The Annals of Statistics
1978, Vol. 6, No. 1, 72–84

LARGE DEVIATIONS OF LIKELIHOOD RATIO STATISTICS WITH APPLICATIONS TO SEQUENTIAL TESTING[1]

BY MICHAEL WOODROOFE

The University of Michigan

We study the tail of the null distribution of the log likelihood ratio statistic for testing sharp hypotheses about the parameters of an exponential family. We show that the classical chisquare approximation is of exactly the right order of magnitude, although it may be off by a constant factor. We then apply our results and techniques to find the error probabilities of a sequential version of the likelihood ratio test. The sequential version rejects if the likelihood ratio statistic crosses a given barrier by a given time. Our approach uses a local limit theorem which takes account of large deviations and integrates the local result by using the theory of Hausdorff measures.

1. Introduction. Let $\mathscr{P} = \{P_\omega : \omega \in \Omega\}$ denote a p-dimensional exponential family, say,

$$\frac{dP_\omega}{d\mu} = \exp\{\omega' x - \psi(\omega)\}\,, \qquad x \in R^p,\ \omega \in \Omega\,,$$

with respect to a sigma finite measure μ on $\mathscr{B}(R^p)$. Here $'$ denotes transpose, and Ω denotes the natural parameter space of the family. *We suppose throughout this paper that* Ω *is an open subset of* R^p *and that* ψ *is strictly convex on* Ω. The former condition is a minor restriction, but the latter is not since it may always be achieved by an appropriate reparameterization.

Recall that if $X \sim P_\omega$, then $E_\omega(X) = \nabla\psi(\omega)$ and $\mathrm{Cov}_\omega(X) = \nabla^2\psi(\omega)$, where ∇ and ∇^2 denote gradient and Hessian, respectively. Let $\Gamma = \nabla\psi(\Omega)$ be the set of possible expectations of the family and observe that

$$\phi(x) = \sup\nolimits_\Omega \omega' x - \psi(\omega) \tag{1}$$

is finite for all $x \in \Gamma$. In fact, the supremum in (1) is attained uniquely at $\hat\omega = \hat\omega(x)$, where $\nabla\psi(\hat\omega) = x$.

Next let Ω_0 be a q-dimensional, C^2 submanifold of Ω with $0 \leqq q < p$ and let

$$\phi_0(x) = \sup\nolimits_{\Omega_0} \omega' x - \psi(\omega)\,, \qquad x \in \Gamma\,. \tag{2}$$

If $X_1, X_2, \cdots$ are i.i.d. with common distribution P_ω for some unknown $\omega \in \Omega$, then the (logarithm of the inverse of the) likelihood ratio statistic for testing $H_0 : \omega \in \Omega_0$ is

$$\Lambda_n = n[\phi(n^{-1}S_n) - \phi_0(n^{-1}S_n)]\,,$$

Received December 1976; revised June 1977.

[1] Research supported by the National Institutes of Health under GM 20507.

AMS 1970 *subject classifications*. Primary 62E20; Secondary 62G10.

Key words and phrases. Large deviations, local limit theorems, Hausdorff measures, exponential families, sequential tests, conditional probabilities.

where $S_n = X_1 + \cdots + X_n$, $n \geqq 1$. Of course, in order for Λ_n to be meaningful, we need an additional condition. *We suppose throughout the paper that for some integer* n_0, $n^{-1}S_n \in \Gamma$ w.p. 1 *for all* $n \geqq n_0$. Then Λ_n is well defined for all $n \geqq n_0$.

It is well known that the null distribution of $2\Lambda_n$ converges weakly to a chi-square distribution on $r = p - q$ degrees of freedom. Here we will supplement this information by finding the exact rate at which $P_\omega^\infty\{\Lambda_n \geqq n\varepsilon\}$ tends to zero as $n \to \infty$ for fixed $\varepsilon > 0$ and for $\varepsilon = \varepsilon_n \to 0$ with $n\varepsilon_n \to \infty$ when $\omega \in \Omega_0$. In fact, we shall prove the following (under conditions which are detailed in Section 2): if $\omega \in \Omega_0$, then for sufficiently small $\varepsilon > 0$,

$$P_\omega^\infty\{\Lambda_n \geqq n\varepsilon\} \sim c(n\varepsilon)^{\frac{1}{2}r-1}e^{-n\varepsilon} \tag{3}$$

as $n \to \infty$, where $r = p - q$ and $c = c(\omega, \varepsilon)$ is a positive constant which is defined in Section 3. The conditions require the existence and uniqueness of the maximum likelihood estimator under the assumption that H_0 is true and the applicability of the local limit theorem.

It is of interest to compare (3) with the estimate suggested by the asymptotic distribution of Λ_n, namely

$$P\{\tfrac{1}{2}\chi_r^2 \geqq n\varepsilon\} \sim \Gamma(\tfrac{1}{2}r)^{-1}(n\varepsilon)^{\frac{1}{2}r-1}e^{-n\varepsilon}$$

as $n \to \infty$. It is especially interesting that this incorrect estimate tends to zero at exactly the same rate as the correct one (3), although the constants c and $\Gamma(\frac{1}{2}r)^{-1}$ may differ. We will also prove (under the same conditions): if $\varepsilon = \varepsilon_n \to 0$ with $n\varepsilon_n \to \infty$, then

$$P_\omega^\infty\{\Lambda_n \geqq n\varepsilon\} \sim \Gamma(\tfrac{1}{2}r)^{-1}(n\varepsilon)^{\frac{1}{2}r-1}e^{-n\varepsilon} \tag{4}$$

as $n \to \infty$. That is, the chi-square approximation is accurate for deviations of order $o(n)$!

We shall also consider a sequential version of the likelihood ratio test. Let $m \geqq n_0$ be an integer and let

$$t = t_a = \inf\{n \geqq m : \Lambda_n \geqq a\} \tag{5}$$

for $a > 0$. If $N > m$ is an integer, we may then form a sequential test of H_0 by taking $\tau = \min\{t, N\}$ observations and deciding for H_0 if, and only if, $t > N$. We prove (under slightly more restrictive conditions): if $\omega \in \Omega_0$, $m \sim a\delta_2^{-1}$, and $N \sim a\delta_1^{-1}$, where $0 < \delta_1 < \delta_2$, then

$$P_\omega^\infty\{t \leqq N\} \sim Ka^{\frac{1}{2}r}e^{-a} \tag{6}$$

as $a \to \infty$, where $K = K(\omega, \delta_1, \delta_2)$ is a positive constant which is defined in Section 4.

We are aware of several related papers. Borovkov and Rogozin (1965) have developed some general techniques for estimating probabilities of large deviations, and we shall use their results. In the special case that H_0 is simple, (3) is an easy consequence of their work. Efron and Truax (1968) have given an expression for the power of the likelihood ratio test against nonlocal alternatives, again in the case of a simple hypothesis; and Brown (1971) has given an expression for $\log P_\omega^\infty\{\Lambda_n \geqq n\varepsilon\}$ in a more general context than the one which we consider.

The form of the stopping time $t = t_a$ was inspired by the papers of Armitage (1957) and Schwarz (1962, 1968). We include an example which shows how our results may be used to approximate the error probabilities of Schwarz' approximation to optimal Bayesian sequential tests. Analogues of relation (6) have been developed by Woodroofe (1976) and Berk (1976) in the context of a normal distribution with unknown mean and known variance and by Lai and Siegmund (1977) in the context of a one-parameter exponential family.

2. Preliminaries. Our main results (3), (4), and (6) require the computation of probabilities for fixed, but arbitrary, $\omega \in \Omega_0$, say $\omega = \omega^0$. Some simplification may be obtained by supposing

$$\omega^0 = 0\,, \quad \nabla\psi(0) = 0\,, \qquad \text{and} \qquad \psi(0) = 0\,. \tag{7}$$

Moreover, there is no loss of generality in making this assumption, since it may be achieved by a simple reparameterization. Hence, *we suppose throughout that* (7) *holds*.

Equation (7) and the italicized assumptions in the introduction are standing assumptions. They will not be repeated in the statements of our lemmas and theorems. In addition, it will be convenient to have names for two optional assumptions.

M: for all $x \in \Gamma$, the supremum in (2) is attained uniquely at a point $\hat\omega_0 = \hat\omega_0(x) \in \bar\Omega_0 \cap \Omega$. Here $^-$ denotes closure in R^p.

L: for some integer n, S_n has a bounded continuous density f_0^n with respect to Lebesgue measure on $\mathscr{B}(R^p)$.

In condition M, it is easy to see that the uniqueness of $\hat\omega$ and $\hat\omega_0$ imply their continuity.

If condition L is satisfied and if n_1 denotes the minimum value of n for which S_n has a bounded continuous density, then S_n has a bounded continuous density for all $n \geqq n_1$. The result of Borovkov and Rogozin (1965) may now be stated.

PROPOSITION 1. *If conditional L is satisfied, then as $n \to \infty$*

$$f_0^n(nx) \sim (2\pi n)^{-\frac{1}{2}p}|\Sigma(x)|^{-\frac{1}{2}}e^{-n\phi(x)}\,, \tag{8}$$

where $\Sigma(x) = \nabla^2\psi[\hat\omega(x)]$. *The limit in* (8) *is attained uniformly on compact subsets of* Γ.

We will need some properties of the Hausdorff measures H_p^r, $0 \leqq r \leqq p$, in R^p. For $G \subset R^p$, let $|G|$ denote the diameter of G and for $B \subset R^p$ let

$$H_p^r(B) = \lim_{\varepsilon\to 0} 2^{-r}\Delta_r \inf\{\textstyle\sum_{i=1}^\infty |G_i|^r : B \subset \bigcup_{i=1}^\infty G_i,\ |G_i| < \varepsilon,\ i \geqq 1\}\,,$$

where $\Delta_r = \Gamma(\frac{1}{2})^r\Gamma(\frac{1}{2}r + 1)^{-1}$ denotes the volume of the unit ball in R^r. Then $H_p{}^r$ is an outer measure and all Borel sets are $H_p{}^r$-measurable. $H_p{}^0$ is counting measure and $H_p{}^p$ is Lebesgue measure L^p.

The following result is specialized from Federer (1969, Theorems 3.2.5 and 3.2.22). Let $p, q \geqq 1$, let $W \in \mathscr{B}(R^p)$, and let $g: R^p \to R^q$ be continuously differentiable on a neighborhood of W. Further, let

$$Dg(x) = \left[\frac{\partial g_i}{\partial x_j}\right] \qquad (q \times p),$$

and

$$J_g(x) = |Dg(x)Dg(x)'| : q \leqq p, \\ = |Dg(x)'Dg(x)| : q \geqq p.$$

Finally, let $n(g, y) =$ cardinality $\{x \in W: g(x) = y\}$ for $y \in R^q$.

PROPOSITION 2. *Let g be as above. If $p < q$ and if $f: g(W) \to R^1$ is a bounded measurable function, then*

$$\int_W f \circ g \cdot J_g \, dL^p = \int_{g(W)} f(y) n(g, y) \, dH_q{}^p(y). \tag{9}$$

If $p \geqq r \geqq q$ and if $f: W \to R^1$ is a nonnegative measurable function, then

$$\int_W f \cdot J_g \, dH_p{}^r = \int_{g(W)} [\int_{g^{-1}\{y\}} f(x) \, dH_p{}^{r-q}(x)] \, dL^q(y). \tag{10}$$

It follows from (9) that if $S \subset R^q$ is a nice p-dimensional surface in R^q, then $H_q{}^p(S)$ is just the p-dimensional surface areas of S as defined, for example, by Edwards (1973), pages 330–344.

We may now state the following corollary.

COROLLARY 1. *Suppose that condition L is satisfied and let u be a positive bounded measurable function on Γ. If $0 < \varepsilon_n \to \varepsilon \geqq 0$ and $n\varepsilon_n \to \infty$, then for any $\delta > 0$*

$$\begin{aligned} &\int_{\Lambda_n \geqq n\varepsilon_n} u(n^{-1}S_n) \, dP_0{}^\infty \\ &\quad = \frac{n^{\frac{1}{2}p-1}}{(2\pi)^{\frac{1}{2}p}} \cdot e^{-n\varepsilon_n} \left\{\int_0^{n\delta} I\left(\varepsilon_n, \frac{y}{n}\right) e^{-y} \, dy[1 + o(1)] + O(e^{-n\delta})\right\} \end{aligned} \tag{11}$$

as $n \to \infty$, where

$$I(\varepsilon, s) = \int_{\phi = \varepsilon + s, \phi_0 \leqq s} u(x) j(x)^{-1} \, dH_p{}^{p-1}(x), \tag{12}$$

$$j(x) = |\Sigma(x)|^{\frac{1}{2}} ||\hat{\omega}(x)||. \tag{13}$$

This is effectively Theorem 2 of Borovkov and Rogozin (1965), except that the Jacobian term $J_\phi(x) = ||\hat{\omega}(x)||$ was apparently omitted there. The corollary may be formally verified by combining (8) and (10) and making an appropriate change of variables.

3. Large deviations. In this section we give precise statements of (3) and (4). The definition of c and the quantification of "sufficiently small" in (3) require some preliminary notation.

We first suppose that Ω_0 is a q-dimensional, C^2 submanifold of Ω with $0 \leqq q < p$. We may then choose local coordinates $\alpha : \Theta \to \Omega_0$ with $\alpha(0) = 0$. Here Θ is an open subset of R^q and α is a one-to-one, twice continuously differentiable function on Θ for which $D\alpha(\theta)$ is of full rank for all $\theta \in \Theta$, if $q \geqq 1$. If $q = 0$, then Θ is a singleton.

Suppose now that condition M is satisfied and let $\Gamma_1 = \{x \in \Gamma : \hat{\omega}_0(x) \in \alpha(\Theta)\}$. Then Γ_1 is open in R^p and $0 \in \Gamma_1$. Moreover, for $x \in \Gamma_1$, $\hat{\omega}_0(x) = \alpha[\hat{\theta}(x)]$, where $\hat{\theta}(x)$ is the (unique) point in Θ at which $g(x, \theta) = x'\alpha(\theta) - \psi \circ \alpha(\theta)$ attains its maximum. If $q \geqq 1$, let $Q(x, \theta) = -\nabla_\theta{}^2 g(x, \theta)$, so

$$Q(x, \theta) \doteq D\alpha(\theta)'\nabla^2\psi[\alpha(\theta)]D\alpha(\theta) - \sum_{i=1}^p \left\{x_i - \frac{\partial\psi}{\partial\omega_i}[\alpha(\theta)]\right\} \nabla^2\alpha_i(\theta)$$

for $x \in \Gamma_1$ and $\theta \in \Theta$; and let Γ_2 be the set of $x \in \Gamma_1$ for which $Q(x, 0)$ is positive definite. If $q = 0$, let $\Gamma_2 = \Gamma_1$. In either case, Γ_2 is open in R^p and $0 \in \Gamma_2$.

To quantify "sufficiently small," let $B_\varepsilon = \{x \in \Gamma : \phi(x) \leqq \varepsilon\}$ for $\varepsilon > 0$ and observe that each B_ε is a compact neighborhood of $0 \in R^p$. Next let

$$\varepsilon_1 = \sup\{\phi(x) : \hat{\omega}_0(x) = 0\}, \qquad \varepsilon_2 = \sup\{\varepsilon : B_\varepsilon \subset \Gamma_2\}, \qquad \text{and}$$
$$\varepsilon_0 = \min\{\varepsilon_1, \varepsilon_2\}.$$

Then, $0 < \varepsilon_0 \leqq \infty$ and for $0 < \varepsilon < \varepsilon_0$,

$$M_\varepsilon = \{x \in \Gamma : \hat{\omega}_0(x) = 0, \phi(x) = \varepsilon\} \tag{14}$$

is an $(r - 1)$-dimensional submanifold of Γ, where $r = p - q$.

THEOREM 1. *Suppose that Ω_0 is a q-dimensional, C^2 submanifold of Ω, where $0 \leqq q < p$, and that conditions M and L are satisfied. If u is a positive, bounded, continuous function on Γ and $0 < \varepsilon < \varepsilon_0$, then as $n \to \infty$,*

$$\int_{\Lambda_n \geqq n\varepsilon} u(n^{-1}S_n)\, dP_0{}^\infty \sim c(u, \varepsilon)(n\varepsilon)^{\frac{1}{2}r-1}e^{-n\varepsilon}, \tag{15}$$

where

$$c(u, \varepsilon) = (2\pi)^{-\frac{1}{2}r}\varepsilon^{-\frac{1}{2}r+1} \int_{M_\varepsilon} u(x)k(x)j(x)^{-1}\, dH_p{}^{r-1}(x)$$

with

$$k(x) = |Q(x, 0)|^{\frac{1}{2}}|D\alpha(0)'D\alpha(0)|^{-\frac{1}{2}}$$

if $q \geqq 1$ and $k(x) = 1$ if $q = 0$. Here j is as in (13). *Moreover, the limit in* (15) *is attained uniformly on compact subintervals of* $(0, \varepsilon_0)$.

REMARKS. 1. Of course, (15) contains (3) as a special case with $c = c(1, \varepsilon)$.

2. If $\varepsilon > \varepsilon_1$, then the left side of (15) is of exponentially smaller order than $e^{-n\varepsilon}$. For then $\phi - \phi_0 \geqq \varepsilon$ implies $\phi \geqq \varepsilon'$ for some $\varepsilon' > \varepsilon$, so the result follows by applying Theorem 1 to $H_0 : \omega = 0$, or directly from Theorem 2 of [3].

3. The definition of ε_0 depends on the local coordinates α through $\alpha(\Theta)$. It is desirable to choose α in such a manner that $\alpha(\Theta)$ is as large as possible.

4. If Ω_0 is a linear hypothesis, say $\Omega_0 = \{A\theta : \theta \in \Theta\}$, where Θ is an open subset of R^q and A is of full rank, then $Q(x, \theta) = A'\nabla^2\psi[A\theta]A$ is positive definite

for all $x \in \Gamma$ and all $\theta \in \Theta$. It follows that $\Gamma_1 = \Gamma_2 = \Gamma$ and that $\varepsilon_0 = \sup \{\phi(x) : \hat{\omega}_0(x) = 0\}$. Moreover, it is easily seen that $\hat{\omega}_0(x) = 0$ if, and only if, $A'x = 0$, so that $\varepsilon_0 = \infty$ if $\Gamma \cap$ null space (A') is unbounded.

THEOREM 2. *Suppose Ω_0 is a q-dimensional, C^2 submanifold of Ω with $0 \leqq q < p$ and that conditions M and L are satisfied. If $\varepsilon = \varepsilon_n \to 0$ with $n\varepsilon_n \to \infty$, then as $n \to \infty$,*

$$P_0^\infty\{\Lambda_n \geqq n\varepsilon\} \sim \Gamma(\tfrac{1}{2}r)^{-1}(n\varepsilon)^{\frac{1}{2}r-1}e^{-n\varepsilon} . \tag{16}$$

The main steps in the proof of Theorem 1 are outlined in the appendix. The proof of Theorem 2 is similar and will be omitted.

EXAMPLE 1. Suppose that $Y = (Y^1, Y^2)$ where Y^1 and Y^2 are independent exponentially distributed random variables with means $1/\theta_1$ and $1/\theta_2$ for some unknown values of $\theta_1 > 0$ and $\theta_2 > 0$ and that we wish to test H_0: $\theta_1 = \theta_2$. In this case the null distribution of Λ_n is independent of the common mean. We study it when $\theta^0 = (1, 1)'$. The transformation $x^i = y^i - 1$ and $\omega_i = 1 - \theta_i$, $i = 1, 2$, then reduces the problem to one in which (7) is satisfied with $\Gamma = (-1, \infty)^2$ and $\Omega = (-\infty, 1)^2$. It is easy to check that conditions M and L are satisfied and that $\varepsilon_0 = \infty$. Moreover, simple calculations yield

$$c(1, \varepsilon) = \{\pi\varepsilon^{-1}(1 - e^{-\varepsilon})(2 - e^{-\varepsilon})\}^{-\frac{1}{2}} .$$

For $\varepsilon \leqq 2$, $c(1, \varepsilon)$ is remarkably close to $c(1, 0) = \pi^{-\frac{1}{2}}$. The ratio $c(1, \varepsilon)/c(1, 0)$ decreases to a minimum of .9547+ near $\varepsilon = .46$ and then increases. At $\varepsilon = 2$, its value is only 1.1138. Of course, $c(1, \varepsilon) \to \infty$ as $\varepsilon \to \infty$.

4. A sequential likelihood ratio test. Let $g = \phi - \phi_0$, so that $\Lambda_n = ng(n^{-1}S_n)$, and let

$$t = t_a = \inf \{n \geqq m : ng(n^{-1}S_n) \geqq a\} \tag{17}$$

for $a > 0$, where $m \geqq n_0$. Further, let $N > m$ be an integer. Then we may form a sequential test of H_0 by taking $\tau = \min \{t, N\}$ observations and deciding in favor of H_0 if, and only if, $t > N$. We will study the probability of a type I error, $\beta = P_0^\infty\{t \leqq N\}$ as $a \to \infty$ under the assumption that $m \sim a\delta_2^{-1}$ and $N \sim a\delta_1^{-1}$, where $0 < \delta_1 < \delta_2 < \varepsilon_0$. *We suppose throughout this section that condition L is satisfied.*

To motivate our approach, we observe that

$$P_0^\infty\{t = n\} = \int_{\Lambda_n \geqq a} u_{n,a}(n^{-1}S_n)\, dP_0^\infty \tag{18}$$

where

$$u_{n,a}(y) = P\{t > n - 1 \mid n^{-1}S_n = y\} .$$

The right side of equation (18) is of the same form as the left side of equation (15) with $u = u_{n,a}$ and $n\varepsilon = a$. Thus we need an estimate for the conditional probability $u_{n,a}$.

Recall that S_n has a bounded continuous density f_0^n for $n \geqq n_1$. Let h be any density with respect to $\mu = P_0$ and define $h_{n,k}$ by

$$h_{n,k}(x_1, \cdots, x_k \mid y) = f_0^{n-k}[ny - \sum_{i=1}^k x_i]/f_0^n(ny)$$

if $f_0^n(ny) > 0$ and $h_{n,k}(x_1, \cdots, x_k \mid y) = \prod_{i=1}^k h(x_i)$ if $f_0^n(ny) = 0$. Here $x_i \in R^p$, $i = 1, \cdots, k$, $y \in \Gamma$, $k \leqq n - n_1$ and $n > n_1$. Then $h_{n,k}$ is a conditional density for $X_1, \cdots, X_k$ given $n^{-1}S_n$ for $1 \leqq k \leqq n - n_1$ and $n > n_1$. The corresponding distribution will be denoted by $R_{n,k}$, so that $R_{n,k}(y, dx_1, \cdots, dx_k) = h_{n,k}(x_1, \cdots, x_k \mid y)\, d\mu(x_1) \cdots d\mu(x_k)$.

LEMMA 1. *As $n \to \infty$, $R_{n,k}(y, \cdot)$ converges strongly to $P_{\hat\omega}^k$, where $\hat\omega = \hat\omega(y)$, for $k = 1, 2, \cdots$; and the convergence is uniform on compact subsets of Γ.*

PROOF. By Proposition 1 and some simple algebra, $h_{n,k}(x_1, \cdots, x_k \mid y) \to \prod_{i=1}^k \exp\{\hat\omega' x_i - \psi(\hat\omega)\}$, uniformly (in y) on compact subsets of Γ for fixed $x_1, \cdots, x_k$. The lemma follows easily.

For $1 \leqq k \leqq n - n_1$ and $n > n_1$, let $R_{n,k}^*(y, dx_1, \cdots, dx_k) = R_{n,k}[y, (dx_1 + y) \cdots (dx_k + y)]$, so that $R_{n,k}^*$ is a version of the conditional distribution of $X_i - n^{-1}S_n$, $i = 1, \cdots, k$, given $n^{-1}S_n$. Further, let $T_k(x_1, \cdots, x_k) = x_1 + \cdots + x_k$ for $x_i \in R^p$, $i = 1, \cdots, k$. Then

$$u_{n,a}(y) = R_{n,k}^*[y, \{(n-j)g[y - (n-j)^{-1}T_j] - a < 0, 1 \leqq j \leqq n - m\}$$

defines a convenient version of $P\{t > n - 1 \mid n^{-1}S_n\}$ for $n > m \geqq n_1$.

In the remainder of this section we suppose that Ω_0 is a C^2 manifold of dimension $q < p$ and that condition M is satisfied. Then $\nabla g = \hat\omega - \hat\omega_0$ is continuous and $\nabla g = 0$ if, and only if, $g = 0$. Let $\Gamma_0 = \{x \in \Gamma : g(x) = 0\}$.

LEMMA 2. *If $n = n_a$ and y_a vary in such a manner that $y_a \to y \in \Gamma - \Gamma_0$ and $ng(y_a) = a + z + o(1)$ with $z \in R^p$, then $\lim u_{n,a}(y_a) = w(y, z)$ as $a \to \infty$, where*

$$w(y, z) = P_{\hat\omega}^\infty\{\nabla g(y)'[ky - S_k] < kg(y) - z, \text{ for all } k \geqq 1\}\,. \tag{19}$$

Lemma 2 follows from Lemma 1 and the expansion $(n-j)g[y_a - (n-j)^{-1}T_j] - a = -\nabla g(y)'T_j - jg(y) + z + o(1)$ as $a \to \infty$ for each fixed j by an argument similar to that given in Section 6 of [11]. We omit details.

Observe that $y = \nabla\psi(\hat\omega)$ in (19), so that $w(y, z) > 0$ for $0 < z < g(y)$. It is also relevant that w is continuous on $(\Gamma - \Gamma_0) \times R^p$.

The main results of this section may now be stated.

LEMMA 3. *If $n = n_a \to \infty$ as $a \to \infty$ and $\varepsilon_a = an^{-1} \to \varepsilon$, where $0 < \varepsilon < \varepsilon_0$, then*

$$P_0^\infty\{t = n\} \sim c(\bar w, \varepsilon) a^{\frac{1}{2}r - 1} e^{-a} \tag{20}$$

where

$$c(\bar w, \varepsilon) = (2\pi)^{-\frac{1}{2}r} \varepsilon^{-\frac{1}{2}r+1} \int_{M_\varepsilon} \bar w k j^{-1}\, dH_p^{r-1}$$

is as in Theorem 1 and

$$\bar w(x) = \int_0^\infty w(x, y) e^{-y}\, dL^1(y)\,. \tag{21}$$

THEOREM 3. *Suppose that Ω_0 is a C^2 manifold of dimension $q < p$ and that* *conditions M and L are satisfied. If* $m \sim a\delta_2^{-1}$ *and* $N \sim a\delta_1^{-1}$, *where* $0 < \delta_1 < \delta_2 < \varepsilon_0$, *then as* $a \to \infty$

$$P_0^\infty\{t \leqq N\} \sim K(\delta_1, \delta_2) a^{\frac{1}{2}r} e^{-a}$$

where

$$K(\delta_1, \delta_2) = \int_{\delta_1}^{\delta_2} c(\bar{w}, \varepsilon)\varepsilon^{-2}\, d\varepsilon\,.$$

Lemma 3 is proved by showing that $u_{n,a}$ may be replaced by w in equation (18). This is discussed further in the appendix. Theorem 3 then follows by substituting (20) into

$$P_0^\infty\{t \leqq N\} = \sum_{k=m}^{N} P_0^\infty\{t = k\}\,.$$

REMARK 5. If $\varepsilon_0 = \infty$ and if $P_0^\infty\{\Lambda_n \geqq a\} <$ const. $a^{\frac{1}{2}r-1}e^{-a}$ for all $n \geqq m_0$ and a sufficiently large, then one may replace the condition $m \sim a\delta_2^{-1}$ by $m \geqq m_0$. Relation (21) is then true with $\delta_2 = \infty$.

EXAMPLE 2. If X has the p-variate normal distribution with unknown mean $\omega \in R^p$ and covariance matrix I_p, and if $\Omega_0 = \{A\theta : \theta \in R^q\}$ is linear, then

$$c(\bar{w}, \varepsilon) = \Gamma(\tfrac{1}{2}r)^{-1} \exp\{-2 \textstyle\sum_{k=1}^{\infty} \Phi[-(\tfrac{1}{2}k\varepsilon)^{\frac{1}{2}}]\} \tag{22}$$

for $\varepsilon > 0$, where Φ denotes the standard normal distribution function. It is clearly sufficient to verify this assertion in the special case that $H_0: \omega_1 = \cdots = \omega_r = 0$; and a moment's reflection shows that it suffices to consider the case $r = p$. In this case $g(x) = \phi(x) = \frac{1}{2}||x||^2$, so

$$\begin{aligned} w(x, y) &= P_x^\infty\{x'[kx - S_k] < \tfrac{1}{2}k||x||^2 - y,\, k \geqq 1\} \\ &= P\{Z_1 + \cdots + Z_k - \tfrac{1}{2}k||x|| < -||x||^{-1}y,\, k \geqq 1\}\,, \end{aligned}$$

where $Z_1, Z_2, \cdots$ are i.i.d. standard univariate normal random variables. The integral defining $\bar{w}$ may now be evaluated as in Section 5 of [12]; and (22) follows.

5. One-sided hypotheses.

In this section we suppose that Ω_0 is of the form $\Omega_0 = \{\omega \in \Omega : \Delta(\omega) \leqq 0\}$, where $\Delta : \Omega \to R^1$ is a continuous function for which $\Omega_0 \neq \Omega$ and

$$\Omega_0^* = \{\omega \in \Omega : \Delta(\omega) = 0\} \tag{23}$$

is a $(p-1)$-dimensional, C^2 manifold; and we will use the notation

$$\phi_0^*(x) = \sup_{\Omega_0^*} [\omega' x - \psi(\omega)]\,.$$

LEMMA 4. *Let* Ω_0 *be as above. If* $x \in \Gamma$ *and* $\phi(x) > \phi_0(x)$, *then* $\phi_0(x) = \phi_0^*(x)$.

The lemma follows easily from the convexity of ψ.

Let $\Lambda_n^* = n(\phi - \phi_0^*)(n^{-1}S_n)$ be the likelihood ratio statistic for testing H_0^*: $\omega \in \Omega_0^*$ and let $A = \{x \in \Gamma : \phi(\omega) > \phi_0(\omega)\}$. If u is any bounded measurable function on Γ and if $\varepsilon > 0$, then

$$\int_{\Lambda_n \geqq n\varepsilon} u(n^{-1}S_n)\, dP_0^\infty = \int_{\Lambda_n^* \geqq n\varepsilon} u_A(n^{-1}S_n) dP_0^\infty\,, \tag{24}$$

where $u_A = u \cdot I_A$. While u_A need not satisfy the hypotheses of Theorem 1, equation (24) is suggestive.

In the following theorems, α denotes local coordinates for Ω_0^*; and M_ε, ε_0 and k are constructed as in Section 3, but with Ω_0 replaced by Ω_0^*.

THEOREM 1'. *Let Ω_0 be as above and suppose that $0 \in \Omega_0^*$. Suppose also that Ω_0^* satisfies condition M and that condition L is satisfied. If u is any positive, bounded, continuous function on Γ and $0 < \varepsilon < \varepsilon_0$, then*

$$\int_{\Lambda_n \geqq n\varepsilon} u(n^{-1}S_n)\, dP_0^\infty \sim c^*(u, \varepsilon)(n\varepsilon)^{-\frac{1}{2}} e^{-n\varepsilon} \tag{25}$$

as $n \to \infty$, where $c^*(u, \varepsilon) = c(u_A, \varepsilon)$ and $c(\cdot, \cdot)$ is as described in Theorem 1, but with Ω_0 replaced by Ω_0^*. Moreover, the limit in (25) is attained uniformly on compact subintervals of $(0, \varepsilon_0)$.

THEOREM 2'. *Suppose that the hypotheses of Theorem 1' are satisfied. If $0 < \varepsilon = \varepsilon_n \to 0$ with $n\varepsilon \to \infty$ as $n \to \infty$, then*

$$P_0^\infty\{\Lambda_n \geqq n\varepsilon\} \sim \frac{1}{2\pi^{\frac{1}{2}}}(n\varepsilon)^{-\frac{1}{2}} e^{-n\varepsilon}\,. \tag{26}$$

THEOREM 3'. *Suppose that the hypotheses of Theorem 1' are satisfied and define t by (5). If $m \sim a\delta_2^{-1}$ and $N \sim a\delta_1^{-1}$, where $0 < \delta_1 < \delta_2 < \varepsilon_0$, then*

$$P_0^\infty\{t \leqq N\} \sim K^*(\delta_1, \delta_2) a^{\frac{1}{2}} e^{-a}$$

where

$$K^*(\delta_1, \delta_2) = \int_{\delta_1}^{\delta_2} c^*(\bar{w}, \varepsilon)\varepsilon^{-2}\, d\varepsilon$$

and $\bar{w}$ is as described in Lemma 3, but with Ω_0 replaced by Ω_0^*.

REMARKS. 6. Typically, AM_ε is a singleton.

7. If the hypotheses of Theorem 1' are satisfied, then the asymptotic null distribution of Λ_n is that of $\frac{1}{2}\max\{0, Z\}^2$, where $Z \sim N(0, 1)$. Thus, it is again the case that $P_0^\infty\{\Lambda_n \geqq n\varepsilon\}$ and $P\{W \geqq n\varepsilon\}$ tend to zero at the same rate for $0 < \varepsilon < \varepsilon_0$ and are asymptotic if $\varepsilon = \varepsilon_n \to 0$ with $n \to \infty$ as $n\varepsilon \to \infty$.

8. The proofs of Theorems 1', 2', and 3' are similar to those of Theorems 1, 2, and 3 and will be omitted. Theorem 1' may be deduced from Theorem 1 by approximating u_A by continuous functions.

EXAMPLE 3. Let $Y_1, Y_2, \cdots$ be i.i.d. normal random vectors with unknown mean $\theta \in R^p$ and covariance matrix I_p and consider H_0: $||\theta|| \geqq \delta$, where $\delta > 0$ is specified. The distribution of Λ_n then depends only on $||\theta||$. We compute it when $\theta = \theta^0 = (\delta, 0, \cdots, 0)'$. The transformation $x = y - \theta^0$ and $\omega = \theta - \theta^0$ reduces the problem to one to which Theorem 1' applies; and straightforward computations yield

$$c^*(1, \varepsilon) = \frac{1}{2\pi^{\frac{1}{2}}}(1 - \delta^{-1}(2\varepsilon)^{\frac{1}{2}})^{\frac{1}{2}(p-1)}$$

for $0 < \varepsilon < \frac{1}{2}\delta^2$. In this case $c^*(1, \varepsilon)$ and $c^*(1, 0) = 1/2\pi^{\frac{1}{2}}$ are quite different even for moderate values of p and ε.

We conclude with an example which shows how our results may be used to compute power and efficiency as well as size.

EXAMPLE 4. Let $X_1, X_2, \cdots$ be i.i.d. normally distributed random vectors with unknown mean $\omega \in R^p$ and covariance matrix I_p. We consider the problem of testing $H_0\colon \omega = 0$ vs. $H_1\colon ||\omega|| \geqq \delta$, where $\delta > 0$ is specified and the region $0 < ||\omega|| < \delta$ is regarded as an indifference region. Let

$$\Lambda_n{}^0 = \tfrac{1}{2}n\left\|\frac{S_n}{n}\right\|^2 \quad \text{and} \quad \Lambda_n{}^1 = \tfrac{1}{2}n \cdot \max\left\{0, \left[\delta - \left\|\frac{S_n}{n}\right\|\right]\right\}^2$$

denote the likelihood ratio statistics for testing H_0 and H_1, respectively; and for $a > 0$, let

$$t_i = t_i{}^a = \inf\{n \geqq 1 : \Lambda_n{}^i \geqq a\}\,.$$

We will study the test which takes $t = \min\{t_0, t_1\}$ observations and decides in favor of H_0 if, and only if, $\Lambda_t{}^0 < \Lambda_t{}^1$. Thus, the test continues sampling as long as $n\delta - (2an)^{\frac{1}{2}} < ||S_n|| < (2an)^{\frac{1}{2}}$. Observe that $t \leqq N = 8a\delta^{-2}$ and that $t_1 \geqq 2a\delta^{-2}$.

It follows from the work of Schwarz (1962, 1968) that this test approximates an optimal Bayesian test with respect to a general prior and loss structure. In Schwarz' Bayesian formulation $a = \log c^{-1}$, where c is the cost of a single observation relative to the cost of a wrong decision.

We will study the power function $\beta(\omega) = P_\omega{}^\infty\{\Lambda_t{}^0 \geqq \Lambda_t{}^1\}$ as $a \to \infty$. It is easily seen that $\beta(\omega) = \beta^*(||\omega||)$, where β^* is continuous and strictly increasing. Thus the maximum error probabilities are $\beta(0)$ and $1 - \beta^*(\delta)$. A simple variation on Theorem 3 shows that

$$\beta(0) \sim P_0{}^\infty\{t_0 \leqq N\} \sim K_0 a^{\frac{1}{2}p} e^{-a}\,, \tag{27}$$

where $K_0 = K(\frac{1}{8}\delta^2, \infty)$ is as in Example 2. Similarly,

$$1 - \beta^*(\delta) \sim K_1 a^{\frac{1}{2}} \cdot e^{-a}\,,$$

where

$$K_1 = \frac{1}{2\pi^{\frac{1}{2}}} \int_{\frac{1}{8}\delta^2}^{\frac{1}{2}\delta^2} [1 - \delta^{-1}(2\varepsilon)^{\frac{1}{2}}]^{\frac{1}{2}(p-1)} \kappa(\varepsilon)\varepsilon^{-2}\, d\varepsilon$$

with $\kappa = \exp\{-2\sum_{k=1}^\infty \Phi[-(\frac{1}{2}k\varepsilon)^{\frac{1}{2}}]\}$, as in (22).

It is straightforward to compute K_0 and K_1 numerically.

It is also straightforward to approximate the expected sample size $E_\omega(t)$. For example, if $||\omega|| > \frac{1}{2}\delta$, then $E_\omega(t) = a||\omega||^{-2} + ||\omega||^{-1}\nu(\omega) + o(1)$ as $a \to \infty$, where $\nu(\omega)$ is the expected excess over the boundary; and a similar expression may be obtained when $||\omega|| < \frac{1}{2}\delta$. By combining the approximations to the error probabilities with those for the expected sample size, one may compute approximate efficiencies of a fixed sample size test or other sequential test with respect to the test described in this example. For example, if f_0 denotes the fixed sample size required to obtain the same error probabilities, then it is easy to approximate $E_\omega(t)/f_0$ with an error which is $o(a^{-1})$ as $a \to \infty$. I hope to pursue this topic in greater generality in a subsequent manuscript.

APPENDIX

We will outline the proofs of Theorem 1 and Lemma 3. For simplicity, we suppose throughout that $q \geqq 1$.

To prove Theorem 1, it will suffice to show that

$$I(\varepsilon, s) \sim (2s)^{\frac{1}{2}q}\Delta_q \int_{M_\varepsilon} ukj^{-1}\,dH_p^{r-1} \tag{28}$$

as $s \to 0$ uniformly on compact subintervals of $(0, \varepsilon_0)$, where $I(\varepsilon, s)$ is as in Corollary 1. For $0 < \varepsilon < \varepsilon_0$, let $L_s = \{x \in \Gamma : \phi(x) = \varepsilon + s,\ \phi_0(x) \leqq s\}$ be the domain of integration in (12). Then $Q(x, 0)$ is positive definite on $L_0 = M_\varepsilon$, so there is a $\delta > 0$ for which $Q(x, 0) \geqq \delta I_q$ for all $x \in M_\varepsilon$. Let Θ_0 be a convex neighborhood of $0 \in \Theta$. Then, given $\eta > 0$, we may cover M_ε with a finite number of open (in R^p) sets $G_1, \cdots, G_m$ for which

$$\hat\theta(x) \in \Theta_0 \qquad \text{and} \qquad \|Q[x, t\hat\theta(x)] - Q(z, 0)\| \leqq \delta\eta \tag{29}$$

for all $x \in G_i$, $z \in G_i \cap M_\varepsilon$, $0 \leqq t \leqq 1$, and $i = 1, \cdots, m$. Here $\|A\|^2 = \operatorname{tr}(A'A)$. Let $\gamma_1, \cdots, \gamma_m$ be a partition of unity on $G = \bigcup_{i=1}^m G_i$ for which $\gamma_i = 0$ on $G - G_i$, $i = 1, \cdots, m$. Then, for s sufficiently small,

$$I(\varepsilon, s) = \int_{L_s} uj^{-1}\,dH_p^{p-1} = \sum_{i=1}^m \int_{L_s} \gamma_i uj^{-1}\,dH_p^{p-1}. \tag{30}$$

We estimate a typical term in (30). Pick $z_i \in G_i \cap M$ and let $Q_i = Q(z_i, 0)$. Then $\phi_0(x) \geqq \frac{1}{2}(1 - \eta)\hat\theta(x)'Q_i\hat\theta(x)$ for $x \in G_i$ by (29). Let $s_\eta = 2s/(1 - \eta)$ and $L_s^* = \{x \in \Gamma : \phi(x) = \varepsilon + s,\ \hat\theta(x)'Q_i\hat\theta(x) \leqq s_\eta\}$, so that $L_s \cap G_i \subset L_s^* \cap G_i$. We then have

$$\begin{aligned}\int_{L_s} \gamma_i uj^{-1}\,dH_p^{p-1} &\leqq \int_{L_s^*} \gamma_i uj^{-1}\,dH_p^{p-1} \\ &= \int_{z'Q_iz \leqq s_\eta} \Big[\int_{\phi=\varepsilon+s,\hat\theta=z} \gamma_i uj^{-1}J_{\hat\theta}^{-1}\,dH_p^{r-1}\Big]\,dL^q(z) \\ &= s_\eta^{\frac{1}{2}q}|Q_i|^{-\frac{1}{2}} \int_{z'z\leqq 1} I^*[s, Q_i^{-\frac{1}{2}}z]\,dL^q(z)\end{aligned} \tag{31}$$

where

$$I^*(s, z) = \int_{\phi=\varepsilon+s,\hat\theta=s_\eta^{\frac{1}{2}}z} \gamma_i uj^{-1}J_{\hat\theta}^{-1}\,dH_p^{r-1} \to \int_{M_\varepsilon} \gamma_i uj^{-1}J_{\hat\theta}^{-1}\,dH_p^{r-1}$$

as $s \to 0$. Here the second step follows directly from equation (10), and the final one may be justified by using (9).

By combining (30) and (31) and letting $\eta \to 0$, we find

$$\limsup\, (2s)^{-\frac{1}{2}q}I(\varepsilon, s) \leqq \Delta_q \int_{M_\varepsilon} uj^{-1}J_{\hat\theta}^{-1}|Q(x, 0)|^{-\frac{1}{2}}\,dH_p^{r-1}(x) \tag{32}$$

as $s \to 0$; and a similar argument will show that $\liminf\, (2s)^{-\frac{1}{2}q}I(\varepsilon, s)$ is at least as big as the right side of (32). Finally, a simple application of the chain rule yields $|Q(x, 0)|^{-\frac{1}{2}}J_{\hat\theta}(x)^{-1} = k(x)$ for $x \in M_\varepsilon$, so the right sides of (28) and (32) are the same. This shows that (28) holds for a fixed ε, $0 < \varepsilon < \varepsilon_0$, and the uniformity may be established by repeating the above argument with ε replaced by a convergent sequence ε_n.

We will now sketch the proof of Lemma 3. It is easy to see that if $n = n_a \to \infty$ as $a \to \infty$ with $\varepsilon_a = an^{-1} \to \varepsilon$, where $0 < \varepsilon < \varepsilon_0$, then

$$P_0^\infty\{t = n\} \sim (2\pi)^{-\frac{1}{2}p} \cdot n^{\frac{1}{2}p-1} \cdot e^{-a} \cdot \int_0^{n\delta} I_{n,a}(\varepsilon_a, n^{-1}y)e^{-y}\,dL^1(y)$$

as $a \to \infty$ for any $\delta > 0$, where $I_{n,a}(\varepsilon, s)$ is as in (12), but with u replaced by $u_{n,a}$ (cf. (11) and (18)). The analysis of $I_{n,a}(\varepsilon_a, n^{-1}y)$ begins as in (29)—(31). Only the treatment of $I^*_{n,a}$ requires substantial comment. Let $0 \leqq y < \infty$, $z \in R^p$, and $s = n^{-1}y$. Further, let $H_a = H_a(y, z) = \{x : \phi(x) = \varepsilon_a + s, \hat{\theta}(x) = s_\eta^{\frac{1}{2}}z\}$, so that

$$I^*_{n,a}(s, z) = \int_{H_a} u_{n,a}\gamma_i j^{-1}J_{\hat{\theta}}^{-1}\, dH_p{}^{r-1}$$

in equation (31). A simple application of Lemma 2 shows that

$$u_{n,a}(x) - w\left\{x, y\left[1 - \frac{z'Q(x, 0)z}{1-\eta}\right]\right\} \to 0$$

as $a \to \infty$, uniformly with respect to $x \in H_a = H_a(y, z)$ and (y, z) in any compact subset of $[0, \infty) \times R^p$; and it follows that

$$\lim I^*_{n,a}(s, Q_i^{-\frac{1}{2}}z) = \int_{M_\varepsilon} \gamma_i j^{-1}J_{\hat{\theta}}^{-1}w[x, \xi]\, dH_p{}^{r-1},$$

where

$$\xi = \xi(x, y, z) = y[1 - (1-\eta)^{-1}z'Q_i^{-\frac{1}{2}}Q(x, 0)Q_i^{-\frac{1}{2}}z].$$

The remainder of the proof proceeds as in Theorem 1. By letting $a \to \infty$ and $\eta \to 0$ (in that order), we find

$$\limsup n^{\frac{1}{2}q}I_{n,a}(\varepsilon_a, n^{-1}y) \leqq (2y)^{\frac{1}{2}q}\int_{z'z\leqq 1}\left[\int_{M_\varepsilon} w[x, y(1 - z'z)]kj^{-1}\, dH_p{}^{r-1}\right] dL^q(z) \tag{33}$$

uniformly on compacta with respect to y; and a similar argument will show that $\liminf n^{\frac{1}{2}q}I_{n,a}(\varepsilon_a, n^{-1}y)$ is at least as big as the right side of (33). The limit in (33) is then integrated to obtain Lemma 3.

REFERENCES

[1] ARMITAGE, P. (1957). Restricted sequential procedures. *Biometrika* **44** 9-26.

[2] BERK, R. (1976). Asymptotically optimal sequential tests. Unpublished manuscript, Dept. of Statistics, Rutgers Univ.

[3] BOROVKOV, A. A. and ROGOZIN, B. A. (1965). On the multidimensional central limit theorem. *Theor. Probability Appl.* **10** 55-62.

[4] BROWN, L. (1971). Nonlocal asymptotic optimality of appropriate likelihood ratio tests. *Ann. Math. Statist.* **42** 1206-1240.

[5] EDWARDS, C. H. (1973). *Advanced Calculus of Several Variables.* Academic Press, New York.

[6] EFRON, B. and TRUAX, D. (1968). Large deviation theory in exponential families. *Ann. Math. Statist.* **39** 1402-1425.

[7] FEDERER, H. (1969). *Geometric Measure Theory.* Springer, Berlin.

[8] LAI, T. L. and SIEGMUND, D. O. (1977). A nonlinear renewal theorem with applications to sequential analysis I. *Ann. Statist.* **5** 946-955.

[9] SCHWARZ, G. (1962). Asymptotic shapes for Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224-236.

[10] SCHWARZ, G. (1968). Asymptotic shapes for sequential testing of truncation parameters. *Ann. Math. Statist.* **39** 2038-2043.

[11] WOODROOFE, M. (1976). A renewal theorem for curved boundaries and moments of first passage times. *Ann. Probability* **4** 67-80.

[12] WOODROOFE, M. (1976). Frequentist properties of Bayesian sequential tests. *Biometrika* **63** 101-110.

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109