The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 2044–2055

ON THE DISTRIBUTION OF THE LOG LIKELIHOOD RATIO TEST STATISTIC WHEN THE TRUE PARAMETER IS "NEAR" THE BOUNDARIES OF THE HYPOTHESIS REGIONS[1]

BY PAUL I. FEDER

Yale University

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed observations each having a density function $f(x, \theta)$ where $\theta \varepsilon \Theta$, a subset of Euclidean k-space. Consider the likelihood ratio statistic for the test of $H_1: \theta \varepsilon \omega_1$ vs. $H_2: \theta \varepsilon \omega_2$ where ω_1 and ω_2 are disjoint subsets of Θ.

In 1938 Wilks [8] proved his classical result on the asymptotic distribution of $-2 \log \lambda$, where

$$\lambda = \sup_{\theta \varepsilon \omega_1} \prod_{j=1}^n f(X_j, \theta) / \sup_{\theta \varepsilon \Theta} \prod_{j=1}^n f(X_j, \theta).$$

He showed that if ω_1 is an r-dimensional hyperplane in Euclidean k-space and ω_2 its complement, then if θ_0, the true state of nature, is in ω_1, $-2 \log \lambda$ has an asymptotic chi square distribution with $k - r$ degrees of freedom.

In 1943 Wald ([7], section 14) showed under somewhat stronger uniformity conditions that if ω_1 behaves locally like an r-dimensional hyperplane, $\omega_2 = \Theta - \omega_1$, and the true state of nature is a *sequence* converging to ω_1 at the rate $n^{-\frac{1}{2}}$, then asymptotically $-2 \log \lambda$ behaves like a *noncentral* chi squared random variable. In 1959 Silvey [6] obtained similar results by the use of Lagrange multipliers.

In 1954 Chernoff [1] generalized the Wilks result to deal with cases where ω_1 and ω_2 are not necessarily hyperplanes and their complements. He showed that if θ_0 (wlog taken to be 0) is a boundary point of both ω_1 and ω_2 (i.e. $\theta_0 \varepsilon \bar{\omega}_1 \cap \bar{\omega}_2$), and both ω_1 and ω_2 are approximable at $\theta_0 = 0$ by positively homogeneous sets (cones) C_1 and C_2, then under regularity conditions essentially those needed to prove the asymptotic normality of the maximum likelihood estimator (mle)

$$(1) \quad L\{-2 \log \lambda^*\} \to L\{\inf_{\theta \varepsilon C_1} (Z - \theta)' J (Z - \theta) - \inf_{\theta \varepsilon C_2} (Z - \theta)' J (Z - \theta)\}$$

where

$$\lambda^* = \sup_{\theta \varepsilon \omega_1} \prod_{j=1}^n f(X_j, \theta) / \sup_{\theta \varepsilon \omega_2} \prod_{j=1}^n f(X_j, \theta),$$

$$J(\theta) = E_\theta \|(\partial \log f(X, \theta)/\partial \theta_i)(\partial \log f(X, \theta)/\partial \theta_j)\|$$

is the $k \times k$ Fisher information matrix with $J \equiv J(0)$ assumed strictly positive

Received 20 February 1968.

[1] This paper is based on a part of the author's doctoral dissertation submitted at Stanford University (September, 1967). The research was primarily supported by National Science Foundation Grant GP-5705 at Stanford University and was partially supported by the Army, Navy, Air Force and NASA under a contract administered by the Office of Naval Research.

definite, and Z is normally distributed with mean 0 and covariance J^{-1}. Note that the statistic λ as used by Wilks is min $(\lambda^*, 1)$.

This asymptotic distribution is precisely the distribution of the likelihood ratio statistic for the test of $\theta \varepsilon C_1$ vs. $\theta \varepsilon C_2$ based on one observation from a $N(\theta, J^{-1})$ distribution with $\theta_0 = 0$.

This paper studies the behavior of $-2 \log \lambda^*$ when θ_0 is *near* the boundaries of ω_1 and ω_2 in the sense that as in [7] the true state of nature is a sequence of points θ_{0n} (not necessarily in ω_1 or ω_2) such that $\theta_{0n} = \theta_0 + o(1)$ where $\theta_0 \varepsilon \bar{\omega}_1 \cap \bar{\omega}_2$. Without loss of generality θ_0 is taken to be 0.

Two cases of interest are discussed in the main theorem.

(a) $d(\theta_{0n}, \omega_i) = O(n^{-\frac{1}{2}})$, $i = 1, 2$ (where $d(\theta, \omega)$ is the Euclidean distance from the point θ to the set ω)

(b) max $\{d(\theta_{0n}, \omega_1), d(\theta_{0n}, \omega_2)\}$ is large when compared with $n^{-\frac{1}{2}}$.

Case (a) gives rise to a noncentral version of (1). More specifically, the asymptotic distribution of $-2 \log \lambda^*$ is like that of the likelihood ratio statistic for the test of $\theta \varepsilon -\gamma_{1n} + C_1$ vs. $\theta \varepsilon -\gamma_{2n} + C_2$ based on one observation from a $N(\theta, J^{-1})$ distribution with 0 the true state of nature. C_1 and C_2 are positively homogeneous sets and γ_{1n}, γ_{2n} are suitably defined k-vectors. The set $v + C$ denotes the translate of C by the vector v (i.e. $\{v + w: w \varepsilon C\}$). This result unifies the Chernoff and Wald extensions of the original Wilks result.

Case (b) leads to a degenerate limiting distribution by the use of a different normalization than in (a).

These results are more precisely stated in Section 3. In Section 4 an illustrative example is presented in which $k = 2$ and $\omega_1 = \{\theta: \theta_2 \geqq \theta_1^2\}$, $\omega_2 = \{\theta: \theta_2 < -\theta_1^2\}$. The asymptotic distribution of $-2 \log \lambda^*$ is examined for various sequences θ_{0n}, each converging to 0.

2. Preliminary results. Let $X_{n1}, X_{n2}, \cdots, X_{nn}$ be independent and identically distributed observations having density $f(x, \theta_{0n})$. Assume that $\theta_{0n} = o(1)$.

The following notation is used throughout:

(a) $L(X^{(n)}, \theta) \equiv \prod_{\alpha=1}^{n} f(X_{n\alpha}, \theta)$ denotes the likelihood function.

(b) $\hat{\theta}$ is the unrestricted mle.

(c) $\hat{\theta}_\varphi$ is the mle restricted to $\varphi \subset \Theta$.

(d) $|\cdot|$ is a vector norm.

(e) $\|\cdot\|$ denotes a matrix.

(f) $\partial g(\theta)/\partial\theta$ represents the $k \times 1$ column vector whose ith component is $\partial g(\theta)/\partial\theta_i$.

(g) $d(\theta, \omega)$ is the Euclidean distance from the point θ to the set ω.

(h) $I(\theta, \psi)$ denotes the Kullback-Leibler distance (or information) between $f(x, \theta)$ and $f(x, \psi)$ and is defined as $\int \log [f(x, \theta)/f(x, \psi)]f(x, \theta)\, d\mu(x)$. It is well known that $I(\theta, \psi) \geqq 0$, with equality if and only if $f(x, \theta) = f(x, \psi)$ except for a set having P_θ measure 0.

The calculus of O_p and o_p is used without any explanation. The reader is referred to Pratt [5] for a rigorous discussion of the properties of these quantities. Loosely speaking, one can operate with them as with O and o.

The following regularity conditions will be imposed. Assumptions (R2)–(R4) are essentially conditions (a), (b), and (c) of [1] and guarantee the asymptotic normality of the mle. Additional assumptions are needed to handle technical difficulties that arise in the consideration of the *triangular array* $X_{n1}, X_{n2}, \cdots, X_{nn}$.

(R1) If $\{\theta_{0n}\}$ is any sequence such that $\theta_{0n} = o(1)$, then $\hat{\theta} = o_p(1)$ and $\hat{\theta}_\varphi = o_p(1)$ where φ is any subset of Θ such that $0 \varepsilon \bar{\varphi}$.

There exists a neighborhood N of $\theta = 0$ such that for all $\theta \varepsilon N$

(R2) $\partial \log f(\cdot, \theta)/\partial\theta_i$, $\partial^2 \log f(\cdot, \theta)/\partial\theta_i\partial\theta_j$ exist and

$$\sup_{|\theta| \leq r} \left| \partial^2 \log f(x, \theta)/\partial\theta_i\partial\theta_j - \partial^2 \log f(x, 0)/\partial\theta_i\partial\theta_j \right| < H(x)g(r)$$

where $E_\theta H(X) < M$ and $g(r)$ approaches 0 as $r \to 0$.

(R3) $$|\partial f(x, \theta)/\partial\theta_i| < F(x), \qquad |\partial^2 f(x, \theta)/\partial\theta_i\partial\theta_j| < F(x)$$

where $E_\theta F(X) < \infty$.

(R4) $J(\theta) \equiv \| E_\theta\{ \partial \log f(X, \theta)/\partial\theta_i \, \partial \log f(X, \theta)/\partial\theta_j\} \|$ is finite and strictly positive definite.

(R5) $$\int (\partial^2 \log f(x, 0)/\partial\theta_i\partial\theta_j) F(x) \, d\mu(x) < \infty,$$
$$\int (\partial^2 \log f(x, 0)/\partial\theta_i\partial\theta_j)^2 F(x) \, d\mu(x) < \infty,$$
$$\int (\partial^2 \log f(x, 0)/\partial\theta_i\partial\theta_j) f(x, 0) \, d\mu(x) < \infty, \quad i, j = 1, \cdots, k.$$

(R6) For every $\delta > 0$, $\liminf_{\theta \to 0} \inf_{|\psi| > \delta} I(\theta, \psi) > 0$.

REMARKS: (i) Condition (R3) is needed to invoke the Lebesgue dominated convergence theorem to justify the differentiation of $\int f(x, \theta) \, d\mu(x)$ twice under the integral sign. This implies that

$$E_\theta \{(\partial/\partial\theta) \log f(X, \theta)\} = 0 \qquad \text{and}$$
$$E_\theta\{(\partial \log f(X, \theta)/\partial\theta_i)(\partial \log f(X, \theta)/\partial\theta_j)\} = -E_\theta\{\partial^2 \log f(X, \theta)/\partial\theta_i\partial\theta_j\}.$$

(ii) It can be shown by an application of the Lebesgue dominated convergence theorem, conditions (R2) – (R5) imply that if $\theta \varepsilon N$, $\eta \varepsilon N$ then $I(\eta, \theta)$ can be differentiated twice with respect to θ under the integral sign. This implies

$$(\partial/\partial\theta)I(\eta, \theta) = -E_\eta\{(\partial/\partial\theta) \log f(X, \theta)\},$$
$$\|\partial^2 I(\eta, \theta)/\partial\theta_i\partial\theta_j\| = \|-E_\eta\{\partial^2 \log f(X, \theta)/\partial\theta_i\partial\theta_j\}\|.$$

(iii) Condition (R6) is a local identifiability condition around $\theta = 0$.

LEMMA 1. *Under conditions* (R1) – (R5)

$$\text{(a)} \quad n^{\frac{1}{2}}(\hat{\theta} - \theta_{0n}) = n^{\frac{1}{2}} J^{-1} A + o_p(1) \tag{2}$$

where $A \equiv A(\theta_{0n}) = n^{-1} \sum_{\alpha=1}^{n} (\partial/\partial\theta) \log f(X_{n\alpha}, \theta_{0n})$

$$\text{(b)} \quad L\{n^{\frac{1}{2}}(\hat{\theta} - \theta_{0n})\} \to N(0, J^{-1}). \tag{3}$$

PROOF. The law of large numbers and central limit theorem for double sequences are applied to the classical method of proof of asymptotic normality of the mle. Q.E.D.

Let $S_n = \{\theta: |\theta| < \delta_n\}$ with δ_n converging to 0, but sufficiently slowly so that $\theta_{0n} = o(\delta_n)$, $\hat{\theta} = o_p(\delta_n)$. Define

$$g_n(\theta) = E_{\theta_{0n}}\{\log [f(X, \theta)/f(X, \theta_{0n})]\}$$

and

$$\hat{g}_n(\theta) = n^{-1} \textstyle\sum_{\alpha=1}^n \log[f(X_{n\alpha}, \theta)/f(X_{n\alpha}, \theta_{0n})].$$

Within the sequence S_n of shrinking neighborhoods, $g_n(\theta)$ and $\hat{g}_n(\theta)$ behave like two paraboloids, with maxima at θ_{0n} and $\hat{\theta}$ respectively and second derivative matrices uniformly close to $-J$. More precisely

LEMMA 2. *Let* $\delta_n = o(1)$ *be any sequence such that* $\theta_{0n} = o(\delta_n)$ *and* $\hat{\theta} = o_p(\delta_n)$. For $|\theta| \leqq \delta_n$

$$g_n(\theta) = -\tfrac{1}{2}(\theta - \theta_{0n})'[J + o(1)](\theta - \theta_{0n}), \tag{4}$$

$$\partial g_n(\theta)/\partial\theta = -[J + o(1)](\theta - \theta_{0n}), \tag{5}$$

$$\hat{g}_n(\theta) - \hat{g}_n(\hat{\theta}) = -\tfrac{1}{2}(\theta - \hat{\theta})'[J + o_p(1)](\theta - \hat{\theta}) \tag{6}$$

with $o(1)$ *and* $o_p(1)$ *applying uniformly in* θ *for* $|\theta| \leqq \delta_n$.

PROOF. Equations (4) and (5) follow immediately from the expansion of $g_n(\theta)$ and $\partial g_n(\theta)/\partial\theta$ in Taylor series about θ_{0n}, and the observation that $g_n(\theta_{0n}) = -I(\theta_{0n}, \theta_{0n}) = 0$, $\partial g(\theta_{0n})/\partial\theta = E_{\theta_{0n}}\{\partial \log f(X, \theta_{0n})/\partial\theta\} = 0$ for n sufficiently large, and $\partial^2 g_n(\theta)/\partial\theta_i\partial\theta_j = \partial^2 g_n(\theta_{0n})/\partial\theta_i\partial\theta_j + \rho_n = -J_{ij} + \epsilon_n + \rho_n$ for n sufficiently large, where $\epsilon_n = o(1)$ and $\rho_n \leqq E_{\theta_{0n}}\{H(X)\delta_n\} \leqq M\delta_n$ for n sufficiently large

Equation (6) similarly follows from the Taylor series expansion of $\hat{g}_n(\theta)$ about $\hat{\theta}$ by noting that $\partial\hat{g}_n(\hat{\theta})/\partial\theta = 0$ with large probability (wlp) as $n \to \infty$ and $\partial^2\hat{g}_n(\theta)/\partial\theta_i\partial\theta_j = n^{-1} \sum_{\alpha=1}^n \partial^2 \log f(X_{n\alpha}, \theta_{0n})/\partial\theta_i\partial\theta_j + R_n = -J(\theta_{0n}) + o_p(1) + R_n = -J + \epsilon_n + o_p(1) + R_n$ where $R_n \leqq \delta_n n^{-1} \sum_{\alpha=1}^n H(X_{n\alpha})$, whenever $|\theta| \leqq \delta_n$. Thus wlp as $n \to \infty$, $\epsilon_n + o_p(1) + R_n$ is *uniformly* small for all $|\theta| \leqq \delta_n$. Summarizing the above yields Lemma 2. Q.E.D.

The rate of convergence of $\hat{\theta}_\varphi$ to θ_{0n} will now be considered. Define ψ_{0n} as the closest point in $\bar{\varphi}$ to θ_{0n}, in the sense of Kullback-Leibler information.

LEMMA 3. *If* $d(\theta_{0n}, \varphi) = O(s_n)$ *with* $s_n = o(1)$, *then*

$$\psi_{0n} - \theta_{0n} = O(s_n), \tag{7}$$

$$\hat{\theta}_\varphi - \theta_{0n} = O_p(\max [s_n, n^{-\frac{1}{2}}]). \tag{8}$$

PROOF. From equation (4) and the fact that $\theta_{0n} = o(1)$, it is readily seen that $g_n(0) = o(1)$. Thus $0 \geqq g_n(\psi_{0n}) \geqq g_n(0) = o(1)$, and (R6) implies $\psi_{0n} = o(1)$. Choose δ_n in Lemma 2 sufficiently large so that $|\psi_{0n}| + |\theta_{0n}| + s_n = o(\delta_n)$. Let η_n be any point in $\bar{\varphi}$ closest in Euclidean distance to θ_{0n}. By hypothesis $\eta_n - \theta_{0n} = O(s_n)$. Thus $\eta_n \in S_n$ for n sufficiently large, and so

$$0 \geqq g_n(\psi_{0n}) \geqq g_n(\eta_n) = O(s_n^2),$$

the last equality following directly from (4). Equation (4) immediately implies (7).

Equation (8) will now be derived. Since $\hat{\theta}_\varphi$ is the restricted mle,

$$0 \leqq \mathcal{g}_n(\hat{\theta}_\varphi) - \mathcal{g}_n(\psi_{0n}) = (\hat{\theta}_\varphi - \psi_{0n})'\{n^{-1} \textstyle\sum_{\alpha=1}^n (\partial/\partial\theta) \log f(X_{n\alpha}, \psi_{0n})\}$$
$$(9) \qquad + \tfrac{1}{2}(\hat{\theta}_\varphi - \psi_{0n})' \|n^{-1} \textstyle\sum_{\alpha=1}^n (\partial^2 \log f(X_{n\alpha}, \psi_{0n})/\partial\theta_i\partial\theta_j\|(\hat{\theta}_\varphi - \psi_{0n})$$
$$+ |\hat{\theta}_\varphi - \psi_{0n}|^2 o_p(1).$$

Equations (5), (7), and remark (ii) imply $E_{\theta_{0n}}\{\partial \log f(X, \psi_{0n})/\partial\theta\} = \partial g_n(\psi_{0n})/\partial\theta = O(s_n)$. By arguments similar to those used to prove the asymptotic normality of the mle, one can show

$$L\{n^{-\frac{1}{2}} \textstyle\sum_{\alpha=1}^n [(\partial/\partial\theta) \log f(X_{n\alpha}, \psi_{0n}) - E_{\theta_{0n}}\{(\partial/\partial\theta) \log f(X, \psi_{0n})\}]\} \to N(0, J).$$

In particular,

$$(10) \qquad n^{-1} \textstyle\sum_{\alpha=1}^n (\partial/\partial\theta) \log f(X_{n\alpha}, \psi_{0n}) = O_p(\max [n^{-\frac{1}{2}}, s_n]).$$

Denote $\max [n^{-\frac{1}{2}}, s_n]$ by u_n. Combining equations (9) and (10) and noting

$$\|n^{-1} \textstyle\sum_{\alpha=1}^n \partial^2 \log f(X_{n\alpha}, \psi_{0n})/\partial\theta_i\partial\theta_j\| = -J + o_p(1),$$

it follows that

$$(11) \quad 0 \leqq (\hat{\theta}_\varphi - \psi_{0n})'O_p(u_n) - \tfrac{1}{2}(\hat{\theta}_\varphi - \psi_{0n})'J(\hat{\theta}_\varphi - \psi_{0n}) + |\hat{\theta}_\varphi - \psi_{0n}|^2 o_p(1).$$

Thus

$$(12) \qquad \hat{\theta}_\varphi - \psi_{0n} = O_p(u_n).$$

Equation (8) follows directly from (7) and (12). This completes the proof of Lemma 3. Q.E.D.

Remark. In analogy with results on the rate of convergence of $\hat{\theta}$ to θ_{0n}, one might expect that $\hat{\theta}_\varphi - \psi_{0n} = O_p(n^{-\frac{1}{2}})$ rather than $O_p(u_n)$ as stated in equation (12). It is interesting to note that this is not true in general.

For example let $\varphi = \{\theta : \theta_2 \geqq -|\theta_1|\}$. Suppose $|\theta_{0n}| = s_n$ where $s_n = o(1)$, $n^{-\frac{1}{2}} = o(s_n)$. Further, suppose that the data consist of $N(\theta, I)$ random variables and θ_{0n} is within distance $o(n^{-\frac{1}{2}})$ of the negative θ_2-axis.

In this case K-L distance is merely half Euclidean distance. It is well known that for the exponential family, $\hat{\theta}_\varphi$ is that element of $\bar{\varphi}$ which is closest (in K-L distance) to $\hat{\theta}$. Thus, in this instance, $\hat{\theta}_\varphi$ is any point in $\bar{\varphi}$ closest to $\hat{\theta}$ in the Euclidean sense. Since θ_{0n} and $\hat{\theta}$ will be on opposite sides of the θ_2-axis with probability approximately $\frac{1}{2}$, $\hat{\theta}_\varphi - \psi_{0n} \neq O_p(n^{-\frac{1}{2}})$.

However, if φ is *convex*, then $\{\partial g_n(\psi_{0n})/\partial\theta\}'(\theta - \theta_{0n}) \leqq 0$ for $\theta \in \varphi \cap N$ and n sufficiently large. From equation (9) and the central limit theorem it then follows that $\hat{\theta}_\varphi - \psi_{0n} = O_p(n^{-\frac{1}{2}})$.

3. The asymptotic distribution of $-2 \log \lambda^*$**.** The asymptotic distribution of $-2 \log \lambda^*$ will be derived for the case when θ_{0n}, the underlying state of nature,

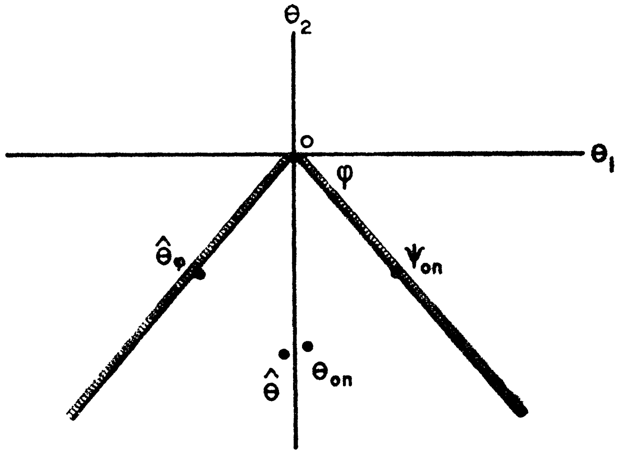

DIAGRAM 1

is "near" the boundaries of both hypothesis spaces. As was indicated in Section 1, the limiting distribution depends on the manner of convergence of θ_{0n} to 0.

First, the following definitions will be introduced. Definition 1 appears in [1] and Definition 3 in [4].

DEFINITION 1. A set C is *positively homogeneous* if $\theta \varepsilon C$ implies $k\theta \varepsilon C$ for all $k > 0$.

Let $\{\xi_n\}$ be a sequence of points in $\bar{\varphi}$ such that $\xi_n \to \theta_0$ and let $C^{(n)} \equiv \xi_n + C$ denote the translate of the set C by the vector ξ_n.

DEFINITION 2. *A set φ is sequentially approximable at θ_0 with respect to $\{\xi_n\}$ by the positively homogeneous set C* if for every $\eta_n = o(1)$

$$\sup_{x\varepsilon\varphi, D_n} \inf_{y\varepsilon C^{(n)}} |y - x| = o(\eta_n), \qquad \sup_{y\varepsilon C^{(n)}, D_n} \inf_{x\varepsilon\varphi} |y - x| = o(\eta_n).$$

where $$D_n = \{z;\ |z - \xi_n| < \eta_n\}.$$

Intuitively, this says that around θ_0, φ and C behave similarly.

DEFINITION 3. The *Levy distance* between two cdf's F and G, is defined to be

$$\delta_L(F, G) = \inf\{\delta: F(x - \delta) - \delta \leqq G(x) \leqq F(x + \delta) + \delta, \quad \text{for all} \quad x\}.$$

REMARK. (See [4], Section 9.) It is well known that $\delta_L(\cdot, \cdot)$ is a metric and convergence in this metric is equivalent to convergence in distribution.

Before stating the main theorem it is necessary to introduce some further notation and to prove a preliminary lemma.

1. $F_n^*(x) = P_n\{-2 \log \lambda^* \leqq x\}$, where $P_n(\cdot)$ is the probability measure corresponding to the parameter θ_{0n}.
2. $Q(w) = w'Jw$, where w is a k-vector and $J = J(0)$ is the Fisher information matrix.
3. $g(z, \tau_1, \tau_2) = \inf_{\theta\varepsilon C_1} Q(z + \tau_1 - \theta) - \inf_{\theta\varepsilon C_2} Q(z + \tau_2 - \theta)$.
4. $Z_n = n^{\frac{1}{2}} J^{-1} A(\theta_{0n})$ with distribution induced by $P_n(\cdot)$.

5. $G_n(x, \tau_1, \tau_2) = P_n\{g(Z_n, \tau_1, \tau_2) \leqq x\}$, $G(x, \tau_1, \tau_2) = P\{g(Z, \tau_1, \tau_2) \leqq x\}$ where $L\{Z\} = N(0, J^{-1})$.

LEMMA 4. $\sup_{|\tau_1| \leqq c, |\tau_2| \leqq c} \delta_L[G_n(\cdot, \tau_1, \tau_2), G(\cdot, \tau_1, \tau_2)] \to 0$.

PROOF. Given $\epsilon > 0$ there exists an $n_1(\epsilon)$ and a $K = K(\epsilon)$ such that for $n > n_1(\epsilon)$, $P\{|Z_n| > K\} < \epsilon$, $P\{|Z| > K\} < \epsilon$. In the compact region $\{|z| \leqq K, |\tau_1| \leqq c, |\tau_2| \leqq c\}$, $g(z, \tau_1, \tau_2)$ is *uniformly continuous*. Thus there exists an $\eta = \eta(\epsilon)$ such that $|g(z, \tau_1, \tau_2) - g(z, \tau_1', \tau_2')| < \epsilon$ for $|\tau_1| \leqq c$, $|\tau_2| \leqq c$, $|\tau_1'| \leqq c$, $|\tau_2'| \leqq c$, $|z| \leqq K$, $|\tau_1 - \tau_1'| \leqq \eta$, $|\tau_2 - \tau_2'| \leqq \eta$. Therefore, when $z, \tau_1, \tau_2, \tau_1', \tau_2'$ obey these constraints

$$G_n(x, \tau_1, \tau_2) \equiv P\{g(Z_n, \tau_1, \tau_2) \leqq x\} \leqq P\{g(Z_n, \tau_1, \tau_2) \leqq x, |Z_n| \leqq K\} + \epsilon$$
$$\leqq P\{g(Z_n, \tau_1', \tau_2') \leqq x + \epsilon, |Z_n| \leqq K\} + \epsilon \leqq G_n(x + \epsilon, \tau_1', \tau_2') + \epsilon.$$

Interchanging τ_1, τ_2 and τ_1', τ_2' and replacing x by $x - \epsilon$, we have

$$G_n(x - \epsilon, \tau_1', \tau_2') \leqq G_n(x, \tau_1, \tau_2) + \epsilon.$$

Thus

$$\delta_L[G_n(\cdot, \tau_1, \tau_2), G_n(\cdot, \tau_1', \tau_2')] \leqq \epsilon \quad \text{for} \quad n \geqq n_1(\epsilon). \tag{13}$$

Similarly

$$\delta_L[G(\cdot, \tau_1, \tau_2), G(\cdot, \tau_1', \tau_2')] \leqq \epsilon. \tag{14}$$

There exists a finite set $\{(\tau_{11}, \tau_{21}), (\tau_{12}, \tau_{22}), \cdots, (\tau_{1m}, \tau_{2m})\}$ such that $|\tau_{1i}| \leqq c$, $|\tau_{2i}| \leqq c$, $i = 1, 2, \cdots, m$, and for every (τ_1, τ_2) with $|\tau_1| \leqq c$, $|\tau_2| \leqq c$, there exists a $(\tau_1', \tau_2') \varepsilon \{(\tau_{11}, \tau_{21}), \cdots, (\tau_{1m}, \tau_{2m})\}$ with $|\tau_1 - \tau_1'| \leqq \eta$, $|\tau_2 - \tau_2'| \leqq \eta$. Since $L\{Z_n\} \to L\{Z\}$ and $g(\cdot, \tau_1, \tau_2)$ is continuous in z, $L\{g(Z_n, \tau_{i1}, \tau_{i2})\} \to L\{g(Z, \tau_{i1}, \tau_{i2})\}$ for each i. Hence

$$\delta_L[G_n(\cdot, \tau_{1i}, \tau_{2i}), G(\cdot, \tau_{1i}, \tau_{2i})] \leqq \epsilon \quad \text{for} \quad n > n_2(\epsilon, \tau_{11}, \tau_{21}, \cdots, \tau_{1m}, \tau_{2m}), \qquad i = 1, \cdots, m. \tag{15}$$

By the triangle inequality

$$\delta_L[G_n(\cdot, \tau_1, \tau_2), G(\cdot, \tau_1, \tau_2)] \leqq \delta_L[G_n(\cdot, \tau_1, \tau_2), G_n(\cdot, \tau_1', \tau_2')]$$
$$+ \delta_L[G_n(\cdot, \tau_1', \tau_2'), G(\cdot, \tau_1', \tau_2')] \quad + \delta_L[G(\cdot, \tau_1', \tau_2'), G(\cdot, \tau_1, \tau_2)].$$

Let $n_0 = \max(n_1, n_2)$. For $n > n_0$, equations (13), (14), and (15) imply $\delta_L[G_n(\cdot, \tau_1, \tau_2), G(\cdot, \tau_1, \tau_2)] \leqq 3\epsilon$ for all $|\tau_1| \leqq c$, $|\tau_2| \leqq c$. This completes the proof of Lemma 4. Q.E.D.

THEOREM 1. *Under regularity conditions* (R1) *to* (R6), *the asymptotic behavior of* $-2 \log \lambda^*$ *is as follows:*

Case 1. *If* $d(\theta_{0n}, \omega_i) = O(n^{-\frac{1}{2}})$ $i = 1, 2$, *and* ω_1, ω_2 *are sequentially approximable at* 0 *with respect to* $\{\xi_{in}\}$, $i = 1, 2$, *by disjoint positively homogeneous sets* C_1 *and* C_2, *then*

$$\delta_L[F_n^*, G(\cdot, \gamma_{1n}, \gamma_{2n})] \to 0 \tag{16}$$

uniformly in θ_{0n} *such that* $|\gamma_{in}| \leqq c$, *where* $\gamma_{in} = n^{\frac{1}{2}}(\theta_{0n} - \xi_{in})$, $i = 1, 2$.

Case 2. *If* $d(\theta_{0n}, \omega_i) = O(s_n)$, $i = 1, 2$ *where* $s_n \to 0$, $n^{\frac{1}{2}}s_n \to \infty$, *and* ω_1, ω_2 *are sequentially approximable at* 0 *with respect to* $\{\xi_{in}\}$ *by disjoint positively homogeneous sets* C_i, $i = 1, 2$, *then*

$$(17) \qquad -(2/ns_n^2) \log \lambda^* = \inf_{\theta \varepsilon C_1} Q(\gamma_{1n} - \theta) - \inf_{\theta \varepsilon C_2} Q(\gamma_{2n} - \theta) + o_p(1)$$

where $\gamma_{in} = s_n^{-1}(\theta_{0n} - \xi_{in})$, $i = 1, 2$, *and the* $o_p(1)$ *term is uniformly small for all* θ_{0n} *such that* $|\gamma_{in}| \leqq c$.

Before proceeding to the proof of the theorem, it may be of interest to make the following remarks:

(a) In Case 1, if $L\{g(Z, \tau_1, \tau_2)\}$ is continuous for all $|\tau_1| \leqq c$, $|\tau_2| \leqq c$, then

$$(18) \qquad \sup_x |F_n^*(x) - G(x, \gamma_{1n}, \gamma_{2n})| \to 0$$

uniformly in θ_{0n} such that $|\gamma_{in}| \leqq c$, $i = 1, 2$.

(b) If the hypotheses are strengthened to assert that $\gamma_{in} \to \gamma_i$, $i = 1, 2$, then $-2 \log \lambda^*$ or $-2 \log \lambda^*/ns_n^2$ has a limiting distribution which is obtained by substituting γ_i for γ_{in} in equation (16) or (17) respectively.

(c) Case 1 with $n^{\frac{1}{2}}\theta_{0n} \to 0$ (and $\xi_{1n} = \xi_{2n} = 0$ for all n) includes the Chernoff result, which deals with the special case where $\theta_{0n} \equiv 0$.

(d) Suppose C_1 is an r-dimensional hyperplane in k-dimensional Euclidean space and C_2 its complement. If $n^{\frac{1}{2}}\theta_{0n} = o(1)$, then a limiting chi squared distribution with $k - r$ degrees of freedom is obtained, just as in the original Wilks result [8]. If $n^{\frac{1}{2}}\theta_{0n} = O(1)$, then a noncentral chi squared distribution results, as stated in Wald ([7], section 14). For example, if $n^{\frac{1}{2}}\theta_{0n} \to \gamma_0$, then

$$L\{-2 \log \lambda^*\} \to \chi'^2(k - r; \kappa) \quad \text{with} \quad \kappa = \tfrac{1}{2} \inf_{\theta \varepsilon C_1} Q(\gamma_0 - \theta).$$

We now proceed to the proof.

Proof. *Case* 1. From Lemma 3, $\hat{\theta}_{\omega_1} - \theta_{0n} = O_p(n^{-\frac{1}{2}})$ and $\hat{\theta}_{\omega_2} - \theta_{0n} = O_p(n^{-\frac{1}{2}})$. From Lemma 1, $\hat{\theta} - \theta_{0n} = J^{-1}A(\theta_{0n}) + o_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}})$. From Lemma 2, equation (6),

$$\log L(X^{(n)}, \theta) \equiv n\hat{g}_n(\theta) = n\hat{g}_n(\hat{\theta}) - \tfrac{1}{2}n(\theta - \hat{\theta})'[J + o_p(1)](\theta - \hat{\theta})$$

with $o_p(1)$ applying uniformly in θ for $|\theta| \leqq \delta_n$. Thus

$$\begin{aligned} -2 \log \lambda^* &\equiv -2[\log L(X^{(n)}, \hat{\theta}_{\omega_1}) - \log L(X^{(n)}, \hat{\theta}_{\omega_2})] \\ &= n[\inf_{\theta \varepsilon \omega_1} (\hat{\theta} - \theta)'[J + o_p(1)](\hat{\theta} - \theta) \\ &\quad - \inf_{\theta \varepsilon \omega_2} (\hat{\theta} - \theta)'[J + o_p(1)](\hat{\theta} - \theta)] \\ &= n[\inf_{\theta \varepsilon \omega_1} Q(J^{-1}A + \theta_{0n} - \theta) - \inf_{\theta \varepsilon \omega_2} Q(J^{-1}A + \theta_{0n} - \theta)] \\ &\quad + r(X^{(n)}, \theta_{0n}) \end{aligned}$$

where $A \equiv A(\theta_{0n})$ and $r(X^{(n)}, \theta_{0n}) = o_p(1)$. Shift the origin to ξ_{in}. Thus

$$\begin{aligned} -2 \log \lambda^* &= \inf_{\theta \varepsilon \omega_1} Q[n^{\frac{1}{2}}J^{-1}A + n^{\frac{1}{2}}(\theta_{0n} - \xi_{1n}) - n^{\frac{1}{2}}(\theta - \xi_{1n})] \\ &\quad - \inf_{\theta \varepsilon \omega_2} Q[n^{\frac{1}{2}}J^{-1}A + n^{\frac{1}{2}}(\theta_{0n} - \xi_{2n}) - n^{\frac{1}{2}}(\theta - \xi_{2n})] + r(X^{(n)}, \theta_{0n}). \end{aligned}$$

Define

$$-2 \log \lambda^*_{\tau_1,\tau_2} = \inf_{\theta\varepsilon\omega_1} Q[n^{\frac{1}{2}}J^{-1}A + \tau_1 - n^{\frac{1}{2}}(\theta - \xi_{1n})] - \inf_{\theta\varepsilon\omega_2} Q[n^{\frac{1}{2}}J^{-1}A + \tau_2 - n^{\frac{1}{2}}(\theta - \xi_{2n})] + r(X^{(n)}, \theta_{0n}),$$

where τ_1 and τ_2 are any vectors such that $|\tau_1| \leqq c$, $|\tau_2| \leqq c$. Since ω_i is sequentially approximable by C_i, $i = 1, 2$, and $J^{-1}A + n^{-\frac{1}{2}}\tau_i = O_p(n^{-\frac{1}{2}})$,

$$-2 \log \lambda^*_{\tau_1,\tau_2} = \inf_{\theta\varepsilon C_1^{(n)}} Q[n^{\frac{1}{2}}J^{-1}A + \tau_1 - n^{\frac{1}{2}}(\theta - \xi_{1n})] - \inf_{\theta\varepsilon C_2^{(n)}} Q[n^{\frac{1}{2}}J^{-1}A + \tau_2 - n^{\frac{1}{2}}(\theta - \xi_{2n})] + u\rho(J^{-1}A, c) + r(X^{(n)}, \theta_{0n})$$

where $|u| \leqq 1$ and $\rho(J^{-1}A, c) = o_p(1)$.

Let $\theta^* = \theta - \xi_{1n}$, $\theta^{**} = \theta - \xi_{2n}$. Then

$$-2 \log \lambda^*_{\tau_1,\tau_2} = \inf_{\theta^*\varepsilon C_1} Q[n^{\frac{1}{2}}J^{-1}A + \tau_1 - \theta^*] - \inf_{\theta^{**}\varepsilon C_2} Q[n^{\frac{1}{2}}J^{-1}A + \tau_2 - \theta^{**}] + u\rho(J^{-1}A, c) + r(X^{(n)}, \theta_{0n}).$$

Define $F_n(x, \tau_1, \tau_2) = P_n\{-2 \log \lambda^*_{\tau_1,\tau_2} \leqq x\}$. Obviously,

$$\sup_{|\tau_1| \leqq c, |\tau_2| \leqq c} \delta_L[F_n(\cdot, \tau_1, \tau_2), G_n(\cdot, \tau_1, \tau_2)] \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Thus, from Lemma 4 and the triangle inequality,

$$\sup_{|\tau_1| \leqq c, |\tau_2| \leqq c} \delta_L[F_n(\cdot, \tau_1, \tau_2), G(\cdot, \tau_1, \tau_2)] \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \tag{19}$$

Since $F_n^*(x) = F_n(x, \gamma_{1n}, \gamma_{2n})$, the substitution of γ_{1n}, γ_{2n} into equation (19) yields

$$\delta_L[F_n^*, G(\cdot, \gamma_{1n}, \gamma_{2n})] \rightarrow 0. \tag{20}$$

Since this is true for *every* sequence $\{\theta_{0n}\}$ such that $|\gamma_{in}| \leqq c$, the result in Case 1 follows.

The proof of Case 2 is similar to that of Case 1 and is omitted.

For the sake of completeness, the behavior of $n^{-1} \log \lambda^*$ will be discussed for the case when 0 is bounded away from at least one of the hypothesis spaces. This is in the spirit of results obtained by Cox [2] and others, if not explicitly mentioned by them.

Suppose that 0 is the true state of nature and that ψ_i is the closest point to 0 in ω_i, $i = 1, 2$, in the sense of Kullback-Leibler distance.

THEOREM 2. *If for every* $\epsilon > 0$ *there exist neighborhoods* $U_{1\epsilon}$, $U_{2\epsilon}$ *such that* $\psi_1 \varepsilon U_{1\epsilon}$, $\psi_2 \varepsilon U_{2\epsilon}$ *and*

$$E_0\{\sup_{\theta'\varepsilon U_{i\epsilon}} \log [f(X, \theta')/f(X, \psi_i)]\} < \epsilon, \qquad i = 1, 2,$$

then

$$n^{-1} \sum_{\alpha=1}^n \log f(X_\alpha, \hat{\theta}_{\omega_i}) = n^{-1} \sum_{\alpha=1}^n \log f(X_\alpha, \psi_i) + o_p(1), \quad i = 1, 2. \tag{21}$$

In particular

$$n^{-1} \log \lambda^* = n^{-1} \sum_{\alpha=1}^{n} \log [f(X_\alpha, \psi_1)/f(X_\alpha, \psi_2)] + o_p(1) \tag{22}$$
$$= I(0, \psi_2) - I(0, \psi_1) + o_p(1).$$

4. Example. The following example illustrates the dependence of the asymptotic distribution of $-2 \log \lambda^*$ upon the manner of convergence of θ_{0n} to 0.

Let $k = 2$ and ω_1, ω_2 be the regions $\theta_2 \geqq \theta_1^2$ and $\theta_2 < -\theta_1^2$ respectively.

(i) Suppose $\theta_{0n} \equiv 0$. This is the case dealt with by Chernoff [1]. It is easily verified that ω_1 and ω_2 are sequentially approximable at 0 with respect to $\{\xi_{1n} \equiv 0\}$ and $\{\xi_{2n} \equiv 0\}$ by the positively homogeneous sets C_1 and C_2, where $C_1 = \{\theta_2 \geqq 0\}$, $C_2 = \{\theta_2 < 0\}$. Thus, asymptotically $-2 \log \lambda^*$ behaves like the likelihood ratio statistic for the test of $\theta_2 \geqq 0$ vs. $\theta_2 < 0$ based on one observation from a $N(0, J^{-1})$ distribution. Obviously, $\gamma_{1n} = \gamma_{2n} = 0$ and

$$L\{-2 \log \lambda^*\} \rightarrow L\{\inf_{\theta_2 \geqq 0} (Z - \theta)'J(Z - \theta) - \inf_{\theta_2 < 0} (Z - \theta)'J(Z - \theta)\}$$

where $L\{Z\} = N(0, J^{-1})$.

There exists a diagonal matrix D and an orthogonal matrix $\Delta \equiv (\Delta^{(1)}, \Delta^{(2)})$ such that $J = \Delta' D^2 \Delta$. Transform the parameter space so that $\varphi = \Gamma J^{\frac{1}{2}} \theta$, where Γ is the orthogonal matrix

$$\begin{pmatrix} \Delta^{(1)\prime} D \Delta / (J_{11})^{\frac{1}{2}} \\ \Delta^{(2)\prime} D^{-1} \Delta / (J^{22})^{\frac{1}{2}} \end{pmatrix}$$

and $J = (J_{ij})$, $J^{-1} = (J^{ij})$. If $W = \Gamma J^{\frac{1}{2}} Z$ then $L\{W\} = N(0, I)$. It is easy to

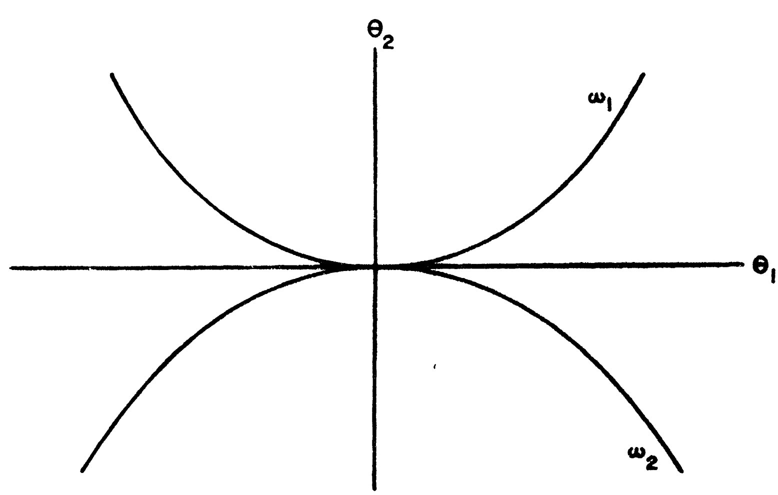

DIAGRAM 2

show that

$$\inf_{\theta_2 \geqq 0} (Z - \theta)'J(Z - \theta) - \inf_{\theta_2 < 0} (Z - \theta)'J(Z - \theta)$$
$$= \inf_{\varphi_2 \geqq 0} (W - \varphi)'(W - \varphi) - \inf_{\varphi_2 < 0} (W - \varphi)'(W - \varphi).$$

Thus $L\{-2 \log \lambda^*\} \to L\{U\}$ where $U = -W_2^2$ if $W_2 \geqq 0$ and $U = W_2^2$ if $W_2 < 0$. This is the distribution of a random variable which is $+\chi^2(1)$ with probability $\frac{1}{2}$ and $-\chi^2(1)$ with probability $\frac{1}{2}$.

(ii) Suppose $\theta_{0n} = (n^{-\frac{1}{3}}, 0)$. The regions ω_1 and ω_2 are sequentially approximable at 0 with respect to $\{\xi_{1n} \equiv (n^{-\frac{1}{3}}, n^{-\frac{2}{3}})'\}$ and $\{\xi_{2n} \equiv (n^{-\frac{1}{3}}, -n^{-\frac{2}{3}})'\}$ by the positively homogeneous sets C_1 and C_2 (defined as in (1)). In this case $\gamma_{1n} = (0, -n^{-\frac{1}{6}})'$, $\gamma_{2n} = (0, n^{-\frac{1}{6}})'$ and so $\gamma_1 = \gamma_2 = (0, 0)'$. One can conclude from equation (16) that $-2 \log \lambda^*$ has the same asymptotic distribution as in (1) above.

(iii) Suppose $\theta_{0n} = (0, n^{-\frac{1}{2}})'$. As in (1), the regions ω_1 and ω_2 are sequentially approximable at 0 with respect to $\{\xi_{1n} \equiv 0\}$ and $\{\xi_{2n} \equiv 0\}$ by C_1 and C_2. Obviously $\gamma_{1n} = \gamma_{2n} = (0, 1)' \equiv a = \gamma_1 = \gamma_2$. Thus

$$L\{-2 \log \lambda^*\} \to L\{\inf_{\theta_2 \geqq 0} Q[Z + a - \theta] - \inf_{\theta_2 < 0} Q[Z + a - \theta]\}$$

where $L\{Z\} = N(0, J^{-1})$. Perform the transformation $\varphi = \Gamma J^{\frac{1}{2}}\theta$ and let $W = \Gamma J^{\frac{1}{2}}Z$ where Γ is defined as in (i). Then

$$\inf_{\theta_2 \geqq 0} Q[Z + a - \theta] - \inf_{\theta_2 < 0} Q[Z + a - \theta] = \inf_{\varphi_2 \geqq 0} (W + \Gamma J^{\frac{1}{2}}a - \varphi)'(W + \Gamma J^{\frac{1}{2}}a - \varphi) - \inf_{\varphi_2 < 0} (W + \Gamma J^{\frac{1}{2}}a - \varphi)'(W + \Gamma J^{\frac{1}{2}}a - \varphi)$$

where $L\{W\} = N(0, I)$ and $\Gamma J^{\frac{1}{2}}a = (J_{12}/J_{11}^{\frac{1}{2}}, 1/(J^{22})^{\frac{1}{2}})'$. Thus, asymptotically $-2 \log \lambda^*$ behaves like the random variable defined as

$$-(W_2 + 1/(J^{22})^{\frac{1}{2}})^2 \quad \text{if} \quad W_2 \geqq -1/(J^{22})^{\frac{1}{2}},$$
$$(W_2 + 1/(J^{22})^{\frac{1}{2}})^2 \quad \text{if} \quad W_2 < -1/(J^{22})^{\frac{1}{2}},$$

where $L\{W_2\} = N(0, 1)$. This is a noncentral analogue of the distribution in (i).

(iv) Suppose $\theta_{0n} = (n^{-\frac{1}{4}}, 0)'$. The regions ω_1 and ω_2 are sequentially approximable at 0 with respect to $\{\xi_{1n} \equiv (n^{-\frac{1}{4}}, n^{-\frac{1}{2}})'\}$ and $\{\xi_{2n} \equiv (n^{-\frac{1}{4}}, -n^{-\frac{1}{2}})\}$. This implies $\gamma_{1n} = (0, -1)' \equiv -a = \gamma_1$ and $\gamma_{2n} = (0, 1)' \equiv a = \gamma_2$. Thus,

$$L\{-2 \log \lambda^*\} \to L\{\inf_{\theta_2 \geqq 0} Q(Z - a - \theta) - \inf_{\theta_2 < 0} Q(Z + a - \theta)\}$$

where $L\{Z\} = N(0, J^{-1})$.

By performing the same change of coordinates as previously, it is easily seen that

$$L\{-2 \log \lambda^*\} \to L\{\inf_{\varphi_2 \geqq 0} (W - \Gamma J^{\frac{1}{2}}a - \varphi)'(W - \Gamma J^{\frac{1}{2}}a - \varphi) - \inf_{\varphi_2 < 0} (W + \Gamma J^{\frac{1}{2}}a - \varphi)'(W + \Gamma J^{\frac{1}{2}}a - \varphi)\}$$

where $L\{W\} = N(0, I)$ and $\Gamma J^{\frac{1}{2}}a = (J_{12}/J_{11}^{\frac{1}{2}}, 1/(J^{22})^{\frac{1}{2}})'$. Thus, asymptotically $-2 \log \lambda^*$ behaves like the random variable defined as

$$\begin{aligned} (W_2 - 1/(J^{22})^{\frac{1}{2}})^2 &\quad \text{if} \quad W_2 < -1/(J^{22})^{\frac{1}{2}}, \\ -4W_2/(J^{22})^{\frac{1}{2}} &\quad \text{if} \quad -1/(J^{22})^{\frac{1}{2}} \leqq W_2 < 1/(J^{22})^{\frac{1}{2}}, \\ -(W_2 + 1/(J^{22})^{\frac{1}{2}})^2 &\quad \text{if} \quad W_2 \geqq 1/(J^{22})^{\frac{1}{2}}. \end{aligned}$$

(v) Suppose $\theta_{0n} = (0, n^{-\frac{1}{4}})$. Then $s_n = n^{-\frac{1}{4}}$ (where s_n is defined in Case 2 of Theorem 1), ω_1 and ω_2 are sequentially approximable at 0 with respect to $\{\xi_{1n} \equiv 0\}$ and $\{\xi_{2n} \equiv 0\}$ by C_1 and C_2, and $\gamma_{1n} = \gamma_{2n} = (0, 1)' \equiv a = \gamma_1 = \gamma_2$. From Case 2 of Theorem 1,

$$-2n^{-\frac{1}{2}} \log \lambda^* \to \inf_{\theta_2 < 0} [a - \theta]' J [a - \theta]$$

in probability.

5. Acknowledgment. The author wishes to express his appreciation to Professor Herman Chernoff of Stanford University under whose guidance this research was conducted, for his extremely valuable advice and suggestions.

REFERENCES

[1] Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25** 573–578.

[2] Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *J. Roy. Statist. Soc. Ser. B* **24** 406–424.

[3] Feder, P. I. (1967). On the likelihood ratio statistic with applications to broken line regression. Technical Report No. 23, Department of Statistics, Stanford Univ.

[4] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Theorems for Sums of Independent Random Variables*. Addison-Wesley, Reading.

[5] Pratt, J. W. (1959). On a general concept of "In probability". *Ann. Math. Statist.* **30** 549–558.

[6] Silvey, S. D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.* **30** 389–407.

[7] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

[8] Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9** 60–62.