The Annals of Probability
1988, Vol. 16, No. 1, 375–396

BOUNDARY CROSSING PROBLEMS FOR SAMPLE MEANS[1]

BY TZE LEUNG LAI

Stanford University

Motivated by several classical sequential decision problems, we study herein the following type of boundary crossing problems for certain nonlinear functions of sample means. Let $X_1, X_2, \ldots$ be i.i.d. random vectors whose common density belongs to the k-dimensional exponential family $h_\theta(x) = \exp\{\theta' x - \psi(\theta)\}$ with respect to some nondegenerate measure ν. Let $\overline{X}_n = (X_1 + \cdots + X_n)/n$, $\hat{\theta}_n = (\nabla\psi)^{-1}(\overline{X}_n)$, and let $I(\theta, \lambda) = E_\theta \log\{h_\theta(X_1)/h_\lambda(X_1)\}$ (= Kullback–Leibler information number). Consider stopping times of the form

$$T_c(\lambda) = \inf\left\{n: I\left(\hat{\theta}_n, \lambda\right) \geq n^{-1} g(cn)\right\}, \qquad c > 0,$$

where g is a positive function such that $g(t) \sim \alpha \log t^{-1}$ as $t \to 0$. We obtain asymptotic approximations to the moments $E_\theta T_c^r(\lambda)$ as $c \to 0$ that are uniform in θ and λ with $|\lambda - \theta|^2/c \to \infty$. We also study the probability that $\overline{X}_{T_c(\lambda)}$ lies in certain cones with vertex $\nabla\psi(\lambda)$. In particular, in the one-dimensional case with $\lambda > \theta$, we consider boundary crossing probabilities of the form

$$P_\theta\left\{\hat{\theta}_n \geq \lambda \text{ and } I\left(\hat{\theta}_n, \lambda\right) \geq n^{-1} g(cn) \text{ for some } n\right\}.$$

Asymptotic approximations (as $c \to 0$) to these boundary crossing probabilities are obtained that are uniform in θ and λ with $|\lambda - \theta|^2/c \to \infty$.

1. Introduction. In this paper we study a class of boundary crossing problems for sample means that are related to several classical problems in sequential statistical methodology. Important advances in these statistical problems were made during the past two decades for normally distributed random variables by using continuous-time approximations and by analyzing the corresponding free boundary problems involving the heat equation, but much remains to be done for other distributions and towards a more basic understanding of the problems. The analysis of these statistical decision problems has led us to a fundamental class of boundary crossing problems for sample means from a multidimensional exponential family. Our main results are stated in Section 2, and the derivation of these results is given in Sections 3 and 4, where some further results and methods for boundary crossing problems are also presented.

We now review briefly some of these sequential decision problems in the context of normal random variables to illustrate that they are intrinsically related to a unifying class of boundary crossing problems for the Wiener process,

Received November 1985; revised December 1986.

[1]This research was supported by the National Science Foundation and the Army Research Office.

AMS 1980 *subject classifications*. Primary 60F10, 60G40; secondary 62L05, 62L15.

Key words and phrases. Sequential analysis, stopping time, boundary crossing probabilities, exponential family, Kullback–Leibler information, uniform strong laws.

which we describe in

LEMMA 1. *Let $\alpha > 0$ and let $w(t)$, $t \geq 0$, be a Wiener process with drift coefficient μ. Let f be a nonnegative function on $(0, \infty)$ such that*

$$f(t) \sim \left(2\alpha t \log t^{-1}\right)^{1/2}, \quad \text{as } t \to 0, \tag{1.1}$$

and

$$\sup_{t \geq a} f(t)/t < \infty, \quad \text{for all } a > 0. \tag{1.2}$$

(i) *Let $\tau(f) = \inf\{t\colon |w(t)| \geq f(t)\}$ and let $\lambda(f) = \sup\{t\colon |w(t)| \leq f(t)\}$. Then as $|\mu| \to \infty$,*

$$E_\mu \tau(f) \sim E_\mu \lambda(f) \sim \left(2\alpha \log \mu^2\right)/\mu^2. \tag{1.3}$$

(ii) *Suppose furthermore that there exist real numbers β and γ such that*

$$f(t) = \left\{2t\left(\alpha \log t^{-1} - \beta \log\log t^{-1} - \gamma + o(1)\right)\right\}^{1/2}, \quad \text{as } t \to 0. \tag{1.4}$$

Then as $\mu \to \infty$,

$$\begin{aligned} &P_\mu\{w(t) \leq -f(t) \text{ for some } t\} \\ &\quad = P_{-\mu}\{w(t) \geq f(t) \text{ for some } t\} \\ &\quad \sim \pi^{-1/2} e^{\gamma} \alpha^{1/2-\alpha} \Gamma(2\alpha)\left(2\mu^2\right)^{-\alpha}\left(\log \mu^2\right)^{\beta-\alpha+1/2}. \end{aligned} \tag{1.5}$$

The special case $\alpha = 1$ of Lemma 1 was obtained by Lai, Robbins and Siegmund (1983) in their analysis of the optimal stopping rule that minimizes $\int_{-\infty}^{\infty} R(\mu; \tau) e^{-c\mu^2}\, d\mu$, $c > 0$, among all stopping rules $\tau \leq \frac{1}{2}$ for the Wiener process $w(t)$, where

$$\begin{aligned} R(\mu; \tau) &= \mu E_\mu\left\{\tau + (1 - 2\tau) I_{\{w(\tau) < 0\}}\right\}, \quad \text{if } \mu \geq 0, \\ &= |\mu| E_\mu\left\{\tau + (1 - 2\tau) I_{\{w(\tau) > 0\}}\right\}, \quad \text{if } \mu < 0. \end{aligned} \tag{1.6}$$

This optimal stopping problem arose as the normalized limit of Anscombe's (1963) model for the optimal determination of the length of a clinical trial involving paired normal data to select the better of two treatments for N patients. Earlier, by an asymptotic analysis of the associated free boundary problem involving the heat equation, Chernoff and Petkau (1981) showed that the optimal stopping rule is of the form $\tau(f)$ in Lemma 1, where $f(t) = 0$ for $t \geq \frac{1}{2}$ and

$$f(t) = \left\{2t\left(\log t^{-1} - \tfrac{1}{2}\log\log t^{-1} - \tfrac{1}{2}\log 16\pi + o(1)\right)\right\}^{1/2}, \quad \text{as } t \to 0. \tag{1.7}$$

For the special case $\alpha = 1$ in Lemma 1, the probability in (1.5) is of a smaller order of magnitude than the expected sample size in (1.3) only when $\beta < \frac{3}{2}$; in this case, by (1.6),

$$R(\mu; \tau(f)) \sim \left(2 \log \mu^2\right)/|\mu|, \quad \text{as } |\mu| \to \infty.$$

If $\beta > \frac{3}{2}$, then the probability in (1.5) is of a larger order of magnitude than

$E_\mu \tau(f)$ and

$$R(\mu; \tau(f)) \sim e^{\gamma}(\log \mu^2)^{\beta - 1/2}/(2\pi^{1/2}|\mu|), \quad \text{as } |\mu| \to \infty.$$

Thus, stopping boundaries with the asymptotic behavior (1.7), or more generally (1.4) with $\beta < \frac{3}{2}$, give an asymptotically optimal balance between the expected sampling time (1.3) and the probability (1.5) of wrong decision, as noted by Lai, Robbins and Siegmund (1983).

An important development in the subject of Bayes sequential tests of composite hypotheses is Chernoff's (1961, 1965) pioneering work on testing the sign of a normal mean. A renormalization of the original problem led to the continuous-time optimal stopping problem of minimizing $\int_{-\infty}^{\infty} \tilde{R}(\mu; \tau)\, d\mu$ among all stopping rules for the Wiener process $w(t)$, where

$$(1.8) \qquad \begin{aligned} \tilde{R}(\mu; \tau) &= E_\mu\{\tau + \mu I_{\{w(\tau) < 0\}}\}, \quad \text{if } \mu \geq 0, \\ &= E_\mu\{\tau + |\mu| I_{\{w(\tau) > 0\}}\}, \quad \text{if } \mu < 0. \end{aligned}$$

The optimal stopping rule is of the form $\tau(f)$ in Lemma 1, where

$$(1.9) \qquad f(t) = \{3t[\log t^{-1} - (\log 8\pi)/3 + o(1)]\}^{1/2}, \quad \text{as } t \to 0,$$

[cf. Chernoff (1965)]. Note that in Lemma 1, $|\mu|$ times the probability in (1.5) is of smaller (larger) order of magnitude than $E_\mu \tau(f)$ if $1 - 2\alpha < (>) - 2$, or equivalently, if $\alpha > (<)\frac{3}{2}$. When $\alpha = \frac{3}{2}$, $|\mu|$ times the probability in (1.5) is of smaller (larger) order of magnitude than $E_\mu \tau(f)$ if $\beta < (>) 2$. The Bayes stopping boundary (1.9) corresponds to $\alpha = \frac{3}{2}$ and $\beta = 0$.

We can prove Lemma 1 by using the same argument as the proof of the special case $\alpha = 1$ given by Lai, Robbins and Siegmund (1983), pages 57–58. The argument is based on an asymptotic formula of Jennen and Lerche (1981) on the first exit density of the Wiener process through a moving boundary. While these results for the Wiener process can be applied to sums of i.i.d. normal $N(\theta, 1)$ random variables $X_1, X_2, \ldots$ in the sequential decision problems via the space–time transformation

$$(1.10) \qquad t = cn, \qquad w(t) = c^{1/2} \sum_1^n X_i, \qquad \mu = c^{-1/2}\theta,$$

their extensions to nonnormal random variables require a different and more general approach, which will be presented in Sections 3 and 4. This approach can be applied not only to the Wiener process but also to sample sums from an exponential family and can be easily generalized to higher dimensions, as will be described in the next section.

2. Main results. In this section we provide a useful analog of Lemma 1 for sample sums from the Koopman–Darmois (exponential) family of distributions and also generalize the result to the multidimensional case. Let $X, X_1, X_2, \ldots$ be i.i.d. random vectors whose common density (with respect to some nondegenerate measure ν) belongs to the k-dimensional exponential family

$$h_\theta(x) = \exp\{\theta' x - \psi(\theta)\}, \tag{2.1}$$

where $\theta = (\theta_1, \dots, \theta_k)'$, $x = (x_1, \dots, x_k)'$ and $'$ denotes the transpose of a vector. Then $E_\theta X = \nabla\psi(\theta)$, $\mathrm{Cov}_\theta X = \nabla^2\psi(\theta)$, and the Kullback–Leibler information number is given by

$$\begin{aligned} I(\theta, \lambda) &= E_\theta \log\{h_\theta(X)/h_\lambda(X)\} \\ &= (\theta - \lambda)'\nabla\psi(\theta) - (\psi(\theta) - \psi(\lambda)) \\ &= \int_0^1 (1 - t)(\lambda - \theta)'\{\nabla^2\psi(t\lambda + (1 - t)\theta)\}(\lambda - \theta)\, dt. \end{aligned} \tag{2.2}$$

Here and in the sequel, we use the notation $\nabla\psi$ to denote the gradient vector $(\partial\psi/\partial\theta_1, \dots, \partial\psi/\partial\theta_k)'$ and $\nabla^2\psi$ to denote the Hessian matrix $(\partial^2\psi/\partial\theta_i\,\partial\theta_j)_{1 \le i,\, j \le k}$. Moreover, let $S_n = X_1 + \cdots + X_n$, $\overline{X}_n = S_n/n$, and let

$$\Theta = \mathrm{Int}\left\{\theta\colon \int \exp(\theta' x)\, d\nu(x) < \infty\right\}$$

be the interior (Int) of the natural parameter space, which is a convex subset of R^k. Noting that $\nabla\psi$ is a diffeomorphism on Θ, the maximum likelihood estimate $\hat{\theta}_n$ of θ at stage n is given by

$$\hat{\theta}_n = (\nabla\psi)^{-1}(\overline{X}_n), \quad \text{if } \overline{X}_n \in \nabla\psi(\Theta). \tag{2.3}$$

To begin with, consider the case of one-dimensional normal random variables with mean θ and variance 1. Introducing the transformation (1.10), we obtain from Lemma 1 its discrete-time analog in

COROLLARY 1. *Let $X_1, X_2, \dots$ be i.i.d. normal random variables with mean θ and variance* 1. *Let f be a nonnegative function on* $(0, \infty)$ *satisfying* (1.1) *and* (1.2). *For $c > 0$, define $T_c = \inf\{n\colon c^{1/2}|S_n| \ge f(cn)\}$ and $L_c = \sup\{n\colon c^{1/2}|S_n| \le f(cn)\}$. Let $\delta_c < \Delta_c$ be positive numbers such that*

$$\begin{aligned} &\delta_c \to 0 \quad \text{and} \quad \delta_c/c \to \infty, \\ &\Delta_c \to \infty \quad \text{and} \quad \Delta_c = o(|\log c|), \quad \text{as } c \to 0. \end{aligned} \tag{2.4}$$

(i) *As $c \to 0$, $E_\theta T_c \sim E_\theta L_c \sim 2\alpha\theta^{-2}\log(\theta^2/c)$ uniformly in θ with $\delta_c \le \theta^2 \le \Delta_c$.*

(ii) *Let $g(t) = f^2(t)/2t$. Suppose furthermore that for some β,*

$$g(t) \ge \alpha \log t^{-1} - \beta \log\log t^{-1}, \quad \text{as } t \to 0. \tag{2.5}$$

Then as $c \to 0$,

$$P_\theta\{c^{1/2}S_n \le -f(cn) \text{ for some } n\} = O\left((c/\theta^2)^\alpha\{\log(\theta^2/c)\}^{\beta - \alpha + 1/2}\right), \tag{2.6}$$

uniformly in $\theta \ge \delta_c^{1/2}$.

In Corollary 1, the restriction $\theta^2 \ge \delta_c$ ensures that $\theta^2/c \to \infty$ as $c \to 0$ by (2.4), while the restriction $\theta^2 \le \Delta_c$ ensures that $(\log c^{-1}\theta^2)/\theta^2 \to \infty$ by (2.4).

Defining $\tau(f)$ and $\lambda(f)$ as in Lemma 1, note that by (1.10),

$$c^{-1}\tau(f) \le T_c \le L_c + 1 \le c^{-1}\lambda(f) + 1,$$

and therefore Corollary 1(i) follows from Lemma 1(i). In view of (1.10), $P_\theta\{c^{1/2}S_n \le -f(cn) \text{ for some } n\} \le P_\mu\{w(t) \le -f(t) \text{ for some } t\}$, so Corollary 1(ii) follows from Lemma 1(ii).

We now extend these results from the normal case to the exponential family of distributions (2.1). Note that for the normal distribution of Corollary 1, $\psi(\theta) = \theta^2/2$ and the Kullback–Leibler information number is given by $I(\theta, \lambda) = (\theta - \lambda)^2/2$. The maximum likelihood estimate of θ at stage n is $\overline{X}_n$. Moreover,

$$c^{1/2}S_n \le -f(cn) \Leftrightarrow \overline{X}_n \le 0 \quad \text{and} \quad I(\overline{X}_n, 0) \ge n^{-1}g(cn),$$

where $g(t) = f^2(t)/2t$. This equivalence provides a basic idea in the generalization of Corollary 1 to the exponential family. Since $\nabla^2\psi(\theta) = \mathrm{Cov}_\theta X$ need no longer be constant as in the normal case, we will restrict θ to a convex subset A of Θ such that

$$\inf_{\theta \in A_\rho} \lambda_{\min}(\nabla^2\psi(\theta)) > 0, \qquad \sup_{\theta \in A_\rho} \lambda_{\max}(\nabla^2\psi(\theta)) < \infty, \tag{2.7}$$
$$\text{and } \nabla^2\psi \text{ is uniformly continuous on } A_\rho \text{ for some } \rho > 0,$$

where $A_\rho = \{x \in R^k: \inf_{\theta \in A}|x - \theta| < \rho\}$ denotes the ρ-neighborhood of A, and $\lambda_{\min}, \lambda_{\max}$ denote the minimum and maximum eigenvalues of a symmetric matrix. Note that in the case of a (nonsingular) normal distribution, we can take $A = \Theta(= R^k = A_\rho)$. In analogy with T_c and L_c of Corollary 1, we define for $c > 0$ and $\lambda \in \Theta$,

$$T_c(\lambda) = \inf\{n: \hat{\theta}_n \in A_\rho \text{ and } I(\hat{\theta}_n, \lambda) \ge n^{-1}g(cn)\}, \tag{2.8}$$

$$L_c(\lambda) = \sup\Big(\{n: \hat{\theta}_n \in A_\rho \text{ and } I(\hat{\theta}_n, \lambda) \le n^{-1}g(cn)\} \cup \{n: \overline{X}_n \notin \nabla\psi(A_\rho)\}\Big), \tag{2.9}$$

where $\hat{\theta}_n = (\nabla\psi)^{-1}(\overline{X}_n)$, as given by (2.3).

First consider the one-dimensional exponential family. Here $\nabla\psi = d\psi/d\theta$, $d^2\psi/d\theta^2 = \lambda_{\min}(\nabla^2\psi) = \lambda_{\max}(\nabla^2\psi)$, and the convex set A is an interval. The quantities $\hat{\theta}_n$ and $I(\theta, \lambda)$ are basic ingredients in the following generalization of Lemma 1 and Corollary 1 to the one-dimensional exponential family. Letting $g(t) = f^2(t)/2t$, note that f satisfies (1.1) and (1.2) iff

$$g(t) \sim \alpha \log t^{-1}, \quad \text{as } t \to 0, \tag{2.10}$$

$$\sup_{t \ge a} g(t)/t < \infty, \quad \text{for all } a > 0. \tag{2.11}$$

THEOREM 1. *Let $X_1, X_2, \ldots$ be i.i.d. random variables having common density $h_\theta(x) = \exp(\theta x - \psi(\theta))$ with respect to some nondegenerate measure ν. Let A be a subinterval of the natural parameter space Θ such that* (2.7) *holds. Let $\alpha > 0$ and let g be a nonnegative function on $(0, \infty)$ satisfying* (2.10) *and*

(2.11). *For* $c > 0$ *and* $\lambda \in \Theta$, *define* $T_c(\lambda)$, $L_c(\lambda)$ *by* (2.8) *and* (2.9), *where* $\hat{\theta}_n$ *is given by* (2.3) *and* $I(\theta, \lambda)$ *is the Kullback–Leibler information number defined in* (2.2). *Let* $\delta_c < \Delta_c$ *be positive numbers satisfying* (2.4).

(i) *For every fixed* $r > 0$, *as* $c \to 0$,

$$E_\theta T_c^r(\lambda) \sim E_\theta L_c^r(\lambda) \sim \{\alpha \log(c^{-1}|\lambda - \theta|^2)/I(\theta, \lambda)\}^r, \tag{2.12}$$

uniformly in $\theta \in A$ *and* $\lambda \in A_\rho$ *with* $\delta_c \le |\lambda - \theta|^2 \le \Delta_c$. *In fact,* $L_c(\lambda) \ge T_c(\lambda) - 1$, *and for every fixed* $0 < \eta < 1$,

$$P_\theta\{T_c(\lambda) \le (1-\eta)\alpha \log(c^{-1}|\lambda - \theta|^2)/I(\theta, \lambda)\} \to 0, \tag{2.13}$$

$$P_\theta\{L_c(\lambda) \ge (1+\eta)\alpha \log(c^{-1}|\lambda - \theta|^2)/I(\theta, \lambda)\} \to 0, \tag{2.14}$$

as $c \to 0$, *uniformly in* $\theta \in A$ *and* $\lambda \in A_\rho$ *such that* $\delta_c \le |\lambda - \theta|^2 \le \Delta_c$.

(ii) *Suppose that condition* (2.10) *is strengthened into the asymptotic expansion*

$$g(t) = \alpha \log t^{-1} - \beta \log\log t^{-1} - \gamma + o(1), \quad \text{as } t \to 0, \tag{2.15}$$

where β *and* γ *are real constants. Then as* $\delta \downarrow 0$ *and* $c \downarrow 0$ *such that* $\delta^2/c \to \infty$ *but* $\delta \log c \to 0$,

$$\begin{aligned} &P_\theta\{\hat{\theta}_n \in A_\rho, \hat{\theta}_n \ge \theta + \delta \text{ and } I(\hat{\theta}_n, \theta + \delta) \ge n^{-1}g(cn) \text{ for some } n\} \\ &\sim P_\theta\{\hat{\theta}_n \in A_\rho, \hat{\theta}_n \le \theta - \delta \text{ and } I(\hat{\theta}_n, \theta - \delta) \ge n^{-1}g(cn) \text{ for some } n\} \\ &\sim \pi^{-1/2} e^{\gamma} \alpha^{1/2-\alpha}\Gamma(2\alpha)(2\psi''(\theta)\delta^2/c)^{-\alpha}\{\log(\delta^2/c)\}^{\beta-\alpha+1/2}, \end{aligned} \tag{2.16}$$

uniformly in $\theta \in A$. [*Note that* $\psi''(\theta)\delta^2 \sim 2I(\theta, \theta \pm \delta)$ *uniformly in* $\theta \in A$.]

(iii) *Suppose that condition* (2.10) *is replaced by*

$$g(t) \ge \alpha \log t^{-1} - \beta \log\log t^{-1}, \quad \text{as } t \to 0, \tag{2.17}$$

for some real constant β. *Define the interval*

$$\begin{aligned} J_\lambda(\theta) &= [\lambda, \infty), \quad \text{if } \lambda > \theta, \\ &= (-\infty, \lambda], \quad \text{if } \lambda < \theta. \end{aligned} \tag{2.18}$$

Then as $c \to 0$,

$$\begin{aligned} &P_\theta\{\hat{\theta}_n \in J_\lambda(\theta) \cap A_\rho \text{ and } I(\hat{\theta}_n, \lambda) \ge n^{-1}g(cn) \text{ for some } n\} \\ &\quad = O\left(\{c^{-1}|\lambda - \theta|^2\}^{-\alpha}\{\log(c^{-1}|\lambda - \theta|^2)\}^{\beta-\alpha+1/2}\right), \end{aligned} \tag{2.19}$$

uniformly in $\theta \in A$ *and* $\lambda \in A$ *with* $|\lambda - \theta|^2 \ge \delta_c$.

We now generalize Theorem 1 to the k-dimensional exponential family. First consider the boundary crossing probabilities (2.16) and (2.19). A multivariate extension of the event $\{\hat{\theta}_n \in J_\lambda(\theta)\}$ has the form

$$\{\overline{X}_n - \nabla\psi(\lambda) \in \mathscr{C}_p(\lambda - \theta)\}, \tag{2.20}$$

where $0 < p < 1$ and

$$\mathscr{C}_p(y) = \{u \in R^k: u'y \ge p|u|\,|y|\} \tag{2.21}$$

is a symmetric cone with vertex 0 and subtending an angle $\cos^{-1}p$ about an axis in the direction of the vector y. Note that if M is a positive definite symmetric matrix and $r = \lambda_{\max}(M)/\lambda_{\min}(M)$, then

$$Mx \in \mathscr{C}_{rp}(M^{-1}y) \Rightarrow x \in \mathscr{C}_p(y) \Rightarrow Mx \in \mathscr{C}_{r^{-1}p}(M^{-1}y). \tag{2.22}$$

In the one-dimensional case, the cone $\mathscr{C}_p(y)$ with $0 < p < 1$ reduces to the half-interval $[0,\infty)$ or $(-\infty, 0]$ according as $y > 0$ or $y < 0$, and therefore the event (2.20) is the same as $\{(\nabla\psi)^{-1}(\overline{X}_n)(=\hat{\theta}_n) \in J_\lambda(\theta)\}$. This observation and (2.22) lead to the following multivariate generalization of parts (ii) and (iii) of Theorem 1.

THEOREM 2. *Let $X_1, X_2, \dots$ be i.i.d. random vectors whose common density (with respect to some nondegenerate measure ν) belongs to the k-dimensional exponential family* (2.1). *Let A be a convex subset of the natural parameter space such that* (2.7) *holds. Let $\alpha > 0$ and let g be a nonnegative function on $(0,\infty)$ satisfying* (2.11) *and* (2.17). *Fix $0 < p < 1$, and define the cone $\mathscr{C}_p(y)$ for $y \in R^k$ by* (2.21). *Let $I(\theta, \lambda)$ be the Kullback–Leibler information number defined in* (2.2) *and let $\hat{\theta}_n$ be the maximum likelihood estimate defined in* (2.3). *Let $\{\delta_c: c > 0\}$ be a set of positive numbers satisfying* (2.4).

(i) *As $c \to 0$,*

$$\begin{aligned} &P_\theta\{\hat{\theta}_n \in A_\rho, \overline{X}_n - \nabla\psi(\lambda) \in \mathscr{C}_p(\lambda - \theta) \text{ and} \\ &\qquad I(\hat{\theta}_n, \lambda) \ge n^{-1}g(cn) \text{ for some } n \ge 1\} \\ &\quad = O\Big(\{c^{-1}|\lambda - \theta|^2\}^{-\alpha}\{\log(c^{-1}|\lambda - \theta|^2)\}^{\beta - \alpha + k/2}\Big), \end{aligned} \tag{2.23}$$

uniformly in $\theta \in A$ and $\lambda \in A$ with $|\lambda - \theta|^2 \ge \delta_c$.

(ii) *Let $k \ge 2$. Suppose that condition* (2.17) *is strengthened into the asymptotic expansion* (2.15). *Then as $c \to 0$ and $\lambda \to \theta$ such that $|\lambda - \theta|^2/c \to \infty$ but $|\lambda - \theta|\log c \to 0$,*

$$\begin{aligned} &P_\theta\Big\{\hat{\theta}_n \in A_\rho, (\nabla^2\psi(\theta))^{-1/2}(\overline{X}_n - \nabla\psi(\lambda)) \in \mathscr{C}_p\big((\nabla^2\psi(\theta))^{1/2}(\lambda - \theta)\big) \\ &\qquad\qquad \text{and } I(\hat{\theta}_n, \lambda) \ge n^{-1}g(cn) \text{ for some } n \ge 1\Big\} \\ &\quad \sim \pi^{-1/2}e^{\gamma}\alpha^{k/2-\alpha}\Gamma(2\alpha)\Big\{\int_p^1 2t^{-2\alpha}(1 - t^2)^{(k-3)/2}\,dt/\Gamma\big(\tfrac{1}{2}(k-1)\big)\Big\} \\ &\qquad \times \{2c^{-1}(\lambda - \theta)'\nabla^2\psi(\theta)(\lambda - \theta)\}^{-\alpha}\{\log(c^{-1}|\lambda - \theta|^2)\}^{\beta - \alpha + k/2}, \end{aligned} \tag{2.24}$$

uniformly in $\theta \in A$.

The proof of Theorem 2 involves an asymptotic analysis of certain mixtures of likelihood ratios and is given in Section 4. In Section 3, we make use of certain uniform strong laws to prove the following multivariate generalization of Theorem 1(i).

THEOREM 3. *Let $\alpha > 0$ and let g be a nonnegative function on $(0, \infty)$ satisfying* (2.10) *and* (2.11). *Let $X_1, X_2, \dots, A, \hat{\theta}_n, I(\theta, \lambda)$ be the same as in Theorem* 2. *For $c > 0$ and $\lambda \in \Theta$, define $T_c(\lambda)$, $L_c(\lambda)$ by* (2.8) *and* (2.9). *Let $\delta_c < \Delta_c$ be positive numbers satisfying* (2.4). *Fix $r > 0$ and $0 < \eta < 1$. Then* (2.12), (2.13) *and* (2.14) *still hold* (*in the present k-dimensional case*) *uniformly in $\theta \in A$ and $\lambda \in A_\rho$ such that $\delta_c \le |\lambda - \theta|^2 \le \Delta_c$.*

Statistical applications of the above theorems are presented in Lai (1987a, b). The uniformity of the convergence in these theorems is of particular interest in these applications since it allows us to integrate the asymptotic relations (2.12), (2.19) and (2.23) with respect to (prior) measures on θ in evaluating the Bayes risks. Moreover, the asymptotic relations (2.12) (with $r = 1$) and (2.19) [or (2.23) in the multidimensional case] provide the correct balance between the probability of wrong decision and the expected duration of sampling, for stopping boundaries of the form $I(\hat{\theta}_n, \lambda) \ge n^{-1}g(cn)$.

3. Uniform strong laws and the proof of Theorem 3. Since $E_\theta X = \nabla\psi(\theta)$, it follows from (2.1) that for $t \in R^k$,

$$E_\theta \exp\{t'(X - E_\theta X)\} = \exp\{\psi(\theta + t) - \psi(\theta) - t'\nabla\psi(\theta)\}. \tag{3.1}$$

In view of (2.7) and (3.1),

$$\sup_{\theta \in A} \sup_{|t| < \rho} E_\theta \exp\{t'(X - E_\theta X)\} < \infty, \tag{3.2}$$

which implies that for all $r > 0$,

$$\sup_{\theta \in A} E_\theta |X - E_\theta X|^r < \infty. \tag{3.3}$$

The uniform L_r-boundedness (3.3) leads to the following two uniform strong laws, which we use for the proof of Theorem 3.

LEMMA 2 [Chow and Lai (1975)]. *For every $r \ge 1$, there exists an absolute constant C_r such that for all i.i.d. k-dimensional random vectors $X, X_1, X_2, \dots$ with $EX = 0$ and $E|X|^{r+1} < \infty$, and for all $\varepsilon > 0$,*

$$E\left[\sup\{n: |n^{-1}S_n| \ge \varepsilon\}\right]^r \le C_r\left\{\left(\varepsilon^{-2}E|X|^2\right)^r + \varepsilon^{-(r+1)}E|X|^{r+1}\right\}.$$

LEMMA 3. *Let $X, X_1, X_2, \dots$ be i.i.d.* (*one-dimensional*) *random variables with a common distribution function $F \in \mathscr{F}$ such that $E_F X = 0$ and $\sup_{F \in \mathscr{F}} E_F|X|^r < \infty$ for some $r > 2$. Then for every fixed $\xi > 0$, as $c \downarrow 0$,*

$$\sup_{F \in \mathscr{F}} P_F\left\{|S_n| \ge \xi\left[n \log\left((cn)^{-1}\right)\right]^{1/2} \text{ for some } n \le n_c\right\} \to 0, \tag{3.4}$$

where n_c are positive integers such that $n_c \to \infty$ and $cn_c \to 0$ as $c \downarrow 0$.

PROOF. Since $E_F X = 0$ and $\sup_{F \in \mathscr{F}} E_F|X|^r < \infty$,

$$E_F|S_n|^r = O(n^{r/2}), \quad \text{uniformly in } F \in \mathscr{F}. \tag{3.5}$$

Let $m_c = [\log n_c]$. Then

$$P_F\Big\{|S_n| \geq \xi\big[n\log\big((cn)^{-1}\big)\big]^{1/2} \text{ for some } n \leq n_c\Big\}$$
$$\leq \sum_{k=0}^{m_c} P_F\Big\{\max_{e^k \leq j \leq e^{k+1}} |S_j| \geq \xi\big[e^k\log\big((ce^k)^{-1}\big)\big]^{1/2}\Big\}$$
$$= O\bigg(\sum_{k=0}^{m_c} \{\log(e^{-k}/c)\}^{-r/2}\bigg) \quad \text{[by (3.5)]}$$
$$= O\bigg(\sum_{j=0}^{m_c} \{\log(1/c) - m_c + j\}^{-r/2}\bigg)$$
$$= O\bigg(\sum_{j=0}^{m_c} \Big\{\log\big((cn_c)^{-1}\big) + j\Big\}^{-r/2}\bigg),$$

uniformly in $F \in \mathscr{F}$. Since $r/2 > 1$ and $cn_c \to 0$, (3.4) follows. □

Proof of Theorem 3. Let

$$D = \sup_{\theta \in A_\rho} \lambda_{\max}\big(\nabla^2\psi(\theta)\big), \qquad d = \tfrac{1}{2}\inf_{\theta \in A_\rho} \lambda_{\min}\big(\nabla^2\psi(\theta)\big).$$

Then

$$\text{(3.6a)} \qquad d|\theta - \lambda| \leq |\nabla\psi(\theta) - \nabla\psi(\lambda)| \leq D|\theta - \lambda|$$

and

$$\text{(3.6b)} \qquad d|\theta - \lambda|^2 \leq I(\theta, \lambda) \leq D|\theta - \lambda|^2, \quad \text{for all } \theta, \lambda \in A_\rho,$$

by (2.2). We first show that as $c \to 0$,

$$\text{(3.7)} \qquad P_\theta\big\{T_c(\lambda) \leq \tfrac{1}{4}dD^{-1}\alpha\log\big(c^{-1}|\theta - \lambda|^2\big)/I(\theta, \lambda)\big\} \to 0,$$

uniformly in $\theta \in A$ and $\lambda \in A_\rho$ with $|\theta - \lambda|^2 \geq \delta_c$.

Let $\theta \in A$ and $\lambda \in A_\rho$ with $|\theta - \lambda|^2 \geq \delta_c$. By (3.6b), if $\hat{\theta}_n \in A_\rho$, then

$$I\big(\hat{\theta}_n, \lambda\big) \leq D|\hat{\theta}_n - \lambda|^2 \leq 2D\big\{|\hat{\theta}_n - \theta|^2 + d^{-1}I(\theta, \lambda)\big\},$$

and therefore by (2.10), as $c \to 0$,

$$P_\theta\big\{T_c(\lambda) \leq \tfrac{1}{4}dD^{-1}\alpha\log\big(c^{-1}|\theta - \lambda|^2\big)/I(\theta, \lambda)\big\}$$
$$\leq P_\theta\big\{\hat{\theta}_n \in A_\rho \text{ and } 2D|\hat{\theta}_n - \theta|^2 \geq n^{-1}g(cn)/3 \text{ for some}$$
$$\text{(3.8)} \qquad n \leq \tfrac{1}{4}dD^{-1}\alpha\log\big(c^{-1}|\theta - \lambda|^2\big)/I(\theta, \lambda)\big\}$$
$$\leq P_\theta\bigg\{|S_n - n\nabla\psi(\theta)| \geq \xi\big[n\log\big((cn)^{-1}\big)\big]^{1/2} \text{ for some } n \leq \frac{\alpha\log\big(c^{-1}\delta_c\big)}{4D\delta_c}\bigg\},$$

where $0 < \xi < d(\alpha/6D)^{1/2}$; the last inequality in (3.8) follows from (3.6a) and (2.10). From (3.3), Lemma 3 and (3.8), (3.7) follows.

Let $0 < \eta < 1$. Take a small positive constant $\varepsilon < 1$ (to be specified later) such that for all $x, y \in A_\rho$ with $|x - y|^2 \le \varepsilon$,

$$1 - \tfrac{1}{4}\eta \le I(x, y)/\left\{\tfrac{1}{2}(x - y)'\nabla^2\psi(y)(x - y)\right\} \le 1 + \tfrac{1}{4}\eta. \tag{3.9}$$

We shall consider the two cases $\delta_c \le |\theta - \lambda|^2 \le \varepsilon^2$ and $\varepsilon^2 < |\theta - \lambda|^2 \le \Delta_c$ separately. Let $f_c(\theta, \lambda) = \alpha \log(c^{-1}|\theta - \lambda|^2)/I(\theta, \lambda)$, $m^* = \frac{1}{4}dD^{-1}f_c(\theta, \lambda)$, $\bar{n} = (1 + \eta)f_c(\theta, \lambda)$, $\underline{n} = (1 - \eta)f_c(\theta, \lambda)$. We assume that η is sufficiently small so that $1 - \eta > \frac{1}{4}dD^{-1}$.

CASE 1. $\delta_c \le |\theta - \lambda|^2 \le \varepsilon^2$. We first note by (2.10) that for $m^* \le i \le \underline{n}$, $i^{-1}g(ci) \le 5Dd^{-1}I(\theta, \lambda)$, provided that ε and c are sufficiently small. Since $I(\theta, \lambda) \le D|\theta - \lambda|^2 \le D\varepsilon^2$ by (3.6b), it follows from (3.9) that for $m^* \le i \le \underline{n}$ (and with ε and c sufficiently small)

$$\begin{aligned} &|\hat{\theta}_i - \theta|^2 \le \varepsilon i^{-1}g(ci)(\le 5D^2d^{-1}\varepsilon^2) \Rightarrow |\hat{\theta}_i - \lambda|^2 \le \varepsilon,\ \hat{\theta}_i \in A_\rho \\ &\quad\Rightarrow I(\hat{\theta}_i, \lambda) \le \tfrac{1}{2}(1 + \tfrac{1}{4}\eta)(\hat{\theta}_i - \lambda)'\nabla^2\psi(\lambda)(\hat{\theta}_i - \lambda) \text{ and } \hat{\theta}_i \in A_\rho. \end{aligned} \tag{3.10}$$

Noting that

$$\begin{aligned} (\hat{\theta}_i - \lambda)'\nabla^2\psi(\lambda)(\hat{\theta}_i - \lambda) \le D|\hat{\theta}_i - \theta|^2 + (\theta - \lambda)'\nabla^2\psi(\lambda)(\theta - \lambda) \\ + 2D|\hat{\theta}_i - \theta|\,|\theta - \lambda|, \end{aligned} \tag{3.11}$$

and that $\frac{1}{2}\underline{n}(\theta - \lambda)'\nabla^2\psi(\lambda)(\theta - \lambda) \le (1 - \frac{1}{4}\eta)^{-1}\underline{n}I(\theta, \lambda)$ by (3.9), we obtain from (3.10) that for $m^* \le i \le \underline{n}$,

$$|\hat{\theta}_i - \theta|^2 \le \varepsilon i^{-1}g(ci) \Rightarrow iI(\hat{\theta}_i, \lambda) \le (1 - \tfrac{1}{2}\eta)g(ci), \tag{3.12}$$

provided that ε and c are sufficiently small. From (3.12), it then follows that for sufficiently small ε, as $c \to 0$,

$$\begin{aligned} &P_\theta\{m^* \le T_c(\lambda) \le \underline{n}\} \\ &\quad\le P_\theta\left\{\hat{\theta}_i \in A_\rho \text{ and } |\hat{\theta}_i - \theta|^2 > \varepsilon i^{-1}g(ci) \text{ for some } m^* \le i \le \underline{n}\right\} \\ &\quad\le P_\theta\Big\{|S_i - i\nabla\psi(\theta)| \ge \tfrac{1}{2}d\left[\varepsilon\alpha i \log\left((ci)^{-1}\right)\right]^{1/2} \\ &\qquad\qquad \text{for some } i \le \alpha\delta_c^{-1}\log(c^{-1}\delta_c)\Big\} \quad [\text{by (3.6a) and (2.10)}] \\ &\quad\to 0, \quad \text{uniformly in } \theta \in A \text{ and } \lambda \in A_\rho \text{ with } \delta_c \le |\theta - \lambda|^2 \le \varepsilon^2, \end{aligned} \tag{3.13}$$

by Lemma 3. From (3.7) and (3.13), the desired conclusion (2.13) on $T_c(\lambda)$ follows.

Let $L = \sup\{n\colon |\bar{X}_n - \nabla\psi(\theta)| \ge \varepsilon|\theta - \lambda|\}$. By (3.3) and Lemma 2, for every $r \ge 1$, there exists $M_r > 0$ such that

$$E_\theta L^r \le M_r(\varepsilon|\theta - \lambda|)^{-2r}, \quad \text{for all } \lambda, \theta \in A \text{ with } |\theta - \lambda|^2 \le \varepsilon^2. \tag{3.14}$$

In view of (2.7), we can choose ε sufficiently small such that for every $\theta \in A$ the sphere $\{x\colon |x - \nabla\psi(\theta)| < \varepsilon\}$ is contained in $\nabla\psi(A_\rho)$, as can be shown by arguments similar to those used to prove the inverse function theorem [cf. Spivak (1965), pages 35–38]. For $n > L$, $|\bar{X}_n - \nabla\psi(\theta)| < \varepsilon|\theta - \lambda| \le \varepsilon^2$ and therefore $\overline{X}_n \in \nabla\psi(A_\rho)$; moreover, by (3.6a),

$$|\hat{\theta}_n - \theta|^2 \le d^{-2}|\overline{X}_n - \nabla\psi(\theta)|^2 < d^{-2}\varepsilon^2|\theta - \lambda|^2(\le d^{-2}\varepsilon^4). \tag{3.15}$$

Hence, if ε is sufficiently small, an argument similar to (3.10) and (3.11) then shows that

$$\begin{aligned} n > L \Rightarrow I(\hat{\theta}_n, \lambda) &\ge \tfrac{1}{2}(1 - \tfrac{1}{4}\eta)(\hat{\theta}_n - \lambda)'\nabla^2\psi(\lambda)(\hat{\theta}_n - \lambda) \text{ and } \hat{\theta}_n \in A_\rho \\ \Rightarrow I(\hat{\theta}_n, \lambda) &\ge (1 - \tfrac{1}{2}\eta)I(\theta, \lambda) \text{ and } \hat{\theta}_n \in A_\rho. \end{aligned} \tag{3.16}$$

From (3.16) and (2.10), (2.11), it follows that by choosing ε sufficiently small,

$$L_c(\lambda) < \max\{L + 1, \bar{n}\}, \tag{3.17}$$

and therefore

$$P_\theta\{L_c(\lambda) \ge \bar{n}\} \le P_\theta\{L + 1 \ge \bar{n}\} \le E_\theta(L+1)/\bar{n} \to 0, \quad \text{as } c \to 0, \tag{3.18}$$

by (3.14), uniformly in $\theta \in A$ and $\lambda \in A_\rho$ with $\delta_c \le |\theta - \lambda|^2 \le \varepsilon^2$. This proves (2.14). In view of (3.14) and (3.17), the desired conclusion (2.12) follows from (2.13), (2.14) and the fact that $T_c(\lambda) - 1 \le L_c(\lambda)$.

CASE 2. $\varepsilon^2 < |\theta - \lambda|^2 \le \Delta_c$. In this case, instead of L defined above, we consider

$$\tilde{L} = \sup\{n: |\overline{X}_n - \nabla\psi(\theta)| \ge \varepsilon^2\}.$$

By (3.3) and Lemma 2, for every $r \ge 1$, there exists $M_r > 0$ uch that $E_\theta\tilde{L}^r \le M_r\varepsilon^{-2r}$ for all $\theta \in A$. For $n > \tilde{L}$, $|\overline{X}_n - \nabla\psi(\theta)| < \varepsilon^2$ and therefore $\hat{\theta}_n \in A_\rho$; moreover, by (3.6a), $|\hat{\theta}_n - \theta| \le d^{-1}|\overline{X}_n - \nabla\psi(\theta)| < d^{-1}\varepsilon^2$ while $|\theta - \lambda| > \varepsilon$. Hence, if ε is sufficiently small, then it follows from (2.2) that

$$\begin{aligned} &n > \tilde{L} \Rightarrow \hat{\theta}_n \in A_\rho, \\ &(1 - \tfrac{1}{2}\eta)I(\theta, \lambda) \le I(\hat{\theta}_n, \lambda) \le (1 + \tfrac{1}{2}\eta)I(\theta, \lambda). \end{aligned} \tag{3.19}$$

Therefore (3.17) and (3.18) still hold in this case with $\tilde{L}$ replacing L, noting that by (3.6b)

$$\log(c^{-1}|\theta - \lambda|^2)/I(\theta, \lambda) \ge D^{-1}\log(c^{-1}|\theta - \lambda|^2)/|\theta - \lambda|^2 \to \infty, \tag{3.20}$$

as $c \to 0$, uniformly in $\theta \in A$ and $\lambda \in A_\rho$ with $\varepsilon^2 < |\theta - \lambda|^2 \le \Delta_c$, by (2.4). Moreover, by (3.19) and (3.20),

$$P_\theta\{m^* \le T_c(\lambda) \le \underline{n}\} \le P_\theta\{\tilde{L} \ge m^*\} \le E_\theta\tilde{L}/m^* \to 0, \tag{3.21}$$

as $c \to 0$, uniformly in $\theta \in A$ and $\lambda \in A_\rho$ with $\varepsilon^2 < |\theta - \lambda|^2 \le \Delta_c$. □

4. Boundary crossing probabilities and the proof of Theorems 1 and 2. To highlight the basic ideas in the analysis of the boundary crossing probabilities in Theorems 1 and 2, we focus on the one-dimensional case (Theorem 1) for which the notation is considerably simpler, and then briefly indicate how the methods can be extended to the multidimensional setting of Theorem 2. In the one-dimensional case, we only consider boundary crossing probabilities of the form

$$P_\theta\{\hat{\theta}_n \in A_\rho, \hat{\theta}_n \ge \theta + \delta \text{ and } I(\hat{\theta}_n, \theta + \delta) \ge n^{-1}g(cn) \text{ for some } n \ge 1\}, \tag{4.1}$$

with $\delta > 0$, as an obvious modification of the argument can be used to study

$$P_\theta\{\hat{\theta}_n \in A_\rho, \hat{\theta}_n \le \theta - \delta \text{ and } I(\hat{\theta}_n, \theta - \delta) \ge n^{-1}g(cn) \text{ for some } n \ge 1\}.$$

We first outline the main steps in the asymptotic analysis of (4.1). Introduce the stopping time

$$(4.2) \qquad N = \inf\{n: \hat{\theta}_n \in A_\rho, \hat{\theta}_n \ge \theta + \delta \text{ and } I(\hat{\theta}_n, \theta + \delta) \ge n^{-1}g(cn)\},$$

($\inf \varnothing = \infty$), and note that (4.1) can be expressed as $\sum_{i=0}^{\infty} P_\theta\{b^i \le N < b^{i+1}\}$ for every $b > 1$. Let

$$(4.3) \qquad B(\theta, r) = \{x: |x - \theta| < r\},$$

$$(4.4) \qquad \varepsilon_n = \{Kn^{-1}\log((cn)^{-1})\}^{1/2},$$

where K is some sufficiently large number as specified in Lemma 4. For certain values of m (such that ε_m is much larger than δ), we make use of Lemma 4 to reduce the analysis of $P_\theta\{m \le N < bm\}$ to that of $P_\theta(E_m)$, where

$$(4.5) \qquad E_m = \{m \le N < bm\} \cap \{\hat{\theta}_N \in B(\theta + \delta, 2\varepsilon_m)\}.$$

Letting $\lambda = \theta + \delta$, a key idea in the analysis of $P_\theta(E_m)$ is the representation

$$(4.6) \qquad P_\theta(E_m) = \int_{E_m} \left\{\prod_{i=1}^{N} (h_\theta(X_i)/h_\lambda(X_i))\right\} L_N^{-1}(\lambda, 4\varepsilon_m)\, dQ_m,$$

where the measure Q_m is defined by

$$(4.7) \qquad Q_m(E) = \int_{A_\rho \cap B(\lambda, 4\varepsilon_m)} P_u(E)\, du, \quad \text{for all events } E,$$

and

$$(4.8) \qquad L_n(\lambda, r) = \int_{A_\rho \cap B(\lambda, r)} \prod_{i=1}^{n} (h_u(X_i)/h_\lambda(X_i))\, du.$$

The representation (4.6) of $P_\theta(E_m)$ follows from the (Radon–Nikodym) change-of-measure formula

$$P_\theta(E_m) = \int_{E_m} (dP_\theta/dP_\lambda)(dP_\lambda/dQ_m)\, dQ_m.$$

For other values of m, we obtain bounds on $P_\theta\{m \le N < bm\}$ by first finding bounds for the likelihood ratio $\prod_{i=1}^{N} h_\theta(X_i)/h_\lambda(X_i)$ (with $\lambda = \theta + \delta$) in the representation

$$(4.9) \qquad P_\theta\{m \le N < bm\} = \int_{\{m \le N < bm\}} \prod_{i=1}^{N} (h_\theta(X_i)/h_\lambda(X_i))\, dP_\lambda,$$

and then applying Lemma 4 to reduce the analysis of $P_\lambda\{m \le N < bm\}$ to that of $P_\lambda(E_m)$, which we estimate by using the representation

$$(4.10) \qquad P_\lambda(E_m) = \int_{E_m} L_N^{-1}(\lambda, 4\varepsilon_m)\, dQ_m.$$

A basic step in the analysis of the integrals in (4.6) and (4.10) is the approximation, in Lemma 6, of the mixture (4.8) of likelihood ratios by some multiple of $n^{-1/2}\exp(nI(\hat{\theta}_n, \lambda))$. This involves the classical ideas of the Laplace method for the asymptotic evaluation of an integral. Making use of Lemma 6 and the limiting behavior of the overshoot $NI(\hat{\theta}_N, \theta + \delta) - g(cN)$ given by Lemma 5, Lemmas 7 and 8 provide bounds and estimates of $P_\theta\{m \le N < bm\}$, which are then used to prove the desired conclusions for the boundary crossing probability (4.1).

The same notation and assumptions as Theorem 1 will be used in Lemmas 6–8 below. Lemmas 4 and 5, however, are given in the more general multidimensional context of Theorem 2.

LEMMA 4. *With the same notation and assumptions as in Theorem 2, given $\alpha > 0$, there exists $K > 0$ such that*

$$P_\theta\{\hat{\theta}_n \in A_\rho \text{ and } |\hat{\theta}_n - \theta|^2 \ge Kn^{-1}|\log cn| \text{ for some } n \le m\} = O\big((cm)^{2\alpha}\big), \quad \text{as } cm \to 0, \tag{4.11}$$

uniformly in $\theta \in A_{\rho/2}$.

PROOF. By (3.6a), $|\hat{\theta}_n - \theta|^2 \le d^{-2}|\overline{X}_n - \nabla\psi(\theta)|^2$ for $\theta, \hat{\theta}_n \in A_\rho$. Let x_j denote the jth component of $x \in R^k$. Since $|\overline{X}_n - \nabla\psi(\theta)|^2 \le k \max_{1\le j\le k}(\overline{X}_n - \nabla\psi(\theta))_j^2$, it suffices to show that there exists K sufficiently large such that for every fixed j,

$$\sum_{i=0}^{M} P_\theta\Big\{\max_{2^i \le n \le 2^{i+1}} (S_n - n\nabla\psi(\theta))_j^2 \ge Kd^2 2^i \log\big((c2^i)^{-1}\big)\Big\} = O\big((c2^M)^{2\alpha}\big),$$

as $c2^M \to 0$, uniformly in $\theta \in A_{\rho/2}$. As in (3.1), the moment generating function of $(X - \nabla\psi(\theta))_j$ is given by

$$E_\theta \exp\{u(X - \nabla\psi(\theta))_j\} = \exp\{\psi(\theta_1, \dots, \theta_{j-1}, \theta_j + u, \theta_{j+1}, \dots, \theta_k) - \psi(\theta) - u(\nabla\psi(\theta))_j\}.$$

Let $\eta_r = (d^2|\log(cr)|)^{1/2}$. In view of (2.7), we can apply standard martingale and exponential inequalities [cf. Chow and Teicher (1978)] to show that if K is sufficiently large, then

$$\sup_{\theta \in A_{\rho/2}} P_\theta\Big\{\max_{r \le n \le 2r} (S_n - n\nabla\psi(\theta))_j \ge K^{1/2} r^{1/2}\eta_r\Big\} \le \sup_{\theta \in A_{\rho/2}} \{\exp(-K^{1/2}\eta_r^2)\}\big(E_\theta \exp\{r^{-1/2}\eta_r(X - \nabla\psi(\theta))_j\}\big)^{2r} = O\big((cr)^{2\alpha}\big),$$

$$\sup_{\theta \in A_{\rho/2}} P_\theta\Big\{\max_{r \le n \le 2r} (n\nabla\psi(\theta) - S_n)_j \ge K^{1/2} r^{1/2}\eta_r\Big\} = O\big((cr)^{2\alpha}\big),$$

as $cr \to 0$. Hence the desired conclusion follows. □

LEMMA 5. *With the same notation and assumptions as in Theorem* 2, *as* $m \to \infty$ *and* $x \to 0$ *such that* $|x|\log m \to 0$,

$$P_\theta\{\hat{\theta}_n \notin A_\rho \text{ for some } n \geq m\} \to 0 \tag{4.12}$$

and

$$P_\theta\{|nI(\hat{\theta}_n, \theta + x) - (n-1)I(\hat{\theta}_{n-1}, \theta + x)| \geq \varepsilon \text{ for some } m \leq n \leq m^2\} \to 0, \tag{4.13}$$

for every $\varepsilon > 0$, *uniformly in* $\theta \in A_{\rho/2}$.

PROOF. In view of (3.2), we can use an argument similar to that used in the proof of Theorem 3 [in particular (3.19) and (3.21)] to prove (4.12). Moreover, it follows from (3.2) that there exists $K > 0$ sufficiently large so that as $m \to \infty$,

$$\sum_{n=m}^{\infty} P_\theta\{|X_n - \nabla\psi(\theta)| \geq K \log n\} \to 0 \tag{4.14}$$

and

$$\sum_{n=m}^{\infty} P_\theta\{|S_n - n\nabla\psi(\theta)| \geq n^{3/4}\} \to 0, \tag{4.15}$$

uniformly in $\theta \in A_\rho$. Suppose that λ, $\hat{\theta}_n$ and $\hat{\theta}_{n-1}$ belong to A_ρ. Then by (2.2),

$$\begin{aligned} &nI(\hat{\theta}_n, \lambda) - (n-1)I(\hat{\theta}_{n-1}, \lambda) \\ &\quad = I(\hat{\theta}_{n-1}, \lambda) + n\{I(\hat{\theta}_n, \hat{\theta}_{n-1}) + (\hat{\theta}_{n-1} - \lambda)'(\nabla\psi(\hat{\theta}_n) - \nabla\psi(\hat{\theta}_{n-1}))\}. \end{aligned} \tag{4.16}$$

Noting that $\nabla\psi(\hat{\theta}_n) = \overline{X}_n$, it follows from (3.6a) and (3.6b) that

$$\begin{aligned} |\hat{\theta}_{n-1} - \lambda| &\leq |\theta - \lambda| + d^{-1}|\overline{X}_{n-1} - \nabla\psi(\theta)|, \\ I(\hat{\theta}_n, \hat{\theta}_{n-1}) &\leq Dd^{-2}|\overline{X}_n - \overline{X}_{n-1}|^2, \\ \nabla\psi(\hat{\theta}_n) - \nabla\psi(\hat{\theta}_{n-1}) &= \overline{X}_n - \overline{X}_{n-1}, \\ |\overline{X}_n - \overline{X}_{n-1}| &\leq \{|X_n - \nabla\psi(\theta)| + |\overline{X}_n - \nabla\psi(\theta)|\}/(n-1). \end{aligned} \tag{4.17}$$

From (4.12) and (4.14)–(4.17), (4.13) follows. □

LEMMA 6. *Let* $X_1, X_2, \ldots$ *be i.i.d. random variables having common density* $h_\theta(x) = \exp(\theta x - \psi(\theta))$ *with respect to some nondegenerate measure* ν. *Let* A *be a subinterval of the natural parameter space such that* (2.7) *holds. Define* $B(\theta, r)$ *by* (4.3) *and* $L_n(\lambda, r)$ *by* (4.8).

(i) *There exists* $\xi > 0$ *such that for all* $n \geq 1$, $\lambda \in A_{\rho/2}$ *and* $r \geq n^{-1/2}$,

$$L_n(\lambda, r) \geq \xi n^{-1/2}\exp(nI(\hat{\theta}_n, \lambda)), \quad \text{on } \{\hat{\theta}_n \in A_\rho \cap B(\lambda, r/2)\}. \tag{4.18}$$

(ii) *Let* r_n *be positive numbers such that* $r_n \to 0$ *but* $n^{1/2}r_n \to \infty$ *as* $n \to \infty$. *Then as* $n \to \infty$,

$$L_n(\lambda, r_n) \sim (2\pi/n\psi''(\lambda))^{1/2}\exp(nI(\hat{\theta}_n, \lambda)), \tag{4.19}$$

on $\{\hat{\theta}_n \in A_\rho \cap B(\lambda, r_n/2)\}$, *uniformly in* $\lambda \in A_{\rho/2}$.

Proof. Let $\lambda \in A_{\rho/2}$ and assume that $\hat{\theta}_n \in A_\rho \cap B(\lambda, r/2)$. For $u \in A_\rho \cap B(\lambda, r)$, we obtain from Taylor's expansion of $\varphi(u) = (u - \lambda)\overline{X}_n - (\psi(u) - \psi(\lambda))$ about the maximizing value $u = \hat{\theta}_n$ [with $\psi'(\hat{\theta}_n) = \overline{X}_n$] that

$$\prod_{i=1}^{n} (h_u(X_i)/h_\lambda(X_i)) = \exp\{(u - \lambda)S_n - n(\psi(u) - \psi(\lambda))\} \tag{4.20}$$
$$= \exp\left\{nI(\hat{\theta}_n, \lambda) - \tfrac{1}{2}n\psi''(\theta_n^*)(u - \hat{\theta}_n)^2\right\},$$

where θ_n^* lies between $\hat{\theta}_n$ and u and therefore belongs to A_ρ. Letting $\tilde{r} = \min\{r, \rho\}$ and $D = \sup_{\theta \in A_\rho}\psi''(\theta)$, it follows from (4.8) and (4.20) that,

$$L_n(\lambda, r) \geq \exp\left(nI(\hat{\theta}_n, \lambda)\right)\int_0^{\tilde{r}} \exp\left(-\tfrac{1}{2}Dnt^2\right) dt,$$

proving (i). To prove (ii), note that $B(\lambda, r) \subset A_\rho$ if $r < \rho/2$. For $r = r_n(\to 0)$, it follows from (4.8) and (4.20) that as $n \to \infty$,

$$L_n(\lambda, r_n) = \exp\left(nI(\hat{\theta}_n, \lambda)\right)\int_{\lambda - \hat{\theta}_n - r_n}^{\lambda - \hat{\theta}_n + r_n} \exp\left\{-\tfrac{1}{2}(1 + o(1))n\psi''(\hat{\theta}_n)t^2\right\} dt, \tag{4.21}$$

where the $o(1)$ term is uniform in $\lambda \in A_{\rho/2}$, in view of (2.7). Since $|\lambda - \hat{\theta}_n| < r_n/2$, (4.19) follows from (4.21). □

Lemma 7. *With the same notation and assumptions as in Theorem* 1(iii), *let* $\delta > 0$ *and define* N *by* (4.2). *Let* $d = \frac{1}{2}\inf_{\theta \in A_\rho}\psi''(\theta)$, $D = \sup_{\theta \in A_\rho}\psi''(\theta)$, $\tilde{d} = \frac{1}{2}d(D^{-1}\alpha)^{1/2}$. *Then there exist positive constants* $B > 1 > q$ *such that if* $cm \leq q$,

$$P_\theta\{m \leq N < 2m\} \tag{4.22}$$
$$\leq B(cm)^\alpha|\log cm|^{\beta + 1/2}\exp\left\{-dm\delta^2 - \tilde{d}\left(m\delta^2|\log cm|\right)^{1/2}\right\},$$

for all $\theta \in A_\rho$ *with* $\theta + \delta \in A_{\rho/2}$. *Moreover, for all* $m \geq 1$ *and for all* $\theta \in A_\rho$ *with* $\theta + \delta \in A_\rho$,

$$P_\theta\{m \leq N < \infty\} \leq \exp(-dm\delta^2). \tag{4.23}$$

Proof. First note that

$$\prod_{i=1}^{N} (h_\theta(X_i)/h_{\theta+\delta}(X_i)) = \exp\{-\delta S_N - N(\psi(\theta) - \psi(\theta + \delta))\} \tag{4.24}$$
$$= \exp\left\{-\delta N(\overline{X}_N - \psi'(\theta + \delta)) - NI(\theta + \delta, \theta)\right\}.$$

By the definition of N, $\overline{X}_N = \psi'(\hat{\theta}_N) \geq \psi'(\theta + \delta)$. Since $I(\theta + \delta, \theta) \geq d\delta^2$ by (3.6b), (4.23) follows from (4.9) (with $\lambda = \theta + \delta$) and (4.24).

In view of (3.6a) and (3.6b),

$$\left(\overline{X}_N - \psi'(\theta + \delta)\right)^2 \geq d^2\left(\hat{\theta}_N - \theta - \delta\right)^2 \tag{4.25}$$
$$\geq d^2D^{-1}I\left(\hat{\theta}_N, \theta + \delta\right) \geq d^2D^{-1}N^{-1}g(cN).$$

From (4.24), (4.25), (2.17), and the fact that $I(\theta + \delta, \theta) \geq d\delta^2$, it follows that on the event $\{m \leq N < 2m\}$,

$$\prod_{i=1}^{N} (h_\theta(X_i)/h_{\theta+\delta}(X_i)) \tag{4.26}$$
$$\leq \exp\left\{-\delta m^{1/2}\left(d^2 D^{-1}\alpha|\log cm|\right)^{1/2}(1 + o(1)) - dm\delta^2\right\},$$

as $cm \to 0$, uniformly in $\theta \in A_\rho$ and $\theta + \delta \in A_\rho$. In view of (4.9) and (4.26), (4.22) will follow if it can be shown that as $cm \to 0$,

$$P_{\theta+\delta}\{m \leq N < 2m\} = O\left((cm)^\alpha|\log cm|^{\beta+1/2}\right), \tag{4.27}$$
$$\text{uniformly in } \theta + \delta \in A_{\rho/2}.$$

Since $\hat{\theta}_N \in A_\rho$, it follows from Lemma 4 that

$$P_{\theta+\delta}\{m \leq N < 2m, \hat{\theta}_N \notin B(\theta + \delta, 2\varepsilon_m)\} = O\left((cm)^{2\alpha}\right), \tag{4.28}$$

uniformly in $\theta + \delta \in A_{\rho/2}$, where ε_m and $B(\lambda, r)$ are defined in (4.3) and (4.4). Defining Q_m by (4.7) (with $\lambda = \theta + \delta$) and E_m by (4.5) (with $b = 2$), we obtain by (4.10) and Lemma 6(i) that

$$\begin{aligned} P_{\theta+\delta}&\{m \leq N < 2m, \hat{\theta}_N \in B(\theta + \delta, 2\varepsilon_m)\} \\ &= P_{\theta+\delta}(E_m) \\ &= \int_{E_m} L_N^{-1}(\theta + \delta, 4\varepsilon_m)\, dQ_m \\ &\leq \xi^{-1}\int_{\{m \leq N < 2m\}} N^{1/2}\exp(-g(cN))\, dQ_m \\ &\leq 2\xi^{-1}(2m)^{1/2}(2cm)^\alpha|\log cm|^\beta \int_{A_\rho \cap B(\theta+\delta, 4\varepsilon_m)} du, \end{aligned} \tag{4.29}$$

where the last inequality follows from (2.17) and (4.7). Since $B(\theta + \delta, 4\varepsilon_m)$ has width $8(Km^{-1}|\log cm|)^{1/2}$, the desired conclusion (4.27) follows from (4.28) and (4.29). □

LEMMA 8. *With the same notation and assumptions as in Theorem* 1(ii), *let* $b > 1$, $\delta > 0$, *and define* N *by* (4.2). *Let*

$$m(c, \delta) = 1/\{\delta^2\log(c^{-1}\delta^2)\}. \tag{4.30}$$

Then as $c \to 0$ *and* $\delta \to 0$ *such that* $\delta^2/c \to \infty$ *but* $\delta \log c \to 0$,

$$P_\theta\{b^{-1}m(c, \delta) \leq N < bm(c, \delta)\}$$
$$\sim \pi^{-1/2}e^\gamma\alpha^{1/2-\alpha}\left(2\psi''(\theta)\delta^2/c\right)^{-\alpha}\left(\log \delta^2/c\right)^{\beta-\alpha+1/2}\int_{(2\alpha b^{-1}\psi''(\theta))^{1/2}}^{(2\alpha b\psi''(\theta))^{1/2}} s^{2\alpha-1}e^{-s}\, ds,$$

uniformly in $\theta \in A$.

PROOF. First note that $cm(c, \delta) \to 0$ and $\log((cm(c, \delta))^{-1}) \sim \log(c^{-1}\delta^2)$. By (4.4), $\varepsilon_{m(c,\delta)} \sim K^{1/2}\delta\log(c^{-1}\delta^2) \to 0$, so $\delta = o(\varepsilon_{m(c,\delta)})$. For simplicity, write

$m = m(c, \delta)$ and define E_m as in (4.5). From Lemma 4 and the fact that $\delta = o(\varepsilon_m)$ and $\hat{\theta}_N \in A_\rho$, it follows that

$$(4.31) \qquad P_\theta\{m \le N < bm,\ \hat{\theta}_N \notin B(\theta + \delta, 2\varepsilon_m)\} = O\big((cm)^{2\alpha}\big),$$

and therefore

$$(4.32) \qquad P_\theta\{m \le N < bm\} = P_\theta(E_m) + O\big((cm)^{2\alpha}\big).$$

By (4.6) and Lemma 6(ii), as $c \to 0$ and $\delta \to 0$ such that $\delta^2/c \to \infty$ and $\delta \log c \to 0$,

$$(4.33) \quad P_\theta(E_m) \sim (\psi''(\theta)/2\pi)^{1/2} \int_{E_m} \left\{ \prod_{i=1}^{N} \frac{h_\theta(X_i)}{h_{\theta+\delta}(X_i)} \right\} N^{1/2} e^{-NI(\hat{\theta}_N, \theta+\delta)}\, dQ_m,$$

uniformly in $\theta \in A$, where Q_m is defined in (4.7) with $\lambda = \theta + \delta$. Let

$$U = \delta N(\bar{X}_N - \psi'(\theta + \delta)) + NI(\theta + \delta, \theta) + \big(NI(\hat{\theta}_N, \theta + \delta) - g(cN)\big).$$

Then by the definition of N,

$$(4.34) \qquad U \ge 0, \quad \text{so } e^{-U} \le 1.$$

Moreover, by (4.24), (4.33) and (4.34),

$$(4.35) \quad \begin{aligned} P_\theta(E_m) &\sim (\psi''(\theta)/2\pi)^{1/2} \int_{E_m} N^{1/2} \exp(-U - g(cN))\, dQ_m \\ &\sim (\psi''(\theta)/2\pi)^{1/2} e^{\gamma} |\log cm|^{\beta} c^{\alpha} \int_{E_m} N^{\alpha+1/2} e^{-U}\, dQ_m, \end{aligned}$$

in view of (2.15). Since $\theta \in A$ and $\varepsilon_m \sim K^{1/2}\delta \log(c^{-1}\delta^2) \to 0$ as $\delta \to 0$ and $c \to 0$ such that $\delta^2/c \to \infty$ and $\delta \log c \to 0$, we can assume that $B(\theta + \delta, 4\varepsilon_m) \subset A_\rho$, and therefore in view of (4.34) and the fact that $1 \le N/m < b$ on E_m, we obtain from (4.7) that

$$(4.36) \quad \begin{aligned} &\int_{E_m} N^{\alpha+1/2} e^{-U}\, dQ_m \\ &\quad \sim m^{\alpha+1/2} \delta \log(c^{-1}\delta^2) \\ &\qquad \times \int_{-4K^{1/2}}^{4K^{1/2}} \left\{ \int_{E_m} (N/m)^{\alpha+1/2} e^{-U}\, dP_{\theta+\delta+t\delta\log(c^{-1}\delta^2)} \right\} dt. \end{aligned}$$

Without loss of generality, we assume that K (in Lemma 4) is large enough so that

$$(4.37) \qquad K > 2\alpha / \inf_{x \in A} \psi''(x).$$

Let $0 < \eta$ (sufficiently small). An argument similar to the proof of Theorem 3 (Case 1) shows that as $c \to 0$ and $\delta \to 0$ such that $\delta^2/c \to \infty$ but $\delta \log c \to 0$,

$$(4.38) \quad \begin{aligned} &P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\{m \le N < bm\} \to 0, \quad \text{uniformly in } \theta \in A, \\ &\text{and in} \quad -4K^{1/2} \le t \le (2\alpha/b\psi''(\theta))^{1/2} - \eta \\ &\text{or} \quad (2\alpha/\psi''(\theta))^{1/2} + \eta \le t \le 4K^{1/2}. \end{aligned}$$

Moreover, uniformly in $\theta \in A$ and $(2\alpha/\psi''(\theta))^{1/2} \geq t \geq (2\alpha/b\psi''(\theta))^{1/2}$,

$$(4.39)\qquad P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\left\{\left|\frac{N}{m} - \frac{2\alpha}{t^2\psi''(\theta)}\right| \leq \eta^2\right\} \to 1,$$

and we now make use of Lemma 5 to show also that

$$(4.40)\qquad P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\{|U - 2\alpha/t| \leq \eta\} \to 1.$$

To prove (4.40), first note that $(N-1)I(\hat{\theta}_{N-1}, \theta+\delta) < g(cN)$ if $\hat{\theta}_{N-1} \geq \theta + \delta$ and $\hat{\theta}_{N-1} \in A_\rho$. Moreover, an argument similar to (3.18) can be used to show that

$$(4.41)\qquad P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\{\hat{\theta}_n < \theta + \delta \text{ for some } n \geq m/2\} \to 0.$$

From Lemma 5, (4.39) and (4.41), it then follows that

$$(4.42)\qquad P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\{NI(\hat{\theta}_N, \theta+\delta) - g(cN) \geq \eta/3\} \to 0.$$

On $\{\hat{\theta}_N \in A_\rho \cap B(\theta+\delta, 2\varepsilon_m)\}$, since $\overline{X}_N = \psi'(\hat{\theta}_N) \geq \psi'(\theta+\delta)$,

$$(4.43)\qquad \overline{X}_N - \psi'(\theta+\delta) \sim \psi''(\theta)(\hat{\theta}_N - \theta - \delta) \sim \{2\psi''(\theta)I(\hat{\theta}_N, \theta+\delta)\}^{1/2}.$$

Since $\varepsilon_m \sim K^{1/2}\delta\log(c^{-1}\delta^2) \to 0$, it then follows from (4.12), (4.37), (4.39) and Lemma 4 that

$$(4.44)\qquad P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\{\hat{\theta}_N \in A_\rho \cap B(\theta+\delta, 2\varepsilon_m)\} \to 1,$$

uniformly in $\theta \in A$ and $(2\alpha/\psi''(\theta))^{1/2} \geq t \geq (2\alpha/b\psi''(\theta))^{1/2}$. From (4.39), (4.42)–(4.44) and (2.15), it then follows that

$$(4.45)\qquad P_{\theta+\delta+t\delta\log(c^{-1}\delta^2)}\{|\delta N(\overline{X}_N - \psi'(\theta+\delta)) - 2\alpha/t| \leq \eta/3\} \to 1.$$

From (4.42), (4.45) and the fact that $m\delta^2 \to 0$, (4.40) follows.

In view of (4.34), (4.38)–(4.40) and (4.44), it follows from (4.36) that

$$(4.46)\qquad \begin{aligned}&\int_{E_m} N^{\alpha+1/2}e^{-U}\,dQ_m\\ &\sim m^{\alpha+1/2}\delta\log(c^{-1}\delta^2)\int_{(2\alpha/b\psi''(\theta))^{1/2}}^{(2\alpha/\psi''(\theta))^{1/2}}\left(\frac{2\alpha}{t^2\psi''(\theta)}\right)^{\alpha+1/2}\exp(-2\alpha/t)\,dt.\end{aligned}$$

Using a change of variable $s = 2\alpha/t$ in (4.46), we obtain from (4.35) and (4.30) that $P_\theta(E_m)$ is asymptotically equivalent to

$$\pi^{-1/2}e^{\gamma}\alpha^{1/2-\alpha}c^{\alpha}(2\psi''(\theta)\delta^2)^{-\alpha}(\log c^{-1}\delta^2)^{\beta-\alpha+1/2}\int_{(2\alpha\psi''(\theta))^{1/2}}^{(2\alpha b\psi''(\theta))^{1/2}} s^{2\alpha-1}e^{-s}\,ds.$$

A similar analysis of $P_\theta\{b^{-1}m \leq N < m\}$ then completes the proof. □

PROOF OF THEOREM 1(iii). Take $\sigma > 0$ and define $m = m(c,\delta)$ as in (4.30). Assume that $\delta^2 \geq \delta_c$. Then $cm \leq c/\{\delta_c\log(c^{-1}\delta_c)\} \to 0$, so by Lemma 7, there exist positive constants $q < 1 < B$ and $d, \tilde{d}$ such that (4.22) and (4.23) hold, and therefore

$$(4.47)\qquad \sum_{j:\,2^j\le \sigma m} P_\theta\{2^j \le N < 2^{j+1}\} \le Bc^\alpha \sum_{j:\,2^j\le \sigma m} r_j(c),$$

$$(4.48)\qquad \sum_{j:\,\sigma m\le 2^j\le q/c} P_\theta\{2^j \le N < 2^{j+1}\} \le Bc^\alpha \sum_{j:\,\sigma m\le 2^j\le q/c} s_j(c,\delta),$$

where

$$r_j(c) = 2^{\alpha j}|\log(c2^j)|^{\beta+1/2}$$

and

$$s_j(c,\delta) = r_j(c)\exp\left\{-\tilde{d}\left(2^j\delta^2|\log(c2^j)|\right)^{1/2}\right\}.$$

Since $r_j(c)/r_{j+1}(c) \sim 2^{-\alpha}$, it then follows that as $c \to 0$,

$$(4.49)\qquad \begin{aligned}\sum_{j:\,2^j\le \sigma m} r_j(c) &\le (\sigma m)^\alpha|\log(c\sigma m)|^{\beta+1/2}\sum_{i=0}^{\infty}2^{-\alpha i}(1+o(1))\\ &\le \left(1-2^{-\alpha/2}\right)^{-1}(\sigma m)^\alpha|\log cm|^{\beta+1/2}.\end{aligned}$$

If σ is chosen large enough so that $2^\alpha\exp\{-(2^{1/2}-1)\sigma^{1/2}\tilde{d}\} < \frac{1}{2}$, we have by (4.30) that for all small c, $s_{j+1}(c,\delta)/s_j(c,\delta) \le \frac{1}{2}$ for $\delta^2 \ge \delta_c$ and $\sigma m \le 2^j < q/c$ (assuming that q is sufficiently small), and therefore

$$(4.50)\qquad \sum_{j:\,\sigma m\le 2^j\le q/c} s_j(c,\delta) \le 4(\sigma m)^\alpha|\log cm|^{\beta+1/2}\exp(-\sigma^{1/2}\tilde{d}).$$

By (4.23),

$$(4.51)\qquad P_\theta\{N \ge q/c\} \le \exp\left\{-dq(\delta^2/c)\right\}.$$

Since $(cm)^\alpha|\log cm|^{\beta+1/2} \sim (c/\delta^2)^\alpha(\log\delta^2/c)^{\beta-\alpha+1/2}$, the desired conclusion (2.19) follows from (4.47)–(4.51). □

Proof of Theorem 1(ii). Take $b > 1$ arbitrarily large. Apply (4.47) and (4.49) with $\sigma = b^{-1}$, (4.48) and (4.50) with $\sigma = b$, together with Lemma 8 to obtain the desired asymptotic relation (2.16). □

We now show how the preceding argument can be modified to prove the multidimensional generalization in Theorem 2. In the k-dimensional setting, $B(\theta, r)$ defined by (4.3) represents an open ball with center θ and radius r, and we define ε_n, Q_m, $L_n(\lambda, r)$ by (4.3), (4.7) and (4.8) as before. A straightforward modification of the proof of Lemma 6 shows that there exists $\xi > 0$ such that for all $n \ge 1$, $\lambda \in A_{\rho/2}$ and $r \ge n^{-1/2}$,

$$(4.52)\qquad L_n(\lambda, r) \ge \xi n^{-k/2}\exp\left(nI(\hat{\theta}_n, \lambda)\right),\quad \text{on } \left\{\hat{\theta}_n \in A_\rho \cap B(\lambda, r/2)\right\};$$

moreover, if $r_n \downarrow 0$ but $n^{1/2}r_n \to \infty$, then as $n \to \infty$,

$$(4.53)\qquad L_n(\lambda, r_n) \sim (2\pi/n)^{k/2}\left(\det\nabla^2\psi(\lambda)\right)^{-1/2}\exp\left(nI(\hat{\theta}_n, \lambda)\right),$$

on $\{\hat{\theta}_n \in A_\rho \cap B(\lambda, r_n/2)\}$, uniformly in $\lambda \in A_{\rho/2}$.

PROOF OF THEOREM 2(i). Define

$$N_1 = \inf\{n: \hat{\theta}_n \in A_\rho, \bar{X}_n - \nabla\psi(\lambda) \in \mathscr{C}_p(\lambda - \theta) \text{ and } I(\hat{\theta}_n, \lambda) \geq n^{-1}g(cn)\}$$

($\inf \varnothing = \infty$). In analogy with (4.24), we now have

$$\prod_{i=1}^{n}(h_\theta(X_i)/h_\lambda(X_i)) = \exp\{-n(\lambda - \theta)'(\bar{X}_n - \nabla\psi(\lambda)) - nI(\lambda, \theta)\}. \quad (4.54)$$

By the definition of N_1, (2.21), (3.6a) and (3.6b),

$$\begin{aligned}(\lambda - \theta)'(\bar{X}_{N_1} - \nabla\psi(\lambda)) &\geq p|\lambda - \theta|\,|\bar{X}_{N_1} - \nabla\psi(\lambda)| \\ &\geq p|\lambda - \theta|\{d^2D^{-1}N^{-1}g(cN)\}^{1/2}, \quad \text{as in (4.25)}.\end{aligned} \quad (4.55)$$

In view of (4.52), (4.54) and (4.55), we can use the same argument as in the proof of Lemma 7 and Theorem 1(iii) to complete the proof, noting that the volume of $B(\lambda, 4\varepsilon_m)$ is a constant multiple of ε_n^k. □

PROOF OF THEOREM 2(ii). Define

$$N_2 = \inf\{n: \hat{\theta}_n \in A_\rho, (\nabla^2\psi(\theta))^{-1/2}(\bar{X}_n - \nabla\psi(\lambda)) \in \mathscr{C}_p((\nabla^2\psi(\theta))^{1/2}(\lambda - \theta)) \text{ and } I(\hat{\theta}_n, \lambda) \geq n^{-1}g(cn)\}$$

($\inf \varnothing = \infty$). Analogous to (4.55), we obtain by (2.22) that

$$(\lambda - \theta)'(\bar{X}_{N_2} - \nabla\psi(\lambda)) \geq p(d/D)^{1/2}|\lambda - \theta|\{d^2D^{-1}N^{-1}g(cN)\}^{1/2}. \quad (4.56)$$

Hence, using an argument similar to the proof of Theorem 1(ii), we need only prove the following analog of Lemma 8: Let $b > 1$,

$$\delta^2 = (\lambda - \theta)'\nabla^2\psi(\theta)(\lambda - \theta), \quad (4.57)$$

and define $m(c, \delta)$ by (4.30). Then as $c \to 0$ and $\delta \to 0$ such that $\delta^2/c \to \infty$ but $\delta \log c \to 0$,

$$\begin{aligned}P_\theta\{b^{-1}m(c,\delta) \leq N_2 < bm(c,\delta)\} &\sim \pi^{-1/2}e^\gamma\alpha^{k/2-\alpha}(2\delta^2/c)^{-\alpha}(\log\delta^2/c)^{\beta-\alpha+k/2}\{\Gamma(\tfrac{1}{2}(k-1))\}^{-1} \\ &\quad \times \int_p^1 2t^{-2\alpha}(1-t^2)^{(k-3)/2}\left\{\int_{(2\alpha b^{-1})^{1/2}t}^{(2\alpha b)^{1/2}t} s^{2\alpha-1}e^{-s}\,ds\right\}dt.\end{aligned} \quad (4.58)$$

To prove (4.58), write $m = m(c, \delta)$ for simplicity and define

$$U = N_2(\lambda - \theta)'(\bar{X}_{N_2} - \nabla\psi(\lambda)) + N_2 I(\lambda, \theta) + (N_2 I(\hat{\theta}_{N_2}, \lambda) - g(cN_2)).$$

In view of (4.56), (4.34) still holds. Let $V_\theta = \nabla^2\psi(\theta)$ and let $E_m = \{m \leq N_2 < bm\} \cap \{\hat{\theta}_{N_2} \in B(\lambda, 2\varepsilon_m)\}$. In analogy with (4.35), we now have

$$\begin{aligned}P_\theta(E_m) &\sim (2\pi)^{-k/2}(\det V_\theta)^{1/2}\int_{E_m} N_2^{k/2}\exp(-U - g(cN))\,dQ_m \\ &\sim (2\pi)^{-k/2}(\det V_\theta)^{1/2}e^\gamma|\log cm|^\beta c^\alpha\int_{E_m} N_2^{\alpha+k/2}e^{-U}\,dQ_m,\end{aligned} \quad (4.59)$$

by (4.6), (4.53) and (4.54). Moreover, in analogy with (4.36),

$$\int_{E_m} N_2^{\alpha+k/2} e^{-U}\, dQ_m$$

$$\sim m^{\alpha+k/2}\{\delta \log(c^{-1}\delta^2)\}^k \det\left(V_\theta^{-1/2}\right) \tag{4.60}$$

$$\times \int_{|V_\theta^{-1/2}x| \le 4K^{1/2}} \left\{ \int_{E_m} (N_2/m)^{\alpha+k/2} e^{-U}\, dP_{\lambda+(\delta \log c^{-1}\delta^2)V_\theta^{-1/2}x} \right\} dx.$$

Assume, without loss of generality, that $K > 2\alpha/\inf_{\theta\in A}\lambda_{\min}(V_\theta)$. Let $0 < \eta$ (sufficiently small). Noting that for $\theta \in A$ and $2\alpha/b \le |x|^2 \le 2\alpha$, as $c \to 0$ and $\lambda \to \theta$ such that $|\lambda - \theta|^2/c \to \infty$ but $|\lambda - \theta|\log c \to 0$,

$$\left\{V_\theta^{1/2}(\lambda - \theta)\right\}'\left\{V_\theta^{-1/2}\left[\nabla\psi\left(\lambda + \delta(\log c^{-1}\delta^2)V_\theta^{-1/2}x\right) - \nabla\psi(\lambda)\right]\right\}$$

$$\sim (\lambda - \theta)'V_\theta^{1/2}x(\delta \log c^{-1}\delta^2),$$

$$\left|V_\theta^{-1/2}\left[\nabla\psi\left(\lambda + \delta(\log c^{-1}\delta^2)V_\theta^{-1/2}x\right) - \nabla\psi(\lambda)\right]\right| \sim (\delta \log c^{-1}\delta^2)|x|,$$

it can be shown by an argument similar to the proof of Theorem 3 (Case 1) that

$$P_{\lambda+(\delta \log c^{-1}\delta^2)V_\theta^{-1/2}x}\left\{\left|\frac{N_2}{m} - \frac{2\alpha}{|x|^2}\right| \le \eta^{1/2}\right\}$$

$$\to 1, \qquad x \in \mathscr{C}_{p+\eta}\left(V_\theta^{1/2}(\lambda - \theta)\right), \tag{4.61}$$

$$\to 0, \qquad x \notin \mathscr{C}_{p-\eta}\left(V_\theta^{1/2}(\lambda - \theta)\right),$$

the convergence being uniform in $\theta \in A$ and x belonging to the regions indicated and such that $\eta \le |V_\theta^{-1/2}x| \le 4K^{1/2}$. Furthermore, uniformly in $\theta \in A$ and $x \in \mathscr{C}_{p+\eta}(V_\theta^{1/2}(\lambda - \theta))$ with $\eta \le |V_\theta^{-1/2}x| \le 4K^{1/2}$,

$$P_{\lambda+(\delta \log c^{-1}\delta^2)V_\theta^{-1/2}x}\left\{\left|U - 2\alpha|x|^{-1}\varphi\left(x, V_\theta^{1/2}(\lambda - \theta)\right)\right| \le \eta^{1/4}\right\} \to 1, \tag{4.62}$$

where $\varphi(x, y) = x'y/(|x|\,|y|)$ is the cosine of the angle between the vectors x and y. The proof of (4.62) uses (4.61) and an argument similar to the proof of (4.40), noting that on $\{\hat\theta_{N_2} \in A_\rho \cap B(\lambda, 2\varepsilon_m)\}$,

$$(\lambda - \theta)'\left(\bar X_{N_2} - \nabla\psi(\lambda)\right)$$

$$\sim (\lambda - \theta)'\nabla^2\psi(\theta)\left(\hat\theta_{N_2} - \lambda\right)$$

$$= |V_\theta^{1/2}(\lambda - \theta)|\,\left|V_\theta^{1/2}\left(\hat\theta_{N_2} - \lambda\right)\right|\varphi\left(V_\theta^{1/2}\left(\hat\theta_{N_2} - \lambda\right), V_\theta^{1/2}(\lambda - \theta)\right)$$

$$\sim \delta\left(2I\left(\hat\theta_{N_2}, \lambda\right)\right)^{1/2}\varphi\left(V_\theta^{1/2}\left(\hat\theta_{N_2} - \lambda\right), V_\theta^{1/2}(\lambda - \theta)\right), \quad \text{by (4.57)}.$$

Moreover, an argument similar to the proof of Theorem 3 also shows that $P_{\lambda+(\delta \log c^{-1}\delta^2)V_\theta^{-1/2}x}\{m \le N_2 < bm\} \to 0$ uniformly in $\theta \in A$ and $|x| \le (2\alpha/b)^{1/2} - \eta$ or $(2\alpha)^{1/2} + \eta \le |x| \le 4K^{1/2}/\lambda_{\min}(V_\theta^{-1/2})$. Hence, it follows from

(4.34) and (4.60)–(4.62) (where η can be arbitrarily small) that

$$\left(\int_{E_m} N_2^{\alpha+k/2} e^{-U}\, dQ_m\right) \Big/ \left\{ m^{\alpha+k/2} (\delta \log c^{-1}\delta^2)^k \det(V_\theta^{-1/2}) \right\}$$

$$\sim \int_{\substack{x \in \mathscr{C}_p(V_\theta^{1/2}(\lambda-\theta)):\\ 2\alpha/b \le |x|^2 \le 2\alpha}} (2\alpha/|x|^2)^{\alpha+k/2} \exp\{-2\alpha|x|^{-1}\varphi(x, V_\theta^{1/2}(\lambda-\theta))\}\, dx$$

$$(4.63) \quad = \int_{r=(2\alpha/b)^{1/2}}^{(2\alpha)^{1/2}} \int \cdots \int_{\substack{0 \le \omega_1 \le \cos^{-1} p;\ 0 \le \omega_{k-1} \le 2\pi\\ 0 \le \omega_2, \dots, \omega_{k-2} \le \pi}} \left\{ (2\alpha/r^2)^{\alpha+k/2} \times e^{-(2\alpha/r)\cos\omega_1} r^{k-1} \prod_{i=1}^{k-2} (\sin\omega_i)^{k-i-1} \right\} d\omega_1 \cdots d\omega_{k-1}\, dr$$

$$= \frac{2\pi^{(k-1)/2}}{\Gamma((k-1)/2)} \int_0^{\cos^{-1} p} \int_{(2\alpha/b)^{1/2}}^{(2\alpha)^{1/2}} (2\alpha)^{\alpha+k/2} r^{-2\alpha-1} e^{-(2\alpha/r)\cos\omega_1} \times (\sin\omega_1)^{k-2}\, dr\, d\omega_1,$$

since $\int_0^\pi (\sin\omega)^j\, d\omega = \pi^{1/2}\Gamma(\frac{1}{2}(j+1))/\Gamma(\frac{1}{2}j+1)$. Using a change of variables $t = \cos\omega_1$, $s = 2\alpha t/r$ in the above integral, we can apply (4.59) and (4.63) to evaluate the component $P_\theta\{m \le N_2 < bm\}$ of (4.58). This and a similar analysis of the other component $P_\theta\{b^{-1}m \le N_2 < m\}$ then complete the proof of the desired conclusion (4.58). □

REFERENCES

ANSCOMBE, F. J. (1963). Sequential medical trials. *J. Amer. Statist. Assoc.* **58** 365–383.

CHERNOFF, H. (1961). Sequential tests for the mean of a normal distribution. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 79–91. Univ. California Press.

CHERNOFF, H. (1965). Sequential test for the mean of a normal distribution III (small t). *Ann. Math. Statist.* **36** 28–54.

CHERNOFF, H. and PETKAU, J. (1981). Sequential medical trials involving paired data. *Biometrika* **68** 119–132.

CHOW, Y. S. and LAI, T. L. (1975). Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. *Trans. Amer. Math. Soc.* **208** 51–72.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

JENNEN, C. and LERCHE, H. R. (1981). First exit densities of Brownian motion through one-sided moving boundaries. *Z. Wahrsch. verw. Gebiete* **55** 133–148.

LAI, T. L. (1987a). Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* To appear.

LAI, T. L. (1987b). Adaptive treatment allocation and the multi-armed bandit problem. *Ann. Statist.* **15** 1091–1114.

LAI, T. L., ROBBINS, H. and SIEGMUND, D. (1983). Sequential design of comparative clinical trials. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday* (M. H. Rizvi, J. S. Rustagi and D. Siegmund, eds.) 51–68. Academic, New York.

SPIVAK, M. (1965). *Calculus on Manifolds.* Benjamin, New York.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305