THE EFFICIENCY OF SOME NONPARAMETRIC COMPETITORS OF THE t-TEST

BY J. L. HODGES, JR., AND E. L. LEHMANN[1]

University of California, Berkeley

0. Summary. Consider samples from continuous distributions $F(x)$ and $F(x - \theta)$. We may test the hypothesis $\theta = 0$ by using the two-sample Wilcoxon test. We show in Section 1 that its asymptotic Pitman efficiency, relative to the t-test, never falls below 0.864. This result also holds for the Kruskal-Wallis test compared with the F-test, and for testing the location parameter of a single symmetric distribution.

A number of alternative notions of asymptotic efficiency are compared in Section 2. In this connection, certain difficulties arise because power is not necessarily a convex function of sample size. As an alternative to the Pitman notion of asymptotic efficiency, we consider in Section 3 one based on the speed with which power at a fixed alternative tends to 1. In particular we obtain, for the sign test relative to the t in normal populations, the limit as $n \to \infty$ of the sequence of power efficiency functions. It is noted that certain interchanges of limit passages are not always possible.

1. Minimum Pitman efficiency of the Wilcoxon and sign tests. For comparing the large sample power of two sequences of tests, the concept of asymptotic relative efficiency was developed by Pitman [1]. An exposition of his work, together with some extensions, was recently given in [2] and [3]. Applications to a number of specific problems are made in [4] and [5].

Let $\beta_N(\theta)$ and $\beta_N^*(\theta)$ denote the power functions of two tests, say A and A^*, based on the same set of N observations, against a parametric family of alternatives labeled by θ, and let θ_0 be the value of θ specified by the hypothesis. We shall assume that all tests are at level of significance α. Let β be a specified power with $\alpha < \beta < 1$. Consider a sequence of alternatives θ_N such that

$$\beta_N(\theta_N) \to \beta, \qquad \text{as } N \to \infty, \tag{1.1}$$

and a sequence $N^* = h(N)$ such that

$$\beta_{N^*}^*(\theta_N) \to \beta, \qquad \text{as } N \to \infty. \tag{1.2}$$

Then if

$$e_{A^*,A} = \lim_{N\to\infty} \frac{N}{N^*} \tag{1.3}$$

Received February 25, 1955.

[1] Prepared with partial support of the Office of Naval Research. This paper may be reproduced in whole or in part for any purpose of the United States Government.

exists, and is independent of α, β and the particular sequences $\{\theta_N\}$ and $\{h(N)\}$ chosen, then $e_{A^*,A}$ is defined to be the *relative asymptotic efficiency* of the test A^* with respect to the test A. Under weak assumptions (1.1) implies that $\theta_N \to \theta_0$, and in the most common cases it turns out that θ_N tends to θ_0 at the rate $N^{-1/2}$. Usually the N observations constitute a sample, or are divided into two samples of sizes m and n with $m + n = N$. In the latter case we assume that m/n tends to some limit ρ, $0 < \rho < \infty$, as N tends to ∞. In many problems, including those we study, $e_{A^*,A}$ is independent of ρ.

Pitman gave a method for obtaining the limit (1.3), and evaluated it for a number of problems. Consider in particular the case of samples $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ from continuous distributions F and G and the hypothesis $H: F = G$. We shall be concerned with the narrower alternatives that G differs from F only by a shift, so that $G(u) = F(u - \theta)$ for all u. The discussion applies to both the one-sided case $\theta > \theta_0 = 0$ and the two-sided case $\theta \neq \theta_0 = 0$. If F is a normal distribution, the appropriate test is Student's t-test. A nonparametric test proposed by Wilcoxon is based on the rank sum of the Y's among the set of N-ordered observations. Pitman computed the relative asymptotic efficiency of the Wilcoxon test relative to the t-test as

$$(1.4) \qquad e_{w,t} = 12\sigma^2 \left[\int f^2(x)\, dx \right]^2,$$

where f is the probability density of the distribution F, and σ^2 is the common variance of the X's and Y's. Some particular values given by Pitman are $e_{w,t} = 3/\pi \sim .95$ when f is a normal density, $e_{w,t} = 1$ for the case of a uniform distribution, and $e_{w,t} = 81/64$ when $f(x) = x^2 e^{-x}/\Gamma(3)$ for $x \geqq 0$. All of these values are surprisingly high and raise the question as to how low e can actually drop. We shall prove, below, the following theorem.

THEOREM 1. *Let N^* satisfy* (1.2) *where the tests A and A^* are the* (*two-sample*) *t-test and Wilcoxon test, respectively, for testing against shift of a continuous distribution F. Then* (*a*)

$$(1.5) \qquad \liminf_{N\to\infty} N/N^* \geqq 108/125 = 0.864$$

whatever F may be.

Furthermore, (b) *the lower limit is attained for the distribution with density* (1.12, 1.13). *for which e = .864.*

PROOF. It was shown by Andrews [5] that if F is continuous, and

$$(1.6) \qquad \lim_{\theta\to\infty} \int \frac{1}{\theta} [F(x + \theta) - F(x)]\, dF(x) = c,$$

then the efficiency, given by (1.3), exists and is $12c^2\sigma^2$. This proof also shows that quite generally

$$(1.7) \qquad \liminf_{N\to\infty} \frac{N}{N^*} \geqq 12\sigma^2 \left[\liminf_{\theta\to 0} \frac{1}{\theta} \int [F(x + \theta) - F(x)]\, dF(x) \right]^2.$$

By Fatou's lemma, the right-hand side is greater than or equal to

$$12\sigma^2\left\{\int\left[\liminf_{\theta\to 0}\frac{F(x+\theta)-F(x)}{\theta}\right]dF(x)\right\}^2. \tag{1.8}$$

It follows further, from the Decomposition Theorem of De La Vallée Poussin (see [6], p. 127), that when F has a singular component,

$$\lim\,[F(x+\theta)-F(x)]/\theta=\infty$$

on a set of positive F-measure, so that (1.8), and hence e, is infinite. We may therefore assume that except on a set of F-measure zero, the density $F(x) = F'(x)$ exists, in which case (1.8) becomes

$$12\sigma^2\left[\int f^2(x)\,dx\right]^2. \tag{1.9}$$

If $\sigma^2=\infty$, then it follows from (1.6) that $e=\infty$, so that we may assume σ^2 to be finite. Since (1.9) is invariant under a change of location or sale, we may take $\sigma^2 = 1$, and the problem of minimizing (1.9) then reduces to that of minimizing

$$\int f^2(x)\,dx \tag{1.10}$$

subject to the conditions

$$\int xf(x)\,dx=0;\qquad \int f(x)\,dx=\int x^2f(x)\,dx=1;\qquad f(x)\geqq 0 \text{ for all } x. \tag{1.11}$$

According to the method of undetermined multipliers, it is sufficient to minimize

$$\int [f^2(x)+2b(x^2-a^2)f(x)]\,dx.$$

For nonnegative f, this is achieved by setting

$$f(x)=b(a^2-x^2),\qquad \text{if } x^2\leqq a^2, \tag{1.12}$$

and $f(x) = 0$ otherwise. The constants a and b are determined from (1.11) to be

$$s=\sqrt{5},\qquad b=\tfrac{3}{20}\sqrt{5}, \tag{1.13}$$

and with these values, (1.9) becomes equal to 108/125, which is therefore a lower bound to (1.7). Since for the density (1.12) the limit of (1.6) may be taken under the integral sign, it is seen that the efficiency exists in this case, and equals the lower bound, which therefore cannot be improved.

To the extent that the above concept of efficiency adequately represents what happens for the sample sizes and alternatives arising in practice, this result shows that use of the Wilcoxon test instead of the Student's t-test can never entail a serious loss of efficiency for testing against shift. (On the other hand, it is obvious from (1.4) that the Wilcoxon test may be infinitely more efficient than the t-test.)

It should be mentioned that there are rank tests: that of Fisher and Yates, which has been discussed by Hoeffding [7], Terry [8], and Dwass [9]; and that of van der Waerden [10], for which the asymptotic efficiénty relative to the t-test is 1 when F is normal, and is conjectured to be >1 when F is not normal. Should this be correct, then for these tests the lower bound .864 in (1.5) would be replaced by the even better value 1.

The conclusion of Theorem 1 also applies to the H-test of Kruskal and Wallis [11] for testing equality of k distributions $F_1, \cdots, F_k$, which are assumed to differ only in location. This follows from the fact that Andrews' work, quoted above, was carried out for this more general problem, and that in particular formulae (1.6) and (1.7) hold for all values of k.

Another application is to the case of a single sample $X_1, \cdots, X_N$ from a distribution $F(x - \theta)$, where F is symmetric about 0. The hypothesis to be tested is $H:\theta = 0$, and if F is known to be normal, the one-sample t-test is appropriate. The Wilcoxon test for this problem is based on the rank sum of the positive X's among the ordered absolute values $|X_1|, \cdots, |X_N|$. Pitman showed that (1.4) also applies in this case, and the considerations of Andrews can be used to generalize this again to (1.6) and (1.7).

A particularly simple test of the hypothesis $H:\theta = 0$ in the one-sample problem is the sign test, based on the number of positive observations. For asymptotic efficiency of the sign test, relative to the t-test, Pitman obtained the result

$$e_{s,t} = 4\sigma^2 f^2(0), \tag{1.14}$$

which is valid whenever the derivative $F'_{(0)} = f(0)$ of F at the origin exists. A particular value given by Pitman is $e = 2/\pi$ in case of a normal distribution. In the present case there is, of course, no positive lower bound, since $e = 0$ when $f(0) = 0$. If the distribution F is assumed to possess a unimodal density (in the weak sense that $0 \leqq |x| < |x'|$ implies $f(x') \leqq f(x)$), then it is easily seen that $e \geqq \frac{1}{3}$, the value $\frac{1}{3}$ being attained for the case of a rectangular distribution. For let $f(0) = 1$ without loss of generality, since (1.14) is invariant under a change of scale. Then we must minimize

$$\int (x^2 - a^2)f(x)\,dx$$

subject to $0 \leqq f(x) \leqq 1$, and this is achieved by putting $f(x) = 1$ when $|x| \leqq a$ and $f(x) = 0$ otherwise.

It may be questioned whether the high efficiency of Wilcoxon relative to t established by Theorem 1 is the result of the particular alternatives considered. It is therefore of interest to make the comparison for other than shift alternatives. We shall now consider what may be called mixture or contamination alternatives. In the two-sample problem this takes the form

$$X_1, \cdots, X_m : F,$$
$$Y_1, \cdots, Y_n : (1 - \theta)F + \theta G.$$

In the one-sample problem, the form is

$$Z_1, \cdots, Z_N: (1 - \theta)F + \theta G, \qquad F(z) + F(-z) = 1.$$

In both cases we take $G \leqq F$ and test the hypothesis $\theta = 0$.

Mixture alternatives may be reasonable in many situations. For example, a treatment may be effective in only a proportion θ of the population of subjects. Thus, cancer operations are effective only if metastasis has not occurred; vitamin therapy is useful only if there is a vitamin deficiency.

If we let (m and) n tend to infinity (at the same rate), while θ tends to 0, with F and G fixed, we can compute the limiting efficiency of the Wilcoxon test relative to the t-test (or, equivalently, to the test based on $\bar{Y} - \bar{X}$) from Pitman's formula

$$e = \lim \left[\frac{\sigma_N(\theta_0)}{\sigma_N^*(\theta_0)} \frac{\mu_N^{*\prime}(\theta_0)}{\mu_N'(\theta_0)}\right]^2,$$

where T_N and T_N^* are the statistics on which the tests are based and it is assumed that

$$\frac{T_N - \mu_N(\theta)}{\sigma_N(\theta)}, \qquad \frac{T_N^* - \mu_N^*(\theta)}{\sigma_N^*(\theta)}$$

have the limiting distribution $N(0, 1)$. If $T_N = \bar{Y} - \bar{X}$ and mnT_N^* is the Mann-Whitney form of the Wilcoxon statistic, one obtains

$$\begin{aligned}
\sigma_N^{*2}(0) &= (1/12)(1/m + 1/n), \\
\sigma_N^2(0) &= \sigma^2(1/m + 1/n), \\
\mu_N^*(\theta) &= P\{X \leqq Y\} = (1 - \theta)\int F\,dF + \theta \int F\,dG, \\
\mu_N(\theta) &= E(Y) - E(X) = \theta\left[\int x\,dG - x\,dF\right].
\end{aligned}$$

It follows that

$$e = 12\sigma^2\left(\frac{\int (F - G)\,dF}{\int xd\,(G - F)}\right)^2 = 12\sigma^2\left(\frac{\int (F - G)\,dF}{\int (F - G)\,dx}\right)^2, \tag{1.15}$$

where, as before, σ^2 is the variance of an observation from F. The equality of the denominators in (1.15) follows by viewing each as an expression for the (signed) area between F and G. The above computation can be made rigorous by the methods of [5].

It is clear that $\int(F - G)\,dF \leqq \frac{1}{2}$, while $\int(F - G)\,dx$ may be arbitrarily large if G is far to the right of F. Thus (1.15) has no positive lower bound, which corresponds to our intuition that the Wilcoxon test, like any rank test, is insensitive to the size of large deviations.

If we particularize to $G(x) = F(x - \Delta)$, Δ fixed, we obtain

$$12\sigma^2\left[\int\left(\frac{F(x) - F(x-\Delta)}{\Delta}\right)dF(x)\right]^2. \tag{1.16}$$

As $\Delta \to 0$, (1.16) agrees (under suitable regularity conditions) with (1.4), so there is no finite upper bound to (1.15). We observe that

$$\int [F(x) - F(x - \Delta)]\, dF(x) = \Pr\{|X_1 - X_2| \leqq \Delta\},$$

and that $\Pr\{|X_1 - X_2| \leqq \Delta\}/\Delta$ will be a decreasing function of Δ whenever $X_1 - X_2$ is unimodal. Thus in particular, if X itself is unimodal, the efficiency decreases as Δ increases [12], so that the performance of Wilcoxon relative to t is often less good against contamination with a shift than against shift itself. For example, if F is normal and $\Delta = \sigma$, (1.16) has the value 0.812; for $\Delta = 2\sigma$ it is 0.533; and it tends to 0 as $\Delta \to \infty$.

We remark that the greater sensitivity of the t-test to contamination is not an unmixed blessing, as the contamination may, in some cases, represent gross errors of observation rather than the true effect of the treatment. In fact, insensitivity to large deviations is one of the advantages of nonparametric tests.

2. Alternative notions of asymptotic efficiency. The result obtained above suggests that if Pitman efficiency is taken as a guide, one may prefer the Wilcoxon test to the t-test in almost all problems of testing against shift. But how reliable is Pitman efficiency? Dixon [13], [14] has emphasized that a comprehensive efficiency comparison of two tests cannot be made with a single number. Suppose that a test A of level α and using N observations has power $\beta_A(N, \alpha, \theta)$ against alternative θ. If test A^*, also of level α, requires N^* observations to produce the same power at the same alternative, we define the efficiency of A^* relative to A in these circumstances to be the ratio N/N^*, and denote it by $e_{A^*,A}(N, \alpha, \theta)$. The complete comparison of A^* with A would require the evaluation of this "power efficiency function" for all values of its three arguments.

We note that the definition of N^* just given is not quite complete. There usually will not exist an integer N^* such that $\beta_{A^*}(N^*, \alpha, \theta) = \beta_A(N, \alpha, \theta)$, but rather an N_0 such that $\beta_{A^*}(N_0, \alpha, \theta) < \beta_A(N, \alpha, \theta) < \beta_{A^*}(N_0 + 1, \alpha, \theta)$. Dixon suggests that N^* be defined by inverse interpolation of $\beta_{A^*}(N^*, \alpha, \theta)$ as a function of N^*; specifically, he proposes polynomial interpolation of N^* against $\Phi^{-1}(\beta_{A^*}(N^*, \alpha, \theta))$ in [13], [14]. We feel that this method, while yielding "smooth" results, lacks any operational or functional meaning. Instead, we prefer to define N^* to be $N_0 + p$, where the test A^* has power $\beta_A(N, \alpha, \theta)$ if its number of observations is randomly chosen with probability p of being $N_0 + 1$ and probability $1 - p$ of being N_0. Thus, our N^* is the expected number of observations required with test A^* to match the power of test A, when randomizing between consecutive integers. (Our definition implies linear interpolation.)

We note in this connection a curious fact. For some tests, and specifically for the t-test against normal shift, $\beta(N, \alpha, \theta)$ is not always convex as a function of N.

Thus, if we wish to attain a stated power for stated α and δ with smallest expected number of observations, we would *not* randomize between consecutive integers! However, as our main objective is to define an N^* which gives the desired power, and the randomization is introduced only out of necessity, we shall use the definition given.

It might be felt that the question of the definition of N^* is too trivial to require so much discussion, and indeed if N is large this is so. But efficiency comparisons are often made for small N and here (especially with β large) the precise definition of N^* becomes important. To illustrate the point we present, below, the efficiency figures given by Dixon [14] for Wilcoxon against t for normal shift of amount δ, equal samples of 5, $\alpha = 4/126$, and the corresponding values as computed by our definition. (We are not able to obtain a worthwhile figure for $\delta = 4$, since the value of β_w is not given by Dixon to enough decimal places.) It is seen that Dixon's conclusion that the "power efficiency decreases slightly for more distant alternatives" is dependent on his method of interpolation for N^*. With our definition, the efficiency rises as

δ	.5	1	1.5	2	2.5	3	3.5	4	4.5
β_w	.072	.210	.431	.674	.858	.953	.988	.998	.9996
e (Dixon's paper)	.97	.97	.96	.95	.94	.94	.93	.92	.91
e (this paper)	.968	.978	.961	.956	.960	.960	.964	.976 ± .01	—

δ is increased beyond about 3, and appears never to fall below about 0.96, while the efficiency as computed by Dixon reaches .91 at $\delta = 4.5$ and seems still to be dropping. Similar results hold for the sign test as discussed in Section 3.

Depending as it does on three arguments, the function $e_{A^*,A}$ is difficult of complete evaluation, and interest has centered on finding simpler quantities which will serve to represent its general behavior. It is obvious that the Pitman efficiency denoted above by $e_{A^*,A}$ is $\lim_{N\to\infty} e_{A^*,A}(N, \alpha, \theta_A)$, where θ_N satisfies (1.1).

A second kind of efficiency limit is considered by Dixon [13], who evaluates for the sign test compared with the t-test the limit $e_{s,t}(N, \alpha, \infty)$ (which he denotes by E_∞). This limit would be of interest if we were concerned with small N, moderate α, but β very near to 1. (He also obtains $e_{s,t}(N, \alpha, 0)$ and finds that $\lim_{N\to\infty} e_{s,t}(N, \alpha, 0)$ is, for his problem, equal to the Pitman efficiency.)

It is clear that a wide choice of limiting values of $e_{A^*,A}(N, \alpha, \theta)$ might be defined, many of them pertinent in one situation or another. We wish next to call attention to one possibility which is in a sense intermediate between those of Pitman and Dixon and which seems to help to round out some comparisons. Instead of letting $\theta \to 0$ as does Pitman, or $\theta \to \infty$ as does Dixon, we hold θ as well as α fixed and let $N \to \infty$. This limit we denote by $e_{A^*A}(\infty, \alpha, \theta)$. It is closely related to the "index" of Chernoff [15], differing mainly in that Chernoff requires that $\alpha \to 0$, so that α and $1 - \beta$ remain of the same order. Our limit is presumably pertinent when one is interested in large samples and the region of high power, but its main interest seems to reside in the fact that it can, in some cases, be computed and serves to give the limit as $N \to \infty$ of sequences of efficiency curves of the form computed by Dixon for small N, permitting interpolation for moderate N.

3. Limiting efficiencies for the sign test. All of the tests we shall now consider (sign, normal, t) arise in both one-sided and two-sided versions. However, it is true for all of them that as the power tends to 1, the probability of type II error for the one-sided test of level α is asymptotically equivalent to that for the corresponding two-sided test of level 2α. The reason for this is simply that the two tests have identical critical values, and that one of the two tails in the two-sided test is dominant. This consideration simplifies the efficiency comparisons made below.

We are interested in the limiting behavior as $N \to \infty$ of the probability of second-kind error of the sign test. Suppose X is binomial for N trials with success probability p. We may test $H: p = r$ against the alternatives $p < r$. The test accepts if $X \geqq a_N$, where $a_N = rN - c\sqrt{N} + d_N$. That d_N is bounded follows easily from the fact that the error of the normal approximation to the binomial is of order $1/\sqrt{N}$ (see, for example, [16], p. 129). (Using this critical value, the level of significance tends to $\Phi(-c)$.) The probability of second-kind error is then

$$P_{\text{II}} = \sum_{x \geqq a_N} \pi(x),$$

$$\text{where } \pi(x) = \binom{N}{x} p^x (1-p)^{N-x}.$$

We can study the behavior of P_{II} by separately considering the initial term $\pi(a_N)$, and the ratio of the sum to this initial term.

LEMMA 3.1. *If $N \to \infty$ and $a/N \to r > p$, then*

$$\sum_{x \geqq a} \pi(x)/\pi(a) \to r(1-p)/(r-p).$$

PROOF. Since $R(x) = \pi(x+1)/\pi(x)$ is strictly decreasing, $[R(a)]^c > \pi(a+c)/\pi(a) > [R(a+b-1)]^c$, where $0 < c \leqq b$. Summing for $0 \leqq c \leqq b$ we get

$$\frac{1-[R(a)]^{b+1}}{1-R(a)} > \frac{\sum_a^{a+b} \pi(x)}{\pi(a)} > \frac{1-[R(a+b-1)]^{b+1}}{1-R(a+b-1)}. \tag{3.1}$$

As $N \to \infty$, $b \to \infty$, and $a/N \to r$, we have $R(a) \to (1-r)/r \cdot p/(1-p) < 1$, so that the upper bound in (3.1) tends to $r(1-p)/(r-p)$. If, in addition, we require $b/N \to \infty$, the lower bound has the same limit. Since $R(x)$ is decreasing for $x > a$, we have $\sum_{x>a+b} \pi(x)/\sum_a^{a+b} \pi(x) \to 0$, from which the result follows.

LEMMA 3.2. *If $a_N = rN - c\sqrt{N} + d_N$ with d_N bounded, then*

$$\pi(a_N) \sim \frac{\exp[-c^2/2r(1-r)]}{\sqrt{N}\sqrt{2\pi}\sqrt{r(1-r)}} \left[\left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r}\right]^N \left[\frac{r(1-p)}{p(1-r)}\right]^{c\sqrt{N}-d_N}$$

as $N \to \infty$.

The proof consists in using Stirling's formula and simplifying.

Combining Lemmas 3.1 and 3.2, we see that

$$\sqrt[N]{P_{\text{II}}} \to \left(\frac{p}{r}\right)^r \left(\frac{1-p}{1-r}\right)^{1-r} \qquad \text{as } N \to \infty. \tag{3.2}$$

Note that the limit depends on the hypothesis r and alternative p, but not on α. Since it happens in each of the three problems dealt with in this section that $\sqrt[N]{P_{\text{II}}}$ tends to a positive limit as $N \to \infty$, we shall give to this limit a name, referring to it as the *base* of P_{II}. The base is essentially the quantity ρ discussed by Chernoff [15]. In fact, the limit in (3.2) is the value obtained in [15] for ρ in the binomial case. However, Chernoff's ρ involves $\alpha \to 0$, whereas our α is fixed, and as he considers a much more general problem his results are less sharp. A similar remark applies to the normal test, below.

We shall also need the bases for the normal and t-tests. Consider the problem of testing that the mean of a normal population of unit variance is zero, against the alternative that the mean is $\delta > 0$. From a sample $X_1, \cdots, X_N$ we compute $N\bar{X} = \sum X_i$ and reject if $\sqrt{N}\bar{X} > K$, where $\Phi(K) = 1 - \alpha$. The power is $\beta(N, \alpha, \delta) = 1 - \Phi(K - \sqrt{N}\,\delta)$. If we fix α and δ and let $N \to \infty$, $1 - \beta = P_{\text{II}}$ is equivalent (in the sense of ratio) to

$$(1/\sqrt{N}\delta)\cdot(1/\sqrt{2\pi})\exp[-(\tfrac{1}{2})(\sqrt{N}\delta - K)^2].$$

The limit of $\sqrt[N]{P_{\text{II}}}$ is thus $\exp[-(\frac{1}{2})\delta^2]$, as given by Chernoff. This is our base, say $b_N(\alpha, \delta)$, which turns out to depend on δ but not on α.

Now suppose that the variance is unknown. We estimate it by $s^2/(N-1)$, where $s^2 = \sum(X_i - \bar{X}_N)^2 : \chi^2_{N-1}$, form $t_N = \sqrt{N}\bar{X}_N/(s/\sqrt{N-1})$, and reject if $t_N > K_N$, where $K_N \to K$. The power is

$$\beta_N^*(\delta) = P\left\{\sqrt{N}\bar{X}_N > \frac{sK_N}{\sqrt{N-1}}\,\middle|\,\delta\right\}$$
$$= P\left\{\sqrt{N}(\bar{X}_N - \delta) > -\delta\sqrt{N} + \frac{sK_N}{\sqrt{N-1}}\,\middle|\,\delta\right\}.$$

Thus

$$1 - \beta_N^*(\delta) = \int_0^\infty \Phi\left(-\delta\sqrt{N} + \frac{sK_N}{\sqrt{N-1}}\right) p_{\chi_{N-1}}(s)\,ds.$$

We first consider an upper bound. Break the integration at $(N-1)^{2/3}$ to get $1 - \beta_N^*(\delta) < \Phi(-\delta\sqrt{N} + K_N(N-1)^{1/6}) + P(\chi_{N-1} > (N-1)^{2/3})$. The first term has base $\exp[-(\frac{1}{2})\delta^2]$ as before; the second has base 0, since (writing $N - 1 = m$)

$$\int_{m^{2/3}}^\infty c_m u^{m-2} \exp(-\tfrac{1}{2}u^2)\,du < \exp(-\tfrac{1}{4}m^{3/4}(\sqrt{2})^{m-1}$$

$$\int_{m^{2/3}}^\infty c_m \left(\frac{u}{\sqrt{2}}\right)^{m-2} \exp\left[-\tfrac{1}{2}\left(\frac{u}{\sqrt{2}}\right)^2\right] d\left(\frac{u}{\sqrt{2}}\right) < 2^{(m-1)/2}\exp(-\tfrac{1}{4}m^{3/4}).$$

Take the $1/(m+1)$ power and pass to limit to get 0.

A straightforward calculus argument permits us to verify the

LEMMA. *If base* $\{a_N\} \leqq$ *base* $\{b_N\}$, *then base* $\{a_N + b_N\}$ = *base* $\{b_N\}$.

With the aid of this lemma we see that our upper bound has the same base $\exp[-(\frac{1}{2})\delta^2]$ as does P_{II} for the normal test. But since the normal test is more powerful than the t, it follows that the base of the t-test is also P_{II}.

We now apply these results to make limiting efficiency statements for fixed δ with $N \to \infty$. Suppose, for a standard test, that $\sqrt[N]{1 - \beta(\delta)} \to A(\delta)$ as $N \to \infty$; while for a second test, suppose that $\sqrt[N]{1 - \beta^*(\delta)} \to A^*(\delta)$. If we define $N^*(N)$ as in Section 2, it is easy to see that $N^*(N)/N \to (\log A^*(\delta))/(\log A(\delta))$. Thus, for the t-test compared to the normal, $e_{t,g}(\infty, \alpha, \delta) = 1$ for all α, δ. It follows that the comparison of sign to t will be the same as that of sign to normal; and as the latter is simpler, we shall examine it.

Let $X_1, \cdots, X_N$ be a sample from a normal population of unit variance. We may test the hypothesis that $E(X_i) = 0$ against the alternative that $E(X_i) = \delta > 0$ by using $\bar{X}_N$, in which case $\sqrt[N]{1 - \beta_N(\delta)} \to \exp[-(\frac{1}{2})\delta^2]$ as seen above. We could also employ the sign test, rejecting the hypothesis if too many of the X_i are positive. The number of positive signs is binomial, with $p = \frac{1}{2}$ under the hypothesis, $p = \Phi(\delta)$ under the alternative. Therefore, for the sign test, we have from (3.2) the base $2\sqrt{\Phi(\delta)[1 - \Phi(\delta)]}$. Thus for the sign test relative to the normal (and hence to the t),

$$e(\infty, \alpha, \delta) = \frac{2 \log 2 + \log \Phi(\delta) + \log [1 - \Phi(\delta)]}{-\delta^2}. \tag{3.3}$$

This quantity is seen to be independent of α but dependent on δ. As $\delta \to 0$, $e(\infty, \alpha, \delta) \to 2/\pi$, which agrees with the Pitman efficiency. As $\delta \to \infty$, $e(\infty, \alpha, \delta) \to \frac{1}{2}$. A few values of (3.3) are shown in the table. It is notable that (3.3) is very flat for δ in the range of interest, thus giving results in good agreement with those obtained from the simpler Pitman limit.

TABLE

δ	$1 - \Phi(\delta)$	$e_{s,t}(\infty, \alpha, \delta)$
0	.50	.637
.253	.40	.636
.524	.30	.634
1.645	.05	.614
3.090	.001	.578
3.719	.0001	.566
∞	0	.500

The curve (3.3) may be regarded as the limit as $N \to \infty$ of the power efficiency function, values of which for the sign test relative to the t-test have been given by Dixon [13]. It appears from Dixon's charts that for fixed α, the actual power efficiency curve decreases smoothly toward its limit (3.3), making it possible to interpolate for intermediate N and thus to obtain rough values of the power of the t-test from binomial tables.

As a curiosity we finally examine the limit of $e(N, \alpha, \delta)$ as $\delta \to \infty$ with N, α fixed—a limit which Dixon denotes by E_∞. Our tool is an analog of basis: we recall that for the normal test, $1 - \beta_N(\delta) = \exp[-(\frac{1}{2})N\delta^2]\cdot f(N, \alpha, \delta)$, where $\log f(N, \alpha, \delta) = 0(N\delta^2)$. Restricting ourselves to even N, we can use formula (10) of [17] to show that for the t-test

$$1 - \beta_N(\delta) = \exp[-(1 - x)N\delta^2/4]\cdot g(N, \alpha, \delta),$$

where again $\log g(N, \alpha, \delta) = 0(N\delta^2)$ and x is the α-point on the beta distribution.

Lemma 3.3. *Suppose that two tests A and B are available for each sample size N, and that*

$$1 - \beta_A(N, \alpha, \delta) = \exp(-a_N\delta^2)f(N, \alpha, \delta),$$

$$1 - \beta_B(N, \alpha, \delta) = \exp(-bN\delta^2)g(N, \alpha, \delta),$$

where $\log f(N, \alpha, \delta)$ *and* $\log g(N, \alpha, \delta)$ *are* $o(\delta^2)$.

Suppose that $\beta_B(N, \alpha, \delta)$ is strictly monotonely increasing in N, $\beta_B(1, \alpha, \delta) = \alpha$, $\beta_B(N, \alpha, \delta) \to 1$ as $N \to \infty$. Then $e_{A,B}(N, \alpha, \infty) = (N_0 + 1)/N$, where N_0 is the greatest integer less than a_N/b.

Proof. We shall first assume that a_N/b is not an integer, so that there exists an integer N_0 with $N_0 < a_N/b < N_0 + 1$. Examining the ratio $[1 - \beta_A(N)]/[1 - \beta_B(m)] = [f(N)/g(m)]\exp[(bm - a_N)\delta^2]$, we see that for all sufficiently large δ,

$$\beta_B(N_0, \alpha, \delta) < \beta_A(N, \alpha, \delta) < \beta_B(N_0 + 1, \alpha, \delta).$$

Recalling our definition of efficiency, we see that if $p(\delta)$ is defined by

$$p(\delta)\beta_B(N_0, \alpha, \delta) + [1 - p(\delta)]\beta_B(N_0 + 1, \alpha, \delta) = \beta_A(N, \alpha, \delta), \tag{3.4}$$

then

$$e_{A,B}(N, \alpha, \delta) = \frac{N_0 + 1 - p(\delta)}{N}.$$

If we solve (3.4) for $p(\delta)$ and let $\delta \to \infty$, we find that $p(\delta) \to 0$. Therefore $e_{A,B}(N, \alpha, \infty) = (N_0 + 1)/N$.

In the remaining case, in which a_N/b is an integer, let $N_0 + 1 = a_N/b$. A similar analysis then produces the same limiting formula.

It is convenient to introduce the convention that $[u]$ denotes the greatest integer less than u. Then we see that

$$e_{t,g}(N, \alpha, \infty) = \frac{\left[\frac{1 - x}{2} N\right] + 1}{N}.$$

It is notable that this limiting efficiency is a discontinuous function of α. Given any N, there exists an $\alpha_0(N)$ such that for $\alpha < \alpha_0(N)$, $e_{t,g}(N, \alpha, \infty) = 1$. But if we fix α and let $N \to \infty$, $e_{t,g}(N, \alpha, \infty)$ tends to a limit less than 1. Thus,

$$\lim_{N\to\infty} \lim_{\delta\to\infty} e_{t,g}(N, \alpha, \delta) \neq \lim_{\delta\to\infty} \lim_{N\to\infty} e_{t,g}(N, \alpha, \delta).$$

Now consider the behavior of $\beta_s(N, \alpha, \delta)$ as $\delta \to \infty$. For an individual observation, the probability p of a positive sign is $\Phi(\delta) \sim 1 - (1/\delta)\varphi(\delta)$ for large δ. Examination of Lemmas 3.1 and 3.2 shows that the assumptions of Lemma 3.3 are met by the sign test with a_m the critical value for the number of positive signs. Therefore

$$e_{s,t}(N, \alpha, \infty) = \frac{\left[\frac{2a_N}{1 - x}\right] + 1}{N}.$$

This formula is not comparable to the E_∞ of Dixon, since our definition of N^* is not the same as his.

REFERENCES

[1] E. J. G. PITMAN, Lecture notes on nonparametric statistical inference, Columbia University, 1949.

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

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

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

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

[6] S. SAKS, *Theory of the Integral*, Warsaw-Lvov, 1937.

[7] W. HOEFFDING, "'Optimum' nonparametric tests," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951.

[8] M. E. TERRY, "Some rank order tests which are most powerful against specific parametric alternatives," *Ann. Math. Stat.*, Vol. 23 (1944), pp. 346–366.

[9] M. DWASS, "The large sample power of rank order tests in the two-sample problem," *Ann. Math. Stat.*, this issue.

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

[11] W. H. KRUSKAL AND W. A. WALLIS, "Use of ranks in one-criterion analysis of variance," *Jour. Am. Stat. Assoc.*, Vol. 47 (1952), pp. 583–621.

[12] J. L. HODGES, JR. AND E. L. LEHMANN, "Matching in paired comparisons," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 787–791.

[13] W. J. DIXON, "Power functions of the sign test and power efficiency for normal alternatives," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 467–473.

[14] W. J. DIXON, "Power under normality of several nonparametric tests," *Ann. Math. Stat.*, Vol. 25, (1954), pp. 610–614.

[15] H. CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.

[16] J. V. USPENSKY, *Introduction to Mathematical Probability*, McGraw-Hill Book Company, Inc., 1937, New York.

[17] W. L. NICHOLSON, "A computing formula for the power of the analysis of variance test," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 607–610.