SIGN AND WILCOXON TESTS FOR LINEARITY[1]

By Richard A. Olshen[2]

Yale University

1. Introduction and summary. This paper introduces two tests of linearity against convexity in regression. In the first, the test statistic is the number of positive signs of second differences computed from certain of the observations. In the second, a Wilcoxon statistic is computed from those differences. Possible competitors of these tests are the usual least-squares t-test applied to regression coefficients, Mood's median test [12], and Hill's R test [6]. Certainly the first of these is to be preferred when errors are independent and normally distributed with common variance, and the alternative is quadratic regression. The sign test to be introduced here is simpler to compute than any of these other three tests, and the Wilcoxon test is also rather simple to compute. Both can be criticized in that their test statistics are calculated from certain randomly chosen observations.

The tests based on second differences are compared with the t-test when the alternative is quadratic regression and errors are continuously and symmetrically distributed. To be precise, in the model

$$Y_i = gX_i^2 + bX_i + a + \epsilon_i \tag{1}$$

for $i = 1, \cdots, N$, we shall compare tests of $H_0: g = 0$ against $H_1: g > 0$; a, b, and g are unspecified, and the ϵ's are independent, with identical distributions which are symmetric about their mean value of zero and have (unknown) variance σ^2. The criterion whereby tests are compared is Pitman efficiency, which is defined as follows.

Suppose θ is an unknown real parameter of a probability distribution H_θ. Suppose further that for each positive integer N, A_N and A_N^* are two size α $(0 < \alpha < 1)$ tests of the null hypothesis $\theta = \theta_0$ against the alternative $\theta > \theta_0$ based on a random sample of size N from H_θ. Let $\beta_N(\theta)$ and $\beta_N^*(\theta)$ be the respective power functions, β be a fixed number in $(\alpha, 1)$, ξ_N be a sequence of numbers for which $\xi_N \downarrow \theta$, and $M_1(\xi_N)[M_2(\xi_N)]$ be the least integer for which $\beta_{M_1}(\xi_N) \geqq \beta[\beta^*_{M_2}(\xi_N) \geqq \beta]$. The Pitman efficiency of A_N relative to A_N^* for the sequence of alternatives ξ_N is defined to be the $\lim_{N\to\infty} M_2(\xi_N)/M_1(\xi_N)$ provided that limit exists and does not depend on α and β beyond the requirement $0 < \alpha < \beta < 1$.

This definition differs from some others which are commonly used (see, for

[1] This research was supported in part by a National Science Foundation Cooperative Graduate Fellowship and in part by the Army, Navy, Air Force, and NASA under a contract administered by the Office of Naval Research.

[2] Now at Stanford University

example, [8]). Various technical facts concerning the Pitman efficiency of one-sample tests are discussed in an appendix, which may be of independent interest.

2. A simple design. In this section and throughout the paper, for $X_1 < X_2 < X_3$, and respective observations Y_1, Y_2, Y_3, the second difference based on these points is $Y_1 - 2Y_2 + Y_3$. Suppose, now, that there are equal numbers (say n) of observations at $X = -k$, $X = 0$, and $X = k$ for some real number k. Form n independent second differences, $\Delta_1, \cdots, \Delta_n$, by choosing one observation at random from each set for each second difference. Each Δ_i has expectation $2gk^2$, and variance $\tau^2 = 6\sigma^2$. (One easily sees that in the model two second differences are independent if, and only if, they do consist of distinct observations. Excessive computational difficulties would accompany consideration of dependent second differences.) The sign test with this design rejects H_0 for large values of

$$S_n = \sum_{i=1}^n I[\Delta_i > 0],$$

where $I[\cdot]$ is the indicator function of the event $[\cdot]$. The Wilcoxon test rejects H_0 for large values of

$$W_n = \sum_{i<j} I[\Delta_i + \Delta_j > 0].$$

While S_n is obviously a special case of the usual one-sample sign statistic, W_n is not precisely a special case of the one-sample Wilcoxon rank-sum, which can be expressed as $W_n + S_n$. Yet W_n and $W_n + S_n$ are equivalent for purposes of computing Pitman efficiency (see the Appendix), and the more tractable expression has been chosen for use here.

The least-squares estimate of g is $(2nk^2)^{-1}\bar{\Delta}$, where $\bar{\Delta} = n^{-1}\sum_{i=1}^n \Delta_i$. And the least-squares t-test of H_0 rejects for large values of

$$t_n = n^{\frac{1}{2}}\bar{\Delta}/(6\hat{\sigma}^2)^{\frac{1}{2}} =_{\text{def}} n^{\frac{1}{2}}\bar{\Delta}/(\hat{\tau}^2)^{\frac{1}{2}}.$$

$\hat{\sigma}^2$ is the usual residual mean-squared appropriate to an analysis of variance of the Y's. Of course when the ϵ's are normal, t_n has a t distribution with $3n - 3$ degrees of freedom, which is central when $g = 0$. A natural competitor of S_n, W_n, and t_n is the one-sample t-test based on the Δ's, which rejects H_0 for large values of

$$\tilde{t}_n = n^{\frac{1}{2}}\bar{\Delta}/[\sum_{i=1}^n (\Delta_i - \bar{\Delta})^2/n - 1]^{\frac{1}{2}} =_{\text{def}} n^{\frac{1}{2}}\bar{\Delta}/(\tilde{\tau}^2)^{\frac{1}{2}}.$$

When the ϵ's are normal $\tilde{t}_n$, in contrast to t_n, has a t distribution with only $n - 1$ degrees of freedom. Plainly the sign, Wilcoxon, and one-sample t-tests of H_0 against H_1 considered here are equivalent for purposes of computing Pitman efficiency to those of the one-sample location problem for F and θ, where F is the common distribution function of the second differences of the error terms in (1), and $\theta = 2gk^2$. The following results are essentially due to Pitman. (Technical details are deferred to the Appendix.) Note that a test statistic and the test based on it are labeled by the same symbol.

(2) THEOREM. *If $\sigma^2 < \infty$ and F is continuous at the origin and has a right derivative ($F'(0)$) there, then for any sequence of alternatives the Pitman efficiency of S_n to $\tilde{t}_n$ is $4\tau^2[F'(0)]^2$.*

(3) THEOREM. *If F is continuous and $\sigma^2 < \infty$, then the Pitman efficiency of W_n to $\bar{t}_n$ exists for every sequence of alternatives. When F is absolutely continuous with density f, that efficiency is $12\tau^2[\int_{-\infty}^{\infty} f^2(u)\, du]^2$, which can of course be infinite. The efficiency is always infinite when ($\sigma^2 < \infty$ and) F is continuous but not absolutely continuous.*

Now computations of Pitman efficiency relative to t_n and $\bar{t}_n$ involve $\hat{\tau}^2$ and $\tilde{\tau}^2$ only insofar as each of them tends in probability to τ^2. (Of course $\sigma^2 < \infty$ implies the convergence is actually with probability 1 and in L_1.) A consequence of this fact is the following.

(4) THEOREM. *The conclusions of* (2) *and* (3) *hold with $\bar{t}_n$ replaced by t_n in the statements.*

3. Observations at integer points. In this section the tests of Section 2 are studied in a more general setting. Consider an experimental design consisting of n observations at each of the integer points from $-r$ to r, so that there are $2r + 1$ different values of X. Because three observations are needed for each second difference, it is convenient to assume that $2r + 1 \equiv 0 \pmod 3$, that is, $r \equiv 1 \pmod 3$. Suppose the model (1) is appropriate and that F is the common distribution function of the second differences of the error terms. Form $n' = (2r + 1)n/3$ independent second differences, $\Delta_1, \cdots, \Delta_{n'}$ (again by random selection) from the observed values for a set of X's, one each at

$$-r + i, \quad -r + (n'/n) + i, \quad -r + 2(n'/n) + i$$

$$\text{for} \quad i = 0, \cdots, (n'/n) - 1.$$

Each Δ_i has expectation $2g(n'/n)^2$ and variance $6\sigma^2$.

As before, the sign test rejects H_0 for large values of

$$S_{n'} = \sum_{i=1}^{n'} I[\Delta_i > 0],$$

while the Wilcoxon test rejects for large values of

$$W_{n'} = \sum_{i<j=1}^{n'} I[\Delta_i + \Delta_j > 0].$$

The comparisons of $S_{n'}$ and $W_{n'}$ to least-squares t are facilitated by first comparing them to the one-sample t-test, $\bar{t}_{n'}$, based on the Δ's. It is in fact obvious that (2) and (3) hold verbatim for the designs being considered here provided n is replaced by n'—recall that $\tau^2 = 6\sigma^2$.

When $r > 1$, the least-squares t-test cannot be described as conveniently as it was in the last section, which was simply the case $r = 1$ because Pitman efficiency is invariant under changes in scale. Least-squares t rejects H_0 for large values of

$$t_{n'} = c(r, n')\hat{g}/(\hat{\sigma}^2)^{\frac{1}{2}},$$

where, as before, $\hat{\sigma}^2$ is the residual mean square appropriate to an analysis of variance of the Y's; $c(r, n')$ is a constant depending only on r and n', and

$$\hat{g} = \sum (3n'X^2 - \sum X^2)Y/[3n' \sum X^4 - (\sum X^2)^2]$$

is the least-squares estimate of g. The summations extend over all $3n'$ observations.

It follows from the definition of Pitman efficiency that for a given sequence of alternatives the efficiency of $S_{n'}(W_{n'})$ with respect to $t_{n'}$ is its efficiency with respect to $\bar{t}_{n'}$ multiplied by the efficiency of $\bar{t}_{n'}$ with respect to $t_{n'}$. The techniques of Section A1 of the Appendix make it easy to prove the following assertion.

(5) LEMMA. *If $\sigma^2 < \infty$, then for every sequence of alternatives the Pitman efficiency of $\bar{t}_{n'}$ to $t_{n'}$ is*

$$10(2r + 1)^4/81(4r^4 + 8r^3 + r^2 - 3r).$$

As has been mentioned, Section 2 was really about the case $r = 1$, and it was implicit there that the efficiency of $\bar{t}_{n'}$ with respect to $t_{n'}$ is 1. That value is unique to $r = 1$. For only when $r = 1$ is $\hat{g}$ an average of the Δ's. The loss of Pitman efficiency of $\bar{t}_{n'}$ with respect to $t_{n'}$ when $r > 1$ corresponds to the inefficiency of such an average as a point estimate of g. (5) and the remarks of the last three paragraphs combine to prove the Theorems (6) and (7).

(6) THEOREM. *Suppose F satisfies the assumptions of* (2). *Then for every sequence of alternatives the Pitman efficiency of $S_{n'}$ to $t_{n'}$ is*

$$80(2r + 1)^4(F'(0))^2\sigma^2/27(4r^4 + 8r^3 + r^2 - 3r).$$

(7) THEOREM. *If F satisfies the assumptions of* (3), *then for every sequence of alternatives the Pitman efficiency of $W_{n'}$ to $t_{n'}$ is*

$$80\sigma^2(2r + 1)^4[\int_{-\infty}^{\infty} f^2(u)\, du]^2/9(4r^4 + 8r^3 + r^2 - 3r)$$

if F is absolutely continuous with density f. The efficiency is ∞ otherwise.

Results of Hodges and Lehmann ([8], pp. 324–7) provide lower bounds for the efficiencies given in (6) and (7). The number given in (7) is always at least $8(2r + 1)^4/75(4r^4 + 8r^3 + r^2 - 3r)$. While the number given in (6) can be 0, it is not less than $10(2r + 1)^4/243(4r^4 + 8r^3 + r^2 - 3r)$ when F has a unimodal density. Naturally both these lower bounds agree with Hodges and Lehmann's bounds (108/125 for $W_{n'}$ to $t_{n'}$, $\frac{1}{3}$ for $S_{n'}$ to $t_{n'}$ with unimodal density) in case $r = 1$. When r is large, the efficiency of $\bar{t}_{n'}$ to $t_{n'}$ is nearly 40/81, and so both efficiencies and their lower bounds are roughly half their values when $r = 1$.

4. Most efficient sign test. The sign test of Section 3 can be generalized as follows. Suppose that the model (1) is appropriate and that Y's are observed at N (not necessarilydistinct) values of X. Form independent second differences $\Delta_1, \Delta_2, \cdots$ from the Y's subject only to the requirement that the expectation of each second difference does not depend on b, so if a difference is based on observations at $X_1 < X_2 < X_3$, then $X_1 - 2X_2 + X_3 = 0$. A test of H_0 against H_1 could be based on the statistic

$$S = S(w, N, \Delta_1, \Delta_2, \cdots) = \sum w_i I[\Delta_i > 0], \tag{8}$$

where the w's are nonnegative weights, and the summation extends over all second differences formed. The test would reject H_0 for large values of S. In view of the discussion of Section A1 and A2 in the Appendix, it is reasonable to compare tests of the type described by means of their efficacies (if they exist), where in the present context the efficacy of S is

$$[(d/dg)E_g\{S\}|_{g=0}]^2/\mathrm{Var}_0\{S\}. \tag{9}$$

The derivative in (9) is taken from the right; $E_g\{\cdot\}$ and $\mathrm{Var}_g\{\cdot\}$ denote the expectation and variance of the random variable in brackets when g is applicable. The "most efficient sign test" is defined as that test of the form (8) which maximizes the expression (9).

Suppose as in previous sections that F is the common distribution function of the second differences of the error terms. Assume that F is continuous at 0 and (to avoid trivialities) that it has a positive right derivative $F'(0)$ there. Then (9) is precisely

$$(F'(0))^2(\sum w_i(X_{1,i}^2 - 2X_{2,i}^2 + X_{3,i}^2))^2/\tfrac{1}{4}\sum w_i^2 \tag{10}$$

where the ith second difference is computed from observations at $X_{1,i} < X_{2,i} < X_{3,i}$, and naturally the sums extend over all differences formed. The Schwarz inequality applied to the numerator of (10) implies that the whole expression is not more than

$$4(F'(0))^2D =_{\mathrm{def}} 4(F'(0))^2\sum (X_{1,i}^2 - 2X_{2,i}^2 + X_{3,i}^2)^2$$

with equality if and only if w_i is proportional to $X_{1,i}^2 - 2X_{2,i}^2 + X_{3,i}^2$ for each i. The proof of the following lemma is now complete.

(11) LEMMA. *If $F'(0)$ is positive, then the most efficient sign test uses second differences formed so as to maximize the criterion D. The weight of the* ith *second difference is proportional to* $X_{1,i}^2 - 2X_{2,i}^2 + X_{3,i}^2$.

(11) is closely related to a result of Cox and Stuart ([2], Section 3).

There are often computational problems associated with the most efficient test. For example, suppose that equal numbers of observations are taken at $-4, -3, -2, \cdots, 4$ (corresponding in the notation of Section 3 to $r = 4$). Then the most efficient test forms second differences from observations at $(-4, 0, 4)$ and either $(-1, 1, 3)$ or $(-3, -1, 1)$ (or both provided independence is maintained). The weights are in the ratio 4:1, and power computations are complicated. The criterion D computed for the most efficient test is in the ratio 272:243 to D computed for the simpler test using identically distributed differences.

5. Acknowledgments. I am grateful to F. J. Anscombe, Byron W. Brown, Herman Chernoff, and Lincoln Moses for their advice on several earlier versions of this paper, and to L. J. Savage, who was of great help in bringing the paper to its final form.

APPENDIX

A1. Pitman's method. The Appendix contains some technical details essential in the body of the paper. The first topic covered is a reformulation of Pitman's method for computing Pitman efficiency as shown, for example, in the paper by Noether [14].

Assume, as in Section 1, that θ is an unknown real parameter of a probability distribution H_θ, and that for each N, A_N is a size α $(0 < \alpha < 1)$ test of the null hypothesis $\theta = \theta_0$ against the alternative $\theta > \theta_0$ based on a random sample $Z_1, \cdots, Z_N$ from H_θ. A_N rejects the null hypothesis for large values of the statistic $T_N = T_N(Z_1, \cdots, Z_N)$. In what follows, derivatives are taken from the right, and $\theta_N = \theta_0 + kN^{-\frac{1}{2}}$, where k is an arbitrary positive constant. We study first the power of A_N at the alternative θ_N.

Suppose there are two real-valued functions $\psi_N(\theta)$ and $\sigma_N(\theta)$ with the following properties:

(A) $\psi_N'(\theta_0)$ exists for all N;

(B) $\lim_{N\to\infty} [\psi_N'(\theta_0)/\sigma_N(\theta_0)N^{\frac{1}{2}}] = c > 0$;

(C) for each x, the distribution function of $[T_N - \psi_N(\theta)]/\sigma_N(\theta)$ tends to the standard normal distribution function $\Phi(x)$ uniformly in θ (and hence uniformly in θ and x ([15], p. 35)) for $\theta_0 \leqq \theta \leqq \theta_0 + d$ for some $d > 0$;

(D) $\lim_{N\to\infty} \sigma_N(\theta_N)/\sigma_N(\theta_0) = 1$;

and an additional assumption to be specified.

Often, but not always, $\psi_N(\theta) = E_\theta\{T_N\}$, $\sigma_N^2(\theta) = \operatorname{Var}_\theta\{T_N\}$.

Now the existence of $\psi_N'(\theta_0)$ implies $[\psi_N(\theta) - \psi_N(\theta_0)]/\theta - \theta_0 = \psi_N'(\theta_0) + g(N, \theta)$, where for fixed N $g(N, \theta)$ is $o(1)$ as $\theta \downarrow \theta_0$. The final assumption, then, is

(E) $g(N, \theta_N)/N^{\frac{1}{2}}\sigma_N(\theta_0) \to 0$ as $N \to \infty$.

Assumption (B) implies that $\psi_N'(\theta_0)$ is ultimately positive, though not necessarily that $\limsup \psi_N'(\theta_0) > 0$. It is interesting to note that assumption (C) is not redundant in the following sense. There is a compact interval I and a family of distribution functions $F_{N,y}(x)$ (one for each pair (N, y), N a positive integer and $y \varepsilon I$) with the following properties. For each (N, y), $F_{N,y}(x)$ has mean 0 and variance 1; $F_{N,y}(x) \to \Phi(x)$ for each y (thus uniformly in x); $F_{N,y}(x)$ is uniformly continuous in the pair (y, x) for each fixed N. Yet, the convergence of $F_{N,y}(x)$ to $\Phi(x)$ is not uniform in (y, x). For example, let $I = [0, 1]$. For $y \varepsilon I$, let $F_{N,y}$ be the distribution function of a standardized gamma variable with parameter the minimum of $(N + 1)^{\frac{1}{2}}$ and $y(N + 1)^{\frac{1}{2}} + y^{-1}(N + 1)^{-\frac{1}{2}}$. Then for each fixed k $F_{N,kN^{-\frac{1}{2}}} \nrightarrow \Phi$, but clearly the distributions $F_{N,y}$ possess the three cited properties.

With assumptions (A)–(E), Noether's arguments can be extended to show that the power of A_N at the alternative θ_N tends to $1 - \Phi(z_\alpha - kc)$, where z_α is defined by $1 - \Phi(z_\alpha) = \alpha$. Those arguments show also that if $\xi_N \downarrow \theta_0$, $\xi_N - \theta_0 = o(N^{-\frac{1}{2}})$, then the power of A_N at ξ_N tends to α, while if $N^{-\frac{1}{2}} = o(\xi_N - \theta_0)$, then the power of A_N at ξ_N tends to 1.

Further consequences of Noether's arguments are these. Suppose A_N (based on T_N) and A_N^* (based on T_N^*) are two sequences of tests satisfying A–E for each fixed k. If $\xi_N \downarrow \theta_0$ is any sequence of alternative hypotheses, then in the notation of Section 1, $M_1(\xi_N) \sim k_1/\xi_N^2$ and $M_2(\xi_N) \sim k_2/\xi_N^2$ for some fixed positive numbers

k_1 and k_2. Moreover, the Pitman efficiency of A_N relative to A_N^* for the sequence ξ_N is

$$\lim_{N\to\infty} [\text{efficacy} \quad \{A_N\}/\text{efficacy} \quad \{A_N^*\}],$$

where efficacy $\{A_N\} = [\psi_N'(\theta_0)]^2/\sigma_N^2\{\theta_0\}$, and efficacy $\{A_N^*\}$ is defined analogously.

Obviously Noether's generalization of Pitman's method can be rephrased in the context of this section.

A2. Application of Pitman's method to sign, Wilcoxon, and t. Let $Z_1, \cdots, Z_N$ be independently and identically distributed with (left-continuous) distribution function F satisfying $F(x - \theta) = 1 - F((\theta - x)+)$ for every x and some unknown θ. The problem consists of testing the null hypothesis $\theta = 0$ against the alternative $\theta > 0$. Assume first that F has finite variance τ^2.

Computations of the efficacies of the t and sign tests for this problem are well known and are omitted. Suffice it to say that for the former, $T_N = N^{\frac{1}{2}}\bar{Z}_N/\hat{\tau}_N$, $\psi_N(\theta) = N^{\frac{1}{2}}\theta/\tau$, $\sigma_N(\theta) \equiv 1$, and for the latter, $T_N = \sum_{n=1}^N I[Z_n > 0]$, $\psi_N(\theta) = NF(\theta)$, $\sigma_N(\theta) = [NF(\theta)F((-\theta)+)]^{\frac{1}{2}}$. $\bar{Z}_N$ and $\hat{\tau}_N$ are the sample mean and sample standard deviations of the Z's. While it is known that (for any sequence of alternatives) the Pitman efficiency of the sign test to the t-test is $4\tau^2[F'(0)]^2$ whenever F is continuous at 0 and has a right derivative $(F'(0))$ there, Noether's version of Pitman's method cannot be used directly to prove this assertion unless F' is assumed to exist in some interval $[0, \delta]$ and to be continuous at 0. A discussion of the case $\tau^2 = \infty$ is contained in the next section.

Computation of the efficacy of the Wilcoxon test for the problem under consideration is, to the best of my knowledge, not available in the literature. Thus, several technical points concerning that computation are discussed here. To facilitate the discussion assume: (i) That F is continuous; (ii) that $\tau^2 < \infty$; and (iii) that

$$\lim_{\delta\downarrow 0} \int_{-\infty}^{\infty} \delta^{-1}[F(u + 2\delta) - F(u)]\, dF(u)$$

exists and is finite—call that limit d.

The Wilcoxon test rejects the null hypothesis for large values of

$$\sum_{n\leqq m=1}^{N} I[Z_n + Z_m > 0]$$
$$= \sum_{n<m=1}^{N} I[Z_n + Z_m > 0] + \sum_{n=1}^{N} I[Z_n > 0] =_{\text{def}} W_N + S_N.$$

Now it follows from facts mentioned in the previous section and a theorem of Hoeffding ([10], Theorem 7.3) that (for any sequence of alternatives) the test which rejects for large values of W_N has the same Pitman efficiency relative to any test satisfying (A)–(E) as does the Wilcoxon test. So we study the test based on W_N.

Let $\psi_N(\theta) = E_\theta\{W_N\}$ and $\sigma_N^2(\theta) = \text{Var}_\theta\{W_N\}$. Compute easily $\psi_N(\theta) = \binom{N}{2} \int_{-\infty}^{\infty} F(u + 2\theta)\, dF(u)$, and notice that $\sigma_N^2(\theta)$ is a continuous function of θ. In particular $\sigma_N^2(0) = N(N - 1)(N - 2)/24$. Hence, in the notation of

Section A1, D is satisfied. Moreover,

$$\psi_N'(\theta) = \lim_{\delta \downarrow 0} \binom{N}{2} \int_{-\infty}^{\infty} \delta^{-1}[F(u + 2\theta + 2\delta) - F(u + 2\theta)]\, dF(u)$$

when this limit exists and is finite. Assumption (iii) implies $\psi_N'(0)$ exists, and arguments of Hodges and Lehmann ([8], p. 326) show that therefore F is absolutely continuous, and $d > 0$. Consequently (A) and (B) and also (E) are satisfied. Hodges and Lehmann's arguments together with the following lemma give a necessary and sufficient condition for (iii). The condition is that the density of F be square summable. The number x is said to be a Lebesgue point of the function h if

$$\lim_{\delta \to 0} \delta^{-1} \int_x^{x+\delta} |h(t) - h(x)|\, dt = 0$$

at x ([13], p. 255).

(12) LEMMA. *Let G be an absolutely continuous distribution function with density g. Then for fixed γ,*

$$(13) \quad \lim_{\delta \to 0} \int_{-\infty}^{\infty} \delta^{-1}[G(x + \gamma + \delta) - G(x + \gamma)]g(x)\, dx = \int_{-\infty}^{\infty} g(x + \gamma)g(x)\, dx$$

if $-\gamma$ is a Lebesgue point of $g*g^-$, where $g^-(x) = g(-x)$, and $*$ denotes convolution. The limit integral is finite and continuous for all γ if and only if g is square summable.

PROOF. Suppose $\delta > 0$, and let W and Z be two independent random variables with the property that $W - \gamma$ and Z have distribution function G. Then

$$\delta^{-1} \Pr\{-\delta < W - Z < 0\} = \delta^{-1} \int_{-\infty}^{\infty} [G(x + \gamma + \delta) - G(x + \gamma)]g(x)\, dx$$
$$= \delta^{-1} \int_{-\delta}^{0} (g*g^-)(x - \gamma)\, dx.$$

So if $-\gamma$ is a Lebesgue point of $g*g^-$, then ([13], p. 255) as $\delta \downarrow 0$ the above three identical quantities tend to $(g*g^-)(-\gamma) = \int_{-\infty}^{\infty} g(u + \gamma)g(u)\, du$.

The final step in proving the lemma in case $\delta > 0$ is accomplished by noting the equivalence of these three propositions.

(14) $g*g^-$ is continuous.

(15) $(g*g^-)(0)$ is a finite number.

(16) g is square summable.

(14) implies (15) implies (16) is trivial. For a proof that (16) implies (14) see, for example, the book by Hewitt and Stromberg ([5], p. 398). Incidentally, that

$$\lim_{\delta \downarrow 0} \delta^{-1} \int_{-\delta}^{0} (g*g^-)(x - \gamma)\, dx = (g*g^-)(-\gamma)$$

when g is continuous is even a consequence of the first mean value theorem for integrals.

The proof for $\delta < 0$ is now obvious.

(12) extends Lemma 3(a) of [9]. Since this paper was submitted for publication, a different proof that (13) holds for $\gamma = 0$ when g is square summable has

been given by Mehra and Sarangi ([11], p. 101). The lemma can be used to extend various recent results in nonparametric theory.

In order to show that (i), (ii) and (iii) imply that the Pitman efficiency of Wilcoxon to t can be computed by the methods of the previous section, there remains the task of showing that W_N satisfies C. This is accomplished by the following theorem, which has obvious multivariate generalizations.

(17) THEOREM. *Suppose that* $Z_1, \cdots, Z_N$ *are independent, identically distributed random variables, each with distribution function* $G(x - \theta)$, *where* $\theta \varepsilon \Theta$, *a compact set of real numbers. (The dependence of the Z's on* θ *is suppressed.) Suppose further that u is a real-valued symmetric function of m real variables for which*

(iv) $E\{u(Z_1, \cdots, Z_m)\} = \eta(\theta)$,

(v) $E\{|u(Z_1, \cdots, Z_m) - \eta(\theta)|^{2+\delta}\}$ *is uniformly bounded for* $\theta \varepsilon \Theta$, *and*

(vi) $\mathrm{Var}\{E\{u(Z_1, \cdots, Z_m) \mid Z_1\}\} = \zeta(\theta)$ *is continuous and uniformly bounded away from* 0 *on* Θ.

For $N \geqq m$ *let*

$$U_N = U_N(Z_1, \cdots, Z_N) = \binom{N}{m}^{-1} \sum_{1 \leqq i_1 < \cdots < i_m \leqq N} u(Z_{i_1}, \cdots, Z_{i_m}).$$

Assume finally that:

(vii) $\mathrm{Var}\{U_N\}$ *is a continuous function of* θ.

Then the distribution function of

$$U_N^* = (U_N - \eta(\theta))/[\mathrm{Var}\{U_N\}]^{\frac{1}{2}}$$

tends to $\Phi(x)$ *uniformly in the pair* (x, θ), $x \varepsilon [-\infty, \infty]$, $\theta \varepsilon \Theta$.

PROOF. This theorem amounts to little more than reading arguments of Hoeffding ([10], Theorems 7.1 and 7.2) in view of a theorem by Parzen ([15], (10.1)). Rather than studying U_N^* directly, first study $\tilde{U}_N = N^{\frac{1}{2}}(U_N - \eta(\theta))$. Thus, let $\Psi(Z_n, \theta) = E\{u(Z_n, Z_{i_2}, \cdots, Z_{i_m}) \mid Z_n\} - \eta(\theta)$, where $i_2, \cdots, i_m \neq n$, and set $Y_N(\theta) = mN^{-\frac{1}{2}} \sum_{n=1}^{N} \Psi(Z_n, \theta)$. Now it follows from (v) that $\mathrm{Var}\{u(Z_1, \cdots, Z_m)\}$ is uniformly bounded on Θ, and so ([10], (5.13), (7.9), (7.10), (7.12)) imply $E\{(\tilde{U}_N - Y_N(\theta))^2\}$ tends to 0 uniformly on Θ. The variables $\Psi(Z_n, \theta)$, $n = 1, \cdots, N$, are independent and identically distributed. Applying (10.1) of [15], the conditions of which are satisfied as a consequence of (v), the distribution function of $Y_N(\theta)$, $H_{N,\theta}(x)$, tends to $\Phi(x/m(\zeta(\theta))^{\frac{1}{2}})$ uniformly in (x, θ). An obvious "uniform" application of a theorem of Slutsky ([3], Section 20.6) implies that the distribution function of $\tilde{U}_N$ tends to $\Phi(x/m(\zeta(\theta))^{\frac{1}{2}})$ uniformly in (x, θ). Now according to Theorem 5.2 of [10], $N \mathrm{Var}\{U_N\}$ decreases as N increases, and $\lim_{N\to\infty} N \mathrm{Var}\{U_N\} = m^2\zeta(\theta)$. Dini's theorem ([5], p. 205) and (vii) imply the convergence is uniform on Θ. Another "uniform" application of the theorem of Slutsky implies the desired conclusion for U_N^*.

The verification that W_N satisfies the assumptions of this theorem is easy and omitted. Collecting results of this section and Section A1, the following fact is evident. If F satisfies (i), (ii), and (iii), then for any sequence of alternatives (tending to 0), the Pitman efficiency of Wilcoxon to t is $12\tau^2[\int_{-\infty}^{\infty} f^2(x)\,dx]^2$, where τ^2 is the variance of F, and f is its density. In case (i) and (ii) hold but (iii) does not, then arguments of Hodges and Lehmann ([8], p. 326) can be employed to show that (again, for any sequence of alternatives) the Pitman efficiency of Wilcoxon to t exists and equals ∞. These conclusions are analogous to those of Andrews [1] for two and larger sample problems.

A3. $\tau^2 = \infty$. Suppose the assumptions of the first paragraph of Section A2 hold except now assume that $\tau^2 = \infty$. For comparisons of the t-test with the sign test and the Wilcoxon test to make sense, it is necessary that the t-test be consistent. I hope to present more information on that consistency in the near future, and at that time to prove several conjectures stated at the end of this section. The following remarks must suffice for now.

As before, the t-test rejects the null hypothesis $\theta = 0$ for large values of $t_N =_{\text{def}} N^{\frac{1}{2}}\bar{Z}_N/\hat{\tau}_N$.

(18) THEOREM. *Assume that* $\bar{Z}_N$ *does not obey the weak law of large numbers, that is,* $\bar{Z}_N$ *does not tend in probability to 0. Then the t-test is not consistent in the following sense. For no fixed* $\theta^* > 0$ *does there exist a sequence of constants* d_N *for which* $\Pr_0[t_N > d_N] \to 0$ *and* $\Pr_{\theta^*}[t_N > d_N] \to 1$ *as* $N \to \infty$. ($\Pr_\theta[\cdot]$ *is the probability of the event* $[\cdot]$ *when* θ *is applicable*).

PROOF. Let θ^* be an arbitrary fixed positive number. The symmetry of F implies $\Pr_0[t_N > 0]$ is uniformly bounded away from 0. So it is enough to prove that $\Pr_{\theta^*}[t_N > 0] \nrightarrow 1$ as $N \to \infty$.

Now $\Pr_{\theta^*}[t_N > 0] = \Pr_{\theta^*}[\bar{Z}_N > 0] = \Pr_0[\bar{Z}_N > -\theta^*]$. Suppose these identical quantities do tend to 1 as $N \to \infty$. By the symmetry of F, then, $\Pr_0[\bar{Z}_N < \theta^*] \to 1$, and so $\Pr_0[-\theta^* < \bar{Z}_N < \theta^*] \to 1$. Let H_N be the distribution function of $\bar{Z}_N$ under the hypothesis $\theta = 0$. By what has been shown and Helly's theorem ([4], p. 261), H_N has a subsequence which converges in distribution, and every limit in distribution of a subsequence of H_N is supported on a subset of $[-\theta^*, \theta^*]$. Because $\bar{Z}_N$ does not tend in probability to 0, at least one limit in distribution of a subsequence of H_N is nondegenerate. Call that limit H. Now H is necessarily infinitely divisible ([4], p. 556). But there are no nondegenerate infinitely divisible distributions supported on bounded sets ([4], p. 174), which contradicts the assumption $\Pr_0[\bar{Z}_N > -\theta^*] \to 1$ and thus completes the proof.

I believe that the converse to this theorem is false. Yet there are weaker sufficient conditions for the consistency of the t-test than the condition $\tau^2 < \infty$. It can be shown, for example, (using the arguments of [4], pp. 232–234) that if $F(x - \theta)$ has density $f(x - \theta) \sim 1/|x - \theta|^3$ as $|x| \to \infty$, then the t-test is consistent.

The following conjectures can be supported by plausibility arguments. Each is believed to hold for every sequence of alternatives and for those F for which $\tau^2 = \infty$ but t is consistent. If F is continuous, the Pitman efficiency of Wilcoxon to t exists and is infinite. If $F'(0)$ exists and is positive, the Pitman efficiency of sign to t exists and is infinite. However, if $F'(0) = 0$, the Pitman efficiency of sign to t, when it exists, can be 0 or ∞.

REFERENCES

[1] ANDREWS, F. C. (1954). Asymptotic behavior of some rank tests for analysis of variance. *Ann. Math. Statist.* **25** 724–736.

[2] COX, D. R. and STUART, A. (1955). Some quick sign tests for trend in location and dispersion. *Biometrika* **41** 80–95.

[3] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[4] FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

[5] HEWITT, E. and STROMBERG, K. (1965). *Real and Abstract Analysis.* Springer, New York.

[6] HILL, B. M. (1962). A test of linearity versus convexity of a median regression curve. *Ann. Math. Statist.* **33** 1096–1123.

[7] HODGES, J. L., JR. (1963). Lectures. Unpublished.

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

[9] HODGES, J. L., JR. and LEHMANN, E. L. (1961). Comparison of the normal scores and Wilcoxon tests. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 307–317. Univ. of California Press.

[10] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

[11] MEHRA, K. L. and SARANGI, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments. *Ann. Math. Statist.* **38** 90–107.

[12] MOOD, A. MCF. (1950). *Introduction to the Theory of Statistics.* McGraw-Hill, New York.

[13] NATANSON, I. P. (1961). *Theory of Functions of a Real Variable* **1**. Ungar, New York.

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

[15] PARZEN, E. (1954). On uniform convergence of families of sequences of random variables. *Univ. Calif. Publ. Statist.* **2** No. 2, 23–54.