STOCHASTIC COMPARISON OF TESTS

By R. R. Bahadur

Indian Statistical Institute, Calcutta

1. Introduction. It is shown in [1], in a special case, that the study (as random variables) of the levels attained when two alternative tests of the same hypothesis are applied to given data affords a method of comparing the performances of the tests in large samples. It is the object of the present paper to show that this method, which may be called stochastic comparison, is quite generally applicable. It is shown here, in particular, that in a given statistical context there is usually a wide class of tests such that, if test 1 and test 2 are in the class, the asymptotic efficiency of 1 relative to 2 is well defined and readily calculable. The argument is stated and discussed in general terms in Sections 2, 3 and 4, and illustrative examples are given in Section 5. The examples include comparison of the Wald-Wolfowitz test and the Smirnov test for two samples, and of the Kruskal-Wallis test and the F test for k samples.

2. Standard sequences. Consider an abstract sample space S of points s, and suppose that s is distributed in S according to some one of a given set $\{P_\theta\}$ of probability measures P_θ, where θ is an abstract parameter taking values in a set Ω. Let Ω_0 be a subset of Ω, and let H denote the hypothesis that $\theta \in \Omega_0$.

Let n be an index that takes the values 1, 2, 3, $\cdots$. For each n, let T_n be a real valued statistic defined on S. We shall say that $\{T_n\}$ is a *standard sequence* (for testing H) if the following three conditions are satisfied.

I. There exists a continuous probability distribution function F such that, for each $\theta \in \Omega_0$,

$$\lim_{n\to\infty} P_\theta(T_n < x) = F(x) \quad \text{for every} \quad x. \tag{1}$$

II. There exists a constant a, $0 < a < \infty$, such that

$$\log\,[1 - F(x)] = -\frac{ax^2}{2}\,[1 + o(1)] \quad \text{as} \quad x \to \infty. \tag{2}$$

III. There exists a function b on $\Omega - \Omega_0$, with $0 < b < \infty$, such that, for each $\theta \in \Omega - \Omega_0$,

$$\lim_{n\to\infty} P_\theta\left(\left|\frac{T_n}{n^{1/2}} - b(\theta)\right| > x\right) = 0 \quad \text{for every} \quad x > 0. \tag{3}$$

The following is a typical example of a standard sequence. Let S be the set of all sequences $s = (x_1, x_2, \cdots \text{ad inf})$ with real coordinates x_n, let Ω be a set of distribution functions $\theta(x)$ on the real line such that $\mu(\theta) = \int_{-\infty}^{\infty} x\,d\theta \geq 0$ and $\int_{-\infty}^{\infty} x^2\,d\theta < \infty$, and let P_θ denote the product measure $\theta \times \theta \times \cdots$ on S.

Received May 11, 1959; revised November 5, 1959.

Let H be the hypothesis that $\mu = 0$. For each n, let T_n be the t statistic based on the first n co-ordinates of s. Then I is satisfied with

$$F = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)\, dx;$$

this F satisfies II with $a = 1$ (cf. para. 1 in Section 5); and III is satisfied with $b(\theta) = \mu(\theta)/\sigma(\theta)$, where $\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \, d\theta$. In this example the index n denotes the sample size, and n has essentially the same role in other examples.

Returning to the general case, suppose that $\{T_n\}$ is a standard sequence. Then T_n has the asymptotic distribution F if H obtains, but otherwise $T_n \to \infty$ in probability. Consequently, large values of T_n are significant when T_n is regarded as a test statistic for H. Accordingly, for any given s, we define $1 - F(T_n(s))$ to be the level attained by T_n in the given case ($n = 1, 2, \cdots$).

In general, $1 - F(T_n(s))$ is only an approximate level, i.e. for given n and s, it does not equal the probability of T_n being as large or larger than $T_n(s)$ when H obtains. However, the study of such levels seems legitimate and useful. In numerous cases of interest, in practice only approximate levels are used; perhaps because the exact null distribution of T_n is not tabulated and too difficult to compute, or because n is so large that it is believed unnecessary to refer to the exact distribution, or even because the "exact level attained by T_n" does not exist, i.e. for the given n the distribution of T_n varies with θ as θ varies over Ω_0. Even in the cases where exact levels exist and are used (or at least in principle could be used) for every n, one hopes that conclusions based on comparisons of approximate levels would provide at least an indication of what to expect in comparisons of exact levels. At present exact levels can be compared in only a few cases, e.g. the cases discussed in [1], because sufficiently precise estimates of the relevant tail probabilities are not available. This point is discussed further in remarks 8 and 10 of Section 4.

Now let us regard the level attained by T_n in a given case as a random variable defined on S. It is convenient to describe the behaviour of this random variable as $n \to \infty$ in terms of K_n, where

$$K_n(s) = -2 \log [1 - F(T_n(s))]. \tag{4}$$

Then, for each θ in Ω_0,

$$\lim_{n \to \infty} P_\theta(K_n < v) = \Pr(\chi_2^2 < v) = 1 - e^{-\frac{1}{2}v} \quad \text{for every} \quad v > 0, \tag{5}$$

where χ_2^2 denotes a chi-square variable with 2 degrees of freedom. Again, with

$$c(\theta) = \begin{cases} 0 & \text{for} \quad \theta \in \Omega_0 \\ a[b(\theta)]^2 & \text{for} \quad \theta \in \Omega - \Omega_0 \end{cases} \tag{6}$$

we have that, for any given θ in Ω,

$$K_n/n = c + \epsilon_n \tag{7}$$

where $\epsilon_n(s, \theta) \to 0$ in probability as $n \to \infty$.

To prove the propositions just stated, first consider a θ in Ω_0. Let y and z be given constants, $0 < y < z < 1$. Since F is a continuous distribution function, there exist numbers a and b such that $F(a) = z - y$, $F(b) = z$. Let $A_n = \{s: T_n < a\}$, $B_n = \{s: F(T_n) < z\}$, and $C_n = \{s: T_n < b\}$. Then $A_n \subset B_n \subset C_n$ for every n, and hence $z - y \leqq \liminf P_\theta(B_n) \leqq \limsup P_\theta(B_n) \leqq z$ by (1). Since y and z are arbitrary, we have $\lim P_\theta(B_n) = z$ for all z in $(0, 1)$. (5) now follows from (4), and it follows from (5) that (7) is satisfied with $c = 0$.

Now consider a θ in $\Omega - \Omega_0$. For any x, let $f(x)$ be the $o(1)$ term on the right side of (2), $-1 \leqq f \leqq \infty$. It then follows from (4) that K_n/n is identical with $a(T_n^2/n)[1 + f(T_n)]$. It is plain from this identity and (2), (3), and (6) that (7) is satisfied, and this completes the proof.

In view of (7) we shall call $c(\theta)$ the *asymptotic slope* of the tests based on $\{T_n\}$ (or simply the slope of $\{T_n\}$) when θ obtains.

The sequence $\{T_n\}$ will be said to be *strongly consistent* if condition III is satisfied with (3) replaced by

$$P_\theta\left(\lim_{n \to \infty} T_n/n^{\frac{1}{2}} = b(\theta)\right) = 1, \tag{8}$$

and if (8) also holds with $b = 0$ for each θ in Ω_0. It is readily seen that if $\{T_n\}$ is strongly consistent the ϵ_n in (7) $\to 0$ with probability one.

In concluding this section it may be worthwhile to note that the statistic $K_n^{\frac{1}{2}}$ is equivalent to T_n in the following technical sense: (i) $\{K_n^{\frac{1}{2}}\}$ is a standard sequence, (ii) for each θ in Ω, the slope of $\{K_n^{\frac{1}{2}}\}$ equals that of $\{T_n\}$, and (iii) for any given n and s, the level attained by $K_n^{\frac{1}{2}}$ equals the level attained by T_n. Since the level attained by $K_n^{\frac{1}{2}}$ is found by referring $K_n^{\frac{1}{2}}$ to the upper tail of a fixed distribution independent of F, $\{K_n^{\frac{1}{2}}\}$ is (so to speak) a normalised version of $\{T_n\}$. The normalised version of $\{K_n^{\frac{1}{2}}\}$ is $\{K_n^{\frac{1}{2}}\}$ itself.

3. Comparison of standard sequences. Suppose now that

$$\{T_n^{(1)}\} = \{T_1^{(1)}, T_2^{(1)}, \cdots\} \quad \text{and} \quad \{T_n^{(2)}\} = \{T_1^{(2)}, T_2^{(2)}, \cdots\}$$

are two standard sequences defined on S, and let $F^{(i)}(x)$, a_i, and $b_i(\theta)$ be the functions and constants prescribed by conditions I, II, and III for sequence i $(i = 1, 2)$. Consider an arbitrary but fixed θ in $\Omega - \Omega_0$ and suppose that s is distributed according to P_θ. It is argued in this section that

$$\varphi_{1,2}(\theta) = c_1(\theta)/c_2(\theta) \tag{9}$$

then serves as the asymptotic efficiency of sequence 1 relative to sequence 2, where $c_i = a_i b_i^2$ is the slope of sequence i, $i = 1, 2$.

First consider the comparison of attained levels for a given sample size n. In a given instance, i.e. for a given s in S, it would be fair to say that the test based on $T_n^{(i)}$ is less successful than that based on $T_n^{(j)}$ if the level attained by $T_n^{(i)}$ exceeds the level attained by $T_n^{(j)}$, i.e., if $K_n^{(i)} < K_n^{(j)}$, where the K_n are defined by (4), $(i, j = 1, 2)$. Since $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are standard sequences, it follows from (7) and (9) that

$$K_n^{(1)}/K_n^{(2)} \to \varphi_{1,2} \tag{10}$$

in probability as $n \to \infty$. Consequently, with probability tending to one, $T_n^{(1)}$ is less successful than $T_n^{(2)}$ if $\varphi < 1$ and more successful if $\varphi > 1$. If $\varphi = 1$, the two tests are equally successful up to terms of the order considered here.

To compare the sample sizes required to attain the same level, for each i let $m_1^{(i)}, m_2^{(i)}, \cdots, m_r^{(i)}, \cdots$ be a sequence of positive integers such that

$$\lim_{r \to \infty} m_r^{(i)} = \infty \qquad (i = 1, 2). \tag{11}$$

For simplicity in notation, let $K_n^{(i)}(s)$ be written as $K^{(i)}(n, s)$ and $T_n^{(i)}(s)$ as $T^{(i)}(n, s)$. We may then say that $m_r^{(1)}$ and $m_r^{(2)}$ are asymptotically equivalent sample sizes for sequences 1 and 2 respectively if

$$K^{(1)}(m_r^{(1)}, s)/K^{(2)}(m_r^{(2)}, s) \to 1 \tag{12}$$

in probability as $r \to \infty$. In view of the argument of the preceding paragraph, the defining condition (12) means that, with probability tending to 1 as $r \to \infty$, $T^{(1)}(m_r^{(1)})$ and $T^{(2)}(m_r^{(2)})$ are equally successful test statistics up to terms of the order considered here. Now, since (11) is satisfied, we can apply (7) to $K^{(i)}(n)$ with n restricted to the sequence $\{m_r^{(i)}\}$, $(i = 1, 2)$. This application shows that $m_r^{(1)}$ and $m_r^{(2)}$ are asymptotically equivalent sample sizes if and only if

$$\lim_{r \to \infty} \{m_r^{(2)}/m_r^{(1)}\} = \varphi_{1.2}\,. \tag{13}$$

It should perhaps be noted here that asymptotically equivalent sample sizes always exist, e.g. $m_r^{(1)} = r$ and $m_r^{(2)} =$ the integral part of $r\varphi + 1$.

Now let us consider the case when both the sequences being compared are strongly consistent. It is plain that in this case (10) is valid as a pointwise limit for almost all s in S. Similarly, (11) and (13) suffice for the validity of (12) as a pointwise limit for almost all s, but in the present case a considerably stronger interpretation of φ can also be given, as follows. For any positive real number v, and for any s in S, let $N_i(v, s)$ denote "the sample size required in order that $K_n^{(i)}$ attains the value v." N_i is not well defined, but surely $N_i^- \leq N_i \leq N_i^+$, where $N_i^- =$ the least n such that $K_{n+1}^{(i)} \geq v$ and $N_i^- = \infty$ if no such n exists, and $N_i^+ =$ the least m such that $K_n^{(i)} > v$ for all $n \geq m$, and $N_i^+ = \infty$ if no such m exists. Now define $R^-(v, s) = N_2^-/N_1^+$ and $R^+(v, s) = N_2^+/N_1^-$, with the convention that $\infty/\infty = 1$ (say). Then, except for a set of points s of P_θ measure zero, we have

$$\lim_{v \to \infty} R^- = \lim_{v \to \infty} R^+ = \varphi_{1.2}\,. \tag{14}$$

To establish (14), choose and fix an s for which

$$K_n^{(i)}/n \to c_i \qquad \text{as } n \to \infty \quad (i = 1, 2). \tag{15}$$

Since the set of all such points s has probability one, it will suffice to establish (14) for the chosen s. It is clear from (15) that $0 < K_n^{(i)} < \infty$ for all sufficiently large n, and that $K_n^{(i)} \to \infty$ as $n \to \infty$. Consequently, $1 \leq N_i^- \leq N_i^+ < \infty$ for all sufficiently large v, and $N_i^- \to \infty$ as $v \to \infty$. We observe next that $K^{(i)}(N_i^-) <$

$\nu \leq K^{(i)}(N_i^- + 1)$, $K^{(i)}(N_i^+ - 1) \leq \nu < K^{(i)}(N_i^+)$ provided only that $2 \leq N_i^+ < \infty$. It follows from these relations by application of (15) that

$$c_i N_i^-/\nu \to 1, \quad c_i N_i^+/\nu \to 1 \qquad (i = 1, 2) \tag{16}$$

as $\nu \to \infty$. It follows from (16), as desired, that (14) is satisfied.

4. Discussion. The following remarks are by way of discussion of the preceding two sections.

1. The notion of a standard sequence is by no means essential to stochastic comparison. Suppose for example that $\{T_n^{(i)}\}$ satisfies condition I with $F^{(i)}$ and condition III with b_i, ($i = 1$ and 2), that $F^{(1)} = F^{(2)} = F$, and that the common limiting distribution function F is strictly increasing. For each s and n, let $L_n^{(i)}$ be the level attained by $T_n^{(i)}$. Then $L_n^{(i)} < L_n^{(j)}$ if and only if $T_n^{(i)} < T_n^{(j)}$, ($i, j = 1, 2$). It follows, exactly as in Section 3, that b_1^2/b_2^2 serves as the asymptotic efficiency of sequence 1 relative to sequence 2. In particular, when a given non-null θ obtains, sequence 1 is asymptotically inferior to, or equivalent to, or superior to sequence 2 according as $b_1^2(\theta) <$, or $=$, or $> b_2^2(\theta)$.

This last criterion was suggested by Anderson and Goodman ([2], pp. 108–109) in the context of chi-square and likelihood ratio tests of certain contingency tables. Their suggestion seems to be the first explicit reference to stochastic comparison in the literature.

2. Suppose that in Sections 2 and 3 the index n is restricted to a subset of the positive integers. It is easily seen that the various definitions and conclusions remain valid in this case, except possibly for (14). However, the proof of (14) goes through if the following condition is satisfied: with $j_1 < j_2 < \cdots$ the sequence of values of n, $j_r/j_{r+1} \to 1$ as $r \to \infty$. This condition also ensures the existence of asymptotically equivalent sample sizes in the sense of (13).

3. In the paragraph preceding (14) in Section 3, the random variables R and R^+ are well defined even if neither of the two standard sequences is strongly consistent. It would be interesting to know whether (14) holds in this case with the almost everywhere limits replaced by limits in probability.

4. Suppose that it is desired to make an asymptotic comparison of two sequences of tests, which happen to be based on standard sequences of real valued statistics. The verification that this last is the case, and the determination of the respective asymptotic slopes, requires little knowledge of the exact distributions of the individual members of each sequence of statistics. Consequently, the method of this paper is much more readily applicable than comparisons based explicitly on power functions (cf., e.g., [3], [4], [5], [6], [7]), since the latter comparisons necessarily require detailed knowledge of the exact distributions of individual statistics at least in the non-null case. This remark is supported by the examples given in the following section.

5. Although stochastic comparison as formulated in Sections 2 and 3 makes no reference to power function considerations, there is a formal connection between the asymptotic slope of a standard sequence and the asymptotic power of the corresponding sequence of tests. This connection is discussed in the appendix

to this paper It is pointed out in Appendix 1 that φ can be regarded as the asymptotic relative efficiency when the power is held fixed (or at least bounded away from 0 and 1) and the resulting test sizes are compared. This fact (cf. also the last sentence of remark 6 below) is stated here not so much as a justification of stochastic comparison as a comment on the numerical value of φ.

6. Suppose that Ω is a metric space, and that $\Omega - \Omega_0$ is dense in Ω_0. Let $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ be standard sequences, and let $\varphi_{1,2}(\theta)$ be defined by (9). Suppose that, for each θ_0 in Ω_0, $\varphi_{1,2}$ has a limit as $\theta \to \theta_0$ through values in $\Omega - \Omega_0$, and call this limit $\psi_{1,2}(\theta_0)$.

Limiting efficiency functions such as ψ have a special role in any asymptotic theory of comparison, for the following reason: if the experimentation is undertaken mainly for the purpose of testing H, large sample sizes will in practice occur in the non-null case only if θ is in the neighbourhood of some point in Ω_0. It is therefore of some interest that alternative methods of asymptotic comparison often lead to the same limiting efficiency function. In particular, as is shown in Appendix 2, it is quite generally true that the limiting efficiency ψ derived in the preceding paragraph coincides with Pitman's efficiency function in cases where Pitman's theory is also applicable.

7. Given a parameter space Ω of points θ and a hypothesis H concerning the value of θ, suppose that $\{T_n^{(i)}\}$ is a standard sequence (for testing H) defined on a sample space $S^{(i)}$ of points $s^{(i)}$, $i = 1, 2$. Let $S = S^{(1)} \times S^{(2)}$ be the set of all pairs $s = (s^{(1)}, s^{(2)})$, and for each θ in Ω let P_θ be any probability measure on S which is consistent with the marginal distributions of the $s^{(i)}$. Then both sequences $\{T_n\}$ are standard sequences defined on S, and the arguments of Section 3 apply verbatim.

In other words, stochastic comparison can be applied even in cases where the two sequences are not defined on the same sample space to begin with, e.g., if $S^{(1)}$ and $S^{(2)}$ are the spaces of alternative experiments. It follows, in particular, that if $\{T_n^{(i)}\}$ is a natural or optimum sequence on $S^{(i)}$ then $\varphi_{1,2}$ is, in a sense, the asymptotic efficiency of experiment 1 relative to that of experiment 2. This application is discussed in more detail in [8]. In this application, the limiting efficiency $\psi_{1,2}$ corresponds to the relative "amount of information per observation" in the theory of estimation.

8. The formulation of Section 2 can be generalised so as to include the case when for each n the level attained by the statistic T_n is defined in terms of a distribution function depending on n. One such generalisation is the following. Let $\{T_n\}$ be a sequence of real valued statistics such that conditions I and III are satisfied. For each n, let $F_n(x)$ be a distribution function, to be thought of as the null distribution function of T_n, such that the following condition II* is satisfied:

II*. (i) $\lim_{n\to\infty} F_n(x) = F(x)$, and (ii) there exists a function f on $(0, \infty)$ into $(0, \infty)$ such that, for any given sequence $\{u_n\}$ of positive constants u_n such that $\lim_{n\to\infty} \{u_n^2/n\} = z$, where $0 < z < \infty$, we have

$$2n^{-1} \log [1 - F_n(u_n)] = -f(z)[1 + o(1)] \qquad \text{as } n \to \infty.$$

For each s and n, let $L_n(T_n) = 1 - F_n(T_n)$, and let $K_n = -2 \log L_n(T_n)$. It can then be shown that (5) and (7) continue to hold, with c defined by

$$(6^*) \qquad c(\theta) = \begin{cases} 0 & \text{for } \theta \varepsilon \Omega_0 \\ f[b^2(\theta)] & \text{for } \theta \varepsilon \Omega - \Omega_0 . \end{cases}$$

The proofs, though not entirely trivial, are omitted.

In the special case when F satisfies condition II, and $F_n = F$ for each n, II* is also satisfied with $f(z) = az$, and the formula (6*) reduces to (6).

Let us say that $\{T_n\}$ is a standard sequence *in the strict sense* if there exists a sequence $\{F_n\}$ such that $P(T_n < x \mid \theta) = F_n(x)$ for each n, x, and θ in Ω_0, and such that I, II*, and III are satisfied by $\{T_n\}$ and $\{F_n\}$. In this case $c(\theta)$ defined by (6*) serves as the *exact slope* of the tests based on $\{T_n\}$, and this can be compared with other slopes (exact or otherwise) as in Section 3. It is clear, however, that determination of the exact slope (assuming that it exists) is as difficult in concrete cases as exact analyses based on power function considerations.

9. For $i = 1$ and 2, let $\{T_n^{(i)}\}$ be a sequence satisfying I, II, II* and III. Suppose $\varphi_{1.2}$ is the efficiency function derived in Section 3, and $\psi_{1.2}$ the limiting efficiency function derived from $\varphi_{1.2}$. Let $\varphi_{1.2}^{\bullet}$ be the efficiency function obtained by comparison of the exact slopes, and $\psi_{1.2}^{\bullet}$ the limiting efficiency function derived from $\varphi_{1.2}^{\bullet}$. In the examples available at present where the conditions of this paragraph are satisfied, $\varphi_{1.2}$ differs from $\varphi_{1.2}^{\bullet}$ at every non-null θ, but $\psi_{1.2} = \psi_{1.2}^{\bullet}$ at every null θ_0. (cf. example 1 in Section 5). It is not difficult to formulate general sufficient conditions in order that $\psi_{1.2} = \psi_{1.2}^{\bullet}$, but perhaps it would be more useful to discover and study further examples of sequences which possess exact slopes.

10. In the examples of stochastic comparison given in the following section the level attained by a statistic T_n is defined as in Section 2. As was stated in Section 2, this procedure is generally inexact. It is therefore of some importance to consider whether it is really useful to compute $\varphi(\theta) = c_1(\theta)/c_2(\theta)$, and to study φ as a function of θ, unless it is known that c_1 and c_2 are the exact slopes of the sequences being compared. A categorical answer to this question should await the study of further examples, and of certain theoretical problems. The author's opinion at present is that conclusions based on an inexact φ are necessarily tentative, but that such conclusions may well prove useful, especially in cases (e.g. examples 2 and 3 in Section 5) where no exact methods of comparison are available at present. Some of the issues involved here are mentioned in the following paragraphs.

The formal content of this paper is essentially descriptive. Given a standard sequence (or more generally, a sequence satisfying I, II* and III) a description of the asymptotic behaviour of the sequence is given in Section 2, and it is pointed out in Section 3 that two such descriptions admit a direct and intuitively plausible comparison. Consequently, whether or not c_1 and c_2 are exact slopes, $\varphi = c_1/c_2$ is an exact relative efficiency in the sense that it is based on an accurate description of what happens in the limit when the prescribed methods of computing levels are used.

If the prescribed methods of computing levels are inexact, the plausibility and usefulness of the present descriptions is diminished by the following considerations. The usual inexact methods (e.g. referring a contingency table chi-square to the chi-square distribution) are not intended for computation of very low probabilities. Consequently, if a non-null θ obtains, and n is sufficiently large, the chances are that the prescribed methods will be abandoned, or at least that the levels computed thereby will not be taken seriously. A related consideration is that the inexact slope c of a statistic T_n can scarcely be said to describe the actual performance of T_n, since c incorporates computational errors of unknown magnitude and direction. Consequently, if in a given case $\varphi(\theta) = c_1/c_2 = \frac{1}{2}$ (say), it cannot be concluded that $T_n^{(2)}$ is really twice as efficient as $T_n^{(1)}$, or even that $T_n^{(2)}$ is really more efficient than $T_n^{(1)}$, when θ obtains. There are examples showing that this objection to the comparison of inexact slopes is not purely hypothetical, i.e., that the values of an inexact φ can indeed be misleading (cf., e.g., the last part of example 2 in Section 5).

There is, however, some reason to think that the numerical value of an inexact φ can be very misleading only if θ is far from Ω_0. In particular, the limiting efficiency ψ derived from an inexact φ often coincides with the limiting efficiency functions derived by exact methods of comparison (cf. remarks 6 and 9). It is perhaps fair to say that such value as a given method of asymptotic comparison may have stems mainly from the limiting efficiency function obtainable by that method. If so, the comparison of inexact slopes affords, or at least promises, a very short cut to the main conclusions of exact analyses.

5. Examples. It is convenient to note at the outset of this section that the following distribution functions F satisfy condition II of Section 2: $F^{(1)}(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp[-\frac{1}{2}t^2]\,dt$, with $a = 1$; $F^{(2)}(x) = P(\chi_k < x)$, where χ_k^2 denotes a chi-square variable with k d.f. ($1 \leq k < \infty$), also with $a = 1$; and $F^{(3)}(x) = 1 - 2\sum_{r=1}^{\infty} (-1)^{r-1} \exp[-2r^2x^2]$, with $a = 4$.

That $F^{(1)}$ satisfies (2) with $a = 1$ follows from ([9], p. 166). To treat $F^{(2)}$, let m be a positive integer such that $2m \geq k$. For any $x > 0$ we then have

$$2[1 - F^{(1)}(x)] = P(\chi_1 > x) \leq P(\chi_k > x)$$
$$= 1 - F^{(2)}(x) \leq P(\chi_{2m} > x) = P(Z \leq m - 1)$$

where Z is a Poisson variable with mean $\frac{1}{2}x^2$. It follows from the result for $F^{(1)}$ and from a direct calculation, respectively, that the lower and upper bounds for $1 - F^{(2)}$ are both of the form $\exp[-\frac{1}{2}x^2(1 + o(1))]$; hence $1 - F^{(2)}$ is also of the same form, i.e. (2) is satisfied with $a = 1$. The verification in the case of $F^{(3)}$ is straightforward from the definition of $F^{(3)}$.

In the examples of stochastic comparison that follow, every sequence $\{T_n\}$ introduced in a particular context is a standard sequence in that context, and the asymptotic null distribution is either $F^{(1)}$, or $F^{(2)}$ (for some $k = 1, 2, \cdots$), or $F^{(k)}$. Except possibly for sequence 1 in example 3 and sequence 3 in example 5, each $\{T_n\}$ is strongly consistent. Throughout the remainder of this section G denotes the standardized normal distribution function, i.e. $G(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp[-t^2/2]\, dt$.

EXAMPLE 1. Let $s = (x_1, x_2, \cdots \text{ ad inf})$, where the x_n are independent random variables with each x_n distributed according to $G([x - \mu]/\sigma)$, where μ and $\sigma > 0$ are entirely unknown. Let $\theta = (\mu, \sigma)$, and let H be the hypothesis that $\mu = 0$.

For each $n = 1, 2, \cdots$ let U_n = the number of positive x_j's in the set $\{x_1, x_2, \cdots, x_n\}$, and let $T_n^{(1)} = |2U_n - n|/n^{1/2}$. For each $n = 2, 3, \cdots$ let $T_n^{(2)}$ denote the t statistic based on $\{x_1, x_2, \cdots, x_n\}$. Then for any $\theta = (\mu, \sigma)$ with $\mu \neq 0$, the efficiency of sequence 1 relative to sequence 2 is

$$\varphi_{1,2}(\theta) = [2G(\Delta) - 1]^2/\Delta^2 \qquad \text{where } \Delta = \mu/\sigma \tag{17}$$

This efficiency function is different from the ones derived in [1] by comparison of exact levels, and also from the one derived in [7] by comparison of power functions. However, we have $\psi_{1,2}(\sigma) = \lim_{\mu \to 0} \varphi_{1,2}(\mu, \sigma) = 2/\pi$ for every σ, and the same is true of the other efficiency functions cited. It should be noted that $\varphi_{1,2}$ is a decreasing function of $|\Delta|$.

Now suppose that σ is known and for each n let $T_n^{(3)}$ be Kolmogorov's statistic, i.e. $T_n^{(3)} = (n)^{1/2} \sup_x |K_n(x) - G(x/\sigma)|$ where K_n denotes the distribution function with masses $1/n$ at each of the points $x_1, x_2, \cdots$, and x_n. It follows from Kolmogorov's theorem, and the Glivenko-Cantelli theorem, that the asymptotic slope of $\{T_n^{(3)}\}$ is $4\delta^2$, where $\delta = \sup_x |G(x - \Delta) - G(x)|$. It follows hence that

$$\varphi_{3,2}(\theta) = [2G(\Delta/2) - 1]^2/(\Delta/2)^2 \tag{18}$$

Since we always have $\varphi_{1,3} = \varphi_{1,2}/\varphi_{3,2}$, it follows from (17) and (18) that $\varphi_{1,3}(\theta) < 1$, that $\psi_{1,3}(\sigma) = 1$ for all σ, and that $\varphi_{1,3} \to 0$ as $|\Delta| \to \infty$.

EXAMPLE 2. Let $s = (x_1, x_2, \cdots \text{ ad inf})$ where the x_n are independently distributed according to $G(x/\sigma)$, where σ is entirely unknown. H is the hypothesis $\sigma = 1$.

Let $T_n^{(1)}$ be Kolmogorov's statistic based on $\{x_1, x_2, \cdots, x_n\}$ and let $T_n^{(2)} = |(2\sum_1^n x_i^2)^{1/2} - (2n)^{1/2}|$. It is then found that $\psi_{1,2} = \lim_{\sigma \to 1} \varphi_{1,2}(\sigma) = 1/(\pi e) \doteq 12/100$, but the function $\varphi_{1,2}$ need not be given here.

Next, let $T_n^{(3)}$ be the sequence obtained by normalizing the best estimate of σ^2, i.e. $T_n^{(3)} = |(\sum_1^n x_i^2 - n)/(2n)^{1/2}|$. We then have

$$\varphi_{2,3}(\sigma) = 4/(1 + \sigma)^2. \tag{19}$$

That $\varphi_{2,3}$ is not $\equiv 1$ is due entirely to the fact that the common asymptotic distribution function for sequences 2 and 3 (i.e. $F^{(k)}$ with $k = 1$) does not provide the exact levels attained for a given n.

EXAMPLE 3. Let $F_1(x)$ and $F_2(x)$ be probability distribution functions on the real line, such that $dF_j = f_j(x)\, dx$ where f_j is a continuous function of x, except possibly at a finite number of points, $(j = 1, 2)$. Let $s = (x_{(\infty)}^{(1)}; x_{(\infty)}^{(2)})$ where $x_{(\infty)}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \cdots \text{ ad inf})$ and $x_{(\infty)}^{(2)} = (x_1^{(2)}, x_2^{(2)}, \cdots \text{ ad inf})$ are independent sequences of independent random variables, with $x_n^{(j)}$ distributed according to F_j, $(n = 1, 2, \cdots; j = 1, 2)$. Let $\theta = (F_1, F_2)$ and let H be the hypothesis that $F_1(x) \equiv F_2(x)$.

Let k and l be given positive integers with $k < l$, and write $p = k/l$, $q = 1 - p$. Assume henceforth in this example that n is restricted to the set $l, 2l, 3l, \cdots$. For each n, let $m_1 = m_1(n) = np$ and $m_2 = m_2(n) = nq$.

For each n, let $s_n = (x_1^{(1)}, x_2^{(1)}, \cdots, x_{m_1}^{(1)}; x_1^{(2)}, x_2^{(2)}, \cdots, x_{m_2}^{(2)})$. Let U_n denote the statistic of Stevens and Wald and Wolfowitz [10] when the datum is s_n, i.e. U_n = the total number of runs of superscripts 1 or 2 when the n elements of s_n are arranged in ascending order. It follows from the results given in [10] that, for any θ, U_n/n converges in probability to $2pq[1 - \gamma]$, where

$$\gamma = (pq) \int_{-\infty}^{\infty} \frac{[f_1(x) - f_2(x)]^2}{pf_1(x) + qf_2(x)} \, dx. \tag{20}$$

This form of the consistency theorem of Wald and Wolfowitz is due to Pitman [3]. Now let $T_n^{(1)} = [\mu(n) - U_n]/\sigma(n)$ where μ and σ^2 are the mean and variance of U when H obtains. It then follows by referring to the results in [10] for the null case that $\{T_n^{(1)}\}$ is a standard sequence, and that its slope is

$$c_1(\theta) = \gamma^2, \tag{21}$$

where γ is given by (20).

Next, let $T_n^{(2)}$ be the statistic of Smirnov, i.e.

$$T_n^{(2)} = (npq)^{\frac{1}{2}} \sup_x | K_n^{(1)}(x) - K_n^{(2)}(x) |,$$

where $K_n^{(j)}$ is the distribution function with masses $1/m_j$ at each of the points $x_1^{(j)}, \cdots, x_{m_j}^{(j)}$, $(j = 1, 2)$. It follows from the theorem of Smirnov and the Glivenko-Cantelli theorem that the slope of $\{T_n^{(2)}\}$ is

$$c_2(\theta) = 4pq\delta^2, \qquad \text{where } \delta = \sup_x | F_1(x) - F_2(x) | \tag{22}$$

Consequently, the efficiency of the Wald-Wolfowitz test relative to the Smirnov test is

$$\varphi_{1,2}(\theta) = \gamma^2/4pq\delta^2. \tag{23}$$

It is seen from (20) and (23) that if $(f_1 - f_2)^2/\min\{f_1, f_2\}$ is integrable, $\varphi \to 0$ as $p \to 0$ or 1, i.e. the relative efficiency of sequence 1 is very small if the two sample sizes m_1, m_2 are very different. It is also seen from (20) and (22) that if F_1 and F_2 are both members of a sufficiently smooth parametric family of distribution functions, and if F_1 is close to F_2, then φ will again be nearly zero, for γ will then be of the order of magnitude of δ^2. This is the case, for example, if (a) $F_1 = G(x)$, $F_2 = G(x - \Delta)$, and $|\Delta|$ is small, or if (b) $F_1 = G(x)$, $F_2 = G(x/\sigma)$, and σ is nearly 1.

We observe next that regularity conditions are essential to the arguments of the preceding paragraph. Thus if (c) $f_1 = 1$ on (0, 1) and 0 elsewhere, and $f_2 = f_1(x - \Delta)$, we have $\varphi = 1/(4pq)$ for all Δ. A different irregular case is (d)

$f_1 = 1$ on $(0, 1)$ and 0 elsewhere, and $f_2 = 1 + \sin(2\pi kx)$ on $(0, 1)$ and 0 elsewhere, where k is a positive integer. In this case $\delta = 1/(\pi k)$ and $\varphi = k^2 \cdot \lambda(p)$, where λ is a positive constant independent of k. By taking k sufficiently large we see that sequence 1 can be much more efficient than sequence 2 even though F_2 is close to F_1 in the sense that δ is small.

Suppose for a moment that in case (d) it is required to discriminate between F_1 and F_2 (with a given large k) on the basis of a single observation x. It is then clear that discriminant functions such as x or $|x|$ are practically useless (because δ is small) and also that, in comparison to the optimum criterion of Neyman and Pearson, x and $|x|$ are very inefficient (because $f_2(x)/f_1(x)$ is far from monotonic in x or $|x|$). We shall show that φ can be very large only in the rather extreme cases where both these conditions are satisfied. More precisely, it will be shown that in general

$$\varphi_{1,2}(\theta) \leqq (1/4pq) \min \{\xi^2, (\delta_*/2\delta)^2\} \tag{24}$$

where δ_* is the L_1 distance between f_1 and f_2 $(0 < \delta_* \leqq 2)$, and ξ is essentially the least upper bound to the number of times that the graph of $y = f_2(x)/f_1(x)$ crosses any line $y =$ const.. It follows from (24), in particular, that if $p = q = \frac{1}{2}$ and f_2/f_1 is monotonic then necessarily $\varphi \leqq 1$.

To establish (24), let $\theta = (F_1, F_2)$ and p be given, with $dF_j = f_j\,dx$, and let γ be defined by (20). Define $g(x) = pf_1(x)/pf_1(x) + qf_2(x)$ if $f_1(x) + f_2(x) > 0$ and $g = p$ (say) otherwise. Let $F_0 = pF_1 + qF_2$. Then

$$\int_{-\infty}^{\infty} g\,dF_0 = \int_{-\infty}^{\infty} pf_1\,dx = p.$$

Consequently

$$\begin{aligned}\gamma &= (pq) \int_{-\infty}^{\infty} [f_1 - f_2/pf_1 + qf_2]^2\,dF_0 \\ &= (pq) \int_{-\infty}^{\infty} [(g - p)/pq]^2\,dF_0 \\ &= \int_{-\infty}^{\infty} [(g - p)/pq]\cdot g \cdot dF_0 \\ &= \int_{-\infty}^{\infty} (f_1 - f_2)\cdot g \cdot dx \\ &= \mu_1 - \mu_2,\end{aligned} \tag{25}$$

where $\mu_j = \int_{-\infty}^{\infty} g\,dF_j$ $(j = 1, 2)$. It follows from (25), by a well known representation of the expected value of a random variable, that

$$\gamma = \int_0^1 [G_2(y) - G_1(y)]\,dy, \tag{26}$$

where $G_j(y) = \Pr.(g(x) \leqq y \mid F_j)$, $j = 1, 2$.

Now let P_j denote the probability measure on the Borel sets of the real line corresponding to F_j, $j = 0, 1, 2$. It is evident from (26) that

$$\gamma \leqq \sup_y \{G_2(y) - G_1(y)\} \leqq \sup_A \{P_2(A) - P_1(A)\}.$$

As is easily seen, $\sup_A \{P_2(A) - P_1(A)\} = \frac{1}{2}\int_{-\infty}^{\infty} |f_1 - f_2| \, dx = \frac{1}{2}\delta_*$. Hence $\gamma \leqq \delta_*/2$.

Next, for any interval I on the real line define $\zeta(I)$ as follows: $\zeta = 0$ if $P_0(I) = 0$ or if $P_0(I) = 1$; $\zeta = 1$ if $0 < P_0(I) < 1$ but I is unbounded; and $\zeta = 2$ in the remaining case. It is then easily seen that $P_2(I) - P_1(I) \leqq \zeta(I)\cdot\delta$ for all I, where δ is given by (22). Now choose and fix a y, $0 < y < 1$, and an $\epsilon > 0$. Let $A = \{x: g(x) < y\}$. Let $\eta(y, \epsilon)$ be the infimum of $\zeta(I_1) + \zeta(I_2) + \cdots + \zeta(I_k)$ over all finite collections $\{I_1, I_2, \cdots, I_k\}$ of disjoint intervals I_r such that with $B = I_1 + I_2 + \cdots + I_k$ we have $P_2(A) - P_1(A) \leqq P_2(B) - P_1(B) + \epsilon$. It is then clear that $P_2(A) - P_1(A) \leqq \eta(y, \epsilon)\cdot\delta + \epsilon$. Since ϵ is arbitrary, we have that $P_2(A) - P_1(A) \leqq \eta(y)\cdot\delta$, where $\eta(y) = \lim_{\epsilon\to 0} \eta(y, \epsilon)$, $0 \leqq \eta(y) \leqq \infty$. Assuming that η is a measurable function function of y, it follows from the present definition of A and (26) that $\gamma \leqq \delta\cdot\int_0^1 \eta(y)\, dy$. In any case, if ξ is the essential supremum of $\eta(y)(\xi \leqq \infty)$ we have $\gamma \leqq \xi\cdot\delta$. Thus $\gamma \leqq \min\{\xi\cdot\delta, \delta_*/2\}$, and (24) now follows from (23).

The argument of the preceding paragraph is valid without any restrictions on f_1 and f_2, but the final result is nontrivial (i.e. $\xi < \infty$) only under certain conditions. The reader may verify, in particular, that if there exists a set B with $P_0(B) = 1$ such that g or $-g$ is non-decreasing on B then $\xi \leqq 1$. More generally, if for some $k = 1, 2, \cdots$ it is possible to find disjoint intervals $I_1, \cdots, I_k$ such that $\sum P_0(I_r) = 1$ and such that g is essentially monotonic on each I_r then $\xi \leqq k$.

In concluding this discussion of example 3, let us note that the numerical value of γ depends on F_1 and F_2 only through the error probabilities in the Neyman-Pearson theory of testing F_1 against F_2 given x. This dependence can be made explicit as follows. For any subset A of the real line let $\alpha(A) = P_1(A)$ and $\beta(A) = 1 - P_2(A)$. α and β are then the errors of the first and second kind in using A as the critical region. For any z, $0 < z < \infty$, let $r(z) = \alpha(A_z) + \beta(A_z)$ where $A_z = \{x: f_2(x) \geqq zf_1(x)\}$. It follows from (26) by a straightforward calculation that

$$\gamma = \int_0^\infty \{1 - r(z)\}[pq/(p + qz)^2] \, dz. \tag{27}$$

It follows from the preceding paragraph that the slope of the Wald-Wolfowitz test remains unchanged if each observation in the two samples is subjected to a $1 - 1$ transformation before being supplied to the statistician. This transformation need not be continuous or monotonic—all that is required is that the distributions of the transformed variable also satisfy the conditions stated at the outset of this example.

EXAMPLE 4. Let $s = (x_{(\infty)}^{(1)}; x_{(\infty)}^{(2)}; \cdots; x_{(\infty)}^{(k)})$ be k independent sequences

$z_{(\infty)}^{(j)} = (z_1^{(j)}, z_2^{(j)}, \cdots$ ad inf) of independent random variables $z_m^{(j)}$ ($m = 1, 2, \cdots; j = 1, 2, \cdots, k$). Let $F(z)$ be a continuous distribution function such that

$$\int_{-\infty}^{\infty} z\, dF = 0, \qquad \int_{-\infty}^{\infty} z^2\, dF = 1. \tag{28}$$

Let $\sigma > 0$ and $\mu_1, \mu_2, \cdots, \mu_k$ be constants and suppose that $z_m^{(j)}$ is distributed according to $F([z - \mu_j]/\sigma)$, ($m = 1, 2, \cdots; j = 1, 2, \cdots, k$). Here $\theta = (F; \mu_1, \mu_2, \cdots, \mu_k; \sigma)$. H is the hypothesis that $\mu_1 = \mu_2 = \cdots = \mu_k$

For each $n = k, k + 1, \cdots$ let $m_1(n), \cdots, m_k(n)$ be positive integers such that $n = m_1 + m_2 + \cdots + m_k$. It is assumed that

$$\lim_{n\to\infty} m_j/n = p_j, \qquad \text{where } 0 < p_j < 1 \ (j = 1, 2, \cdots, k) \tag{29}$$

For each n, let $s_n = (z_1^{(1)}, \cdots, z_{m_1}^{(1)}; \cdots; z_1^{(k)}, \cdots z_{m_k}^{(k)})$, and let $T_n^{(1)}$ be the square root of the statistic of Kruskal and Wallis ([11], Section 2). It then follows from the results given in [11] that $\{T_n^{(1)}\}$ is a standard sequence, with slope c_1 defined as follows. Let

$$\Delta_{rs} = (\mu_r - \mu_s)/\sigma \qquad \alpha_{rs} = \int_{-\infty}^{\infty} [F(z + \Delta_{rs}) - F(z)]\, dF, \qquad \beta_j = \sum_{r=1}^{k} p_r \alpha_{jr}, \tag{30}$$

for all r, s and $j = 1, 2, \cdots, k$. Then

$$c_1(\theta) = 12\left(\sum_{j=1}^{k} p_j\beta_j^2\right). \tag{31}$$

Next, let $T_n^{(2)}$ denote the square root of the usual analysis of variance statistic based on s_n. Then $\{T_n^{(2)}\}$ is also a standard sequence, and we have

$$\varphi_{1,2} = 12\left(\sum_{j=1}^{k} p_j\beta_j^2\right) \Big/ \left(\sum_{j=1}^{k} p_j\gamma_j^2\right), \tag{32}$$

where

$$\gamma_j = \sum_{r=1}^{k} p_r\Delta_{jr} \qquad \text{for } j = 1, 2, \cdots, k \tag{33}$$

Suppose now that $dF = f(z)\, dz$, and that f is sufficiently regular so that $\int_{-\infty}^{\infty} [F(z + \Delta) - F(z)]\, dF = \Delta \int_{-\infty}^{\infty} f(z)\, dF + \Delta \cdot o(1)$ as $\Delta \to 0$. It then follows easily from (30), (32), and (33) that, for any $\theta_0 = (F; \mu, \mu, \cdots, \mu, \sigma)$

$$\psi_{1,2}(\theta_0) = \lim_{\theta\to\theta_0} \varphi_{1,2}(\theta) = 12\left[\int_{-\infty}^{\infty} f\, dF\right]^2. \tag{34}$$

It is shown in [7] that ψ is never less than .864. On the other hand, since $|\alpha_{rs}| \leq 1$ we have $|\beta_j| \leq 1$ and hence $\varphi_{1,2} \leq 12/(\sum_1^k p_j\gamma_j^2)$. Consequently $\varphi_{1,2} \to 0$ as

$\max_j \|\gamma_j\| \to \infty$, i.e. as the mean of at least one sequence $x^{(j)}$ becomes very different from the weighted mean of the others.

EXAMPLE 5. Let $s = (v_1, v_2, \cdots \text{ad inf})$ be a sequence of independent and identically distributed random variables $v_n = (x_n, y_n)$, where x and y have a bivariate normal distribution. H is the hypothesis $\rho = 0$, where ρ is the correlation between x and y.

For each n, let r_n denote the sample product moment correlation based on $s_n = (v_1, v_2, \cdots, v_n)$. Let $T_n^{(1)} = \frac{1}{2}|\log [(1 + r_n)/(1 - r_n)]|$, and $T_n^{(2)} = (n-2)^{\frac{1}{2}} \cdot |r_n^2/1 - r_n^2|^{\frac{1}{2}}$. We then have

$$\varphi_{1,2}(\theta) = \frac{(1-\rho^2)}{4\rho^2}\left[\log\left(\frac{1+\rho}{1-\rho}\right)\right]^2 \tag{35}$$

It is easily seen that $\varphi_{1,2}$ is a decreasing function of $|\rho|$, varying from 1 to 0.

Next, let a_n = the median x value in s_n, and b_n = the median y value in s_n. Let f_{1n} = the number of pairs v_i in s_n with $x_i > a_n$, $y_i > b_n$; f_{2n} the number with $x_i < a_n$, $y_i > b_n$, f_{3n} the number with $x_i < a_n$, $y_i < b_n$; and f_{4n} the number with $x_i > a_n$, $y_i < b_n$. Let $T_n^{(3)}$ = the square root of the chi-square statistic based on the 2×2 table of the four frequencies f. Then

$$\psi_{3,2} = (4/\pi^2)\cdot(1-\rho^2)\cdot\left[\frac{\sin^{-1}\rho}{\rho}\right]^2, \qquad \text{and } \psi_{3,2}(\theta_0) = 4/\pi^2 \doteq 41/100. \tag{36}$$

Now let r_n' be Spearman's coefficient of rank correlation based on s_n, and let r_n'' be the difference sign covariance, i.e.

$$r_n'' = \sum_{i,j=1}^{n} \text{sgn}\,(x_i - x_j)\cdot\text{sgn}\,(y_i - y_j)/n(n-1),$$

where sgn $(z) = +1, 0,$ or -1 accordingly as $z >, =,$ or < 0. Let $T_n^{(4)} = |r_n'|/\sigma'(n)$ and $T_n^{(5)} = |r_n''|/\sigma''(n)$ where σ' and σ'' are the standard deviations of r' and r'' respectively when $\rho = 0$. We then have, by using formulae given in [12],

$$\begin{aligned} \varphi_{4,2}(\theta) &= (9/\pi^2)\cdot(1-\rho^2)\cdot\left[\frac{\sin^{-1}(\rho/2)}{(\rho/2)}\right]^2, \\ \varphi_{5,2}(\theta) &= (9/\pi^2)\cdot(1-\rho^2)\cdot\left[\frac{\sin^{-1}\rho}{\rho}\right]^2. \end{aligned} \tag{37}$$

It follows from (37) that $\psi_{4,2}(\theta_0) = \psi_{5,2}(\theta_0) \doteq 91/100$. It also follows that $\varphi_{4,5}$ is a decreasing function of $|\rho|$, varying from 1 to 4/9.

APPENDICES

The argument of this paper depends entirely on the practical principle that if the null hypothesis does not obtain, and if in a given instance test statistic 1 attains the level L_1 while statistic 2 attains L_2, statistic 2 is superior in that instance if and only if $L_2 < L_1$. As might be expected, this principle is closely related to comparisons based on power function considerations. The connection is discussed in the following appendices.

Appendix 1. It will be shown here that the asymptotic slope of a standard sequence is a functional on the family of power functions associated with the sequence of statistics (proposition 2), and that slopes are consistent with power in the following sense: if the power of the test based on $T_n^{(1)}$ never exceeds that of the corresponding test based on $T_n^{(2)}$, then $\varphi_{1,2} \leq 1$ (proposition 3). These conclusions are useful analytical tools in applications such as the one mentioned in remark 7 of Section 4.

Consider a sample space S of points s, a set $\{P_\theta : \theta \varepsilon \Omega\}$ of alternative distributions on S, and a hypothesis $H: \theta \varepsilon \Omega_0$. Let $\{T_n\}$ be a (not necessarily standard) sequence of real valued statistics such that, for each θ in Ω_0,

$$\text{(i)} \qquad \lim_{n\to\infty} P_\theta(T_n < x) = F(x) \quad \text{for every } x,$$

where F is a probability distribution function, and such that, for each θ in $\Omega - \Omega_0$,

$$\text{(ii)} \qquad T_n \to \infty \quad \text{in probability.}$$

For any given constant α, $0 < \alpha < 1$, and each n, the size α test (of H) based on T_n is then defined to be the following procedure: reject H if and only if $1 - F(T_n) \leq \alpha$. In general, this test is not literally of size α, i.e.

$$P_\theta(1 - F(T_n) \leq \alpha) \neq \alpha$$

for each n and each θ in Ω_0, but the present definition seems legitimate and useful in view of the reasons stated in Section 2. For any θ in Ω and any n, let $1 - \beta_n(\alpha \mid \theta)$ denote the power of the size α test based on T_n when θ obtains, i.e. $\beta_n(\alpha \mid \theta) = P_\theta(F(T_n) < 1 - \alpha)$.

Now consider a fixed θ in $\Omega - \Omega_0$ and a fixed α. It is easily seen from (ii) that $\beta_n \to 0$ as $n \to \infty$. It can be shown in certain cases that in fact $n^{-1} \log \beta_n \to -r/2$, where r is a positive constant depending on θ (and possibly also on α). In such cases, if r_1 and r_2 are the constants associated with two sequences $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$, r_1/r_2 is the asymptotic efficiency of sequence 1 relative to sequence 2 in the following sense: r_1/r_2 is the (limiting) ratio of sample sizes required to attain an assigned (arbitrarily small) probability of an error of type two. This method of comparison is due to Hodges and Lehmann [7]. A very similar method was devised earlier by Chernoff [6]. The method is, however, quite difficult to apply because precise estimates of β_n are required.

An alternative analysis which suggests itself is the dual of the preceding one, i.e. to let θ and β be fixed, say $\beta_n(\alpha_n \mid \theta) = \beta_0$, where $0 < \beta_0 < 1$, and to study the rate at which α_n must then tend to zero. This approach was mentioned by Cochran ([13], p. 323). It might appear at first sight that this second method would be just as difficult as the first, but that is not the case. In the present formulation, α and β are not really interchangeable. Indeed, in the definition of the power function, we have already exploited the lack of symmetry between Ω_0 and $\Omega - \Omega_0$ by replacing the set of null distribution functions

$$\{P_\theta(T_n < x) : \theta \in \Omega_0, n = 1, 2, \cdots\}$$

by the single distribution function $F(x)$. It follows from proposition 2 below that if $\{T_n\}$ is a standard sequence then $n^{-1} \log \alpha_n \to -c/2$, where $c(\theta)$ is the slope of $\{T_n\}$ as defined in Section 2. Consequently, $\varphi = c_1/c_2$ serves as the relative efficiency of two standard sequences in this method of comparison.

A third method of comparison of power functions, due to Pitman [2], depends essentially on fixing both α and β, say $\beta_n(\alpha \mid \theta_n) = \beta_0$, and studying the rate at which θ_n must then tend to some null value. It will be shown in Appendix 2, under essentially the same general conditions as are usually required for application of this method (cf., e.g., [4]), that asymptotic efficiency in Pitman's sense coincides with ψ, the limit of φ as θ tends to a null value.

We proceed to establish the connection between the slope of a standard sequence and the family of power functions associated with the sequence. In the following propositions 1–4 we consider an arbitrary but fixed θ in $\Omega - \Omega_0$.

PROPOSITION 1. *Suppose that $\{T_n\}$ is a standard sequence with slope $c(\theta)$. For any sequence $\{\alpha_n\}$ of values α_n in $(0, 1)$, let*

(iii) $$v_n = 2 \log (1/\alpha_n).$$

Then

(iv) $$\liminf_{n\to\infty} \{v_n/n\} < c(\theta) \text{ implies } \liminf_{n\to\infty} \{\beta_n(\alpha_n \mid \theta)\} = 0,$$

and

(v) $$\limsup_{n\to\infty} \{v_n/n\} > c(\theta) \text{ implies } \limsup_{n\to\infty} \{\beta_n(\alpha_n \mid \theta)\} = 1.$$

PROOF. Let K_n be defined by (4). It then follows from the definition of β_n and (iii) that

(vi) $$\beta_n(\alpha_n \mid \theta) = P_\theta(K_n < v_n).$$

As is shown in Section 2, $K_n/n \to c$ in probability. It follows hence from (vi) that (iv) and (v) are valid.

As an immediate consequence of proposition 1 we have

PROPOSITION 2. *If*

(vii) $$0 < \liminf_{n\to\infty} \{\beta_n(\alpha_n \mid \theta)\} \leq \limsup_{n\to\infty} \{\beta_n(\alpha_n \mid \theta)\} < 1,$$

then

(viii) $$\lim_{n\to\infty} \{v_n/n\} = c(\theta).$$

It should be observed that there may exist no sequence $\{\alpha_n\}$ such that (vii) is satisfied. We shall then say that $\{T_n\}$ is degenerate at θ. Although degeneracy can scarcely occur in the applications, it is necessary to take it into account in the general case.

Next, let $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ be two standard sequences, and let $1 - \beta_n^{(i)}(\alpha \mid \theta)$ denote the power function of the size α test based on $T_n^{(i)}$, $i = 1, 2$. For each n, let $\delta_n(1, 2 \mid \theta) = \sup_\alpha [\beta_n^{(2)}(\alpha \mid \theta) - \beta_n^{(1)}(\alpha \mid \theta)]$. It is easily seen (e.g. from (vi)) that $0 \leq \delta \leq 1$. Let us say that $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$ at θ if

$$\text{(ix)} \qquad \lim_{n\to\infty} \delta_n(1, 2 \mid \theta) = 0.$$

PROPOSITION 3. *If $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$ at θ, then $\varphi_{1,2}(\theta) \leq 1$.*

PROOF. Suppose first that $\{T_n^{(2)}\}$ is not degenerate. Let $\{\alpha_n\}$ be a sequence such that (vii) holds with $\beta = \beta^{(2)}$, and let v_n be defined by (iii). Then $c_2(\theta) = \lim (v_n/n)$ by proposition 2. Since $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$, we have

$$\liminf \beta_n^{(1)}(\alpha_n \mid \theta) \geq \liminf \beta_n^{(2)}(\alpha_n \mid \theta) > 0.$$

Hence $c_1(\theta) \leq \liminf (v_n/n)$ by (iv). Thus $c_1(\theta) \leq c_2(\theta)$.

Suppose now that $\{T_n^{(2)}\}$ is degenerate at θ. Let t be a non-negative random variable with a continuous distribution function (e.g. a chi-square with 1 d.f.), independent of s, and let $\{\lambda_n\}$ be a sequence of positive constants with $\lim_{n\to\infty} \lambda_n = 0$, (e.g. $\lambda_n = 1/n$). Define $T_n^{(3)} = [K_n^{(2)} + \lambda_n \cdot t]^{\frac{1}{2}}$. It is readily seen that $\{T_n^{(3)}\}$ is a standard sequence on the space S^* of points (s, t), that $c_3 = c_2$, and that $\beta_n^{(3)}(\alpha \mid \theta) = P_\theta(K_n^{(2)} + \lambda_n \cdot t < v)$, where $v = 2 \log (1/\alpha)$. Since $\lambda_n \cdot t > 0$, and since $\beta_n^{(2)}(\alpha \mid \theta) = P_\theta(K_n^{(2)} < v)$, it follows that $\{T_n^{(3)}\}$ dominates $\{T_n^{(2)}\}$ and hence also $\{T_n^{(1)}\}$. For each n, the distribution of $K_n^{(2)} + \lambda_n \cdot t$ is continuous when θ obtains, so that $\{T_n^{(3)}\}$ is not degenerate. Hence $c_1 \leq c_3 (= c_2)$ by the preceding paragraph. This completes the proof.

The following is a partial converse of proposition 3.

PROPOSITION 4. *If $\varphi_{1,2} < 1$, then $\{T_n^{(2)}\}$ dominates $\{T_n^{(1)}\}$.*

PROOF. For any α we have

$$\text{(x)} \qquad \begin{aligned} \beta_n^{(2)}(\alpha \mid \theta) &= P_\theta(K_n^{(2)} < v) \text{ by (vi)} \\ &= P_\theta(K_n^{(2)} < v, K_n^{(1)} < v) + P_\theta(K_n^{(2)} < v, K_n^{(1)} \geq v) \\ &\leq P_\theta(K_n^{(1)} < v) + P_\theta(K_n^{(2)} < K_n^{(1)}) \\ &= \beta_n^{(1)}(\alpha \mid \theta) + P_\theta(K_n^{(2)} < K_n^{(1)}) \text{ by (vi).} \end{aligned}$$

Since $K_n^{(1)}/K_n^{(2)} \to \varphi$ in probability (cf. (10)), and since $\varphi < 1$, $P_\theta(K_n^{(2)} < K_n^{(1)}) \to 0$ as $n \to \infty$. It follows hence from (x), as desired, that (ix) is satisfied.

It follows from propositions 3 and 4 that $\varphi_{1,2} < 1$ if and only if sequence 2 dominates sequence 1 but sequence 1 does not dominate sequence 2. It also follows that $\varphi = 1$ if and only if (a) each sequence dominates the other, or (b) neither sequence dominates the other. It can be shown by simple examples that contingency (b) does occur, i.e. in the general case, domination induces only a partial ordering of the class of all standard sequences.

Appendix 2. In this appendix we discuss ψ, the limit of φ as θ tends to a null value, in a special context. Suppose that Ω is an interval on the real line, and that $H:\theta = \theta_0$, where θ_0 is a point in Ω. Let $\{U_n\}$ be a sequence of statistics on S such that the following conditions are satisfied for each θ in Ω. (A) $\mu_n(\theta) = E[U_n \mid \theta]$ and $\sigma_n^2(\theta) = \text{Var}(U_n \mid \theta)$ exist, $0 < \sigma_n^2 < \infty$; (B) $V_n(s, \theta) = [U_n - \mu_n(\theta)]/\sigma_n(\theta)$ is asymptotically normally distributed with zero mean and unit variance; (C) with $\Delta_n(\theta) = [\mu_n(\theta) - \mu_n(\theta_0)]/\sigma_n(\theta_0)$, $\lim_{n\to\infty} \Delta_n(\theta)/n^{\frac{1}{2}} = b(\theta)$(say), where $b \neq 0$ for $\theta \neq \theta_0$; (D) $[\sigma_n(\theta)/\sigma_n(\theta_0)]/n^{\frac{1}{2}} \to 0$ as $n \to \infty$; and (E) there exist an even positive integer k and a positive constant λ such that $[b(\theta)]^2 = \lambda\cdot(\theta - \theta_0)^k [1 + o(1)]$ as $\theta \to \theta_0$.

Suppose that $\{U_n^{(1)}\}$ and $\{U_n^{(2)}\}$ are two sequences satisfying conditions (A)–(E), and let $\Delta_n^{(i)}(\theta)$, $b_i(\theta)$, k_i, and λ_i be the corresponding parametric functions and constants, $i = 1, 2$. Define $T_n^{(i)} = | V_n^{(i)}(s, \theta_0) |$. It is then readily seen from (A)-(D) that $\{T_n^{(i)}\}$ is a standard sequence with slope $[b_i(\theta)]^2$. Hence $\varphi_{1,2} = [b_1/b_2]^2$. It now follows from (E) that

$$\psi_{1,2}(\theta_0) = \lim_{\theta\to\theta_0} \varphi_{1,2}(\theta) = \begin{cases} 0 & \text{if } k_1 > k_2 \\ \lambda_1/\lambda_2 & \text{if } k_1 = k_2 \\ \infty & \text{if } k_1 < k_2. \end{cases} \tag{xi}$$

It follows from (C) that we also have

$$\psi_{1,2}(\theta_0) = \lim_{\theta\to\theta_0} \lim_{n\to\infty} [\Delta_n^{(1)}(\theta)/\Delta_n^{(2)}(\theta)]^2. \tag{xii}$$

The right side of (xii) is closely related to Pitman's formula for the relative limiting efficiency, and becomes identical with the latter under certain additional conditions. Suppose, for example, that $k_1 = k_2 = 2$, that $\Delta_n^{(i)}$ is a continuously differentiable function of θ, $d\Delta_n^{(i)}/d\theta = \Lambda_n^{(i)}(\theta)$ say, and that condition (C) is satisfied uniformly in a neighbourhood of θ_0 by both sequences. In this case, by first interchanging the order of the two limits in (xii), and then using the differentiability conditions, we obtain

$$\psi_{1,2}(\theta_0) = \lim_{n\to\infty} [\Lambda_n^{(1)}(\theta_0)/\Lambda_n^{(2)}(\theta_0)]^2. \tag{xiii}$$

Suppose next that α and β_0 are constants, $0 < \alpha < 1 - \beta_0 < 1$, and $\{\theta_n\}$ is a sequence in Ω such that

$$\lim_{n\to\infty} [\Lambda_n^{(i)}(\theta_n)/\Lambda_n^{(i)}(\theta_0)] = 1, \qquad \lim_{n\to\infty} \beta_n^{(i)}(\alpha \mid \theta_n) = \beta_0, \qquad (i = 1, 2). \tag{xiv}$$

It then follows from (xiii) and the first part of (xiv) that

$$\psi_{1,2}(\theta_0) = \lim_{n\to\infty} [\Lambda_n^{(1)}(\theta_n)/\Lambda_n^{(2)}(\theta_n)]^2. \tag{xv}$$

Since the right side of (xv) is Pitman's formula, we see from the second part of (xiv) that ψ is the asymptotic efficiency of sequence 1 relative to 2 in Pitman's sense. As far as calculation of ψ is concerned, however, (xii), (xiii) or (xv) are not required since ψ is already given by (xi).

Appendix 3. Under certain conditions the slope of a standard sequence $\{T_n\}$ can be expressed as the limit, as $n \to \infty$, of n^{-1} times the expected ratio of the power of the test based on T_n to its size, with the size chosen at random according to a certain fixed distribution. This representation of a slope seems to be of some interest, partly because slopes are considered in Sections 2 and 3 of the paper without reference to testing at a preassigned level.

Suppose, for example, that T_n is a sequence such that in the null case T_n is asymptotically normally distributed with zero mean and unit variance (conditions I and II, with $a = 1$), and that in the non-null case

$$\text{(xvi)} \qquad \lim_{n\to\infty} E\left[\frac{T_n}{\sqrt{n}} - b\right]^2 = 0,$$

where b is the parametric function specified in Condition III. For any given $u > 0$, consider the following test: reject H if and only if $|T_n| \geqq u$. Let α be the (approximate) size and γ_n the power ($\gamma = 1 - \beta$) of this test, and let $\rho_n = \gamma_n/\alpha$, i.e.

$$\text{(xvii)} \qquad \alpha(u) = 2\int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt, \qquad \gamma_n(u \mid \theta) = P(|T_n| \geqq u \mid \theta),$$
$$\rho_n(u \mid \theta) = \frac{\gamma_n(u \mid \theta)}{\alpha(u)}.$$

Let U be a random variable taking values in $(0, \infty)$ according to

$$\text{(xviii)} \qquad P(U \leqq u) = \int_0^u (4t)\left\{\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx\right\} dt.$$

We then have

$$\text{(xix)} \qquad c(\theta) = \lim_{n\to\infty} \frac{1}{n} E[\rho_n(U \mid \theta)]$$

for every non-null θ. It follows from (xix), in particular, that in the non-null case $K_n/E(\rho_n) \to 1$ in probability.

To verify (xix), we note that $c = ab^2 = b^2$, and that b^2 is the limit of $n^{-1}E(T_n^2)$, by (xvi). Since $E[T_n^2 \mid \theta] = \int_0^\infty P(T_n^2 \geqq t \mid \theta)\, dt$, it follows that

$$\text{(xx)} \qquad c(\theta) = \lim_{n\to\infty} \frac{1}{n}\int_0^\infty P(|T_n| \geqq \sqrt{t} \mid \theta)\, dt.$$

The desired conclusion follows from (xvii) and (xviii) by a change of variable in the integral on the right side of (xx).

The formula for c obtained above is perhaps the simplest one in a class of such formulae. To obtain another member of the class, we note from (xvi) and $b > 0$ that b is the limit of $n^{-\frac{1}{2}}E(|T_n|)$. It follows hence that

$$\text{(xxi)} \qquad c(\theta) = \frac{2}{\pi} \lim_{n\to\infty} \frac{1}{n} E^2[\rho_n(V \mid \theta)],$$

where V is distributed in $(0, \infty)$ according to

$$P(V \leq v) = \int_0^v \left\{ \int_t^\infty e^{-\frac{1}{2}x^2}\, dx \right\} dt. \tag{xxii}$$

REFERENCES

[1] BAHADUR, R. R., "Simultaneous comparison of the sign and optimum tests of a normal mean," Hotelling Festschrift, Stanford University Press, Stanford, Calif., 1960, to be published.

[2] ANDERSON, T. W. AND GOODMAN, LEO A., "Statistical inference about Markov Chains," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 89–109.

[3] PITMAN, E. J. G., *Lecture notes on nonparametric inference*, Columbia University, New York, 1949, Unpublished.

[4] NOETHER, GOTTFRIED E., "A theorem of Pitman," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 64–68.

[5] HOEFFDING, WASSILY, AND ROSENBLATT, JOAN RAUP, "The efficiency of tests," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 52–63.

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

[7] HODGES, J. L. AND LEHMANN, E. L., "The efficiency of some nonparametric competitors of the t test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.

[8] BAHADUR, R. R., "On the asymptotic efficiency of tests and estimates," *Sankhyā*, 1960, To appear.

[9] FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 2nd ed., John Wiley and Sons, New York, 1957.

[10] WALD, A. AND WOLFOWITZ, J., "On a test whether two samples are from the same population," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 147–162.

[11] KRUSKAL, WILLIAM K., "A nonparametric test for the several sample problem," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 525–540.

[12] KENDALL, M. G., *Rank Correlation Methods*, 2nd ed., Charles Griffin and Co., London, 1955

[13] COCHRAN, WILLIAM G., "The χ^2 test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 315–345.