COMPARISON OF THE NORMAL SCORES AND WILCOXON TESTS

J. L. HODGES, JR. AND E. L. LEHMANN
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

For testing the equality of two distributions F and G on the basis of samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ from these distributions, a number of procedures are available. If the tests are to be powerful against shift alternatives given by

$$G(y) = F(y - \theta), \tag{1.1}$$

the most commonly proposed tests are

(a) Student's t test;

(b) Wilcoxon's two-sample test based on the sum $s_1 + \cdots + s_n$ of ranks of the Y's;

(c) The Normal scores test. This test has been proposed in two asymptotically equivalent versions, the test statistic in both cases being of the form

$$h(s_1) + \cdots + h(s_n) \tag{1.2}$$

with large values significant against the alternatives $\theta > 0$.

(i) The function

$$h(s) = E(W^{(s)}), \tag{1.3}$$

where $W^{(1)} < \cdots < W^{(m+n)}$ are the order statistics of a sample of size $m + n$ from a standard normal distribution was introduced by Fisher and Yates in the introduction to table XX of [3], who also gave a table of (1.3). These authors propose replacing the variables X_i and Y_j in the t-statistic by the function (1.3) of their ranks and applying to these values the usual analysis of variance, which amounts to using as critical value that appropriate to the t-test. The corresponding rank test (in which the critical value is obtained from the distribution of ranks rather than that of t), was proposed by Hoeffding [5] and discussed further by Terry [8], who also gave a table of percentage points.

(ii) The closely related function

$$h(s) = \Phi^{-1}\left(\frac{s}{m + n + 1}\right), \tag{1.4}$$

where Φ denotes the cumulative distribution function of the standard normal

This research was sponsored in part by the Office of Naval Research under Contracts. Nonr-222(43) and 2842(00).

distribution, was proposed by van der Waerden [9]. Further discussion of this test and tables of percentage points were given by van der Waerden and Nievergelt [10]. The asymptotic equivalence of (i) and (ii) was established for a wide class of cases by Chernoff and Savage [2].

A convenient method for making large-sample comparisons between two tests was developed by Pitman [7], who defined the relative asymptotic efficiency of two sequences of tests as the limiting inverse ratio of sample sizes necessary to achieve the same power β against the same alternative at the same significance level α and showed that for a large class of problems this efficiency is independent of α and β. For details see section 2.

It was shown by Pitman that the relative asymptotic efficiency $e_{W,t}(F)$ of the Wilcoxon to the t-test against the alternatives (1.1) is $3/\pi \sim .955$ when F is a normal distribution, and in [4] that for all F (throughout the paper we shall assume that all distributions F considered are continuous)

$$e_{W,t}(F) \geqq .864, \tag{1.5}$$

the lower bound being attained for a distribution with parabolic density.

The relative asymptotic efficiency $e_{N,t}(F)$ of the normal scores test to the t-test against the alternatives (1.1) with F a normal distribution is well known to equal 1. It was proved by Chernoff and Savage [2] that

$$e_{N,t}(F) \geqq 1 \tag{1.6}$$

for all F satisfying certain regularity conditions. The method used in [4] to prove (1.5) can be extended in the present case to show that (1.6) holds in fact for all F.

While then neither $e_{W,t}$ nor $e_{N,t}$ can ever be very small, it is easy to find distributions for which they are arbitrarily large or even infinite.

The above results suggest that on the basis of power, at least for large samples, both the Wilcoxon and normal scores test are preferable to the t-test for general use. The purpose of the present paper is to determine, on the same basis, which of the two rank tests is preferable in various circumstances.

2. A basic limit formula

The comparison of two tests of a hypothesis $\theta = 0$ on the basis of Pitman's asymptotic relative efficiency is usually made (see for example, Noether [6]) for sequences of test statistics $\{S_N\}$, $\{T_N\}$, whose expectations

$$\mu(\theta) = E_\theta(S_N), \qquad \nu(\theta) = E_\theta(T_N) \tag{2.1}$$

(or other appropriate norming constants) are assumed to have finite derivatives $\mu'(0)$, $\nu'(0)$ at $\theta = 0$. For the comparison we wish to make here, such derivatives do not always exist or may be infinite, and we begin therefore with a slight generalization of the usual approach.

LEMMA 1. *Let* $\{U_N\}$ *be a sequence of test statistics,* $\{a_N\}$ *a sequence of positive* *numbers, and* $\{\eta_N\}$ *a sequence of real-valued functions such that, when* θ_N *is the true parameter value,*

$$\frac{U_N - \eta_N(\theta_N)}{a_N} \tag{2.2}$$

tends in law to the standard normal distribution for every sequence $\{\theta_N\}$ *with* $\theta_N \to 0$. *Consider any sequence of critical regions*

$$\frac{U_N - \eta_N(0)}{a_N} > k_N, \tag{2.3}$$

where

$$k_N \to k = \Phi^{-1}(1 - \alpha), \tag{2.4}$$

that is, any sequence (2.3) *which asymptotically has significance level* α. *Let* $\beta_N(\theta)$ *denote the power of the Nth test of the sequence against the alternative* θ *and let* $\{\theta_N\}$ *be any sequence tending to* 0.

Then

$$\beta_N(\theta_N) \to \beta \tag{2.5}$$

if and only if

$$\frac{\eta_N(\theta_N) - \eta_N(0)}{a_N} \to c, \tag{2.6}$$

where

$$\beta = 1 - \Phi(k - c). \tag{2.7}$$

That the limiting distribution of (2.2) is normal is sufficient for the problem to be considered here but clearly not necessary for the validity of the lemma, which only requires a continuous limiting distribution.

PROOF.

$$\begin{aligned} \beta_N(\theta_N) &= P_{\theta_N}\left\{\frac{U_N - \eta_N(0)}{a_N} > k_N\right\} \\ &= P_{\theta_N}\left\{\frac{U_N - \eta_N(\theta_N)}{a_N} > k_N - \frac{\eta_N(\theta_N) - \eta_N(0)}{a_N}\right\}. \end{aligned} \tag{2.8}$$

Since $[U_N - \eta_N(\theta_N)]/a_N$ has the continuous limit law Φ, the convergence is uniform. If therefore $[\eta_N(\theta_N) - \eta_N(0)]/a_N \to c$, $\beta_N(\theta_N) \to 1 - \Phi(k - c)$. Conversely, if $[\eta_N(\theta_N) - \eta_N(0)]/a_N$ does not converge, there exist two subsequences converging to different limits c_1 and c_2. For these subsequences, $\beta_N(\theta_N)$ would converge to different values β_1 and β_2, which implies that $\beta_N(\theta_N)$ does not converge.

Consider now two sequences of tests. Let S_N, T_N be two test statistics based on N observations, $N = 1, 2, \cdots$, and suppose that

$$\frac{S_N - \mu_N(\theta_N)}{b_N} \quad \text{and} \quad \frac{T_N - \nu_N(\theta_N)}{c_N} \tag{2.9}$$

tend to $N(0, 1)$ whenever $\theta_N \to 0$. To compare the tests, fix $0 < \alpha < \beta < 1$ and find any sequence $\theta_N \to 0$ such that the power $\beta_N(\theta_N)$ of the test based on T_N tends to β. For each N, let an integer r_N be determined so that S_{r_N} performs about as well as T_N; more specifically, so that the power $\beta_N^*(\theta_N)$ of the test based on S_{r_N} also tends to β. (Note that β_N^* is based on r_N rather than N observations.)

LEMMA 2. *Under the above assumptions,*

$$\frac{\nu_N(\theta_N) - \nu_N(0)}{\mu_{r_N}(\theta_N) - \mu_{r_N}(0)} \frac{b_{r_N}}{c_{r_N}} \to 1 \quad \text{as} \quad N \to \infty. \tag{2.10}$$

PROOF. By assumption, $\beta_N(\theta_N) \to \beta$. Hence by lemma 1

$$\frac{\nu_N(\theta_N) - \nu_N(0)}{c_N} \to c = \Phi^{-1}(1 - \alpha) - \Phi^{-1}(1 - \beta). \tag{2.11}$$

Here $0 < c < \infty$ since $0 < \alpha < \beta < 1$. The corresponding application of lemma 1 with S_{r_N} playing the role of U_N, shows that

$$\frac{\mu_{r_N}(\theta_N) - \mu_{r_N}(0)}{b_{r_N}} \to c \tag{2.12}$$

and the result follows.

In applications of the above lemma to the two-sample problem, the total sample size N is the sum of the individual sample sizes

$$N = m_N + n_N. \tag{2.13}$$

It turns out, however, that for the statistics to be considered the assumed limiting distribution for S_N is valid regardless how the total sample size N is divided between m_N and n_N provided only m_N/n_N is bounded away from 0 and ∞ as $N \to \infty$. Since the corresponding remark also applies to the division of r_N into its two components m_N' and n_N', lemma 2 remains valid with the only additional restriction

$$0 < \varliminf \frac{m_N}{n_N} \leqq \varlimsup \frac{m_N}{n_N} < \infty, \qquad 0 < \varliminf \frac{m_N'}{n_N'} \leqq \varlimsup \frac{m_N'}{n_N'} < \infty. \tag{2.14}$$

3. Bounds for $e_{W,N}(F)$

Equivalent to the Wilcoxon statistic is the statistic

$$S_N = \frac{1}{mn} \sum \sum W_{ij}, \tag{3.1}$$

where W_{ij} is 1 or 0 as $X_i < Y_j$ or $X_i > Y_j$. It was shown in [1] and [2] that $[S_N - \mu(\theta_N)]/b_N$ tends to the standard normal distribution if $\theta_N \to 0$ where

$$\mu(\theta) = P_\theta(X_i < Y_j) = \int F(x + \theta)\, dF(x) \tag{3.2}$$

and, if $m/N \to \lambda$,

$$b_N^2 = \frac{N+1}{12mn} \sim \frac{1}{12\lambda(1-\lambda)N}. \tag{3.3}$$

For the Normal scores statistic T_N given by (1.2) and (1.3), it was shown by Chernoff and Savage that $[T_N - \nu_N(\theta)]/c_N$ tends to the standard normal distribution if $\theta_N \to 0$, where

$$\nu_N(\theta) = \int \Phi^{-1}\left\{\frac{m}{N} F(x+\theta) + \frac{n}{N} F(x)\right\} dF(x) \tag{3.4}$$

and

$$c_N^2 = \frac{m}{nN} \sim \frac{\lambda}{(1-\lambda)N}. \tag{3.5}$$

Substituting in (2.10), it follows that

(3.6)

$$\frac{\int\left[\Phi^{-1}\left\{\frac{m}{N} F(x+\theta_N) + \frac{n}{N} F(x)\right\} - \Phi^{-1}\{F(x)\}\right] dF(x)}{\int [F(x+\theta_N) - F(x)]\, dF(x)} \frac{1}{\lambda\sqrt{12}}\left(\frac{N}{r_N}\right)^{1/2} \to 1.$$

As a first application of this formula, we shall exhibit a class of distributions F for which

$$e_{W,N}(F) = \lim \frac{N}{r_N} = 0, \tag{3.7}$$

so that the asymptotic efficiency of the Normal scores test relative to the Wilcoxon test is infinite. Consider a distribution F having a density f such that

$$\begin{aligned} f(x) &= 0 && \text{for } x < a \\ f(x) &\to d > 0 && \text{as } x \to a_+. \end{aligned} \tag{3.8}$$

We shall show that for such a distribution,

$$\lim_{\theta\to 0_+} \frac{1}{\theta} \int \left\{\Phi^{-1}\left[\frac{m}{N} F(x+\theta) + \frac{n}{N} F(x)\right] - \Phi^{-1}[F(x)]\right\} dF(x) = \infty. \tag{3.9}$$

From this and (3.6), (3.7) follows, provided only

$$\frac{1}{\theta}\int [F(x+\theta) - F(x)]\, dF(x) \tag{3.10}$$

is bounded away from ∞ as $\theta \to 0_+$, as is easy to check for example in the rectangular and exponential case.

By Fatou's theorem, it will suffice to show that

$$\int \lim_{\theta\to 0_+} \frac{1}{\theta}\left\{\Phi^{-1}\left[\frac{m}{N} F(x+\theta) + \frac{n}{N} F(x)\right] - \Phi^{-1}[F(x)]\right\} dF(x) = \infty. \tag{3.11}$$

Choose an interval $(a, a + \Delta)$ within which $f(x)$ is positive. As the integrand above is nonnegative, it will suffice to consider the integral over $(a, a + \Delta)$. Write

$$(3.12)\quad \frac{1}{\theta_N}\left\{\Phi^{-1}\left[\frac{m}{N}F(x+\theta_N)+\frac{n}{N}F(x)\right]-\Phi^{-1}[F(x)]\right\}$$

$$\equiv \frac{\Phi^{-1}\left[\frac{m}{N}F(x+\theta_N)+\frac{n}{N}F(x)\right]-\Phi^{-1}[F(x)]}{\left[\frac{m}{N}F(x+\theta_N)+\frac{n}{N}F(x)\right]-F(x)}\ \frac{\frac{m}{N}[F(x+\theta_N)-F(x)]}{\theta_N}.$$

As $N \to \infty$, the first factor on the right, which is a difference quotient, tends to

$$(3.13)\qquad \frac{d}{du}\Phi^{-1}(u)\Big|_{u=F(x)} = \frac{1}{\varphi\{\Phi^{-1}[F(x)]\}},$$

while the second factor tends to $\lambda f(x)$. We see that (3.9) is at least

$$(3.14)\qquad \int_a^{a+\Delta} \frac{\lambda f^2(x)\,dx}{\varphi\{\Phi^{-1}[F(x)]\}} = \int_{-\infty}^{\Phi^{-1}[F(a+\Delta)]} \lambda f(x)\,dy,$$

where $\Phi(y) = F(x)$. As $y \to \infty$, $x \to a_+$, and hence $f(x) \to d$. The last displayed integral is therefore infinite and (3.8) is proved.

It follows from these considerations that for a large class of distributions—including the rectangular and exponential—the Normal scores test at least for large samples is very much more efficient than the Wilcoxon test. This phenomenon is made intuitive by realizing that the Normal scores test pays in the limit infinitely more attention to the rankings at the extremes than does the Wilcoxon test. When F has a density that drops discontinuously to zero at either extreme, the rankings at that extreme will contain much more information than those in the central part where f is positive. We do not know how large n must be for the Normal scores test to be considerably better than the Wilcoxon for such a distribution as the rectangular.

Can we similarly find distribution F for which $e_{W,N}(F) = \infty$, so that the Wilcoxon test is very much more efficient than the Normal scores test? To see that this is not possible, let us note that by the mean value theorem, $[\Phi^{-1}(v) - \Phi^{-1}(u)]/(v-u)$ is equal to $1/\varphi[\Phi^{-1}(\xi)]$ for some intermediate ξ and that therefore

$$(3.15)\qquad \Phi^{-1}(v) - \Phi^{-1}(u) \geqq \sqrt{2\pi}(v-u) \quad \text{for all} \quad u < v.$$

Applying this inequality to the numerator of the first fraction in (3.6), this numerator is seen to be $\geqq$

$$(3.16)\qquad \frac{m}{N}\sqrt{2\pi}\int [F(x+\theta_N) - F(x)]\,dF(x)$$

so that the first factor of (3.6) is at least $m\sqrt{2\pi}/N$. It therefore follows from (3.6) that

$$(3.17)\qquad \overline{\lim_{N\to\infty}}\ \frac{m\sqrt{2\pi}}{N\lambda\sqrt{12}}\left(\frac{N}{r_N}\right)^{1/2} \leqq 1,$$

and since $m/\lambda N \to 1$ that

(3.18) $$\lim \frac{N}{r_N} \leqq \frac{12}{2\pi} = \frac{6}{\pi} \sim 1.91.$$

The relative asymptotic efficiency of the Wilcoxon to the Normal scores test therefore satisfies

(3.19) $$e_{W,N}(F) \leqq \frac{6}{\pi} \sim 1.91 \qquad \text{for all} \quad F.$$

4. The range of $e_{W,N}(F)$

In order to compute $e_{W,N}(F)$ for various distributions F, we need to obtain the limit of the first factor in (3.6). Dividing numerator and denominator by θ_N, and assuming that we can differentiate under the integral sign, (3.6) leads to the result, essentially contained in the work of Chernoff and Savage, that

(4.1) $$e_{W,N}(F) = 12 \left(\frac{\int f^2(x)\,dx}{\int \frac{f^2(x)\,dx}{\varphi\{\Phi^{-1}[F(x)]\}}} \right)^2,$$

where f is the density of F.

The following lemma gives conditions under which the limit passage under the integral sign is permissible in the denominator and the numerator respectively.

LEMMA 3.

(a) *Let F be a continuous cumulative distribution function, differentiable in each of the open intervals $(-\infty, a_1), (a_1, a_2), \cdots, (a_{s-1}, a_s), (a_s, \infty)$. If the derivative f of F is bounded in each of these intervals, then*

(4.2) $$\lim_{\theta \to 0} \int \frac{1}{\theta} [F(x + \theta) - F(x)] f(x)\,dx = \int f^2(x)\,dx.$$

(b) *If in addition to the assumptions made under* (a), *the function*

$$f(x)/\varphi\{\Phi^{-1}[F(x)]\}$$

is bounded as $x \to \pm\infty$, then

(4.3)

$$\lim_{\theta \to 0} \frac{1}{\theta} \int \left\{ \Phi^{-1}\left[\frac{m}{N} F(x + \theta) + \frac{n}{N} F(x)\right] - \Phi^{-1}[F(x)] \right\} dF(x) = \int \frac{f^2(x)\,dx}{\varphi\{\Phi^{-1}[F(x)]\}}$$

and the convergence is uniform in m, n, and N.

PROOF. (a) Consider first the case $s = 0$, so that F is differentiable for all x with a derivative f which is never greater than M. Then for any $\theta > 0$

(4.4) $$\frac{F(x + \theta) - F(x)}{\theta} \leqq M$$

by the mean value theorem. Any sequence of integrands

$$f(x)[F(x + \theta_N) - F(x)]/\theta_N$$

is therefore bounded by the integrable function $Mf(x)$, and the result follows from the bounded convergence theorem. If $s = 1$, and the bound on f is M_1 in $(-\infty, a_1)$ and M_2 in (a_1, ∞), we have that for any $\theta > 0$

$$\frac{1}{\theta}[F(x+\theta) - F(x)] \leqq \begin{cases} M_1, & x + \theta \leqq a_1, \\ M_2, & x \geqq a_1, \\ \max(M_1, M_2), & x \leqq a_1 \leqq x + \theta. \end{cases} \tag{4.5}$$

Hence

$$\frac{1}{\theta}[F(x+\theta) - F(x)]f(x) \leqq \max(M_1, M_2)f(x) \tag{4.6}$$

for all x and $\theta > 0$, and the result follows as before. The proof clearly extends to any s and to the case $\theta < 0$.

(b) Consider again first the case $s = 0$, and let

$$g(\theta) = \Phi^{-1}\left[\frac{m}{N}F(x+\theta) + \frac{n}{N}F(x)\right]. \tag{4.7}$$

Then for any $\theta > 0$, $g(\theta) \leqq \Phi^{-1}[F(x+\theta)]$ and

$$\frac{g(\theta) - g(0)}{\theta} \leqq \frac{\Phi^{-1}[F(x+\theta)] - \Phi^{-1}[F(x)]}{\theta}. \tag{4.8}$$

By the mean value theorem, the right side of this inequality is equal to $f(x+\xi)/\varphi\{\Phi^{-1}[F(x+\xi)]\}$ for some $0 < \xi < \theta$. This is bounded for $x + \xi$ in any finite interval since f is bounded, and near $\pm\infty$ by assumption (b).

The extension to the case $s > 0$ is quite similar to that in (a). If for example $s = 1$, then for any $x < a < y$

$$\begin{aligned} &\frac{\Phi^{-1}[F(y)] - \Phi^{-1}[F(x)]}{y - x} \\ &\qquad \leqq \max\left\{\frac{\Phi^{-1}[F(y)] - \Phi^{-1}[F(a)]}{y - a}, \frac{\Phi^{-1}[F(a)] - \Phi^{-1}[F(x)]}{a - x}\right\} \end{aligned} \tag{4.9}$$

and the argument for $s = 0$ can be applied to each of the fractions under the maximum sign.

When the regularity conditions of (a) and (b) are satisfied, we see—by dividing the numerator and the denominator of the first fraction in (3.6) by θ_N—that $\lim (N/r_N)$ and hence $e_{W,N}(F)$ is given by (4.1).

Using this formula, we shall now show that all values in the open interval $(0, 6/\pi)$ are taken on by $e_{W,N}(F)$ for suitable F. For this purpose consider the following family of distributions

$$F_{a,\epsilon}(x) = \begin{cases} \Phi(x), & 0 \leqq x \leqq \epsilon, \\ \Phi(y), & x > \epsilon, \end{cases} \tag{4.10}$$

where $y = \epsilon + a(x - \epsilon)$, and $F_{a,\epsilon}$ is defined by symmetry for $x \leqq 0$. The associated density is

$$f_{a,\epsilon}(x) = \begin{cases} \varphi(x) & \text{for } |x| \leqq \epsilon, \\ a\varphi(y) & \text{for } |x| > \epsilon. \end{cases} \tag{4.11}$$

To check the regularity conditions, we note that F satisfies condition (a) with $s = 1$. Condition (b) is trivially satisfied since

$$\frac{f(x)}{\varphi\{\Phi^{-1}[F(x)]\}} = \begin{cases} \varphi(x)/\varphi(x) = 1 & \text{for } |x| \leqq \epsilon, \\ a\varphi(y)/\varphi(y) = a & \text{for } |x| > \epsilon. \end{cases} \tag{4.12}$$

In order to obtain the efficiency $e_{W,N}(F)$, we compute

$$\int f^2(x)\,dx = \frac{1}{2\sqrt{\pi}}\left\{\left[\Phi(\epsilon\sqrt{2}) - \frac{1}{2}\right] + a[1 - \Phi(\epsilon\sqrt{2})]\right\} \tag{4.13}$$

and

$$\int \frac{f^2(x)\,dx}{\varphi\{\Phi^{-1}[F(x)]\}} = \left[\Phi(\epsilon) - \frac{1}{2}\right] + a[1 - \Phi(\epsilon)], \tag{4.14}$$

to get

$$e_{W,N}(F_{a,\epsilon}) = \frac{3}{\pi}\left(\frac{\left[\Phi(\epsilon\sqrt{2}) - \frac{1}{2}\right] + a[1 - \Phi(\epsilon\sqrt{2})]}{\left[\Phi(\epsilon) - \frac{1}{2}\right] + a[1 - \Phi(\epsilon)]}\right)^2. \tag{4.15}$$

As $a \to \infty$, this tends to

$$\frac{3}{\pi}\left(\frac{1 - \Phi(\epsilon\sqrt{2})}{1 - \Phi(\epsilon)}\right)^2 \tag{4.16}$$

which tends to 0 as $\epsilon \to \infty$. On the other hand, as $a \to 0$, $e_{W,N}(F_{a,\epsilon})$ tends to

$$\frac{3}{\pi}\left(\frac{\Phi(\epsilon\sqrt{2}) - \frac{1}{2}}{\Phi(\epsilon) - \frac{1}{2}}\right)^2 \tag{4.17}$$

which tends to $6/\pi$ as $\epsilon \to 0$. By continuity, all intermediate values are also taken on.

The only question concerning the range of $e_{W,N}(F)$ that still remains in doubt is whether the upper end point $6/\pi$ is attained. We shall now show that this is the case for all distributions F possessing an even density f such that

$$\lim_{x \to 0} f(x) = \infty \tag{4.18}$$

and such that for all $A > 0$

$$\int_0^A f^2(x)\,dx = \infty, \qquad \int_A^\infty f^2(x)\,dx < \infty, \qquad \int_A^\infty \frac{f^2(x)\,dx}{\varphi\{\Phi^{-1}[F(x)]\}} < \infty. \tag{4.19}$$

We have shown at the end of section 3 that the first factor of (3.6) is always at least $m\sqrt{2\pi}/N$ and from this deduced that $e_{W,N}(F) \leqq 6/\pi$. For a distribution F satisfying the stated conditions we shall now prove that the first factor of (3.6) can be made less than $(m/N)\sqrt{2\pi}(1+\epsilon)$ for any given $\epsilon > 0$ by choosing θ_N sufficiently small. Together with the earlier inequality this shows that the first factor tends to $\lambda\sqrt{2\pi}$ as $N \to \infty$ and hence that $e_{W,N}(F) = 6/\pi$.

To prove the desired inequality, let $\epsilon > 0$ and let A, θ be positive numbers so small that $0 \leqq x < x + \theta \leqq A$ implies

$$\frac{\Phi^{-1}\left[\frac{m}{N}F(x+\theta) + \frac{n}{N}F(x)\right] - \Phi^{-1}\{F(x)\}}{\frac{m}{N}[F(x+\theta) - F(x)]} \leqq \left(1 + \frac{\epsilon}{2}\right)\sqrt{2\pi}. \tag{4.20}$$

This can be achieved since the left side by the mean value theorem is equal to $1/\varphi\{\Phi^{-1}[F(\xi)]\}$ for some ξ between x and $x + \theta$. The first factor of (3.6) with θ instead of θ_N is then at most

$$\frac{\left[\frac{m}{N}\left(1+\frac{\epsilon}{2}\right)\sqrt{2\pi}\int_0^A [F(x+\theta) - F(x)]\,dF(x) + \int_A^\infty [\Phi^{-1}\{F(x+\theta)\} - \Phi^{-1}\{F(x)\}]\,dF(x)\right]}{\int_0^A [F(x+\theta) - F(x)]\,dF(x) + \int_A^\infty [F(x+\theta) - F(x)]\,dF(x)}. \tag{4.21}$$

As $\theta \to 0$, we have by assumption that

$$\frac{1}{\theta}\int_0^A [F(x+\theta) - F(x)]\,dF(x) \geqq \int_0^A \lim \frac{1}{\theta}[F(x+\theta) - F(x)]\,dF(x) = \infty, \tag{4.22}$$

while the integrals from A to ∞ in the numerator and the denominator remain bounded. It follows that for θ sufficiently small, the quantity (4.21) is less than $(m/N)(1+\epsilon)\sqrt{2\pi}$, and hence that for N sufficiently large, the first factor of (3.6) is also less than $(m/N)(1+\epsilon)\sqrt{2\pi}$, as was to be proved.

5. Choice between the tests

We conclude with an attempt to throw some light on the question of choice between the two tests. Table I shows values of $e_{W,N}(F)$ for a number of well-known distributions F.

TABLE I

F	$e_{W,N}(F)$
Rectangular	0
Exponential	0
Normal	$3/\pi \sim .955$
Logistic	$\pi/3 \sim 1.05$
Double exponential	$3\pi/8 \sim 1.18$
Cauchy	1.413

Qualitatively we may venture to guess that the Normal scores test is preferable when the distribution has an abrupt tail, like the rectangular; that they are about equally good with a bell-shaped density with a thin tail; and that the Wilcoxon test will perform relatively better when the tails are heavy so that the information is mainly to be found in the central rankings. We note that the sequence (4.10) used to give $e_{W,N} \to 1.91$ is one with very heavy tails.

These considerations suggest that a normal distribution contaminated with gross errors might favor the Wilcoxon. To investigate this, let us consider an F_ρ obtained by mixing a standard normal with a small fraction ρ of a normal with expectation 0 and standard deviation 3. By numerical integration we find the values for $\rho = .01$ and $\rho = .02$, as shown in table II.

TABLE II

ρ	$e_{W,N}(F_\rho)$
0	.955
.01	.979
.02	.997

It appears that the Wilcoxon test will be better in large samples if something like one observation in forty is a "gross error" as defined above.

REFERENCES

[1] F. C. ANDREWS, "Asymptotic behavior of some rank tests for analysis of variance," *Ann. Math. Statist.*, Vol. 25 (1954), pp. 724–735.

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

[3] R. A. FISHER and F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research*, Edinburgh and London, Oliver and Boyd, 1938.

[4] J. L. HODGES, JR. and E. L. LEHMANN, "The efficiency of some nonparametric competitors of the *t*-test," *Ann. Math. Statist.*, Vol. 27 (1956), pp. 324–335.

[5] W. HOEFFDING, " 'Optimum' nonparametric tests," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 83–92.

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

[7] E. J. G. PITMAN, Lecture notes on nonparametric statistical inference, given at Columbia University, 1949 (unpublished).

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

[9] B. L. VAN DER WAERDEN, "Ein neuer Test für das Problem der zwei Stichproben," *Math. Ann.*, Vol. 126 (1953), pp. 93–107.

[10] B. L. VAN DER WAERDEN and E. NIEVERGELT, *Tables for Comparing Two Samples by X-test and Sign Test*, Berlin, Springer, 1956.