THE TREATMENT OF TIES IN SOME NONPARAMETRIC TESTS[1]

By Joseph Putter

University of California, Berkeley

1. Introduction. Most of nonparametric testing theory is usually presented under the assumption that all the samples involved are drawn from continuous distributions, and that tied observations can therefore be ignored or treated in any convenient way, without affecting the performance characteristic of the test. In practice, however, this assumption is not a realistic one, and the distributions involved are in general to be regarded as discontinuous, either because of intrinsic reasons (integer-valued or otherwise discrete random variables) or because of limitations on the precision of measurements. Therefore, usually, ties will occur with positive probability, and the way they are treated does affect the performance characteristic of the test. The problem of ties has therefore to be considered, in particular with a view to preserving the nonparametric character of the test, and to making sure of setting it up on the desired level of significance.

The usual practice in attacking the problem has been to consider the conditional distributions of the statistics concerned given that the number of observations in each tied group is a fixed constant. This, however, was never explicitly made clear, and these conditional distributions, as well as their variances and other characteristics, are referred to as distributions (or variances, etc.) "when ties are present." In this category belong Kendall's work on ties in rank correlation theory, and Kruskal's theorem concerning a generalized Wilcoxon test (see Section 8).

In this paper, we attack the problem from the standpoint of the ties being random variables. Our main concern is the comparison between the "randomized" and the "nonrandomized" way of treating the ties. In Sections 3 and 4 we consider the one-sided sign test, and show that randomization reduces both the exact power and the asymptotic efficiency of the test. In Sections 5–8 we consider the Wilcoxon test. For small samples the nonrandomized treatment of ties presents practical difficulties, but the asymptotic (large sample) problem can be handled. Again, it is shown that randomization results in reduced efficiency.

2. Notation and theorems used. We shall use the notation $\mathfrak{N}(a, b)$ for normal random variables (with mean a and variance b), and $\mathfrak{B}(n, p)$ for binomials. The symbol $\xrightarrow{P}$ will denote convergence in probability, and $\xrightarrow{L}$ convergence in law (convergence of distributions).

Received March 16, 1954.

[1] Work supported in part by the Office of Naval Research, and submitted in partial satisfaction of the requirements for the degree of Ph.D. at the University of California, Berkeley.

To compare the asymptotic performances of two consistent tests, we shall use Pitman's concept of asymptotic relative efficiency. The concept is presented in Pitman's lecture notes and also by Noether [11]. In particular, we shall use the following theorem.

THEOREM A (Pitman, as quoted in Noether [11], pp. 241–242). *Let H be a hypothesis specifying the value θ_0 of a population parameter θ, and A the one-sided alternative $\theta > \theta_0$. Let $\{\tau_{in}\}$, $i = 1, 2; n = 1, 2, \cdots$, be two sequences of tests of H against A, on the same level of significance α. Let τ_{in} consist of rejecting H when $\bar{S}_{in} > k_{in}$, where $\bar{S}_{in}$ are statistics and k_{in} appropriate constants. Let $\psi_{in}(\theta)$ and $\sigma_{in}(\theta)$ be functions such that $\psi'_{in}(\theta)$ exists in the neighborhood of θ_0, and let the following conditions be satisfied as $n \to \infty$:*

$$\frac{\psi'_{in}(\theta_n)}{\psi'_{in}(\theta_0)} \to 1, \qquad \theta_n = \theta_0 + \frac{a}{n^{1/2}}, \qquad a \text{ a positive constant}; \tag{2.1}$$

$$\frac{\sigma_{in}(\theta_n)}{\sigma_{in}(\theta_0)} \to 1; \tag{2.2}$$

$$\frac{H_i(n)}{n^{1/2}} \to c_i, \qquad H_i(n) = \frac{\psi'_{in}(\theta_0)}{\sigma_{in}(\theta_0)}, \qquad c_i \text{ a positive constant}; \tag{2.3}$$

and either

$$\frac{\bar{S}_{in} - \psi_{in}(\theta)}{\sigma_{in}(\theta)} \xrightarrow{L} \mathfrak{N}(0, 1) \tag{2.4}$$

uniformly in θ in the neighborhood of θ_0, or

$$\frac{\bar{S}_{in} - \psi_{in}(\theta_n)}{\sigma_{in}(\theta_n)} \xrightarrow{L} \mathfrak{N}(0, 1). \tag{2.5}$$

Then the asymptotic relative efficiency of $\{\tau_{2n}\}$ with respect to $\{\tau_{1n}\}$ is $\lim_{n\to\infty} H_2^2(n)/H_1^2(n)$.

(Noether defines $\psi_{in}(\theta) = E(\bar{S}_{in} \mid \theta)$ and $\sigma_{in}(\theta) = \sigma(\bar{S}_{in} \mid \theta)$, but it is easily seen from Pitman's proof of the theorem that this specification is not necessary.)

To handle the uniform convergence required in condition (2.4), we shall use the following theorem.

THEOREM B (*Parzen* [12], p. 35). *A necessary and sufficient condition for a sequence of distributions $F_n = F_n^{(\theta)}$ to converge to a distribution F uniformly in θ is that*

$$f_n^{(\theta)}(t) \to f(t) \text{ uniformly in } \theta, \tag{2.6}$$

where $f_n^{(\theta)}$ and f denote the respective characteristic functions. The convergence (2.6) *is then jointly uniform in θ and t for every finite t-interval.*

(Theorem B is a particular case of Parzen's Theorem 7c.)

THE SIGN TEST

3. Randomized and nonrandomized test. Let $Z_1, \cdots, Z_n$ be independent and identically distributed random variables. Denote the number of positive

Z_k's by N_+, of negative Z_k's by N_-, and of zeros among the Z_k's by N_0. The sign test consists of rejecting the hypothesis

$$H: \quad P(Z_k > 0) = P(Z_k < 0),$$

against the alternative

$$A: \quad P(Z_k > 0) > P(Z_k < 0),$$

say, whenever N_+ is too large.

In practice Z_k frequently is of the form $X_k - Y_k$, where X_k and Y_k are independent. If the distribution functions of X_k and Y_k are continuous, then $P(Z_k = 0) = 0$. In this case, under the hypothesis, N_+ is $\mathcal{B}(n, \frac{1}{2})$, which gives us the cutoff point.

In the general (discontinuous) case, denote

$$P(Z_k > O \mid H) = p_+, \qquad P(Z_k = O \mid H) = p_0,$$

$$P(Z_k > O \mid A) = q_+, \qquad P(Z_k = O \mid A) = q_0, \qquad P(Z_k < O \mid A) = q_-.$$

Consider the conditional distribution of N_+ given that $N_0 = n_0$. Under H,

$$P(N_+ = x \mid n_0) = p_H(x) = \binom{n - n_0}{x} (\tfrac{1}{2})^{n-n_0};$$

under A,

$$P(N_+ = x \mid n_0) = p_A(x) = \binom{n - n_0}{x} \left(\frac{q_-}{1 - q_0}\right)^{n-n_0} \left(\frac{q_+}{q_-}\right)^x,$$

$x = 0, 1, \cdots, n - n_0$. Thence

$$\frac{p_A(x)}{p_H(x)} = c(n_0) \left(\frac{q_+}{q_-}\right)^x,$$

which is a strictly increasing function of x. Therefore, by the Neyman-Pearson lemma, the unique most powerful test based on N_+ and N_0 is given by

$$N_+ > k(N_0), \tag{3.1}$$

where the cutoff point $k(n_0)$ is, of course, the one corresponding to $\mathcal{B}(n - n_0, \frac{1}{2})$.

It is obvious that $k(N_0)$ is not a linear function of N_0. Thence the test (3.1) does not coincide with the test

$$N_+ + \tfrac{1}{2}N_0 > k, \tag{3.2}$$

which was proposed, e.g., by Dixon and Mood [3]. In fact, the distribution of $N_+ + \frac{1}{2}N_0$ under H depends on the unknown parameter p_0, so that the cutoff point k cannot be well defined. The usual practice seems to be to take for k the cutoff point corresponding to $\mathcal{B}(n, \frac{1}{2})$. This, as was shown by Hemelrijk [4], results in lowering the level of significance and consequently also the power of the test. However, the difficulty caused by the dependence of k on p_0 can be obviated when asymptotic properties are considered, and we shall return to the matter in the next section.

The test (3.1), which amounts to "omitting the ties from the observations" and which was suggested, e.g., by Dixon and Massey [2], is, as we have seen, the unique most powerful test based on N_+ and N_0. However, another customary procedure is one based on "randomization": after observing the Z_k's, we perform N_0 independent random experiments, assigning each of the N_0 zeros among the Z_k's a positive or negative sign with equal probabilities ($= \frac{1}{2}$). We thus get, say, N_+^r additional positives. The random variable $N_+^R = N_+ + N_+^r$ is, under H, $\mathcal{B}(n, \frac{1}{2})$, and we can apply the test

$$N_+^R > k \tag{3.3}$$

without worrying about the unknown p_0.

Consider, again, the conditional situation given that $N_0 = n_0$. Denote by $p(y)$ the frequency distribution of $\mathcal{B}(n_0, \frac{1}{2})$. The joint (conditional) frequency distribution of N_+ and N_+^r is $p_H(x)p(y)$ under H, and $p_A(x)p(y)$ under A. The ratio of the two expressions is $p_A(x)/p_H(x)$, so that (3.1) is also the unique most powerful test based on N_+, N_0, *and* N_+^r. We have thus proved the following theorem.

THEOREM 1. *The nonrandomized test* (3.1) *is uniformly more powerful* (*against the one-sided alternative* A) *than the randomized test* (3.3).

As a numerical example, we give in Table I the powers of the two tests for $n = 10$, against the alternative $q_+/q_- = 2$. Since the power of either test depends on q_0, we tabulate the conditional power given $N_0 = n_0$, for all values of n_0. The tests are considered on the .05 level. To keep this level exact (and to get a valid comparison between the tests), we modify the tests in the usual way. For example, the test (3.3) is now formulated as follows: Reject H with probability 1 if $N_+^R > k$; reject H with probability φ if $N_+^R = k$; accept H otherwise. In our particular case, $k = 8$ and $\varphi = .893$.

TABLE I

n_0	0	1	2	3	4	5	6	7	8	9	10
Power of randomized test (3.3)	.278	.241	.208	.177	.150	.127	.106	.089	.074	.061	.050
Power of nonrandomized test (3.1)	.278	.244	.232	.216	.184	.171	.158	.119	.088	.067	.050

In particular, against the alternative $q_0 = q_- = \frac{1}{4}$, $q_+ = \frac{1}{2}$, the power of (3.3) is .195, while that of (3.1) is .221.

4. Asymptotic properties. For large sample sizes n, it is convenient to use the normal approximation to the binomial, and we shall now compare the performances of the randomized and nonrandomized tests when this approximation is used.

For the randomized test statistic N_+^R we have, under H,

$$\frac{2N_+^R - n}{n^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1), \tag{4.1}$$

which gives us the usual normal approximation to (3.3). Under A, N_+^R is

$$\mathcal{B}(n, q_+ + \tfrac{1}{2}q_0),$$

and hence

$$\frac{N_+^R - n(q_+ + \frac{1}{2}q_0)}{[n(q_+ + \frac{1}{2}q_0)(q_- + \frac{1}{2}q_0)]^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1). \tag{4.2}$$

It is easily seen from (4.1) and (4.2) that the test (3.3) is consistent against the one-sided alternative A.

It is more difficult to derive the normal approximation to (3.1). Since in any case *the* normal approximation of a test is not an obviously definable concept, we shall derive a nonrandomized asymptotic test by starting from (3.2). The joint distribution of N_+, N_0, and N_- is trinomial; hence the test statistic

$$N_+' = N_+ + \tfrac{1}{2}N_0$$

is asymptotically normal. More precisely, under H,

$$\frac{2N_+' - n}{[n(1 - p_0)]^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1).$$

Since $N_0/n \xrightarrow{P} p_0$, we have

$$\frac{2N_+' - n}{(n - N_0)^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1), \tag{4.3}$$

which gives us an asymptotic test independent of p_0. Under A,

$$\frac{N_+' - n(q_+ + \frac{1}{2}q_0)}{(n[(q_+ + \frac{1}{2}q_0)(q_- + \frac{1}{2}q_0) - \frac{1}{4}q_0])^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1), \tag{4.4}$$

and, again, the nonrandomized test corresponding to (4.3) is consistent against A.

We now compare the asymptotic performances of the two tests in terms of Pitman's concept of asymptotic relative efficiency. In the notation of Theorem A, put

$$\theta = q_+ + \tfrac{1}{2}q_0, \qquad \theta_0 = \tfrac{1}{2}.$$

THEOREM 2. *Let $\{A_\theta, \theta > \frac{1}{2}\}$ be a family of alternatives for which $q_0 = p_0$. Then the asymptotic relative efficiency of the randomized test* (3.3) *with respect to the (nonrandomized) test based on the statistic*

$$T_n = \frac{2N_+' - n}{(n - N_0)^{1/2}}$$

is $1 - p_0$.

PROOF. Put

$$\bar{S}_{1n} = T_n(1 - p_0)^{1/2}, \qquad \bar{S}_{2n} = N_+^R,$$
$$\psi_{1n}(\theta) = (2\theta - 1)n^{1/2}, \qquad \psi_{2n}(\theta) = n\theta,$$
$$\sigma_{1n}(\theta) = (4\theta(1 - \theta) - p_0)^{1/2}, \qquad \sigma_{2n}(\theta) = (n\theta(1 - \theta))^{1/2}.$$

Conditions (2.1)–(2.3) obviously hold, and we proceed to verify (2.4) and/or (2.5).

For $i = 2$, the convergence (2.4) holds by (4.2). From the usual proof of binomial convergence to the normal, it is easily seen that the corresponding convergence of characteristic functions is uniform in θ in the neighborhood of θ_0, and hence, by Theorem B, so is (4.2), and condition (2.4) holds. For $i = 1$, we have

$$\frac{\bar{S}_{1n} - \psi_{1n}(\theta)}{\sigma_{1n}(\theta)} = \frac{2(N_+ - \theta n)}{(n[4\theta(1 - \theta) - p_0])^{1/2}}\left(\frac{n(1 - p_0)}{n - N_0}\right)^{1/2}$$
$$+ (2\theta - 1)\left(\frac{n(1 - p_0)}{4\theta(1 - \theta) - p_0}\right)^{1/2}\left(\left(\frac{n}{n - N_0}\right)^{1/2} - \frac{1}{(1 - p_0)^{1/2}}\right).$$

Now,

$$\frac{2(N_+ - \theta n)}{(n[4\theta(1 - \theta) - p_0])^{1/2}} \xrightarrow{L} \mathfrak{N}(0, 1)$$

by (4.4), and this convergence, as before, can easily be shown to be uniform in θ in the neighborhood of θ_0. We have

$$\left(\frac{n(1 - p_0)}{n - N_0}\right)^{1/2} \xrightarrow{P} 1 \quad \text{and} \quad \left(\frac{n}{n - N_0}\right)^{1/2} - \frac{1}{(1 - p_0)^{1/2}} \xrightarrow{P} 0$$

independently of θ, and

$$(2\theta_n - 1)\left(\frac{n(1 - p_0)}{4\theta_n(1 - \theta_n) - p_0}\right)^{1/2} \to 2a.$$

Hence condition (2.5) holds. Our result now follows from Theorem A.

THE WILCOXON TEST

5. Notation and known results. We shall use the following notation in connection with the Wilcoxon test. $(X_1, \cdots, X_n)$ is a sample of n independent observations from a distribution $F(z)$, and $(Y_1, \cdots, Y_m)$ is a sample of m independent observations from a distribution $G(z)$. If all the $m + n$ observations in the pooled sample are different, we rank them in ascending order of magnitude, assigning the rank 1 to the smallest observation. We denote by S_{nm} the sum of the ranks assigned to the X's. The Wilcoxon test of the hypothesis $F = G$ consists of rejecting the hypothesis when S_{nm} is too large.

The mean, variance, and asymptotic distribution of S_{nm}, in the case when F and G are continuous (and therefore the probability of getting two or more equal observations is 0), are known, and are summarized below.

When $F = G$, every possible ordering of the pooled sample occurs with the same probability $1/(n + m)!$, and the distribution of S_{nm} can be derived from this fact alone. We shall denote any statistic with this probability distribution by S^0_{nm}. From Mann-Whitney [10] we have

$$ES^0_{nm} = \frac{n(n + m + 1)}{2} = \mu_{nm}, \quad \text{say}; \tag{5.1}$$

$$\sigma^2(S^0_{nm}) = \frac{nm(n + m + 1)}{12} = \sigma^2_{nm}, \quad \text{say}; \tag{5.2}$$

$$T^0_{nm} = \frac{S^0_{nm} - \mu_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, 1) \qquad \text{as } \frac{1}{n} + \frac{1}{m} \to 0. \tag{5.3}$$

In general, when F and G are any two (continuous) distributions, we have, from Mann-Whitney [10],

$$ES_{nm} = \mu_{nm} + nm\theta, \tag{5.4}$$

$$\theta = \theta(F, G) = P(X_1 > Y_1) - \tfrac{1}{2} = \int_{-\infty}^{\infty} G(z)\, dF(z) - \tfrac{1}{2}, \tag{5.5}$$

$$\begin{aligned} \sigma^2(S_{nm}) = \sigma^2_{nm} + nm[(\theta - \lambda_1)(n - 1) \\ + (\theta - \lambda_2)(m - 1) - \theta^2(n + m - 1)], \end{aligned} \tag{5.6}$$

$$\lambda_1 = \lambda_1(F, G) = \tfrac{1}{3} - \int_{-\infty}^{\infty} F^2(z)\, dG(z), \tag{5.7}$$

$$\lambda_2 = \lambda_2(F, G) = \tfrac{1}{3} - \int_{-\infty}^{\infty} [1 - G(z)]^2\, dF(z). \tag{5.8}$$

When $n \to \infty$ while $m/n = c$ is held constant, we have, from Lehmann [9],

$$\frac{S_{nm} - ES_{nm}}{\sigma(S_{nm})} \xrightarrow{L} \mathfrak{N}(0, 1). \tag{5.9}$$

(That $\sigma(S_{nm})$ is the correct norming factor can be seen from Hoeffding [6], Theorem 5.2.)

For the case of discontinuous F and G, which we shall consider in the following sections, we adopt the following notation. We assume the common discontinuities of F and G (which are the only ones that matter) to be finite in number, and denote them by ξ_k, $k = 1, \cdots, K$. Their locations are not assumed known, and are irrelevant to our considerations. We define

$p_k = P(X_1 = \xi_k)$, $\quad q_k = P(Y_1 = \xi_k)$;

U_k = the number of X's which are equal to ξ_k;

V_k = the number of Y's which are equal to ξ_k;

$W_k = U_k + V_k$;

$U = (U_1, \cdots, U_K)$, $\quad V = (V_1, \cdots, V_K)$, $\quad W = (W_1, \cdots, W_K)$.

We shall write $\sum_k$ for $\sum_{k=1}^K$.

6. The treatment of ties. When F and G are continuous, the probability of getting tied (equal) observations is 0, so that this event may be ignored. In the discontinuous case, however, ties occur with positive probability, and when they do occur, the pooled sample can no longer be uniquely ordered. The problem arises, therefore, of how the Wilcoxon test is to be defined in such a case.

An obvious solution to the problem, proposed by many writers on the subject, is, again, "randomization": each group of equal observations is ordered at random, giving every possible ordering (within the group) the same probability. This results in an ordering of the pooled sample, and the sum of the ranks of the X's can now be defined. The only difference from the continuous case is that this new random variable is defined over a different sample space, because its value depends not only on the observed X's and Y's but also on the outcome of the randomization procedure. We shall denote this sum of the "randomized" ranks of the X's also by S_{nm}.

Again, if $F = G$, every possible (randomized) ordering of the pooled sample is equally probable, and hence S_{nm} is distributed as S^0_{nm}. The Wilcoxon test can therefore be applied using S_{nm}, with the same cutoff point as in the continuous case. The main objection to this procedure seems to be that the outcome of the test (rejection or acceptance of the hypothesis) is thus made to depend not only on the observations but also on an additional, and more or less irrelevant, random experiment. We are thus led to look for a test which is

(i) distribution-free under the hypothesis;
(ii) dependent on the observations only; and
(iii) as close as possible to the original Wilcoxon test.

We leave the precise meaning of this last requirement unspecified for the moment, and shall elaborate the point later on.

For the remainder of this section, we shall need to consider only the case when $F = G$, and it will be convenient to assume that F is purely discontinuous. In this case, the ordering of the pooled sample is given by the nonzero component of the two vectors U and V, as long as the observations alone are considered. Hence any rank (order) statistic which depends on the observations only can be expressed in terms of U and V. Requirement (ii) means, therefore, that the rejection region R of the test will be a region in the $2K$-dimensional sample space of the random vector $(U_1, \cdots, U_K, V_1, \cdots, V_K)$.

In this sample space, the vector W is a sufficient statistic for the vector parameter $(p_1, \cdots, p_K)$, i.e., the conditional probability

$$P(u \mid w) = P(U = u, V = w - u \mid W = w)$$
$$= \frac{n!}{u_1! \cdots u_K!} \cdot \frac{m!}{(w_1 - u_1)! \cdots (w_K - u_K)!} \Big/ \frac{(m+n)!}{w_1! \cdots w_K!}$$

is independent of the p_k's. Hence, if the size α of R, that is,

$$P(R) = \sum_w P(W = w)\, P(R \mid W = w)$$
$$= \sum_w \frac{(m+n)!}{w_1! \cdots w_K!} p_1^{w_1} \cdots p_K^{w_K} \sum_{(u, w-u) \in R} P(u \mid w),$$

is to be independent of the p_k's (requirement (i)), we must have $P(R \mid W = w) = \alpha$ for every w, which is the usual condition on distribution-free tests when a sufficient statistic with a complete family is involved. But since for every fixed w we have only a finite set of probabilities $P(u \mid w)$, and these sets vary with w, it will in general be impossible to find a region R with exact size α. However, this difficulty can be obviated, e.g., by considering regions which include some sample points not definitely but with certain given probabilities. Thus it appears that some random element outside the observations is unavoidable, unless we do not insist on the exact size α. But in practice this consideration is unimportant, because one is usually quite content to stop just short of the given size α.

Suppose that various regions R of the required type and of exact size α (produced by the above, or any other, device) are available. Denote by R_0 the rejection region $[S_{nm} > a]$, of the same size α, given by the "randomized" Wilcoxon test. Then R_0 is defined in a different sample space, which can be described as the result of splitting each point of the (u, v)-space into several points corresponding to the possible outcomes of the randomization procedure. We shall view the sets R as sets in this "extended" sample space, too.

We have $P(R) = P(R_0) = \alpha$, or $P(R \cap \bar{R}_0) = P(\bar{R} \cap R_0)$, where the notation $\bar{A}$ stands for the complement of A. Now, one possible interpretation of requirement (iii) above is to choose R so as to minimize $P(R \cap \bar{R}_0)$. This may be justified as follows. Suppose F is really continuous, and the ties occur only because of insufficient precision of measurement. The randomized test is, in a sense, approximately equivalent to the (Wilcoxon) test which we would use if our measurements were precise, because the effect of the randomization procedure is similar to the effect of replacing each discontinuity by an interval of uniform distribution (cf. Section 7). It is therefore reasonable to try to minimize the probability of getting a result (rejection or acceptance of the hypothesis) different from the result of the randomized test. But this probability, when the hypothesis is true, is $P(R \cap \bar{R}_0) + P(\bar{R} \cap R_0) = 2P(R \cap \bar{R}_0)$.

We thus want to minimize $P(R \cap [S_{nm} \leqq a])$, which will be achieved if we minimize

$$P(R \cap [S_{nm} \leqq a] \mid W = w) = \sum_{(u,w-u)\varepsilon R} P(u \mid w)\, P(S \leqq a \mid U = u, V = w - u)$$

for every w. This is to be done under the condition

$$(6.1) \qquad \sum_{(u,w-u)\varepsilon R} P(u \mid w) = P(R \mid W = w) = \alpha.$$

In a manner analogous to the proof of the Neyman-Pearson lemma, it is easily seen that the "optimum" region is obtained by the following procedure. For every vector w, we order all the possible vectors $(u, v) = (u, w - u)$ by the magnitude of $P(S \leqq a \mid U = u, V = w - u)$. We take that vector $(u, w - u)$ for which this probability is smallest, then that vector for which it is the next smallest, etc., until the (conditional) size α, as in (6.1), is reached. Doing this for all w, we get the desired R.

Unfortunately, the tabulations required for this test are much too extensive. We can approximate the test if, instead of rejecting the hypothesis when $P(S_{nm} \leqq a \mid U = u, V = w - u)$ is too small, we reject it when $E(S_{nm} \mid U = u, V = w - u)$ is too large. The two tests will probably not differ too much.

The statistic

$$S'_{nm} = E_r(S_{nm} \mid U, V), \tag{6.2}$$

where E_r denotes expected value under randomization, is the sum of the midranks of the X's, where the midrank of an observation is defined as the mean rank of all the observations equal to that observation, or, more precisely,

$$\text{midrank } (X) = \frac{N_1 + N_2 + 1}{2},$$

where N_1 is the number of observations smaller than X, and N_2 is the number of observations (including X) not larger than X. The statistic S'_{nm} has been proposed by many writers as a test statistic to replace S_{nm} when ties are present. However, by the preceding considerations, the cutoff point has to depend on W, and the tabulation involved is prohibitive. A few cases have been tabulated in [13], but they can merely serve as an indication of the task involved in more exhaustive tabulation.

Kruskal [7] derived the conditional asymptotic distribution of S'_{nm} given fixed $W = w(n, m)$ which fulfill a certain convergence condition (cf. Section 8) for the case $F = G$. In the next two sections we shall derive the (unconditional) asymptotic distribution of S'_{nm} in general, and discuss some consequences.

7. The asymptotic distribution of S'_{nm}. We now drop the assumption that F and G are purely discontinuous. Consider the conditional distribution of S_{nm} given a fixed pooled sample of X's and Y's. For this fixed sample, let $U = u$, $V = v$. Denote by r the sum of the ranks of those X's which are not equal to any ξ_k (and which are therefore, with probability 1, untied), and by r_k the number of those observations (X's and Y's) which are smaller than ξ_k.

Under the randomization procedure which generates S_{nm}, those observations which are equal to ξ_k are assigned the ranks $r_k + 1, r_k + 2, \cdots, r_k + u_k + v_k$ at random, with every ordering equally probable. Hence the sum of the ranks of those X's which are equal to ξ_k is $u_k r_k + S^0_{u_k, v_k}$, and

$$S_{nm} = r + \textstyle\sum_k (u_k r_k + S^0_{u_k, v_k}).$$

Therefore, by (6.2) and (5.1),

$$S'_{nm} = r + \textstyle\sum_k (u_k r_k + \mu_{u_k, v_k}),$$

$$S_{nm} = S'_{nm} + \textstyle\sum_k (S^0_{u_k, v_k} - \mu_{u_k, v_k}),$$

where the $K + 1$ terms on the right are (conditionally) independent. Since this holds for every fixed sample, we can write

$$S_{nm} = S_{nm} + \sum_k (S^0_{U_k, V_k} - \mu_{U_k, V_k}), \tag{7.1}$$

where the terms on the right are conditionally independent given U and V.

Obviously, we have

$$ES'_{nm} = EE_r(S_{nm} \mid U, V) = ES_{nm}.$$

To calculate $\sigma^2(S_{nm})$, we note that, by (7.1),

$$\begin{aligned} \sigma^2(S_{nm}) &= E(S_{nm} - ES_{nm})^2 \\ &= EE_r[(S_{nm} - ES_{nm})^2 \mid U, V] \\ &= EE_r\{[S'_{nm} - ES_{nm} + \textstyle\sum_k (S^0_{U_k, V_k} - \mu_{U_k, V_k})]^2 \mid U, V\} \\ &= \sigma^2(S'_{nm}) + \tfrac{1}{12}\textstyle\sum_k EU_kV_k(U_k + V_k + 1) \\ &= \sigma^2(S'_{nm}) + \frac{nm}{12}\sum_k p_kq_k[(n-1)p_k + (m-1)q_k + 3], \end{aligned}$$

or

$$\sigma^2(S'_{nm}) = \sigma^2(S_{nm}) - \frac{nm}{12}\sum_k p_kq_k[(n-1)p_k + (m-1)q_k + 3]. \tag{7.2}$$

In particular, when $F = G$,

$$\sigma^2(S'_{nm}) = \sigma^2_{nm} - \frac{nm}{12}\sum_k p_k^2[(n+m-2)p_k + 3]. \tag{7.3}$$

Of some interest, when $F = G$, is also the conditional variance $\sigma^2(S'_{nm} \mid W)$. Since the conditional distribution of S_{nm} given W is still that of S^0_{nm}, this variance can be computed in a manner similar to the preceding argument, giving (when $F = G$)

$$\sigma^2(S'_{nm} \mid W) = \sigma^2_{nm} - \frac{nm}{12(n+m)(n+m-1)}\sum_k W_k(W_k^2 - 1). \tag{7.4}$$

This is the variance given by Kruskal [7], and in a more cumbersome form by Hemelrijk [5].

Since ES_{nm} and $\sigma^2(S_{nm})$ as given by (5.4) and (5.6) refer to the continuous case, we shall touch on the modifications required for discontinuous F and G. By Lemma 5.1 of Lehmann [9], there exist two continuous distributions F^* and G^* under which the distribution of S_{nm} is the same as under F and G. These continuous distributions are obtained, essentially, by replacing the discontinuities by intervals of uniform distribution. We define

$$\begin{aligned} \theta^* &= \theta^*(F, G) = \theta(F^*, G^*), \\ \lambda_j^* &= \lambda_j^*(F, G) = \lambda_j(F^*, G^*), \qquad j = 1, 2, \end{aligned}$$

referring to the definitions (5.5), (5.7), and (5.8). From (5.4) and (5.6), we now have

$$(7.5)\qquad ES_{nm} = \mu_{nm} + \theta^* nm,$$

$$(7.6)\qquad \begin{aligned}\sigma^2(S_{nm}) &= \sigma^2_{nm} + nm[(\theta^* - \lambda_1^*)(n-1)\\ &\quad + (\theta^* - \lambda_2^*)(m-1) - \theta^{*2}(n+m-1)].\end{aligned}$$

For later use, we compute θ^* in terms of F and G. Denote by B the real line with the points ξ_k excluded. We have

$$\begin{aligned}\theta^* + \tfrac{1}{2} &= \int_{-\infty}^{\infty} G^*(z)\, dF^*(z) = \int_B G(z)\, dF(z) + \sum_k \int_0^1 [G(\xi_k - 0) + tq_k]p_k\, dt\\ &= \int_B G(z)\, dF(z) + \sum_k p_k\, G(\xi_k - 0) + \tfrac{1}{2}\sum_k p_k q_k = P(X_1 > Y_1)\\ &\qquad + \tfrac{1}{2}P(X_1 = Y_1),\end{aligned}$$

or

$$(7.7)\qquad \theta^* = P(X_1 > Y_1) + \tfrac{1}{2}P(X_1 = Y_1) - \tfrac{1}{2}.$$

We now give a theorem connecting the asymptotic distribution of S'_{nm} with that of S_{nm}. Note that the symbol $\sigma^2_{U_k,V_k}$ will stand for $\frac{1}{12}\, U_k V_k(U_k + V_k + 1)$, as in (5.2), and will have nothing to do with the variances of U_k and V_k. This refers, of course, to all the symbols with U_k and V_k as subscripts.

THEOREM 3. *If, for a pair of distributions* (F, G), *and possibly under some restrictions concerning the relation between* n *and* m, *we have*

$$(7.8)\qquad \frac{S_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, b^2),$$

$$(7.9)\qquad \frac{\sigma^2_{U_k,V_k}}{\sigma^2_{nm}} \xrightarrow{P} b_k^2$$

as $1/n + 1/m \to 0$, *then, under the same conditions,*

$$\frac{S'_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, \bar{b}^2),$$

where

$$\bar{b}^2 = b^2 - \textstyle\sum_k b_k^2.$$

PROOF. Subtracting ES_{nm} from both sides of (7.1) and dividing by σ_{nm}, we have

$$(7.10)\qquad \frac{S_{nm} - ES_{nm}}{\sigma_{nm}} = \frac{S'_{nm} - ES_{nm}}{\sigma_{nm}} + \sum_k \frac{\sigma_{U_k,V_k}}{\sigma_{nm}} T^0_{U_k,V_k},$$

where T^0 is defined by (5.3). The U_k and V_k are $\mathcal{B}(n, p_k)$ and $\mathcal{B}(m, q_k)$, respectively.

Let $d > 0$ be a fixed number which we shall specify later, and define

$$R_{nm} = \left[\left|\frac{U_k}{n} - p_k\right| < d, \quad \left|\frac{V_k}{m} - q_k\right| < d, \quad \left|\frac{\sigma_{U_k, V_k}}{\sigma_{nm}} - b_k\right| < d, \quad \text{all } k \right]$$

We have, by (7.9),

$$P(R_{nm}) \to 1 \quad \text{as} \quad 1/n + 1/m \to 0. \tag{7.11}$$

Define

$h(t) \quad = e^{-t^2/2}$;

$h_{nm}(t) \ = \text{the characteristic function of } T^0_{nm}$;

$f_{nm}(t) \ = \text{the characteristic function of } \dfrac{S_{nm} - ES_{nm}}{\sigma_{nm}}$;

$f^{uv}_{nm}(t) \ = \text{the conditional characteristic function of } \dfrac{S_{nm} - ES_{nm}}{\sigma_{nm}} \text{ given } U = u, V = v$;

$g_{nm}(t) \ = \text{the characteristic function of } \dfrac{S'_{nm} - ES_{nm}}{\sigma_{nm}}$;

$g^{uv}_{nm}(t) \ = \text{the conditional characteristic function of } \dfrac{S'_{nm} - ES_{nm}}{\sigma_{nm}} \text{ given } U = u, V = v$;

$A \quad = \prod_k h(tb_k), \qquad a_k = \dfrac{\sigma_{u_k, v_k}}{\sigma_{nm}}$

All integrals will be taken in the (u, v)-space, with respect to the probability measure in that space.

By (7.10), we have

$$g^{uv}_{nm}(t) = \frac{f^{uv}_{nm}(t)}{\prod_k h_{u_k, v_k}(ta_k)},$$

or

$$g^{uv}_{nm}(t) - h(t\bar{b}) = \frac{1}{A}\,[f^{uv}_{nm}(t) - h(tb)] + \frac{1}{A}\, g^{uv}_{nm}(t) \left[\prod_k h(tb_k) - \prod_k h_{u_k, v_k}(ta_k)\right]$$

Hence

$$|g_{nm}(t) - h(t\bar{b})| = \left|\int [g^{uv}_{nm}(t) - h(t\bar{b})]\right|$$
$$\leqq \int_{\bar{R}_{nm}} |g^{uv}_{nm}(t) - h(t\bar{b})| + \frac{1}{A}\left|\int_{R_{nm}} [f^{uv}_{nm}(t) - h(tb)]\right| + \frac{1}{A}\int_{R_{nm}} \left|\prod_k h(tb_k) - \prod_k h_{u_k, v_k}(ta_k)\right|.$$

Using the definition of R_{nm}, the property (7.11), the condition (7.8), and the fact that, by (5.3), $h_{nm}(t) \to h(t)$, each of the three expressions involved can be shown to converge to 0 as $1/n + 1/m \to 0$, and the theorem is proved.

8. Consequences of Theorem 3. Since the asymptotic distribution of S_{nm} is known, Theorem 3 enables us to investigate the asymptotic behavior of S'_{nm} and to compare the tests based on the two statistics.

THEOREM 4. *If $F = G$, then*

$$\frac{S'_{nm} - \mu_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, 1 - \sum_k p_k^3), \text{ as } \frac{1}{n} + \frac{1}{m} \to 0.$$

Therefore, if $s_{nm} = s_{nm}(U, V)$ is any sequence of positive statistics satisfying

$$\frac{s_{nm}^2}{\sigma_{nm}^2} \xrightarrow{P} 1 - \sum_k p_k^3,$$

then

$$T'_{nm} = \frac{S'_{nm} - \mu_{nm}}{s_{nm}} \xrightarrow{L} \mathfrak{N}(0, 1) \text{ as } \frac{1}{n} + \frac{1}{m} \to 0.$$

PROOF. By (5.9), we know that (7.8) holds with $b = 1$. We also have

$$\frac{\sigma_{U_k,V_k}^2}{\sigma_{nm}^2} = \frac{U_k V_k (W_k + 1)}{nm(n + m + 1)}$$

$$= \frac{U_k}{n}\frac{V_k}{m}\left(\frac{W_k}{n+m} + \frac{n + m - W_k}{(n+m)(n+m+1)}\right) \xrightarrow{P} p_k^3$$

Hence the theorem follows from Theorem 3.

Theorem 4 gives us test statistics whose asymptotic distributions, under the hypothesis $F = G$, are independent of F, and which can therefore be used to obtain asymptotically distribution-free tests. The rejection region of such a test will be $[T'_{nm} > a]$, where a is given by

$$(2\pi)^{-1/2} \int_a^\infty e^{-t^2/2}\, dt = \alpha.$$

We shall refer to tests of this type as "the nonrandomized tests," and to the Wilcoxon test, based on the randomized S_{nm}, as "the randomized test."

Convenient choices for the norming factor s_{nm} are given, e.g., by

$$s_{nm}^2 = \sigma_{nm}^2 - \tfrac{1}{12} \textstyle\sum_k U_k V_k (W_k + 1), \tag{8.1}$$

or

$$s_{nm}^2 = \sigma_{nm}^2 - \frac{nm}{12(n+m)(n+m-1)} \sum_k W_k (W_k^2 - 1). \tag{8.2}$$

The norming factor given by (8.2), which is, by (7.4), the conditional standard deviation $\sigma(S'_{nm} \mid W)$ under the hypothesis, was suggested by Kruskal and Wallis [8]. In this case, Kruskal [7] proved that the conditional distribution of

T'_{nm} (given W) tends to $\mathfrak{N}(0, 1)$ if the W's are fixed vectors such that s_{nm}/σ_{nm} converges to a positive limit.

We now turn to the case $F \neq G$ (the alternative of the test). In the continuous case, it has been shown by Mann-Whitney [10], van Dantzig [1], and Lehmann [9] that if m/n is held constant, then the Wilcoxon test is consistent (i.e., its power tends to 1 as $n \to \infty$) against all alternatives under which $P(X_1 > Y_1) > \frac{1}{2}$. We proceed to derive the analogous consistency property for the discontinuous case.

THEOREM 5. *Let $m/n = c$ be fixed, and*

$$P(X_1 > Y_1) + \tfrac{1}{2}P(X_1 = Y_1) > \tfrac{1}{2}. \tag{8.3}$$

Then the randomized test is consistent. If, moreover, the norming factor s_{nm} satisfies

$$\frac{s_{nm}}{\sigma_{nm}} \xrightarrow{P} h > 0, \tag{8.4}$$

then the nonrandomized test is also consistent.

REMARK. The condition (8.4) is always satisfied if s_{nm} is defined by either (8.1) or (8.2). For (8.1), we have

$$h^2 = 1 - \frac{1}{1 + c} \sum_k p_k q_k (p_k + c q_k),$$

and for (8.2),

$$h^2 = 1 - \frac{1}{(1 + c)^3} \sum_k (p_k + c q_k)^3,$$

and both quantities are positive, unless F and G are both degenerate and identical, which is obviously impossible under (8.3).

PROOF OF THEOREM. By (5.9) and Lemma 5.1 of [9], we have

$$\frac{S_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, b^2), \tag{8.5}$$

where $b = \lim_{n\to\infty} \sigma(S_{nm})/\sigma_{nm}$ is, by (7.6), a function of c and of the parameters θ^*, λ_1^*, and λ_2^*. The rejection region of the randomized test is $[T_{nm} > a]$, where $T_{nm} = (S_{nm} - \mu_{nm})/\sigma_{nm}$. But, by (7.5), we have

$$T_{nm} = \frac{S_{nm} - ES_{nm}}{\sigma_{nm}} - \frac{\theta^* nm}{\sigma_{nm}},$$

and, by (5.2), $nm/\sigma_{nm} \to \infty$ as $n \to \infty$. Hence, by (7.7), the randomized test is consistent.

Also, from (8.5) it follows, by Theorem 3, that

$$\frac{S'_{nm} - ES_{nm}}{\sigma_{nm}} \xrightarrow{L} \mathfrak{N}(0, \bar{b}^2),$$

where

$$\bar{b}^2 = b^2 - \frac{1}{1+c} \sum_k p_k q_k (p_k + c q_k).$$

Hence

$$\frac{S'_{nm} - ES_{nm}}{s_{nm}} \xrightarrow{L} \mathfrak{N}\left(0, \frac{\bar{b}^2}{\bar{h}^2}\right),$$

and the consistency of the nonrandomized test follows by the same argument as above, which completes the proof of the theorem.

9. Asymptotic efficiency. We shall now compare the randomized and the nonrandomized tests in terms of Pitman's concept of asymptotic relative efficiency. We shall restrict ourselves to the case of purely discontinuous distributions. Under a host of conditions (necessary to insure that the conditions of Theorem A are satisfied), it will be shown that the nonrandomized test is asymptotically more efficient than the randomized test, and that its asymptotic efficiency does not depend on the choice of the norming factor s_{nm}. The parameter θ will be $\theta^*(F, G) = P(X_1 > Y_1) + \frac{1}{2}P(X_1 = Y_1) - \frac{1}{2}$, and hence $\theta_0 = 0$.

LEMMA. *Let* Z_i $(i = 1, 2, \cdots, r)$ *be* $\mathfrak{B}(na_i, p_i)$, *and put* $\bar{Z}_i = Z_i - na_i p_i$. *Then*

$$\prod_{i=1}^r Z_i = n^{r-1}\left[n + \sum_{i=1}^r \frac{\bar{Z}_i}{a_i p_i}\right] \prod_{i=1}^r a_i p_i + o_p(n^{r-1/2}).$$

Here the notation $f_n = o_p(g_n)$ stands, as usual, for $f_n/g_n \xrightarrow{P} 0$. The proof of the lemma consists of expanding the product $\prod Z_i = \prod(\bar{Z}_i + na_i p_i)$ and noting that $\bar{Z}_i/n^{1/2}$ converges in law (to a normal).

THEOREM 6. *Let* $m/n = c$ *be fixed, and* F *be a purely discontinuous distribution. Let* $\{G_\theta, 0 \leqq \theta \leqq \theta_1\}$ *be a family of purely discontinuous distributions having the same discontinuities* ξ_k *as* F, *with jumps* $q_k(\theta)$. *Let* $\{G_\theta\}$ *have the following properties*:

(1) $G_0 = F$;

(2) $q_k(\theta) > q > 0$;

(3) $\theta^*(F, G_\theta) = \theta$;

(4) *the convergence* $(S_{nm} - ES_{nm})/\sigma_{nm} \xrightarrow{L} \mathfrak{N}(0, b^2(\theta))$, *given by* (8.5), *is uniform in* θ;

(5) *the functions* $q_k(\theta)$ *are continuous at* $\theta = 0$.

Let $s_{nm} = s_{nm}(U, V)$ *be continuous functions of* U *and* V, *having, under* (F, G_θ), *finite variances and satisfying the following conditions*:

(6) $s_{nm}/n^{1/2} = \sum_k \alpha_k(\theta)\bar{U}_k + \sum_k \beta_k(\theta)\bar{V}_k + \gamma(\theta)n + o_p(n^{1/2})$, *where*
$\bar{U}_k = U_k - np_k$, $\bar{V}_k - mq_k(\theta)$;

(7) $\gamma^2(0) = (c(1 + c)/12)(1 - \sum_k p_k^3)$:

(8) $\gamma(\theta)$ *is differentiable, and* $\gamma'(\theta)$ *is continuous at* $\theta = 0$;

(9) *at least one of the* $2K$ *inequalities* $c\theta\alpha_k(\theta) \neq \gamma(\theta)\bar{\alpha}_k(\theta)$, $c\theta\beta_k(\theta) \neq \gamma(\theta)\bar{\beta}_k$ *holds, where (arranging the* ξ_k*'s so that* $\xi_k < \xi_{k+1}$*)*

$$\bar{\alpha}_k(\theta) = 1 + c[\textstyle\sum_{j<k} q_j(\theta) + \frac{1}{2}q_k(\theta)], \qquad \bar{\beta}_k = \textstyle\sum_{j>k} p_j + \frac{1}{2}p_k .$$

Under these conditions, the asymptotic relative efficiency of the randomized test with respect to the nonrandomized test is $1 - \sum_k p_k^3$.

REMARKS. (i) Conditions (6) and (9) are satisfied if s_{nm} is given either by (8.1) or by (8.2). For (8.1), for example, using the lemma in this section, we have

$$\frac{12}{c}\frac{s_{nm}^2}{n^3} = 1 + c - \sum_k p_k q_k(\theta)[p_k + cq_k(\theta)] - \sum_k q_k(\theta)[2p_k + cq_k(\theta)]\frac{\bar{U}_k}{n} - \sum_k p_k[p_k + 2cq_k(\theta)]\frac{\bar{V}_k}{m} + o_p(n^{-1/2}),$$

and using the Taylor expansion for $(a + bx)^{1/2}$ we get (6). The same method works for (8.2).

(ii) Condition (7) is necessary, by Theorem 4, to make s_{nm} an admissible norming factor. It is satisfied for the choices (8.1) and (8.2) if

$$\left[\frac{d}{d\theta}\left(1 - \sum_k q_k^3(\theta)\right)\right]_{\theta=0} = 0,$$

which is analogous to the condition $q_0 = p_0$ in Theorem 2.

PROOF OF THEOREM. In terms of Theorem A, put

$$\bar{S}_{1n} = T'_{nm}, \qquad \bar{S}_{2n} = S_{nm} - \mu_{nm},$$

$$\psi_{1n}(\theta) = \frac{c\theta}{\gamma(\theta)} n^{1/2}, \qquad \psi_{2n}(\theta) = \theta nm,$$

$$\sigma_{1n}(\theta) = \frac{c(1 + c)}{12n^3\gamma^2(\theta)} \sigma([S'_{nm} - \psi_{1n}(\theta)s_{nm}] \mid \theta),$$

$$\sigma_{2n}(\theta) = \sigma(S_{nm} \mid \theta).$$

We have then

$$H_1^2(n) = \frac{12cn}{(1 + c)(1 - \sum_k p_k^3)}, \qquad H_2^2(n) = \frac{n^2m^2}{\sigma_{nm}^2}.$$

The verification of (2.1)–(2.3) is routine, and we proceed to verify (2.4). The convergences involved are all uniform in θ; except for the one required by condition (4), this follows from condition (2) and Theorem B. (All the usual binomial convergences, when put in terms of characteristic functions, are seen to be uniform as long as the probability parameters are bounded away from 0 and 1.)

We have

$$\frac{\bar{S}_{2n} - \psi_{2n}(\theta)}{\sigma_{2n}(\theta)} = \frac{S_{nm} - ES_{nm}}{\sigma(S_{nm})},$$

which, by (5.9), verifies (2.4) for $i = 2$. Also, we have

$$\frac{S_{1n} - \psi_{1n}(\theta)}{\sigma_{1n}(\theta)} = \frac{S'_{nm} - \mu_{nm} - \psi_{1n}(\theta)s_{nm}}{\sigma_{1n}(\theta)s_{nm}}.$$

Arranging the ξ_k's in ascending order of magnitude, we have

$$S'_{nm} = \sum_k U_k \left(\sum_{j<k} W_j + \frac{W_k + 1}{2} \right)$$

It follows, by the lemma in this section, that

$$\frac{S'_{nm}}{n} = \sum_k \bar{\alpha}_k(\theta)\bar{U}_k + \sum_k \bar{\beta}_k \bar{V}_k + n\bar{\gamma}(\theta) + o_p(n^{1/2}),$$

where

$$\bar{\gamma}(\theta) = \textstyle\sum_k p_k\{\sum_{j<k} [p_j + cq_j(\theta)] + \frac{1}{2}[p_k + cq_k(\theta)]\}.$$

Hence

$$\frac{1}{n}[S'_{nm} - \mu_{nm} - \psi_{1n}(\theta)s_{nm}] = \sum_k \left[\bar{\alpha}_k(\theta) - \frac{c\theta}{\gamma(\theta)}\alpha_k(\theta) \right] \bar{U}_k + \sum_k \left[\bar{\beta}_k - \frac{c\theta}{\gamma(\theta)}\beta_k(\theta) \right] \bar{V}_k + o_p(n^{1/2}).$$

Because of (9), this expression is asymptotically normal, and it is easily shown that (2.4) holds for $i = 1$. Hence, by Theorem A, the asymptotic relative efficiency of the randomized test with respect to the nonrandomized test is

$$\lim_{n\to\infty} \frac{H_2^2(n)}{H_1^2(n)} = 1 - \sum_k p_k^3,$$

and the theorem is proved.

10. Acknowledgment. The author wishes to thank Professor E. L. Lehmann for his suggestion of the problem and for his helpful interest in the work.

REFERENCES

[1] D. van Dantzig, "On the consistency and the power of Wilcoxon's two-sample test," *Nederl. Akad. Wetensch., Proc.*, Vol. 54 (1951), pp. 1–8.

[2] W. J. Dixon and F. J. Massey, Jr., *An introduction to statistical analysis*, McGraw-Hill Book Co., 1951, p. 248.

[3] W. J. Dixon and A. M. Mood, "The statistical sign test," *J. Amer. Stat. Assn.*, Vol. 41 (1946), pp. 557–566.

[4] J. Hemelrijk, "A theorem on the sign test when ties are present," *Nederl. Akad. Wetensch., Proc.*, Vol. 55 (1952), pp. 322–326.

[5] J. Hemelrijk, "Note on Wilcoxon's two sample test when ties are present," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 133–135.

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

[7] W. H. KRUSKAL, "A nonparametric test for the several-sample problem," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 525–540.

[8] W. H. KRUSKAL AND W. A. WALLIS, "Use of ranks in one-criterion variance analysis," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 583–621.

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

[10] H. B. MANN AND D. 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.

[11] G. E. NOETHER, "Asymptotic properties of the Wald-Wolfowitz test of randomness," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 231–246.

[12] E. PARZEN, "On uniform convergence of families of sequences of random variables," *Univ. California Publ. Stat.*, Vol. 2 (1954), pp. 23–54.

[13] *Probability Tables for the Wilcoxon Test When There are Ties*, National Bureau of Standards Report No. 1859, Washington, D. C.