

AD-A092 634

STANFORD UNIV CA DEPT OF STATISTICS

F/6 12/1

REPEATED LIKELIHOOD RATIO TESTS FOR CURVED EXPONENTIAL FAMILITE--ETC(U)

SEP 80 S P LALLEY

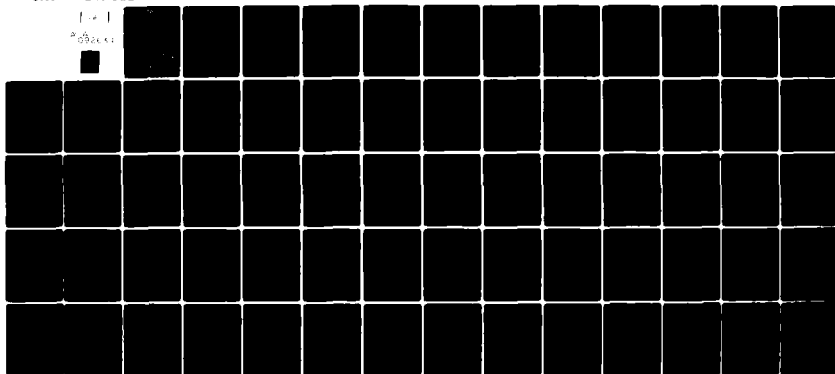
N00014-77-C-0306

NL

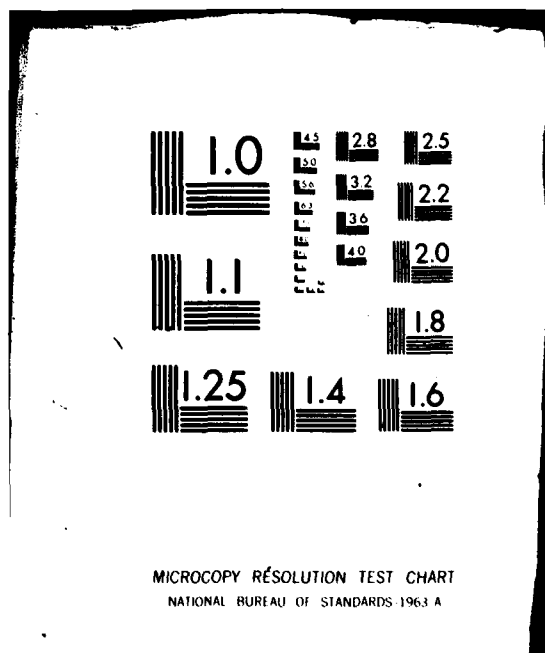
UNCLASSIFIED

TR-10

1-1
032443



END
DATE
FILMED
-1-81
DTIC



⑥ LEVEL II

REPEATED LIKELIHOOD RATIO TESTS
FOR CURVED EXPONENTIAL FAMILIES

BY

STEVEN PAUL LALLEY

TECHNICAL REPORT NO. 10
SEPTEMBER 1, 1980

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
OFFICE OF NAVAL RESEARCH

DTIC
ELECTE
DEC 9 1980
S D
B

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

80 12 03 013

AD A092634

DDC FILE COPY

REPEATED LIKELIHOOD RATIO TESTS
FOR CURVED EXPONENTIAL FAMILIES

by

Steven Paul Lalley

TECHNICAL REPORT NO. 10

September 1, 1980

Prepared under Contract

N00014-77-C-0306 (NR-042-373)

for the Office of Naval Research

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

REPEATED LIKELIHOOD RATIO TESTS
FOR CURVED EXPONENTIAL FAMILIES

Steven Paul Lalley, Ph.D.
Stanford University, 1980

A class of repeated significance tests for curved hypotheses in multiparameter exponential families is studied, and asymptotic formulae for the significance levels of such tests are obtained. Special attention is given the important case of comparing Bernoulli success probabilities.

Accession For	
RTIS GSA&I	☒
RTIS TAR	☐
Unannounced	☐
Justification	☐
By	
Distribution/	
Availability Codes	
Avail and/or	
Dist	Special
A	

ACKNOWLEDGMENTS

Professor David Siegmund has been a constant source of encouragement, ideas, problems, and advice in my years at Stanford. For all his help, and all I have learned from him, I am most grateful.

Throughout the past year I have had enlightening conversations with many people concerning subjects related to this work: I would particularly like to thank Robert Berk, T. L. Lai, and Michael Woodroffe.

I am indebted to Wanda Miller for her patience in typing the manuscript, and to Mike Steele and Brad Efron for reading it.

TABLE OF CONTENTS

Chapter	Page
1. Introduction	1
2. Example: Testing for the Equality of Two Bernoulli Parameters	3
3. Preliminaries Concerning Exponential Families and Statement of the Main Result	15
4. Normal Approximation in Several Dimensions	30
5. Expansion of Likelihood Functions	37
6. The Collapsing Argument	45
7. Proof of Theorem 2	54
REFERENCES	61

REPEATED LIKELIHOOD RATIO TESTS FOR CURVED EXPONENTIAL FAMILIES

1. Introduction

We study here the significance levels of repeated likelihood ratio tests for nested hypotheses in multiparameter exponential families. In various hypothesis testing situations, the peculiar constraints of medical research (and the even more peculiar constraints of non-medical research) occasionally preclude the determination of a sample size in advance of experimental results: various authors (notably Armitage [1], [2], and Schwartz [11]) have argued that in many such cases a reasonable option is provided by certain simple stopping rules based on the behavior of a (generalized) likelihood ratio statistic. Unfortunately, determining the operating characteristics of such procedures remains a difficult issue, although in recent years important advances have been made by Woodroffe [14], [15], [16], Siegmund [12], and Lai & Siegmund [9], [10].

Let $\{P_\theta, \theta \in \Omega\}$ be an exponential family of probability measures on $\mathbb{R}^p$:

$$(1.1) \quad (dP_\theta/dP_0)(x) = \exp\{\theta^T x - \psi(\theta)\} \quad .$$

The natural parameter space Ω is assumed to be an open subset of $\mathbb{R}^p$, and ψ is assumed to be strictly convex on Ω . Suppose that Ω_1 is a smooth relatively closed q_1 -dimensional submanifold of Ω , and that Ω_0 is a smooth relatively closed q_0 -dimensional submanifold of Ω_1 ,

where $0 < q_0 < q_1 \leq p$: these are to be the null and alternative hypotheses, respectively. (In the terminology of Efron [5], $\{P_\theta : \theta \in \Omega_1\}$ are "curved exponential families.") Let $X_1, X_2, \dots$ be i.i.d. from P_θ , and let $e^{\wedge_n}$ be the generalized likelihood ratio statistic for testing

$$H_0 : \theta \in \Omega_0 \quad \text{v.} \quad H_1 : \theta \in \Omega_1 .$$

The repeated likelihood ratio test will be based on the stopping rule $T \wedge m_1$, where

$$(1.2) \quad T = T_a = \min\{n \geq m_0 : \wedge_n > a\} ;$$

if $T \leq m_1$, H_0 should be rejected, whereas if $T > m_1$, H_0 should not be rejected.

The main result of this work is that for $m_0 \sim \epsilon_2^{-1} a$ and $m_1 \sim \epsilon_1^{-1} a$, and $\theta_0 \in \Omega_0$

$$(1.3) \quad P_{\theta_0} \{T_a \leq m_1\} \sim C a^{(q_1 - q_0)/2} e^{-a}$$

as $a \rightarrow \infty$, provided a certain host of regularity conditions are satisfied. The constant C , which depends on θ_0 , ϵ_1 , and ϵ_2 , will take the unpleasant form of a surface integral in $\mathbb{R}^p$, which may, however, be evaluated numerically in many cases of statistical interest.

2. Example: Testing for the Equality of Two Bernoulli Parameters

Suppose we observe a sequence $\{(X_i, Y_i) : i = 1, 2, \dots\}$ of i.i.d. random vectors taking values in the set $\{0, 1\}^2$, with

$$(2.1) \quad P_{p_1, p_2} \{X_1 = e_1, Y_1 = e_2\} = p_1^{e_1} p_2^{e_2} (1 - p_1)^{(1-e_1)} (1 - p_2)^{(1-e_2)}$$

where $e_1, e_2 \in \{0, 1\}$. The parameters p_1 and p_2 are unknown; we wish to test the hypothesis $p_1 = p_2$.

Imagine that the variables X_i, Y_i are success indicators in a clinical trial. Patients suffering from a particular disorder arrive infrequently at a clinic where they may be treated according to one of two procedures: because of the nature of the disorder the patients must be treated immediately, and a response (success or failure) is apparent within a relatively short period of time (compared to interarrival times). If the disorder is serious, sequential experimentation to compare the efficacies of the two procedures may be appropriate.

Such a situation was considered by Siegmund and Gregory [13], who proposed several sequential procedures for testing the hypothesis $p_1 = p_2$. One of these was a sequential version of the generalized likelihood ratio test, which had previously been studied in different contexts by Armitage [1], [2], Schwartz [11], Siegmund [12], and Woodroffe [15], [16]. This test is easily described. Let

$$(2.2) \quad H(x) = x \log x + (1-x) \log(1-x)$$

$$I(x,y) = H(x) + H(y) - 2H((x+y)/2)$$

$$\Lambda_n = n I(\bar{x}_n, \bar{y}_n)$$

$$T = T_a = \min\{n \geq m_0 : \Lambda_n > a\}$$

where

$$\bar{x}_n = n^{-1} \sum_{j=1}^n X_j$$

$$\bar{y}_n = n^{-1} \sum_{j=1}^n Y_j.$$

The variable Λ_n is the logarithm of the generalized likelihood ratio statistic, which is commonly employed as a test statistic in fixed sample procedures. In using it as the basis for a sequential test, one observes pairs (X_i, Y_i) until the time $T \wedge m_1$ (m_1 being some fixed patient horizon), rejecting the hypothesis $p_1 = p_2$ iff $T \leq m_1$.

The problem of computing significance levels and power functions for the test procedure just described is not nearly so easy as for the fixed-sample generalized likelihood ratio test, whose asymptotic theory has been thoroughly developed. Siegmund and Gregory [13] have derived heuristically an asymptotic formula for the Type I error probability; their formula agrees formally with a result of Woodroffe [15] which was proved under assumptions too stringent to include this problem as an admissible case. This formula is contained in Theorem 1 below.

THEOREM 1. Let ϵ_1 and ϵ_2 be fixed constants such that $(2 \log 2)^{-1} < \epsilon_1 < \epsilon_2$. Suppose that $m_0 = a\epsilon_1 + o(a)$ and $m_1 = a\epsilon_2 + o(a)$ (recall that $T_a = \min\{n > m_0 : \wedge_n > a\}$). Then as $a \rightarrow \infty$

$$(2.3) \quad a^{-1/2} e^a P_{p,p} \{T_a \leq m_1\} \rightarrow C(p; \epsilon_1, \epsilon_2)$$

for every $p \in (0,1)$. For $0 < p \leq 1/2$, $C(p; \epsilon_1, \epsilon_2) = C(2p - 1/2; \epsilon_1, \epsilon_2)$ and

$$(2.4) \quad C(p; \epsilon_1, \epsilon_2)$$

$$= \pi^{-1/2} \int_{\xi \in (0, 2p) \cap \{\xi: \epsilon_2^{-1} \leq I(\xi, 2p-\xi) \leq \epsilon_1^{-1}\}} v(\xi, 2p-\xi) [I(\xi, 2p-\xi)]^{-1/2} \cdot [p(1-p)/\xi(1-\xi)(2p-\xi)(1+\xi-2p)]^{1/2} d\xi,$$

and

$$(2.5) \quad v(p_1, p_2) = \lim_{a \rightarrow \infty} E_{p_1, p_2} e^{-(\wedge_T - a)}.$$

That the limit in (2.5) exists (except for a countable set of (p_1, p_2) for which $p_1 \neq p_2$) is a consequence of Theorem 1 of Lai and Siegmund [9]. In fact, Woodroffe [16] has obtained an integral formula for the function $v(p_1, p_2)$ which is explicit enough to allow numerical integration.

The restriction on the initial sample size m_0 is rather peculiar and deserves some comment. Notice that the function $I(p_1, p_2)$ is bounded for $(p_1, p_2) \in [0,1]^2$: it achieves a maximum of $2 \log 2$ at the points $(0,1)$ and $(1,0)$. Thus $\wedge_n > a$ can occur only if

$n > a/2 \log 2$. Moreover, if $\Lambda_n > a$ for some n close to $a/2 \log 2$, then $(\bar{x}_n, \bar{y}_n)$ must be close to either $(0,1)$ or $(1,0)$; since in most conceivable applications neither p_1 nor p_2 would be close to zero or one, it would be somewhat unsettling to terminate the experiment on the basis of such an anomolous sample. A larger initial sample size protects against this possibility.

In the following sections an analogue of Theorem 1 will be formulated and proved under the assumption that the observations are from a multiparameter exponential family. This theorem will have one major shortcoming: namely, it will be necessary to impose even more stringent requirements on the initial sample size. (For the problem discussed in this section, the hypotheses of Theorem 2 would require $\epsilon_1 > (\log 2)^{-1}$, i.e., that the initial sample size be twice as large as Theorem 1 requires it to be.) The mathematical difficulty which necessitates the stronger conditions stems from the fact that large deviations theorems need not in general be uniform near the "boundary" of an exponential family. Fortunately, this difficulty disappears in many concrete cases of practical importance: for instance, whenever the mean parameter space is all of $\mathbb{R}^p$; and also in multinomial families.

We will give a (somewhat sketchy) proof of Theorem 1 for those cases where

$$(2.6) \quad N(\epsilon_1, \epsilon_2; p) = \{(r, 2p-r) : 0 \leq r \leq 2p \text{ and}$$

$$\epsilon_2^{-1} \leq I(r, 2p-r) \leq \epsilon_1^{-1}\}$$

is contained in the open square $(0,1) \times (0,1)$. The argument has two steps: first we show that only those sample paths for which $(\bar{x}_T, \bar{y}_T)$ is "near" N contribute substantially to the probability in (2.3); then we perform a local analysis near N .

PROPOSITION 1. As $a \rightarrow \infty$

$$(2.7) \quad P_{p,p} \{T_a \leq a \epsilon_2^{-1}; \text{dist}((\bar{x}_T, \bar{y}_T), N) > a^{-1/2} \log a\} = o(a^{-k} e^{-a})$$

for every $k > 0$.

NOTE: In adapting the arguments presented here to the more general problem discussed in the following sections the primary difficulty is in obtaining analogues of Proposition 1 (cf. Section 6). It is because of these difficulties that the more stringent assumptions on initial sample size are necessary.

PROOF. This is based on the "fundamental identity of sequential analysis," viz.,

$$(2.8) \quad P_{p,p}(A) = \int_A L_T dQ = \int_0^1 \int_0^1 (E_{p_1, p_2} 1_A L_T) dp_1 dp_2 ,$$

where

$$(2.9) \quad Q(B) = \int_0^1 \int_0^1 P_{p_1, p_2}(B) dp_1 dp_2$$

and

$$(2.10) \quad L_n = (n+1)^2 \binom{n}{nx_n} \binom{n}{ny_n} p^{n(\bar{x}_n + \bar{y}_n)} (1-p)^{n(2-\bar{x}_n - \bar{y}_n)} .$$

(Here $Q(B)$ is defined for all events $B \in \mathcal{F}((X_1, Y_1), (X_2, Y_2), \dots)$, and (2.8) holds for all A in the "stopped" σ -algebra of events A such that $A \cap \{T \leq n\} \in \mathcal{F}((X_1, Y_1), \dots, (X_n, Y_n))$ for all n).

Stirling's formula (cf. Feller [6], Chapter II, inequality (9.15)) provides a (crude) upper bound for L_n :

$$(2.11) \quad L_n \leq C n^4 e^{-\Lambda_n} e^{-\xi_n}$$

where

$$(2.12) \quad \xi_n = n[2H((\bar{x}_n + \bar{y}_n)/2) - (\bar{x}_n + \bar{y}_n)\log(p/(1-p)) - 2\log(1-p)]$$

for some constant $C > 0$. Now

$$H(\omega) = \omega \log(p/(1-p)) - 2 \log(1-p)$$

is a strictly convex, smooth, nonnegative function of $\omega \in (0,1)$ which is zero for $\omega = p$ and satisfies $H''(\omega) > 0$. Thus there is a constant $C^* > 0$ such that $(x,y) \in (0,1)^2$ and

$$(2.3) \quad \text{dist}((x,y), \{(r, 2p-r) : 0 \leq r \leq 2p\}) > \delta$$

implies

$$(2.14) \quad 2H((x+y)/2) - (x+y)\log(p/(1-p)) - 2 \log(1-p) > C^* \delta^2.$$

Clearly (2.8), (2.11), and (2.14) imply that for every $\delta > 0$

$$(2.15) \quad P_{p,p} \{T_a \leq a \epsilon_2^{-1}; \text{dist}((\bar{x}_T, \bar{y}_T), \{(r, 2p-r) : 0 \leq r \leq 2p\}) > \delta a^{-1/2} \log a\} \\ = o(a^{-k} e^{-a})$$

for every $k > 0$, since $\Lambda_T > a$.

Similarly it may be shown that if $T \leq a\epsilon_2^{-1}$,
 $\text{dist}((\bar{x}_T, \bar{y}_T), N) > a^{-1/2} \log a$, but

$$\text{dist}((\bar{x}_T, \bar{y}_T), \{(r, 2p-r), 0 \leq r \leq 2p\}) \leq \delta a^{-1/2} \log a$$

for some (sufficiently small) $\delta > 0$, then

$$(2.16) \quad \Lambda_T - a \geq C^{**} \log a ;$$

this together with (2.8) and (2.11) imply

$$(2.17) \quad P_{p,p} \{T \leq a\epsilon_2^{-1}; \text{dist}((\bar{x}_T, \bar{y}_T), N) > a^{-1/2} \log a;$$

$$\text{dist}((\bar{x}_T, \bar{y}_T), \{(r, 2p-r), 0 \leq r \leq 2p\}) \leq \delta a^{-1/2} \log a\}$$

$$= o(a^{-k} e^{-a})$$

for all $k > 0$. This and (2.15) imply (2.7). ///

For the next step of the proof we will again exploit the fundamental identity of sequential analysis, but with a new probability measure, which we will again refer to as Q . Let

$$(2.18) \quad Q(B) = \int_{r=0}^{2p} P_{r, 2p-r}(B) dr / (2p) ;$$

then

$$(2.19) \quad dP_{p,p}^{(n)} / dQ^{(n)} = L_n = \left[\int_{r=0}^{2p} r^{\bar{n}x_n(1-r)} n^{\bar{n}x_n} (2p-r)^{\bar{n}y_n} \cdot (1+r-2p)^{n-\bar{n}y_n} dr / (2p) \right]^{-1}$$

(where $P_{p,p}^{(n)}$ and $Q^{(n)}$ denote the restrictions of $P_{p,p}$ and Q to $\mathfrak{F}((X_1, Y_1), \dots, (X_n, Y_n))$).

PROPOSITION 2. Suppose that for some $(r, 2p-r) \in N$,

$$(2.20) \quad \text{dist}((\bar{x}_n, \bar{y}_n), (r, 2p-r)) \leq n^{1/2+1/7}.$$

Then as $n \rightarrow \infty$

$$(2.21) \quad L_n \sim e^{-\Lambda_n \left(\frac{n}{\pi}\right)^{1/2}} (2p) ((2r(1-r))^{-1} + (2(2p-r)(1+r-2p))^{-1})^{1/2} \\ \cdot \exp \left\{ \left(\frac{n}{2}\right) \begin{bmatrix} \bar{x}_n - r \\ \bar{y}_n + r - 2p \end{bmatrix}^T M_r \begin{bmatrix} \bar{x}_n - r \\ \bar{y}_n + r - 2p \end{bmatrix} \right\}$$

where

$$(2.22) \quad M_r = \begin{bmatrix} (r(1-r))^{-1} & ((2p-r)(1+r-2p))^{-1} \\ (r(1-r))^{-1} & ((2p-r)(1+r-2p))^{-1} \end{bmatrix} \\ - (p(1-p))^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Relation (2.21) holds uniformly for $(\bar{x}_n, \bar{y}_n)$ satisfying (2.20) with $(r, 2p-r) \in N$.

The proof of this is omitted: it is a straightforward but tedious exercise in the use of Laplace's method of asymptotic expansion.

The strategy for the rest of the proof is to show that for each $(r, 2p-r) \in N$, $(\Lambda_T - a)$ and

$$\sqrt{T} \begin{pmatrix} \bar{x}_T - r \\ \bar{y}_T + r - 2p \end{pmatrix}$$

are approximately independent under $E_{r,2p-r}$ as $a \rightarrow \infty$. For then the Central Limit Theorem, the Nonlinear Renewal Theorem of Lai and Siegmund [9], and the fact that

$$(2.23) \quad \frac{T_a}{a} P_{r,2p-r} \rightarrow \frac{1}{I(r,2p-r)}$$

will make possible the evaluation of $E_{r,2p-r} a^{-1/2} e^{-a} L_T 1_A$, where

$$(2.24) \quad A = \{T_a \leq a e^{-1}; \text{dist}((\bar{x}_T, \bar{y}_T), N) \leq a^{-1/2} \log a\}.$$

Nearly all of the technical difficulties associated with this program are obviated by the following inequalities.

LEMMA 1. Let S_n have a binomial distribution $BI(n,p)$ under P_p , $p \in [0,1]$. Then for each $k > 0$, $\delta > 0$, and $\alpha > 0$

$$(2.25) \quad \max_{0 \leq p \leq 1} P_p \{|S_n - np| > \delta n^{1/2} \log n\} = o(n^{-k})$$

and

$$(2.26) \quad \max_{0 \leq p \leq 1} P_p \{|S_n - np| > \delta n^{1/2 + \alpha}\} = o(e^{-\delta n^\alpha}).$$

PROOF. Using the Markov inequality,

$$\begin{aligned}
P_p \{S_n - np < n^{1/2} f(n)\} &\leq E_p \exp\{-\beta n^{-1/2}(S_n - np)\}/e^{-\beta f(n)} \\
&= e^{\beta p n^{1/2}} (1 - p + p e^{-\beta n^{-1/2}})^n / e^{-\beta f(n)} \\
&= \exp\{\beta p n^{1/2} + n(-\beta p n^{-1/2} + O(\frac{1}{n}))\} / e^{-\beta f(n)} = O(1)/e^{-\beta f(n)}
\end{aligned}$$

for $\beta > 0$, $f(n) < 0$. It is clear from the Taylor series expansion that the $O(1)$ term is uniform in p . The reverse inequality may be obtained similarly. ///

COROLLARY 1. As $a \rightarrow \infty$

$$(2.27) \quad \max_{(r, 2p-r) \in N} P_{r, 2p-r} \{T_a \notin [n_2, n_3]\} = o(e^{-a^{1/4}})$$

$$(2.28) \quad \max_{(r, 2p-r) \in N} P_{r, 2p-r} \{|\zeta_n - \zeta_{n_1}| > C/\log a,$$

$$\text{some } n \in [n_2, n_3]\} = o(e^{-a^{1/16}})$$

$$\begin{aligned}
(2.29) \quad \max_{(r, 2p-r)} P_{r, 2p-r} \{|\bar{x}_n - \bar{x}_{n_1}| + |\bar{y}_n - \bar{y}_{n_1}| > C a^{-1/2} \log a\} \\
= o(e^{-a^{1/32}}) ;
\end{aligned}$$

here

$$(2.30) \quad n_1 = n_1(a, r) = \llbracket a/I(r, 2p-r) - a^{1/2 + \eta} \rrbracket$$

$$(2.31) \quad n_2 = n_2(a, r) = \llbracket a/I(r, 2p-r) - a^{1/2 + \eta/2} \rrbracket$$

$$(2.32) \quad n_3 = n_3(a, r) = \llbracket a/I(r, 2p-r) + a^{1/2 + \eta/2} \rrbracket$$

$$(2.34) \quad \zeta_n = \zeta_n(r) = n \begin{pmatrix} \bar{x}_n - r \\ \bar{y}_n + r - 2p \end{pmatrix}^T M_r \begin{pmatrix} \bar{x}_n - r \\ \bar{y}_n + r - 2p \end{pmatrix}$$

and $\eta \in (0, 1/32)$ is some fixed constant. Relations (2.28) and (2.29) hold for all $C > 0$.

The corollary is an easy consequence of the preceding lemma.

Define

$$(2.35) \quad A_r = A \cap \{T_a \in [n_2, n_3]\} \\ \cap \{|\zeta_T(r) - \zeta_{n_1}(r)| \leq 1/\log a\} \\ \cap \{|\bar{x}_T - \bar{x}_{n_1}| + |\bar{y}_T - \bar{y}_{n_1}| \leq a^{-1/2} \log a\} \\ \cap \{|\bar{x}_{n_1} - r| + |\bar{y}_{n_1} + r - 2p| \leq a^{-1/2 + \eta/4}\} ;$$

by Proposition 2

$$(2.36) \quad 1_A a^{-1/2} e^a L_T = o(e^{C \log^2 a})$$

for some $C > 0$, so the Corollary implies that

$$(2.37) \quad a^{-1/2} e^a P_{p,p}(A) = \int_0^{2p} E_{r, 2p-r} (1_A a^{-1/2} e^a L_T) dr / 2p \\ = \int_0^{2p} E_{r, 2p-r} (1_{A_r} a^{-1/2} e^a L_T) dr / 2p + o(1) .$$

Now by Proposition 2 and (2.35),

$$\begin{aligned}
(2.38) \quad 1_{A_r} a^{-1/2} e^a L_T &\sim e^{-(\wedge_T - a)} \left(\frac{1}{\pi} I(r, 2p-r) \right)^{1/2} (2p) \\
&\cdot (2r(1-r)^{-1} + (2(2p-r)(1+r-2p))^{-1})^{1/2} \\
&\cdot \exp\{\zeta_{n_1}(r)/2\} 1_{A_r},
\end{aligned}$$

and this holds uniformly for $r \in N$ on A_r .

The "asymptotic independence" argument is completed by the following result.

PROPOSITION 3. For each $r \in N$ such that the random variable

$I(r, 2p-r) + (X_1 - r)\partial I/\partial p_1|_{(r, 2p-r)} + (Y_1 + r - 2p)\partial I/\partial p_2|_{(r, 2p-r)}$
has a nonlattice distribution when $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \sim P_{r, 2p-r}$,

$$(2.39) \quad E_{r, 2p-r} [e^{-(\wedge_T - a)} | \mathfrak{F}_{n_1}] = v(r, 2p-r) \frac{P_{r, 2p-r}}{0} \cdot 0.$$

This result is implicit in the proof of the Nonlinear Renewal Theorem given by Lai and Siegmund [9].

It is relatively easy to deduce Theorem 1 from (2.37)-(2.39). Uniform integrability problems may be handled by using (2.36), the Lemma, the Corollary, and the Berry-Esseen Theorem (for random vectors). The details of these arguments are straightforward but tedious, and will be omitted: they would, perhaps, serve only to obscure the basic argument. The bloodthirsty reader should rest assured that his appetite for raw, gory arguments (and detailed obscurity) will almost certainly be satisfied by the end of this work.

3. Preliminaries Concerning Exponential Families and Statement of the Main Result

Let $X_1, X_2, \dots$ be an i.i.d. sequence of random vectors each with law P_θ . We will not distinguish between P_θ , the measure on $\mathbb{R}^p$, and P_θ , the measure on the σ -algebra $\mathcal{J}(X_1, X_2, \dots)$. Recall that if $X \sim P_\theta$, then

$$(3.1) \quad E_\theta X = \nabla_\theta \psi(\theta) = \mu_\theta$$

$$\text{cov}_\theta X = \nabla_\theta^2 \psi(\theta) = \mathbb{I}(\theta) ;$$

since we have assumed ψ to be strictly convex, $\mathbb{I}(\theta)$ is perforce positive definite for each $\theta \in \Omega$. Assume that $\mu_0 = 0$.

Let $\Gamma = \{\mu_\theta; \theta \in \Omega\}$ be the mean parameter space. Because $\mathbb{I}(\theta) = \nabla_\theta^2 \psi(\theta)$ is strictly positive definite, the map

$$(3.2) \quad \mu_\cdot : \Omega \rightarrow \Gamma \quad \text{by}$$

$$\theta \mapsto \mu_\theta$$

is a diffeomorphism (this is the Inverse Function Theorem of Calculus). Thus although Γ need not be convex, it is an open subset of $\mathbb{R}^p$.

The (nonnegative) function

$$(3.3) \quad \phi(x) = \sup_{\theta \in \Omega} (\theta^T x - \psi(\theta))$$

is the "convex dual" of ψ . For $x \in \Gamma$ the supremum in (3.3) is uniquely attained at that θ for which $\mu_\theta = x$; henceforth this θ will be referred to as $\hat{\theta}(x)$. It is evident that for $x \in \Gamma$

$$(3.4) \quad \nabla_x \phi(x) = \hat{\theta}(x)$$

$$\nabla_x^2 \phi(x) = \hat{\Phi}(\hat{\theta}(x))^{-1}.$$

Moreover, since $0 \in \Omega$ (by (1.1)), the set

$$(3.5) \quad K_b = \{x \in \mathbb{R}^P : \phi(x) \leq b\}$$

is compact for each $b > 0$.

Recall that Ω_1 is a smooth, relatively closed q_1 -dimensional submanifold of Ω and Ω_0 is a smooth relatively closed q_0 -dimensional submanifold of Ω_1 . Since $\theta \rightarrow \mu_\theta$ is a diffeomorphism, $\Gamma_1 = \{\mu_\theta : \theta \in \Omega_1\}$ are smooth relatively closed submanifolds of Γ . (NOTE: A convenient and elementary source of information concerning the topological and geometric concepts used here is Guillemin and Pollack [8].) Define convex functions ϕ_0 and ϕ_1 by

$$(3.6) \quad \phi_1(x) = \sup_{\theta \in \Omega_1} (\theta^T x - \psi(\theta)) ;$$

the log generalized likelihood ratio statistic is then

$$(3.7) \quad \Lambda_n = n(\phi_1(S_n/n) - \phi_0(S_n/n)) .$$

It is apparent that the behavior of the functions ϕ_0 and ϕ_1 will play a crucial role in all that follows.

Unfortunately, for a given $x \in \mathbb{R}^P$ the supremum in (3.6) need not be uniquely attained. For $\theta_0 \in \Omega_1$, a necessary condition for

$$(3.8) \quad \theta_0^T x - \psi(\theta_0) = \sup_{\theta \in \Omega_1} (\theta^T x - \psi(\theta))$$

is

$$(3.9) \quad x - \mu_{\theta_0} \in T_{\Omega_1}(\theta_0)^\perp$$

where $T_{\Omega_1}(\theta_0)$ is the space of tangent vectors to Ω_1 in $\mathbb{R}^P$.

(Throughout the paper we will use the notation $TN(y)$ to denote the space of tangent vectors to N at y : i.e., if N is a q -dimensional submanifold of $\mathbb{R}^P$, then for $y \in N$

$$(3.10) \quad TN(y) = \{v \in \mathbb{R}^P : \exists \text{ smooth } g : [-1,1] \rightarrow N \\ \text{with } g(0) = y \text{ and } g'(0) = v\} .$$

For each $y \in N$ $TN(y)$ is a q -dimensional vector subspace of $\mathbb{R}^P$.) Let

$$(3.11) \quad U_1 = \{x \in \Gamma : \text{the supremum in (3.6) is attained uniquely} \\ \text{at some point } \hat{\theta}_1(x) \in \Omega_1\} .$$

LEMMA 1. For each $\theta \in \Omega_1$ the affine space $\mu_\theta + T_{\Omega_1}(\theta)^\perp$ intersects Γ_1 transversally at μ_θ . Furthermore, for each $\theta \in \Omega_1$ there is a neighborhood $N(\mu_\theta)$ of μ_θ (open in Γ) such that $N(\mu_\theta) \subset U_1$ and such that

$$\hat{\theta}_1 : N(\mu_\theta) \rightarrow \Omega_1$$

is a smooth submersion. If $x \in \Gamma$ has an open neighborhood $N_x \subset U_1$ such that $\hat{\theta}_1 : N_x \rightarrow \Omega_1$ is a smooth submersion, then

$$(3.12) \quad \nabla_x \phi_1(x) = \hat{\theta}_1(x) ,$$

$$(3.13) \quad v^T \nabla_x^2 \phi_1(x) v > 0 \quad \forall v \in T_{\Omega_1}(\hat{\theta}_1(x)) ,$$

and

$$(3.14) \quad \omega^T \nabla_x^2 \phi_1(x) \omega = 0 \quad \forall \omega \in T\Omega_1(\hat{\theta}_1(x))^\perp.$$

NOTE: Let N_1 and N_2 be smooth submanifolds of $\mathbb{R}^p$, and let $y \in N_1 \cap N_2$. Then N_1 and N_2 are said to intersect transversally at y if the tangent spaces together span $\mathbb{R}^p$, i.e., if

$$(3.15) \quad TN_1(y) + TN_2(y) = \mathbb{R}^p.$$

A map $g : N_1 \rightarrow N_2$ is said to be submersive at $x \in N_1$ if for every $v \in TN_2(g(x))$ there exists a smooth $f : [-1,1] \rightarrow N_1$ with $f(0) = x$ such that $(g \circ f)'(0) = v$; i.e., if dg_x maps $TN_1(x)$ onto $TN_2(g(x))$.

PROOF OF THE LEMMA. Since $\dim(\mu_\theta + T\Omega_1(\theta)^\perp) + \dim(T\Gamma_1(\mu_\theta)) = p$, the transversality of $\mu_\theta + T\Omega_1(\theta)^\perp$ and Γ_1 at μ_θ will follow from showing

$$(3.16) \quad T\Omega_1(\theta)^\perp \cap T\Gamma_1(\mu_\theta) = \{0\}.$$

For (3.16), suppose $g : [-1,1] \rightarrow \Gamma_1$ is a smooth map such that $g(0) = \mu_{\theta_0}$; then since $\nabla\psi : \Omega \rightarrow \Gamma$ is a diffeomorphism, there is a smooth $f : [-1,1] \rightarrow \Omega_1$ such that $f(0) = \theta_0$ and $g(t) = \nabla_{\theta} \psi(f(t))$. Now

$$\begin{aligned} g'(0) &= \nabla_{\theta}^2 \psi(f(0)) \cdot f'(0) \\ &= \ddagger(\theta_0) \cdot f'(0). \end{aligned}$$

Since $f'(0) \in T\Omega_1(\theta_0)$ and $\ddagger(\theta_0)$ is positive definite, it is impossible for $g'(0) \perp T\Omega_1(\theta_0)$. This proves (3.16).

Fix $\theta_0 \in \Omega_1$; there exists a neighborhood $N(\mu_{\theta_0})$ in Γ such that if $x \in N(\mu_{\theta_0})$, then

$$(3.17) \quad \sup\{(\theta^T x - \psi(\theta)) : \theta \in \Omega_1 \text{ and } \mu_\theta \notin N(\mu_{\theta_0})\} \\ \leq \sup\{(\theta^T x - \psi(\theta)) : \theta \in \Omega_1 \text{ and } \mu_\theta \in N(\mu_{\theta_0})\}$$

and such that the supremum on LHS (3.17) is not attained. Now $N(\mu_{\theta_0})$ may be chosen small enough that the affine spaces $\mu_\theta + T\Omega_1(\theta)^\perp$ give a "smooth fibration" of $N(\mu_{\theta_0})$: i.e., for each $x \in N(\mu_{\theta_0})$ there is a unique θ_x for which $x \in \mu_{\theta_x} + T\Omega_1(\theta_x)^\perp$, and such that the map $x \rightarrow \theta_x$ is smooth and submersive. (This fact relies on the fact that the spaces $\mu_\theta + T\Omega_1(\theta)^\perp$ intersect Γ_1 transversally at μ_θ , together with the fact that the map $\theta \rightarrow T\Omega_1(\theta)$ is a smooth mapping into the set of q_1 -dimensional vector subspaces of $\mathbb{R}^P$.) But the necessary condition (3.9) together with (3.16) implies that for $x \in N(\mu_{\theta_0})$, $\theta_x = \hat{\theta}_1(x)$.

Next, suppose that $x \in \Gamma$ has an open neighborhood $N_x \subset U_1$ such that $\hat{\theta}_1 : N_x \rightarrow \Omega_1$ is a smooth submersion. Then

$$\begin{aligned} (\partial/\partial x_j)\phi_1(x) &= (\partial/\partial x_j)(\hat{\theta}_1(x)^T x - \psi(\hat{\theta}_1(x))) \\ &= (\hat{\theta}_1(x))_j + \sum_{k=1}^P [(\partial/\partial x_j)(\hat{\theta}_1(x))_k] \cdot x_k \\ &\quad - \sum_{k=1}^P [(\partial/\partial x_j)(\hat{\theta}_1(x))_k] \cdot (\partial\psi/\partial \theta_k)(\hat{\theta}_1(x)) = (\hat{\theta}_1(x))_j \end{aligned}$$

since by (3.9) $x - \nabla_\theta \psi(\hat{\theta}_1(x)) \perp T\Omega_1(\hat{\theta}_1(x))$. It now follows from (3.12) that

$$(3.18) \quad \nabla_x^2 \phi_1(x) = ((\partial/\partial x_j)(\hat{\theta}_1(x))_k)_{j,k=1,\dots,p} ;$$

since $x \rightarrow \hat{\theta}_1(x)$ is submersive this matrix includes $T\Omega_1(\hat{\theta}_1(x))$ in its range, and clearly $T\Omega_1(\hat{\theta}_1(x))^\perp$ is contained in the kernel. Thus $\nabla_x^2 \phi_1(x)$ is invertible on $T\Omega_1(\hat{\theta}_1(x))$. On the other hand it is non-negative definite on $\mathbb{R}^p$ since ϕ_1 is convex. Consequently it must be strictly positive definite on $T\Omega_1(\hat{\theta}_1(x))$. It may also be shown using (3.18) that $\nabla_x^2 \phi_1(x) \upharpoonright T\Omega_1(\hat{\theta}_1(x))^\perp = 0$. ///

Lemma 1 gives a partial indication of the importance of two topological regularity properties: namely, transversality conditions and the submersiveness of the MLE maps. Another reason the transversality conditions figure in the analysis stems from the following purely topological fact, which will be exploited in Section 6.

LEMMA 2. Suppose V is an open subset of $\mathbb{R}^p$, and N_1, N_2 are relatively closed submanifolds of V . Let K be a compact subset of V such that if $x \in N_1 \cap N_2 \cap K$, then N_1 and N_2 meet transversally at x . Then for any $\epsilon > 0$ there exist $\delta > 0$, $\delta^* > 0$, and $a_0 > 0$ such that for any $a \geq a_0$ and any $y \in V$

$$(3.19) \quad \text{dist}(y, N_1 \cap N_2 \cap K) > \epsilon/a$$

$$\text{dist}(y, N_1 \cap K) < \delta/a$$

implies

$$(3.20) \quad \text{dist}(y, N_2) > \delta^*/a .$$

In other words a point cannot be far from the intersection without being far from one or the other of the two manifolds. This is manifestly untrue of manifolds which intersect nontransversally: e.g.,

$$N_1 = \{(x,y) \in \mathbb{R}^2 : y = x^2\}$$

$$N_2 = \{(x,y) \in \mathbb{R}^2 : y = 0\} \quad .$$

PROOF. We will give only a rough outline of the argument. Suppose first that N_1 and N_2 are affine subspaces of $\mathbb{R}^p$: the existence of δ and δ^* follows from the construction of disjoint angular corridors around N_1 and N_2 as illustrated by Figure 3.1

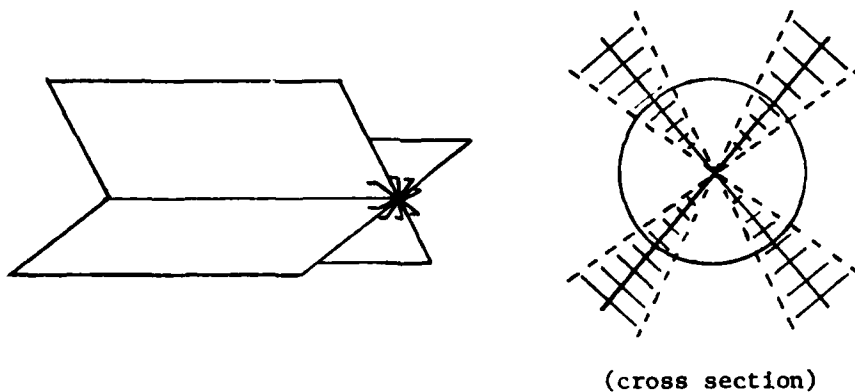


Figure 3.1

In the general case, N_1 and N_2 may be approximated to first order by the appropriate translates of their tangent spaces; if N_1 and N_2 intersect transversally at x , then the tangent spaces $TN_1(x)$ and $TN_2(x)$ intersect transversally. Angular corridors may then be constructed as before. Thus for each $x \in N_1 \cap N_2 \cap K$ there is a closed neighborhood U_x in $\mathbb{R}^p$ and constants $\delta_x, \delta_x^*, a_0(x)$ such that for $a \geq a_0(x)$ and $y \in V$

$$\text{dist}(y, N_1 \cap N_2 \cap K \cap U_x) > \epsilon/a$$

and

$$\text{dist}(y, N_1 \cap K \cap U_x) < \delta_x/a$$

imply

$$\text{dist}(y, N_2) > \delta_x^*/a .$$

The lemma now follows from a compactness argument (since N_1 and N_2 are relatively closed in V and K is compact, $N_1 \cap N_2 \cap K$ is compact). ///

The conclusion of the main theorem depends heavily on the assumption that the MLE maps behave nicely near a certain critical manifold, and also that the manifold Γ_1 not contort itself too strenuously in certain regions of Γ . Let $\theta_0 \in \Omega_0$, and define $\epsilon_0(\theta_0)$ to be the largest extended real such that the following three conditions are satisfied:

- I. For every ϵ , $0 < \epsilon < \epsilon_0$, the set $\{x \in \Gamma : \phi(x) - \theta_0^T x + \psi(\theta_0) \leq \epsilon\}$ is compact.
- II. For each $x \in \Gamma_1 \cap (\mu_{\theta_0} + \mathcal{M}_0(\theta_0)^\perp)$ such that $\phi_1(x) - \phi_0(x) < \epsilon_0$ there exists a neighborhood N_x of x in Γ such that $N_x \subset U_0$ and $\hat{\theta}_0$ is a smooth submersion on N_x , with $\hat{\theta}_0(x) = \theta_0$.
- III. For each $x \in \Gamma_1 \cap (\mu_{\theta_0} + \mathcal{M}_0(\theta_0)^\perp)$ such that $\phi_1(x) - \phi_0(x) < \epsilon_0$, Γ_1 intersects $(\mu_{\theta_0} + \mathcal{M}_0(\theta_0)^\perp)$ transversally at x , and Γ_1 intersects the level surface $\{y \in \Gamma : \phi(y) - \theta_0^T y = \phi(x) - \theta_0^T x\}$ transversally at x .

Note that there is always a positive ϵ such that $\{x \in \Gamma : \phi(x) - \theta_0^T x + \psi(\theta_0) \leq \epsilon\}$ is compact, since $\{x \in \mathbb{R}^p : \phi(x) - \theta_0^T x + \psi(\theta_0) \leq \epsilon\}$ is compact (cf. (3.5), and reparametrize the exponential family). That $\epsilon_0(\theta_0) > 0$ may be deduced from this and Lemma 1. It should be noticed that in the special case $\Omega = \Omega_1$ condition III is automatically satisfied, and in case $\Gamma = \mathbb{R}^p$, condition I is automatically satisfied.

THEOREM 2. Suppose $0 < \epsilon_1 < \epsilon_2 < \epsilon_0(\theta_0)$; recall that

$$(3.21) \quad T_a = T = \min\{n \geq a\epsilon_2^{-1} : \wedge_n > a\}.$$

Then as $a \rightarrow \infty$

$$(3.22) \quad P_{\theta_0}\{T_a \leq a\epsilon_1^{-1}\} \sim a^{(q_1 - q_0)/2} \cdot e^{-a} \cdot c(\epsilon_1, \epsilon_2; \theta_0)$$

where

$$\begin{aligned}
(3.23) \quad c(\epsilon_1, \epsilon_2; \theta_0) &= \int_{M(\epsilon_1, \epsilon_2; \theta_0)} v(y) (2\pi(\phi(y) - \phi_0(y)))^{-(q_1 - q_0)/2} \\
&\cdot [\det(H_2(y)) / \det(\sharp(\hat{\theta}(y))(H_1(y) + H_3(y)))]^{1/2} \\
&\cdot \sigma(dy)
\end{aligned}$$

$$(3.24) \quad M(\epsilon_1, \epsilon_2; \theta_0) = \{y \in \Gamma_1 \cap (\mu_{\theta_0} + \mathcal{M}_0(\theta_0)^\perp) : \epsilon_1 \leq \phi(y) - \phi_0(y) \leq \epsilon_2\}$$

$$(3.25) \quad v(y) = \lim_{a \rightarrow \infty} E_{\hat{\theta}(y)} e^{-(\wedge_T - a)}$$

$$(3.26) \quad H_1(y) = \sharp(\hat{\theta}(y))^{-1} \underline{P} H_2(y)^{-1} \underline{P} \sharp(\hat{\theta}(y))^{-1}$$

$$(3.27) \quad H_2(y) = \underline{P} \sharp(\hat{\theta}(y))^{-1} \underline{P} \uparrow (\mathcal{M}_0(\theta_0)^\perp \cap T\Gamma_1(y))$$

$$(3.28) \quad H_3(y) = \sharp(\hat{\theta}(y))^{-1} - \nabla_y^2 \phi_1(y) + \nabla_y^2 \phi_0(y) \quad .$$

$\underline{P}$ is the orthogonal projection operator onto the space $\mathcal{M}_0(\theta_0)^\perp \cap T\Gamma_1(y)$, and σ is the volume element measure for the manifold-with-boundary $M(\epsilon_1, \epsilon_2; \theta_0)$.

Many comments are in order. First, conditions I and III (transversality) imply that $M(\epsilon_1, \epsilon_2; \theta_0)$ really is a compact manifold-with-boundary: this is a consequence of the Implicit Function Theorem. Second, the " $\det H_2(y)$ " which appears in the numerator may be confusing: $H_2(y)$ is a (positive definite) operator on $T\Gamma_1(y) \cap \mathcal{M}_0(\theta_0)^\perp$, and the determinant is simply meant to be the product of its eigenvalues on $T\Gamma_1(y) \cap \mathcal{M}_0(\theta_0)^\perp$. Third, it remains

to be seen that the integral is finite, and in fact that the integrand is defined (cf. (3.25)). Notice that $\phi - \phi_0$ is a continuous function which is bounded away from zero on $M(\epsilon_1, \epsilon_2; \theta_0)$. The other two factors of importance require more care.

LEMMA 3. For every $y \in M(\epsilon_1, \epsilon_2; \theta_0)$ the matrix $H_1(y) + H_3(y)$ is strictly positive definite on $\mathbb{R}^p$.

PROOF. Since $\ddagger(\theta)$ is everywhere strictly P.D. it is clear that $H_1(y)$ is N.N.D. on $\mathbb{R}^p$ and strictly P.D. on $\mathcal{M}_0(\theta_0)^\perp \cap T\Gamma_1(y)$. Also if $y \in M(\epsilon_1, \epsilon_2; \theta_0)$, then since $\epsilon_2 < \epsilon_0(\theta_0)$ and $\epsilon_0(\theta_0)$ satisfies condition II, it follows from Lemma 1 that $\nabla_y^2 \phi_0(y)$ is N.N.D. on $\mathbb{R}^p$ and strictly P.D. on $\mathcal{M}_0(\theta_0)$.

Now consider $\nabla_y^2(\phi - \phi_1)(y)$. Since $\phi - \phi_1$ is a nonnegative function on $\mathbb{R}^p$ which is zero on Γ_1 , $\nabla_y^2(\phi - \phi_1)(y)$ is N.N.D. on $\mathbb{R}^p$ whenever $y \in \Gamma_1$, by Taylor's Theorem. Furthermore, if $y \in \Gamma_1$, then $\nabla_y^2 \phi_1(y)$ is zero on the vector subspace $\mathcal{M}_1(\hat{\theta}(y))^\perp$ (cf. (3.14) of Lemma 1). Thus $\nabla_y^2(\phi - \phi_1)(y)$ is strictly P.D. on $\mathcal{M}_1(\hat{\theta}(y))^\perp$ for each $y \in \Gamma_1$. Now since $(y + \mathcal{M}_1(\hat{\theta}(y))^\perp)$ intersects Γ_1 transversally at y (Lemma 1 again) it follows that $\nabla_y^2(\phi - \phi_1)(y)$ is strictly P.D. on $T\Gamma_1(y)$.

But $T\Gamma_1(y)^\perp + \mathcal{M}_0(\theta_0) + (\mathcal{M}_0(\theta_0)^\perp \cap T\Gamma_1(y)) = \mathbb{R}^p$, so $H_1(y) + H_3(y)$ is strictly P.D. on $\mathbb{R}^p$. ///

As for $v(y)$, the existence of the limit in (3.25) is a consequence of a general theorem of Lai and Siegmund [9]. In order that their theorem be applicable, however, a certain random walk

associated with the process $\{\wedge_n\}$ must be nonlattice. In the next lemma it is shown that this is so for almost every $y \in M(\epsilon_1, \epsilon_2; \theta_0)$.

LEMMA 4. For all $\epsilon_1 < \epsilon_2 < \epsilon_0(\theta_0)$,

$$(3.29) \quad \sigma\{y \in M(\epsilon_1, \epsilon_2; \theta_0) : (\phi_1(y) - \phi_0(y)) + (X_1 - y)^T \nabla_y (\phi_1 - \phi_0)(y)$$

has a nonlattice distribution when $X_1 \sim P_{\hat{\theta}(y)}\} = 0$,

where $\sigma(\cdot)$ denotes the volume element measure on $M(\epsilon_1, \epsilon_2; \theta_0)$.

PROOF. Call a point $\omega \in \mathbb{R}^p$ a support point of the exponential family if for some $\theta \in \Omega$ $P_\theta\{u : |u - \omega| < \epsilon\} > 0$ for each $\epsilon > 0$. Clearly if ω is a support point, then $P_{\theta^*}\{u : |u - \omega| < \epsilon\} > 0$ for every $\theta^* \in \Omega$ and $\epsilon > 0$. Because the covariance matrices $\mathbb{I}(\theta)$ are strictly P.D., there exist support points $\omega_1, \dots, \omega_p$ which form a (vector space) basis for $\mathbb{R}^p$.

Let

$$I(y) = \phi_1(y) - \phi_0(y), \quad y \in \mathbb{R}^p.$$

A necessary condition for $I(y) + (X_1 - y)^T \nabla_y I(y)$ to have a lattice distribution (under $P_{\hat{\theta}(y)}$) is that for any pair $\{\omega_i, \omega_j\}$ of the atoms there exist a rational number $q_{\{i,j\}}$ such that either

$$(3.30) \quad q_{\{i,j\}}(I(y) + (\omega_i - y)^T \nabla_y I(y)) = I(y) + (\omega_j - y)^T \nabla_y I(y)$$

or

$$(3.31) \quad q_{\{i,j\}}(I(y) + (\omega_j - y)^T \nabla_y I(y)) = I(y) + (\omega_i - y)^T \nabla_y I(y).$$

Suppose that there is a $y = y_0 \in M(\epsilon_1, \epsilon_2; \theta_0)$ such that conditions (3.30)-(3.31) are satisfied for $y = y_0$ and some particular set $\{q_{\{i,j\}}\}$ of $\binom{p}{2}$ rational numbers: we will show that there is a neighborhood $N(y_0)$ of y_0 such that if $y \in N(y_0) \cap \Gamma_1 \cap (\mu_{\theta_0} + \mathcal{N}_0(\theta_0)^\perp)$, then (3.30)-(3.31) are not satisfied for y and the same set of rationals $\{q_{\{i,j\}}\}$. By the countability of the rationals this will prove (27).

Suppose the indices are labelled so that alternative (3.30) holds. Consider the first order Taylor series of

$$(3.32) \quad f_{\{i,j\}}(y) = I(y)(1 - q_{\{i,j\}}) + (\omega_i - q_{\{i,j\}} \quad \omega_j - y(1 - q_{\{i,j\}}))^T \nabla I(y)$$

around $y = y_0$:

$$(3.33) \quad Tf_{\{i,j\}}(y) = (\omega_i - q_{\{i,j\}} \quad \omega_j - y_0(1 - q_{\{i,j\}}))^T \cdot \nabla_y^2 I(y_0) \cdot (y - y_0) \quad .$$

Since $\nabla_y^2 \phi_0(y_0) \upharpoonright \mathcal{N}_0(\theta_0)^\perp = 0$, Lemma 1 and the transversality condition III imply that $\nabla_y^2 I(y_0) \upharpoonright (\mathcal{N}_0(\theta_0)^\perp \cap T\Gamma_1(y_0))$ is strictly P.D. Consequently, because $\omega_1, \dots, \omega_p$ is a basis for $\mathbb{R}^p$, it follows that for each $u \in (\mathcal{N}_0(\theta_0)^\perp \cap T\Gamma_1(y_0))$ satisfying $|u| = 1$, there is a pair $\{i,j\}$ such that

$$(3.34) \quad Tf_{\{i,j\}}(y_0 + tu) \neq 0$$

whenever $t \neq 0$. Since $Tf_{\{i,j\}}$ is the principal term in $f_{\{i,j\}}$, and

since $B = \{u \in (\mathbb{T}\Omega_0(\theta_0))^\perp \cap \mathbb{T}\Gamma_1(y_0) : |u| = 1\}$ is compact, there is a $\delta > 0$ such that

$$f_{\{i,j\}}(y) \neq 0$$

for any $y \in \Gamma_1 \cap (\mu_{\theta_0} + \mathbb{T}\Omega_0(\theta_0))^\perp$ satisfying $0 < |y - y_0| < \delta$. This proves the lemma. ///

Although this lemma together with the result of Lai and Siegmund shows that the function $v(y)$ is well-defined for almost every y ($d\sigma$), there is as yet no hope of evaluating it. However, an important result of M. Woodroffe has $v(y)$ expressed as an integral involving only the characteristic function of the random variable considered in Lemma 4. Woodroffe's theorem makes possible the evaluation of the constant appearing in Theorem 2 by numerical integration in many cases of statistical interest: his paper [16] contains not only a proof of the theorem but several interesting examples of its use. (NOTE: Actually Woodroffe's theorem carries certain hypotheses concerning the smoothness of the underlying distributions which are unnecessary, as an elementary modification of his proof shows).

Theorem 2 generalizes another theorem of Woodroffe (Theorem 3 of [15]) which essentially covers the case $\Omega = \Omega_1$, but under smoothness conditions on the distributions $\{P_\theta\}$ which rule out all problems involving categorical data. His proof seems to be very much tied to these assumptions, and bears no resemblance to the approach used in this work.

For the proof of Theorem 2 we will assume that $0 \in \Omega_0$ and $\theta_0 = 0$. For arbitrary $\theta_0 \in \Omega_0$ we may always reduce to this case by reparametrizing and recentering the exponential family.

4. Normal Approximation in Several Dimensions

Certain refinements of the multidimensional Central Limit Theorem play a key role in the analysis which follows. Of primary importance are (1) bounds on the probability of moderate deviations of the sample mean, and (2) uniformity in the convergence of $n^{-1/2}(S_n - n\mu_\theta)$ (under P_θ) over compact subsets of the natural parameter space.

Let $\dagger$ be a symmetric positive definite matrix on $\mathbb{R}^p$, and let $Q_\dagger$ be the Gaussian measure on $\mathbb{R}^p$ with mean zero and covariance $\dagger$, i.e.,

$$(4.1) \quad Q_\dagger(A) = (\det \dagger)^{-1/2} (2\pi)^{-p/2} \int_A \exp\{-y^T \dagger^{-1} y/2\} dy$$

for Borel sets A . In addition, let α_n be the class of p -dimensional half-spaces

$$(4.2) \quad A_\ell(a) = \{x \in \mathbb{R}^p : \ell^T x > a\}$$

where

$$|\ell| = 1$$

and

$$0 \leq a \leq n^{1/6}/\log n.$$

PROPOSITION 1. Let K be any compact subset of Ω (the natural parameter space of the exponential family $\{P_\theta\}$, which is assumed to be an open set of $\mathbb{R}^p$). Then

$$(4.3) \quad \lim_{n \rightarrow \infty} \sup_{\theta \in K} \sup_{A \in \mathcal{G}_n} P_{\theta} \{ n^{-1/2} (S_n - n\mu_{\theta}) \in A \} / Q_{\mathbb{J}(\theta)}(A) = 1$$

and

$$(4.4) \quad \lim_{n \rightarrow \infty} \inf_{\theta \in K} \inf_{A \in \mathcal{G}_n} P_{\theta} \{ n^{-1/2} (S_n - n\mu_{\theta}) \in A \} / Q_{\mathbb{J}(\theta)}(A) = 1$$

The proof of this is a rather tedious modification of the proof of Cramér's Theorem (cf. Feller [7], Chapter XVI, Section 7). The only real novelty is the uniformity in θ . However, the third moments $\{E_{\theta} |\ell^T (X_1 - \mu_{\theta})|^3; |\ell| = 1\}$ are uniformly bounded away from ∞ , and the second moments $\{E_{\theta} |\ell^T (X_1 - \mu_{\theta})|^2; |\ell| = 1\}$ are uniformly bounded away from zero, for θ in any compact subset of Ω (recall that the covariance matrices $\mathbb{J}(\theta)$ were assumed to be positive definite on Ω). Thus the Berry-Esseen Theorem provides a bound for the error in the normal approximation to the P_{θ} -distribution of $n^{-1/2} (\ell^T (S_n - n\mu_{\theta}))$ which is uniform in θ and ℓ . Moreover, the errors in the Taylor series expansions used in the proof are all uniformly small, again by the compactness of K .

COROLLARY 1. Let K be a compact subset of Ω . Then for all $\delta > 0$ and $k > 0$

$$(4.5) \quad \max_{\theta \in K} P_{\theta} \{ |S_n - n\mu_{\theta}| > \delta n^{1/2} \log n \} = o(n^{-k})$$

and for all ϵ such that $0 < \epsilon < 1/6$

$$(4.6) \quad \max_{\theta \in K} P_{\theta} \{ |S_n - n\mu_{\theta}| > \delta \cdot n^{\frac{1}{2}} + \epsilon \} = O(\exp\{-\delta^2 n^{3\epsilon/2}\}) ,$$

as $n \rightarrow \infty$.

The proof is straightforward.

COROLLARY 2. Let K be a compact subset of Ω , and suppose $\theta \rightarrow A(\theta)$ is a continuous function of $\theta \in K$ with values in the group of symmetric positive definite $p \times p$ matrices. Then

$$(4.7) \quad E_{\theta} 1_{B(n,\theta)} \exp\{(S_n - n\mu_{\theta})^T (\hat{I}(\theta)^{-1} - A(\theta)) (S_n - n\mu_{\theta}) / 2n\} \\ \rightarrow (\det A(\theta) \det \hat{I}(\theta))^{-1/2}$$

as $n \rightarrow \infty$; $B(n,\theta)$ is the event

$$(4.8) \quad B(n,\theta) = \{ |S_n - n\mu_{\theta}| \leq n^{1/2} a_n \} ,$$

and $\{a_n\}$ is any sequence of constants such that $a_n \rightarrow \infty$ and $a_n = O(n^{1/6}/\log n)$. Furthermore, the convergence in (4.7) is uniform for $\theta \in K$, for each sequence $\{a_n\}$.

PROOF. This is accomplished in two stages, using Theorem 1 to establish the uniform integrability of the random variables, and Bhattacharya's multidimensional extension of the Berry-Esseen Theorem for the integration.

BHATTACHARYA'S THEOREM (cf. Bhattacharya and Rao [3], Corollary 15.2). Suppose $X_1, X_2, \dots$ are i.i.d. random vectors in $\mathbb{R}^p$ with mean zero, covariance I , and finite absolute third moment $\rho_3 = E|X_1|^3$.

Then for each bounded Borel measurable function $f: \mathbb{R}^p \rightarrow \mathbb{R}$

$$\begin{aligned} & |E f(S_n/n^{1/2}) - \int f(y) e^{-|y|^2/2} dy / (2\pi)^{p/2}| \\ & \leq c_1 M(f) \rho_3 n^{-1/2} + 2\omega(f; c_2 \rho_3 n^{-1/2}) \end{aligned}$$

where

$$\begin{aligned} M(f) &= \sup_{x, y \in \mathbb{R}^p} |f(x) - f(y)| \\ \omega(f; \epsilon) &= \sup_{u \in \mathbb{R}^p} \int_{y \in \mathbb{R}^p} \left[e^{-|y|^2/2} / (2\pi)^{p/2} \right] \\ & \quad \cdot \sup\{|f(x_1 + u) - f(x_2 + u)| : |x_1 - y| \vee |x_2 - y| < \epsilon\} dy \end{aligned}$$

$$S_n = X_1 + \dots + X_n,$$

and c_1, c_2 are universal constants (which may depend on the dimension p).

The idea is to apply this result to the functions

$$f_{\theta, b}(y) = g_{\theta}(y) 1\{g_{\theta}(y) < b\}$$

where

$$g_{\theta}(y) = \exp\{y^T (I - \ddagger(\theta))^{1/2} A(\theta) \ddagger(\theta)^{1/2} y / 2\}.$$

It is clear that for fixed b ,

$$\sup_{\theta \in K} M(f_{\theta, b}) < \infty$$

and

$$\lim_{\epsilon \downarrow 0} \sup_{\theta \in K} \omega(f_{\theta, b}; \epsilon) = 0 ,$$

and that

$$\lim_{b \uparrow \infty} \int f_{\theta, b}(y) e^{-|y|^2/2} dy / (2\pi)^{p/2} = (\det A(\theta) \sharp(\theta))^{-1/2}$$

uniformly for $\theta \in K$. Thus to prove (4.7), it suffices to show that for any $\epsilon > 0$ there is a b so large that

$$(4.9) \quad \sup_{\theta \in K} E_{\theta} 1_{B(n, \theta)} 1_{\{g_{\theta}(\sharp(\theta))^{-1/2} (S_n - n\mu_{\theta}) n^{-1/2} \geq b\}} \\ \cdot g_{\theta}(\sharp(\theta))^{-1/2} (S_n - n\mu_{\theta}) n^{-1/2} < \epsilon .$$

Let

$$\delta = 1 \wedge \inf \{x^T A(\theta) x / x^T \sharp(\theta)^{-1} x : \theta \in K, |x| = 1\} :$$

since K is compact, $\delta > 0$. Then for each $\theta \in K$ there is a polyhedron $R(\theta)$ such that for every $y \in R(\theta)$

$$y^T (\sharp(\theta)^{-1} - A(\theta)) y \leq 2$$

and

$$y^T \sharp(\theta)^{-1} y > 2(1 - \delta/2)^{-1} .$$

This follows from the definition of δ by piecing together patches of hyperplanes along the level surface $\{y \in \mathbb{R}^p : y^T (\sharp(\theta)^{-1} - A(\theta)) y = 2\}$ (cf. Figures 4.1 and 4.2). Furthermore, since $\sharp(\theta)$ and $A(\theta)$ are continuous in θ , $R(\theta)$ may be chosen "continuously" in θ :

in particular, it may be assumed that there is a finite integer m such that for each $\theta \in K$, $R(\theta)$ has no more than m distinct faces.

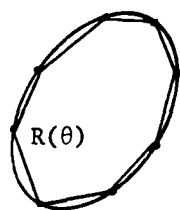


Figure 4.1

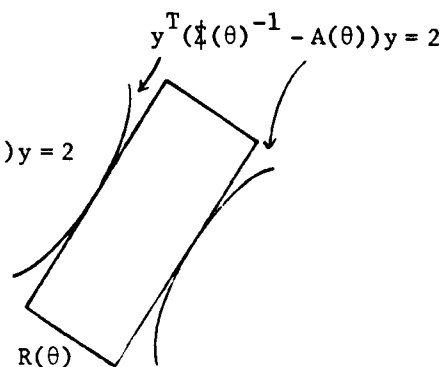


Figure 4.2

Each polyhedron $R(\theta)$ separates $\mathbb{R}^p$ into a bounded component and an unbounded component, which will be denoted $R^{\text{INT}}(\theta)$ and $R^{\text{EXT}}(\theta)$. Now Theorem 1 and a crude bound on the tail of the cumulative normal distribution function imply that for $b < a < c$, a_n , $\theta \in K$, and sufficiently large n

$$P_{\theta}\{(S_n - n\mu_{\theta})n^{-1/2} a^{-1} \in R^{\text{EXT}}(\theta)\} \leq 2m \cdot \exp\{-a^2/(1-\delta/2)\}.$$

But

$$\begin{aligned}
(4.10) \quad & E_{\theta} L_{B(n, \theta)} 1\{g_{\theta}(\dagger(\theta))^{-1/2} (S_n - n\mu_{\theta})n^{-1/2} \geq b\} \\
& \cdot g_{\theta}(\dagger(\theta))^{-1/2} (S_n - n\mu_{\theta})n^{-1/2} \\
& \leq \sum_{k=0}^{\infty} P_{\theta}\{(S_n - n\mu_{\theta})n^{-1/2} \in (\log^{1/2} b)(1 - \delta/4)^{-k/2} R^{\text{EXT}}(\theta); \\
& \quad |S_n - n\mu_{\theta}|n^{-1/2} < a_n\} \\
& \cdot \max\{g_{\theta}(\dagger(\theta))^{-1/2} y : \\
& \quad y \in (\log^{1/2} b)(1 - \delta/4)^{-k/2} R^{\text{EXT}}(\theta); \\
& \quad y \notin (\log^{1/2} b)(1 - \delta/4)^{-(k+1)/2} R^{\text{EXT}}(\theta)\} \\
& \leq 2m \sum_{k=0}^{\infty} \exp\{(\log b)[(1 - \delta/4)^{-(k+1)} - (1 - \delta/4)^{-k}(1 - \delta/2)^{-1}]\}
\end{aligned}$$

for sufficiently large n , and all $\theta \in K$. The series on RHS (4.10) can be made arbitrarily small by choosing b large; this proves (4.9), and thus (4.7). ///

5. Expansion of likelihood Functions

Let N be a smooth, compact, r -dimensional submanifold of Γ , and let $f(\cdot)$ be a smooth, strictly positive probability density on N (with respect to the "volume element" measure $\sigma(\cdot)$). Define

$$(5.1) \quad Q(A) = \int_N P_{\hat{\theta}(y)}(A) f(y) \sigma(dy) ;$$

thus Q is a probability measure on the σ -algebra $\mathcal{J}(X_1, X_2, \dots)$.

Furthermore, the measures P_{θ} and Q , when restricted to the σ -algebras $\mathcal{J}(X_1, \dots, X_n)$ (these restricted measures will be denoted $P_{\theta}^{(n)}$ and $Q^{(n)}$ respectively), are mutually absolutely continuous, and

$$(5.2) \quad dP_0^{(n)} / dQ^{(n)} = \left[\int_N e^{\hat{\theta}(y)^T S_n^{-n} \psi(\hat{\theta}(y))} f(y) \sigma(dy) \right]^{-1} .$$

The objective of this section is the derivation of a more tractable expression for $dP_0^{(n)} / dQ^{(n)}$ when S_n/n is near N , as $n \rightarrow \infty$.

It will be convenient to have some notation available for various matrices which will occur. Recall that the tangent space $TN(y)$ to N at y is the vector subspace of $\mathbb{R}^P$ defined by

$$(5.3) \quad TN(y) = \{v \in \mathbb{R}^P : \exists \text{ smooth } g : [-1, 1] \rightarrow N$$

$$\text{with } g(0) = y \text{ and } g'(0) = v\} ;$$

thus $TN(y)$ is an r -dimensional vector space, for every $y \in N$. Let $P_{TN(y)}$ denote the orthogonal projection operator from $\mathbb{R}^P$ to $TN(y)$, and let

$$(5.4) \quad H_1(y) = \ddagger(\hat{\theta}(y))^{-1} P_{TN(y)} H_2(y)^{-1} P_{TN(y)} \ddagger(\hat{\theta}(y))^{-1}$$

$$(5.5) \quad H_2(y) = P_{TN(y)} \sharp(\hat{\theta}(y))^{-1} P_{TN(y)} \uparrow TN(y) .$$

Note that $H_2(y)$ considered as an operator on $TN(y)$ is invertible, since $\sharp(\hat{\theta}(y))$ is invertible; furthermore, since $TN(y)$ is vector-space isomorphic to $\mathbb{R}^r$, $H_2(y)$ may be interpreted as an operator on $\mathbb{R}^r$, and $\det H_2(y)$ is then unambiguously defined.

PROPOSITION 1. Suppose $y_1 \in N$, and

$$(5.6) \quad S_n/n = y_1 + hn^{-1/2}$$

where $|h| < \delta n^{-1/7}$. Then as $n \rightarrow \infty$

$$(5.7) \quad dP_0^{(n)}/dQ^{(n)} \sim e^{-n\phi(S_n/n)} f(y_1)^{-1} (n/2\pi)^{r/2} \\ \cdot \exp\{h^T(\sharp(\hat{\theta}(y_1))^{-1} - H_1(y_1))h/2\} \\ \cdot \det(H_2(y_1))^{1/2} .$$

This relation holds uniformly for $y_1 \in N$ and $|(S_n/n) - y_1| < \delta n^{1/7-1/2}$, for every $\delta > 0$.

The proof is a relatively straightforward exercise in the use of Laplace's method. The basic idea is that for large n the only part of N which contributes to the integral in (5.2) is a small patch around y_1 (essentially of radius $n^{1/7-1/2} \log n$), and that the integral over this patch is approximately equal to the integral over the tangent space at y_1 . Because N is compact, all of the errors are uniformly small.

Since the relation (5.7) is crucial to all that follows, the argument will be given in detail. Taylor's Theorem yields

$$\begin{aligned}
 (5.8) \quad \hat{\theta}(y)^T x - \psi(\hat{\theta}(y)) - \phi(x) &= (y - y_1)^T \ddagger(\hat{\theta}(y_1))^{-1} (x - y_1) \\
 &\quad - \frac{1}{2} (y - y_1)^T \ddagger(\hat{\theta}(y_1))^{-1} (y - y_1) \\
 &\quad - \frac{1}{2} (x - y_1)^T \ddagger(\hat{\theta}(y_1))^{-1} (x - y_1) \\
 &\quad + o(|y_1 - y|^3) + o(|x - y_1|^3)
 \end{aligned}$$

by way of the identities

$$(5.9) \quad y_1 = \nabla_{\theta} \psi(\hat{\theta}(y_1))$$

$$\hat{\theta}(y_1) = \nabla_y \phi(y_1)$$

$$\ddagger(\hat{\theta}(y_1)) = \nabla_{\theta}^2 \psi(\hat{\theta}(y_1))$$

$$\ddagger(\hat{\theta}(y_1))^{-1} = \nabla_y^2 \phi(y_1) \quad .$$

Because N is compact, the remainder terms in (5.5) are uniformly small, i.e., there exist $C, \delta^* > 0$ such that whenever $y_1, y \in N$, $|y - y_1| < \delta^*$, and $|x - y_1| < \delta^*$,

$$(5.10) \quad o(|x - y_1|^3) < C|x - y_1|^3$$

$$o(|y - y_1|^3) < C|y - y_1|^3 \quad .$$

Relation (5.8) may now be used to integrate over the set of $y \in N$ which are within $n^{1/7-1/2} \log n$ of y_1 . If $x \in \Gamma$, and $|x - y_1| \leq \delta n^{1/7-1/2}$, then

$$\begin{aligned}
 (5.11) \quad & \int_N \exp\{n[\hat{\theta}(y)^T x - \psi(\hat{\theta}(y)) - \phi(x)]\} \\
 & 1\{|y - y_1| \leq n^{1/7-1/2} \log n\} f(y) \sigma(dy) \\
 & \sim \exp\{-n(x - y_1)^T \hat{\Phi}(\hat{\theta}(y_1))^{-1} (x - y_1)/2\} \\
 & \cdot f(y_1) \\
 & \cdot \int_N \exp\{n(y - y_1)^T \hat{\Phi}(\hat{\theta}(y_1))^{-1} (x - y_1)\} \\
 & \cdot \exp\{-n(y - y_1)^T \hat{\Phi}(\hat{\theta}(y_1))^{-1} (y - y_1)/2\} \\
 & 1\{|y - y_1| \leq n^{1/7-1/2} \log n\} \sigma(dy) \\
 & \sim \exp\{-n(x - y_1)^T \hat{\Phi}(\hat{\theta}(y_1))^{-1} (x - y_1)/2\} \\
 & \cdot f(y_1) \\
 & \cdot \int_{TN(y_1)} \exp\{n y^T \hat{\Phi}(\hat{\theta}(y_1))^{-1} (x - y_1)\} \\
 & \cdot \exp\{-n y^T \hat{\Phi}(\hat{\theta}(y_1))^{-1} y/2\} m(dy)
 \end{aligned}$$

where $m(dy)$ denotes the Lebesgue measure on $TN(y_1)$. Moreover, the last relation is valid uniformly for $y_1 \in N$ and $|x - y| \leq \delta n^{1/7-1/2}$, since the curvature form of N at y_1 is uniformly bounded

(in operator norm) for $y_1 \in N$ (because N is compact). The last integral may be evaluated exactly by completing the square, leaving

$$(5.12) \int_N \exp\{n[\hat{\theta}(y)^T x - \psi(\hat{\theta}(y)) - \phi(x)]\}$$

$$1\{|y - y_1| \leq n^{1/7-1/2} \log n\} f(y) \sigma(dy)$$

$$\sim \exp\{-n(x - y_1)^T \ddagger(\hat{\theta}(y_1))^{-1} (x - y_1)/2\}$$

$$\cdot f(y_1) \cdot (2\pi/n)^{r/2} \cdot (\det H_2(y_1))^{-1/2}$$

$$\cdot \exp\{n(x - y_1)^T H_1(y_1)(x - y_1)/2\} \quad .$$

Note that since $|x - y_1| < \delta n^{-1/2+1/7}$ the last quantity is never smaller than $\exp\{-C' n^{2/7}\}$. Thus to complete the proof of (5.7) it is sufficient to show that the integral over $\{y \in N : |y - y_1| > n^{1/7-1/2} \log n\}$ is of smaller order of magnitude.

Now the function

$$u(y) = \hat{\theta}(y)^T x - \psi(\hat{\theta}(y)) - \phi(x)$$

is a nonpositive function of $y \in \Gamma$ whose only zero is at $y = x$. The Taylor series expansion (5.8) for $u(y)$ shows that there exist constants $C^* > 0$ and n_0 such that for $n \geq n_0$,

$$|y_1 - y| > n^{1/7-1/2} \log n$$

$$|x - y_1| \leq \delta n^{1/7-1/2}$$

and

$$y_1 \in N$$

imply

$$(5.13) \quad \hat{\theta}(y)^T x - \psi(\hat{\theta}(y)) - \phi(x) \leq -C^* n^{-1+2/7} \log^2 n .$$

Consequently,

$$(5.14) \quad \int_N \exp\{n[\langle \hat{\theta}(y) | x \rangle - \psi(\hat{\theta}(y)) - \phi(x)]\} \\ 1\{|y - y_1| > n^{1/7-1/2} \log n\} f(y) \sigma(dy) \\ \leq \exp\{-C^* n^{2/7} \cdot \log^2 n\} .$$

This completes the proof of (5.7). ///

Let

$$M_0 = \{x \in \mathbb{R}^p : \phi(x) < \epsilon_0 \text{ and } u^T x = 0 \forall u \in T_{\Omega_0}(0)\} .$$

Suppose that N is a smooth compact r -dimensional sub-manifold of

$M_0 \cap \Gamma_1$ with boundary, e.g.,

$$N = \{x \in M_0 \cap \Gamma_1 : \epsilon_1 \leq \phi_1(x) - \phi_0(x) \leq \epsilon_2 < \epsilon_0\} .$$

It is possible to mimic the preceding analysis to obtain asymptotic expressions for likelihood functions, but only when S_n/n is not too near the boundary ∂N : near ∂N , N is not well approximated by its tangent space, but instead by a half-space of its tangent space.

As before, let $f(\cdot)$ be a smooth, strictly positive probability density on N with respect to the volume element measure $\sigma(\cdot)$, and let

$$Q(A) = \int_N P_{\hat{\theta}(y)}(A) f(y) \sigma(dy) .$$

PROPOSITION 2. Suppose $y_1 \in N$ and

$$(5.15) \quad S_n/n = y_1 + hn^{-1/2}, \quad \text{and} \quad \text{dist}(S_n/n, \partial N) > n^{1/7-1/2} \log n$$

with $|h| < \delta n^{1/7}$. Then as $n \rightarrow \infty$,

$$(5.16) \quad dP_0^{(n)}/dQ^{(n)} \sim e^{-n\phi(S_n/n)} f(y_1)^{-1} (n/2\pi)^{r/2} \\ \cdot \exp\{h^T (\hat{\Phi}(\hat{\theta}(y_1))^{-1} - H_1(y_1))h/2\} \\ \cdot \det(H_2(y_1))^{1/2}.$$

This may be rewritten as

$$(5.17) \quad dP_0^{(n)}/dQ^{(n)} \sim e^{-\Lambda_n} f(y_1)^{-1} (n/2\pi)^{r/2} \\ \cdot \exp\{h^T (\hat{\Phi}(\hat{\theta}(y_1))^{-1} - H_1(y_1))h/2\} \\ \cdot \exp\{-h^T H_3(y_1)h/2\} \\ \cdot (\det H_2(y_1))^{1/2}$$

where

$$(5.18) \quad H_3(y) = \nabla_y^2 \phi(y) - \nabla_y^2 \phi_1(y) + \nabla_y^2 \phi_0(y).$$

Relations (5.16) and (5.17) are valid uniformly on the event

$$\{\text{dist}(S_n/n, N) < \delta n^{1/7-1/2}, \text{dist}(S_n/n, \partial N) > n^{1/7-1/2} \log n\}.$$

Recall that $\Lambda_n = n(\phi_1(S_n/n) - \phi_0(S_n/n))$ is the log-generalized likelihood ratio statistic for testing H_0 v. H_1 . It is worth noting

a consequence of (5.17): for each compact $K \subset N \setminus \partial N$ and each $\delta > 0$, there is a $C > 0$ such that

$$\text{dist}(S_n/n, K) < \delta n^{-1/2} \log n$$

implies

$$(5.19) \quad dP_0^{(n)}/dQ^{(n)} \leq e^{-\Lambda n} \exp\{C \log^2 n\}.$$

The proof of (5.16) is essentially the same as the proof of (5.7), and (5.17) follows simply from (5.16): note that

$$\begin{aligned} n\phi(S_n/n) - n\phi_1(S_n/n) \\ = (1/2n)(S_n - ny_1)^T [\nabla_y^2 \phi(y_1) - \nabla_y^2 \phi_1(y_1)](S_n - ny_1) \\ + O(n^{-2}|S_n - ny_1|^3) \end{aligned}$$

and

$$n\phi_0(S_n/n) = (1/2n)(S_n - ny_1)^T \nabla_y^2 \phi_0(y_1)(S_n - ny_1) + O(n^{-2}|S_n - ny_1|^3)$$

since $\phi(y_1) = \phi_1(y_1)$, $\nabla_y \phi(y_1) = \nabla_y \phi_1(y_1) = \hat{\theta}(y_1)$, $\phi_0(y_1) = 0$, and $\nabla_y \phi_0(y_1) = 0$ for $y_1 \in M_0 \cap \Gamma_1$. That the $O(\cdot)$ terms are uniformly small follows from the compactness of N . ///

6. The Collapsing Argument

The limiting constants in the conclusion of Theorem 2 occur as integrals over certain submanifolds of the mean parameter space: this corresponds to the fact that the bulk of $P_0\{\wedge_n \geq n\epsilon\}$ and $P_0\{T \leq a\epsilon_1^{-1}\}$ is accounted for by sample paths for which S_n/n and S_T/T are near the critical manifolds. The purpose of this section is to prove that the probabilities in question actually do shrink to integrals on the critical manifolds (hence the term "collapsing argument": it is not meant to suggest any structural deficiency in the proof itself).

Let

$$T_a = \{\inf_{n \geq a} \epsilon_2^{-1} : \wedge_n > a\}$$

$$M_0 = \{x \in \Omega_0(0)^\perp : \phi(x) < \epsilon_0\}$$

$$N_\epsilon = \{x \in M_0 \cap \Gamma_1 : \phi(x) = \phi_1(x) = \epsilon\} ;$$

$$\tilde{N}_\epsilon = \{x \in M_0 : \phi(x) = \epsilon\} .$$

PROPOSITION 1. Assume that $0 < \epsilon$, $\epsilon_2 < \epsilon_0$ and $0 < \epsilon_1 < \epsilon < \epsilon_2$. Then for every δ , $k > 0$

$$(6.1) \quad P_0\{\wedge_n \geq n\epsilon; \text{dist}(S_n/n, N_\epsilon) > \delta n^{-1/2} \log n\} = o(n^{-k} e^{-n\epsilon}) \quad \text{as } n \rightarrow \infty,$$

and

$$(6.2) \quad P_0\{T_a \leq a\epsilon_1^{-1}; \text{dist}(S_T/T, N_{a/T}) > \delta a^{-1/2} \log a\} = o(a^{-k} e^{-a})$$

as $a \rightarrow \infty$.

The proof will proceed by a series of crude estimates based on a likelihood ratio identity. The first step is to show that only those sample paths for which the sample means S_n/n fall in a certain compact subset of $\mathbb{R}^p$ are of any consequence.

LEMMA 1. There is a compact set $K \subset \Gamma$ such that $x \in K$ implies

$\phi(x) < \epsilon_0$; also

$$(6.3) \quad \{x \in \Gamma : \phi(x) \leq \max(\epsilon, \epsilon_2)\} \subset K^\circ$$

and

$$(6.4) \quad P_0\{\exists n \geq m : S_n/n \notin K\} = o(e^{-m\epsilon_3})$$

as $m \rightarrow \infty$, for some $\epsilon_3 > \max(\epsilon, \epsilon_2)$.

PROOF. Choose ϵ_3, ϵ_4 so that

$$\max(\epsilon, \epsilon_2) < \epsilon_3 < \epsilon_4 < \epsilon_0.$$

Then the sets

$$K_3 = \{x \in \Gamma : \phi(x) \leq \epsilon_3\}$$

$$K_4 = \{x \in \Gamma : \phi(x) \leq \epsilon_4\}$$

are compact (this is the reason for condition I on ϵ_0), and K_3 is contained in the interior of K_4 . Thus there is a compact K contained

in the interior of K_4 and containing K_3 in its interior, whose boundary ∂K is a polyhedron (i.e., K is the intersection of finitely many half-spaces). The existence of K may be deduced from the finite-dimensional Krein-Milman Theorem, the compactness of ∂K_4 , and the convexity of ϕ .

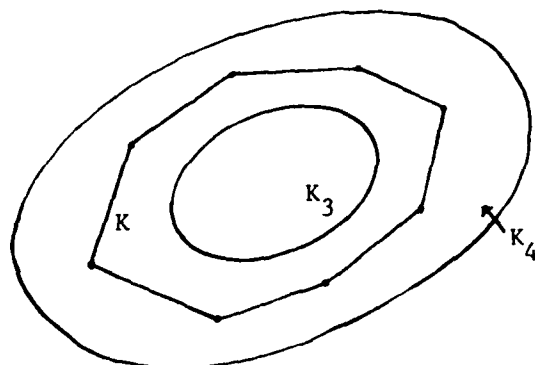


Figure 6.1

Now Chernoff's large deviation theorem for random variables (cf. [4]) gives exponential bounds for $P_0\{S_n/n \notin H\}$ where H is any half-space; since K is the intersection of finitely many half-spaces, and since

$$P_0\{\exists n \geq m : S_n/n \notin K\} \leq \sum_{n=m}^{\infty} P\{S_n/n \notin K\} \quad ,$$

(6.4) follows easily. ///

Let $f(\cdot)$ be a smooth probability density on Γ (with respect to Lebesgue measure on $\mathbb{R}^p$) which is strictly positive on K . Define

$$(6.5) \quad Q(A) = \int_{\Gamma} P_{\hat{\theta}(y)}(A) f(y) dy ;$$

then $Q^{(n)}$ and $P_0^{(n)}$ (the restrictions of P_0 and Q to the σ -algebra $\mathcal{B}(X_1, \dots, X_n)$) are mutually absolutely continuous, and

$$(6.6) \quad dP_0^{(n)}/dQ^{(n)} = \left[\int_{\Gamma} e^{\hat{\theta}(y)^T S_n - n\psi(\hat{\theta}(y))} f(y) dy \right]^{-1} .$$

LEMMA 2. There is a constant $C > 0$ such that

$$(6.7) \quad [dP_0^{(n)}/dQ^{(n)}] 1_{\{S_n/n \in K\}} \leq C \cdot n^{2p} e^{-n\phi(S_n/n)} .$$

PROOF. To obtain an upper bound for (6.6), one may replace the domain of integration Γ by a p -dimensional cube of side $1/n^2$ centered at S_n/n . Since K is compact and $f(\cdot)$ has a strictly positive minimum on K , (6.7) follows routinely. ///

To prove Proposition 1 it now suffices to show that there is a constant $\beta > 0$ such that if $x \in K$, $\epsilon_1 \leq b \leq \epsilon_2$, and

$$(6.8) \quad \phi_1(x) - \phi_0(x) \geq b$$

and

$$(6.9) \quad \text{dist}(x, N_b) > \delta n^{-1/2} \log n ,$$

then

$$(6.10) \quad \phi(x) \geq \beta n^{-1} \log^2 n + b .$$

For then Lemma 2 would imply

$$(6.11) \quad P_0\{\wedge_n \geq n\epsilon; \text{dist}(S_n/n, N_\epsilon) > \delta n^{-1/2} \log n; S_n/n \in K\}$$

$$\leq C n^{2p} \int_{\{.\}} e^{-n\epsilon} e^{-\beta \log^2 n} dQ$$

and

$$(6.12) \quad P_0\{T_a \leq a\epsilon_1^{-1}; \text{dist}(S_T/T, N_{a/T}) > \delta n^{-1/2} \log n; S_T/T \in K\}$$

$$\leq C(a\epsilon_1^{-1})^{2p} \int_{\{.\}} e^{-a} e^{-\beta \log^2(a\epsilon_2^{-1})} dQ .$$

LEMMA 3. For every $\delta > 0$ there exist $\gamma > 0$ and $\beta > 0$ such that if $x \in K$ and $\epsilon < \gamma$

$$(6.13) \quad \text{dist}(x, \Gamma_1) > \delta \epsilon$$

implies

$$(6.14) \quad \phi(x) - \phi_1(x) > \beta \epsilon^2 .$$

PROOF. First recall that $\phi - \phi_1$ is a non-negative function of $x \in \Gamma$ which is zero only for $x \in \Gamma_1$; consequently, it suffices to consider only those $x \in K$ for which

$$(6.15) \quad \text{dist}(x, \Gamma_1) < \eta ,$$

where η is a small positive number of our choosing. Now η may be chosen so small that for $x \in K$ satisfying (6.15) $x \in U_1$ and the MLE map $\hat{\theta}_1$ is submersive in a neighborhood of x (cf. Lemma 1, Section 3), and also small enough that for $x \in K$ satisfying (6.15) the two-term Taylor

series for $(\phi - \phi_1)$ around $\mu_{\hat{\theta}_1}(x)$ is accurate enough that

$$\begin{aligned}
 (6.16) \quad \phi(x) - \phi_1(x) &\geq \phi(\mu_{\hat{\theta}_1}(x)) - \phi_1(\mu_{\hat{\theta}_1}(x)) \\
 &\quad + (x - \mu_{\hat{\theta}_1}(x))^T \nabla(\phi - \phi_1)(\mu_{\hat{\theta}_1}(x)) \\
 &\quad + (1/4)(x - \mu_{\hat{\theta}_1}(x))^T \nabla^2(\phi - \phi_1)(\mu_{\hat{\theta}_1}(x))(x - \mu_{\hat{\theta}_1}(x)) .
 \end{aligned}$$

Since $\mu_{\hat{\theta}_1}(x) \in \Gamma_1$ it follows from Lemma 1 of Section 3 that

$$\phi(\mu_{\hat{\theta}_1}(x)) = \phi_1(\mu_{\hat{\theta}_1}(x))$$

and

$$\nabla\phi(\mu_{\hat{\theta}_1}(x)) = \nabla\phi_1(\mu_{\hat{\theta}_1}(x)) = \hat{\theta}(x) .$$

Thus to prove (6.14) it suffices to show that for $x \in K$ satisfying

(6.15) there exists a $\beta > 0$ such that

$$(x - \mu_{\hat{\theta}_1}(x))^T \nabla^2(\phi - \phi_1)(\mu_{\hat{\theta}_1}(x))(x - \mu_{\hat{\theta}_1}(x)) \geq \beta(\text{dist}(x, \Gamma_1))^2 .$$

Notice first that $x - \mu_{\hat{\theta}_1}(x) \in \mathcal{N}_1(\hat{\theta}_1(x))^\perp$, and recall that $\nabla^2 \phi_1(\mu_{\hat{\theta}_1}(x)) \upharpoonright \mathcal{N}_1(\hat{\theta}_1(x))^\perp = 0$ (this is (3.14) of Lemma 1, Section

3). Thus by the compactness of K and the fact that

$\nabla^2 \phi(y) = \ddagger(\hat{\theta}(y))^{-1}$ is everywhere P.D., there exists a $\beta > 0$ such

that whenever $x \in K$ satisfies (6.15),

$$(6.17) \quad (x - \mu_{\hat{\theta}_1}(x))^T \nabla^2(\phi - \phi_1)(\mu_{\hat{\theta}_1}(x))(x - \mu_{\hat{\theta}_1}(x)) \geq \beta |x - \mu_{\hat{\theta}_1}(x)|^2 .$$

Since $\mu_{\hat{\theta}_1}(x) \in \Gamma_1$, it is clear that

$$(6.18) \quad |x - \mu_{\hat{\theta}_1}(x)| > \text{dist}(x, \Gamma_1) .$$

This proves (6.14). ///

LEMMA 4. For each $\delta > 0$ there exist $\gamma > 0$, $\delta^* > 0$, and $\beta > 0$ such that if $x \in K$, $\epsilon \leq \gamma$, and

$$(6.19) \quad \text{dist}(x, M_0) > \delta \epsilon$$

but

$$(6.20) \quad \text{dist}(x, \Gamma_1) < \delta^* ,$$

then

$$(6.21) \quad \phi_0(x) \geq \beta \epsilon^2 .$$

PROOF. $\phi_0(x)$ is a convex non-negative function of $x \in K$ which is zero iff $x \in M_0$. Thus it suffices to consider only those $x \in K$ for which $\text{dist}(x, M_0) < \eta$: here $\eta > 0$ is a constant of our choosing.

Recall that M_0 and Γ_1 intersect transversally whenever they intersect (this by condition III in the definition of ϵ_0). By Lemma 2 of Section 3 there exist $\delta^* > 0$, $\eta > 0$ small enough that if $x \in K$ satisfies (6.20) and

$$(6.22) \quad \text{dist}(x, M_0) < \eta ,$$

then

$$(6.23) \quad \text{dist}(x, M_0 \cap \Gamma_1) < \eta^* .$$

Here $\eta^* > 0$ has been chosen small enough that if $x \in K$ satisfies (6.23), then $x \in U_0$ and the MLE map $\hat{\theta}_0$ is a smooth submersion in a neighborhood of x , and in addition the two-term Taylor series for ϕ_0 around $\underline{P}_{\mathbb{T}\Omega_0(0)}^\perp(x)$ is accurate enough that

$$(6.24) \quad \phi_0(\underline{Px}) + (x - \underline{Px})^T \nabla \phi_0(\underline{Px}) + (1/4)(x - \underline{Px})^T \nabla^2 \phi_0(\underline{Px})(x - \underline{Px}) \\ = (1/4)(x - \underline{Px})^T \nabla^2 \phi_0(\underline{Px})(x - \underline{Px}) \leq \phi_0(x)$$

$(\underline{P} = \underline{P}_{\mathbb{T}\Omega_0(0)}^\perp)$ denotes the orthogonal projection onto $\mathbb{T}\Omega_0(0)^\perp$.

By Lemma 1 of Section 3 $\nabla^2 \phi_0(y)$ is strictly P.D. on $\mathbb{T}\Omega_0(0)$ for any y such that $\hat{\theta}_0(y) = 0$; since $x - \underline{Px} \in \mathbb{T}\Omega_0(0)$ the lemma follows from (6.18) and an obvious compactness argument. ///

LEMMA 5. For every $\delta > 0$ there exist $\delta^* > 0$, $\beta > 0$, and $\gamma > 0$ such that if $\epsilon \leq \gamma$, and $x \in \Gamma$ satisfies

$$(6.25) \quad \text{dist}(x, M_0 \cap \Gamma_1) < \delta^* \epsilon,$$

$$(6.26) \quad \text{dist}(x, N_b) > \delta \epsilon$$

and

$$(6.27) \quad \phi(x) > b,$$

then

$$(6.28) \quad \phi(x) > b + \beta \epsilon.$$

Here $N_b = \{y \in M_0 \cap \Gamma_1 : \phi(y) = b\}$, and (6.28) holds for all b such that $\epsilon_1 \leq b \leq \epsilon_2$ (for the same β).

PROOF. It follows from Lemma 2 of Section 3 together with condition III in the definition of ϵ_0 that there exist δ^* , β^* , and γ such that if $x \in \Gamma$ satisfies (6.19)-(6.20), then

$$(6.29) \quad \text{dist}(x, \{y \in \Gamma : \phi(y) = b\}) > \beta^* \epsilon .$$

If x also satisfies (6.27), then

$$(6.30) \quad \text{dist}(x, \{y \in \Gamma : \phi(y) \leq b\}) > \beta^* \epsilon .$$

Relation (6.28) follows directly, since $\nabla \phi$ is nonzero on the level surface $\{y \in \Gamma : \phi(y) = b\}$. That it holds uniformly for $\epsilon_1 \leq b \leq \epsilon_2$ follows from an easy compactness argument. ///

It now follows from Lemmas 3-5 that if $x \in K$ satisfies (6.8)-(6.9), then it satisfies (6.10) (since $\phi \geq \phi_1 \geq \phi_0 \geq 0$). This proves Proposition 1. ///

7. Proof of Theorem 2

The proof of Theorem 2 will be based on the likelihood ratio identity

$$(7.1) \quad P_0(A) = \int_M [E_{\hat{\theta}(y)} 1_A L_T] \sigma_M(dy) / \sigma(M)$$

where

$$(7.2) \quad M = M(\epsilon_5, \epsilon_6; 0) = \{y \in \Gamma_1 \cap \Pi_0(0)^\perp : \epsilon_5 \leq \phi_1(y) - \phi_0(y) \leq \epsilon_6\},$$

$$(7.3) \quad 0 < \epsilon_5 < \epsilon_1 < \epsilon_2 < \epsilon_6 < \epsilon_0,$$

$$(7.4) \quad L_T = d(P_0 \upharpoonright \mathcal{F}_T) / d(Q \upharpoonright \mathcal{F}_T),$$

and

$$(7.5) \quad Q(B) = \int_M P_{\hat{\theta}(y)}(B) \sigma_M(dy) / \sigma(M).$$

Recall that M is a compact manifold-with-boundary, so its total surface area $\sigma(M)$ is finite. The relation (6.1) holds for all events $A \in \mathcal{F}_T$ ($\mathcal{F}_T$ is the "stopped" sigma algebra of events A for which $A \cap \{T \leq n\} \in \mathcal{F}(X_1, \dots, X_n)$ for every n) and (7.5) holds for all $B \in \mathcal{F}(X_1, X_2, \dots)$.

According to Proposition 1 of Section 6,

$$(7.6) \quad P_0\{T_a \leq a\epsilon_1^{-1}; \text{dist}(S_T, M(\epsilon_1, \epsilon_2; 0)) > \delta \cdot a^{-1/2} \log a\} = o(a^{-k} e^{-a})$$

as $a \rightarrow \infty$, for all $k, \delta > 0$. Consequently, it suffices to show that

$$(7.7) \quad P_0(A) \sim \text{RHS (3.22)}$$

for the event

$$(7.8) \quad A = \{T \leq a\epsilon_1^{-1} ; a^{-1/2} \log a \geq \text{dist}(S_T/T, M(\epsilon_1, \epsilon_2; 0))\} .$$

The plan of the proof, then, will be to evaluate (7.1) for A defined in (7.8) by exploiting the asymptotic formula for $1_A L_T$ provided by Proposition 2 of Section 5.

NOTE: Throughout the rest of this section, A will be the event defined by (7.8).

The manifold M divides neatly into three zones, in each of which the integrand behaves differently. These are

$$(7.9) \quad M_1 = \{y \in M : \epsilon_1 + a^{-1/2 + \eta} \leq \phi_1(y) - \phi_0(y) \leq \epsilon_2 - a^{-1/2 + \eta}\}$$

$$M_2 = \{y \in M : \phi_1(y) - \phi_0(y) \leq \epsilon_1 + a^{-1/2 + \eta} \quad \text{or}$$

$$\phi_1(y) - \phi_0(y) \geq \epsilon_2 - a^{-1/2 + \eta}\}$$

$$\cap \{y \in M : \text{dist}(y, M(\epsilon_1, \epsilon_2; 0)) < a^{-1/2 + \eta}\}$$

$$M_3 = \{y \in M : \text{dist}(y, M(\epsilon_1, \epsilon_2; 0)) \geq a^{-1/2 + \eta}\}$$

where $0 < \eta < 1/32$ is some fixed constant. It is clear that

$M_1 \uparrow M(\epsilon_1, \epsilon_2; 0)^\circ$ and that $M_2 \uparrow \partial M(\epsilon_1, \epsilon_2; 0)$ (thus $\sigma(M_2) \uparrow 0$ as $a \rightarrow \infty$).

It will develop that $\int_{M_1} \sim \text{RHS (3.22)}$ and that $\int_{M_2}$ and $\int_{M_3}$ are of smaller order of magnitude.

For $y \in M$, define

$$(7.10) \quad n_1 = n_1(a, y) = \ll (a/(\phi_1(y) - \phi_0(y))) - a^{1/2 + \eta} \rrbracket$$

$$n_2 = n_2(a, y) = \ll (a/(\phi_1(y) - \phi_0(y))) - a^{1/2 + \eta/2} \rrbracket$$

$$n_3 = n_3(a, y) = \ll (a/(\phi_1(y) - \phi_0(y))) + a^{1/2 + \eta/2} \rrbracket$$

and

$$(7.11) \quad \zeta_n(y) = (S_n - ny)^T (\hat{\Gamma}(\hat{\theta}(y))^{-1} - H_1(y) - H_3(y)) (S_n - ny) / 2n,$$

where $H_1(y)$ and $H_3(y)$ are as in (3.26) and (3.28).

LEMMA 1. As $a \rightarrow \infty$,

$$(7.12) \quad \max_{y \in M} P_{\hat{\theta}(y)} \{ |S_n - ny| > C a^{1/2 + \beta}, \text{ some } n \in [a\epsilon_2^{-1}, a\epsilon_1^{-1}] \} \\ = o(e^{-a^\beta}),$$

$$(7.13) \quad \max_{y \in M_1} P_{\hat{\theta}(y)} \{ T_a \notin [n_2(a, y), n_3(a, y)] \} = o(e^{-a^{\eta/4}}),$$

$$(7.14) \quad \max_{y \in M} P_{\hat{\theta}(y)} \{ |\zeta_n^{(y)} - \zeta_{n_1}^{(y)}| > C/\log a, \text{ some } n \in [n_2, n_3] \} \\ = o(e^{-a^{1/16}}),$$

and

$$(7.15) \quad \max_{y \in M} P_{\hat{\theta}(y)} \{ |(S_n/n) - (S_{n_1}/n_1)| > C a^{-1/2} \log a \} = o(e^{-a^{1/32}}),$$

for all $C > 0$, $0 < \beta < 1/6$.

This is a routine consequence of Corollary 1, Section 4; the proof is left to the reader.

Next define events

$$\begin{aligned}
 (7.16) \quad A_y &= \{ |(S_{n_1}/n_1) - y| < a^{-1/2 + \eta} \} \\
 A_y^* &= \{ |(S_{n_1}/n_1) - (S_T/T)| < a^{-1/2} \log a \} \\
 &\quad \cap \{ |\zeta_{n_1}(y) - \zeta_T(y)| < 1/\log a \} \\
 A_y^{**} &= A_y^* \cap \{ T_a \in [n_2(a, y), n_3(a, y)] \} .
 \end{aligned}$$

By Proposition 2, Section 5

$$\begin{aligned}
 (7.17) \quad L_T 1_A 1_{A_y} 1_{A_y^*} &\sim e^{-\Lambda_T} \exp\{\zeta_{n_1}(y)\} (T/2\pi)^{(q_1 - q_0)/2} \\
 &\quad \cdot (\det H_2(y))^{1/2} \sigma(M) 1_A 1_{A_y} 1_{A_y^*}
 \end{aligned}$$

and

$$\begin{aligned}
 (7.18) \quad L_T 1_A 1_{A_y} 1_{A_y^{**}} &\sim e^{-\Lambda_T} \exp\{\zeta_{n_1}(y)\} \sigma(M) 1_A 1_{A_y} 1_{A_y^{**}} \\
 &\quad \cdot (a/2\pi(\phi_1(y) - \phi_0(y)))^{(q_1 - q_0)/2} \\
 &\quad \cdot (\det H_2(y))^{1/2} ;
 \end{aligned}$$

furthermore, these relations are valid uniformly for $y \in M$ and uniformly on the events $A_y \cap A_y^*$, and $A_y \cap A_y^{**}$, respectively.

LEMMA 2. As $a \rightarrow \infty$,

$$(7.19) \quad P_{\hat{\theta}(y)} \{ |E_{\hat{\theta}(y)} [1_A 1_{A_y^{**}} e^{-(\wedge_T - a)} | \mathfrak{F}_{n_1(a,y)}] - v(y) | > \epsilon \} \rightarrow 0$$

for every $\epsilon > 0$ and every $y \in M(\epsilon_1, \epsilon_2; 0)$ such that the random variable

$$(\phi_1(y) - \phi_0(y)) + (X_1 - y)^T \nabla(\phi_1 - \phi_0)(y)$$

has a nonlattice distribution when $X_1 \sim P_{\hat{\theta}(y)}$ (cf. Lemma 4, Section 3).

The σ -algebra $\mathfrak{F}_{n_1}$ is the one generated by $X_1, \dots, X_{n_1}$. It should be noticed that the convergence indicated by (7.19) need not be uniform in y . Fortunately, the rv's are bounded.

Lemma 2 is very much related to the nonlinear renewal theorem of Lai and Siegmund [9]. Although the statement of their theorem does not imply Lemma 2, their proof does: in fact, they obtain an unconditional limit theorem by first proving a conditional statement, which in our case becomes

$$(7.20) \quad E_{\hat{\theta}(y)} [e^{-(\wedge_T - a)} | \mathfrak{F}_{n_1}] - v(y) \rightarrow 0 \quad \text{A.S. } (P_{\hat{\theta}(y)})$$

Since $1_A 1_{A_y^{**}} \rightarrow 1$ (cf. Lemma 1) A.S. $(P_{\hat{\theta}(y)})$, (19) follows from this.

The key to evaluating $E_{\hat{\theta}(y)} 1_A 1_{A_y} 1_{A_y^{**}} L_T$ is now provided by Corollary 2 of Section 4. This allows that uniformly for $y \in M$,

$$(7.21) \quad E_{\hat{\theta}(y)} 1_{A_y} \exp\{\zeta_{n_1(a,y)}\} \rightarrow (\det \mathfrak{I}(\hat{\theta}(y)) \cdot (H_1(y) + H_3(y)))^{-1/2}$$

as $a \rightarrow \infty$. Consequently by (7.18), (7.19) and the Dominated Convergence Theorem

$$(7.22) \quad \int_{M_1} a^{-(q_1-q_0)/2} e^a E_{\hat{\theta}(y)} [1_{A \cap A_y^{**}} L_T] \sigma_M(dy) / \sigma_M(M) \\ \rightarrow C(\epsilon_1, \epsilon_2; 0) .$$

Recall from Proposition 2 of Section 5 (5.19) that

$$(7.23) \quad 1_A L_T \leq e^{-\Lambda_T} \exp\{C \log^2 n\}$$

for some constant C . By Lemma 1

$$\max_{y \in M_1} P_{\hat{\theta}(y)}(A \setminus (A_y \cap A_y^{**})) = o(e^{-a^{\eta/4}}) ;$$

thus

$$(7.24) \quad \int_{M_1} a^{-(q_1-q_0)/2} e^a E_{\hat{\theta}(y)} [1_{A \setminus (A_y \cap A_y^{**})} L_T] \sigma_M(dy) / \sigma_M(M) \\ \rightarrow 0 \quad \text{as } a \rightarrow \infty .$$

For $y \in M_2$, (7.17) and a compactness argument show that there is a constant $C^* > 0$ such that

$$a^{-(q_1-q_0)/2} e^a L_T 1_{A \cap A_y^*} \leq C^* \exp\{\zeta_{n_1}(y)\} 1_{A_y}$$

and (7.21) implies that there is a constant $C^{**} > 0$ such that

$$E_{\hat{\theta}(y)} a^{-(q_1-q_0)/2} e^a L_T 1_{A \cap A_y^*} \leq C^{**} .$$

This, (7.23), and the fact that

$$\max_{y \in M_2} P_{\hat{\theta}(y)}(A \setminus (A_y \cap A_y^*)) = o(e^{-a^{\eta/4}})$$

imply

$$(7.25) \quad \int_{M_2} a^{-(q_1 - q_0)/2} e^{-a} E_{\hat{\theta}(y)} L_T 1_A \sigma_M(dy) / \sigma_M(M) \rightarrow 0,$$

since $\sigma_M(M_2) \rightarrow 0$.

Finally, for $y \in M_3$,

$$(7.26) \quad A \cap \{|S_n/n - y| < a^{-1/2 + \eta/2} \text{ for all } n \in [a\epsilon_2^{-1}, a\epsilon_1^{-1}]\} = \emptyset.$$

By Lemma 1

$$\max_{y \in M_3} P_{\hat{\theta}(y)}\{|S_{n_1}/n_1 - y| > a^{-1/2 + \eta/2} \text{ for some } n \in [a\epsilon_2^{-1}, a\epsilon_1^{-1}]\}$$

$$= o(e^{-a^{\eta/4}});$$

this (7.23), and (7.26) imply

$$(7.27) \quad \int_{M_3} a^{-(q_1 - q_0)/2} e^a E_{\hat{\theta}(y)} [1_A L_T] \sigma_M(dy) / \sigma_M(M) \rightarrow 0.$$

The relations (7.1), (7.22), (7.24), (7.25), and (7.27) prove

Theorem 2. ///

REFERENCES

- [1] Armitage, P. (1967). Some developments in the theory and practice of sequential medical trials. Proc. 5th Berkeley Symposium 4, 791-804.
- [2] Armitage, P. (1975). Sequential Medical Trials, 2nd edition. Oxford: Blackwell.
- [3] Bhattacharya, R., and Ranga Rao, R. (1976). Normal Approximation and Asymptotic Expansions. John Wiley and Sons, New York.
- [4] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Stat. 23, 493-507.
- [5] Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with discussion). Ann. Statist. 3, 1189-1242.
- [6] Feller, W. (1968). An Introduction to Probability Theory and its Applications, Volume 1, 3rd edition. John Wiley and Sons, New York.
- [7] Feller, W. (1971). An Introduction to Probability Theory and its Applications, Volume 2, 2nd edition. John Wiley and Sons, New York.
- [8] Guillemin, V. and Pollack, A. (1974). Differential Topology. Prentice-Hall, Englewood Cliffs, New Jersey.
- [9] Lai, T.L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. Ann. Stat. 5, 946-954.
- [10] Lai, T.L. and Siegmund, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. Ann. Stat. 7, 60-76.
- [11] Schwartz, G. (1962). Asymptotic shapes of Bayes' sequential testing regions. Ann. Math. Stat. 33, 224-236.
- [12] Siegmund, D. (1977). Repeated significance tests for a normal mean. Biometrika 64, 177-190.
- [13] Siegmund, D. and Gregory, P. (1979). A sequential clinical trial for testing $p_1 = p_2$. Stanford University Technical Report.

- [14] Woodroofe, M. (1976). Frequentist properties of Bayesian sequential tests. Biometrika 63, 101-110.
- [15] Woodroofe, M. (1978). Large deviations of likelihood ratios with applications to sequential testing. Ann. Stat. 6, 72-84.
- [16] Woodroofe, M. (1979). Repeated likelihood ratio tests. Biometrika 66, 453-463.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER No. 10	2. GOVT ACCESSION NO. AD-A092 634	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) REPEATED LIKELIHOOD RATIO TESTS FOR CURVED EXPONENTIAL FAMILIES.		5. TYPE OF REPORT & PERIOD COVERED 9 TECHNICAL REPORT.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(S) STEVE ALLEY 10) ST 1/Lalley		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-77-C-0306
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPARTMENT OF STATISTICS STANFORD UNIVERSITY STANFORD, CALIFORNIA 11) 2		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-373
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH STAT. & PROBABILITY PROGRAM, CODE 436 ARLINGTON, VIRGINIA 22217		12. REPORT DATE SEPTEMBER 1, 1980
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) T-4 12) 67		13. NUMBER OF PAGES 62
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE AND SALE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Curved Exponential Family, Repeated Significance Test, Nonlinear Renewal Theory, Transversality.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A class of repeated significance tests for curved hypotheses in multi-parameter exponential families is studied, and asymptotic formulae for the significance levels of such tests are obtained. Special attention is given the important case of comparing Bernoulli success probabilities.		

DD FORM 1473

JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

33

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)