The Annals of Statistics
1982, Vol. 10, No. 1. 44–64

ASYMPTOTIC LOGNORMALITY OF *P*-VALUES[1]

By Diane Lambert and W. J. Hall

Carnegie-Mellon University and University of Rochester

Sufficient conditions for asymptotic lognormality of exact and approximate, unconditional and conditional P-values are established. It is pointed out that the mean, which is half the Bahadur slope, and the standard deviation of the asymptotic distribution of the log transformed P-value together, but not the mean alone, permit approximation of both the level and power of the test. This provides a method of discriminating between tests that have Bahadur efficiency one. The asymptotic distributions of the log transformed P-values of the common one- and two-sample tests for location are derived and compared.

1. Introduction. A P-value or observed significance level is used to assess the strength of the evidence against a null hypothesis H; hence its size under alternative models is important. An appropriate P-value, exact or approximate, is stochastically at least as large as a uniform [0, 1] random variable U under H (Bahadur and Raghavachari, 1970; Kempthorne and Folks, 1971) and stochastically smaller than U under alternative models. Precisely how much smaller depends on the finite sampling distribution G_n of the P-value. Unfortunately, G_n (or equivalently, the power function) is often intractable. Various summaries of G_n, such as its mean (Dempster and Schatzoff, 1965) and median (Joiner, 1969), have been proposed to describe and compare P-values, but these too are difficult to compute.

Bahadur (1960a) has overcome these complications for the sample mean, sign and t P-values under normal alternatives by approximating G_n by a lognormal df. The present paper extends his results to more general exact and approximate, unconditional and conditional P-values under more general alternative distributions. Sufficient conditions for asymptotic lognormality are given in Section 4. The main theorems for P-values based on sums, including rank and permutation sums, are in Section 5. Applications to one and two sample tests for location appear in Section 6.

The parameters of the asymptotic distribution are interpreted in Section 3. The asymptotic mean is the slope (our "slope" is half of Bahadur's slope) or the almost sure exponential rate at which the P-value approaches zero as the sample size increases. The relationship between the slope and asymptotic level of a test has been discussed previously (Bahadur, 1960b, 1967, 1971). Here we argue that the slope and asymptotic standard deviation together better approximate the level and power of the test than does the slope alone.

Power, possibly asymptotic, of tests of a given size is the usual basis for evaluating the performance of a test in the Neyman-Pearson framework. In practice, however, a P-value approach to testing without a fixed significance level α is as common and more versatile. With this approach, the mean and standard deviation of the asymptotic distribution of the log P-value are of interest in their own right as summary measures of test performance.

At present, the quality of the lognormal approximation to the finite sample distribution of a P-value P_n is uncertain. The proofs of asymptotic lognormality suggest that a modified P-value $\sqrt{n}P_n$ may be more stable with increasing n than P_n itself, and hence more

Received February 1979; revised July 1981.

[1] Supported by the U.S. Army Research Office through the University of Rochester.

AMS 1970 *subject classifications*. Primary 62F20; secondary 62G20.

Key words and phrases. P-value, Bahadur efficiency, slope, one-sample tests, two-sample tests.

appropriate for comparing P-values based on different sample sizes. Other calculations (e.g., Lambert, 1978) suggest that an additional square root transformation applied to the log P-value may improve the approximation or that $-\Phi^{-1}(P_n)$ may be closer to normality. We intend to study such approximations in a later paper.

2. Definition of *P*-values. Let $\mathscr{S}$ be a set of points s, Θ a set of parameters θ, and suppose that the distribution of s is determined by θ. There is given $\Theta_0 \subset \Theta$, and the null hypothesis H under test is $\theta \in \Theta_0$. Let n be an index restricted to certain positive integral values, and for each n let $T_n(s)$ be a test statistic, large values being significant. The index n may be thought of as the cost of observing T_n. A possible advantage to the present locution is that one- and two-sample cases can be treated simultaneously. In the latter case, assume that each admissible n has an associated partition into positive integers, say $n = n_1 + n_2$, such that $n_1/n \to \lambda$, $0 < \lambda < 1$, as $n \to \infty$ and assume T_n is really a statistic U_{n_1,n_2} for each admissible n.

For each $\theta \in \Theta$, denote the left continuous distribution function (df) of T_n by $F_n(\cdot; \theta)$ and let $\bar{F}_n = 1 - F_n$. Then the P-value P_n is defined to be $P_n = \sup_{\theta\in\Theta_0}\bar{F}_n(T_n; \theta)$; P_n may be called *exact* to distinguish it from an *approximate P-value* for which F_n is replaced by an approximation. An approximate P-value is of particular interest when F_n is unknown or untabulated. A P-value is called *conditional* if the null distribution is conditional on a statistic $V_n(s)$, e.g. as with a permutation test P-value.

3. Uses of asymptotic lognormality. Typically (regularity conditions are given later), if P_n is any P-value, exact or approximate, unconditional or conditional, based on a test statistic T_n that is asymptotically normal (AN) under P_θ, then P_n is asymptotically lognormal $(-nc(\theta), n\tau^2(\theta))$ under P_θ. The parameter c is the a.s. limit of $-n^{-1} \log P_n$, termed the *half-slope* by Bahadur and the *slope* in this paper. If P_n is an approximate P-value, then c is the (Bahadur) *approximate* (half-) *slope*. The approximate slope need not approximate the exact slope (e.g., see Gleser, 1964) but it is the parameter of interest for an approximate P-value.

The joint distribution of P-values P_n and P_n' calculated from the same sample is often also asymptotically bivariate lognormal. As Bahadur (1960a) suggested, this permits the P-value of one test to be predicted from the P-value of a second, perhaps more easily implemented, test. It also permits study of the interrelationship of two tests, but this is not pursued further here.

The parameters c and τ provide an asymptotic description of test behavior. The slope c, which is the exponential rate at which the P-value approaches zero, has been extensively studied (e.g., Bahadur, 1971). The slope alone is commonly used to compare tests. Yet recommending a test because the standardized form of its P-value $-n^{-1} \log P_n$ has a large asymptotic mean without regard to its variance τ^2/n seems as unwise as favoring an unbiased estimator without knowing its variance. Also, τ often varies in θ, and hence the slope itself cannot adequately summarize the distribution of P_n.

The parameters c and τ together provide a means of relating sample size to both the level α and asymptotic power $\beta \in (0, 1)$ as $\alpha \to 0$ with θ fixed. Consider the test that rejects the null hypothesis when $P_n \le \alpha$. If P_n is an exact P-value, this test is of size $\le\alpha$; it is of size α if, in addition, P_n is exactly uniformly distributed. (If P_n is an approximate P-value, the test is only of nominal size α.) In any case, the power of the test is $P_\theta(P_n \le \alpha) = P_\theta(Z_n \le \gamma(\alpha))$ where $\gamma(\alpha) = \sqrt{n}(n^{-1} \log \alpha + c)/\tau$ and $Z_n = \gamma(P_n)$. Since Z_n is $AN(0, 1)$, it follows that the power approaches β if and only if $\gamma(\alpha)$ approaches $b \equiv \Phi^{-1}(\beta)$. Assuming $c > 0$, and writing $A = -\log \alpha$, it is easily seen that the condition for asymptotic power β is equivalent to

$$n = \frac{A}{c}\left\{1 + \frac{\tau b}{(Ac)^{1/2}} + o(A^{-1/2})\right\}. \tag{1}$$

It follows from (1) that $n \sim A/c$ as $\alpha \to 0$. This asymptotic relation is well-known (see, e.g., Bahadur, 1967) and does not require convergence in distribution of the standardized P-value or specification of β. It is clear from (1) that the choice of β affects only b, not A or c, and hence represents a second order effect. (That β is strictly between 0 and 1 implies b is finite and conversely.) In particular, for two exact P-values P_{1n} and P_{2n} (with uniform null distributions and asymptotically lognormal nonnull distributions) it is possible to choose sample sizes n_1 and n_2 so that the corresponding tests are both of size α, $n_1/n_2 \to c_2/c_1$ as $\alpha \to 0$, but the power of test i converges to β_i for $i = 1, 2$, where β_1 and β_2 are arbitrary values in (0, 1).

From another perspective, if one uses the formula $A = nc$ to choose a sample size n to achieve a small test size α, the resulting log P-value is approximately normally distributed around $-nc$ and the power is approximately ½ since

$$P_\theta(P_n \leq \alpha) = P_\theta(Z_n \leq \sqrt{n}(c - n^{-1}A)/\tau) \to \tfrac{1}{2}.$$

By contrast, the formula $A = nc - \sqrt{n}b\tau$ for choosing n yields, as $n \to \infty$ or $\alpha \to 0$,

$$P_\theta(P_n \leq \alpha) = P_\theta(Z_n \leq b) \to \beta.$$

(The formula should not be relied on for β near one, since an accurate approximation in the tails of the lognormal distribution would be required.) In summary, the slope and standard deviation τ together, but not the slope alone, relate sample size to both the asymptotic level and asymptotic power of a test.

If P_{1n}, P_{2n} have the same slope c, then the parameter τ can be used to distinguish them. As shown in equation (1) above, the P-value with the smaller τ asymptotically requires the smaller sample size to attain level α and power $\beta > \tfrac{1}{2}$. Similarly, the P-value with the smaller τ will be small with higher probability, confining attention to powers exceeding ½.

In view of the present lack of information about how relevant slopes are to finite sample P-values, the parameter τ might also be useful for comparing P-values with unequal slopes. If, for a given nonnull θ, $\tau_1 < \tau_2$ and c_1 is smaller, but not much smaller, than c_2, then P_{1n} may be preferable to P_{2n}. For example, suppose $c_2 > c_1 > 10/n$ so that a "highly significant" P-value is expected in either case ($e^{-10} < 10^{-4}$). If

$$(c_1\tau_2 - c_2\tau_1)/(\tau_2 - \tau_1) > -n^{-1}\log(\sqrt{n}\alpha)$$

then, based on the lognormal approximations to $\sqrt{n}P_{in}$, P_{1n} is more likely to be less than α than is P_{2n}.

The lemmas in Section 5 that assert asymptotic lognormality are perhaps as useful as the lognormality result itself. These lemmas show that in several cases under modest regularity conditions $n^{-1}\log P_n$ behaves like an average of i.i.d r.v.'s in large samples. That is,

$$-n^{-1}\log P_n = c + n^{-1}\textstyle\sum_1^n u(X_i) + o_p(n^{-1/2}),$$

where c is the slope at an alternative P_θ, $\theta \in \Theta_1$; $X_1, \cdots, X_n$ is a random sample from P_θ; and $u(X_i)$ has mean zero and variance τ^2 under P_θ. The rate of convergence can often be improved; see Section 5 below and Lambert (1978). The quantity $u(X_i)$ measures the extent to which the observation X_i changes the P-value. In other words, $-n^{-1}\log P_n$ is approximately equal to its "model" value, which is the slope c, modified by a factor determined by the average influences of the data. Such an expansion is useful in robustness studies of P-values and tests (Lambert, 1981).

4. Sufficient conditions for asymptotic lognormality of *P*-values. Bahadur (1960b, 1967) has formulated certain conditions for the existence of slopes. Such formulations are often trivial - Bahadur himself calls them "prescriptions" - but they do exhibit the various components of the problem. A similar formulation can be given for the existence of the parameters c and τ of the asymptotic lognormal distribution.

Choose and fix a nonnull θ, $\theta \in \Theta - \Theta_0$. Assume that there exist constants $b(\theta)$ and $\sigma(\theta)$, $-\infty < b < \infty$ and $0 < \sigma < \infty$, such that

(i) $$\sqrt{n}\{T_n - b(\theta)\} \quad \text{is} \quad AN(0, \sigma^2(\theta))$$

under θ as $n \to \infty$. This condition is satisfied by suitable versions of all statistics under consideration, plus uncountably many others, and the associated b and σ are known, at least in principle. Let $\mathscr{I}$ be an open interval containing $b(\theta)$.

For each n, let G_n be a nondecreasing function defined on the real line, $0 \le G_n \le 1$; G_n is to be thought of as the exact left-continuous null df of T_n or as an approximation to the exact null df. Let $\bar{G}_n \equiv 1 - G_n$ and define the P-value $P_n(s) \equiv \bar{G}_n(T_n(s))$.

Assume that associated with the given sequence $\{\bar{G}_n\}$ of functions there is a real valued function f defined on $\mathscr{I}$ such that the following holds: if

(ii) $b \in \mathscr{I}$ and $\{b_n\}$ is a sequence of numbers such that $b_n = b + O(n^{-1/2})$

then

(iii) $\log \bar{G}_n(b_n) = -nf(b_n) + o(\sqrt{n})$ as $n \to \infty$.

Assume also

(iv) f is continuously differentiable on $\mathscr{I}$; let $f' = h$.

LEMMA 4.1. *The implication* (ii) $\Rightarrow$ (iii) *and the conditions* (i) *and* (iv) *together imply that* $n^{-1/2}\{\log P_n + nc(\theta)\}$ *is* $AN(0, \tau^2(\theta))$ *where* $c(\theta) = f(b(\theta))$ *and* $\tau(\theta) = \sigma(\theta)h(b(\theta))$.

PROOF. For any real x let $D_n(x) = n^{-1/2}|\log \bar{G}_n(x) + nf(x)|$ if $x \in \mathscr{I}$ and let $D_n(x) = \infty$ otherwise. For any positive constant k, let $E_n(k) = \sup\{D_n(x) : |x - b| \le k/\sqrt{n}\}$. Then, for each k, $E_n(k) \to 0$ as $n \to \infty$, for otherwise the implication (ii) $\Rightarrow$ (iii) is contradicted.

Choose $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$. It follows from (i) that $\sqrt{n}(T_n - b)$ is bounded in probability. Hence, there exists $k > 0$ such that $P(\sqrt{n}|T_n - b| > k) < \varepsilon_1$ for all n. Since there exists m such that $E_n(k) < \varepsilon_2$ for $n \ge m$, it follows that $D_n(T_n) > \varepsilon_2$ implies $\sqrt{n}|T_n - b| > k$ for $n \ge m$. Hence, $P\{D_n(T_n) > \varepsilon_2\} < \varepsilon_1$ for all $n \ge m$. Since ε_1 and ε_2 are arbitrary, $D_n(T_n) \to 0$ in probability. Since $D_n(T_n) < \infty$ implies $D_n(T_n) = n^{-1/2}|\log P_n + nf(T_n)|$ it follows that, for any sequence $\{c_n\}$, $n^{-1/2}(\log P_n + nc_n)$ has an asymptotic distribution iff $n^{1/2}\{c_n - f(T_n)\}$ does, and the asymptotic distributions are then the same.

Since $T_n \to b$ in probability, since (i) holds, and since f is sufficiently smooth, $\sqrt{n}\{f(b) - f(T_n)\}$ is $AN(0, \tau^2)$. Hence, $n^{-1/2}\{\log P_n + nf(b)\}$ is also $AN(0, \tau^2)$. □

Verification of the implication (ii) $\Rightarrow$ (iii) is usually the most difficult step in the proof that a sequence of P-values is asymptotically lognormal. For a simple illustration of Lemma 4.1, suppose the exact null df of the test statistic T_n is approximated by a normal $(0, n^{-1})$ df.

THEOREM 4.1. *If* $\sqrt{n}\{T_n - b(\theta)\}$ *is* $AN(0, 1)$ *under* $\theta \in \Theta - \Theta_0$ *for some positive* $b(\theta)$ *then* $n^{-1/2}\log\{\sqrt{n}\bar{\Phi}(\sqrt{n}T_n)\} - \sqrt{n}b^2(\theta)/2$ *is* $AN(0, b^2(\theta))$ *under* θ.

PROOF. Let $\{b_n\}$ be a sequence of positive numbers satisfying $b_n = b(\theta) + O(n^{-1/2})$. By Mills' ratio (Feller, 1968, page 175)

$$\frac{1}{\sqrt{n}b_n}\left(1 - \frac{1}{nb_n^2}\right) \le \frac{1}{\sqrt{2\pi}} \exp(\tfrac{1}{2}nb_n^2)\bar{\Phi}(\sqrt{n}b_n) \le \frac{1}{\sqrt{n}b_n}.$$

It follows that the implication (ii) $\Rightarrow$ (iii) of Lemma 4.1 holds for $\bar{G}_n(t) = \bar{\Phi}(\sqrt{n}t)$ with $f(b) = b^2/2$. In the statement of the theorem, $\bar{\Phi}$ is multiplied by $\sqrt{n}$ to improve the lognormal approximation to the finite sample distribution of the P-value. □

The implication (ii) $\Rightarrow$ (iii) is verified for other unconditional P-values in Section 5. Because G_n of (ii) is nonrandom, Lemma 4.1 does not apply to conditional P-values for which the exact null df G_n depends on the sample path. However, it is straightforward to generalize Lemma 4.1 to obtain the following Lemma 4.2.

LEMMA 4.2. *Assume that the test statistic T_n satisfies* (i) *and, for each n, G_n is a random function with values in* $[0, 1]$, *i.e.,* $G_n\colon \mathscr{R} \times \mathscr{S} \to [0, 1]$. *Let* $\bar{G}_n = 1 - G_n$ *and define the conditional P-value* $P_n(s) = \bar{G}_n(T_n(s), s)$. *Suppose that associated with* $\{G_n\}$ *and given nonnull θ there is a real valued function f defined on an open interval $\mathscr{I}$ containing $b(\theta)$ such that the following holds: if* (ii) $b \in \mathscr{I}$ *and* $\{b_n\}$ *is a sequence of numbers that satisfies* $b_n = b + O(n^{-1/2})$ *then* (iii′) $\log \bar{G}_n(b_n) = -nf(b_n) + o_p(\sqrt{n})$ *as* $n \to \infty$. *Also assume f satisfies* (iv). *Then* $n^{-1/2}\{\log P_n + nc(\theta)\}$ *is* $AN(0, \tau^2(\theta))$ *where* $c(\theta) = f(b(\theta))$ *and* $\tau(\theta) = \sigma(\theta)h(\theta)$.

The implication (ii) $\Rightarrow$ (iii′) is often difficult to prove. It is verified for the two-sample Fisher-Pitman permutation test in Section 5.

5. Proofs of asymptotic lognormality. Four theorems are given. The implication (ii) $\Rightarrow$ (iii) of Lemma 4.1 is proved first for P-values that are based on sums of i.i.d. r.v.'s and then for one and two sample rank P-values. Finally, the implication (ii) $\Rightarrow$ (iii′) of Lemma 4.2 is proved for two sample conditional permutation P-values.

In each case a large deviation probability of the type $P(\Sigma X_{in} \geq nb_n)$, b_n a constant, is approximated by applying a variant of Cramér's device (Feller, 1971). Since the right tail behavior of the df of ΣX_{in}, F_n say, is of interest, and the central behavior is not, $F_n(dx)$ is first multiplied by $\exp(\beta_n x)$ for a positive β_n and then renormalized by division by $\phi_n(\beta_n)$, the moment generating function of F_n at β_n. For each n, β_n is chosen so that the mean of

$$e^{\beta_n x}F_n(dx)/\phi_n(\beta_n) = F_{b_n}(dx),$$

say, is nb_n, i.e. β_n is chosen so that the given tail probability of F_n may be expressed in terms of the central behavior of F_{b_n}. Next F_{b_n}, which is the df of a sum, is approximated by a normal df by an application of the Liapunov-Berry-Esseen Theorem.

Other authors have used similar techniques to obtain implications like (ii) $\Rightarrow$ (iii). For example, Petrov (1965) proves a stronger version of the implication (ii) $\Rightarrow$ (iii) for i.i.d. r.v.'s with a finite moment generating function on an interval $[0, B]$, $B > 0$. Klotz (1965) and Stone (1967, 1969) consider the case $b_n \equiv b$ for one- and two-sample rank sums respectively and obtain pointwise convergence at rate $o(n)$. The proof in Klotz (1965) is deficient, however, because it does not allow for the possibility of a negative error in the normal approximation of F_{b_n}. Here we obtain convergence at rate $o(\sqrt{n})$ with $b_n = b + o(1)$ for one- and two-sample rank sums and two-sample permutation sums under modest regularity conditions. Uniform convergence at a rate faster than $o(\sqrt{n})$ was considered in Lambert (1978).

The following Lemma 5.1 shows that the validity of (ii) $\Rightarrow$ (iii) depends on the rates of convergence of β_n, $\phi_n(\beta_n)$ and $\phi_n''(\beta_n)/\phi_n(\beta_n)$.

LEMMA 5.1. *Let* $X_{1n}, \cdots, X_{nn}$ *be independent* r.v.'s, $n = 1, 2, \cdots$, X_{in} *having a nondegenerate* df F_{in} *and a finite moment generating function* (mgf) ϕ_{in} *on an interval* $[0, B]$, $B > 0$. *Suppose for each term of a convergent sequence of constants* $\{b_n\}$ *there is a unique* β_n *which maximizes* $e^{nb_n\beta}/\phi_n(\beta)$ *where* $\phi_n = \Pi\phi_{in}$ *is the* mgf *of* ΣX_{in}. *For each i, n construct a* r.v. Z_{in} *with* df F_{ib_n} *defined by*

$$F_{ib_n}(dx) = e^{\beta_n x}F_{in}(dx)/\phi_{in}(\beta_n).$$

Let

$$\sigma_n^2 = \Sigma \operatorname{var}(Z_{in}), \qquad k_n = \sigma_n\beta_n \quad \text{and} \quad \rho_n = \Sigma E\,|Z_{in} - EZ_{in}|^3.$$

Assume $k_n \to \infty$ *and* $\rho_n/\sigma_n^3 = o(1)$ *as* $n \to \infty$. *Then*

$$-\log P(\Sigma X_{in} \geq nb_n) = nb_n\beta_n - \sum \log \phi_{in}(\beta_n) + o(k_n).$$

PROOF. First note that Z_{in} has mgf $\psi_{in}(t) = \phi_{in}(\beta_n + t)/\phi_{in}(\beta_n)$ which is finite on an interval around zero. Hence, all moments of Z_{in} are finite and may be obtained by differentiation of $\psi_{in}(t)$. In particular,

$$\Sigma EZ_{in} = nb_n \quad \text{and} \quad \sigma_n^2 = \frac{\phi_n''(\beta_n)}{\phi_n(\beta_n)} - n^2b_n^2.$$

Let H_n be the df of $\sigma_n^{-1}(\Sigma Z_{in} - nb_n)$. Then

$$\begin{aligned} P(\Sigma X_{in} \geq nb_n) &= \int_{\Sigma x_i \geq nb_n} \cdots \int \Pi F_{in}(dx_i) \\ &= \phi_n(\beta_n) \int_{\Sigma x_i \geq nb_n} \cdots \int e^{-\beta_n \Sigma x_i} \Pi F_{ib_n}(dx_i) \\ &= \phi_n(\beta_n) e^{-nb_n\beta_n} I_n, \end{aligned}$$

where $I_n = \int_0^\infty e^{-k_n z} H_n(dz)$. As in the proof of Theorem 2.2 of Bahadur (1971), for any $\varepsilon > 0$,

$$I_n \geq \int_0^\varepsilon e^{-k_n z} H_n(dz) \geq e^{-k_n \varepsilon}\{H_n(\varepsilon) - H_n(O)\}.$$

It follows by the Liapunov-Berry-Esseen Theorem that $H_n(z) \to \Phi(z)$ for each z, since $\rho_n/\sigma_n^3 \to 0$. Hence, $k_n^{-1} \log I_n$ is asymptotically bounded from below by $-\varepsilon$. Because ε is arbitrary and $0 < I_n < 1$, it follows that $\log I_n = o(k_n)$. □

The implication (ii) ⇒ (iii) for sums of i.i.d. r.v.'s follows easily from Lemma 5.1; the proof is omitted.

THEOREM 5.1. *Suppose* $\{X_i\}$ *is a sequence of* i.i.d. r.v.'*s with a common* df F *and* mgf ϕ *that is finite on an interval* $[0, B]$, $B > 0$. *Let* $\mathscr{I}_B$ *be the interval* $(EX_1, \lim_{t\to\infty} \psi(t))$ *where* $\psi(t) = \phi'(t)/\phi(t)$. *Then for a sequence of constants* $\{b_n\}$ *in* $\mathscr{I}_B$ *satisfying* $b_n = b + o(1)$ *for some* $b \in \mathscr{I}_B$,

$$-\log P(\Sigma X_i \geq nb_n) = n\beta_n b_n - n \log \phi(\beta_n) + o(\sqrt{n})$$

where β_n *is the solution to* $\psi(\beta_n) = b_n$.

The following Theorem 5.2 for one-sample signed rank P-values is also a consequence of Lemma 5.1. First, it is convenient to relate the tail behavior of the quantile function G^{-1} to the number of moments of G.

LEMMA 5.2. *Let* G *be an absolutely continuous* df *with nonnegative support* $\mathscr{S}$, *density* g *and inverse* Q. *If* $\int x^r\, dG(x) < \infty$ *for some* $r > 1$, *then* $Q(1 - 1/n) = o(n^{1/r})$. *If, in addition,* $\lim_{x\to\infty}\{\bar{G}(x)/g(x)\}$ *is finite, then* $\int_{1-1/n}^1 Q(x)\, dx = o(n^{1/r-1})$.

PROOF. Assume $\mathscr{S}$ is unbounded, since the lemma is trivial otherwise. Let $Q(1 - 1/n) = x$. Then

$$Q\left(\frac{n-1}{n}\right) = o(n^{1/r}) \quad \text{iff} \quad x = o(\bar{G}(x)^{-1/r}),$$

or

$$\text{iff } x^r \bar{G}(x) = o(1).$$

The last equality is true since $\int x^r \, dG(x) < \infty$. Similarly,

$$\int_{1-1/n}^{1} Q(x)\, dx = o(n^{1/r-1}) \quad \text{iff} \quad \int_{y \ge x} y\, dG(y) = o(\bar{G}(x)^{1-1/r})$$

or, by an integration by parts,

$$\text{iff} \quad x\bar{G}(x)^{1/r} + \bar{G}(x)^{1/r-1} \int_{y \ge x} \bar{G}(y)\, dy = o(1).$$

An application of L'Hopital's rule to $\int_{y \ge x} \bar{G}(y)\, dy / \bar{G}(x)$ completes the proof. □

THEOREM 5.2. *Given a sequence* $\{Y_i\}$ *of* i.i.d. r.v.'*s that satisfy* $P(Y_1 \le 0) = P(Y_1 \ge 0) = ½$, *let* R_{in} *be the rank of* $|Y_i|$ *in* $|Y_1|, \cdots, |Y_n|$ *and define a signed rank sum statistic* $S_n = \Sigma \text{sgn}(Y_i) a_n(R_{in})$. *Suppose the scores* $a_n(i)$ *are generated by an absolutely continuous increasing* df G *with positive support, density* g, *inverse* Q *and finite third moment as follows*: *either* $a_n(i) = Q(i/(n+1))$ *or* $a_n(i) = EQ(U_{in})$ *where* U_{in} *denotes the ith largest order statistic of a random sample of size n from a uniform* (0, 1) *distribution. Suppose further that* $\lim_{x\to\infty} \bar{G}(x)/g(x) < \infty$. *Finally, let* $\{b_n\}$ *be a sequence of positive constants satisfying* $\sup b_n < \sup_t \int x \tanh(tx)\, dG(x)$ *and* $b_n = b + o(1)$ *for* $b > 0$. *Then if* $a_n(i) = Q(i/(n+1))$,

$$-\log P(S_n \ge nb_n) = n\beta_n b_n - n \int \log \cosh(\beta_n x)\, dG(x) + o(\sqrt{n})$$

where β_n *is the unique solution to* $\int x \tanh(\beta_n x)\, dG(x) = b_n$. *The conclusion also holds for* $a_n(i) = EQ(U_{in})$ *if both g and* $\bar{G}/g$ *are nonincreasing.*

PROOF. Assume for now that $a_n(i) = EQ(U_{in})$; the modifications for $a_n(i) = Q(i/(n+1))$ will be obvious.

As in Klotz (1965), let the r.v.'s $X_{1n}, \cdots, X_{nn}$ be defined by $X_{in} = \text{sgn}(Y_i)E_{in}$, where $E_{in} = EQ(U_{in})$. The mgf of X_{in} is $\phi_{in}(t) = \cosh(tE_{in})$. For each i, n, construct a r.v. Z_{in} corresponding to X_{in} as in Lemma 5.1. Note that Z_{in} assumes values $\pm E_{in}$ with probabilities ½ $\exp(\pm\beta_{nn}E_{in})/\cosh(\beta_{nn}E_{in})$ where $\beta_{nn} > 0$ is uniquely defined by $\Sigma E_{in} \tanh(\beta_{nn}E_{in}) = nb_n$. Also $\Sigma EZ_{in} = nb_n$ and $\sigma_n^2 = \Sigma \text{ var } Z_{in} = \Sigma E_{in}^2 \text{ sech}^2(\beta_{nn}E_{in})$. Applying Lemma 5.1 gives

$$-\log P(S_n \ge nb_n) = n\beta_{nn} b_n - \Sigma \log \cosh(\beta_{nn}E_{in}) + o(\sqrt{n})$$

if (a) $\sigma_n \beta_{nn}$ has exact order $n^{1/2}$ and (b) $\rho_n/\sigma_n^3 \to 0$.

Conditions (a) and (b) both hold if the sequences $\{n^{-1}\sigma_n^2\}$, $\{n^{-1}\rho_n\}$ and $\{\beta_{nn}\}$ converge to positive constants. Since $\Sigma E_{in}^3 \to \int x^3\, dG(x)$ (Hoeffding, 1953), the first two sequences converge to positive constants whenever β_{nn} has a positive limit.

If we define the df G_n by $G_n(x) = \#(E_{in} < x)/n$ and expand $\tanh(\beta_{nn}x)$ around βx, where β is defined by $\int x \tanh(\beta x)\, dGF(x) = b$, then

$$\int x \tanh(\beta x)\, d\{G_n(x) - G(x)\} - (\beta_{nn} - \beta) \int x^2 \text{ sech}^2(\beta_n' x)\, dG_n(x) = b_n - b$$

for some $\beta_n' \in (\beta_{nn}, \beta)$. Hence $\beta_{nn} - \beta = o(1)$ as required.

A Taylor expansion gives

$$-\log P(S_n \ge nb_n) = n\beta_n b_n - \Sigma \log \cosh(\beta_n E_{in}) + O(n(\beta_{nn} - \beta_n)^2) + o(\sqrt{n}).$$

We next show $\beta_{nn} - \beta_n = O(n^{-1/2})$. By definition of β_{nn} and β_n,

$$\int x \tanh(\beta_n x)\, d\{G_n(x) - G(x)\} + (\beta_{nn} - \beta_n) \int x^2 \operatorname{sech}^2(\beta'_{nn} x)\, dG_n(x) = 0.$$

By a Taylor expansion of $E_{in} \tanh(\beta_n E_{in})$ around $Q(i/(n+1))$,

$$\int x \tanh(\beta_n x)\, d\{G_n(x) - G(x)\}$$

$$(2) \qquad = \int_0^{(n-1)/(n+1)} [Q(x_n)\tanh\{\beta_n Q(x_n)\} - Q(x)\tanh\{\beta_n Q(x)\}]\, dx$$

$$+ n^{-1} Q\left(\frac{n}{n+1}\right)\tanh\left(\beta_n Q\left(\frac{n}{n+1}\right)\right) - \int_{(n-1)/(n+1)}^{1} Q(x)\tanh(\beta_n Q(x))\, dx$$

$$+ n^{-1} \sum \left(E_{in} - Q\left(\frac{i}{n+1}\right)\right)\{\tanh(\beta_n E'_{in}) + \beta_n E'_{in} \operatorname{sech}^2(\beta_n E'_{in})\}$$

where $x_n = i/(n+1)$ for $(i-1)/(n+1) < x \le i/(n+1)$ and $E'_{in} \in (E_{in}, Q(i/(n+1)))$. Since $\Sigma | E_{in} - Q(i/(n+1)) | = O(\sqrt{n})$ (Hoeffding, 1973) and $0 < \tanh x + x \operatorname{sech}^2 x < 2$ for $x > 0$, the last term on the right is $O(n^{-1/2})$. By a Taylor expansion of $Q(x_n)\tanh\{\beta_n Q(x_n)\}$, for $x'_n \in (x_n, x)$, the first term on the right is $O(n^{-1}) \int_{y \le (n-1)/(n+1)} Q'(x'_n)\, dx$, which is no larger than $O(n^{-1})Q(n/(n+1))$. By Lemma 5.2, this term and the second and third terms of (2) are $o(n^{-2/3})$. Therefore, if $a_n(i) = E_{in}$ then $\beta_{nn} - \beta_n = O(n^{-1/2})$, whereas if $a_n(i) = Q(i/(n+1))$ then $\beta_{nn} - \beta_n = o(n^{-2/3})$.

Finally, for $a_n(i) = Q(i/n + 1))$, a Taylor expansion similar to the one in equation (1) gives

$$\Sigma \log \cosh\left(\beta_n Q\left(\frac{i}{n+1}\right)\right) = n \int \log \cosh(\beta x)\, dG(x) + o(n^{1/3}).$$

If g is nonincreasing, then $E_{in} \ge Q(i/(n+1))$, and if $\bar{G}/g$ is nonincreasing, then $E_{in} \le Q(i/(n + \tfrac{1}{2}))$ (van Zwet, 1964). Consequently, under these conditions

$$\Sigma \log \cosh(\beta_n E_{in}) = n \int \log \cosh(\beta_n x)\, dG(x) + o(n^{1/3})$$

also, which completes the proof of the theorem. □

In Theorem 5.2 the assumption that g and $\bar{G}/g$ are nonincreasing can be replaced with $\Sigma | E_{in} - Q(i/(n+1)) | = o(\sqrt{n})$. Of course, if $\Sigma | E_{in} - Q(i/(n+1)) | \ne o(\sqrt{n})$, then

$$-\log P(S_n \ge nb_n) = n\beta_n b_n - \Sigma \log \cosh(\beta_n E_{in}) + o(\sqrt{n})$$

and the asymptotic normal distribution of the corresponding log P-value need not be centered at $n\beta b - n \int \log \cosh(\beta x)\, dG(x)$.

We next prove the implications required for asymptotic lognormality of two-sample rank and permutation P-values. Henceforth, if $a \in [0, 1]$ then $\bar{a} = 1 - a$.

LEMMA 5.3. *Given a vector* $\mathbf{Y}_n = (Y_{1n}, \cdots, Y_{nn})$ *of* $n \ge 2$ *ordered constants* ($Y_{in} < Y_{jn}$ *for* $i < j$), *let* $T_n(\lambda_n)$ *be the sum of a simple random sample without replacement of size* $n\lambda_n (0 < \lambda_n < 1)$ *from* $\mathbf{Y}_n$. *Suppose* (i) $\lambda_n - \lambda = O(n^{-1})$ *for some* $\lambda \in (0, 1)$. *Define* $G_n(y) = \#(Y_{in} \le y)/n$ *and assume* (ii) G_n *converges uniformly to a continuous* df G *with inverse* Q *and* (iii) $\int |y|^3\, dG_n(y) \to \int |y|^3\, dG(y) < \infty$. *Let* $\{b_n\}$ *be a sequence of constants* *satisfying* (iv) $b_n = b + o(1)$ *and* (v) *both* b_n *and* b *belong to* $(\int \lambda Q(y)\,dy, \int_{\bar\lambda}^1 Q(y)\,dy)$. *Further assume* (iv) $G_n(A) = 0$ *for all n and some finite A. Then*

$$\text{(vii)}\qquad \begin{aligned} -n^{-1}\log P\{T_n(\lambda_n) \ge nb_n\} &= \beta_{nn} b_n + \bar\lambda_n \log \gamma_{nn} - \lambda_n \log \lambda_n - \bar\lambda_n \log \bar\lambda_n \\ &\quad - n^{-1}\Sigma \log(\gamma_{nn} + e^{\beta_{nn} Y_{in}}) + o(\sqrt{n}) \\ &= nc_n(\lambda_n, b_n) + o(\sqrt{n}) \end{aligned}$$

where β_{nn} *and* γ_{nn} *are determined by*

$$\int (1 + \gamma_{nn} e^{-\beta_{nn} y})^{-1}\, dG_n(y) = \lambda_n, \qquad \int y(1 + \gamma_{nn} e^{-\beta_{nn} y})^{-1}\, dG_n(y) = b_n.$$

If $\mathbf{Y}_n$ *is random and* G_n *satisfies* (ii), (iii) *and* (vi) a.s. $[G]$ *then* (vii) *holds* a.s. $[G]$.

PROOF. Assume for now that $\bar G_n(0) = \bar G(0) = 1$ for all n.

The distribution of $T_n(\lambda_n)$ is the same as the conditional distribution of $\tilde T_n = \Sigma\, U_i Y_{in}$ given $\Sigma\, U_i = n\lambda_n$ where $U_1, \cdots, U_n$ are i.i.d. Bernoulli (p) r.v.'s for any $p \in (0, 1)$. Therefore

$$\text{(3)}\qquad P(\tilde T_n \ge nb_n) = \textstyle\sum_1^n P\left\{T_n\left(\frac{k}{n}\right) \ge nb_n\right\} P(\sum_1^n U_i = k).$$

Lemma 5.1 applies to $P(\tilde T_n \ge nb_n)$ with $X_{in} = U_i Y_{in}$, $\phi_{in}(t) = \bar p + pe^{tY_{in}}$ and

$$Z_{in} = \begin{cases} 0 & \text{with probability } \bar p(\bar p + pe^{\beta_{nn} Y_{in}})^{-1} \\ Y_{in} & \text{with probability } pe^{\beta_{nn} Y_{in}}(\bar p + pe^{\beta_{nn} Y_{in}})^{-1} \end{cases}$$

where β_{nn} is determined by $n^{-1}\,\Sigma\, Y_{in}(1 + \bar p p^{-1} \exp(-\beta_{nn} Y_{in}))^{-1} = b_n$. It is convenient to choose $p_n = (1 + \gamma_{nn})^{-1}$ where γ_{nn} is determined by $\lambda_n = n^{-1}\,\Sigma\,(1 + \gamma_{nn} \exp(-\beta_{nn} Y_{in}))^{-1}$. For this choice of p_n, the expected number of nonzero Z_{in} equals $n\lambda_n$. The existence of β_{nn} and γ_{nn} is guaranteed by (v); their convergence is proved in the appendix, Section 7. Since

$$\rho_n = \Sigma\, E\,|Z_{in} - EZ_{in}|^3 \le \Sigma\, E\,|Y_{in}|^3$$

and

$$\sigma_n^2 = \Sigma \operatorname{var}(Z_{in}) = \Sigma\, Y_{in}^2 \gamma_{nn} e^{\beta_{nn} Y_{in}} (\gamma_{nn} + e^{\beta_{nn} Y_{in}})^{-2},$$

it follows from the proof of Lemma 5.1 that

$$-\log P(\tilde T_n \ge nb_n) = n\beta_{nn} b_n - \sum \log\left(\frac{\gamma_{nn} + e^{\beta_{nn} Y_{in}}}{\gamma_{nn} + 1}\right) + A_n$$

where A_n is nonnegative and has order $o(\sqrt{n})$. If we approximate $\binom{n}{k}$ in $P(\Sigma\, U_i = k)$ by Stirling's formula, we may rewrite equation (3) as

$$\begin{aligned} 1 = \sum \exp\bigg[&\log P\left\{T_n\left(\frac{k}{n}\right) \ge nb_n\right\} + n\beta_{nn} b_n - \sum \log(\gamma_{nn} + e^{\beta_{nn} Y_{in}}) \\ &- k \log\left(\frac{k}{n}\right) - (n-k)\log\left(\frac{n-k}{n}\right) + (n-k)\log \gamma_{nn} + A_n - \log \sqrt{n} + O(1) \bigg]. \end{aligned}$$

For n sufficiently large, the remainder term $r_n(\lambda_n, b_n)$ defined by

$$\begin{aligned} r_n(\lambda_n, b_n) = \log P\{T_n(\lambda_n) \ge nb_n\} + n\beta_{nn} b_n + n\bar\lambda_n \log \gamma_{nn} \\ - n\lambda_n \log \lambda_n - n\bar\lambda_n \log \bar\lambda_n - \sum \log\,(\gamma_{nn} + e^{\beta_{nn} Y_{in}}) \end{aligned}$$

is bounded above by $\log \sqrt{n}$ + a constant, because A_n is nonnegative.

Suppose $r_n(\lambda_n, b_n) < -\,(A_n + \log n)$ for n large. Then $r_n(k/n, b_n) < -\,(A_n + \log n)$ for $1 \le k \le n$. To see this, define $\log P\{T_n(\lambda) \ge nb_n\}$ by linear interpolation if $n\lambda \in (i, i + 1)$,

$0 < i < n$. With this convention, Stone (1969) shows that $r_n(\lambda, b) = o(n)$ uniformly in λ. The convergence is also uniform in b on compact sets. For, β_{nn} and γ_{nn} converge uniformly to continuous functions β and γ of (λ, b) and the sequence $\{\log P[T_n(\lambda) \geq nb]\}$, which is monotonic in b and λ, converges to a continuous function. Hence, $r_n(\lambda_n, b_n) = o(n)$ and $r_n(\lambda_n^*, b_n) \neq o(n)$ if $\lambda_n^* = \lambda^* + O(n^{-1})$ and $\lambda^* \neq \lambda$ (with γ_{nn} and β_{nn} defined as above). In particular, $|r_n(\lambda_n, b_n)| > A_n + \log n$ implies $|r_n(\lambda_n^*, b_n)| > A_n + \log n$ for n large. But $r_n(\lambda_n^*, b_n) < \log \sqrt{n}$ + a constant so $r_n(\lambda_n^*, b_n) < -(A_n + \log n)$ for n large as claimed. Hence,

$$1 = \Sigma \exp\left\{r_n\left(\frac{k}{n}, b_n\right) + A_n - \log \sqrt{n} + O(1)\right\} < n \exp\{-3 \log \sqrt{n} + O(1)\} = o(1),$$

which is a contradiction. Therefore $r_n(\lambda_n, b_n) = o(\sqrt{n})$.

To complete the proof for $\bar{G}_n(0) < 1$ and $\bar{G}_n(A) = 1$ for some $A > -\infty$, shift each Y_{in} to the right by $-A$ units, observe that $P\{T_n(\lambda_n) \geq nb_n\} = P\{T_n(\lambda_n, A) \geq n(b_n + \lambda_n A)\}$, where $T_n(\lambda_n, A)$ is the sum of a simple random sample without replacement of size $n\lambda_n$ from $Y_{in} - A, \cdots, Y_{nn} - A$, and apply the above result.

The restriction that G be supported on $[A, \infty]$ is removed in two steps. First, G_n and G are truncated at A and renormalized. The function $c_n(\lambda_n, b_n, A)$, which is $c_n(\lambda_n, b_n)$ with the truncated G_n and G, is then approximated by a function $c(\lambda_n, b_n, A)$ which is asymptotically equivalent uniformly in A. Second, the restriction to finite A is removed. The first step requires a careful analysis of orders of convergence and different techniques are needed for random and nonrandom $\mathbf{Y}_n$. We first consider nonrandom $\mathbf{Y}_n$. □

LEMMA 5.4. *Let scores $\{Y_{in}\}$ be generated by an absolutely continuous increasing* df *G with density g, inverse Q and finite absolute third moment as follows: either* (viii) $Y_{in} = Q(i/(n+1))$ *or* (ix) $Y_{in} = EQ(U_{in})$ *where U_{in} denotes the ith largest order statistic of a random sample of size n from a uniform* (0, 1) *distribution. Assume $\bar{G}(x)/g(x)$ and $G(-x)/g(-x)$ have finite limits as $x \to \infty$. Let the sequences of constants $\{\lambda_n\}$ and $\{b_n\}$ satisfy conditions* (i) *and* (iv) *through* (v) *of Lemma* 5.3.

Choose and fix an A_0 such that $\bar{G}_n(A_0) > 0$ for all $n \geq 2$. For every $A < A_0$, define a df *G_{nA} according to*

$$G_{nA}(x) = \begin{cases} \dfrac{G_n(x) - G_n(A)}{\bar{G}_n(A)} & x > A \\ 0 & x \leq A. \end{cases}$$

Define G_A from G by analogy. Finally, define functions $c_n(\lambda_n, b_n, A)$ and $c(\lambda_n, b_n, A)$ by

$$c_n(\lambda_n, b_n, A) = \beta_{nA} b_n + \bar{\lambda}_n \log \gamma_{nA} - \lambda_n \log \lambda_n - \bar{\lambda}_n \log \bar{\lambda}_n - \int \log(e^{\beta_{nA} y} + \gamma_{nA})\, dG_{nA}(y),$$

$$c(\lambda_n, b_n, A) = \beta_A b_n + \bar{\lambda}_n \log \gamma_A - \lambda_n \log \lambda_n - \bar{\lambda}_n \log \bar{\lambda}_n - \int \log(e^{\beta_A y} + \gamma_A)\, dG_A(y),$$

where β_{nA} and γ_{nA} are determined by

$$\lambda_n = \int (1 + \gamma_{nA} e^{-\beta_{nA} y})^{-1}\, dG_{nA}(y),\ b_n = \int y\,(1 + \gamma_{nA} e^{-\beta_{nA} y})^{-1}\, dG_{nA}(y)$$

and β_A, γ_A are defined similarly but with G_A replacing G_{nA}. Then

$$c_n(\lambda_n, b_n, A) - c(\lambda_n, b_n, A) = o(n^{-1/2})$$

uniformly in A if $\{Y_{in}\}$ is defined by (viii). *The conclusion is also valid for $\{Y_{in}\}$ defined by* (ix) *if* (x) $Q((i-1)/(n+1)) \leq Y_{in} \leq Q(i/n)$ *for* $2 \leq i \leq n-1$.

REMARK. van Zwet (1964) gives conditions on G which imply (x). The requirement

(x) may be replaced by the weaker condition that $\Sigma | Y_{in} - Q(i/(n+1)) | = o(n^{1/2})$ which is more difficult to verify.

PROOF. By a Taylor expansion and the definitions of β_{nA} and γ_{nA}

$$c_n(\lambda_n, b_n, A) - c(\lambda_n, b_n, A) = -\int \log(\gamma_A + e^{\beta_A y})\, d\{G_{nA}(y) - G_A(y)\} + O((\beta_{nA} - \beta_A)^2) + O((\gamma_{nA} - \gamma_A)^2).$$

We first show $\beta_{nA} - \beta_A = O(n^{-1/2})$ and $\gamma_{nA} - \gamma_A = O(n^{-1/2})$. Here and throughout the proof of the lemma, the $O(\cdot)$ and $o(\cdot)$ terms are taken to be uniform in A as n increases.

As in Lemma 7.1, for both $m = 0$ and $m = 1$,

$$(4)\quad 0 = \int y^m(1 + \gamma_A e^{-\beta_A y})^{-1} d\{G_{nA}(y) - G_A(y)\} - \int \frac{(\gamma_{nA} - \gamma_A)y^m e^{-\beta_A y} - (\beta_{nA} - \beta_A)\gamma_A y^{m+1} e^{-\beta_A y}}{(1 + \gamma_{nA} e^{-\beta_{nA} y})^2}\, dG_{nA}(y) + O((\beta_{nA} - \beta_A)^2) + O((\gamma_{nA} - \gamma_A)^2).$$

Define $n_A = \min\{i\colon Y_{in} \geq A\}$ and $\bar{p}_{nA} = G_n(A)$, and write

$$\int y^m(1 + \gamma_A e^{-\beta_A y})^{-1}\, dG_{nA}(y) = (np_{nA})^{-1} \sum_{i \geq n_A} Q^m\left(\frac{i}{n+1}\right)\left(1 + \gamma_A e^{-\beta_A Q\left(\frac{i}{n+1}\right)}\right)^{-1} + (np_{nA})^{-1} \sum_{i \geq n_A} \left\{Y_{in} - Q\left(\frac{i}{n+1}\right)\right\}\left\{\frac{m(Y'_{in})^{m-1}}{1 + \gamma_A e^{-\beta_A Y'_{in}}} + \frac{\gamma_A \beta_A (Y'_{in})^m e^{-\beta_A Y'_{in}}}{(1 + \gamma_A e^{-\beta_A Y'_{in}})^2}\right\}$$

for some $Y'_{in} \in \left[Y_{in}, Q\left(\frac{i}{n+1}\right)\right]$.

The last sum is $O(n^{-1/2})$ (Hoeffding, 1973). The first and last terms of the first sum may be omitted since each is $o(n^{-2/3})$ by an obvious extension of Lemma 5.2. Therefore,

$$\int y^m(1 + \gamma_A e^{-\beta_A y})^{-1}\, dG_{nA}(y) = p_{nA}^{-1} \int_{\frac{n_A}{n+1}}^{\frac{n-1}{n+1}} Q^m(x_n)\{1 + \gamma_A e^{-\beta_A Q(x_n)}\}^{-1}\, dx + O(n^{-1/2}),$$

where $x_n = i/(n+1)$ if $(i-1)/(n+1) \leq x < i/(n+1)$. A Taylor expansion about x of the right side integrand then gives

$$(5)\quad \int y^m(1 + \gamma_A e^{-\beta_A y})^{-1}\, d\{G_{nA}(y) - G_A(y)\} = O(n^{-1/2})$$

if

$$(6)\quad p_{nA}^{-1} \int_{\frac{n_A}{n+1}}^{\frac{n}{n+1}} Q^m(x)\{1 + \gamma_A e^{-\beta_A Q(x)}\}^{-1}\, dx - \bar{G}(A)^{-1} \int_{G(A)}^{1} Q^m(x)\{1 + \gamma_A e^{-\beta_A Q(x)}\}^{-1}\, dx = O(n^{-1/2}).$$

When $G(A) = 0$ for some finite A, $p_{nA} = \bar{G}(A) = 1$ and the last condition (6) holds by Lemma 5.2. When the support of G is unbounded below, the left side of condition (6) is $O(p_{nA} - \bar{G}(A)) + o(n^{-2/3})$. If $Y_{in} = Q(i/(n+1))$ then n_A is the largest integer in $(n+1)G(A)$; if $Y_{in} = EQ(U_{in})$ then under assumption (x) n_A is between $nG(A) - 1$ and $(n+1)G(A) + 1$. In either case, $p_{nA} = \bar{G}(A) + O(n^{-1})$.

Returning to (4), we now have

$$\int \frac{(\gamma_{nA} - \gamma_A)y^m e^{-\beta_A y} - (\beta_{nA} - \beta_A)\gamma_A y^{m+1} e^{-\beta_A y}}{(1 + \gamma_{nA} e^{-\beta_{nA} y})^2} \, dG_{nA}(y)$$
$$= O((\beta_{nA} - \beta_A)^2) + O((\gamma_{nA} - \gamma_A)^2) + O(n^{-1/2}).$$

Consequently, both $\beta_{nA} - \beta_A$ and $\gamma_{nA} - \gamma_A$ are $O(n^{-1/2})$; cf. the proof of Lemma 7.1.

In summary, $c_n(\lambda_n, b_n, A) - c(\lambda_n, b_n, A) = o(n^{-1/2})$ if

$$(np_{nA})^{-1} \textstyle\sum_{i \geq n_A} \log(\gamma_A + e^{\beta_A Y_{in}}) = \displaystyle\int \log(\gamma_A + e^{\beta_A y}) \, dG_A(y) + o(n^{-1/2}).$$

The last term in the sum on the left may be disregarded since

$$n^{-1} E_{nn} \leq \int_{\frac{n-1}{n}}^{1} Q(x) \, dx = o(n^{-2/3})$$

(Lai and Robbins, 1976). Under assumption (x), the remaining terms of the sum are bounded by

$$p_{nA}^{-1} \int_{\frac{n_A}{n+1}}^{\frac{n}{n+1}} \log(\gamma_A + e^{\beta_A Q(x_n)}) \, dx,$$

where $x_n = i/(n+1)$ when $i/(n+1) \leq x < (i+1)/(n+1)$, and by

$$p_{nA}^{-1} \int_{\frac{n_{A-1}}{n}}^{\frac{n-1}{n}} \log(\gamma_A + e^{\beta_A Q(x_n')}) \, dx,$$

where $x_n' = i/n$ when $(i-1)/n < x \leq i/n$, if n is sufficiently large that $n_A \geq 2$. Since both bounds equal $\int \log(\gamma_A + e^{\beta_A x}) \, dG_A(x) + o(n^{-2/3})$ (cf. the derivation of equation (5)), the proof of the lemma is complete. □

THEOREM 5.3. *Under the conditions of Lemma* 5.4,

$$-\log P(T_n \geq nb_n) = nc(\lambda_n, b_n) + o(n^{-1/2})$$

where

$$c(\lambda_n, b_n) = \beta_n b_n + \bar{\lambda}_n \log \gamma_n - \int \log(\gamma_n + e^{\beta_n y}) \, dG(y) - \lambda_n \log \lambda_n - \bar{\lambda}_n \log \bar{\lambda}_n$$

and β_n, γ_n *are the solutions to*

$$\lambda_n = \int (1 + \gamma_n e^{-\beta_n y})^{-1} \, dG(y); \qquad b_n = \int y(1 + \gamma_n e^{-\beta_n y})^{-1} \, dG(y).$$

PROOF. If $G(A) = 0$ for some finite A then the Theorem is already established in Lemmas 5.3 and 5.4. If the support of G is unbounded, then the theorem is proved by an extension of Section 3 of Stone (1969). The extension will be sketched for completeness.

As in Lemma 5.3, let $T_n(\lambda_n)$ be the sum of a simple random sample taken without replacement from G_n. Let nS_{nA} be the sum of those sampled values which are greater than A $(A < 0)$. Say $n\lambda_{nA}$ sampled values exceed A. Then $P\{T_n(\lambda_n) \geq nb_n\} \leq P(S_{nA} \geq b_n) = EP(S_{nA} \geq nb_n \mid \lambda_{nA})$. For each fixed $\lambda_{nA} \leq \lambda_n$, Lemmas 5.3 and 5.4 imply

$$(7) \qquad -(np_{nA})^{-1} \log P(S_{nA} \geq nb_n \mid \lambda_{nA}) = c(\lambda_{nA}/p_{nA}, b_n, A) + \epsilon_{nA}$$

where $\bar{p}_{nA} = G_n(A)$ and $\epsilon_{nA} = o(n^{-1/2})$ uniformly in A.

Plainly, the sequence $\{\exp(-\, nc(\lambda_{nA}/p_{nA}, b_n, A))\}$ is uniformly integrable. It also converges in probability uniformly in n as $A \to -\infty$. For, $\lambda_{nA}/p_{nA} - \lambda_n = o_p(1)$ uniformly in n (i.e., $\lim_A P[\sup_n |\lambda_{nA}/p_{nA} - \lambda_n| \ge \delta] = 0$ whenever $\delta > 0$) and $\lim_A c(\lambda, b, A) = c(\lambda, b)$ uniformly in λ, b on compact subsets of $\{\lambda, b\colon 0 < \lambda < 1, \lambda \int Q(y)\,dy < b < \int_{\bar\lambda}^1 Q(y)\,dy\}$ since $\beta_A(\lambda, b)$, $\gamma_A(\lambda, b)$ converge uniformly to $\beta(\lambda, b)$, $\gamma(\lambda, b)$ by Lemma 7.1. Hence,

$$P\{T_n(\lambda_n) \ge nb_n\} \le E[\exp\{-nc(\lambda_n, b_n) + o(n^{1/2})\}].$$

The final step is to show that the last inequality may be reversed. Let $S'_{nA} = T_n(\lambda_n) - S_{nA}$. Conditional on λ_{nA}, the sums S_{nA} and S'_{nA} are independent and S'_{nA} is the sum of a simple random sample without replacement from H_{nA} where $H_{nA}(y) = G_n(y)/G_n(A)$ for $y \le A$. Let $b'_n > b_n$. Then

$$P\{T_n(\lambda_n) \ge nb_n\} \ge P\{S_{nA} \ge nb'_n, S'_{nA} \ge -n(b'_n - b_n)\}$$
$$= E[P(S_{nA} \ge nb'_n | \lambda_{nA}) P\{S'_{nA} \ge -n(b'_n - b_n) | \lambda_{nA}\}].$$

Invoking Chebyshev's Inequality, we have

$$P\{S'_{nA} \ge -n(b'_n - b_n) | \lambda_{nA}\} \ge 1 - \epsilon'_{nA},$$

where ϵ'_{nA} is independent of S_{nA} conditional on λ_{nA} and $\lim_n E\epsilon'_{nA} = 0$ uniformly in A. Thus,

$$P\{T_n(\lambda_n) \ge nb_n\} \ge \exp\{-nc(\lambda_n, b'_n) + o(n^{1/2})\}.$$

Letting b'_n decrease to b_n completes the argument since c is continuous. □

The next objective should be to extend Theorem 5.3 to random sequences $\{Y_{in}\}$. This goal is not achievable, however. The assumption $Q((i-1)/(n+1)) \le Y_{in} \le Q(i/n)$ for $2 \le i \le n-1$ in Theorem 5.3 controls the variation between G_n and G to such an extent that the error in approximating $n^{-1}\Sigma \log(\gamma + e^{\beta Y_{in}})$ by $\int \log(\gamma + e^{\beta y})\,dG(y)$ is $o(n^{-1/2})$. For random $\{Y_{in}\}$, however, $n^{-1}\Sigma\log(\gamma + e^{\beta Y_{in}}) - \int\log(\gamma + e^{\beta y})\,dG(y) = O_p(n^{-1/2})$. Therefore, the sum in the following Theorem 5.4 cannot be replaced by an integral.

THEOREM 5.4. *Let $\mathbf{Y}_n = (Y_{1n}, \cdots, Y_{nn})$ be the order statistics of a random sample from a continuous distribution with finite absolute third moment and inverse Q. Let $T_n(\lambda_n)$ be the sum of a simple random sample of size $n\lambda_n$ taken without replacement from $\mathbf{Y}_n$. Suppose (i) $\lambda_n - \lambda = O(n^{-1})$ for some $\lambda \varepsilon (0, 1)$. Let $\{b_n\}$ be a sequence of constants satisfying* (iv) *$b_n = b + o(1)$ and* (v) *both b_n and b belong to $\mathscr{I} = (\lambda \int Q(y)\,dy, \int_{\bar\lambda}^1 Q(y)\,dy)$. Then* a.s. $[G]$

$$-n^{-1}\log P\{T_n(\lambda_n) \ge nb_n | \mathbf{Y}_n\} = C_n(\lambda_n, b_n) + o(n^{-1/2})$$

where

$$C_n(\lambda_n, b_n) = \beta_n b_n + \bar\lambda_n \log \gamma_n - n^{-1}\Sigma \log(\gamma_n + e^{\beta_n Y_{in}}) - \lambda_n \log \lambda_n - \bar\lambda_n \log \bar\lambda_n$$

and β_n, γ_n are determined by

$$\lambda_n = \int (1 + \gamma_n e^{-\beta_n y})^{-1}\,dG(y), \qquad b_n = \int y(1 + \gamma_n e^{-\beta_n y})^{-1}\,dG(y).$$

PROOF. Similar to Lemma 5.4, we first show $c_n(\lambda_n, b_n, A) - C_n(\lambda_n, b_n, A) = O(k_n)$ uniformly in $A < A_0$ where $\bar G(A_0) > 0$, $k_n = n^{-1}\log\log n$,

$$C_n(\lambda_n, b_n, A) = b_n \beta_A + \bar\lambda_n \log \gamma_A - \int \log(\gamma_A + e^{\beta_A y})\,dG_{nA}(y) - \lambda_n \log \lambda_n - \bar\lambda_n \log \bar\lambda_n$$

and β_A, γ_A satisfy

$$\lambda_n = \int (1 + \gamma_A e^{-\beta_A y})^{-1}\,dG_A(y), \quad b_n = \int y(1 + \gamma_A e^{-\beta_A y})^{-1}\,dG_A(y).$$

The df's G_{nA} and G_A are defined in Lemma 5.4. A Taylor expansion similar to the first one in Lemma 5.4 gives

$$c_n(\lambda_n, b_n, A) - C_n(\lambda_n, b_n, A) = O((\beta_{nA} - \beta_A)^2) + O((\gamma_{nA} - \gamma_A)^2).$$

Obviously, $c_n(\lambda_n, b_n, A) - C_n(\lambda_n, b_n, A) = O(k_n)$ if (*) $\beta_{nA}(\lambda, b) - \beta_A(\lambda, b)$ and $\gamma_{nA}(\lambda, b) - \gamma_A(\lambda, b)$ are both $O(k_n^{1/2})$ uniformly in A and in (λ, b) on compact subsets E of $(0, 1) \times \mathscr{I}$. Following the proof of Lemma 7.1, (*) is satisfied if

$$V_n(\lambda, b, A) = \int y^m(1 + \gamma_A(\lambda, b)e^{-\beta_A(\lambda,b)y})^{-1}\, d(G_{nA}(y) - G_A(y)) = O(k_n^{1/2})$$

uniformly in A and in (λ, b) on E for both $m = 0$ and $m = 1$. Here $V_n(\lambda, b, A)$ is an average of i.i.d mean zero, finite variance $s^2(\lambda, b, A)$ r.v.'s, and the Law of the Iterated Logarithm implies $\limsup k_n^{1/2}s(\lambda, b, A)V_n(\lambda, b, A) = 1$ a.s. $[G]$. Hence (*) is satisfied since $\beta_A(\lambda, b)$ and $\gamma_A(\lambda, b)$ are continuous and converge to continuous functions $\beta(\lambda, b)$ and $\gamma(\lambda, b)$ as $A \to -\infty$ (again, as in Lemma 7.1).

A further Taylor expansion gives

$$C_n(\lambda, b, A) - C_n(\lambda, b) = b(\beta_A - \beta) + \bar{\lambda}\log(\gamma_A/\gamma) - \beta_A \int y\, dG_{nA}(y) + \beta \int y\, dG_n(y)$$
$$- \int \log(1 + \gamma_A e^{-\beta_A y})\, dG_{nA}(y) + \int \log(1 - \gamma e^{-\beta y})\, dG_n(y).$$

Several applications of the Glivenko-Cantelli Lemma and the convergence of $\int |y|^3\, dG_n(y)$ to $\int |y|^3\, dG(y)$ give $\lim_{A\to-\infty} C_n(\lambda, b, A) - C_n(\lambda, b) = 0$ uniformly in n and in (λ, b) on E.

The remainder of the proof parallels the proof of Theorem 5.3. Here equation (7) is replaced by

$$(8) \qquad (np_{nA})^{-1}\log P(S_{nA} \geq nb_n \mid \mathbf{Y}_n, \lambda_{nA}) = C_n(\lambda_n, b_n, A) + \epsilon_{nA},$$

where $\epsilon_{nA} = o(n^{-1/2})$ uniformly in A a.s. $[G]$. Then an ordered sample $\mathbf{Y}_n$ is fixed, a sample for which (8) is true, $G_n = G + o(1)$ uniformly and $\int |y|^3\, dG_n(y) = \int |y|^3\, dG(y) + o(1)$. The statement of the theorem is then established for the chosen sample. Since the set of such samples has probability one under G, the proof is now complete. □

6. Applications. We first consider the one-sample problem in which $X_1, X_2, \cdots$ are i.i.d. normal $(\theta, 1)$ r.v.'s and the null hypothesis is $H_0: \theta \leq 0$. Under the alternative, θ is positive. With these conditions, Bahadur (1960a) derived the asymptotic lognormality of the $\bar{X}$, t and sign test exact P-values. The parameters of the limiting distribution are given in Table 1.

The asymptotic lognormality under normal alternatives of many other P-values appropriate for H_0 is available from Theorems 4.1, 5.1, and 5.2. For example, Huber's (1965, 1968) robust test is based on $T_n^* = n^{-1}\Sigma X_i^*$ where $x^* =$ median (a, x, b) and the censoring parameters a, b depend on the choice of alternative θ and the extent of contamination under each hypothesis. Both the exact P-value (cf. Theorem 5.1) and an approximate P-value based on the asymptotic normality of T_n^* (cf. Theorem 4.1) are asymptotically lognormal. The approximate P-value is more practical; the parameters of its limiting distribution are specified in Table 1. Theorems 5.2 and 4.1 imply the asymptotic lognormality of the one sample Wilcoxon exact and approximate P-values; Table 1 gives the parameters corresponding to the exact P-value. Note that the asymptotic standard deviation of $n^{-1/2}\log P_n$ for an exact rank P-value satisfying the conditions of Theorem 5.2 equals the standard deviation of the test statistic multiplied by the parameter β defined in Theorem 5.2.

In the two-sample normal shift problem, the observations $X_{11}, \cdots, X_{1n_1}$ are i.i.d. normal $(\mu, 1)$ and the observations $X_{21}, \cdots, X_{2n_2}$ in a second independent sample are i.i.d. normal $(\mu + \theta, 1)$. Here μ is unspecified $n_2/(n_1 + n_2) = \lambda_n$, say, converges to some $\lambda \in (0, 1)$ at rate

TABLE 1
Parameters (c, τ^2) of the asymptotic distribution of the log P-value
One Sample Tests for Location; Normal $(\theta, 1)$ Alternative

Test	Slope c	τ^2
$\bar{X}$	$\theta^2/2$	θ^2
t	$\frac{1}{2}\log(1+\theta^2)$	$\theta^2/(1+\theta^2)^2$
sign	$\Phi(\theta)\log\Phi(\theta) + \bar{\Phi}(\theta)\log\bar{\Phi}(\theta) + \log 2$	$\Phi(\theta)\bar{\Phi}(\theta)(\log(\Phi(\theta)/\bar{\Phi}(\theta)))^2$
Huber's robust approximate P-value censoring at a, b	$\frac{1}{2}(\mu_{a,b,\theta} - \mu_{a,b,0})^2/\sigma^2_{a,b,0}$	$(\mu_{a,b,\theta} - \mu_{a,b,0})^2\sigma^2_{a,b,\theta}/\sigma^4_{a,b,0}$
Wilcoxon	$\beta(2\Phi(\sqrt{2}\theta) - 1) - \log\cosh\beta$	$4\beta^2\left(\int \Phi^2(x+\theta)\phi(x-\theta)\,dx - \Phi^2(\sqrt{2}\theta)\right)$

$\mu_{a,b,\theta} = \int x^*\phi(x-\theta)\,dx$; x^* = median (a, x, b); $\sigma^2_{a,b,\theta} = \int (x^* - \mu_{a,b,\theta})^2\phi(x-\theta)\,dx$;

β solves $\Phi(\sqrt{2}\theta) = 2\int_0^1 x(1+e^{-2\beta x})^{-1}\,dx$.

$O(n^{-1})$. Under the null hypothesis $\theta \leq 0$ and under the alternative θ is positive.

The asymptotic lognormality of the two-sample Wilcoxon, normal scores and van der Waerden (for which $a_n(i) = \Phi^{-1}(i/(n+1))$ P-values under an alternative θ is implied by Theorem 5.3. The asymptotic standard deviation of any two-sample rank standardized log P-value equals the standard deviation of the test statistic multiplied by the parameter β of Theorem 5.3. The asymptotic variance of the normal scores and van der Waerden test statistics is given by Dupac and Hajek (1969).

The asymptotic lognormality of the conditional P-value of the permutation test based on $\bar{X}_2$ (or equivalently on $\bar{X}_2 - \bar{X}_1$) under normal shift alternatives follows from Theorem 5.4. To see this, first note the constants γ and β of Theorem 5.4 that correspond to λ and the a.s. limit of $n^{-1}\Sigma X_{2i}$ are $\bar{\lambda}e^{\theta^2/2}/\lambda$ and θ respectively. Therefore, the slope of the $\bar{X}_2$ permutation P-value P_n is $\bar{\lambda}\theta^2/2 - \int \log(\bar{\lambda}e^{\theta^2/2-\theta z} + \lambda)(\lambda\phi(z-\theta) + \bar{\lambda}\phi(z))\,dz = c_p(\theta)$, say, where ϕ is the normal (0, 1) pdf. Theorem 5.4 then implies that $\sqrt{n}$ $(\log P_n + c_p(\theta))$ and

$$\sqrt{n}\Big(n^{-1}\Sigma\log(\bar{\lambda}_n + \gamma_n^{-1}\bar{\lambda}_n e^{\beta_n X_{1i}}) - \bar{\lambda}\int \log(\bar{\lambda} + \gamma^{-1}\bar{\lambda}e^{\beta x})\phi(x)\,dx$$
$$+ n^{-1}\Sigma\log(\lambda_n + \gamma_n\lambda_n e^{-\beta_n X_{2i}}) - \lambda\int\log(\lambda + \gamma\lambda e^{-\beta x})\phi(x-\theta)\,dx\Big)$$

have the same limiting distribution, if any. The central limit theorem and a.s. convergence of β_n, γ_n to β, γ complete the proof that P_n is asymptotically lognormal. The parameters of the $\bar{X}_2$ permutation P-value, as well as the parameters for the normal scores, van der Waerden, Wilcoxon, pooled-sample-variance t and parametric $\bar{X}_2 - \bar{X}_1$ P-values, are listed in Table 2.

Numerical values of the exact slopes for the one- and two-sample rank tests mentioned above are reported and discussed in Klotz (1965) and Woodworth (1970). The maximal slope for testing the two-sample nonparametric hypothesis that all n observations have identical arbitrary continuous marginal distributions is derived in Hajek (1974) and tabulated for normal shift alternatives in Jones and Sethuraman (1978) for $\lambda = ½$. It is shown in Bahadur and Raghavachari (1970) that if $\lambda = ½$, the permutation test based on $\bar{X}_2$ attains the maximal slope at normal shift alternatives. It is easily seen from our general formula that this optimal property of the permutation test is valid for each λ, $0 < \lambda < 1$.

TABLE 2
Parameters (c, τ^2) of the asymptotic distribution of the log P-value
Two Sample Tests for Shift; Normal (μ, 1), Normal ($\mu + \theta$, 1) alternative

Test	Slope	τ^2
$\bar{X}_2 - \bar{X}_1$	$\lambda\bar{\lambda}\theta^2/2$	$\lambda\bar{\lambda}\theta^2$
t, pooled variance	$\frac{1}{2}\log(1 + \lambda\bar{\lambda}\theta^2)$	$\lambda\bar{\lambda}\theta^2/(1 + \lambda\bar{\lambda}\theta^2)$
Wilcoxon	$2\beta W_\theta + \bar{\lambda}\log\gamma - \log(\gamma + e^\beta) - \lambda\log\lambda - \bar{\lambda}\log\bar{\lambda}$	$\lambda\bar{\lambda}\beta^2\left(\int \Phi^2(x)\phi(x-\theta)\,dx - \Phi^2(\theta/\sqrt{2})\right)$
Normal scores and van der Waerden approximate	$N_\theta^2/(2\lambda\bar{\lambda})$	$N_\theta^2 V_\theta/(\lambda\bar{\lambda})$
Normal scores and van der Waerden, exact	$\delta N_\theta - \int \log(\nu + e^{\delta y})\phi(y)\,dy - \lambda\log\lambda - \bar{\lambda}\log(\bar{\lambda}/\nu)$	$\lambda\bar{\lambda}\delta^2 V_\theta$
$\bar{X}_2 - \bar{X}_1$ permutation	$\bar{\lambda}\theta^2/2 - \int P_\theta(y)(\lambda\phi(y) + \bar{\lambda}\phi(y-\theta))\,dy$	$(\bar{\lambda}^2 + \lambda^2)\left(\int P_\theta^2(y)\phi(y)\,dy - \left(\int P_\theta(y)\phi(y)\,dy\right)^2\right)$

$W_\theta = \lambda \int (\lambda\Phi(y) + \bar{\lambda}\Phi(y+\theta))\phi(y)\,dy$; β, γ solve $\lambda = \int (1 + \gamma e^{-\beta y})^{-1}\,dy$ and $W_\theta = \int y(1 + \gamma e^{-\beta y})^{-1}\,dy$;

$N_\theta = \lambda \int \Phi^{-1}(\lambda\Phi(y) + \bar{\lambda}\Phi(y+\theta))\phi(y)\,dy$;

$$V_\theta = 2\int_{-\infty}^{\infty}\int_{-\infty}^{t} \frac{\bar{\lambda}\phi(s)\phi(t)\Phi(s-\theta)\Phi(\theta-t) + \lambda\phi(s-\theta)\phi(t-\theta)\Phi(s)\Phi(-t)}{\phi(\Phi^{-1}(\lambda\Phi(s-\theta) + \bar{\lambda}\Phi(s)))\phi(\Phi^{-1}(\lambda\Phi(t-\theta) + \bar{\lambda}\Phi(t)))}\,ds\,dt;$$

δ, ν solve $\lambda = \int (1 + \nu e^{-\delta y})^{-1}\phi(y)\,dy$ and $N_\theta = \int y(1 + \nu e^{-\delta y})^{-1}\phi(y)\,dy$; $P_\theta(y) = \log(\bar{\lambda}e^{\theta y - \theta^2/2} + \lambda)$.

For all the tests considered here under normal alternatives, the τ parameter approaches zero as θ approaches zero, suggesting that slopes may be reliable indicators of P-value behavior for small θ. The value of τ for an optimal test may or may not exceed the value of τ for a nonoptimal test. For example, the asymptotic standard deviation of the parametric $\bar{X}$ log P-value increases linearly in θ, and the asymptotic standard deviation of the t and all the nonparametric tests considered here eventually approach zero as θ increases. But the value of τ for the $\bar{X}_2$ permutation P-value, which is optimal, is strictly smaller than the value of τ for the normal scores P-value. As Groenboom and Oosterhoff (1977) report, the exact slopes of all the rank P-values herein approach log 2 as θ increases; the slope of the $\bar{X}_2$ permutation P-value shares this property. The τ for each of the rank and permutation exact P-values approaches zero for absolutely continuous symmetric alternatives and therefore each of these P-values should exhibit similar behavior for large θ and n.

The approximate distributions of the above one- and two-sample P-values for $n = 100$ and $n_1 = n_2 = 50$ are compared in Figures 1–5. In each case the slope $c(\theta)$ (or approximate mean and median of the $-n^{-1}\log P$-value) and the approximate 2.5 and 97.5 percentiles, namely $c(\theta) \pm .1\Phi^{-1}(.975)\tau(\theta)$, are drawn as a function of θ. The censoring parameters 0, 1.0 and -2.5, 3.5 for Huber's robust test are designed for testing $\theta = 0$ against $\theta = 1$ in the model $P_\theta \in \{(1-\epsilon)N(\theta, 1) + \epsilon G,\ G \text{ arbitrary}\}$ with $\epsilon = .126$ and $\epsilon = .00008$ respectively. As would be expected, the effects of censoring the log likelihood ratio are minimal near $\theta = 0$ and substantial for large θ.

For each rank test, the percentiles have been graphed either for the one-sample case or the two-sample case. In the two-sample case, if $\lambda = .5$ and the score generating function is symmetric about s, then the parameter γ of Theorem 5.3 equals $\exp(\beta s)$, and the slope

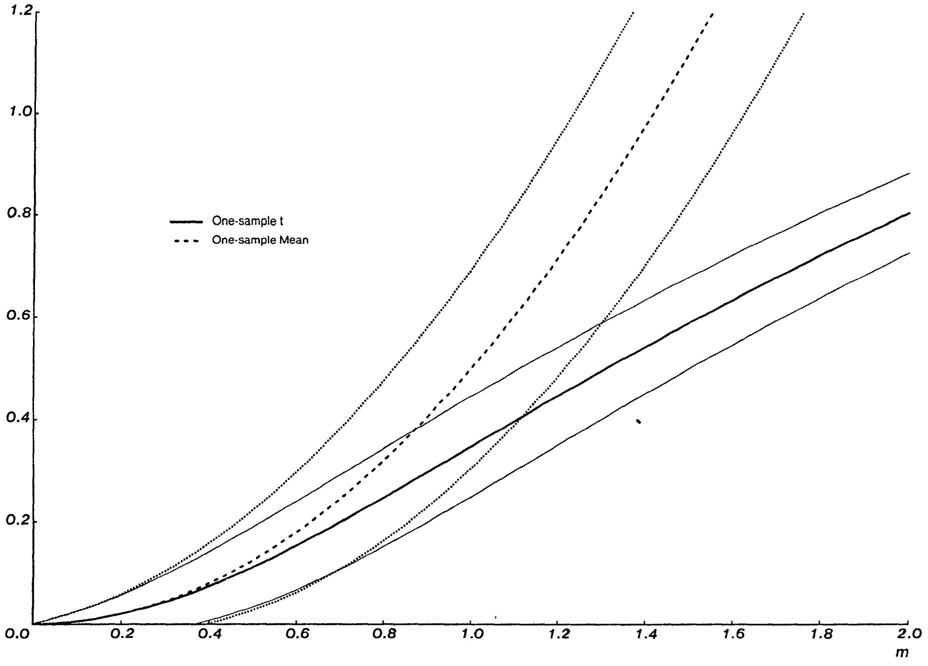

FIG. 1. *Approximate mean,* 2.5 *and* 97.5 *percentiles of the one-sample* (n = 100) −.01 log *P-value for a normal* (m, 1) *alternative.*

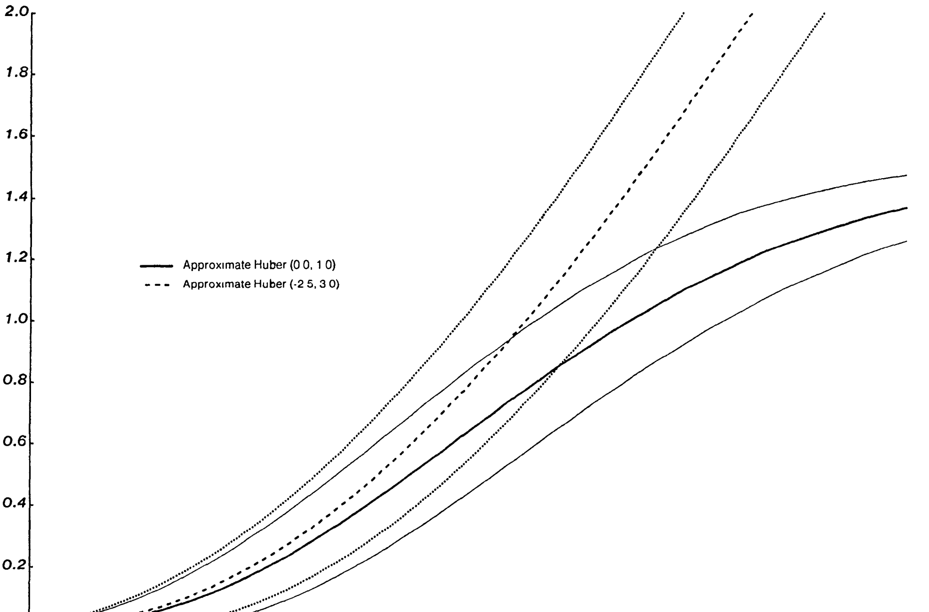

FIG. 2. *Approximate mean,* 2.5 *and* 97.5 *percentiles of the one-sample* (n = 100) −.01 log *P-value for a normal* (m, 1) *alternative.*

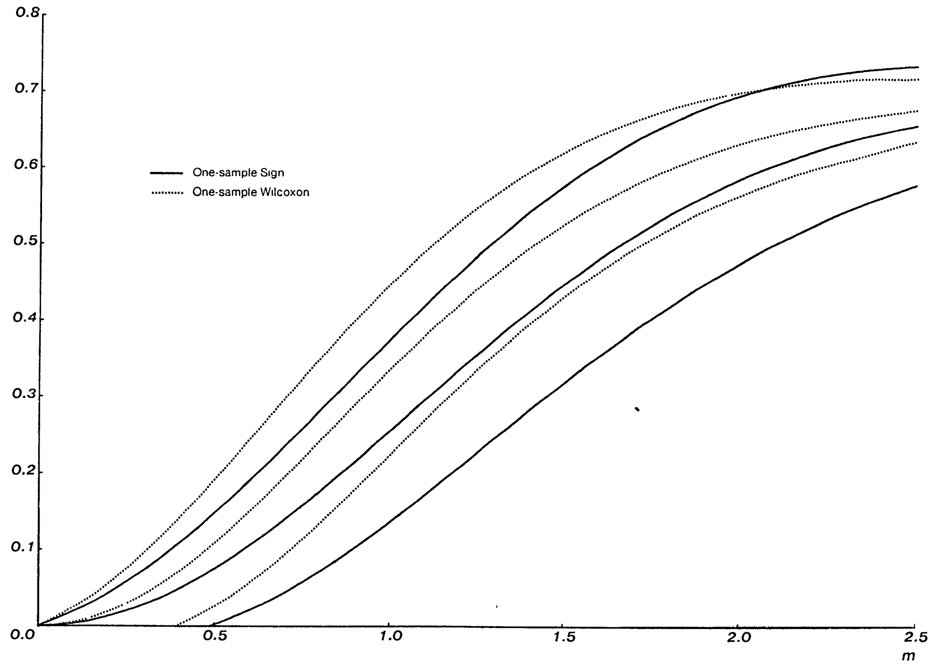

FIG. 3. *Approximate mean, 2.5 and 97.5 percentiles of the one-sample* ($n = 100$) $-.01 \log P$*-value for a normal* $(m, 1)$ *alternative.*

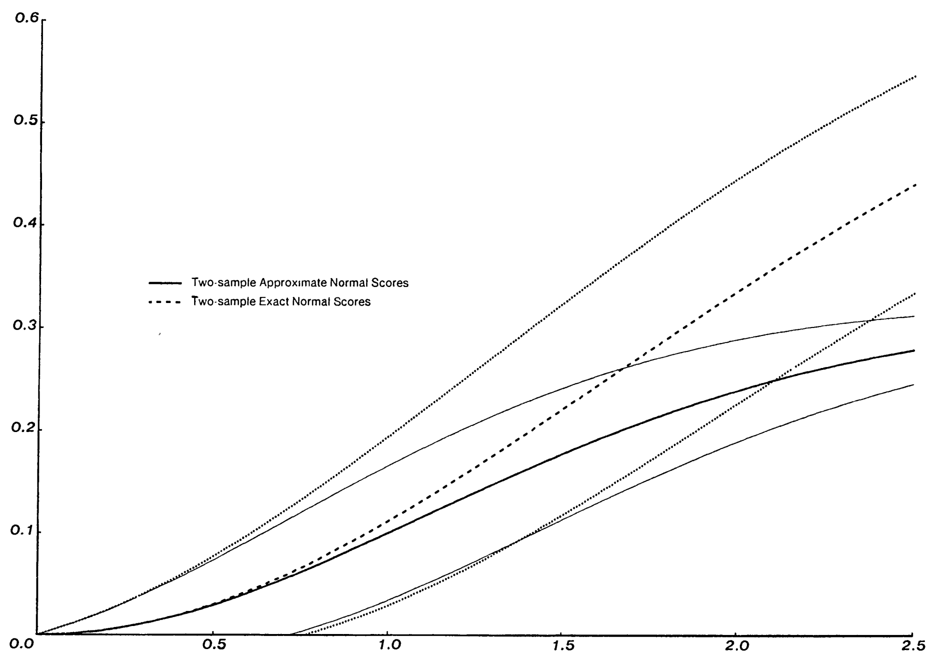

FIG. 4. *Approximate mean, 2.5 and 97.5 percentiles of the two-sample* ($n_1 = n_2 = 50$) $-.01 \log P$*-value for a normal* $(0, 1)$ *normal* $(m, 1)$ *alternative.*

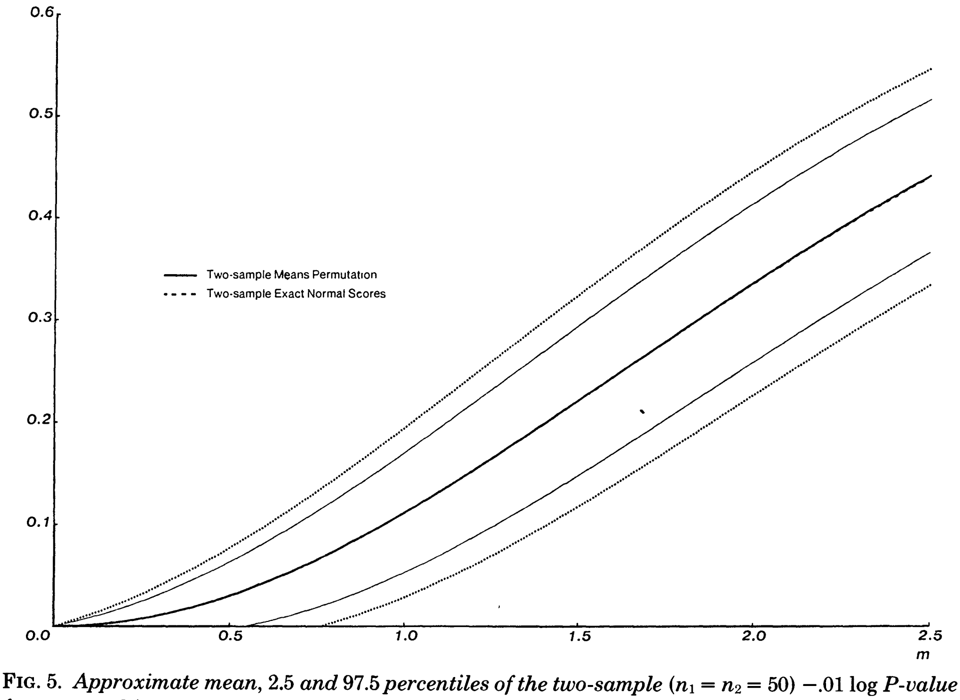

FIG. 5. *Approximate mean,* 2.5 *and* 97.5 *percentiles of the two-sample* ($n_1 = n_2 = 50$) $-.01 \log$ *P-value for a normal* (0, 1), *normal* (*m*, 1) *alternative.*

$c(\theta)$ and variance parameter $\tau^2(\theta)$ of the two-sample rank P-value equal the slope and variance parameter for the corresponding one-sample P-value evaluated at $\theta/2$. It is interesting to note that although the Wilcoxon test has larger slope than the sign (or two-sample median) test at normal alternatives, the Wilcoxon approximate 95th percentile is lower than the sign approximate 95th percentile for large θ. Hence, the slope should be a more reliable measure of test behavior for the Wilcoxon than for the sign test at large θ, but the additional variability in the sign P-value does allow it to assume (exceedingly) small values with higher probability. Also note the difference in both the slopes and the widths of the approximate 95% probability bands for the exact and approximate normal scores log P-values. For $\theta < .625$, both the approximate and exact P-values have similar approximate distributions, but for $\theta > 2.4$, approximately 97.5% of the normal scores exact P-values lie above approximately 97.5% of the normal scores approximate P-values. The properties of the exact test, then, are not necessarily inherited by an approximate procedure. Finally, as discussed above, the approximate 95% probability band for the $\bar{X}_2$ permutation test lies entirely within the corresponding exact normal scores band.

APPENDIX

LEMMA 7.1. *Under conditions* (ii), (iii) *and* (v) *of Lemma* 5.3, *the functions* $\beta_n(\lambda, b)$ *and* $\gamma_n(\lambda, b)$ *defined by*

$$\lambda = \int (1 + \gamma_n e^{-\beta_n y})^{-1} \, dG_n(y), \qquad b = \int y(1 + \gamma_n e^{-\beta_n y})^{-1} \, dG_n(y)$$

converge uniformly on compact sets to functions $\beta(\lambda, b)$ *and* $\gamma(\lambda, b)$ *defined analogously with* G *replacing* G_n.

PROOF. Observe that

$$0 = \int y^m(1 + \gamma_n e^{-\beta_n y})^{-1} \, dG_n(y) - \int y^m(1 + \gamma e^{-\beta y})^{-1} \, dG(y)$$

both for $m = 0$ and $m = 1$. A Taylor expansion gives

$$0 = \int y^m(1 + \gamma e^{-\beta y})^{-1}\, d\{G_n(y) - G(y)\}$$
$$- \int \frac{y^m(\gamma_n - \gamma)e^{-\beta y} - (\beta_n - \beta)\gamma y^{m+1} e^{-\beta y}}{(1 + \gamma_n' e^{\beta_n y})^2}\, dG_n(y)$$
$$+ (\beta_n - \beta)(\gamma_n - \gamma) \int \frac{y^{m+1} e^{-\beta y}}{(1 + \gamma_n' e^{-\beta_n'' y})^2}\, dG_n(y) - \frac{1}{2}(\beta_n - \beta)^2 \int \frac{\gamma_n y^{m+2} e^{-\beta_n' y}}{(1 + \gamma_n' e^{-\beta_n'' y})^2}\, dG_n(y)$$

for some $\gamma_n' \in (\gamma_n, \gamma)$ and $\beta_n', \beta_n'' \in \frac{1}{4}\beta_n, \beta)$. The first integrand on the right is bounded by $|y|$; the last two integrands are $O(|y|^3)$. Hence (iii) and a further expansion of $(1 + \gamma_n' \exp(-\beta_n'' y))^{-2}$ imply

$$o(1) = \int \frac{(\gamma_n - \gamma) y^m e^{-\beta y} - (\beta_n - \beta)\gamma y^{m+1} e^{-\beta y}}{(1 + \gamma e^{-\beta y})^2}\, dG_n(y) + \dot{O}((\beta_n - \beta)^2) + O((\gamma_n - \gamma)^2).$$

Suppose $\beta_n - \beta$ or $\gamma_n - \gamma$ does not converge to zero uniformly. Then $\beta_n - \beta$ and $\gamma_n - \gamma$ must have the same order. In fact, from the last equation,

$$\frac{\gamma_n - \gamma}{\gamma} = (\beta_n - \beta) \int y\, dG_n^*(y), \quad \beta_n - \beta = \frac{(\gamma_n - \gamma)}{\gamma} \frac{\int y\, dG_n^*(y)}{\int y^2 dG_n^*(y)},$$

where $G_n^*(dy) = e^{-\beta y}(1 + \gamma e^{-\beta y})^{-2} G_n(dy) / \int e^{-\beta y}(1 + \gamma e^{-\beta y})^{-2}\, dG_n(y)$.

Hence, $$1 + o(1) = \left(\int y\, dG_n^*(y)\right)^2 \Big/ \int y^2\, dG_n^*(y),$$

which is a contradiction. Therefore, $\beta_n(\lambda, b)$ and $\gamma_n(\lambda, b)$ converge uniformly on compact sets to $\beta(\lambda, b)$ and $\gamma(\lambda, b)$. In particular, $\beta_n(\lambda_n, b_n) \equiv \beta_{nn}$ and $\gamma_n(\lambda_n, b_n) \equiv \gamma_{nn}$ converge to positive constants under the assumption of Lemma 5.3 □

8. Acknowledgements. We are grateful to a referee for the proof that uniform convergence can be replaced by the weaker conditions of Lemma 4.1, for many corrections to the original manuscript and for many suggestions on presentation and interpretation. We also thank Jon Wellner for his critical reading of the thesis (Lambert, 1978) on which this paper is based and Luke-Jon Tierney for several helpful discussions. The computations for Section 6 were carried out with the aid of MACSYMA, a symbolic manipulation program developed at MIT and supported by the National Aeronautics and Space Administration, Office of Naval Research, U.S. Department of Energy, and U.S. Air Force.

REFERENCES

BAHADUR, R. R. (1960a). Simultaneous comparison of the optimum and sign tests of a normal mean. *Contributions to Probability and Statistics* (ed. I. Olkin et al.) 77–88. Stanford University Press.

BAHADUR, R. R. (1960b). Stochastic comparison of tests. *Ann. Math. Statist.* **31** 276–295.

BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*, SIAM, Philadelphia.

BAHADUR, R. R. and RAGHAVACHARI, M. (1970). *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 129–152. University of California Press.

BAHADUR, R. R. and RANGA RAO, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

DAVID, H. A. (1970). *Order Statistics*. Wiley, New York.

DEMPSTER, A. P. and SCHATZOFF, M. (1965). Expected significance level as a sensitivity index for test statistics. *J. Amer. Statist. Assoc.* **60** 420–436.

DUPAC, V. and HAJEK, J. (1969). Asymptotic normality of simple linear rank statistics under alternatives II. *Ann. Math. Statist.* **40** 1992–2017.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications, Vol. 1*, 3rd ed. Wiley, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications, Vol. 2*, 2nd ed. Wiley, New York.

GLESER, L. J. (1964). On a measure of test efficiency proposed by R. R. Bahadur. *Ann. Math. Statist.* **35** 1537–1544.

GROENEBOOM, P. and OOSTERHOFF, J. (1977). Bahadur efficiency and probabilities of large deviations. *Statistica Neerlandica* **31** 1–24.

HAJEK, J. (1974). Asymptotic sufficiency of the vector of ranks in the Bahadur sense. *Ann. Statist.* **2** 75–83.

HOEFFDING, W. (1953). On the distribution of the expected values of the order statistics. *Ann. Math. Statist.* **24** 93–100.

HOEFFDING, W. (1973). On the centering of a simple linear rank statistic. *Ann. Statist.* **1** 54–66.

HUBER, P. J. (1965). A robust version of the probability ratio test. *Ann. Math. Statist.* **36** 1753–1758.

HUBER, P. J. (1968). Robust confidence limits. *Z. Wahrsch. verw. Gebiete* **10** 269–278.

JOINER, B. L. (1969). The median significance level and other small sample measures of test efficacy. *J. Amer. Statist. Assoc.* **64** 971–985.

JONES, D. and SETHURAMAN, J. (1978). Bahadur efficiencies of the Student's t tests. *Ann. Statist.* **6** 559–566.

KEMPTHORNE, O. and FOLKS, L. (1971). *Probability, Statistics and Data Analysis.* Iowa State University Press, Ames.

KLOTZ, J. (1965). Alternative efficiencies for signed rank tests. *Ann. Math. Statist.* **36** 1759–1766.

LAI, T. L. and ROBBINS, H. (1976). Maximally dependent random variables. *Proc. Nat. Acad. Sci. U.S.A.* **73** 286–288.

LAMBERT, D. (1978). *P*-values: asymptotics and robustness. Ph.D. dissertation, Department of Statistics, University of Rochester, Rochester, New York.

LAMBERT, D. (1981). Influence functions for testing. *J. Amer. Statist. Assoc.* **76** 649–657.

PETROV, V. V. (1965). On the probabilities of large deviations for sums of independent random variables. *Theor. Probability Appl.* **10** 287–298.

STONE, M. (1967). Extreme tail probabilities for the null distribution of the two-sample Wilcoxon statistic. *Biometrika* **54** 629–640.

STONE, M. (1969). Approximations to extreme tail probabilities for sampling without replacement. *Proc. Camb. Phil. Soc.* **66** 587–606.

VAN ZWET, W. R. (1964). *Convex Transformations of Random Variables.* Mathematical Center Tracts **7**, Mathematisch Centrum, Amsterdam.

WOODWORTH, G. (1970). Large deviations and Bahadur efficiency of linear rank statistics. *Ann. Math. Statist.* **41** 251–283.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14627

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PA 15213