The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1206–1240

NON-LOCAL ASYMPTOTIC OPTIMALITY OF APPROPRIATE LIKELIHOOD RATIO TESTS

BY L. D. BROWN

Cornell University

CONTENTS

1. Introduction. Suppose $x_1, x_2, \cdots, x_n$ are n independent identically distributed observations on a random variable having distribution F_θ, $\theta \in \Theta$. Suppose it is desired to test the null hypothesis $\theta \in \Theta_0$ versus the alternative $\theta \in \Theta_1 = \Theta - \Theta_0$. Several different methods have been proposed to relate asymptotic performance (as $n \to \infty$) of two (or more) different sequences of tests. It may be said that these methods fall into two broad categories.

First, there are "local" methods such as Pitman efficiency and its generalizations, see Noether (1955), Neyman (1959). In these approaches properties of tests are compared at appropriately chosen sequences of points in the alternative hypothesis (and perhaps also a sequence of points in the null hypothesis). A different alternative point is chosen for each sample size n, and properties are compared as $n \to \infty$. Generally the sequences of points are chosen so that the probabilities of type I error and type II error at the chosen points remain bounded away from zero for at least one of the sequences of tests under comparison. The characteristic of these methods which makes the name "local" appropriate is that the sequence of points in the alternative hypothesis gets arbitrarily close to the null hypothesis.

On the other hand there are the "non-local," or "fixed alternative," methods. In these methods the rate of exponential convergence to zero of the significance level and/or of type II error at a particular point are examined. Denoting probabilities of type I and type II error by α_n and β_n, respectively, one thus looks either at $\lim_{n\to\infty} n^{-1} \log \alpha_n(\theta)$ for fixed $\theta \in \Theta_0$, or more usually $\lim_{n\to\infty} n^{-1} \log \sup_{\theta \in \Theta_0} \alpha_n(\theta)$, and/or at $\lim_{n\to\infty} n^{-1} \log \beta_n(\theta)$ for fixed $\theta \in \Theta_1$. (See Section 2 for explicit definitions of these terms and others used in the introduction.)

While one might consider other measures of rate of approach of α_n and/or β_n to 0, the exponential measurement described above seems to be right for "non-local" asymptotic properties. It is discriminating enough to provide non-trivial com-

Received October 3, 1968; revised January 4, 1971.

parisons between different sequences of tests, but not so discriminating that the comparisons become essentially as difficult and possibly unrewarding as admissibility comparisons for fixed sample sizes.

We shall show under appropriate regularity conditions that using this measure of speed of approach to zero there exists a sequence of tests which is asymptotically optimal.

To be more precise, in Section 2 we construct a test statistic λ_n^* which we prove in Section 5 has properties described in the following paragraphs.

Let $\alpha_n^s = \sup_{\theta \in \Theta_0} \alpha_n(\theta)$ and $\beta_n^s(\theta)$ be the significance level and probability of type II error of any sequence of tests of Θ_0 versus Θ_1. Let α_n^i and $\beta_n^i(\theta)$ be the significance level and probability of type II error of the tests which reject if

$$\lambda_n^* > C_n^i, \qquad i = 1, 2.$$

The following results are valid under appropriate regularity conditions which are given in Theorems 2 and 3, Section 5.

MAIN RESULT 1. *If* $\limsup_{n\to\infty} \alpha_n^s < 1$, *then the constants* C_n^1 *can be chosen so that*

$$\alpha_n^1 \leqq \alpha_n^s \tag{1.1}$$

and for all $\theta \in \Theta_1$

$$\liminf_{n\to\infty} \left(n^{-1} \log \beta_n^s(\theta) - n^{-1} \log \beta_n^1(\theta)\right) \geqq 0. \tag{1.2}$$

For the purposes of discussion we give an alternate, closely related result

MAIN RESULT 2. *Fix* $\theta_1 \in \Theta_1$. *If* $\limsup_{n\to\infty} \beta_n^s(\theta_1) < 1$ *then the constants* C_n^2 *can be chosen so that*

$$\beta_n^2(\theta_1) \leqq \beta_n^s(\theta_1) \tag{1.3}$$

and

$$\liminf_{n\to\infty} \left(n^{-1} \log \alpha_n^s - n^{-1} \log \alpha_n^2\right) \geqq 0. \tag{1.4}$$

There are several other possible alternate forms; they should be clear from the proofs in Sections 4 and 5.

[It is to be emphasized that the central part of Main Result 2 from our present point of view is the fact that the test statistic λ_n^* itself does not depend on the choice of θ_1 or on the values of $\beta_n(\theta_1)$. If this were not required to be the case then this Result would take on a very different character, viz: "If $\limsup_{n\to\infty} \beta_n^s(\theta_1) < 1$ then there exists a sequence of tests such that the appropriate versions of the conclusions (1.3) and (1.4) are satisfied." In fact these tests can be based on the likelihood ratio statistic for testing Θ_0 versus the simple alternative $\{\theta_1\}$. R. Bahadur who pointed out to us this observation (1969) also observed that if $\liminf_{n\to\infty} \beta_n^s(\theta_1) > 0$ then this latter result can be deduced by a relatively simple argument under less stringent regularity conditions than we require in the present

paper. We remark that even if $\liminf_{n\to\infty}\beta_n^s(\theta_1) = 0$ this latter result can be deduced from our Main Result 2 and the construction of λ_n^*—subject of course to the regularity conditions under which Main Result 2 is valid.]

When $\lim \alpha_n^s$ exists and $0 < \lim \alpha_n^s < 1$, a type of comparison such as that in Result 1 is used in Hodges and Lehmann (1956).

More interestingly, when $\lim \beta_n^s(\theta_1)$ exists and $0 < \lim \beta_n^s(\theta_1) < 1$, Result 2 can be made into a statement about exact slopes. In that case Bahadur (1966) has shown that the sequence of likelihood ratio tests of Θ_0 versus Θ_1 having power $1-\beta_n(\theta_1)$ at θ_1 possesses the same optimal property as that described above for the tests based on our statistic λ_n^*.

Hoeffding (1965) has studied rates of convergence of error probabilities in the Multinomial Case. In that case (under appropriate mild assumptions on Θ_0) there is a likelihood ratio test having error probabilities which satisfy the conclusions of Result 1. Hoeffding has shown even more. Namely, if each of the tests in the sequence S are appropriately different from the likelihood ratio test and if $(\log n)^{-1} \log \alpha_n^s \to -\infty$ then strict inequality holds in the analog of (1.2) (in which β_n^1 is of course replaced by the error probabilities of the appropriate likelihood ratio test, say $\beta_n^L(\theta)$) at "most" parameter points in Θ_1 for which $\beta_n^L(\theta_1) \to 0$.

Herr (1967) has studied a particular, common, family of problems involving the multivariate normal distribution. Without completely proving that the appropriate likelihood ratio tests satisfy the analog of Result 1, he has proved a result much like the second quoted result of Hoeffding's.

As must be expected, our statistic is equivalent to the likelihood ratio statistic in the cases considered by Hoeffding and Herr. In general, λ_n^* is n^{-1} times the log of a likelihood ratio statistic for *some* statistical problem. *However* this statistical problem may not be the same as the statistical problem under consideration. λ_n^* will, in general, be the statistic for the "larger" problem of testing, Θ_0^* versus Θ_1^*, say, where $\Theta_0 \subset \Theta_0^*$ and $\Theta_1 \subset \Theta_1^*$. (More exactly, the set of distributions $\{F_\theta\}_{\theta\in\Theta_0}$ satisfy $\{F_\theta\}_{\theta\in\Theta_0} \subset \{F_\theta\}_{\theta\in\Theta_0^*}$; etc.) Usually $\Theta_0 = \Theta_0^*$, however often $\Theta_1 \neq \Theta_1^*$ so that λ_n^* is essentially different from the likelihood ratio statistic, say, λ_n, for testing Θ_0 versus Θ_1. Denote the error probabilities of this test by α_n^λ, $\beta_n^\lambda(\theta)$, etc. When $\Theta_1 \neq \Theta_1^*$ it will often be true that, for example, when $\alpha_n^1 = \alpha_n^\lambda$ there exist many values of $\theta \in \Theta_1$ such that

$$\liminf_{n\to\infty} \left(n^{-1} \log \beta_n^\lambda(\theta) - n^{-1} \log \beta_n^1(\theta)\right) > 0. \tag{1.5}$$

Thus the test based on λ_n^* is asymptotically non-locally better than the test based on λ_n. It is even quite possible that the inequality in (1.5) will hold for all $\theta \in \Theta_1$.

Theorem 1(g) describes a condition which, if violated by Θ_1, will lead to inequality in (1.5). Further results are given in Section 6.

It is often the case that the reduction of an original alternative hypothesis—Θ_1^*—to a smaller one—Θ_1—results from some extra information about the alternative hypothesis. If so, the result mentioned above leads to the following:

HEURISTIC PRINCIPLE 1. *If you have some "extra" information about the alternative hypothesis, forget it! For a given significance level, to use this extra information in a likelihood ratio test may quite likely result in a test with exponentially larger error probabilities. Furthermore, the rate of exponential convergence to zero of the error probabilities when using this extra information (in any sort of a test) cannot be smaller than those of the best test which does not use this information.*

We remark that the applicability of this principle may be somewhat restricted since we have made some assumptions in Sections 2 and 4. In particular, the absolute continuity requirements may restrict this principle. See Section 7.

It should be emphasized that the last sentence of the above principle does not imply that *nothing* can ever be gained by utilizing "extra" information; it only says that the rate of exponential convergence to zero cannot be improved. (For an example see Bahadur (1966) Remark 3b.) In fact, if the extra information is of a very special type, all of the error probabilities can be somewhat decreased.

Let us also remind the reader that the above principle refers only to non-local properties. We do not know in general whether "extra" information can only improve local properties of likelihood ratio tests, or whether, as with non-local properties, the converse is sometimes true.

The application of this principle is illustrated in the examples of Section 6.

It is most important to emphasize that this principle does not apply to "extra" information about the null hypothesis. In fact the opposite result holds for extra information about the null hypothesis, as the following principle indicates.

HEURISTIC PRINCIPLE 2. *If you have any extra information about the nullhypothesis, use it in forming the likelihood ratio test. For a given significance level it cannot increase the rate of exponential convergence to zero of type* II *error, and it may decrease it.*

In principle, the results in Corollaries 1 and 2, Section 6, can be used to describe when specific "extra" information about the alternative (null) hypothesis increases (decreases) the rate of exponential convergence of probabilities of type II error. In practice these results are sometimes easy and sometimes difficult to apply.

Roughly speaking this difference between information about the null hypothesis and the alternative hypothesis is due to the following consideration: Information about the alternative hypothesis is only used to alter the statistic in the testing problem. For our purposes the likelihood ratio statistic which makes the least limitations on the alternative hypothesis consistent with the general assumptions of Theorem 1–3 turns out to be an optimal statistic. On the other hand limitation of the null hypothesis performs two functions. First, it also alters the likelihood ratio statistic. It is clear that a test based on such a modified statistic cannot have better asymptotic non-local properties than the optimal test described above. However, limitation of the null hypothesis at the same time decreases the family of probabilities which contribute to defining the significance level of the test. More precisely, after such a limitation the significance level is now the supremum over a smaller family of error probabilities.

The precise way in which such extra information about the null hypothesis acts is not entirely clear without specific reference to Theorem 1. Consider, for example, the problem of testing Θ_0 versus Θ_1 at level $\alpha_n = e^{-na}$, so that $\lim_{n\to\infty} n^{-1} \log \alpha_n = -a$. Let $\alpha_n(\theta)$, $\theta \in \Theta_0$ denote the probability of type I error at $\theta \in \Theta_0$ of an "optimal test" of Θ_0. Then, a reduction of Θ_0 to Θ_0', say, ($\Theta_0' \subset \Theta_0$) which removes some, *but not all*, of the points, θ, such that $\lim_{n\to\infty} n^{-1} \log \alpha_n(\theta) = -a$ may result in an asymptotic exponential reduction of type II error for some, or all, points in Θ_1. Perhaps more surprisingly, a removal from Θ only of points, θ, for which $\lim n^{-1} \log \alpha_n(\theta) < -a-\varepsilon$, $\varepsilon > 0$, can sometimes result in an asymptotic exponential reduction of type II error for some or all points in Θ_1. (We leave the detailed proof based on Theorem 1 of this result to the interested reader.)

As to the appropriateness of using the rate of exponential convergence of error probabilities to zero as a measure of non-local asymptotic performance we mention that *for each fixed* $\theta_1 \in \Theta_1$ a statistic optimal in this sense has "asymptotic relative efficiency" at least 1 at θ_1. This can be deduced directly from Result 1 or Result 2, much as in Theorem 2 of Bahadur (1966). However, a more interesting conclusion would be that an appropriate sequence of tests based on λ_n^* has asymptotic relative efficiency at least 1 (uniformly on Θ_1).

(To be precise, one would want to show that for any sequence, S, of tests there is a sequence T of tests based on λ_n^* and a function $m(n)$ satisfying

$$\liminf_{n\to\infty} n/m(n) \geqq 1$$

$$\alpha_{m(n)}^T \leqq \alpha_n^S$$

$$\beta_{m(n)}^T(\theta) \leqq \beta_n(\theta) \qquad \text{for all } \theta \in \Theta_1.$$

Since we have not been able to prove (or disprove) such a result, we do not pursue this topic further here.)

Section 2 contains the definitions and assumptions used in the remainder of the paper. Section 3 is devoted to several lemmas including Lemma 4, which is a standard type of result about probabilities of large deviations. The main mathematical result is Theorem 1 in Section 4. This theorem is applied in Section 5 to prove the Main Results quoted in this Introduction. These are Theorems 2 and 3 of Section 5. Theorem 1 is also applied in Section 6—in Corollaries 1 and 2—to prove formal versions of the Heuristic Principles described in this Introduction. The bulk of Section 6 is devoted to examples which illustrate the application of Corollaries 1 and 2. The final section, Section 7, contains an example and some remarks concerning the necessity of the absolute continuity assumptions used elsewhere in the paper. Theorem 4 of Section 7 partially removes one of these assumptions. The reader interested primarily in results and examples rather than proofs may proceed directly from Section 2 to the statement of Theorem 1 and then to either Section 5 or Section 6.

2. Notation and assumptions. Let $\{F_\theta\}_{\theta\in\Theta}$ be a family of probability distributions on a probability space $\mathcal{X}$, $\mathcal{B}$. We assume throughout that the distributions are all

dominated by a σ-finite measure, ν, and we denote their density by $dF_\theta/d\nu = f(\cdot, \theta)$. We also assume $\theta_1 \neq \theta_2$ implies $F_{\theta_1} \neq F_{\theta_2}$.

We suppose $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \varnothing$ and we are interested in testing the null hypothesis $\theta \in \Theta_0$ versus the alternative $\theta \in \Theta_1$. In order to describe the assumptions to be made on the family of densities, we describe an imbedding of Θ into a (perhaps) larger space Θ^* and some related notational conventions. If θ_0, $\theta_1 \in \Theta$, and $f(\cdot, \theta_0)$ and $f(\cdot, \theta_1)$ are not mutually singular densities, define for $0 < \xi < 1$

$$c(\xi, \theta_0, \theta_1) = \left(\int f^{1-\xi}(x, \theta_0) f^{\xi}(x, \theta_1) \nu(dx)\right)^{-1}$$

(Note: $c < \infty$), and define the density corresponding to the new parameter $(\xi, \theta_0, \theta_1)$ by

$$f(x, (\xi, \theta_0, \theta_1)) = c(\xi, \theta_0, \theta_1) f^{1-\xi}(x, \theta_0) f^{\xi}(x, \theta_1).$$

See Lemma 2 for some properties of the function c. If $f(\cdot, \theta_0)$ and $f(\cdot, \theta_1)$ are not mutually singular, define

$$f(x, (0, \theta_0, \theta_1)) = \lim_{\xi \searrow 0} f(x, (\xi, \theta_0, \theta_1)) = c(0, \theta_0, \theta_1) f(x, \theta_0) \operatorname{sgn} f(x, \theta_1)$$

where $c(0, \theta_0, \theta_1) = \left(\int f(x, \theta_0) \operatorname{sgn} f(x, \theta_1) \nu(dx)\right)^{-1}$, and $\operatorname{sgn} a = 1$ if $a > 0$, $= 0$ if $a = 0$. Similarly let

$$f(x, (1, \theta_0, \theta_1)) = \lim_{\xi \nearrow 1} f(x, (\xi, \theta_0, \theta_1)) = c(1, \theta_0, \theta_1) f(x, \theta_1) \operatorname{sgn} f(x, \theta_0).$$

Both limits exist. In fact, if $f(\cdot, \theta_0)$ and $f(\cdot, \theta_1)$ are mutually absolutely continuous $f(x, (0, \theta_0, \theta_1)) = f(x, \theta_0)$ and $f(x, (1, \theta_0, \theta_1)) = f(x, \theta_1)$.

We will assume there exists a locally compact second countable metric space Θ^* with associated probability densities $\{f(\cdot, \theta)\}_{\theta \in \Theta^*}$ which satisfies:

(2.1a) $$\forall \theta \in \Theta \exists \theta' \in \Theta^* \ni f(\cdot, \theta) = f(\cdot, \theta') \text{ a.e. } (\nu)$$

(2.1b) $$\theta_0 \in \Theta_0, \theta_1 \in \Theta_1, 0 \leqq \xi \leqq 1 \Rightarrow \exists \theta \in \Theta^* \text{ such that } f(\cdot, (\xi, \theta_0, \theta_1)) = f(\cdot, \theta) \text{ a.e. } (\nu)$$

(2.1c) $$\theta_1 \neq \theta_2 \text{ implies } f(\cdot, \theta_1) \not\equiv f(\cdot, \theta_2) \text{ a.e. } (\nu)$$

and some additional properties to be described later. (2.1b) is meant to contain the assumption that for all $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$, $f(\cdot, \theta_0)$ and $f(\cdot, \theta_1)$ are not mutually singular. (This is explicitly well contained within Assumption 4.) In other words the densities $\{f(\cdot, \theta)\}_{\theta \in \Theta}$ may be imbedded in a natural way into $\{f(\cdot, \theta)\}_{\theta \in \Theta^*}$ where Θ^* satisfies (2.1b), (2.1c) and some additional properties.

A potentially awkward feature of Θ and Θ^* as so far described is that there may be many different parameter descriptions of the same probability distribution. [e.g., if $\mathcal{X} = [0, 1]$, $\nu =$ Lebesgue measure on $\mathcal{X}$, $f(x, \theta_0) \equiv 1$, $f(x, \theta_1) = 3x^2$, $f(x, \theta_2) = 4x^3$ then $(\frac{1}{2}, \theta_0, \theta_1)$ and $(\frac{1}{3}, \theta_0, \theta_2)$ both describe the density $2x$.] To overcome this notational difficulty we make the following convention, which is to apply to Θ and Θ^*, *as well as all other* parameter spaces defined in this paper:

CONVENTION. *All parameter descriptions of the same probability distribution are identified together.*

Thus we will write $(\xi, \theta_0, \theta_1) = (\xi', \theta_0', \theta_1')$ if and only if $f(\cdot, (\xi, \theta_0, \theta_1)) = f(\cdot, (\xi', \theta_0', \theta_1'))$ a.e. (ν). In the same spirit the statement $(\xi, \theta_0, \theta_1) \in \Theta$ means that there is a $\theta \in \Theta$ such that $f(\cdot, \theta) = f(\cdot, (\xi, \theta_0, \theta_1))$ a.e. (ν). (Note that this convention makes Assumption (2.1c) formally unnecessary, but is not implied by that assumption.) Let $\Theta_0^c, \Theta_1^c, \Theta^c$ denote the closure of $\Theta_0, \Theta_1, \Theta$, in Θ^*, respectively. Let $\Theta_0^* = \Theta_0^c$.

The assumptions we need to make are somewhat more restrictive but of a similar nature to those in Bahadur (1966). That paper contains certain comments which we will not reproduce here about such assumptions. In the main we have chosen the form of the assumptions below on the basis of notational and conceptual simplicity, rather than to give our results the greatest possible generality. However, it is clear that assumptions as weak as those in Bahadur (1966) will not suffice for our purposes. See Section 7 for one comment on this matter.

If Θ^* is compact *define* $\overline{\Theta} = \Theta^* \cup \{\infty\}$ where ∞ is an isolated point of $\overline{\Theta}$. If Θ^* is not compact *define* $\overline{\Theta} = \Theta^* \cup \{\infty\}$ the one-point compactification of Θ^*. (In the case where Θ^* is compact we could just as well have let $\overline{\Theta} = \Theta^*$ but we have pursued the above course in order not to have to separate arguments which follow into two cases depending on whether $\infty \in \Theta$ or not.) We *assume* (for convenience of symbolic presentation and without loss of generality) that the metric ρ on Θ^* has been chosen so that $\lim_{\theta \to \infty} \rho(\theta, \theta') = \infty$ for all $\theta' \in \Theta^*$.

Define $f(x, \infty) = 0$. Note $f(x, \infty)$ is not a probability density. Define $\overline{\Theta}_0$ as the closure of Θ_0^* in $\overline{\Theta}$. Thus $\overline{\Theta}_0 = \Theta_0^*$ or $\Theta_0^* \cup \{\infty\}$. For $d > 0$ and $\theta \neq \infty$ define

$$g(x, \theta, d(= \sup \{f(x, \theta') : \theta' \in \Theta^*, \rho(\theta, \theta') < d\}.$$

Fix a point ψ_0, say, in Θ^* and define

$$g(x, \infty, d) = \sup \{f(x, \theta') : \rho(\psi_0, \theta') > d^{-1}\}.$$

(In the sequel we will sometimes refer to the set $\{\theta : \rho(\psi_0, \theta) > d^{-1}\}$ as the neighborhood of infinity of "radius" d.)

Let $\mathscr{B}_\Theta$ denote the Borel field on Θ^*. Finally, let $\mathscr{M}$ be the topological space of a.e. (ν) equivalence classes of $\mathscr{B}$ measurable functions on $\mathscr{X}$ with the topology of convergence in measure (ν) *on all sets of finite* ν *measure.* (Let η be any finite measure equivalent to ν. Then the above topology is equivalent to the topology of convergence in measure (η).)

ASSUMPTION 1. *There is a locally compact second countable metric space* Θ^* *with metric* ρ *satisfying* (2.1) *and the following conditions.*

(a) *There exists a* $d' > 0$ *such that for* $0 < d < d'$ $g(\cdot, \cdot, d)$ is $\mathscr{B} \times \mathscr{B}_\Theta$ *measurable. Assume also* $f(\cdot, \cdot)$ *is* $\mathscr{B} \times \mathscr{B}_\Theta$ *measurable.*

(b) *The* 1–1 *map* $m: \overline{\Theta} \to f(\cdot, \theta)$ *is a homeomorphism of* $\overline{\Theta}$ *onto* $m(\overline{\Theta}) \subset \mathscr{M}$, *such that* $\theta_i \to \theta$ *if and only if* $f(\cdot, \theta_i) \to f(\cdot, \theta)$ a.e. (ν).

(c) *For all* $\theta \in \overline{\Theta}$, $g(\cdot, \theta, d) \searrow f(\cdot, \theta)$ a.e. (ν) *as* $d \to 0$.

We note that there is a crucial interrelationship between Assumption 1b and the construction of $\overline{\Theta}$ as the one-point compactification of $\overline{\Theta}^*$. To make this clear we begin with an example:

Let $\mathcal{X} = (-\infty, \infty)$, ν = Lebesgue measure, and let $N(\mu, \sigma^2)$, $\sigma^2 > 0$, denote the normal density with mean μ and variance σ^2. Let $N(\mu, 0)$ denote the probability distribution which gives mass one to the point μ. Consider the problem of testing $\Theta_0 = \{0 < \sigma^2 \leqq 1\}$ versus the alternative $\Theta_1 = \{1 < \sigma^2 < \infty\}$. It can be shown after some computation that Assumption 1b is satisfied here if the metric on Θ^* is chosen to be equivalent to the (weak) topology generated by neighborhoods of the form

$$N_{g,h,\varepsilon} = \{f(\cdot, \theta): \theta \in \Theta^* \int |f(x, \theta) - g(x)| h(x)dx < \varepsilon\}$$

where $\varepsilon > 0$, h is an arbitrary continuous function with *compact support*, and $g(\cdot) = f(\cdot, \theta')$, $\theta \in \Theta^*$. (This is the "natural" metric for Θ^* in such situations.) (If Θ_0 is further restricted by $\mu \in K$ where K is compact then all the other assumptions of this section are also satified. With Θ_0 as in the example above it turns out that Assumption 2b is not satisfied. Nevertheless the main conclusions of Theorems 1–3 are satisfied, but a few special minor changes are required in the proofs of these theorems to verify that fact.) On the other hand suppose Θ_0 consists of all densities of the form $\alpha N(\mu_1, \sigma_1^2) + (1-\alpha)N(\mu_1, \sigma_2^2)$ $0 < \sigma_i^2 \leqq 1$, $0 \leqq \alpha \leqq 1$, and Θ_1 is as before. The Assumption 1b is not satisfied because, for example, the sequence of densities described by $\alpha_i = \frac{1}{2}$, $\mu_{1i} = 0$, $\sigma_{1i} = 1$, $\sigma_{2i} = 1/i$ tends (a.e.) to the limit $N(0, 1)/2$ which is not in Θ^* and is also not identically 0.

We note that the "natural" compactification in the above situations is not the one-point compactification we have used but rather contains $\{N(\mu, \sigma^2): 0 \leqq \sigma^2 < \infty\}$ in the first situation above and contains $\{\alpha N(\mu_1, \sigma_1^2) + (1-\alpha)N(\mu_1, \sigma_2^2): 0 \leqq \alpha \leqq 1, 0 \leqq \sigma_i < \infty\}$ in the second situation. From this point of view the two situations may not seem so different, and also the use of a one-point compactification may seem artificial. The second situation appears to us to lead to complications similar to those encountered in the (simpler) example in Section 7, precisely on account of the appearance of $N(\mu, 0)/2 \neq 0$, etc., in the a.e. pointwise limit closure of Θ_0^*. For this reason we are unable to obtain general results of the type of Theorem 1 for such problems.

ASSUMPTION 2. (a) *For each $\theta \in \Theta^*$ there is a $d_1 = d_1(\theta) > 0$ such that*

$$\int g(x, \theta, d_1)\nu(dx) < \infty, \tag{2.2}$$

and

(b) *For each $\theta \in \Theta^*$ there is a $d_2 = d_2(\theta) > 0$ such that*

$$\int g(x, \infty, d_2)f(x, \theta)\nu(dx) < \infty. \tag{2.3}$$

Assumption 2 is sufficient (but much more than necessary) to guarantee that given τ, $0 < \tau < 1$, $\theta \in \Theta^*$, and $\varepsilon > 0$ there exists a $d > 0$ such that

$$\int [g(x, \theta, d)/f(x, \theta_0)]^\tau F_{\theta_0}(dx) < 1 + \varepsilon$$

for all $\theta_0 \in \Theta_0$, which is Assumption 6 of Bahadur (1966); and also given $\theta \in \Theta_1{}^*$ and $\theta_0 \in \Theta_0$ there exists a $d > 0$ such that

$$\int \log^+[g(x, \theta_0, d)/f(x, \theta)]F_\theta(dx) < \infty,$$

which is Assumption 5 of Bahadur's paper.

An important element in all our considerations is the information number

$$K(\theta,\theta') = \int \log(f(x,\theta)/f(x,\theta'))f(x,\theta)\nu(dx)$$

defined $(0 \leqq K \leqq \infty)$ for all $\theta, \theta' \in \Theta^*$. Extend this definition by defining $K(\infty, \theta) = \infty$ and $K(\theta, \infty) = \infty$ for all $\theta \in \Theta^*$.

ASSUMPTION 3. *For all* θ, $\theta' \in \Theta^*$ *such that* $K(\theta, \theta') < \infty$, *there exists a* $d = d(\theta, \theta') > 0$ *such that*

$$\int \log\left(\frac{g(x, \theta, d)}{f(x, \theta')}\right) g(x, \theta, d)\nu(dx) < \infty.$$

Define

$$J(\theta') = \inf_{\theta \in \Theta_0} K(\theta', \theta).$$

Except for some results in Section 7, we will generally have to assume

ASSUMPTION 4. *The probability distributions* $\{F_\theta\}_{\theta \in \Theta_0{}^*}$ *are mutually absolutely continuous and for* $\theta' \in \Theta_1{}^*$, $\theta \in \Theta_0{}^*$, $F_{\theta'}$ *is absolutely continuous with respect to* F_θ. *Without any further loss of generality we then assume* ν *is chosen so that* $f(\cdot, \theta) > 0$ a.e. (ν) *for all* $\theta \in \Theta_0{}^*$.

Except where otherwise specifically noted we assume throughout this paper that Assumptions 1–4 are satisfied.

Let $(x_1, x_2, \cdots)$ be a sequence of independent identically distributed observations on $\mathcal{X}$, each having distribution F_θ. Let T_n be a test (possibly randomized), of Θ_0 versus Θ_1 which depends only on $x_1, \cdots, x_n = x^{(n)}$ and denote by $T = T_1, T_2, \cdots$ a sequence of such tests. Let

$$\alpha_n{}^T(\theta) = \Pr_\theta \{\text{the test } \quad T_n \quad \text{rejects } \quad \Theta_0\}$$

$$\alpha_n{}^T = \sup_{\theta_0 \in \Theta_0} \alpha_n{}^T(\theta_0)$$

and for $\theta \in \Theta_1$

$$\beta_n{}^T(\theta) = \Pr_\theta \{\text{the test } \quad T_n \quad \text{accepts } \quad \Theta_0\}$$

$\alpha_n{}^T$ and $\beta_n{}^T(\theta)$ are the significance level and probability of type II error at $\theta \in \Theta_1$ of the test T_n, respectively.

If A and B are any two sets of parameter points, define the associated likelihood ratio statistic by

$$l_n(A, B) = \frac{\sup \{\prod_{i=1}^n f(x_i, \theta): \theta \in B\}}{\sup \{\prod_{i=1}^n f(x_i, \theta): \theta \in A\}}.$$

Define

$$\lambda_n(A, B) = \frac{1}{n} \log l_n(A, B).$$

For convenience we sometimes write $\lambda_n(\theta_0, \theta_1)$ instead of $\lambda_n(\{\theta_0\}, \{\theta_1\})$, etc. (Both l_n and λ_n are of course functions of $x_1, \cdots, x_n$, but we do not explicitly display this in the notation.) Generally we will take A to be the null hypothesis and B the alternative and we will construct tests of the form: Reject Θ_0 with probability 1 if $\lambda_n > C_n$. The constants C_n are called *critical constants.* Temporarily denote the level of such a test by $\alpha_n{}^C$. The statement $\alpha_n{}^C \overset{*}{=} a$ means $C = \inf\{c: \alpha_n{}^c \leqq a\}$. Note that $\alpha_n{}^C \overset{*}{=} a$ implies $\alpha_n{}^C \leqq a$.

When no mention in the following is made of A and B it is understood that $A = \Theta_0$ and $B = \Theta_1$. Thus $\lambda_n = \lambda_n(\Theta_0, \Theta_1)$.
Also define

$$\lambda_n{}^* = \lambda_n(\Theta_0{}^*, \Theta_1{}^*).$$

It should be noted that in at least one respect the above regularity conditions are not as hard to verify as they may at first appear. In many practical situations, Θ is naturally taken as a locally compact metric space and Θ_0 is closed in Θ. If that is the case then one can always take the family of densities $\{f(\cdot, \theta)\}_{\theta \in \Theta^*}$ to be exactly $\{f(\cdot, \theta)\}_{\theta \in \Theta} \cup \{f(\cdot, (\xi, \theta, \theta_1))\}_{0 \leqq \xi \leqq 1, \theta \in \Theta_0, \theta_1 \in \Theta_1}$ and Θ^* is naturally isomorphic to a quotient space of $\Theta \cup ([0, 1] \times \Theta_0 \times \Theta_1)$. (The "quotient" results from the fact that one must identify all points θ and/or $(\xi, \theta_0', \theta_1)$ which represent the same density. After the identification, $\Theta_0 = \Theta_0{}^*$.) Any of Assumptions 1, 2, 3, and 4 which are satisfied by Θ and the family of densities $\{(\cdot, \theta)\}_{\theta \in \Theta}$ will also be satisfied by the often much larger family $\{f(\cdot, \theta)\}_{\theta \in \Theta^*}$ as defined above. We leave the proof to the reader.

3. Preparatory lemmas. In this section we prove several lemmas which are needed in the proof of Theorem 1 and elsewhere in the paper. The following properties of K and J which are implied by Assumptions 1 and 3 are used at several points in the proof of Theorem 1.

LEMMA 1. *Suppose Assumptions* 1 *to* 3 *are satisfied. Then* $K(\cdot, \cdot)$ *is a lower-semi-continuous extended real-valued function on* $\Theta^* \times \overline{\Theta}$. *For each* $\theta' \in \Theta^*$, $K(\cdot, \theta')$ *is a continuous extended real-valued function on* $\overline{\Theta}$. $J(\cdot)$ *is a continuous function on* Θ^*.

PROOF. Let $\log^+ a = \max(0, \log a)$ and $\log^- a = \min(0, \log a) \leqq 0$, etc.

$$\begin{aligned}
&\int \log^- \left(\frac{f(x, \theta)}{f(x, \theta')}\right) f(x, \theta) \nu(dx) \\
&\quad = -\int \log^+ \left(\frac{f(x, \theta')}{f(x, \theta)}\right) f(x, \theta) \nu(dx) \qquad (3.1) \\
&\quad \geqq -\int \left(\frac{f(x, \theta')}{f(x, \theta)} - 1\right)^+ f(x, \theta) \nu(dx) \\
&\quad \geqq -\int f(x, \theta') \nu(dx) \geqq -1 \qquad (= -1 \quad \text{if} \quad \theta' \in \Theta^*).
\end{aligned}$$

Hence by Fatou's Lemma and Assumption 1b if $\psi_i \to \psi \in \Theta^*$, and $\theta_i \to \theta \in \overline{\Theta}$.

$$\liminf_{i\to\infty} \int \log\left(\frac{f(x,\psi_i)}{f(x,\theta_i)}\right) f(x,\psi_i)\nu(dx) \geqq \int \log\left(\frac{f(x,\psi)}{f(x,\theta)}\right) f(x,\psi)\nu(dx).$$

Hence $K(\cdot,\cdot)$ is a lower-semicontinuous function on $\Theta^* \times \overline{\Theta}$.

Suppose $\theta' \in \Theta^*$ and $\theta_i \to \theta \in \Theta^*$. For any fixed, small, $d > 0$ and all i sufficiently large

$$\log^+\left(\frac{f(x,\theta_i)}{f(x,\theta')}\right) f(x,\theta_i) \leqq \log^+\left(\frac{g(x,\theta,d)}{f(x,\theta')}\right) g(x,\theta,d).$$

Hence Assumptions 3 and 1b imply that $K(\theta_i,\theta') \to K(\theta,\theta')$.

Suppose $\theta_i \to \infty$. Let $S_{ij} \subset \mathcal{X}$ be the set such that $jf(x,\theta') < f(x,\theta_i)$. Assumption 1b guarantees that for all j $\lim_{i\to\infty} \int S_{ij} f(x,\theta_i)\nu(dx) = 1$.

$$\begin{aligned}
&\int \log^+\left(\frac{f(x,\theta_i)}{f(x,\theta')}\right) f(x,\theta_i)\nu(dx) \\
&\quad \geqq \int_{S_{ij}} \log\frac{f(x,\theta_i)}{f(x,\theta')} f(x,\theta_i)\nu(dx) \\
&\quad \geqq (\log j)\int_{S_{ij}} f(x,\theta_i)\nu(dx) \\
&\quad \to \log j.
\end{aligned}$$

Observe from (3.1) that

$$\int \log^-\left(\frac{f(x,\theta_i)}{f(x,\theta')}\right) f(x,\theta_i)\nu(dx) \geqq -1.$$

Since j in the preceding expression is arbitrary we thus have

$$K(\theta_i,\theta') = \int \log\left(\frac{f(x,\theta_i)}{f(x,\theta')}\right) f(x,\theta_i)\nu(dx) \to \infty \qquad \text{as } i \to \infty.$$

We have thus shown that $K(\cdot,\theta')$ is continuous.

Use the previously established lower-semicontinuity of $K(\theta,\cdot)$ on $\overline{\Theta}$. Since $\overline{\Theta}_0$ is closed, for each $\theta \in \Theta^*$ there is a $\psi = \psi(\theta) \in \overline{\Theta}_0$ such that

$$J(\theta) = K(\theta,\psi(\theta)). \tag{3.2}$$

In fact, there is no loss of generality in assuming $\psi \in \Theta_0{}^*$. Let $\theta_i \to \theta \in \Theta^*$. Since $\overline{\Theta}$ is compact there is a subsequence, say $\{\theta_{i'}\}$ such that $\psi(\theta_{i'}) \to \psi_0 \in \overline{\Theta}_0$ and $\lim \inf_{i\to\infty} K(\theta_i,\psi(\theta_i)) = \liminf_{i'\to\infty} K(\theta_{i'},\psi(\theta_{i'}))$. From the lower-semicontinuity of $K(\cdot,\cdot)$ we have

$$\begin{aligned}
\liminf_{i\to\infty} J(\theta_i) &= \liminf_{i\to\infty} K(\theta_i,\psi(\theta_i)) \\
&= \liminf_{i'\to\infty} K(\theta_{i'},\psi(\theta_{i'})) \geqq K(\theta,\psi_0).
\end{aligned}$$

On the other hand $J(\theta_i) \leqq K(\theta_i, \psi_0)$ and $K(\cdot, \psi_0)$ is a continuous function. (If $\psi_0 = \infty$, $K(\cdot, \psi_0) \equiv \infty$.) Hence

$$\limsup_{i\to\infty} J(\theta_i) \leqq \limsup_{i\to\infty} K(\theta_i, \psi_0) = K(\theta, \psi_0).$$

This proves $J(\cdot)$ is a continuous function on Θ^*.

The proof of the lemma is complete.

The following is another very useful result which is easily proved.

LEMMA 2. *If $\theta_0, \theta_1 \in \Theta$ and $f(\cdot, \theta_0)$ and $f(\cdot, \theta_1)$ are not mutually singular densities then $c^{-1}(\xi, \theta_0, \theta_1)$ is a continuous strictly convex function of ξ on $0 \leqq \xi \leqq 1$. For $0 \leqq \xi \leqq 1$, $1 \leqq c(\xi, \theta_0, \theta_1)$. If $f(\cdot, \theta_0)$ and $f(\cdot, \theta_1)$ are mutually absolutely continuous then $c(0, \theta_0, \theta_1) = 1 = c(1, \theta_0, \theta_1)$. For $0 < \xi < 1$ the defining expression for c may be differentiated under the integral sign an arbitrary number of times.*

PROOF. Since $0 \leqq \xi \leqq 1$ implies $a^{1-\xi}b^{\xi} \leqq \max(a, b)$ we have

$$(3.3)\quad 2 \geqq c^{-1}(\xi, \theta_0, \theta_1) = \int f^{1-\xi}(x, \theta_0) f^{\xi}(x, \theta_1) \nu(dx) = \int \exp\left(\xi \log \frac{f(x, \theta_1)}{f(x, \theta_0)}\right) f(x, \theta_0) \nu(dx).$$

Thus c^{-1} is a moment generating function which is finite for $0 \leqq \xi \leqq 1$. The assertions in the lemma concerning continuity and differentiability follow directly from this fact. The strict convexity of c also follows from this fact since a direct computation yields $(d^2/d\xi^2)c^{-1}(\xi, \theta_0, \theta_1) > 0$. The values for $c(0, \theta_0, \theta_1)$ and $c(1, \theta_0, \theta_1)$ may be computed directly from the definition. The "proof" is complete.

We will need some other properties of K which follow directly from the above two lemmas. In the discussion which follows we will use Assumptions 1–4 freely and we also assume $\theta_1 \in \Theta_1$ and $f(\cdot, \theta_1) > 0$ a.e. (ν). This latter assumption will appear again in Theorems 1–3. We assert that for any neighborhood N of infinity $K((\gamma, \theta, \theta_1), \theta)$ is lower-semicontinuous on $\{(\gamma, \theta): \theta \in \Theta_0^*, 0 \leqq \gamma \leqq 1, \theta \notin N \text{ or } (\gamma, \theta, \theta_1) \notin N\}$.

In particular it follows that $K((\gamma, \theta, \theta_1), \theta)$ is a jointly lower-semicontinuous function of γ, θ for $0 \leqq \gamma \leqq 1$, $\theta \in \Theta_0^*$. For any $\theta_2 \in \bar{\Theta}$, $K((\gamma, \theta, \theta_1), \theta_2)$ is a continuous function of γ, θ for $0 \leqq \gamma \leqq 1$, $\theta \in \Theta_0^*$. To verify these assertions begin by defining the map

$$(3.4)\qquad m': (\gamma, \theta, \theta_1) \to f(\cdot, (\gamma, \theta, \theta_1)).$$

With m as defined in Assumption 1b it follows from Assumption 4 that $m^{-1} \circ m'$ is a continuous map of $\{\gamma, \theta; 0 \leqq \gamma \leqq 1, \theta \in \Theta_0^*\}$ into $\Theta^*(m^{-1} \circ m'(\gamma, \theta) = (\gamma, \theta, \theta_1) \in \Theta^*)$. The assertions then follow directly from Lemma 1.

Utilizing the differentiability properties in Lemma 2 and the definitions of K and c, etc., we have

$$K((\gamma, \theta, \theta_1), \theta) = \log(c(\gamma, \theta, \theta_1))$$

$$+c(\gamma,\theta,\theta_1)\gamma\int\log\left(\frac{f(x,\theta_1)}{f(x,\theta)}\right)f^{1-\gamma}(x,\theta)f^{\gamma}(x,\theta_1)\nu(dx)\cdot \tag{3.5}$$

$$=\log c(\gamma,\theta,\theta_1)-\gamma\frac{d}{d\gamma}\left(\log c(\gamma,\theta,\theta_1)\right)$$

and for $0<\gamma<1$, $K((\gamma,\theta,\theta_1),\theta)<\infty$. Differentiating (3.5) yields for $0<\gamma<1$

$$\frac{d}{d\gamma}K((\gamma,\theta,\theta_1),\theta)=-\gamma\frac{d^2}{d\gamma^2}\log c(\gamma,\theta,\theta_1)$$

$$\begin{aligned}&=\gamma[c(\gamma,\theta,\theta_1)\int\log^2\left(\frac{f(x,\theta_1)}{f(x,\theta)}\right)f^{1-\gamma}(x,\theta)f^{\gamma}(x,\theta_1)\nu(dx)\\&\quad-(c(\gamma,\theta,\theta_1)\int\log\left(\frac{f(x,\theta_1)}{f(x,\theta)}\right)f^{1-\gamma}(x,\theta)f^{\gamma}(x,\theta_1)\\&\quad\times\nu(dx))^2]\\&=\gamma\,\mathrm{Var}_{(\gamma,\theta,\theta_1)}\left(\log\frac{f(x,\theta_1)}{f(x,\theta)}\right)\\&>0.\end{aligned} \tag{3.6}$$

We have used Assumption 4 and $f(\cdot,\theta_1)>0$ a.e. (ν) to guarantee that the variance in (3.6) is not zero. Note also from (3.6) that $(d/d\gamma)K((\gamma,\theta,\theta_1),\theta)<\infty$ for $0<\gamma<1$.

We derive from the above facts that

$$\lim_{\gamma\downarrow 0}K((\gamma,\theta,\theta_1),\theta)=0 \tag{3.7}$$

$$\lim_{\gamma\uparrow 1}K((\gamma,\theta,\theta_1),\theta)=K(\theta_1,\theta)$$

and

$$\lim_{\gamma\downarrow 0}K((\gamma,\theta,\theta_1),\theta_1)=K(\theta,\theta_1) \tag{3.8}$$

$$\lim_{\gamma\uparrow 1}K((\gamma,\theta,\theta_1),\theta_1)=0.$$

There are various other closely related facts about K, c, f, etc., which we will use without further proof when needed.

Before proceeding further we need some lemmas concerning probabilities of large deviations of likelihood ratio statistics. Such results are of course the foundation on which theorems on non-local asymptotic optimality are based. Much of what follows is already known and the rest is a simple generalization of proofs in the literature; see for example Chernoff (1952), Bahadur (1966), Rao (1962). For the sake of clarity we begin by stating a result of Chernoff (1952) in the form we will use it in this paper. We note that this result can be somewhat strengthened but we have been unable to make use of stronger versions than the following.

LEMMA 3. *Let $z_1, z_2, \cdots$ be the values of a sequence of independent identically distributed random variables. Let $k_n \to k$. Then*

$$\limsup_{n\to\infty} n^{-1} \log \Pr\{n^{-1}\textstyle\sum_{i=1}^{n} z_i > k_n\} \leqq \inf_{t>0} \log E(e^{tz-tk}). \tag{3.9}$$

If the infimum on the right of (3.9) *is attained for some $t > 0$ on the interior of the region of convergence of the expectation in* (3.9) *then*

$$\lim_{n\to\infty} n^{-1} \log \Pr\{n^{-1}\textstyle\sum_{i=1}^{n} z_i > k_n\} = \inf_{t>0} \log E(e^{tz-tk}). \tag{3.10}$$

We will utilize this result, for example, when

$$z_i = \log \frac{f(x_i, \theta_3)}{f(x_i, \theta_2)}$$

We then have

$$\limsup_{n\to\infty} n^{-1} \log \Pr_\theta\{\lambda_n(\theta_2, \theta_3) > k_n\} \leqq \inf_{t>0} \log E_\theta\left(e^{-tk}\left(\frac{f(x, \theta_3)}{f(x, \theta_2)}\right)^t\right), \tag{3.11}$$

with the limit existing on the left and equality holding under the specified conditions.

The main result we need is the following lemma. The lemma can clearly be further generalized, but we do not need more generality for our purposes.

We assume throughout the remainder of this section that Assumptions 1–4 are satisfied.

LEMMA 4. *Let h_n $n = 1, 2, \cdots$ be a sequence of continuous real-valued functions on $\overline{\Theta}_0$ such that $h_n(\theta) \to h(\theta) < \infty$ uniformly for $\theta \in \overline{\Theta}_0$. Suppose h and $\theta_1 \in \Theta_1$ satisfy $f(\cdot, \theta_1) > 0$ a.e. (ν) and $-K(\theta, \theta_1) < h(\theta) < K(\theta_1, \theta)$ for all $\theta \in \overline{\Theta}_0$.*

Then

$$\begin{aligned} &\lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\inf_{\theta\in\bar{\Theta}_0}(\lambda_n(\theta, \theta_1) - h_n(\theta)) \leqq 0\} \\ &= \sup_{\theta\in\Theta_0} \lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\lambda_n(\theta, \theta_1) - h_n(\theta) \leqq 0\} \\ &= \sup_{\theta\in\Theta_0} \inf_{t>0} \log E_{\theta_1}\left(e^{th(\theta)}\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^t\right). \end{aligned} \tag{3.12}$$

PROOF. Fix $\varepsilon > 0$. Using Assumption 2 and Assumption 1c, for $0 < t < 1$ and $\theta \in \overline{\Theta}_0$

$$E_{\theta_1}\left(\left(\frac{g(x, \theta, d)}{f(x, \theta_1)}\right)^t\right) \searrow E_{\theta_1}\left(\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^t\right) \leqq 1 \qquad \text{as } d \searrow 0. \tag{3.13}$$

Let $h(\theta, d) = \sup\{h(\theta'): \delta(\theta', \theta) < d\}$. Define $h(\infty, d)$ analogously. Then for each $\theta \in \overline{\Theta}_0$ there is a $d = d(\theta) > 0$ such that

$$\begin{aligned} &\inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta,d)}\left(\frac{g(x, \theta, d)}{f(x, \theta_1)}\right)^t\right) \\ &\qquad \leqq (1+\varepsilon) \inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta)}\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^t\right). \end{aligned} \tag{3.14}$$

Let $N_1, N_2, \cdots, N_k$ be a finite covering of $\bar{\Theta}_0$ by neighborhoods of the points $\theta_1', \theta_2', \cdots, \theta_k'$ having radii $d_1 = d(\theta_1')$, $d_2 = d(\theta_2')$, $\cdots$, $d_k = d(\theta_k')$, respectively. Then, using Lemma 3 since $h_n(\theta_i', d_i) \to h(\theta_i', d_i)$ we have

$$\begin{aligned}
&\lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\inf_{\theta\in\Theta_0}(\lambda_n(\theta,\theta_1)-h_n(\theta)) \leqq 0\} \\
&\leqq \lim_{n\to\infty} n^{-1} \log \sum_{i=1}^{k} \Pr_{\theta_1}\left\{n^{-1}\left(\sum_{j=1}^{n} \log\frac{f(x_j,\theta_1)}{g(x_j,\theta_i',d_i)}\right)-h_n(\theta_i',d_i) \leqq 0\right\} \qquad (3.15)\\
&= \lim_{n\to\infty} n^{-1} \log \sup_{1\leqq i\leqq k} \Pr_{\theta_1}\left\{n^{-1}\sum_{j=1}^{n}\left(\log\frac{f(x_j,\theta_1)}{g(x_j,\theta_i',d_i)}\right)\right. \\
&\qquad \left.-h_n(\theta_i',d_i) \leqq 0\right\} \\
&\leqq \sup_{1\leqq i\leqq k} \inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta_i',d_i)}\left(\frac{g(x,\theta_i',d_i)}{f(x,\theta_1)}\right)^t\right).
\end{aligned}$$

Applying (3.14) and (3.15) we find

$$\begin{aligned}
&\lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\inf_{\theta\in\bar{\Theta}_0}(\lambda_n(\theta,\theta_1)-h_n(\theta)) \leqq 0\} \\
&\leqq \sup_{\theta\in\bar{\Theta}_0} \inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta,d(\theta))}\left(\frac{g(x,\theta,d(\theta))}{f(x,\theta_1)}\right)^t\right) \qquad (3.16)\\
&\leqq (1+\varepsilon) \sup_{\theta\in\Theta_0^*} \inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta)}\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right)^t\right),
\end{aligned}$$

Let $\varepsilon \to 0$.

$$\begin{aligned}
&\lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\inf_{\theta\in\bar{\Theta}_0}(\lambda_n(\theta,\theta_1)-h_n(\theta)) \leqq 0\} \qquad (3.17)\\
&\qquad \leqq \sup_{\theta\in\Theta_0^*} \inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta)}\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right)^t\right)
\end{aligned}$$

Trivially,

$$\begin{aligned}
&\Pr_{\theta_1}\{\inf_{\theta\in\bar{\Theta}_0}(\lambda_n(\theta,\theta_1)-h_n(\theta)) \leqq 0\} \qquad (3.18)\\
&\qquad \geqq \sup_{\theta\in\bar{\Theta}_0} \Pr_{\theta_1}\{(\lambda_n(\theta,\theta_1)-h_n(\theta)) \leqq 0\}.
\end{aligned}$$

For this paragraph fix $\theta \in \Theta_0$ and let

$$e(t) = \log E_{\theta_1}\left(e^{th(\theta)}\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right)^t\right).$$

Note that $e(t)$ is well defined and convex on $0 \leqq t \leqq 1$. Since $e^{e(t)}$ is the well-defined moment generating function of $h(\theta)+\log(f(x,\theta)/f(x,\theta_1))$, $e(t)$ can be differentiated inside the integral for $0 < t < 1$ to give

$$\begin{aligned}
e'(t) = h(\theta)+&\left(\int \log\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right)\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right)^t f(x,\theta_1)\nu(dx)\right) \qquad (3.19)\\
&\left(\int\left(\frac{f(x,\theta)}{f(x,\theta_1)}\right)^t f(x,\theta_1)\nu(dx)\right)^{-1}
\end{aligned}$$

Since by Assumption 4 both $f(\cdot, \theta) > 0$ a.e. (ν) and $f(\cdot, \theta_1) > 0$ a.e. (ν) the second factor in the product on the right of (3.19) is continuous for $0 \leqq t \leqq 1$, and tends to one as $t \searrow 0$ or $t \nearrow 1$. (Note that this factor is exactly $c(1-t, \theta, \theta_1)$.) Since $(f(x, \theta)/f(x, \theta_1))^t$ is continuous increasing (decreasing, resp.) in t for $0 \leqq t \leqq 1$ whenever $\log (f(x, \theta)/f(x, \theta_1))$ is positive (negative), it is true in similar fashion that the first factor is also continuous on [0, 1]. Using the definition of K to evaluate this factor at $t = 0$ and $t = 1$ we have

$$\lim_{t \searrow 0} e'(t) = h(\theta) - K(\theta_1, \theta) < 0$$
$$\lim_{t \nearrow 1} e'(t) = h(\theta) + K(\theta, \theta_1) > 0.$$

It follows from the convexity of $e(t)$ that $\inf_{t>0} e(t)$ occurs at some point $t' \in (0, 1)$.

Hence, again using Lemma 3

$$(3.20)\quad \sup_{\theta \in \overline{\Theta}_0} \lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\lambda_n(\theta, \theta_1) - h_n(\theta) \leqq 0\} = \sup_{\theta \in \overline{\Theta}_0} \inf_{0<t<1} \log E_{\theta_1}\left(e^{th(\theta)}\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^t\right).$$

Combining (3.17), (3.18) and (3.20) proves that (3.12) is valid. This completes the proof of the lemma.

Note that, since $f(x, \cdot)$ is continuous a.e. (ν) on $\overline{\Theta}$, the conclusion of the lemma is unchanged if $\inf_{\theta \in \overline{\Theta}_0}$ is replaced by $\inf_{\theta \in \Theta_0}$.

Note also that if $h_n(\theta) = k_n$ and $k_n \to k$ with $\sup_{\theta \in \overline{\Theta}_0}(-K(\theta, \theta_1)) < k < J(\theta_1)$ the conclusion of the lemma can be specialized to

$$(3.21)\quad \lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\lambda_n(\Theta_0, \theta_1) \leqq k_n\} = \sup_{\theta \in \Theta_0} \inf_{0<t<1} \log E_{\theta_1}\left(e^{tk}\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^t\right).$$

Reversing the roles of Θ_0 and Θ_1 in Lemma 2 gives a formally different result. Minor modifications are needed in the regularity conditions. Omitting these, the analog of (3.21), for example, is

$$(3.22)\quad \lim_{n\to\infty} n^{-1} \log \Pr_{\theta_0}\{\lambda_n(\theta_0, \Theta_1) \geqq k_n\} = \sup_{\theta \in \Theta_1} \inf_{0<t<1} \log E_{\theta_0}\left(e^{-tk}\left(\frac{f(x, \theta)}{f(x, \theta_0)}\right)^t\right)$$

where $(\sup_{\theta \in \Theta_1} K(\theta_0, \theta)) < k < \inf_{\theta \in \Theta_1} K(\theta, \theta_0)$. The condition $k < \inf_{\theta \in \Theta_1} K(\theta, \theta_0)$ would normally be obnoxious in usual applications of (3.11); but other, nicer, conditions can be substituted if $\inf_{0<t<1}$ is replaced by $\inf_{t>0}$ and if this infimum is attained in the interior of the region where $E_{\theta_0}(e^{-tk}(f(x, \theta)/f(x, \theta_0))^t)$ is finite. In any case, without the necessity of making any such assumptions we certainly have

LEMMA 5. *Let* $\theta_2 \in \Theta_1^*$ *and suppose* $f(\cdot, \theta_2) > 0$ a.e. (ν). *Let* $\theta_0 \in \Theta_0^*$. *Let* $k_n \to k > 0$. Then

$$(3.23)\quad \limsup_{n\to\infty} n^{-1} \log \Pr_{\theta_2}\{\lambda_n(\theta_0, \Theta_1) > k_n\} \leqq \sup_{\theta \in \Theta_1} \inf_{t>0} \log E_{\theta_2}\left\{e^{-tk}\left(\frac{f(x, \theta)}{f(x, \theta_0)}\right)^t\right\}.$$

PROOF. The proof of this result is similar to the verification of the validity of (3.17) in the proof of Lemma 4. We omit the details.

4. Statement and proof of the fundamental theorem. This section contains the statement and proof of the theorem which provides the bulk of the result claimed in the Introduction. It is extended and amplified in the following sections of the paper.

THEOREM 1. *Suppose Assumptions* 1 *to* 4 *are satisfied. Let* $\{\alpha_n : n = 1, 2, \cdots\}$ *satisfy* $\lim_{n\to\infty} n^{-1} \log \alpha_n = -a < 0$. *Let* $\mathscr{S} = \{S\}$ *be the set of all sequences of tests of* Θ_0 *versus* Θ_1 *satisfying* $\alpha_n{}^S \leqq \alpha_n$. *Define* $\beta_n(\theta) = \inf_{S\in\mathscr{S}} \beta_n{}^S(\theta)$. *Let* $\theta_1 \in \Theta_1$ *satisfy* $J(\theta_1) > a$, *and* $f(\cdot, \theta_1) > 0$ a.e. (ν).

(a) *Then,* $\lim n^{-1} \log \beta_n(\theta_1) = -\beta(\theta_1)$ *exists and* $\infty > \beta(\theta_1) > 0$. [See conclusion (b) for an explicit expression for β.]

For $\theta \in \Theta_0{}^*$ *define* $\zeta(\theta)$ *by* $K((\zeta(\theta), \theta, \theta_1), \theta) = a$.

(b) *Then,* $\zeta(\theta)$ *is uniquely determined, and*

$$\beta(\theta_1) = \inf_{\theta\in\Theta_0{}^*} K((\zeta(\theta), \theta, \theta_1), \theta_1). \tag{4.1}$$

Let Q^* *be the sequence of tests based on* $\lambda_n{}^* = \lambda_n(\Theta_0{}^*, \Theta_1{}^*)$ *with critical constants* $C_n{}^*$ *defined by* $\alpha_n{}^{Q^*} \doteq \alpha_n$.

(c) *Then,* $C_n{}^* \to a$, *and*

(d) $\lim_{n\to\infty} n^{-1} \log \beta_n{}^{Q^*}(\theta_1) = -\beta(\theta_1)$.

Let Q *be the sequence of tests based on* λ_n *with critical constants* C_n *defined by* $\alpha_n{}^Q \doteq \alpha_n$.

(e) *If there is a* $\theta' \in \Theta_1$ *such that* $J(\theta') = a$ *and* $f(\cdot, \theta') > 0$ a.e. (ν), *then* $C_n \to a$.

(f) *If* $C_n \to a$, *a sufficient condition for*

$$\lim_{n\to\infty} n^{-1} \log \beta_n{}^Q(\theta_1) = -\beta(\theta_1) \tag{4.2}$$

is that there exist an $\varepsilon > 0$ *such that for each* $\theta \in \Theta_0$ *there exists a* $\gamma = \gamma(\theta)$, $0 < \gamma < 1$, *satisfying*

$$K((\gamma, \theta, \theta_1), \theta) \geqq a + (\varepsilon - \gamma)^+ \tag{4.3}$$

$$K((\gamma, \theta, \theta_1), \theta_1) \geqq \beta(\theta_1), \quad \text{and} \quad (\gamma, \theta, \theta_1) \in \Theta_1.$$

(g) *If* $C_n \to a$, *a necessary condition for* (4.2) *to hold is that for all* $\theta \in \Theta_0{}^*$ *satisfying*

$$K((\zeta(\theta), \theta, \theta_1), \theta_1) = \beta(\theta_1) \tag{4.4}$$

we have $(\zeta(\theta), \theta, \theta_1) \in \Theta_1{}^c$. *If this condition fails to hold then, in fact,*

$$\lim_{n\to\infty} \inf n^{-1} \log \beta_n{}^Q(\theta_1) > -\beta(\theta_1).$$

PROOF. We now proceed to establish part (e) of the theorem. The first goal is to prove equation (4.12) below.

Suppose $\theta' \in \Theta_1$ satisfies $J(\theta') = a$ and $f(\cdot, \theta') > 0$ a.e. (ν). Fix $\varepsilon > 0$. From Lemma 4,

$$\begin{aligned} &\lim_{n\to\infty} n^{-1} \log \Pr_{\theta'}\{\lambda_n(\Theta_0, \theta') \leqq a-\varepsilon\} \\ &= \sup_{\theta \in \Theta_0} \inf_{0<t<1} \log E_{\theta'}\left(e^{t(a-\varepsilon)} \left(\frac{f(x,\theta)}{f(x,\theta')}\right)^t \right) \\ &= \sup_{\theta \in \bar{\Theta}_0} \inf_{0<t<1} [t(a-\varepsilon) - \log c((1-t), \theta, \theta')]. \end{aligned} \tag{4.5}$$

From Assumptions 1c, 2b and $f(\cdot, \infty) \equiv 0$ it follows that there is a neighborhood N', say, of ∞ and an $\varepsilon_1' > 0$ such that $\inf_{0<t<1}[t(a-\varepsilon) - \log c((1-t), \theta, \theta')] < -\varepsilon_1' < 0$ for $\theta \in N'$.
From (3.5), for $0 < t < 1$,

$$\begin{aligned} &\frac{d}{dt}[t(a-\varepsilon) - \log c((1-t), \theta, \theta')] \\ &\qquad = \left[a - \varepsilon - \frac{K((1-t), \theta, \theta'), \theta) - \log c((1-t), \theta, \theta')}{1-t} \right]. \end{aligned} \tag{4.6}$$

From previously established properties of K and c it follows that the derivative in (4.6) is a continuous function of t, θ for $0 < t < 1$, $\theta \in \Theta_0{}^*$; and that for each $\theta \in \Theta_0{}^*$ there is a $\delta = \delta(\theta) > 0$ such that this derivative is negative for $0 < t < \delta$. It follows that there is a continuous function $\varepsilon_1(\cdot)$ on $\Theta_0{}^*$ such that $\inf_{0<t<1}[t(a-\varepsilon) - \log c((1-t), \theta, \theta')] < -\varepsilon_1(\theta) < 0$. It then follows from $\bar{\Theta}_0$ compact and the above that for some $\varepsilon_1 > 0$

$$\lim_{n\to\infty} n^{-1} \log \Pr_{\theta'}\{\lambda_n(\Theta_0, \theta') \leqq a-\varepsilon\} < -\varepsilon_1 < 0. \tag{4.7}$$

Temporarily denote the sequence of tests with rejection regions $\{\lambda_n(\Theta_0, \theta') > a-\varepsilon\}$ by U. Let $\psi \in \Theta_0$ satisfy $K(\theta', \psi) = J(\theta')$ (ψ exists: See proof of Lemma 1.) Consider the problem of finding the lowest significance level sequence of tests, say V, of $H_0: \psi$ versus $H_1: \theta'$ satisfying $\beta_n{}^V(\theta') \leqq \beta_n{}^U(\theta')$.

According to the Neyman–Pearson Lemma, V is of the form

$$\text{Reject} \quad H_0 \quad \text{if} \quad \lambda_n(\psi, \theta') > k_n.$$

(To be strictly correct, the possibility of randomization at k_n should be included in the above definition. However, it can easily be checked that the following reasoning is correct, even allowing for that possibility.) By definition k_n must satisfy

$$\limsup_{n\to\infty} n^{-1} \log \Pr_{\theta'}\{\lambda_n(\psi, \theta') \leqq k_n\} < -\varepsilon_1. \tag{4.8}$$

Suppose

$$\limsup_{n\to\infty} k_n \geqq a. \tag{4.9}$$

Then for any $\varepsilon > 0$

$$\limsup_{n\to\infty} n^{-1} \log \Pr_{\theta'}\{\lambda_n(\psi, \theta') \leqq k_n\}$$

$$\geqq \lim_{n\to\infty} n^{-1} \log \Pr_{\theta'}\{\lambda_n(\psi, \theta') \leqq a-\varepsilon\}$$

$$= \inf_{0<t<1} \log E_{\theta'}\left(e^{t(a-\varepsilon)}\left(\frac{f(x,\psi)}{f(x,\theta')}\right)^t\right) \tag{4.10}$$

$$= q(\varepsilon) \text{ (say).}$$

From (4.5), (4.6) and properties of K and c it is easy to see that $q(e)$ defined in (4.10) satisfies $\lim_{\varepsilon\searrow 0} q(\varepsilon) = 0$. (This follows for example from the fact that the derivative in (4.6) is always $\geqq -\varepsilon$ and $E_{\theta'}(e^{0(a-\varepsilon)}(f(x,\psi)/f(x,\theta'))^0) = 1$.) Hence the assumption (4.9) leads to a contradiction of (4.8). We therefore have

$$\liminf_{n\to\infty} n^{-1} \log \Pr_{\psi}\{\lambda_n(\psi, \theta') > k_n\}$$

$$= \liminf_{n\to\infty} \inf_{t>0} \log E_{\psi}\left(e^{-tk_n}\left(\frac{f(x,\theta')}{f(x,\psi)}\right)^t\right) \tag{4.11}$$

$$> -a.$$

Since V has a not larger significance level than U

$$\liminf_{n\to\infty} \sup_{\theta\in\Theta_0} n^{-1} \log \Pr_{\theta}\{\lambda_n(\Theta_0, \theta') > a-\varepsilon\}$$

$$\geqq \liminf_{n\to\infty} n^{-1} \log \Pr_{\psi}\{\lambda_n(\psi, \theta') > a-\varepsilon\} \tag{4.12}$$

$$> -a.$$

On the other hand, Lemma 5 of Bahadur (1966) states that

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta_0} n^{-1} \log \Pr_{\theta}\{\lambda_n(\Theta_0, \Theta_1) > a+\varepsilon\} \leqq -(a+\varepsilon) < -a. \tag{4.13}$$

It follows from (4.12), (4.13) and $\lim_{n\to\infty} n^{-1} \log \alpha_n = -a$ that

$$a-\varepsilon < \liminf_{n\to\infty} C_n < \limsup_{n\to\infty} C_n < a+\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $C_n \to a$. This proves part (e) of the theorem.

Choose any $\theta \in \Theta_0{}^*$. Consider parameter points in Θ^* of the form $(\gamma, \theta, \theta_1)$. Since J is continuous and $J(\theta) = 0$ for $\theta \in \Theta_0{}^*$ and $J(\theta_1) > a$, there exists a γ_2, say, such that $\theta_2 = (\gamma_2, \theta, \theta_1)$ satisfies: $\theta_2 \in \Theta^*$, $J(\theta_2) = a$, hence $\theta_2 \in \Theta_1{}^*$ and $f(\cdot, \theta_2) > 0$ a.e. (ν). It follows from part (e), established above, that $C_n{}^* \to a$. This proves part (c) of the theorem.

It is clear from (3.6) and (3.7) that $\zeta(\theta)$ is uniquely determined for each $\theta \in \Theta_0{}^*$, and $0 < \zeta < 1$. Define

$$b = \inf_{\theta\in\Theta_0{}^*} K((\zeta(\theta), \theta, \theta_1), \theta_1). \tag{4.14}$$

(We have yet to prove that $\beta(\theta_1)$ as defined in part (a) of the theorem satisfies $\beta(\theta_1) = b$.) From (3.5), $b < \infty$. Since $\overline{\Theta}$ is compact there is a convergent sequence $\{(\zeta(\theta_i), \theta_i, \theta_1): i = 2, 3, \cdots\}$ in $\overline{\Theta}$ such that

$$K((\zeta(\theta_i), \theta_i, \theta_1), \theta_1) \to b \tag{4.15}$$

and such that also $\{\zeta(\theta_i)\}$ converges in [0, 1] and $\{\theta_i\}$ converges. Define

$$\mu_0 = \lim_{i\to\infty} \theta_i$$
$$\mu_1 = \lim_{i\to\infty} (\zeta_i(\theta), \theta_i, \theta_1).$$

Since $b < \infty$, $\mu_0 = \infty$ is impossible. For, if $\theta_1 \to \infty$ then $K((\zeta(\theta_i), \theta_i, \theta_1), \theta_1) \to \infty$ unless $\zeta_i \to 1$ in such a manner that $\mu_1 \neq \infty$. However, if that is the case then there is a neighborhood N of ∞, $N \subset \overline{\Theta}$, such that $\{(\zeta(\theta_i), \theta_i, \theta_1), \theta_i\} \subset (\overline{\Theta} - N) \times \overline{\Theta}$. From this it follows that $\liminf K((\zeta(\theta_i), \theta_i, \theta_1), \theta_i) \geqq K(\mu_1, \infty) = \infty$, a contradiction; hence, $\mu_0 = \infty$ is impossible.

Note that we now have $\mu_0 \in \Theta_0^*$ and

$$\begin{aligned} \mu_1 &= (\zeta(\mu_0), \mu_0, \theta_1) \\ K(\mu_1, \mu_0) &= a \\ K(\mu_1, \theta_1) &= b. \end{aligned} \tag{4.16}$$

Since $K(\theta_1, \mu_0) > a = K(\mu_1, \mu_0)$, $0 < \zeta(\mu_0) < 1$, and $b > 0$.

[Incidentally, we have proved above that there is always a $\theta \in \overline{\Theta}_0$ such that $K((\zeta(\theta), \theta, \theta_1), \theta_1) = b$. After we show that $\beta(\theta_1) = b$, it then follows trivially that the necessary condition (g) of the theorem is never an empty condition.]

Next we prove

$$\limsup_{n\to\infty} n^{-1} \log \beta_n^{\varphi^*}(\theta_1) \leqq -b. \tag{4.17}$$

After proving (4.17), we then prove (4.27). From these inequalities the proof of parts (a), (b), (d), and (f) of the theorem easily follows.

Define $\eta(\theta)$ for $\theta \in \Theta_0^*$ by $K((\eta(\theta), \theta, \theta_1), \theta_1) = b$. $0 < \eta(\theta) < 1$ for $\theta \in \Theta_0^*$. Note that $K((\eta(\theta), \theta, \theta_1), \theta) \geqq a$. Since $K((\gamma, \theta, \theta_1), \theta_1)$ is a continuous function of γ, θ for $0 \leqq \gamma \leqq 1$, $\theta \in \Theta_0^*$, it follows that $\eta(\theta)$ and $(\eta(\theta), \theta, \theta_1)$ are continuous functions of θ for $\theta \in \Theta_0^*$. Also $\eta(\theta) \to 1$ as $\theta \to \infty$. Hence $\eta(\theta)$ is continuous on $\overline{\Theta}_0$, and $K((\eta(\theta), \theta, \theta_1), \theta) \to \infty$ as $\theta \to \infty$.

Similarly to (3.5) we can write

$$K((\gamma, \theta, \theta_1), \theta_1) = \log c(\gamma, \theta, \theta_1) + (1-\gamma)\frac{d}{d\gamma}(\log c(\gamma, \theta, \theta_1)). \tag{4.18}$$

Adding (4.18) to $(1-\gamma)\gamma^{-1}$ times (3.5) and simplifying gives

$$\log c(\gamma, \theta, \theta_1) = \gamma K((\gamma, \theta, \theta_1), \theta_1) + (1-\gamma)K((\gamma, \theta, \theta_1), \theta). \tag{4.19}$$

Hence

$$\log c(\eta(\theta), \theta, \theta_1) \geqq \eta(\theta)b + (1-\eta(\theta))a. \tag{4.20}$$

Now, let k_n: $n = 1, 2, \cdots$ satisfy $k_n \to a$. (In the present application $k_n = C_n^*$, but the following results do not depend on that fact.) Using (4.20) we can write

$$\begin{aligned} &\Pr_{\theta_1}\{\lambda_n(\Theta_0^*, \Theta_1^*) \leqq k_n\} \\ &\leqq \Pr_{\theta_1}\{\underline{x}: \exists \theta \in \Theta_0^* \ni \lambda_n(\theta, (\eta(\theta), \theta, \theta_1)) \leqq k_n\} \end{aligned}$$

$$
\begin{aligned}
(4.21) \qquad &= \Pr_{\theta_1}\{\underline{x}: \exists \theta \in \Theta_0^* \ni \log c(\eta(\theta), \theta, \theta_1) \\
&\quad + \eta(\theta)\lambda_n(\theta, \theta_1) \leqq k_n\} \\
&= \Pr_{\theta_1}\left\{\sup_{\theta \in \Theta_0}\left[\lambda_n(\theta, \theta_1) - \frac{k_n - \log c(\eta(\theta), \theta, \theta_1)}{\eta(\theta)}\right] \leqq 0\right\} \\
&\leqq \Pr_{\theta_1}\left\{\sup_{\theta \in \Theta_0}\left[\lambda_n(\theta, \theta_1) - (a-b) + \frac{a-k_n}{\eta(\theta)}\right] \leqq 0\right\}.
\end{aligned}
$$

Since $\eta(\theta)$, being continuous on $\overline{\Theta}_0$, is bounded away from 0, and

$$-K((\eta(\theta), \theta, \theta_1), \theta_1) = -b < a-b < a \leqq K((\eta(\theta), \theta, \theta_1), \theta)$$

we may apply Lemma 4, thus getting

$$
\begin{aligned}
(4.22) \quad &\limsup_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\lambda_n(\Theta_0^*, \Theta_1^*) \leqq k_n\} \\
&\qquad \leqq \sup_{\theta \in \Theta_0} \operatorname{in}_{0<t<1} \log E_{\theta_1}\left(e^{t(a-b)}\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^t\right).
\end{aligned}
$$

Using the definition of c and (4.20) we have

$$
\begin{aligned}
(4.23) \qquad &\log E_{\theta_1}\left(e^{(1-\eta(\theta))(a-b)}\left(\frac{f(x, \theta)}{f(x, \theta_1)}\right)^{(1-\eta(\theta))}\right) \\
&= (1-\eta(\theta))(a-b) - \log c(\eta(\theta), \theta, \theta_1) \\
&\leqq -b.
\end{aligned}
$$

This proves that (4.17) is valid. We now proceed to establish (4.27).

Consider the problem of testing the simple hypothesis μ_0 versus the alternative θ_1. The most powerful level α_n test of μ_0 versus θ_1 has rejection region defined by

$$\text{Reject} \quad \mu_0 \quad \text{if } \lambda_n(\mu_0, \theta_1) > k_n''.$$

Since μ_0, $\mu_1 = (\eta(\mu_0), \mu_0, \theta_1)$, and θ_1 form a monotone likelihood ratio family an equivalent test is that given by

$$\text{Reject} \quad \mu_0 \quad \text{if } \lambda_n(\mu_0, \mu_1) > k_n'$$

where k_n' (and k_n'') are chosen to give level α_n. Since $K(\mu_1, \mu_0) = a$ part (e) of the theorem (with $\Theta_0 = \{\mu_0\}$, $\Theta_1 = \{\mu_1\}$) proves $k_n' \to a$. Thus

$$
\begin{aligned}
&\lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\{\lambda_n(\mu_0, \mu_1) \leqq k_n'\} \\
(4.24) \qquad &= \lim_{n\to\infty} n^{-1} \log \Pr_{\theta_1}\left\{\lambda_n(\mu_0, \theta_1) - \frac{k_n' - \log c(\eta(\mu_0), \mu_0, \theta_1)}{\eta(\mu_0)} \leqq 0\right\} \\
&= \inf_{0<t<1} \log E_{\theta_1}\left(e^{t(a-b)}\left(\frac{f(x, \mu_0)}{f(x, \theta_1)}\right)^t\right) \\
&= \inf\left(t(a-b) - \log c(1-t, \mu_0, \theta_1)\right).
\end{aligned}
$$

[Note: $\eta(\mu_0) = \zeta(\mu_0)$.] Using (3.5) and (4.19) we have

$$\frac{d}{dt}\left(t(a-b)-\log c(1-t, \mu_0, \theta_1)\right)\Big]_{t=1-\eta(\mu_0)} \tag{4.25}$$

$$= a-b-\frac{K((\eta(\mu_0), \mu_0, \theta_1), \mu_0)-\log c(\eta(\mu_0), \mu_0, \theta_1)}{\eta(\mu_0)}$$

$$= 0.$$

Hence

$$\inf_{0<t<1}\left(t(a-b)-\log c(1-t, \mu_0, \theta_1)\right) = -b = -K(\mu_1, \theta_1). \tag{4.26}$$

(4.24) and (4.26) prove that

$$\liminf_{n\to\infty} n^{-1}\log \beta_n(\theta_1) \geqq -b. \tag{4.27}$$

Combining (4.27) and (4.17),

$$\lim_{n\to\infty} n^{-1}\log \beta_n(\theta_1) = \lim_{n\to\infty} n^{-1}\log \beta_n^{Q^*}(\theta_1) = -b. \tag{4.28}$$

This proves parts (a), (b), and (d) of the theorem.

To prove part (f), observe that for $\zeta(\theta) \leqq \xi \leqq \eta(\theta)$ it follows from (4.19), that

$$\frac{\log c(\xi, \theta, \theta_1)-a}{\xi} \geqq b-a+\frac{1-\xi}{\xi}\left(K((\xi, \theta, \theta_1), \theta)-a\right).$$

Let $\gamma = \gamma(\theta)$ satisfy the assumption (4.3). For sufficiently large n, (so that $C_n < a+\varepsilon/2$) we have, as in (4.21)

$$\begin{aligned}
&\Pr_{\theta_1}\{\lambda_n(\Theta_0, \Theta_1) \leqq C_n\} \\
&\leqq \Pr_{\theta_1}\{\exists\theta \in \Theta_0 \ni \lambda_n(\theta, (\gamma(\theta), \theta, \theta_1)) \leqq C_n\} \qquad (4.29)\\
&\leqq \Pr_{\theta_1}\left\{\sup_{\theta\in\Theta_0}\left(\lambda_n(\theta, \theta_1)-(a-b)+\frac{a-C_n+(1-\gamma(\theta))(\varepsilon-\gamma(\theta))^+}{\gamma(\theta)}\right)\leqq 0\right\} \\
&\leqq \Pr_{\theta_1}\left\{\sup_{\theta\in\Theta_0}\left(\lambda_n(\theta, \theta_1)-(a-b)-\frac{2|C_n-a|}{\varepsilon}\right) \leqq 0\right\}.
\end{aligned}$$

Since $C_n \to a$, we can proceed as in (4.22) and (4.23) to prove

$$\limsup_{n\to\infty} n^{-1}\log \beta_n^{Q}(\theta_1) \leqq -b. \tag{4.30}$$

(4.2) of (f) then follows from parts (a) and (b) of the theorem.

It remains only to prove part (g) of the theorem. To begin, we suppose $\theta_0' \in \overline{\Theta}_0$ satisfies the condition (4.4), namely

$$K((\zeta(\theta_0'), \theta_0', \theta_1), \theta_1) = b,$$

and that $(\zeta(\theta_0'), \theta_0', \theta_1) \notin \Theta_1{}^c$. In the following we use the fact $C_n \to a$.

Let $\zeta' = \zeta(\theta_0')$, and $0 < \zeta'' < \zeta(\theta_0')$, and let $\theta' = (\zeta(\theta_0'), \theta_0', \theta_1)$, and $\theta'' = (\zeta'', \theta_0', \theta_1)$. We consider the problem of testing $\{\theta''\}$ versus θ_1. We compare the properties of two different sequences of tests for this problem. These are T_1:

(4.31) Reject if $\lambda_n(\theta_0', \Theta_1) > C_n$,

and T_2:

(4.32) Reject if $\lambda_n(\theta_0', \theta') > l$

where l is to be determined later ($l \geqq a$).

We first consider the asymptotic behavior of $\alpha_n^{T_1}$. Application of Lemma 5 yields

$$\text{(4.33)}\quad \limsup_{n\to\infty} n^{-1} \log \alpha_n^{T_1} = \limsup_{n\to\infty} n^{-1} \log \Pr_{\theta''}\{\lambda_n(\theta_0', \Theta_1) > C_n\}$$

$$\leqq \sup_{\theta \in \Theta_1} \inf_{t>0} \log E_{\theta''}\left\{e^{-ta}\left(\frac{f(x,\theta)}{f(x,\theta_0')}\right)^t\right\}$$

$$E_{\theta''}\left\{e^{-ta}\left(\frac{f(x,\theta)}{f(x,\theta_0')}\right)^t\right\}$$

$$\text{(4.34)}\quad = e^{-ta}\int\left(\frac{f(x,\theta)}{(x,\theta_0')}\right)^t c(\zeta'', \theta_0', \theta_1) f^{1-\zeta''}(x,\theta_0') f^{\zeta''}(x,\theta_1)\nu(dx)$$

$$= e^{-ta}\frac{c(\zeta'', \theta_0', \theta_1)}{c^{1-t}(\zeta''/(1-t), \theta_0', \theta_1)}$$

$$\cdot \int f^t(x,\theta) f^{1-t}(x, (\zeta''/(1-t), \theta_0', \theta_1))\nu(dx).$$

If we let $t = t' = (\zeta' - \zeta'')/\zeta'$, so that $\zeta''/(1-t') = \zeta'$, we have from (4.34)

$$\inf_{t>0} \log E_{\theta''}\left\{e^{-ta}\left(\frac{f(x,\theta)}{f(x,\theta_0')}\right)^t\right\}$$

$$\text{(4.35)}\quad \leqq -t'a + \log c(\zeta'', \theta_0', \theta_1) - (\zeta''/\zeta') \log c(\zeta', \theta_0', \theta_1)$$

$$+ \log \int f^{t'}(x,\theta) f^{1-t'}(x,\theta')\nu(dx).$$

Since $1 > t' > 0$ is independent of θ, and $\theta' \notin \Theta_1^c$, and Θ_1^c is compact in $\overline{\Theta}$ the integral which occurs on the right of (4.35) which is a continuous function of θ is uniformly less than one on Θ_1^c. Hence there is an $\varepsilon_1 > 0$ such that

$$\limsup_{n\to\infty} n^{-1} \log \alpha_n^{T_1}$$

$$\leqq \inf_{t>0} \log E_{\theta''}\left\{e^{-ta}\left(\frac{f(x,\theta)}{f(x,\theta_0')}\right)^t\right\}$$

$$\text{(4.36)}\quad \leqq -a(\zeta' - \zeta'')/\zeta' + \log c(\zeta'', \theta_0', \theta_1)$$

$$-(\zeta''/\zeta') \log c(\zeta', \theta_0', \theta_1) - \varepsilon_1$$

$$= -N - \varepsilon_1 \text{ (say)}$$

for all $\theta \in \Theta_1^c$.

Now we investigate $\alpha_n^{T_2}$. Let $\theta = \theta'$ in (4.34).

$$\log E_{\theta''}\left\{e^{-ta}\left(\frac{f(x,\theta')}{f(x,\theta_0')}\right)^t\right\}$$

$$(4.37)\qquad = -ta+\log c(\zeta'', \theta_0', \theta_1)+t\log c(\zeta\ , \theta_0', \theta_1) \\ -\log c(\zeta''+t\zeta', \theta_0', \theta_1).$$

Differentiating (4.37) with respect to t and using (3.5) gives

$$-a+\log c(\zeta', \theta_0', \theta_1)+\frac{\zeta'}{\zeta''+t\zeta'}\left(K(\zeta''+t\zeta', \theta_0', \theta_1), \theta_0'\right) \\ -\log c(\zeta''+t\zeta', \theta_0', \theta_1))$$

which clearly has the value zero if (and only if) $t = t'$.

Substituting $t = t'$ in (4.37),

$$\inf_{t>0} \log E_{\theta''}\left\{e^{-ta}\left(\frac{f(x, \theta')}{f(x, \theta_0')}\right)^t\right\} = -N.$$

From evident continuity properties, there is an $l > a$ such that

$$(4.38)\quad \lim_{n\to\infty} n^{-1}\log \alpha_n^{T_2} \\ = \inf_{t>0} \log E_{\theta''}\left\{e^{-tl}\left(\frac{f(x, \theta')}{f(x, \theta_0')}\right)^t\right\} \geqq -N-\varepsilon_1/2 > -N-\varepsilon_1.$$

Since θ_0', θ'', θ', θ_1 form a monotone likelihood ratio family, the test T_2 is most powerful among level $\leqq \alpha_n^{T_2}$ tests. But $\alpha_n^{T_1} < \alpha_n^{T_2}$ for sufficiently large n. Hence

$$(4.39)\qquad \beta_n^{T_1}(\theta_1) \geqq \beta_n^{T_2}(\theta_1)$$

for all sufficiently large n.

Note that

$$(4.40)\qquad \lim_{n\to\infty} n^{-1}\log \beta_n^{T_2}(\theta_1) \\ = \inf_{t>0} \log E_{\theta_1}\left\{e^{+tl}\left(\frac{f(x, \theta_0')}{f(x, \theta')}\right)^t\right\} \\ > -\beta(\theta_1).$$

(Compare (4.40) with (4.27).)

Combining (4.39) and (4.40)

$$\liminf_{n\to\infty} n^{-1}\log \beta_n^{T_1}(\theta_1) \geqq \lim_{n\to\infty} n^{-1}\log \beta_n^{T_2}(\theta_1) \\ > -\beta(\theta_1).$$

Hence

$$(4.41)\qquad \lim_{n\to\infty} n^{-1}\log \Pr_{\theta_1}\{\lambda_n(\Theta_0, \Theta_1) \leqq C_n\} \\ \geqq \lim_{n\to\infty} n^{-1}\log \Pr_{\theta_1}\{\lambda_n(\theta_0', \Theta_1) \leqq C_n\} \\ = \lim_{n\to\infty} n^{-1}\log \beta_n^{T_1} \\ > -\beta(\theta_1).$$

Since (4.41) contradicts (4.2), this proves that the condition in part (g) is necessary for (4.2) to hold; and if this "necessary" condition is violated then, in fact, (4.41) is satisfied. This completes the proof of part (g).

The proof of the theorem is complete.

5. Main results on asymptotic optimality: statements and proofs. In this section we formally state and, using Theorem 1, prove the Main Results claimed in the Introduction.

Main Result 1 is contained in

THEOREM 2. *Suppose Assumptions* 1–4 *are satisfied. Let S be any sequence of tests satisfying* $\limsup_{n\to\infty}\alpha_n^S < 1$. *Define the sequence,* Q^1, *of tests by the rejection regions*:

$$\text{Reject } \Theta_0 \text{ if } \lambda_n^* > C_n^1$$

where the critical constants C_n^1 *are determined by* $\alpha_n^{Q^1} \doteq \alpha_n^S$. *Then, for each* $\theta_1 \in \Theta_1$ *satisfying* (a) $f(\cdot, \theta_1) > 0$ a.e. (ν), *and* (b) $K(\theta_0, \theta_1) < \infty$ *for some* $\theta_0 \in \Theta_0$,

$$\liminf_{n\to\infty}\left(n^{-1}\log\beta_n^S(\theta_1) - n^{-1}\log\beta_n^{Q^1}(\theta_1)\right) \geqq 0. \tag{5.1}$$

[We remark here to amplify some remarks following the statement of Heuristic Principle 1 in Section 1 that it is possible for the equality to zero to hold in (5.1) and yet $\beta_n^S(\theta_1) < \beta_n^{Q^1}(\theta_1)$ for all n.]

PROOF. For clarification, we remark that the conclusion of Theorem 2 is contained verbatim in Theorem 1 in the special case where $\lim_{n\to\infty} n^{-1}\log\alpha_n^S = -a$ exists (in the sequel we sometimes allow $a = \infty$) and satisfies $\infty > a > 0$ and where $J(\theta_1) > a$. For the case where $J(\theta_1) \leqq a \leqq \infty$ we proceed as follows:

For this paragraph suppose $f(\cdot, \theta_1) > 0$ a.e. (ν) and $J(\theta_1) \leqq a$. Define $\psi \in \Theta_0^*$ by $K(\theta_1, \psi) = J(\theta_1)$. Fix $\varepsilon > 0$. Consider the problem of testing the null hypothesis $\{\psi\}$ versus the alternative $\{\theta_1\}$ at a sequence of significance levels each of which is at most $\alpha_n(\varepsilon)$ where $n^{-1}\log\alpha_n(\varepsilon) \to -(J(\theta_1)-\varepsilon)$. (For this paragraph only, observe the convention, $\infty - \varepsilon = 1/\varepsilon$.) By Theorem 1(b) the most powerful such sequence of tests, say $T(\varepsilon)$, satisfies

$$\lim_{n\to\infty} n^{-1}\log\beta_n^{T(\varepsilon)}(\theta_1) \geqq -K\big(\big(\zeta_\varepsilon(\psi), \psi, \theta_1\big), \theta_1\big)$$

where $K(\zeta_\varepsilon(\psi), \psi, \theta_1), \psi) = J(\theta_1) - \varepsilon$. From the properties of K given in the proof of Theorem 1 it follows that $\lim_{\varepsilon\searrow 0}\zeta_\varepsilon(\psi) = 1$ and thus

$$\lim_{\varepsilon\downarrow 0} K\big(\big(\zeta_\varepsilon(\psi), \psi, \theta_1\big), \theta_1\big) = 0.$$

Since for each $\varepsilon > 0$, $\alpha_n(\varepsilon) > \alpha_n^S$ for sufficiently large n, it follows that $\beta_n^S(\theta_1) \geqq \beta_n^{T(\varepsilon)}$. Letting $\varepsilon \downarrow 0$ we have for the θ_1 of this paragraph

$$\lim_{n\to\infty} n^{-1}\log\beta_n^S(\theta_1) = 0. \tag{5.2}$$

Hence if $0 < a = \lim_{n\to\infty} n^{-1}\log\alpha_n^S \leqq \infty$, Theorem 2 follows either from (5.2) if $J(\theta_1) \leqq a$ or from Theorem 1 (a), (d) if $J(\theta_1) > a$.

Now suppose $\lim n^{-1} \log \alpha_n{}^S = 0$. Then Q^1 has smaller β than the sequence of tests, Q^*, corresponding to levels $\alpha_n = e^{-na}$, for any $a > 0$.

Hence, applying Theorem 1 (a), (b), (d) for fixed values of a satisfying $0 < a < J(\theta_1)$ gives

$$\lim\sup_{n\to\infty} n^{-1} \log \beta_n{}^{Q^1}(\theta_1) \leqq -b_a(\theta_1) \tag{5.3}$$

where b_a is defined by $b_a(\theta_1) = \inf_{\theta_0 \in \Theta_0{}^*} K(\zeta^a(\theta_0), \theta_0, \theta_1), \theta_1)$, and ζ^a is defined by $K((\zeta^a(\theta_0), \theta_0, \theta_1), \theta_0) = a$.

In the proof of Theorem 1 it is shown (following (4.16)) that $b_a(\theta_1) < \infty$ and there is a $\theta_0(a) \in \Theta_0{}^*$ such that

$$K(\theta_0(a), \theta_1) = b_a(\theta_1).$$

We can choose $a_i \searrow 0$ such that $\theta_0(a_i)$ converges, say $\theta_0(a_i) \to \varphi$, $\varphi \in \overline{\Theta}_0$. Using Lemma 1, $K(\theta_0(a_i), \theta_1) \to K(\varphi, \theta_1)$. By construction, and the properties (3.7) and (3.8) of K it follows that

$$K(\varphi, \theta_1) = \inf_{\theta_0 \in \Theta_0{}^*} K(\theta_0, \theta_1).$$

By Assumption (b) of the theorem $K(\varphi, \theta_1) < \infty$. Hence $\varphi \neq \infty$. From (5.3) we now have

$$\lim\sup_{n\to\infty} n^{-1} \log \beta_n{}^{Q^1}(\theta_1) \leqq -K(\varphi, \theta_1). \tag{5.4}$$

Consider the problem of testing (by, say, T) the null hypothesis φ against the alternative θ_1 at a sequence of levels α_n satisfying $\lim\sup_{n\to\infty} \alpha_n < 1$. By Rao (1962) the probabilities of type II error, $\beta_n{}^T(\theta)$, of this test must satisfy

$$\lim\inf_{n\to\infty} n^{-1} \log \beta_n{}^T(\theta_1) \geqq -K(\varphi, \theta_1). \tag{5.5}$$

(5.4) and (5.5) together prove (5.1) in the case where $\lim_{n\to\infty} n^{-1} \log \alpha_n{}^S = 0$. ((5.5) is originally due to C. Stein, see Bahadur (1967) page 316–317; also Bahadur (1969).)

To finish the proof of the theorem, suppose for some S and θ_1 satisfying (a) and (b) (5.1) fails to hold. Then there is a subsequence n_i such that

$$\lim\inf_{i\to\infty} (n_i{}^{-1} \log \beta_{n_i}^S(\theta_1) - n_i{}^{-1} \log \beta_n{}^1(\theta_1)) < 0 \tag{5.6}$$

and such that $a' = -\lim_{i\to\infty} n^{-1} \log \alpha_{n_i}$ exists. (Possibly $a' = \infty$.) There clearly exists a sequence of tests, say S', defined for every n such that

$$\begin{aligned} \alpha_{n_i}^{S'} &= \alpha_{n_i}^S \\ \beta_{n_i}^{S'}(\theta_1) &= \beta_{n_i}^S(\theta_1) \\ \lim_{n\to\infty} n^{-1} \log \alpha_n{}^{S'} &= -a' \end{aligned} \tag{5.7}$$

Since $\lim_{n\to\infty} n^{-1} \log \alpha_n{}^{S'}$ exists we have already established that (5.1) is true for S'. But, from (5.6) and (5.7)

$$\lim\inf_{n\to\infty} (n^{-1} \log \beta_n{}^{S'}(\theta_1) - n^{-1} \beta_n^{Q^1}(\theta_1)) < 0,$$

a contradiction. Hence (5.6) must be false, which proves that (5.1) is true in general.

It should be clear from the proof given above how to prove that if $J(\theta_1) > -\liminf_{n\to\infty} n^{-1} \log \alpha_n^S$ then $\lim_{n\to\infty}(C_n^1 + n^{-1} \log \alpha_n^S) = 0$; and also how to use Theorem 1(b) to construct an asymptotic lower bound for $n^{-1} \log \beta_n^S(\theta_1)$ in terms of $n^{-1} \log \alpha_n^S$ using the formula (4.1). It is also clear that the "necessary condition" of Theorem 1—Theorem 1(e), (g)—can be applied to the general situation by considering separate subsequences $\{n_i\}$ such that $0 < \lim_{i\to\infty} n_i^{-1} \log \alpha_{n_i}^S < \infty$. The statement and proof of such a result follows directly from Theorem 1(e), (g) and the simple fact that one can define a sequence, say S', defined for every n, which satisfies (5.7). We leave the details of the above remarks to the interested reader.

The second Main Result mentioned in the Introduction is contained in

THEOREM 3. *Suppose Assumptions* 1–4 *are satisfied. Let* $\theta_1 \in \Theta_1$ *be any point such that* (a) $f(\cdot, \theta_1) > 0$ a.e. (ν) *and* (b) $J(\theta_1) < \infty$. *Suppose* S *is any sequence of tests satisfying* $\limsup_{n\to\infty} \beta_n^S(\theta_1) < 1$. *Then there exists a sequence of tests,* Q^2, *defined by*

$$\text{Reject } \Theta_0 \text{ if } \lambda_n^* > C_n^2$$

which satisfies

$$\beta_n^{Q^2}(\theta_1) \leqq \beta_n^S(\theta_1) \qquad \text{for all } n \tag{5.8}$$

and

$$\liminf_{n\to\infty} \left(n^{-1} \log \alpha_n^S - n^{-1} \log \alpha_n^{Q^2}\right) \geqq 0. \tag{5.9}$$

PROOF. Since $\limsup_{n\to\infty} \beta_n^S(\theta_1) < 1$ it follows from Rao (1962) (see (5.5)) that $\liminf_{n\to\infty} n^{-1} \log \alpha_n^S \geqq -J(\theta_1) > -\infty$. At first we suppose $\lim_{n\to\infty} n^{-1} \log \alpha_n^S = -a$, $a > 0$, (also $a < \infty$), exists. For $0 < \varepsilon < a$ let T_ε denote the sequence of tests based on λ_n^*, and having critical constants $C_n(\varepsilon)$ satisfying $\alpha_n^{T\varepsilon} \doteq e^{-n(a-\varepsilon)}$. It follows from Theorem 1(b), (d) and the continuity properties of K (particularly (3.7), (3.8)) that

$$\lim_{n\to\infty} n^{-1} \log \beta_n^{T_\varepsilon}(\theta_1) < \liminf_{n\to\infty} n^{-1} \log \beta_n^S(\theta_1). \tag{5.10}$$

From (5.10) it is clear that for each ε there is an $n(\varepsilon)$ such that $n > n(\varepsilon)$ implies $\beta_n^{T_\varepsilon}(\theta_1) < \beta_n^S(\theta_1)$.

In the following we assume that $n(\varepsilon)$ is chosen to be the smallest possible integer for which the above holds. From the definition of the tests T_ε it is clear that $C_n(\varepsilon)$ is a non-increasing function of ε. Hence $\beta_n^{T_\varepsilon}(\theta_1)$ and $n(\varepsilon)$ are also non-increasing functions of ε. For each n, let

$$\varepsilon(n) = \inf\{\varepsilon : n(\varepsilon) \leqq n\}.$$

$\varepsilon(n)$ is a non-increasing function of n. From the above we see that $\varepsilon(n) \searrow 0$ as $n \to \infty$. For the case at hand, define $C_n^2 = C_n(\varepsilon(n))$. The test Q^2 then satisfies both (5.8) and (5.9).

If, instead of the above situation, we have $\lim_{n\to\infty} n^{-1} \log \alpha_n^S = 0$ the theorem is trivial. For, C_n^2 can certainly be chosen so that (5.8) is satisfied, and (5.9) will be true simply because $\alpha_n^S \leqq 1$.

Finally, suppose the theorem is false. Then there exists a subsequence n_i such that $\lim_{i\to\infty} n_i^{-1} \log \alpha_{n_i}^{S} = -a'$ exists and such that for any sequence of tests of the form of Q^2 satisfying (5.8),

$$\liminf_{i\to\infty} \left(n_i^{-1} \log \alpha_{n_i}^{S} - n_i^{-1} \log \alpha_{n_i}^{Q^2}\right) < 0.$$

Since $\limsup_{n\to\infty} \beta_n^{S}(\theta_1) < 1$ it follows from Rao (1962) that $a' \leqq J(\theta_1)$. But then, as in the proof of Theorem 2, we could construct a sequence of tests S' satisfying $\lim_{i\to\infty} n^{-1} \log \alpha_n^{S'} = a'$ and $\limsup_{n\to\infty} \beta_n^{S}(\theta_1) < 1$ and such that the theorem is false for this sequence S'. This contradicts what we have established in the preceding paragraphs of the proof.

This completes the proof of the theorem.

6. Theory and examples concerning the use of extra information. In this section we give some examples to illustrate the general principles concerning extra information which were mentioned at the end of the Introduction. All of the results we use are essentially contained in Theorem 1. In order to emphasize them we state them here as corollaries before beginning the examples.

For the statement of these results we suppose Assumptions 1–4 are satisfied; except that we will relax the assumption $\Theta_0{}^c = \Theta_0{}^*$ in order to state the second result below.

As the "basic" problem we consider a problem of testing $\Theta_0{}^*$ versus $\Theta_1{}^*$, where these spaces satisfy the relevant parts of Assumptions 1–4. By "extra" information we mean information which limits the parameter space to a set Θ', smaller than Θ^*. $\Theta' \subset \Theta'^c \subsetneqq \Theta^*$. Define $\Theta_i{}'^c = \Theta'^c \cap \Theta_i{}^*$, $i = 0, 1$. If $\Theta_i{}'^c \neq \Theta_i{}^*$, then we say we have "extra" information about the null ($i = 0$) or alternative ($i = 1$) hypothesis, respectively.

We state our results only for the situation

$$\lim_{n\to\infty} n^{-1} \log \alpha_n = -a.$$

They can be generalized to other situations by the methods of Section 5. We use Q' as a generic symbol to denote any sequence of tests based on

$$\lambda_n' = \lambda_n(\Theta_0', \Theta_1') = \lambda_n(\Theta_0'^c, \Theta_1'^c)$$

satisfying $\alpha_n{}^{Q'} = \alpha_n$, and where the nature of Θ_0' and Θ_1' should be clear from the context. Q^* denotes the sequence based on $\lambda_n{}^*$ satisfying $\alpha_n{}^{Q^*} = \alpha_n$, and Q'^* denotes the test satisfying $\alpha_n{}^{Q'} \doteq \alpha_n$ based on $\lambda_n'^* = \lambda_n(\Theta_0'^*, \Theta_1'^*)$ where $\Theta_0'^*$ and $\Theta_1'^*$ are formed in the natural way from $\Theta_0'^c$ and $\Theta_1'^c$ as indicated at the end of Section 2.

For fixed a (as above), and $\Theta_1 \in \Theta_1{}^*$, $0 \leqq a \leqq J(\theta_1)$, define (as in Section 4) $\zeta(\cdot)$ on $\Theta_0{}^*$ by $K((\zeta(\theta), \theta, \theta_1), \theta) = a$.

Let S_0 be defined by

$$(6.1)\quad S_0 = \{\theta_0 : \theta_0 \in \Theta_0{}^*, K((\zeta(\theta_0), \theta_0, \theta_1), \theta_1) = \inf_{\theta \in \Theta_0{}^*} K((\zeta(\theta), \theta, \theta_1), \theta_1)\}.$$

In Example 3 we will emphasize the fact that S_0 depends on θ_1 by writing $S_0(\theta_1)$.

Define $\varphi(\theta)$ for $\theta \in \Theta_0{}^*$ by $K((\varphi(\theta), \theta, \theta_1), \theta_1) = \inf_{\theta \in \Theta_0{}^*} K((\zeta(\theta), \theta, \theta_1), \theta_1) = \beta(\theta_1)$. If $\theta \in S_0$ then $\varphi(\theta) = \zeta(\theta)$.

COROLLARY 1. Extra information about the alternative. *Assume* $f(\cdot, \theta_1) > 0$ a.e. (ν).

(a) *Extra information about the alternative can never increase the rate of exponential convergence to* 0 *of* β. *That is*

$$\lim_{n\to\infty} n^{-1} \log \beta_n^{Q'}(\theta_1) \geqq \lim_{n\to\infty} n^{-1} \log \beta_n^{Q^*}(\theta_1). \tag{6.2}$$

(b) *If* $J(\theta_1) > a > 0$ *and if there is a* $\theta' \in \Theta_1{}'^c$ *satisfying* $f(\cdot, \theta') > 0$ a.e. (ν) *and* $J(\theta') = a$ *then there is equality in* (6.2) *if and only if* $\theta_0 \in S_0$ *implies* $(\varphi(\theta_0), \theta_0, \theta_1) \in \Theta_1{}'^c$.

Note. It can also be deduced from part (b), and from the obvious converse to Theorem 1(e), that if equality holds in (6.2) then

$$\lim_{n\to\infty} n^{-1} \log \alpha_n^{Q'}(\theta_0) = -a \qquad \text{for all } \theta_0 \in S_0.$$

Also note that both sides of (6.2) are finite unless $a = 0$ and $K(\theta_0, \theta_1) = \infty$ for all $\theta_0 \in \Theta_0$, in which case both sides of (6.2) are ∞.

PROOF OF COROLLARY. If $J(\theta_1) \leqq a$, $\lim_{n\to\infty} n^{-1} \log \beta_n^{Q^*}(\theta_1) = 0$ so (6.2) is trivial. If $0 < a < J(\theta_1)$ part (a) is a restatement of Theorem 1(d). And, if $0 = a$ part (a) is contained in Theorem 2.

Under the assumption we have made, namely $0 < a < J(\theta_1)$, part (b) is contained in Theorem 1(g). This completes the proof of the corollary.

Note. If $a = 0$ there may or may not be equality in (6.2). If $S_0 \notin \Theta_1{}^c$ then presumably inequality always holds in (6.2). However in the interesting case when $S_0 \in \Theta_1{}^c$ (as in Example 1, below) sometimes there is equality in (6.2) and sometimes there is inequality in (6.2). We do not know precise conditions under which equality holds.

The precise implication of Corollary 2, below, is clarified in Example 3 of this section.

COROLLARY 2. Extra information about the null hypothesis. *Assume* $f(\cdot, \theta_1) > 0$ a.e. (ν).

(a) *Extra information about the null hypothesis can never decrease the rate of exponential convergence to zero of* β. *That is,*

$$\lim_{n\to\infty} n^{-1} \log \beta_n^{Q'^*}(\theta_1) \leqq \lim_{n\to\infty} n^{-1} \log \beta_n^{Q^*}(\theta_1). \tag{6.3}$$

(b) *If* $J(\theta_1) > a > 0$ *there is equality in* (6.3) *if and only if there is a* $\theta_0 \in S_0$ *such that* $\theta_0 \in \Theta_0{}'^c = \Theta_0{}'^*$. *(i.e. if and only if* $S_0 \cap \Theta_0{}'^c \neq \varnothing$.)

PROOF OF COROLLARY. If $a \geqq J(\theta_1)$ both sides of (6.3) are zero. Using Theorem 1(b), (d), if $0 < a < J(\theta_1)$,

$$\lim_{n\to\infty} n^{-1} \log \beta_n^{Q'^*}(\theta_1) = -\inf_{\theta_0 \in \Theta_0{}'^c} K(\varphi(\theta_0), \theta_0, \theta_1), \theta_1)$$

and

$$\lim_{n\to\infty} n^{-1} \log \beta_n^{Q^*}(\theta_1) = -\inf_{\theta_0 \in \Theta_0{}^*} K((\varphi(\theta_0), \theta_0, \theta_1), \theta_1).$$

Since $\Theta_0'^c \subset \Theta_0^*$, (6.3) follows. Part (b) also follows in this situation from the definition of S_0 and continuity properties of K. This completes the proof of the Corollary.

The techniques used in the proofs of Corollaries 1 and 2 can clearly be used to give some sorts of results if the extra information is about both the null and alternative hypotheses simultaneously. However, the situation is more complicated, and we have not been able to develop a statement of a result which adds anything to the reasoning suggested by the relevant parts of Theorem 1.

We begin the examples with a simple, though slightly artificial example illustrating Corollary 1. This example is a particularly good one in that all the likelihood ratio tests and the relevant limiting probabilities are easy to compute. It is the only example presented here in which all these computations are carried out.

EXAMPLE 1. Let $\mathcal{X} = (0, \infty) \times (0, \infty) = \{(x, y): x > 0, y > 0\}$; ν = Lebesgue measure; $\Theta = (0, \infty)$; and $f((x, y), \theta) = e^{-\theta x - \theta^{-1} y}$. Let $\Theta_0 = \{1\} = \{\theta_0\}$ and $\Theta_1 = \Theta - \Theta_0$.

We can take $\Theta^* = (0, \infty) \times (0, \infty) = \{(\varphi_1, \varphi_2): \varphi_1 > 0, \varphi_2 > 0\}$, and

$$f((x, y), \varphi) = \varphi_1 \varphi_2 e^{-\varphi_1 x - \varphi_2 y}.$$

For algebraic and topological purposes we imbed Θ^* in E_2. Clearly, $\theta \in \Theta$ corresponds to the point $(\theta, \theta^{-1}) \in \Theta^*$. Also, it is easy to see that for $\theta_1, \theta_2 \in \Theta^*$, $(\gamma, \theta_1, \theta_2) = \gamma\theta_1 + (1-\gamma)\theta_2$.

Since points $(\gamma, \theta_0, \theta_1) \in \Theta^*$ are on the straight line joining θ_0 and θ_1 in Θ^* and since Θ is a strictly convex curve in Θ^*, none of the points $(\gamma, \theta_0, \theta_1)$, $0 < \gamma < 1$, $\theta_1 \in \Theta$, is in Θ. It is clear (and we will compute explicitly below) that $J(\theta)$ is continuous on Θ so that for any $a > 0$ there is a $\theta' \in \Theta$ such that $J(\theta') = a$. Hence for $0 < a < J(\theta_1)$ inequality holds in (6.2). That is to say, from our asymptotic non-local point of view, λ_n^* is a strictly better statistic than λ_n.

We illustrate the above remarks by describing how to compute the relevant values of α and β and give one numerical example. Direct computation gives

$$(6.3a) \quad K((\varphi_1, \varphi_2); (\varphi_1', \varphi_2')) = (\varphi_1' - \varphi_1)/\varphi_1 + (\varphi_2' - \varphi_2)/\varphi_2 + \log \varphi_1' \varphi_2'.$$

Because of the logarithm which occurs on the right of (6.3a), it is awkward to solve in general for $\zeta(\theta_0)$ given an arbitrary value for $\theta_1 \in \Theta$ and a.

For the specific values $\theta_1 = 4 \in \Theta$ (i.e. $\theta_1 \equiv (4, \frac{1}{4}) \in \Theta^*$) and $a = \log(\frac{3}{2}) - \frac{1}{6} \approx .24$ we have $\zeta = \frac{1}{3}$. Hence $(\zeta, \theta_0, \theta_1) = (2, \frac{3}{4}) \in \Theta^*$. Using (6.3a) and Theorem 1 we then have for the test Q^* which rejects when $\lambda_n^* > a$,

$$n^{-1} \log \alpha_n^{Q^*} = -a$$

$$n^{-1} \log \beta_n^{Q^*}(\theta_1) = -(\tfrac{1}{3} + \log \tfrac{3}{2}) \approx -.74.$$

We know no general method for computing β_n^Q for tests based on λ_n, however for this specific problem we may compute as follows: A straightforward maximization gives $\lambda_n = (\bar{y}_n^{\frac{1}{2}} - \bar{x}_n^{\frac{1}{2}})^2$ where

$$\bar{x}_n = n^{-1} \sum_{i=1}^n x_i, \quad \bar{y}_n = n^{-1} \sum_{i=1}^n y_i.$$

(Also, one can compute

$$\lambda_n^* = (\bar{x}_n + \bar{y}_n - 2) - \log \bar{x}_n \bar{y}_n,$$

so it is clear that $\lambda_n^* \geqq \lambda_n$ with equality only if $\bar{x} = \bar{y} = 1$.) Using Lemma 3 we compute

$$n^{-1} \log \Pr_{\theta_1}\{\bar{y}_n < c\} \to 1 - (c/4) + \log(c/4) \tag{6.4}$$

and

$$n^{-1} \log \Pr_{\theta_1}\{\bar{x}_n > d\} \to 1 - 4d + \log 4d \tag{6.5}$$

for $\frac{1}{4} < d$, $c < 4$. It is not hard to then see that $n^{-1} \log \Pr_{\theta_1}\{\bar{y}_n^{\frac{1}{2}} - \bar{x}_n^{\frac{1}{2}} \leqq a^{\frac{1}{2}}\}$ has the limiting value determined by finding the infimum over c, d satisfying

$$c^{\frac{1}{2}} - d^{\frac{1}{2}} = a^{\frac{1}{2}} \quad \tfrac{1}{4} < d, c < 4 \tag{6.6}$$

of the maximum of (6.4) and (6.5). This, "inf max", occurs when

$$1 - c/4 + \log(c/4) = 1 - 4d + \log 4d. \tag{6.7}$$

Solving (6.6) and (6.7) graphically for the above value of a gives $c \approx .94$, $d \approx .671$ and $n^{-1} \log \beta_n^{Q}(\theta_1) \approx -.69$ as compared to the best value, $-.74$, given above.

In many situations Corollary 1 can be applied when the extra information concerns the variance in normal testing problems. A simple example follows.

EXAMPLE 2. Let $\mathscr{X} = (-\infty, \infty)$; ν = Lebesgue measure; the parameter space $\Theta' = \{(\mu, \sigma): \mu = \sigma, \mu > 0\}$; and let $f(\cdot, (\mu, \sigma))$ be the normal density with mean μ and variance σ^2. Let $\Theta_0' = \{(\mu, \sigma): \mu = 1, (\mu, \sigma) \in \Theta'\} = \{(1, 1)\}$.

We have written Θ' instead of Θ, etc., to facilitate the discussion in the next example. Θ' may be imbedded directly in $\Theta'^* = \{(\mu, \sigma): \sigma > 0\}$ and we take $\Theta_0'^* = \{1, 1\} = \Theta_0$. (Note: The natural parameter space, in which the points of Θ can be properly geometrically visualized, of course actually has coordinates $(\mu/\sigma^2, 1/\sigma^2)$.)

The statistic λ_n can be expressed directly in terms of $\bar{x}$ and $s_x^2 = n^{-1}\Sigma(x_1 - \bar{x})^2$, however the expression is a little messy and we omit it. The main point is that for forming likelihood ratio statistics it is (asymptotically non-locally) strictly better to ignore the information that $\sigma = \mu$. From this point of view, rather than using λ_n, one should use $\lambda_n'^*$ which is easily determined to be

$$\lambda_n'^* = [(\bar{x} - 1)^2 + s_x^2 - \log s_x^2 - 1]/2.$$

For example, a test Q'^* based on $\lambda_n'^*$ satisfying $\lim n^{-1} \log \alpha_n = -a$ has rejection region:

$$\text{Reject if } \lambda_n'^* > a.$$

EXAMPLE 3. Suppose the original parameter space is $\Theta = \Theta'^*$ where Θ'^* is as in Example 2, except suppose that the null hypothesis is

$$\Theta_{01} = \{(\mu, \sigma): \mu = 1 \quad \text{or} \quad \sigma = 1\},$$

and $\Theta_{11} = \Theta - \Theta_{01}$. Since we also want to consider two other related problems we use the subscript one in the above, and we use α_{n1}, β_{n1}, etc. to denote the error probabilities for this problem. The other two problems have Θ as above, but have null hypothesis $\Theta_{02} = \{(\mu, \sigma): \mu = 1\}$ and $\Theta_{03} = \{(\mu, \sigma): \sigma = 1\}$, respectively. For testing Θ_{02} versus $\Theta_{12} = \Theta - \Theta_{02}$ we have the optimal statistic

$$\lambda_{n2}^* = 2^{-1} \log [1 + (\bar{x}-1)^2/s_x^2]$$

which is equivalent to the usual t-statistic. For testing Θ_{03} we have

$$\lambda_{n3}^* = (s_x^2 - \log s_x^2 - 1)/2.$$

For testing $\Theta_{01} = \Theta_{02} \cup \Theta_{03}$ the optimal statistic is $\lambda_{n1}^* = \min(\lambda_{n2}^*, \lambda_{n3}^*)$. The tests which reject when these statistics are $> a$ have levels satisfying $n^{-1} \log \alpha_{ni} \to -a$, $i = 1, 2, 3$, for their respective problems.

To illustrate Corollary 2, suppose one is provided with "extra" information that in fact the null hypothesis is Θ_0' as defined in Example 2. Let $S_{0i}(\cdot)$, $i = 1, 2, 3$, be defined by (6.1) for their respective problems. In theory it is possible to find S_{0i}, explicitly, but the algebra is difficult. However, without carrying the computation far it is easy to see that there are no values of $\theta_1 \in \Theta_1'^*$ such that $(1, 1) \in S_{01}(\theta_1)$ and very few (if any (?)) values of $\theta_1 \in \Theta_1'^*$ such that $(1, 1) \in S_{0i}(\theta_1)$, $i = 2, 3$. Corollary 2 says that, except for these few (if any) values of θ_1, β_n' defined by the test of Example 2 satisfies

$$\lim_{n\to\infty} n^{-1} \log \beta_n'(\theta_1) < \lim_{n\to\infty} n^{-1} \log \beta_{ni}(\theta_1) \qquad i = 1, 2, 3,$$

and thus is asymptotically non-locally better than the three tests of the preceding paragraph.

For the tests based on λ_{n2}^* and λ_{n3}^* there is nothing more to add since these tests are obviously similar over their respective null hypothesis. However the test defined above based on λ_{n1}^* is not similar. In fact, $\alpha_{n1} = \Pr_{\theta_0}\{\lambda_{n3}^* > a\} \times \Pr_{\theta_0}\{\lambda_{n2}^* > a \mid \lambda_{n3}^* > a\}$, so that

$$\lim_{n\to\infty} n^{-1} \log \alpha_{n1} = -a - a(e^a - 1)/2 = -a(e^a + 1)/2.$$

Hence, a test based on λ_{n1}^* satisfying $\lim_{n\to\infty} n^{-1} \log \alpha_{n1}((1, 1)) = -a$ is actually

$$\text{Reject if } \lambda_{n1}^* > b$$

where $b(e^b + 1) = 2a$ $(b < a)$. Corollary 2 *does not* answer the question whether this test has the same asymptotic non-local properties for testing Θ_0' as the optimal test defined in Example 2. We do not know the answer to this question, though we suspect it to be "No".

Example 2 illustrates (among other things) how a convenient non-local asymptotically optimal test may be found by imbedding the original problem in a larger problem. It is to be noted that even where the likelihood ratio test for the original problem is optimal it may still be more convenient to use a more easily computed optimal statistic for a larger problem. This idea can be used, for example, to motivate using appropriate MANOVA tests for non-standard or standard model II ANOVA situations.

In this regard, Example 3 then illustrates that in this imbedding process only the alternative hypothesis should be enlarged.

7. On weakening the absolute continuity assumption. It was briefly mentioned in Section 1, that the regularity assumptions under which Theorems 1–3 are proved are somewhat restrictive. In particular, it would be desirable to relax the assumption that $f(\cdot, \theta_1) > 0$ a.e. (ν) and, further, to also remove Assumption 4 entirely.

However, if either of these assumptions is removed entirely then the conclusion of Theorem 1b is not always true. We deal mainly with the case where Assumption 4 is satisfied. Consider the following extended example.

EXAMPLE 4. Let ν be Lebesgue measure on $\mathcal{X} = (-\infty, \infty)$ and $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$
where

$$\begin{aligned} f(x, \theta_0) &= \tfrac{1}{2} && 0 \leqq x \leqq 2; \\ &= 0 && \text{otherwise.} \\ f(x, \theta_1) &= 2x && 0 \leqq x \leqq 1; \\ &= 0 && \text{otherwise.} \end{aligned}$$

Assumptions 1–4 are clearly satisfied.

$$\begin{aligned} K(\theta_1, \theta_0) &= \textstyle\int_0^1 (\log 2x + \log 2) 2x \, dx \\ &= 2 \log 2 - \tfrac{1}{2} \\ K((\gamma, \theta_0, \theta_1), \theta_0) &= \log 2(1+\gamma) - \gamma/(1+\gamma) \\ K((\gamma, \theta_0, \theta_1), \theta_1) &= (1-\gamma)/(1+\gamma) + \log((1+\gamma)/2). \end{aligned}$$

(It is immediately clear that the relations (3.7) and (3.8), which are fundamental in the proof of Theorem 1, are not valid here.)

Consider the problem of testing Θ_0 versus Θ_1 at a sequence of levels α_n satisfying $\lim_{n\to\infty} n^{-1} \log \alpha_n = -a$.

Suppose $a = \log 2$, for that is the case which causes the most difficulty. To be more precise, suppose $\alpha_n = e^{-na} = (2)^{-n}$. Then $\varphi(\theta_0) = 0$ and (4.1) of Theorem 1(b) predicts $\beta(\theta_1) = 1 - \log 2$. However, it is evident that since

$$\Pr_{\theta_0}\{0 \leqq x_i \leqq 1 : i = 1, 2, \cdots, n\} = (\tfrac{1}{2})^n,$$

the most powerful level α_n test rejects on the basis of n-observations whenever $0 \leqq x_i \leqq 1$, $i = 1, 2, \cdots, n$; and this test has probability zero of type II error, rather than approximately $e^{-n(1 - \log 2)}$ as apparently predicted by the conclusion in Theorem 1(b).

Note that if $\alpha_n = k(n)2^{-n}$ where $\limsup_{n\to\infty} k(n) < 1$ and $\liminf_{n\to\infty} k(n) > 0$

we may apply the previously used result of Rao (1962) to the problem of testing at significance level $k(n)$ the distribution

$$f(x, \theta_0') = 1 \qquad 0 \leqq x \leqq 1;$$
$$= 0 \qquad \text{otherwise;}$$

versus $f(\cdot, \theta_1)$. ($f(\cdot, \theta_0')$ is the conditional density of X given θ_0 and $0 \leqq x \leqq 1$.) We derive that the most powerful level $\alpha_n = k(n)2^{-n}$ test of $\{\theta_0\}$ versus $\{\theta_1\}$ has a probability $\beta_n{}^T$ of type II error satisfying $\lim_{n\to\infty} n^{-1} \log \beta_n{}^T = -(1-\log 2)$, as predicted by Theorem 1(b).

In general, we see that if $K(\theta_1, \theta) > a > 1-\log 2 = K((0, \theta_0, \theta_1), \theta_0)$ then $\varphi(\theta_0) > 0$ and the value of $\beta_n{}^T$ for the best level α_n test satisfies $\lim_{n\to\infty} n^{-1} \cdot \log \beta_n{}^T = -\beta(\theta_1)$ as in Theorem 1(b). If $a < 1-\log 2$ then $\varphi(\theta_0)$ fails to exist and $\beta_n{}^T = 0$ for sufficiently large n. But, if $a = 1-\log 2$ then $\phi(\theta_0) = 0$ and $\lim_{n\to\infty} n^{-1} \log \beta_n{}^T$ may or may not exist, depending on the precise values of α_n. (In this case it will be true that

$$-\infty \leqq \liminf_{n\to\infty} n^{-1} \log \beta_n{}^T \leqq \limsup_{n\to\infty} n^{-1} \log \beta_n{}^T \leqq$$
$$-K((0, \theta_0, \theta_1)\theta_0) = -(1-\log 2).)$$

A rather lengthy, though straightforward, revision using the above ideas of Lemma 2 and the proof of Theorem 1 gives the result below. The result is of more limited applicability than Theorem 1 because of the special additional hypothesis that $\varphi(\mu_0) > 0$ (see below). We omit the proof.

THEOREM 4. *Suppose the hypotheses of Theorem* 1 *are satisfied except that* $f(\cdot, \theta_1) \not\equiv 0$ a.e. (ν). *Suppose also that there is a* $\theta_0 \in \Theta_0$ *such that* $K((0, \theta_0, \theta_1), \theta_0) < a$. *Let*

$$P = \{\theta \in \Theta_0{}^*: K((0, \theta, \theta_1), \theta) \leqq a\}.$$

For $\theta \in P$ *define* $\zeta(\theta)$ *as in Theorem* 1(b) [i.e. $K((\zeta(\theta), \theta, \theta_1), \theta) = a$.] *Define*

$$b(\theta_1) = \inf_{\theta \in P} K((\zeta(\theta), \theta, \theta_1), \theta_1).$$

Suppose there exists a $\mu_0 \in \Theta_0{}^*$ *such that* $\zeta(\mu_0) > 0$ *and* $K((\zeta(\mu_0), \mu_0, \theta_1), \theta_1) = b(\theta_1)$. *Then*

(a) $\lim_{n\to\infty} n^{-1} \log \beta_n(\theta_1) = -\beta(\theta_1)$ *exists*;

(b) $\beta(\theta_1) = b(\theta_1)$;

(c) $C_n{}^* \to a$;

and

(d) $\lim_{n\to\infty} n^{-1} \log \beta_n{}^{Q^*}(\theta_1) = -\beta(\theta_1)$.

Note that, by Lemma 1, P is closed. Easily established continuity properties of K then prove the existence of a $\mu_0 \in P$ such that $K((\zeta(\mu_0), \mu_0, \theta_1), \theta_1) = b$. However, as in Example 4, it is quite possible that $\zeta(\mu_0) = 0$.

Before proceeding further we point out that Theorem 4 can be used to obtain analogs of Theorems 2 and 3 valid under appropriate added restrictions. Also, we point out that to adequately handle the case where $\zeta(\mu_0) = 0$ would require estimates on probabilities of large deviations of a different nature than those used elsewhere in this paper.

In the cases where Assumption 4 is violated, it is easy to see again that Theorem 1(b) must be modified. The distributions of Example 4 can, in fact, demonstrate this if we let the null hypothesis, Θ_0', say, be defined by $\Theta_0' = \{\theta_1\}$ and let $\Theta_1' = \{\theta_0\}$. Since the null hypothesis is simple in this situation, Theorem 4 can be applied directly to this problem by reversing the roles of Θ_0 and Θ_1.

However, in the general case, where the null hypothesis is not simple, the situation may be much more complicated. The reason is that for fixed $\theta_1 \in \Theta_1$ $K((0, \theta_0, \theta_1), \theta_0)$ may take on a continuum of values as θ_0 varies in Θ_0. Thus, a direct extension of Theorem 4 often fails to yield any positive results.

REFERENCES

[1] BAHADUR, R. R. (1966). An optimal property of the likelihood ratio statistic. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 13–26.

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

[3] BAHADUR, R. R. (1969). Private Communication.

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

[5] HERR, D. G. (1967). Asymptotically optimal tests for multivariate normal distributions. *Ann. Math. Statist.* **38** 1829–1844.

[6] 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.

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

[8] NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. *Probability and Statistics. The Harold Cramér Volume.* Wiley, New York, 213–234.

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

[10] RAO, C. R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. Roy. Statist. Soc. Ser. B.* **24** 46–72.