(0) = 0$, $m_1'(0) - m_2'(0) \neq 0$. Let $\rho(t)$ be the inverse of $F(x) = t$ where $F(x) = \int_{-\infty}^{x} f(t)\, dt$.

Evidently, the carrier of $f(x)$ must be $(-\infty, \infty)$ if Assumption B is to be satisfied. (This, of course, is not sufficient for Assumption B.) It is easy to verify that

$$g(t) = Df'(\rho(t)) / f(\rho(t)),$$

$$G(t) = Df(\rho(t)), \qquad \text{where } D = (m_1'(0) - m_2'(0))K^{1/2}(1 - K)^{1/2},$$

and that Conditions C through E hold if

$$\int_{-\infty}^{\infty} (f'(x))^2 / f(x)\, dx < \infty, \tag{9.1}$$

and $f'(x) \rightarrow 0$ as $x \rightarrow -\infty$. (L'Hospital's rule is used, together with the fact that $f(-\infty)$, $f(\infty)$ must equal zero. It is easy to verify (9.1) for the normal density function, for instance. For the usual pathological example, the Cauchy distribution, these conditions can be shown to hold also.)

One can make quite similar remarks about scale-parameter alternatives, where

$$f_i(x, \theta) = \frac{1}{\sigma_i(\theta)} f\left(\frac{x}{\sigma_i(\theta)}\right), \qquad \text{and } \sigma_i(0) = 1, \qquad i = 1, 2. \tag{9.2}$$

II. *Normal alternatives.* Suppose $f_1(x, \theta), f_2(x, \theta)$ are normal density functions, $(m_1(\theta), 1)$, $(m_2(\theta), 1)$ respectively, where $m_1 - m_2 = \theta$. Then an easy calculation shows that $\int_0^1 g^2(t)\,dt = K(1 - K)$. For $\theta > 0$, there is a uniformly most powerful similar test based on $\sum_{i=1}^{m} X_i / m - \sum_{i=m+1}^{N} X_i / n = T_N$. As

$$E'_{N,0}(N \operatorname{Var}_0 T_N)^{-1/2} \to K^{1/2}(1 - K)^{1/2}$$

as $N \to \infty$ and $\theta = \delta N^{-1/2}$, it follows that the large-sample power of the locally best L_h test is, for large h, arbitrarily close to the large-sample power of the test based on T_N.

In an exactly analogous way one can treat the scale-parameter alternative with normal density functions. In that case, also, the locally best rank order test has the same large-sample power as the test based on the F statistic.

III. *The asymptotic efficiency of the Wilcoxon-Mann-Whitney statistic.*[2]

Let $\bar{f}_1(x, \theta) = \bar{f}(x)$, the uniform density on (0, 1), and let $\bar{f}_2(x, \theta) = 2(\theta)\bar{F}(x)\bar{f}(x) + (1 - \theta)\bar{f}(x)$. By Theorem 3.2, the best rank order statistic against this alternative is equivalent to $\Sigma\, a_{Ni}R_i$, which is equivalent to the Wilcoxon-Mann-Whitney statistic [11]. This must then be the best L_h statistic for all $h \geqq 1$. Let f_1, f_2, f be as described in Example I on location-parameter alternatives. We consider the alternative given by density functions f_1, f_2. Let t_N be the locally best rank order statistic against that alternative. By Theorems 8.2 and 8.3 (Corollary) and an elementary calculation, the asymptotic efficiency of $\Sigma\, a_{Ni}R_i$ relative to t_N when the true alternative is f_1, f_2 is the square of

$$\frac{\int_{-\infty}^{\infty} (1 - 2\,F(x))\, f'(x)\, dx}{\left[\int_0^1 (2t - 1)^2\, dt\right]^{1/2} \left[\int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\, dx\right]^{1/2}} = \frac{2\sqrt{3} \int_{-\infty}^{\infty} f^2(x)\, dx}{\left[\int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\, dx\right]^{1/2}}. \tag{9.3}$$

The role of $(\bar{f}_1, \bar{f}_2)$ can be dropped. The number (9.3) is the asymptotic efficiency of $\Sigma\, a_{Ni}R_i$ relative to t_N where the true alternative is (f_1, f_2).

Letting $(\bar{f}_1, \bar{f}_2)$ be normal densities $(m_1(\theta), 1)(m_2, (\theta), 1)$ where $m_1(\theta) - m_2(\theta) = \theta$, (9.3) turns out to be $\sqrt{3/\pi}$, hence the asymptotic efficiency is $3/\pi$. By (II), above, this is also the asymptotic efficiency of $\Sigma\, a_{Ni}R_i$ relative to the t statistic. This result was apparently first given by Pitman in [14]. (See also [12].)

IV. *A rank test for dispersion.* In [12] Mood suggested a rank order statistic (against a dispersion alternative) equivalent to $\Sigma\, a_{Ni}[R_i^2 - (N + 1)R_i]$. This statistic is asymptotically equivalent to

$$\Sigma\, a_{Ni}[(R_i / N)^2 - R_i / N]. \tag{9.4}$$

(See Remark (b)), Theorem 2.1.)

[2] See Remark (c) to Theorem 2.1 for definition of efficiency.

Let $\bar{f}_1(x, \theta) = (1 - \theta)6[\bar{F}(x) - \bar{F}^2(x)]\bar{f}(x) + \theta\bar{f}(x)$,

$$\bar{f}_2(x, \theta) = \bar{f}(x), \text{ the uniform density on } (0, 1).$$

By Theorem 3.2 the best rank order statistic against $(\bar{f}_1, \bar{f}_2)$ is asymptotically equivalent to (9.4). Let t_N be the best rank order statistic against the alternative (f_1, f_2) described by (9.2). Theorems 8.2, 8.3 (Corollary), and an elementary calculation show that the asymptotic efficiency of (9.4) relative to t_N when (f_1, f_2) is the true alternative is the square of

$$\frac{\int_{-\infty}^{\infty} 6(F^2(x) - F(x) + 1)(xf'(x) + f(x))\,dx}{\left(\int_0^1 [6(t^2 - t) + 1]^2\,dt\right)^{1/2}\left(\int_{-\infty}^{\infty} [xf'(x) + f(x)]^2/f(x)\,dx\right)^{1/2}}. \tag{9.5}$$

As in (III), above, the role of $\bar{f}_1, \bar{f}_2$ can be dropped.

If for the alternative (f_1, f_2) as given by (9.2) f is the normal (0, 1) density function and

$$\frac{\sigma_1(\theta)}{\sigma_2(\theta)} = 1 - \theta$$

(hence H_0 implies the variances are the same), then (9.5) becomes

$$\frac{\int_{-\infty}^{\infty} (6F(x) - 6F^2(x) - 1)(1 - x^2)f(x)\,dx}{\int_0^1 (6t^2 - 6t + 1)^2\,dt\Big)\left(\int_{-\infty}^{\infty} (x^2 - 1)^2 f(x)\,dx\right)^{1/2}} = \sqrt{15/2\pi^2}.$$

The integrals in the numerator are evaluated in [8]. By the remark at the end of II, this is also the efficiency of (9.4) relative to the F statistic. This result was given by Mood in [12].

10. Acknowledgment. I am indebted to Prof. Wassily Hoeffding for suggesting the study of L_h statistics and for much help and encouragement.

REFERENCES

[1] A. R. Collar, "On the reciprocation of certain matrices," *Proc. Roy. Soc. Edinburgh*, Vol. 59 (1939), pp. 195–206.

[2] B. V. Gnedenko and A. M. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, trans. by K. L. Chung, Addison-Wesley Press, Inc., Cambridge, Massachusetts, 1954.

[3] W. Hoeffding, "A class of statistics with asymptotically normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 293–325.

[4] W. Hoeffding, "'Optimum' non-parametric tests," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951, pp. 83–92.

[5] W. Hoeffding, "The large-sample power of tests based on permutations of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 169–192.

[6] W. Hoeffding, "On the distribution of the expected values of the order statistics," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 93–100.

[7] W. HOEFFDING AND J. R. ROSENBLATT, "The efficiency of tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 52–63.

[8] H. L. JONES, "Exact lower moments of order statistics in small samples from a normal distribution," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 270–273.

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

[10] E. L. LEHMANN, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 23–43.

[11] H. B. MANN AND R. R. WHITNEY, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 50–60.

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

[13] G. E. NOETHER, "On a theorem of Pitman," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 64–68.

[14] E. J. G. PITMAN, "Notes on non-parametric statistical inference," unpublished notes. (For a published reference to the same, see [13].)

[15] M. H. STONE, *Linear Transformations in Hilbert Space and Their Applications to Analysis, Am. Math. Soc.* Colloquium publications, Vol. 15 (1932).

[16] D. TEICHROEW, "Empirical power functions of non-parametric tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 340–344.