RATES OF CONVERGENCE OF ESTIMATES AND TEST STATISTICS[1]

By R. R. Bahadur

University of Chicago[2]

0. Introduction. This paper contains brief descriptions of certain large sample theories of estimation and testing null hypotheses. The classical asymptotic variance theory of estimation is considered in Section 1; a parallel and closely related development based on probabilities of large deviations in Section 2; and a relatively unexplored viewpoint involving the rate at which the estimate itself approaches the true value in Section 3. Sections 4–7 describe a version of testing in which any given test statistic is evaluated in terms of the rate at which it makes the null hypothesis more and more incredible as the sample size increases when a non-null distribution obtains.

The statistical framework considered throughout the paper is the following: X is an abstract sample space of points x. The probability distribution of x is determined by an abstract parameter θ which takes values in a set Θ. $s = (x_1, x_2, \cdots, \text{ad inf})$ is a sequence of independent observations on x. For each $n = 1, 2, \cdots, T_n = T_n(s)$ is a real valued statistic which depends on s only through $(x_1, \cdots, x_n)$. Most of the propositions stated formally are versions of propositions in [5], [7], [9], and [10]. Sufficient conditions for the validity of the propositions are discussed in an appendix, and all proofs are deferred to the appendix.

PART I. POINT ESTIMATES

$g(\theta)$ is a real valued parametric function defined on Θ. It is required to estimate the value of g.

1. Asymptotic variance. Suppose that T_n is a consistent and asymptotically normal estimate of g with asymptotic variance v/n, i.e., there exists $v(\theta)$, $0 < v < \infty$, such that for each θ

$$(1) \qquad n^{\frac{1}{2}}(T_n(s) - g(\theta))/(v(\theta))^{\frac{1}{2}} \to \mathfrak{N}(0, 1) \text{ in distribution}$$

as $n \to \infty$ when θ obtains. For any $\epsilon > 0$ and any θ let

$$(2) \qquad \alpha_n(\epsilon, \theta) = P_\theta(|T_n(s) - g(\theta)| \geqq \epsilon).$$

Received 17 October 1966.

[1] This is an expository paper based on a special invited address to the Institute of Mathematical Statistics. (Philadelphia, September 1965.)

[2] This work was supported in part by Research Grant No. NSF GP 3707 from the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation, and in part by the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Given δ, $0 < \delta < 1$, let $M = M(\epsilon, \delta, \theta)$ be the sample size required to make $\alpha_n < \delta$, i.e., M is the smallest integer m such that

$$(3) \qquad \alpha_n(\epsilon, \theta) < \delta \quad \text{for all} \quad n \geqq m.$$

We then have

Proposition 1. *For each δ and θ,*

$$(4) \qquad M(\epsilon, \delta, \theta) \sim v(\theta) \cdot [z(\delta)/\epsilon]^2 \quad as \quad \epsilon \to 0,$$

where $z = z(\delta)$ is given by $P(\mathfrak{N}(0, 1) \geqq z) = \delta/2$.

It follows that if $\{T_n^{(i)}\}$ satisfies (1) with $v = v_i(\theta)$, and if $M^{(i)}$ is the sample size required by $\{T_n^{(i)}\}$ to make α_n less than δ $(i = 1, 2)$, then $\lim_{\epsilon\to 0} M^{(2)}(\epsilon, \delta, \theta)/M^{(1)}(\epsilon, \delta, \theta) = v_2(\theta)/v_1(\theta)$ for each δ and θ. Consequently, $v_2(\theta)/v_1(\theta)$ serves as the asymptotic efficiency of $T_n^{(1)}$ relative to $T_n^{(2)}$ when θ obtains.

There is a fully developed theory concerning the existence and construction of estimates which are optimal according to the asymptotic-variance criterion. Cf., e.g., [20], [21], [31], [32], [34]. Some aspects of this theory are outlined in the remaining paragraphs of this section, mainly to facilitate comparisons with subsequent sections.

Suppose that Θ is an open set in the k dimensional Euclidean space of points $\theta = (\theta_1, \cdots, \theta_k)$. Let $I(\theta) = \{I_{ij}(\theta)\}$ be the information matrix when the sample consists of a single observation on x. Suppose that g is a sufficiently smooth function of θ, let $h_i(\theta) = \partial g(\theta)/\partial \theta_i$, and let

$$(5) \qquad \hat{v}(\theta) = \textstyle\sum_{i,j=1}^{k} h_i(\theta) \cdot I^{ij}(\theta) \cdot h_j(\theta).$$

Let T_n be an estimate such that, for some v, (1) holds for all θ.

Proposition 2. *The set of all θ for which*

$$(6) \qquad v(\theta) \geqq \hat{v}(\theta)$$

does not hold is of Lebesgue measure zero.

This proposition is due to LeCam [31], [33]. A simple method of proof is given in [9]. It is now well known but worth recalling here that (in the absence of regularity conditions on T_n itself) the set of points θ where T_n is superefficient (i.e., where (6) does not hold) can be non-empty. The existence of superefficient estimates was discovered by J. L. Hodges [31].

Now let $\hat{T}_n$ be the maximum likelihood estimate of g, i.e., $\hat{T}_n = g(\hat{\theta}_n)$ where $\hat{\theta}_n$ is the maximum likelihood estimate of θ based on $(x_1, \cdots, x_n)$.

Proposition 3. *For each θ, $\hat{T}_n$ is asymptotically normally distributed with mean $g(\theta)$ and variance $\hat{v}(\theta)/n$.*

It has been emphasized by LeCam that rigorous proofs of this classical proposition consist of two very different parts. It is first shown, under regularity conditions of a global sort (cf. the appendix), that $\hat{\theta}_n$ exists and is a consistent estimate of θ. Consequently, $\hat{\theta}_n$ is a consistent root of the likelihood equations. The desired conclusion is now deducible from familiar regularity conditions of a local sort.

It is argued in some recent studies [36], [42] that one should require of an estimate T_n for which (1) holds that (1) hold uniformly in θ. The need for uniformity may be illustrated as follows. Suppose that for given ϵ and δ we wish to determine a sample size n such that, no matter what θ may be, α_n given by (2) is less than δ. The minimum sample size required is essentially $\sup\{M(\epsilon, \delta, \theta)$: θ in $\Theta\}$, and in the absence of uniformity this may be infinite.

It is shown in [9], [36], [39], [42] that uniformity in θ in (1) implies that (6) holds for all θ. An underlying reason for this is that uniformity in θ in (1) implies that v is continuous in θ. If v and $\hat{v}$ are both continuous in θ, the set where (6) does not hold is an open set of measure zero and therefore empty.

The question of how to compare two different estimates $T_n^{(1)}$ and $T_n^{(2)}$ if they both have asymptotic variance $\hat{v}/n$ has been considered recently by Rao [36]. It is shown in [36] by examples that such comparisons may often be feasible and worthwhile.

2. Asymptotic effective variance. Suppose now that $\{T_n\}$ is a consistent (but not necessarily asymptotically normal) estimate of g, i.e., with α_n defined by (2),

$$\alpha_n(\epsilon, \theta) \to 0 \quad \text{as} \quad n \to \infty \tag{7}$$

for each θ and $\epsilon > 0$. It is suggested in [5], [14] that the rate at which (7) occurs for a given ϵ provides criteria for the performance of $\{T_n\}$. When applied to asymptotically normal estimates such criteria are not always in accordance with the criterion of Section 1.

In typical cases, (7) occurs exponentially fast. Suppose for the present that this is the case, i.e., for each θ and ϵ there exists $\gamma(\epsilon, \theta)$, $0 < \gamma < \infty$, such that

$$n^{-1} \log \alpha_n(\epsilon, \theta) \to -\tfrac{1}{2}\gamma(\epsilon, \theta). \tag{8}$$

It then follows that

$$M(\epsilon, \delta, \theta) \sim 2 \log(1/\delta)/\gamma(\epsilon, \theta) \quad \text{as} \quad \delta \to 0. \tag{9}$$

If (8) holds with $\gamma = \gamma_i$ for a sequence $\{T_n^{(i)}\}$, $(i = 1, 2)$, then $M^{(2)}(\epsilon, \delta, \theta)/M^{(1)}(\epsilon, \delta, \theta) \to \gamma_1(\epsilon, \theta)/\gamma_2(\epsilon, \theta)$ as $\delta \to 0$. Consequently, γ_1/γ_2 serves as the asymptotic efficiency of $T_n^{(1)}$ relative to $T_n^{(2)}$ when θ obtains. This efficiency depends in general on ϵ. Unless there is special interest in some particular ϵ, and this interest persists even with large samples and estimates of great precision on hand, it seems reasonable to take $\lim_{\epsilon \to 0} \gamma_1(\epsilon, \theta)/\gamma_2(\epsilon, \theta)$ to be the relative efficiency. In case $T_n^{(i)}$ is asymptotically normal with asymptotic variance v_i/n $(i = 1, 2)$ this last limit is usually (but not always) equal to v_2/v_1.

The following is an amusing reformulation of the above. Given a sequence $\{T_n\}$, for any n, θ, and $\epsilon > 0$ define $\tau_n = \tau_n(\epsilon, \theta)$ by means of the equation

$$P_\theta(|T_n - g| \geqq \epsilon) = P(|\mathfrak{N}(0, 1)| \geqq \epsilon/\tau_n), \qquad 0 \leqq \tau_n \leqq \infty. \tag{10}$$

Since the right-hand side of (10) may be computed exactly by entering a standard normal table with ϵ/τ_n, τ_n might be called the effective standard deviation of T_n

when T_n is regarded as a point estimate of g, θ obtains, and it is required, for some theoretical or practical reason, to compute the left-hand side of (10). In case T_n is exactly normally distributed with mean g then τ_n^2 equals the actual variance of T_n.

It follows from (2) and (10) (cf. [19], p. 166) that, for fixed ϵ and θ,

$$(11) \qquad \log \alpha_n \sim -\epsilon^2/2\tau_n^2 \quad \text{as} \quad n \to \infty$$

and hence from (8) that

$$(12) \qquad n\tau_n^2(\epsilon, \theta) \to \epsilon^2/\gamma(\epsilon, \theta) = w(\epsilon, \theta) \quad \text{say.}$$

In view of (12), w/n may be called the asymptotic effective variance of T_n. In terms of effective variances, (9) is the following analogue of Proposition 1.

PROPOSITION 4. *For each ϵ and θ,*

$$(13) \qquad M(\epsilon, \delta, \theta) \sim w(\epsilon, \theta) \cdot [z(\delta)/\epsilon]^2 \quad as \quad \delta \to 0.$$

If T_n is asymptotically normal with asymptotic variance v/n then, in typical cases,

$$(14) \qquad w(\epsilon, \theta) \to v(\theta) \quad \text{as} \quad \epsilon \to 0.$$

This implies, as noted already, that in typical cases the relative effective-variance efficiency of two asymptotically normal sequences tends to the relative variance efficiency as $\epsilon \to 0$.

It is shown in [5], under certain general conditions, that the maximum likelihood estimate of g is an optimal estimate according to the criterion of asymptotic effective variance. This conclusion may be stated as follows. Suppose that Θ is an open set in the k dimensional Euclidean space of points $\theta = (\theta_1, \cdots, \theta_k)$, and let $\hat{v}$ be defined by (5). Let $\{T_n\}$ be any consistent estimate of g, and for each n, ϵ, and θ let τ_n be given by (10).

PROPOSITION 5. *For each θ,*

$$(15) \qquad \underline{\lim}_{\epsilon \to 0} \underline{\lim}_{n \to \infty} \{n\tau_n^2(\epsilon, \theta)\} \geqq \hat{v}(\theta).$$

Now let $\hat{T}_n = g(\hat{\theta}_n)$ be the maximum likelihood estimate of g, and let $\hat{\tau}_n$ be determined as above by $\hat{T}_n$.

PROPOSITION 6. *For each θ,*

$$(16) \qquad \lim_{\epsilon \to 0} \lim_{n \to \infty} \{n\hat{\tau}_n^2(\epsilon, \theta)\} = \hat{v}(\theta).$$

In other words, $n\hat{\tau}_n^2(\epsilon, \theta) \to$ a limit, $\hat{w}(\epsilon, \theta)$ say, as $n \to \infty$ (at least if ϵ is sufficiently small), and $\hat{w}(\epsilon, \theta) \to \hat{v}(\theta)$ as $\epsilon \to 0$.

It is easily seen that (11) is valid provided only that $\alpha_n > 0$ for all sufficiently large n and (7) holds. In view of (11), Propositions 5 and 6 may be stated as follows. If T_n is a consistent estimate then (for given θ and all sufficiently small ϵ) $\alpha_n(\epsilon, \theta)$ cannot tend to zero at an exponential rate faster than $\exp(-n\epsilon^2/2\hat{v}(\theta))$ and $\hat{\alpha}_n(\epsilon, \theta)$ does tend to zero nearly at this optimal exponential rate.

According to the present criterion there are no superefficient estimates, i.e., (15) must hold at every θ. It follows, in particular, that if T_n is a superefficient estimate according to the criterion of Section 1 then $\{T_n\}$ does not satisfy (14) at parameter points θ where superefficiency holds. In view of (4) and (13), (14) is equivalent to the statement that the limits of $\{\epsilon^2 M(\epsilon, \delta, \theta)/z^2(\delta)\}$ as $\epsilon \to 0$ and $\delta \to 0$ in this order or the reverse order are the same. The uniformity required here seems unrelated to the uniformity in θ discussed in [9], [36], [39], [42], although either uniformity excludes superefficiency in the sense of asymptotic variance.

It is presumably the case that certain estimates other than $\hat{T}_n$ which are efficient according to the criterion of asymptotic variance are also efficient in the present sense, but general theorems to this effect are not yet available. To consider an example, suppose that x takes only three values a, b, and c with respective probabilities θ, θ, and $1 - 2\theta$, where $0 < \theta < \frac{1}{2}$. Let $g(\theta) = \theta$ and let T_n be the minimum chi-square estimate based on $(x_1, \cdots, x_n)$. Then, for ϵ sufficiently small, T_n and $\hat{T}_n$ both have asymptotic effective variance $\hat{w}(\epsilon, \theta)/n$ when θ obtains. A formula for $\hat{w}$ is given in the appendix.

The main tools available at present for finding asymptotic effective variances are Chernoff's theorem [15], [8] and its generalizations [40]. Sanov's theorem [38], [26], [27] is very useful in the multinomial case.

3. Strong convergence. Suppose now that $\{T_n\}$ is a strongly consistent estimate of g, i.e., for each θ,

$$T_n \to g(\theta) \quad \text{with probability one} \tag{17}$$

when θ obtains. For any θ, $\epsilon > 0$, and $s = (x_1, x_2, \cdots \text{ad inf})$ let $N = N(\epsilon, \theta, s)$ be the sample size required to make $|T_n - g| < \epsilon$, i.e., N is the smallest integer m such that

$$|T_n(s) - g(\theta)| < \epsilon \quad \text{for all} \quad n \geqq m \tag{18}$$

and $N = \infty$ if no such m exists. According to (17), $P_\theta(N(\epsilon, \theta, s) < \infty) = 1$ for all ϵ and θ. What else can be said about N, especially for very small ϵ?

Suppose that T_n is asymptotically normal and satisfies the following structural condition. For each θ, there exists a function $h(x, \theta)$ such that

$$E_\theta(h(x, \theta)) = 0, \qquad E_\theta(h^2(x, \theta)) = v(\theta) \tag{19}$$

where $0 < v < \infty$, and such that

$$T_n(s) = g(\theta) + n^{-1}\textstyle\sum_{i=1}^{n} h(x_i, \theta) + R_n(s, \theta) \tag{20}$$

where R_n is asymptotically negligible in the sense that $n^{\frac{1}{2}}R_n \to 0$ in probability and

$$(n^{\frac{1}{2}}/(\log \log n)^{\frac{1}{2}})R_n \to 0 \quad \text{with probability one} \tag{21}$$

when θ obtains. This condition is satisfied by the maximum likelihood estimate and related estimates, by estimates which are functions of the sample moments of one or more real valued characteristics $y = y(x)$, and by certain others [11], of course under appropriate regularity assumptions.

PROPOSITION 7. *For each θ*

$$\overline{\lim}_{\epsilon\to 0} N(\epsilon, s, \theta)/v(\theta)\cdot[(2/\epsilon^2) \log\log (1/\epsilon)] = 1 \tag{22}$$

with probability one when θ obtains.

It would be interesting to know whether $\overline{\lim}$ can be replaced by lim in (22).[3] If so, the present viewpoint would lead to exactly the same numerical relative efficiencies as the viewpoint of Section 1. It is already clear, however, that asymptotic variances play a dominant role here.

It can be shown by examples that asymptotic normality is not enough and that the asymptotic structure assumed here for T_n (cf. (19), (20), (21)) is essential to (22) These examples suggest that if Θ is Euclidean and if (1) holds for all θ then (22) holds with = replaced by $\geqq$ for almost all θ, but the author does not know whether this is indeed the case in general.

PART II. TEST STATISTICS

Θ_0 is a given subset of Θ. It is required to test the null hypothesis that some θ in Θ_0 obtains.

4. The level attained. Exact slope.

For each n let T_n be a test statistic such that large values of T_n are significant.[4] For any θ and t let

$$F_n(t, \theta) = P_\theta(T_n(s) < t) \tag{23}$$

and

$$G_n(t) = \inf\{F_n(t, \theta): \theta \text{ in } \Theta_0\}. \tag{24}$$

For given s, the level attained by T_n is defined to be

$$1 - G_n(T_n(s)) = L_n(s) \text{ say.} \tag{25}$$

In other words, the data $x_1, \cdots, x_n$ being given, $L_n(x_1, \cdots, x_n)$ is the maximum probability, consistent with the hypothesis, of obtaining a value of T_n as large or larger than $T_n(x_1, \cdots, x_n)$.

In many examples, T_n has an exact null distribution, i.e., $F_n(t, \theta)$ is the same for each θ in Θ_0, and the notion of level attained is sometimes restricted to statistics which satisfy this similarity condition. However, in many problems some important statistics do not satisfy the similarity condition, and the present definition of L_n is intended to permit the inclusion of such statistics in the discussion.

[3] Professor V. Strassen has pointed out to the author that the answer is no; the inferior limit is zero with probability one.

[4] If small values of T_n are significant, consider $-T_n$ or any other strictly decreasing function of T_n instead. If both large and small values of T_n are significant it may be that T_n is being used in a manner equivalent e.g. to using $|T_n - \alpha_n|$ where α_n is a constant. If no such reformulation is applicable the present viewpoint is not available.

L_n is of course a random variable; indeed it is a statistic. In typical cases L_n is asymptotically uniformly distributed over $(0, 1)$ in the null case, and

$$(26) \qquad L_n \to 0 \quad \text{as} \quad n \to \infty$$

with probability one in the non-null case. Some authors have argued [3], [5], [6], [7] and continue to argue [10] that the rate at which (26) occurs when a given non-null θ obtains is an indication of the asymptotic efficiency of T_n against that θ.[5]

Given δ, $0 < \delta < 1$, let $N(\delta, s)$ = the least integer m such that

$$(27) \qquad L_n(s) < \delta \quad \text{for all} \quad n \geqq m$$

and let $N = \infty$ if no such m exists. Then, for given s, N is the sample size required in order that T_n becomes (and remains) significant at the level δ. What can be said about the random integer N? In particular, what happens to N as $\delta \to 0$?

In typical cases, (26) occurs exponentially fast. Suppose then that there exists a parametric function $c(\theta)$ defined over the non-null set $\Theta - \Theta_0$ such that $0 < c < \infty$ and such that

$$(28) \qquad n^{-1} \log L_n \to -\tfrac{1}{2}c(\theta) \quad \text{as} \quad n \to \infty$$

with probability one when θ obtains. In accordance with the terminology of [7] let us call c the exact slope of the sequence $\{T_n\}$.

PROPOSITION 8. *If a non-null θ obtains then*

$$(29) \qquad N(\delta, s) \sim 2 \log(1/\delta)/c(\theta) \quad \textit{as} \quad \delta \to 0$$

with probability one.

It follows that if $\{T_n^{(i)}\}$ is a test sequence with exact slope c_i $(i = 1, 2)$ then $N^{(2)}/N^{(1)} \to c_1/c_2$ so that $c_1(\theta)/c_2(\theta)$ serves as the asymptotic efficiency of $T_n^{(1)}$ relative to $T_n^{(2)}$ when θ obtains.

It is in general a non-trivial problem to determine whether a given sequence $\{T_n\}$ has an exact slope and to evaluate it. It is often convenient to attack the problem in two parts, as follows. Suppose that

$$(30) \qquad T_n/n^{\frac{1}{2}} \to b(\theta)$$

with probability one when a non-null θ obtains, where b is a parametric function defined on $\Theta - \Theta_0$ with $0 < b < \infty$. Suppose also that

$$(31) \qquad n^{-1} \log [1 - G_n(n^{\frac{1}{2}}t)] \to -f(t) \quad \text{as} \quad n \to \infty$$

for each $t > 0$ in an open interval which includes each value of b, where f is a continuous function on the interval, with $0 < f < \infty$. It is readily seen that in this case the exact slope exists for each non-null θ and equals $2f(b(\theta))$.

If the given $\{T_n\}$ does not satisfy the two conditions of the preceding para-

[5] It is suggested in [17] that even for fixed n the expected value of L_n is an index of the performance of T_n.

graph, it may well be that $\{T_n'\}$ does, where $T_n' = \varphi_n(T_n)$, with each $\varphi_n(t)$ a strictly increasing function of t. In this case, the levels attained by T_n and T_n' are the same for every n and s, so the above prescription applied to T_n' yields the common exact slope of both sequences $\{T_n'\}$ and $\{T_n\}$.

The difficulty of the problem lies almost entirely in the need to verify (31) and to find f. Methods given in [15], [18], [27], [38], [40] are sometimes applicable. Examples of these and other methods are given (occasionally in the terminology of Section 5 below) in [1], [2], [6], [10], [24], [30].

It is shown in [10], under certain general conditions, that (contrary to the first paragraph on p. 240 of [5]) the likelihood ratio statistic of Neyman and Pearson is an optimal statistic in the sense of exact slopes. This conclusion may be stated as follows: Suppose that for each θ the distribution of the single observation x admits a density function $f(x, \theta)$ with respect to a fixed measure μ. For any θ and θ_0 in Θ let the Kullback-Liebler information number K be defined by

$$K(\theta, \theta_0) = E_\theta(\log [f(x, \theta)/f(x, \theta_0)]). \tag{32}$$

Then $0 \leqq K \leqq \infty$, and $K = 0$ if and only if $P_\theta \equiv P_{\theta_0}$. For each θ in Θ, let

$$J(\theta) = \inf \{K(\theta, \theta_0) : \theta_0 \text{ in } \Theta_0\}. \tag{33}$$

Then J is well-defined over Θ, with $0 \leqq J \leqq \infty$; $J = 0$ on Θ_0; and in typical cases $0 < J < \infty$ on $\Theta - \Theta_0$.

PROPOSITION 9. *If c is the exact slope of a sequence $\{T_n\}$ then $c(\theta) \leqq 2J(\theta)$ for each non-null θ.*

Now let λ_n be the likelihood ratio statistic based on $(x_1, \cdots, x_n)$, i.e.,

$$\lambda_n(s) = \sup \{\textstyle\prod_{i=1}^n f(x_i, \theta) : \theta \text{ in } \Theta_0\} / \sup \{\textstyle\prod_{i=1}^n f(x_i, \theta) : \theta \text{ in } \Theta\}, \tag{34}$$

and let $\hat{T}_n$ be any strictly decreasing function of λ_n, e.g.,

$$\hat{T}_n = -2 \log \lambda_n. \tag{35}$$

PROPOSITION 10. *The exact slope of $\{\hat{T}_n\}$ exists and equals $2J(\theta)$ for each non-null θ.*

It is noteworthy that the exact slopes of certain statistics generally believed to be asymptotically equivalent to likelihood ratio statistics (e.g. chi-square tests of the multinomial) are actually less than $2J(\theta)$ for most non-null values of θ(cf. [1]). An indication that the class of statistics which are optimal in the sense of exact slopes is not very large first appeared in [26]. Certain optimal statistics other than $\hat{T}_n$ are given in [12].

5. Some power function considerations. As might be expected, there are several connections between the rate at which (26) occurs and the power functions of the family of critical regions based on T_n (cf. [5], [6], [7]; cf. also [17]). The main connection between exact slopes and power is perhaps the following: Let $\{T_n\}$ be a sequence with exact slope $c(\theta)$. Consider a particular non-null θ.

Let p be given, $0 < p < 1$, and let $\{k_n\}$ be a sequence of constants such that, with $W_n = \{T_n \geqq k_n\}$,

$$P_\theta(W_n) \to p \quad \text{as} \quad n \to \infty. \tag{36}$$

For each n, let

$$\alpha_n = \sup\{P_{\theta_0}(W_n): \theta_0 \ \text{ in } \ \Theta_0\}. \tag{37}$$

Then α_n is the size of W_n in testing Θ_0 against $\Theta - \Theta_0$, and $P_\theta(W_n)$ is the power of W_n against the given θ. Note that here k_n depends on the given θ; hence α_n does also. It is clear that $\alpha_n \to 0$; it can be shown that the rate is exactly the rate at which $L_n \to 0$.

To be more precise let $M = M(\delta, \theta)$ be the least integer m such that

$$\alpha_n(\theta) < \delta \quad \text{for all} \quad n \geqq m. \tag{38}$$

We then have

Proposition 11. *$n^{-1} \log \alpha_n(\theta) \to -\frac{1}{2}c(\theta)$ as $n \to \infty$. Hence*

$$M(\delta, \theta) \sim 2 \log(1/\delta)/c(\theta) \quad \textit{as} \quad \delta \to 0. \tag{39}$$

It was suggested by Cochran [16] that alternative test statistics might be compared by fixing the power against a specified alternative and looking at the resulting sizes as n increases. It is plain from Proposition 11 that for statistics which have exact slopes this suggestion is immediately feasible, and that the Cochran viewpoint will always lead to precisely the same conclusions as the considerations of Section 4.

The Cochran viewpoint is located at one extreme of the field of interest. The Neyman-Pearson viewpoint, i.e., fixing the size and looking at the power (or rather, in the present asymptotic context, at 1 − power) is located at the diametrically opposite extreme. [25] is an example of successful use of the Neyman-Pearson viewpoint in asymptotics. In general it is much more difficult to use the Neyman-Pearson viewpoint than the Cochran viewpoint.

The middle ground between the two extremes, in which middle ground α_n and $1 - P_\theta(W_n)$ both tend to zero, has been explored in some recent studies. [37] treats the case when α_n goes to zero as fast as some negative power of n, and [26] the case when α_n goes to zero faster than any negative power of n but not exponentially fast.

6. An approximation. The preceding two sections do not depend in any way on the asymptotic properties, if any, of T_n in the null case. Suppose now that T_n has an asymptotic null distribution, i.e., there exist a probability distribution function F such that, for each θ_0 in Θ_0,

$$F_n(t, \theta_0) \to F(t) \quad \text{as} \quad n \to \infty \tag{40}$$

for each t. In this case, it is of interest to consider the approximate level

$$L_n^{(a)}(s) = 1 - F(T_n(s)). \tag{41}$$

In typical cases, $L_n^{(a)} \to 0$ exponentially fast, i.e., for each non-null θ there exists a $c^{(a)}(\theta)$, $0 < c^{(a)} < \infty$, such that

$$(42) \qquad n^{-1} \log L_n^{(a)} \to -\tfrac{1}{2}c^{(a)}(\theta) \quad \text{as} \quad n \to \infty$$

with probability one when θ obtains. We suppose that this is the case, and call $c^{(a)}$ the approximate slope of $\{T_n\}$.

If $c_i^{(a)}$ is the approximate slope of a sequence $\{T_n^{(i)}\}$ $(i = 1, 2)$ then the arguments of Section 4 applied to approximate levels yield $c_1^{(a)}(\theta)/c_2^{(a)}(\theta)$ as the approximate asymptotic efficiency of $T_n^{(1)}$ relative to $T_n^{(2)}$ when θ obtains.

A prescription for verifying (42) and finding $c^{(a)}$ is the following: Suppose that T_n satisfies (30) for some b, and that for some a, $0 < a < \infty$, the limiting null distribution F satisfies

$$(43) \qquad \log [1 - F(t)] \sim -\tfrac{1}{2}at^2 \quad \text{as} \quad t \to +\infty.$$

Then (42) holds with $c^{(a)}(\theta) = a[b(\theta)]^2$. There are many examples of this calculation [1], [7], [23], [28].

The distinction between $c^{(a)}$ and c is important. Indeed there is some question whether it is useful to compute $c^{(a)}$ unless it is known in advance that $c^{(a)}$ is close to c for the θ under consideration; this advance knowledge is almost never available. (A happy exception is $\hat{T}_n$ given by (34) and (35); for this statistic $c^{(a)} = c = 2J$ for all θ, under general conditions.) Even for large n, (41) is a reasonable definition of the approximate level attained by T_n only under certain conditions. One is that (40) hold uniformly in θ_0, or at least that $G_n \to F$.[6] Another is that T_n be not so highly significant that the level attained, if computed at all, will certainly not be computed by an approximation such as (41) since such approximations typically involve relative errors of unknown magnitude and direction when L_n is very small. If a non-null θ obtains this last proviso for the reasonableness of $L_n^{(a)}$ starts disappearing as n increases and is gone entirely by the time $c^{(a)}$ appears in the limit. Examples mentioned in the following section show that the indicated potential unreliability of $c^{(a)}$ is often realized. Conclusions based entirely on approximate slopes must therefore be regarded as tentative.

It is just possible that in certain circumstances $c^{(a)}$ has more relevance than c does to large samples in which the level attained happens not to be nearly zero. D. L. Wallace and the author have recently asked an electronic computer some questions bearing on this. The first few responses of that oracle have been quite intriguing.

It is argued in [7] that it is worthwhile to compute $c^{(a)}$, on a tentative basis, for the following reason. Suppose for simplicity that Θ is an open set in k dimensional Euclidean space of points $\theta = (\theta_1, \cdots, \theta_k)$. In case a non-null θ far from any point of Θ_0 obtains, and T_n is a respectable statistic, the chances are that the computation of $L_n^{(a)}$ or L_n for that matter will soon be abandoned as n

[6] This uniformity is akin to the uniformity in estimation advocated in [36], [42].

increases. Since slopes appear only in the limit as $n \to \infty$, it may be that such interest or application as slopes may have to large samples stems from the values of the slopes in the neighborhood of some null parameter θ_0. In many cases, $c^{(a)}(\theta)$ is a good approximation to $c(\theta)$ for θ in such a neighborhood. More precisely, for given $\theta_0 = (\theta_{10}, \cdots, \theta_{k0})$ in Θ_0, there exists a positive definite matrix $M(\theta_0) = \{m_{ij}(\theta_0)\}$ such that

$$c^{(a)}(\theta) \sim \sum_{i,j=1}^{k}(\theta_i - \theta_{i0})m_{ij}(\theta_0)(\theta_j - \theta_{j0}) \tag{44}$$

and

$$c(\theta) \sim \sum_{i,j=1}^{k}(\theta_i - \theta_{i0})m_{ij}(\theta_0)(\theta_j - \theta_{j0}) \tag{45}$$

as $\theta \to \theta_0$ in any manner. Consequently, if $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are two typical sequences, $c_1^{(a)}/c_2^{(a)} \sim c_1/c_2$ as $\theta \to \theta_0$ in any manner, so that the approximate local relative efficiency is the same as the exact local relative efficiency. It thus seems that approximate slopes afford, or at least promise, a very short cut to the main conclusions from exact slopes; cf., e.g., [1], [2], [7]; cf., however, part C of Section 7 also.

It is of some interest that in many cases the main conclusions available from slopes coincide with main conclusions (i.e., local relative efficiencies) afforded by power function considerations; cf., e.g., Appendix 2 of [7].

It is worth noting that $c_1^{(a)}/c_2^{(a)}$ will in general tend to a limit as $\theta \to \theta_0$ only if θ approaches θ_0 from a fixed direction, and that the limit will depend on the direction. This dependence on direction is closely related to the dependence which appears in comparing asymptotically normal estimates of the vector parameter θ.

Let c be the exact slope of a sequence $\{T_n\}$. Let us say that $\{T_n\}$ is locally optimal if, for each θ_0 in Θ_0, $c(\theta) \sim 2J(\theta)$ as $\theta \to \theta_0$. It seems likely that the class of locally optimal sequences is much wider than the class of sequences which have the exact slope $2J(\theta)$ for each non-null θ, especially if Θ_0 consists of a single point θ_0. In the latter case, $\{T_n\}$ is locally optimal if and only if M is the information matrix in (44) or (45).

7. Some perils of approximation.

A. Let $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ be sequences with approximate slopes $c_1^{(a)}$ and $c_2^{(a)}$ respectively. It can easily be the case that $T_n^{(1)}$ and $T_n^{(2)}$ are equivalent for each n in the sense that $T_n^{(2)} \equiv \varphi_n(T_n^{(1)})$ where φ_n is a strictly increasing function, but $c_1^{(a)}(\theta) \not\equiv c_2^{(a)}(\theta)$, with wide discrepancy for θ far from any point of Θ_0. Cf. examples in [7], [22].

Needless to say the exact slopes of equivalent statistics are always the same.

B. Let $c^{(a)}$ and c be the approximate and exact slopes of a sequence $\{T_n\}$. In general, $c^{(a)} \not\equiv c$, and the discrepancy can be wide for θ far from any point of Θ_0. Suppose, for example, that x is real valued and normally distributed with mean μ and variance σ^2, and that it is required to test $\mu = 0$. Let $T_n = |\text{Students } t|$. Then $c^{(a)} = \Delta^2$, $c = \log(1 + \Delta^2)$, where $\Delta = |\mu|/\sigma$.

C. In the example just considered, and many others, $c^{(a)}$ is a good local approximation to c, but this is not so in general. Suppose that it is required to test independence in a 2×2 table, with all marginal frequencies except the total sample size free. Let T_n be the chi-square statistic based on a sample of size n. Then, in the non-null case, $L_n^{(a)} \to 0$ exponentially fast, but L_n cannot go to zero at a rate faster than $1/n$. Thus $c^{(a)} > 0$, but in effect $c = 0$, for every non-null θ. This example is due to Hoeffding [26].

In the example under consideration, the limit in (40) is not uniform in θ_0 over the entire null set Θ_0. However, as is well known, in referring T_n to the chi-square 1 d.f. table one is not necessarily using the approximation (40)–(41); one may be approximating the exact level attained by the two-sided conditional test of Fisher. It can be shown that this last is an asymptotically optimal test, i.e., with L_n^* the exact conditional level attained by T_n given all the marginals, $n^{-1} \log L_n^* \to -J$ with probability one, and that $c^{(a)}$ is a good local approximation to $2J$.

This example and many others suggest that there is good local approximation whenever each T_n has an exact null distribution (cf. para 2 of Section 4), but it is not known at present whether this last is the case.

D. In the same 2×2 table example as above let T_n be the chi-square statistic as before and let $\hat{T}_n = -2 \log \lambda_n$, with λ_n the likelihood ratio statistic. Suppose that we were to proceed as follows in any 2×2 table problem: given the data we see which of the two statistics T_n and $\hat{T}_n$ is larger, declare it to be the statistic in use, and refer it to the 1 d.f. chi-square table to obtain the level attained by it in the given case. It is plain that we would then be engaged in actual cheating. However, if accused of cheating we could offer approximation as a defense, i.e., we are invariably using the statistic $V_n = \max\{T_n, \hat{T}_n\}$ and referring V_n to its asymptotic null distribution.

Proof that the approximation defense is untenable (whatever the status of such approximations in current theory and practice) is afforded, perhaps, by the following asymptotic considerations: Let $c_1^{(a)}$ be the approximate slope of T_n and $c_2^{(a)} = 2J$ that of $\hat{T}_n$. Let c, $c^{(a)}$ be the exact and approximate slopes of V_n. Then $c^{(a)} = \max\{c_1^{(a)}, c_2^{(a)}\}$ but $c = 0$ for all non-null θ. The exact slope of $\hat{T}_n$ is, as usual, $2J$. The conditional exact slope (given the marginals) of each of the three statistics T_n, $\hat{T}_n$, and V_n is $2J$.

It thus seems that there is no substitute for exact analysis, whatever the sample size.

APPENDIX. Notes on Propositions 1–11.

Proposition 11 is stated and proved in Appendix 1 of [7] under certain restrictions on T_n and for approximate sizes α_n; a statement and proof which is similarly inadequate is given in [5]. Proposition 11 as stated does involve a restriction on T_n, namely the existence of k_n such that (36) holds, but no other conditions are required. To see this, suppose that $\overline{\lim}_{n\to\infty} \{n^{-1} \log \alpha_n\} > -\frac{1}{2}c$. It then follows from (28) that there exists a sequence $m_1 < m_2 < \cdots$ of positive integers m_r

such that, with probability one, $\alpha_{m_r} > L_{m_r}$ for all sufficiently large r. We observe next from (23) and (24) that, for each n, $\alpha_n = 1 - G_n(k_n)$. Since $L_n = 1 - G_n(T_n)$, and since G_n is non-decreasing, it follows that, with probability one, $T_{m_r} > k_{m_r}$ for all sufficiently large r. Hence $P_\theta(T_{m_r} > k_{m_r}) \to 1$ as $r \to \infty$, contradicting (36). Thus $\overline{\lim}_{n\to\infty} \{n^{-1} \log \alpha_n\} \leqq -\frac{1}{2}c$. A similar argument shows that $\underline{\lim}_{n\to\infty} \{n^{-1} \log \alpha_n\} \geqq -\frac{1}{2}c$. This establishes the first part; the second part is immediate from the first.

Proposilion 10 is established in [10] under regularity conditions of a global sort; local regularity conditions associated with the classical null distribution theory of $\hat{T}_n$ are not required. The following is an outline of the proof:

First consider the case where Θ is a finite set, say $\Theta = \{\theta_1, \cdots, \theta_j, \theta_{j+1}, \cdots, \theta_k\}$ with $1 \leqq j < k$, and suppose $\Theta_0 = \{\theta_1, \cdots, \theta_j\}$. Let

$$\text{(i)} \qquad \hat{U}_n = -n^{-1} \log \lambda_n .$$

This statistic is of course equivalent to $\hat{T}_n$ given by (34) and (35). Let $\hat{L}_n$ be the level attained by $\hat{U}_n$. We must show that, for each non-null θ

$$\text{(ii)} \qquad n^{-1} \log \hat{L}_n \to -J(\theta)$$

when θ obtains.

Let $\hat{U}_n(q, p)$ be the statistic $\hat{U}_n$ when the entire parameter space is $\{\theta_p, \theta_q\}$ and the null hypothesis is $\{\theta_p\}$, i.e.,

$$\text{(iii)} \qquad \hat{U}_n(q, p) = \max \{0, n^{-1}\textstyle\sum_{r=1}^n \log [f(x_r, \theta_q)/f(x_r, \theta_p)]\}.$$

Suppose that a non-null θ_q obtains (i.e., $q > j$). It is plain from (iii) and $K \geqq 0$ that then $\hat{U}_n(q, p) \to K(\theta_q, \theta_p)$ with probability one. It follows from the consistency of the maximum likelihood estimate that, with probability one, $\max_{1\leqq i\leqq k}\{\hat{U}_n(i, p)\} = \hat{U}_n(q, p)$ for all sufficiently large n. Hence $\max_{1\leqq i\leqq k}\{\hat{U}_n(i, p)\} = \hat{V}_n(p)$ (say) $\to K(\theta_q, \theta_p)$ with probability one for each p. Since

$$\text{(iv)} \qquad \hat{U}_n = \min_{1\leqq p\leqq j}\{\hat{V}_n(p)\}$$

it follows that

$$\text{(v)} \qquad \hat{U}_n \to J(\theta_q)$$

with probability one when θ_q obtains.

Now let θ_p be a null parameter point (i.e., $1 \leqq p \leqq j$), and let $t > 0$. Then, from (iv) and the definition of $\hat{V}_n$,

$$\text{(vi)} \qquad \begin{aligned} P_{\theta_p}(\hat{U}_n \geqq t) &\leqq P_{\theta_p}(\hat{V}_n(p) \geqq t) \\ &\leqq \textstyle\sum_{i=1}^k P_{\theta_p}(\hat{U}_n(i, p) \geqq t). \end{aligned}$$

Writing $f_i^{(n)}(\cdot) \equiv \prod_{r=1}^n f(x_r, \theta_i)$,

$$\text{(vii)} \qquad \begin{aligned} P_{\theta_p}(\hat{U}_n(i, p) \geqq t) &= P_{\theta_p}(f_p^{(n)} \leqq e^{-nt}\cdot f_i^{(n)}) \\ &\leqq e^{-nt}P_{\theta_i}(\ ,\ ,\ ) \leqq e^{-nt}. \end{aligned}$$

It follows from (vi) and (vii) that $1 - \hat{F}_n(t, \theta_p) \leqq ke^{-nt}$. Since θ_p is arbitrary, we have $1 - \hat{G}_n(t) \leqq ke^{-nt}$. Hence $\hat{L}_n \equiv 1 - \hat{G}_n(\hat{U}_n) \leqq k \exp(-n\hat{U}_n)$. Hence

$$\text{(viii)} \qquad \overline{\lim}_{n\to\infty} \{n^{-1} \log \hat{L}_n\} \leqq -J(\theta)$$

with probability one when θ obtains, by (v).

The general case can be reduced to the case just considered (i.e., finite Θ) by the compactification device of Wald; for details see [10]. Assume henceforth that, in the general case under consideration, (viii) holds. The desired conclusion is immediate from (viii) and Proposition 9 *if* $\{\hat{U}_n\}$ has an exact slope.

It seems difficult to show directly (even in special cases) that $\{\hat{U}_n\}$ has an exact slope; it seems much easier to show that, for *any* $\{T_n\}$,

$$\text{(ix)} \qquad \underline{\lim}_{n\to\infty} \{n^{-1} \log L_n\} \geqq -J(\theta)$$

with probability one when θ obtains; this inequality applied to $\hat{L}_n$, together with (viii), establishes Proposition 10.

A proof of (ix) is given in [10]; a better proof, under a weaker assumption, is given in [13]. The assumption required in [13] is that if θ is a non-null point with $J(\theta) < \infty$ then there exists a sequence $\{\theta_i\}$ of null points such that $K(\theta, \theta_i) \to J(\theta)$ and such that not only $K(\theta, \theta_i) < \infty$ but $E_\theta(\log [f(x, \theta)/f(x, \theta_i)])^2 < \infty$ for each i. It is not known at present whether this last assumption is essential to (ix).

Proposition 9. A version of this proposition appears in [5] but this version involves unnecessary conditions on the framework and on T_n. In fact, no conditions whatsoever are required. It is known [10], [13] that for any sequence $\{T_n\}$ and any θ in Θ,

$$\text{(x)} \qquad \overline{\lim}_{n\to\infty} \{n^{-1} \log L_n\} \geqq -J(\theta)$$

with probability one when θ obtains. Proposition 9 is immediate from (x) and the present hypothesis that (28) holds.

The following is an alternative direct proof of Proposition 9; this proof parallels the proof of Proposition 10 and may therefore be of interest. Let θ and θ_0 be points in Θ. Consider the (possibly randomized) likelihood ratio test of θ_0 against θ based on $(x_1, \cdots, x_n)$ such that the power of this test is p, where $0 < p < 1$ is given. Let $\hat{\alpha}_n$ be the size of this test. Let K be defined by (32), and assume $0 \leqq K < \infty$.

LEMMA. $n^{-1} \log \hat{\alpha}_n \to -K$ *as* $n \to \infty$.

This lemma is due to C. Stein (unpublished). The lemma is stated and proved independently in [5] but in a complicated way involving unnecessary regularity assumptions. It may be noted, in view of Proposition 10, that, except for certain technical details, the lemma provides an illustration of Proposition 11 with $\{\theta, \theta_0\}$ as the parameter space, $T_n = \hat{T}_n$, and $c(\theta) = 2J(\theta) = 2K(\theta, \theta_0)$.

To establish the lemma, note first that the present hypothesis that $K < \infty$ implies that $P_\theta \ll P_{\theta_0}$ on the sample space of a single observation x, say $dP_\theta/dP_{\theta_0} = g(x)$, and that $K = E_\theta(\log g)$. Since the lemma holds trivially if $K = 0$, suppose that $K > 0$. Let $h_n = g(x_1) \cdots g(x_n)$. The (possibly randomized) likelihood ratio test of power p is of the form: Reject θ_0 with probability φ_n, where

$$\begin{aligned} \varphi_n &= 1 \quad \text{if} \quad h_n > \hat{k}_n \\ &= \delta_n \quad \text{if} \quad h_n = \hat{k}_n \\ &= 0 \quad \text{if} \quad h_n < \hat{k}_n \end{aligned}$$

where $\hat{k}_n$ and δ_n are constants, $0 \leqq \delta_n \leqq 1$. Then

$$\text{(xi)} \qquad P_\theta(h_n > \hat{k}_n) \leqq E_\theta(\varphi_n) = p \leqq P_\theta(h_n \geqq \hat{k}_n)$$

for each n. Since $n^{-1} \log h_n \to K$ in probability when θ obtains, and since $0 < p < 1$, it follows from (xi) that

$$\text{(xii)} \qquad n^{-1} \log \hat{k}_n \to K.$$

We observe next that

$$\begin{aligned} \hat{\alpha}_n &= E_{\theta_0}(\varphi_n) \leqq P_{\theta_0}(h_n \geqq \hat{k}_n) = \int_{h_n \geqq \hat{k}_n} dP_{\theta_0}^{(n)} \\ &\leqq \int_{h_n \geqq \hat{k}} (\hat{k}_n{}^{-1} h_n)\, dP_{\theta_0}^{(n)} = \hat{k}_n{}^{-1} P_\theta(h_n \geqq \hat{k}_n) \leqq \hat{k}_n{}^{-1} \end{aligned}$$

for each n. Hence by (xii),

$$\text{(xiii)} \qquad \overline{\lim}_{n\to\infty} \{n^{-1} \log \hat{\alpha}_n\} \leqq -K.$$

Choose and fix $\epsilon > 0$ and let $a_n = \exp(n\epsilon)$. Then, for each n,

$$\begin{aligned} \hat{\alpha}_n &= \int_{X^{(n)}} \varphi_n \, dP_{\theta_0}^{(n)} \geqq \int_{h_n \leqq \hat{k}_n a_n} \varphi_n \, dP_{\theta_0}^{(n)} \\ &\geqq (\hat{k}_n a_n)^{-1} \int_{h_n \leqq \hat{k}_n a_n} \varphi_n \cdot h_n \cdot dP_{\theta_0}^{(n)} = (\hat{k}_n a_n)^{-1} \int_{h_n \leqq \hat{k}_n a_n} \varphi_n \, dP_\theta{}^{(n)} \\ &= (\hat{k}_n a_n)^{-1} [\int_{X^{(n)}} \varphi_n' \, dP_\theta{}^{(n)} - \int_{h_n > \hat{k}_n a_n} \varphi_n \, dP_\theta{}^{(n)}] \\ &\geqq (\hat{k}_n a_n)^{-1} [p - \int_{h_n > \hat{k}_n a_n} dP_\theta{}^{(n)}] = (\hat{k}_n a_n)^{-1} [p - P_\theta(h_n > \hat{k}_n a_n)]. \end{aligned}$$

Since $P_\theta(h_n > {}_n a_n) \to 0$ by (xii), and since $p > 0$, it follows that

$$\text{(xiv)} \qquad \underline{\lim}_{n\to\infty} \{n^{-1} \log \hat{\alpha}_n\} \geqq -K - \epsilon.$$

Since ϵ is arbitrary, the lemma is established.

To establish Proposition 9, let $\{T_n\}$ be a sequence with exact slope c. Choose and fix a non-null θ. If $J(\theta) = \infty$ there is nothing to prove. Suppose then that $J(\theta) < \infty$. Let $\{\theta_{0j}\}$ be a null sequence such that $K(\theta, \theta_{0j}) < \infty$ for each j and $K(\theta, \theta_{0j}) \to J(\theta)$.

Consider a particular θ_{0j} and call it θ_0 for simplicity. Let $\{k_n\}$ be a sequence such that $P_\theta(T_n \geqq k_n) \to p + \delta$, where $0 < \delta < 1 - p$. Let φ_n be the likelihood ratio test of θ_0 against θ of power p. Then $E_\theta(\varphi_n) < P_\theta(T_n \geqq k_n)$ for all sufficiently large n. Hence $E_{\theta_0}(\varphi_n) < P_{\theta_0}(T_n \geqq k_n)$ for all sufficiently large n. Hence, $E_{\theta_0}(\varphi_n) \leqq \alpha_n$ for all sufficiently large n, with α_n given by (37). Hence, by Proposition 11 (with p replaced by $p + \delta$) and Stein's lemma, $c(\theta) \leqq 2K(\theta, \theta_0)$, i.e., $c(\theta) \leqq 2K(\theta, \theta_{0j})$. Since j is arbitrary, $c(\theta) \leqq 2J(\theta)$.

Proposition 8 is an immediate consequence of (28).

Proposition 7 is a straightforward consequence of the Hartman-Wintner version of the law of the iterated logarithm.

Proposition 6 requires all the global and local conditions required by Proposition 3 and in addition the existence of various moment generation functions. This is to be expected, since Proposition 6 asserts that $\hat{T}_n$ is consistent and that $\hat{\alpha}_n \to 0$ at an exponential rate related to the asymptotic variance of $\hat{T}_n$.

Proposition 6 as stated here is a sharper version of a proposition obtained (in detail for $k = 1$ and in outline for $k > 1$) in [5]. What is shown in [5] is that

$$\text{(xv)} \qquad \lim_{\epsilon\to 0} \overline{\lim}_{n\to\infty} \{n\hat{\tau}_n^2\} = \lim_{\epsilon\to 0} \underline{\lim}_{n\to\infty} \{n\hat{\tau}_n^2\} = \hat{v}.$$

The writer thinks that the gap between (xv) and (16) can be closed but at present has proofs of (16) only in certain cases, e.g., $k = 1$, and k arbitrary but X finite.

Suppose first that θ is real valued, and that $g(\theta) = \theta$. Let $u(x, \theta) = (\partial/\partial\theta) \log f(x, \theta)$, and let $\varphi(t\colon \theta_1, \theta)$ denote the moment generating function of $u(x, \theta_1)$ when θ obtains. Let $\epsilon > 0$ and

$$\text{(xvi)} \qquad \begin{aligned} \rho_1(\epsilon, \theta) &= \inf\{\varphi(t\colon \theta + \epsilon, \theta) : t \geqq 0\}, \\ \rho_2(\epsilon, \theta) &= \inf\{\varphi(t\colon \theta - \epsilon, \theta) : t \leqq 0\}. \end{aligned}$$

Then, under general conditions, $0 < \rho_i < 1$ for all sufficiently small ϵ and $\rho_i \to 1$ as $\epsilon \to 0$ $(i = 1, 2)$. Let

$$\text{(xvii)} \qquad \hat{\gamma}(\epsilon, \theta) = 2 \min\{\log (1/\rho_1(\epsilon, \theta)), \log (1/\rho_2(\epsilon, \theta))\}.$$

Then, for all sufficiently small ϵ, $\lim_{n\to\infty} \{n\hat{\tau}_n^2(\epsilon, \theta)\} = \hat{w}(\epsilon, \theta)$ where

$$\text{(xviii)} \qquad \hat{w}(\epsilon, \theta) = \epsilon^2/\hat{\gamma}(\epsilon, \theta).$$

To establish this, choose and fix θ in Θ. For any n and s, let $\dot{L}_n(\cdot, s)$ denote the log-likelihood function based on $(x_1, \cdots, x_n)$. It is first shown, as in [5], by several applications of the Bernstein-Chernoff bound, that the following holds for each sufficiently small $h > 0$. There exists a β, $0 < \beta < 1$, and for each n a measurable event A_n depending only on $(x_1, \cdots, x_n)$, such that

$$\text{(xix)} \qquad P_\theta(A_n) > 1 - \beta^n$$

for all sufficiently large n, and such that, for each n, A_n implies all the following: the supremum of L_n over Θ is attained, and attained at exactly one point in Θ, say $\hat{\theta}_n$; $|\hat{\theta}_n - \theta| < h$; and L_n is a strictly concave function over $(\theta - h, \theta + h)$. Choose and fix a sufficiently small $h > 0$. Let $\epsilon > 0$ be so small that $\epsilon < h$, and ρ_1 and ρ_2 are both $>\beta$. It follows from (xix) that $|P_\theta(\hat{\theta}_n \geqq \theta + \epsilon) - P_\theta(L_n'(\theta + \epsilon, s) \geqq 0)| \leqq \beta^n$, where L_n' is the derivative of L_n with respect to θ. It now follows by an application of Chernoff's theorem that $n^{-1} \log P_\theta(\hat{\theta}_n \geqq \theta + \epsilon) \to \log \rho_1$. A similar argument shows that $n^{-1} \log P_\theta(\hat{\theta}_n \leqq \theta - \epsilon) \to \log \rho_2$. Hence $n^{-1} \log \hat{\alpha}_n \to -\frac{1}{2}\hat{\gamma}$, where $\hat{\gamma}$ is given by (xvi) and (xvii). Now (11) applies to show that, with $\hat{w}$ given by (xviii), $\hat{w}/n$ serves as the asymptotic effective variance of $\hat{T}_n \equiv \hat{\theta}_n$. That $\hat{w} \to \hat{v}$ as $\epsilon \to 0$ may be shown directly, or by appealing to (xv). Suppose now that g is an arbitrary (but sufficiently smooth) function and $g' \neq 0$ at θ. Then g is strictly monotone over $(\theta - h, \theta + h)$ provided $h > 0$ is sufficiently small. Consequently, with $\gamma = g(\theta)$, $|\hat{\alpha}_n(\epsilon, \theta) - P_\gamma(|\hat{\gamma}_n - \gamma| \geqq \epsilon)| < \beta^n$ for all ϵ and all sufficiently large n; hence etc.

Suppose now that $k > 1$, g is sufficiently smooth with $\hat{v}(\theta) > 0$, and X is finite. It is first shown, as in the case considered above, that for each sufficiently small $h > 0$ there exist β and A_n such that (xix) holds for all sufficiently large n, and such that A_n implies the following: L_n is maximized at exactly one point in Θ, say $\hat{\theta}_n = (\hat{\theta}_{1n}, \cdots, \hat{\theta}_{kn})$, and $d(\hat{\theta}_n, \theta) < h$. Now Sanov's theorem applies to show that $\hat{w}$ exists for each sufficiently small ϵ. The formula for $\hat{w}$ obtained thus is rather impenetrable, but (xv) is always available to complete the proof.

Suppose finally that $k > 1$ and X is arbitrary. In case $I(\theta)$ is independent of θ a refinement of the local arguments of [5] yields (16), but the case when $I(\theta)$ varies with θ remains open.

Proposition 5. The proof in [5] of this proposition requires certain unnecessary integrability assumptions on the framework. The proposition is valid provided only that g is continuously differentiable over Θ, and that for any θ_0 in Θ there exists a $k \times k$ positive definite matrix $I(\theta_0) = \{I_{ij}(\theta_0)\}$ such that, with K given by (32), $2K(\theta, \theta_0) \sim (\theta - \theta_0)I(\theta - \theta_0)'$ as $\theta \to \theta_0$. Under additional assumptions this I is necessarily the information matrix but even if it is not (15) holds with $\hat{v}$ defined by (5).

The improvements just described are immediately available from the argument in [5] because (as noted above) Stein's lemma is valid without the integrability assumptions invoked in [5] for this lemma. The argument in [5] proceeds as follows: Choose and fix a θ_0 in Θ. If $\hat{v}(\theta_0) = 0$ there is nothing to prove. Suppose then that $\hat{v}(\theta_0) > 0$, i.e., at least one of the partial derivatives of g at θ_0 is $\neq 0$. Then, with $I = I(\theta_0)$ and $h = (h_1(\theta_0), \cdots, h_k(\theta_0))$, $a = hI^{-1}$ is a non-zero vector. Let $\epsilon > 0$, and let $\theta_1 = \theta_0 + \epsilon \cdot a$. Let λ be a constant, $0 < \lambda < 1$, and let W_n be the critical region $\{|T_n - g(\theta_0)| \geqq \lambda\delta\}$ for testing θ_0 against θ_1, where $\delta = \epsilon \cdot ah'$. Since T_n is a consistent estimate of g, and since $|g(\theta_1) - g(\theta_0)| = \epsilon \cdot ah' + o(\epsilon) = \delta + o(\delta)$ as $\epsilon \to 0$, it follows that, for each sufficiently small ϵ, $P_{\theta_1}(W_n) \to 1$ as $n \to \infty$. Consequently, if ϵ is sufficiently small, $P_{\theta_0}(W_n) >$ the size of the likelihood ratio test of power $\frac{1}{2}$ (say) for all sufficiently large n. It now follows from Stein's lemma that, with α_n defined by (2),

$$\text{(xx)} \qquad \underline{\lim}_{n\to\infty}\{n^{-1}\log \alpha_n(\lambda\delta, \theta_0)\} \geqq -K(\theta_1, \theta_0).$$

By dividing both sides of (xx) by $\lambda^2\delta^2$, and observing that $\lambda\delta$ decreases continuously to zero as ϵ decreases to zero, it follows that

$$\text{(xxi)} \quad \underline{\lim}_{\epsilon\to0}\,\underline{\lim}_{n\to\infty}\{(n\epsilon^2)^{-1}\log \alpha_n(\epsilon, \theta_0)\} \geqq -\lambda^{-2}\,\overline{\lim}_{\epsilon\to0}\{K(\theta_1, \theta_0)/\epsilon^2(ah')^2\}.$$

Since $2K(\theta_1, \theta_0) \sim (\theta_1 - \theta_0)I(\theta_1 - \theta_0)'$ as $\theta_1 \to \theta_0$, it follows from the definitions of θ_1, a, and $\hat{v}$ that

(xxii) $$\lim_{\epsilon\to 0} K(\theta_1, \theta_0)/\epsilon^2(ah')^2 = 1/2\hat{v}(\theta_0).$$

Since λ is arbitrary, we conclude from (xxi) and (xxii) that

(xxiii) $$\underline{\lim}_{\epsilon\to 0}\, \underline{\lim}_{n\to\infty} \{(n\epsilon^2)^{-1} \log \alpha_n(\epsilon, \theta_0)\} \geqq -1/2\hat{v}(\theta_0).$$

In view of (11), (xxiii) is equivalent to the validity of (15) at $\theta = \theta_0$.

Proposition 4 is immediate from (9), (12), and (cf. [19], p. 166) $z^2 \sim 2\log(1/\delta)$ as $\delta \to 0$.

Proposition 3. Cf. the paragraph immediately following the statement of this proposition. General assumptions which guarantee the consistency of maximum likelihood estimates were first formulated by Wald [41]. The following version of Wald's theorem is based mainly on [29], [41] and partly on [10], [35].

Let there be given the abstract sample space X, and abstract parameter space Θ, and the family $\{f(x, \theta)\}$ of probability densities relative to fixed a σ-finite measure μ, and let $P_\theta(A) = \int_A f(x, \theta)\, d\mu$ for $A \subset X$. Let us then say, as in [10], that a compact metric space $\bar{\Theta}$ of points θ, with distance function d, is a suitable compactification of Θ if the following conditions (i)–(iii) are satisfied. (i) Θ is an everywhere dense subset of $\bar{\Theta}$. (ii) For each θ_1 in $\bar{\Theta}$, $g(x, \theta_1, \epsilon) = \sup\{f(x, \theta): \theta \text{ in } \Theta, d(\theta_1, \theta) < \epsilon\}$ is measurable in x for all sufficiently small $\epsilon > 0$. (iii) With $g(x, \theta_1, 0) = \lim_{\epsilon\to 0} g(x, \theta_1, \epsilon)$, $\int_X g(x, \theta_1, 0)\, d\mu \leqq 1$ for all θ_1 in $\bar{\Theta}$. In the following, for any set $\Theta_0 \subset \Theta$ and any extended real valued function $h(\theta)$, we write $h(\Theta_0) = \sup\{h(\theta): \theta \text{ in } \Theta_0\}$. If there exists a suitable compactification of Θ it follows, in particular, that $f(x, \Theta)$ is measurable in x.

Now assume that

(a) there exists a suitable compactification of Θ, say $\bar{\Theta}$;

(b) $E_\theta[\log(f(x, \Theta)/f(x, \theta))] < \infty$ for each θ in Θ;

(c) if θ and θ_1 are points in Θ and $\bar{\Theta}$ respectively, with $\theta \neq \theta_1$, then $\mu\{x: f(x, \theta) \neq g(x, \theta_1, 0)\} > 0$;

(d) Θ is open in $\bar{\Theta}$;

(e) for θ in Θ, $f(x, \theta) = g(x, \theta, 0)$ for all x.

For any n, θ in Θ, and s let $l_n(\theta, s) = \prod_{r=1}^{n} f(x_r, \theta)$. Let $\hat{\Theta}_n(s) = \{\theta: \theta \text{ in } \Theta, l_n(\theta, s) = l_n(\Theta, s)\}$. Then, for given s, $\hat{\Theta}_n$ is the (possibly empty) set of all maximum likelihood estimates of θ based on $(x_1, \cdots, x_n)$. It can be shown that, with probability one, $\hat{\Theta}_n$ is non-empty for all sufficiently large n and that the entire set $\hat{\Theta}_n$ converges to θ.

To establish this, choose and fix a θ in Θ and suppose henceforth that θ obtains. For each n and s let $\Theta_n^*(s) = \{\theta_1 : \theta_1 \text{ in } \Theta, l_n(\theta_1, s) \geqq \frac{1}{2} l_n(\Theta, s)\}$. It follows from (b) that Θ_n^* is non-empty with probability one. Note that $\hat{\Theta}_n \subset \Theta_n^*$ for every n and s. We proceed to show that, with probability one,

(xxiv) $$\sup\{d(\theta, \theta_1): \theta_1 \text{ in } \Theta_n^*\} \to 0.$$

For any θ_1 in $\bar{\Theta}$ let $S(\theta_1, \epsilon) = \{\theta: \theta \text{ in } \bar{\Theta}, d(\theta_1, \theta) < \epsilon\}$. Suppose that $\theta_1 \neq \theta$. It then follows from (b) and (c) by condition (iii) of (a) that there exist $\epsilon_1 > 0$ and $c_1 > 0$ such that

(xxv) $$-\infty \leqq E_\theta(\log[g(x, \theta_1, \epsilon_1)/f(x, \theta)]) < -c_1.$$

In the following, for any set $\Delta \subset \overline{\Theta}$ let $\Delta^0 = \Delta \cap \Theta$. Since $l_n(S^0(\theta_1, \epsilon_1), s) \leqq \prod_{r=1}^n g(x_r, \theta_1, \epsilon_1)$, it follows from (xxv) that, with probability one,

$$\text{(xxvi)} \qquad (1/n) \log (l_n(S^0(\theta_1, \epsilon_1), s)/l_n(\theta, s)] < -c_1$$

for all sufficiently large n.

Now choose and fix $\epsilon > 0$ so small that with $\Delta = \overline{\Theta} - S(\theta, \epsilon)$, Δ^0 is non-empty. The preceding argument applies to each θ_1 in Δ. Since Δ is compact, there exist points $\theta_1, \cdots, \theta_k$ (and corresponding $\epsilon_1, \cdots, \epsilon_k$ and $c_1, \cdots, c_k$ of the above paragraph) such that $\Delta \subset \bigcup_{i=1}^k S(\theta_i, \epsilon_i)$. Hence $\Delta^0 \subset \bigcup_1^k S^0(\theta_i, \epsilon_i)$, and so $l_{\bar{n}}(\Delta^0, s) \leqq l_n(\bigcup_1^k S^0(\theta_i, \epsilon_i), s) = \max \{l_n(S^0(\theta_i, \epsilon_i), s) : 1 \leqq i \leqq k\}$ for every n and s. It now follows from (xxvi) that, with probability one,

$$\text{(xxvii)} \qquad (1/n) \log [l_n(\Delta^0, s)/l_n(\theta, s)] < -c$$

for all sufficiently large n, where $c = \min \{c_1, \cdots, c_k\}$, $0 < c < \infty$. It follows from (xxvii) and the definition of Θ_n^* that, with probability one,

$$\text{(xxviii)} \qquad \Theta_n^* \subset S(\theta, \epsilon)$$

for all sufficiently large n. Since ϵ in (xxviii) is arbitrarily small, (xxiv) holds with probability one. It remains to show that $\hat{\Theta}_n$ is non-empty for all sufficiently large n.

For any set $\Delta \subset \overline{\Theta}$ let $\bar{\Delta}$ denote its closure in $\overline{\Theta}$. Let $\epsilon > 0$ be so small that $\bar{S}(\theta, \epsilon) \subset \Theta$; such an ϵ exists, by (d). Choose and fix an s and n such that (xxviii) holds. Then

$$\text{(xxix)} \qquad \overline{\Theta}_n^* \subset S^0(\theta, \epsilon).$$

By the definition of Θ_n^*, it contains a sequence $\{\theta_1, \theta_2, \cdots\}$ such that $l_n(\theta_j) \to l_n(\Theta)$ as $j \to \infty$. By passing to a subsequence, if necessary, we may suppose that $\theta_j \to \theta_0$ say, where θ_0 is in $\overline{\Theta}_n{}^*$; hence θ_0 is in Θ, by (xxix). It is clear that for any $\epsilon_0 > 0$, $\prod_{r=1}^n g(x_r, \theta_0, \epsilon_0) \geqq l_n(\theta_j, s)$ for all sufficiently large j; hence $\prod_1^n g(x_r, \theta_0, \epsilon_0) \geqq l_n(\Theta)$. Since ϵ_0 is arbitrary, it follows from (e) that $l_n(\theta_0) \geqq l_n(\Theta)$; i.e., θ_0 is a point in $\hat{\Theta}_n$.

The preceding formulation and proof depend heavily on the particular versions of the density functions $f(x, \theta)$ under consideration; this is appropriate, since the existence and properties of maximum likelihood estimates also depend on the versions in use.

It is plain that conditions (d) and (e) can be dispensed with if we presume the existence of maximum likelihood estimates. It can be shown by simple examples (in which (a), (b), (c) hold) that if (d) or (e) does not hold then $\hat{\Theta}_n$ can be empty for every n and s. Conditions (d) and (e) are, however, rather peripheral in the sense that in typical cases they are satisfied if (a) holds. It is interesting that the identifiability condition (c) is automatically satisfied if θ denotes the unknown probability distribution, i.e., if Θ is a set of probability measures on X. This last is a natural parametrization since maximum likelihood always estimates the entire distribution from given data.

It thus seems that (a) and (b) are the main assumptions. It is known [4], [29] that conditions such as (a) and (b) are at once quite restrictive, often hard to verify, and indispensable to any general proof of consistency.

Suppose now that Θ is an open set in k dimensional Euclidean space, that $f(x, \theta)$ (and therefore each $l_n(\theta, s)$) is positive and differentiable in θ. It then follows immediately from the conclusion just established that, with probability one, for all sufficiently large n, each point in $\hat{\Theta}_n$ is a solution of $\partial L_n(\theta, s)/\partial\theta_i = 0$ $(i = 1, \cdots, k)$ where $L_n = \log l_n$. It can be shown that there exists, for each n, a measurable function $\hat{\theta}_n$ of $(x_1, \cdots, x_n)$ with values in Θ such that $\hat{\theta}_n$ is in $\hat{\Theta}_n$ whenever the latter set is non-empty. Assume that g is a continuously differentiable function of θ. Then $\hat{T}_n(s) = g(\hat{\theta}_n(s))$ is measurable, and (with probability one when θ obtains) $\hat{T}_n$ is a maximum likelihood estimate of g for all sufficiently large n and $\hat{T}_n \rightarrow \theta$. The measurability of $\hat{T}_n$ is essential to the statement of Proposition 3.

The additional local regularity conditions required (for the present purpose of completing the proof of Proposition 3, and for other purposes) have been formulated by Cramér, LeCam, Rao, and many others. In particular, the conditions stated at the outset of Section 3 of [9] suffice for the second part of the proof of Proposition 3. These conditions imply, incidentally, that with probability one $\hat{\Theta}_n$ consists of a single point for all sufficiently large n.

Proposition 2 can be established easily by a slight modification of the argument in Section 3 of [9] as follows. Assume that the framework satisfies the local regularity conditions stated at the outset of Section 3 on p. 1550 of [9], and that g is continuously differentiable over Θ. Let θ^0 be a point in Θ, let $(a_1, \cdots, a_k)$ be a fixed non-zero vector, and let $\theta_n{}^0 = \theta^0 + n^{-\frac{1}{2}}a$. By taking D_n to be $\{T_n \geqq g(\theta_n{}^0)\}$ in the argument on p. 1551 of [9] it follows that *if*

$$\text{(xxx)} \qquad \overline{\lim}_{n\to\infty} P(T_n \geqq g(\theta_n{}^0) \mid \theta_n{}^0) \geqq \tfrac{1}{2}$$

then

$$\text{(xxxi)} \qquad (v(\theta^0))^{\frac{1}{2}} \geqq ah'/(aIa')^{\frac{1}{2}}$$

where $h = (h_1, \cdots, h_k)$ is the vector of partial derivatives of g at θ^0, and $I = I(\theta^0)$.

Since $\lim_{n\to\infty} P(T_n \geqq g(\theta) \mid \theta) = \frac{1}{2}$ for each θ, it follows (cf. Section 2 of [9]) that, for given a, there exists a null set N and a sequence $m_1 < m_2 < \cdots$ of positive integers such that, for all θ^0 in $\Theta - N$, $P(T_n \geqq g(\theta_n{}^0) \mid \theta_n{}^0) \rightarrow \frac{1}{2}$ as $n \rightarrow \infty$ through the sequence $\{m_1, m_2 \cdots\}$. Hence, for almost all θ^0 in Θ, (xxx) holds for all non-zero a with rational co-ordinates. Hence, for almost all θ^0, (xxxi) holds for all non-zero rational a and consequently for all non-zero a. If (xxxi) holds for all non-zero a then $\hat{v}(\theta^0)$ cannot be less than $\hat{v}(\theta^0)$.

The conclusion stated and proved in Section 3 of [9] is a consequence of Proposition 2. For, if t_n is an asymptotically normal estimate of θ, then for any $b = (b_1, \cdots, b_k)$, $T_n = t_nb'$ is an asymptotically normal estimate of $g(\theta) = \theta b'$. Hence, with V the asymptotic covariance of $n^{\frac{1}{2}}t_n$, $bVb' \geqq b'I^{-1}b$ for almost all θ, by Proposition 2. Since b is arbitrary, it follows that $V - I^{-1}$ is positive semi-definite for almost all θ.

Proposition 1 is perhaps an unfamiliar formulation but the proof is more or less immediate.

REFERENCES

[1] ABRAHAMSON, I. G. (1965). On the stochastic comparison of tests of hypotheses. Ph.D. dissertation, Univ. of Chicago.

[2] ABRAHAMSON, I. G. (1967). The exact Bahadur efficiencies for the Kolmogorov-Smirnov and Kuiper one- and two-sample statistics. To appear in *Ann. Math. Statist.*

[3] ANDERSON, T. W. and GOODMAN, L. A. (1957). Statistical inference about markov chains. *Ann. Math. Statist.* **28** 89–109.

[4] BAHADUR, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhyā* **20** 207–210.

[5] BAHADUR, R. R. (1960). Asymptotic efficiency of tests and estimates. *Sankhyā* **22** 229–252.

[6] BAHADUR, R. R. (1960). Simultaneous comparison of the optimum and sign tests of a normal mean. *Contributions to Probability and Statistics—Essays in Honor of Harold Hotelling.* Stanford Univ. Press. 79–88.

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

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

[9] BAHADUR, R. R. (1964). On Fisher's bound for asymptotic variances. *Ann. Math. Statist.* **35** 1545–1552.

[10] BAHADUR, R. R. (1965). An optimal property of the likelihood ratio statistic. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** Univ. of California Press.

[11] BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

[12] BAHADUR, R. R. and BICKEL, P. J. (1966). Asymptotic optimality of Bayes statistics. *Sankhyā.* (To appear).

[13] BAHADUR, R. R. and BICKEL, P. J. (1966). Conditional levels in large samples. *Contributions to Statistics and Probability—Essays in memory of S. N. Roy.* Univ. of N. Car. Press. (To appear).

[14] BASU, D. (1956). On the concept of asymptotic efficiency. *Sankhyā* **17** 193–196.

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

[16] COCHRAN, W. G. (1952). The χ^2 goodness of fit test. *Ann. Math. Statist.* **23** 315–345.

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

[18] FELLER, W. (1943). Generalization of a probability theorem of Cramér. *Trans. Amer. Math. Soc.* **54** 361–372.

[19] FELLER, W. (1957). *An Introduction to Probability and Its Applications* (2nd Edition). Wiley, New York.

[20] FERGUSON, T. S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Ann. Math. Statist.* **29** 1046–1062.

[21] FISHER, R. A. (1925). Theory of statistical estimation. *Contributions to Mathematical Statistics* (1950). Wiley, New York.

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

[23] GLESER, L. J. (1966). The comparison of multivariate tests of hypothesis by means of Bahadur efficiency. *Sankhyā Ser. A* **28**, Parts 2 and 3.

[24] HOADLEY, A. B. (1965). The theory of large deviations with statistical applications. Ph.D. dissertation, Univ. of California at Berkeley.

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

[26] HOEFFDING, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408.

[27] HOEFFDING, W. (1965). Large deviations in the multinomial distribution. *Fifth Berkeley Symp. Math. Statist. Prob.* **1**.

[28] JOFFE, A. and KLOTZ, J. (1962). Null distribution and Bahadur efficiency of the Hodges bivariate sign test. *Ann. Math. Statist.* **33** 803–807.

[29] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

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

[31] LECAM, L. (1953). On some asymptotic properties of maximum likelihood and related Bayes estimates. *University of California Publications in Statistics.*

[32] LECAM, L. (1955). On the asymptotic theory of estimation and testing hypotheses. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 129–156. Univ. of California Press.

[33] LECAM, L. (1958). Les propriétés asymptotiques des solutions des Bayes. *Publ. Inst. Statist. Univ. Paris* **7** 17–35.

[34] NEYMAN, J. (1949). Contributions to the theory of the χ^2 test. *Berkeley Symp. Math. Statist. Prob.*, 239–273. Univ. of California Press.

[35] RAO, C. R. (1957). Maximum likelihood estimation for the multinomial distribution. *Sankhyā* **18** 139–148.

[36] RAO, C. R. (1963). Criteria of estimation in large samples. *Sankhyā Ser. A* **25** 189–206.

[37] RUBIN, H. and SETHURAMAN, J. (1965). Probabilities of moderate deviations. Bayes risk efficiency. *Sankhyā Ser. A* **27** 325–356.

[38] SANOV, I. N. (1957). On the probability of large deviations of random variables. *Sel. Transl. Math. Statist. Prob.* **1** 213–244.

[39] SCHMETTERER, L. (1966). On the asymptotic efficiency of estimates. *Research Papers in Statistics* (Neyman Festschrift), 301–317. Wiley, New York.

[40] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316.

[41] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

[42] WOLFOWITZ, J. (1965). Asymptotic efficiency of the maximum likelihood estimator. *Theor. Prob. Appl.* **10** 247–260.