Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

NEARLY OPTIMAL GENERALIZED SEQUENTIAL LIKELIHOOD RATIO TESTS IN MULTIVARIATE EXPONENTIAL FAMILIES*

BY TZE LEUNG LAI AND LI MIN ZHANG
Stanford University

A simple class of generalized sequential likelihood ratio tests is introduced for testing hypotheses in multivariate exponential families. These sequential tests have asymptotically optimal frequentist properties and also provide approximate Bayes solutions with respect to a large class of prior distributions.

1. Introduction. Let $X_1, X_2, \cdots$ be i.i.d. $p \times 1$ random vectors whose common multivariate density (with respect to some nondegenerate dominating measure ν) belongs to the exponential family

$$f_\theta(x) = \exp\{\theta' x - \psi(\theta)\}. \tag{1.1}$$

Thus, $E_\theta X = \nabla\psi(\theta)$, $\mathrm{Cov}_\theta X = \nabla^2\psi(\theta)$, and the Kullback-Leibler information number is given by

$$I(\theta, \lambda) = E_\theta \log\{f_\theta(X)/f_\lambda(X)\} = (\theta - \lambda)'\nabla\psi(\theta) - (\psi(\theta) - \psi(\lambda)). \tag{1.2}$$

Let $S_n = X_1 + \cdots + X_n$, $\bar{X}_n = S_n/n$ and let $\Theta = \{\theta \in \boldsymbol{R}^p : \int \exp(\theta' x) d\nu(x) < \infty\}$ be the natural parameter space. Consider the problem of testing sequentially $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, where Θ_0, Θ_1 are disjoint subsets of Θ such that

$$\Delta = \inf\left\{||\lambda - \theta|| : \theta \in \Theta_0, \lambda \in \Theta_1\right\} > 0\ , \qquad (||\theta|| = \sqrt{\theta'\theta}). \tag{1.3}$$

Let g be a nonnegative function on $(0, \infty)$ such that for some $\xi \in \boldsymbol{R}$,

$$g(t) \sim \log t^{-1} \quad \text{and} \quad g(t) \geq \log t^{-1} + \xi \log\log t^{-1} \quad \text{as} \quad t \to 0. \tag{1.4}$$

---

* This research was supported by the National Science Foundation, the National Security Agency, and the Air Force Office of Scientific Research.

AMS 1980 Subject Classifications: Primary 62H15, 62L10.

Key words and phrases: Bayes risk efficiency, curved exponential families, generalized sequential likelihood ratio tests, Kullback-Leibler information, nonlinear multivariate transformations.

As in Lai (1988b), we shall restrict θ to a convex subset A of Θ such that

$$\inf_{\theta\in A_\rho} \lambda_{\min}(\nabla^2\psi(\theta)) > 0, \qquad \sup_{\theta\in A_\rho} \lambda_{\max}(\nabla^2\psi(\theta)) < \infty,$$
$$\text{and } \nabla^2\psi \text{ is uniformly continuous on } A_\rho \text{ for some } \rho > 0, \tag{1.5}$$

where $A_\rho = \{\lambda \in \boldsymbol{R}^p : \inf_{\theta\in A} \|\theta-\lambda\| < \rho\}$ and $\lambda_{\min}, \lambda_{\max}$ denote the minimum and maximum eigenvalues of a symmetric matrix. Define the stopping rule

$$\begin{aligned} N(g,c) = \inf\Big\{n \geq 1 : \widehat{\theta}_n \in A_\rho, \max_{j=0,1}\Big(\sum_{i=1}^n \log f_{\widehat{\theta}_n}(X_i) \\ - \sup_{\theta\in\Theta_j}\sum_{i=1}^n \log f_\theta(X_i)\Big) \geq g(cn)\Big\}, \end{aligned} \tag{1.6}$$

where $\widehat{\theta}_n$ is the maximum likelihood estimator that maximizes $\sum_{i=1}^n \log f_\theta(X_i)$ $(= n(\theta'\bar{X}_n - \psi(\theta))$ over $\theta \in \Theta$. Noting that $\widehat{\theta}_n = (\nabla\psi)^{-1}(\bar{X}_n)$ if $\bar{X}_n \in \nabla\psi(\Theta)$, we can express the statistics in (1.6) as

$$\begin{aligned} \ell_{n,j} &= \sum_{i=1}^n \log f_{\widehat{\theta}_n}(X_i) - \sup_{\theta\in\Theta_j}\sum_{i=1}^n \log f_\theta(X_i) \\ &= n\left\{(\widehat{\theta}_n'\bar{X}_n - \psi(\widehat{\theta}_n)) - \sup_{\theta\in\Theta_j}(\theta'\bar{X}_n - \psi(\theta))\right\} \\ &= \inf_{\theta\in\Theta_j} nI(\widehat{\theta}_n,\theta), \end{aligned} \tag{1.7}$$

in view of (1.2), at least when $\widehat{\theta}_n \in A_\rho$. When stopping occurs at stage n, we use the terminal decision rule δ^* that rejects H_1 or H_0 according as $\ell_{n,1} \geq \ell_{n,0}$ or $\ell_{n,1} < \ell_{n,0}$. This is a multivariate extension of the generalized sequential likelihood ratio test (GSLRT) proposed by Lai (1988a) in the univariate case $p = 1$, and Lai and Zhang (1993) showed that in the multivariate case such GSLRT has the following asymptotically optimal frequentist and Bayesian properties for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$.

THEOREM 1. *For the test $(N(g,c),\delta^*)$, in which g satisfies (1.4) for some $\xi \in \boldsymbol{R}$, let $\alpha_j = \sup_{\theta\in\Theta_j} P_\theta\{(N(g,c),\delta^*)$ rejects $H_j\}$ $(j = 0,1)$. Let $\mathcal{T}(\alpha_0,\alpha_1)$ be the class of all sequential tests (T,δ) such that $\sup_{\theta\in\Theta_j} P_\theta\{(T,\delta)$ rejects $H_j\} \leq \alpha_j$ for $j = 0,1$. Let Δ be the distance between Θ_0 and Θ_1 as specified in (1.3) and define*

$$J(\theta) = \max\left\{\inf_{\lambda\in\Theta_0} I(\theta,\lambda), \inf_{\gamma\in\Theta_1} I(\theta,\gamma)\right\}. \tag{1.8}$$

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

$$E_\theta N(g,c) \sim |\log c|/J(\theta) \sim \inf_{(T,\delta)\in\mathcal{T}(\alpha_0,\alpha_1)} E_\theta T \quad \text{as} \quad c \to 0,$$

uniformly in $\theta \in A$ *with* $J(\theta) \le D_c$, *for any positive numbers* $D_c \to \infty$ *such that* $D_c = o(\log c)$ *as* $c \to 0$.

(ii)*As* $c \to 0$ *and* $\Delta \to 0$ *such that* $\Delta^2/c \to \infty$,

$$\sup_{\theta \in A} E_\theta N(g,c) \sim \inf_{(T,\delta)\in \mathcal{T}(\alpha_0,\alpha_1)} \sup_{\theta\in A} E_\theta T \sim \left\{\sup_{\theta\in A}(J(\theta))^{-1}\right\} \log(\Delta^2/c).$$

(iii) *Let* G *be a probability distribution on* A. *Let* $r(T,\delta)$ *be the Bayes risk*

$$\begin{aligned} r(T,\delta) =& c\int_\Theta E_\theta T dG + \int_{\Theta_0} \ell(\theta) P_\theta\{Reject\, H_0\}\, dG \\ &+ \int_{\Theta_1} \ell(\theta) P_\theta\{Reject\, H_1\}\, dG, \end{aligned} \tag{1.9}$$

of a test (T,δ) *of* H_0 *versus* H_1. *Suppose that the loss function* ℓ *in* (1.9) *for wrongly rejecting the true hypothesis satisfies* $\sup_{\theta\in\Theta_0\cup\Theta_1} \ell(\theta) < \infty$ *and* $\inf_{\theta\in\Theta_0\cup\Theta_1} \ell(\theta) > 0$, *that* $G(S\cap\Theta_j) > 0$ *for every p-dimensional ball* S *with center belonging to* Θ_j, $j = 0,1$, *and that* $\xi > p/2$ *in* (1.4). *Then as* $c \to 0$,

$$r(N(g,c),\delta^*) \sim c|\log c| \int_A (J(\theta))^{-1} dG(\theta) \sim \inf_{(T,\delta)} r(T,\delta).$$

Theorem 1(iii) shows that $(N(g,c),\delta^*)$ is asymptotically Bayes risk efficient as $c \to 0$ for fixed $\Delta > 0$. In Sections 2 and 3, we shall show that $(N(g,c),\delta^*)$ is still asymptotically Bayes risk efficient as $\Delta \to 0$ when H_0, H_1 are one-sided hypotheses about some real-valued function of θ, and we shall also extend this kind of tests to the case when there is no indifference zone, generalizing Lai's (1988a) theory of nearly optimal sequential tests in univariate exponential families. The derivation of these results in Section 3 uses transformation techniques in multivariate analysis and certain geometric properties of multivariate exponential families.

2. Asymptotically Bayes risk efficient GSLRT when $0A\Delta \to$ **o as** $\mathbf{c} \to \mathbf{0}$ **or when there is no indifference zone.** In this section we consider the Bayes problem of minimizing the Bayes risk (1.9), in which $\ell(\theta) = 1$ for $\theta \in \Theta_0 \cup \Theta_1 \subset A$ (the $0-1$ loss), G is a prior distribution on A and H_0, H_1 are one-sided hypotheses about some real-valued function $z(\theta)$ of the parameter vector θ. Let $z : A_\rho \to \boldsymbol{R}$ and $y : A_\rho \to \boldsymbol{R}^{p-1}$ be continuously differentiable functions such that

$$\zeta : A_\rho \to \boldsymbol{R}^p \text{ is one-to-one, where } \zeta(\theta) = \begin{pmatrix} z(\theta) \\ y(\theta)\end{pmatrix}, \text{ and}$$

$$\begin{aligned} &\sup_{\theta\in A_\rho} \lambda_{\max}\left\{\left(\frac{\partial\zeta}{\partial\theta}\right)\left(\frac{\partial\zeta}{\partial\theta}\right)'\right\} < \infty, \\ &\inf_{\theta\in A_\rho} \lambda_{\min}\left\{\left(\frac{\partial\zeta}{\partial\theta}\right)\left(\frac{\partial\zeta}{\partial\theta}\right)'\right\} > 0. \end{aligned} \tag{2.1}$$

The notation $\partial\zeta/\partial\theta$ is used to denote the Jacobian matrix $(\partial\zeta_i/\partial\theta_j)_{1\le i,j\le p}$. In view of (2.1), the restriction of ζ to A is a diffeomorphism from A onto $\zeta(A)$ and therefore we can regard ζ as a reparameterization of A in lieu of θ. The one-sided hypotheses H_0, H_1 can be conveniently stated in terms of this reparameterization as $H_0 : z \le z_0$ versus $H_1 : z \ge z_0 + \epsilon u(y)$ (with an indifference zone whose width may depend on y) or $H_0 : z < z_0$ versus $H_1 : z > z_0$ (without an indifference zone), with the component vector y of ζ treated as a nuisance parameter.

To test $H_0 : z \le z_0$ versus $H_1 : z \ge z_0 + \varepsilon u(y)$ with cost c per observation and the $0-1$ loss, we again use the GSLRT with stopping rule $N(g, c)$ and terminal decision rule δ^*. In Theorem 1(iii) dealing with the case of a fixed distance Δ between Θ_0 and Θ_1, g is assumed to satisfy (1.4) with $\xi > p/2$. This condition on g still suffices for the Bayes risk efficiency of $(N(g,c),\delta^*)$ when $\varepsilon \to 0$ as $c \to 0$ such that $\varepsilon^2/c \to \infty$, as will be shown in Theorem 2 below. However, in the case of no indifference zone or in the case $\varepsilon^2/c \to \gamma$ (finite), we require a particular choice of g which agrees with the stopping boundary for a continuous-time optimal stopping problem that arises from Wiener process approximations to random walks.

Let $w_\eta(t), t \ge 0$, be a Wiener process with $E(w_\eta(t)) = \eta t$ and $\mathrm{Var}(w_\eta(t)) = t$. Lai (1988a) studied the problems of testing $H : \eta \le -\gamma$ versus $K : \eta \ge \gamma$ and $H' : \eta < 0$ versus $K' : \eta > 0$, with the $0-1$ loss and a cost of t for observing the process for a period of length t, assuming a flat prior (i.e., Lebesgue measure) on $\eta \in \boldsymbol{R}$. Given $\gamma \ge 0$ (the case $\gamma = 0$ corresponds to H' versus K'), the optimal stopping rule is of the form $\tau_{\gamma,\eta} = \inf\{t > 0 : |w_\eta(t)| \ge h_\gamma(t)\}$, and the terminal decision rule is to accept H (or H') iff $w_\eta(\tau_\gamma) < 0$. Lai (1988a) computed numerically the boundaries $h_\gamma(t)$ for certain values of γ, and used these numerical results and an asymptotic analysis of the free boundary problem for the heat equation associated with the optimal stopping problem to derive simple closed-form approximations to $h_\gamma(t)$ for all $\gamma \ge 0$ and $t > 0$. He also showed that the optimal Bayes risk is finite, i.e., for $0 \le \gamma < \infty$,

$$\begin{aligned}\infty > b(\gamma) := \int_{-\infty}^{\infty} E(\tau_{\gamma,\eta})d\eta + \int_{\gamma}^{\infty} P(w_\eta(\tau_{\gamma,\eta}) < 0)\, d\eta \\ + \int_{-\infty}^{-\gamma} P(w_\eta(\tau_{\gamma,\eta}) > 0)\, d\eta.\end{aligned} \tag{2.2}$$

For the problem of testing $H_0 : \theta < 0$ versus $H_1 : \theta > 0$ (or $H_0 : \theta \le -\Delta/2$ versus $H_1 : \theta \ge \Delta/2$) for the mean θ of a univariate normal population with known variance 1, let

$$\eta = c^{-1/2}\theta, \qquad \gamma = c^{-1/2}\Delta/2, \qquad t = cn, \qquad w_\eta(t) = \sqrt{c}S_n. \tag{2.3}$$

Since $\sqrt{c}\theta n = \eta t$, $w_\eta(t)$ is a Wiener process with drift coefficient η and with t restricted to the set $\{c, 2c, \cdots\}$, which becomes dense in $[0,\infty)$ as $c \to 0$. For

$\gamma \geq 0$, let

$$g_\gamma(t) = (h_\gamma(t) + \gamma t)^2/2t. \tag{2.4}$$

Since $I(\theta, \lambda) = (\theta - \lambda)^2/2$ and $\widehat{\theta}_n = \bar{X}_n$ for the normal distribution, it follows from (2.3) and (1.7) that for the above hypotheses on a normal mean,

$$|w_\eta(t)| \geq h_\gamma(t) \Longleftrightarrow (|S_n| + \Delta n/2)^2/2n \geq g_\gamma(cn) \Longleftrightarrow \max_{j=0,1} \ell_{n,j} \geq g_\gamma(cn).$$

Theorem 3 below shows that the GSLRT with stopping rule $T(g_0, c)$ defined in (2.7) is asymptotically Bayes risk efficient as $c \to 0$, not just for testing the hypotheses $H_0 : \theta < 0$ versus $H_1 : \theta > 0$ for the mean θ of a univariate normal distribution, but much more generally for testing $H_0 : z(\theta) < z_0$ versus $H_1 : z(\theta) > z_0$ for real-valued functions $z(\theta)$ of the parameter vector θ of the multivariate exponential family (1.1). Theorem 2(ii) proves an analogous result for the test $(N(g_\gamma, c), \delta^*)$ of $H_0 : z \leq z_0$ versus $H_1 : z \geq z_0 + \varepsilon\sigma(y)$ as $c \to 0$ and $\varepsilon \to 0$ such that $c^{-1/2}\varepsilon/2 \to \gamma$, where $\sigma(y)$ is defined in (2.6) below. In the univariate exponential family with $z(\theta) = \theta$ and $z_0 = \theta_0, \sigma(y)$ reduces to $(d^2\psi(\theta)/d\theta^2|_{\theta=\theta_0})^{-1/2}$. This factor of ε in H_1 was inadvertently omitted in the statement (but not the proof) of Theorem 1(iii) of Lai (1988a). While Theorems 2 and 3 focus on the $0-1$ loss, their proofs and results can be extended to more general loss functions of the form $\ell(\theta) = \beta \,|\, z(\theta) - z_0|^\alpha$ $(\alpha > -1, \beta > 0)$, using the ideas of Lai (1988c) in the case of univariate exponential families with $z(\theta) = \theta$.

Since ζ defines a reparameterization of $\theta \in A$, we can express a prior distribution G of θ with support in A as a prior distribution of $\zeta(= (z, y))$. We shall assume that for some $d > 0$, $[z_0 - d, z_0 + d] \subset z(A)$ and that the prior distribution G, as a distribution of (z, y), satisfies

G has density function π_G with respect to Lebesgue measure in the region $[z_0 - d, z_0 + d] \times y(A)(\subset \zeta(A))$, (2.5a)

$$\pi_G(z, y) \to \pi_G(z_0, y) \text{ as } z \to z_0, \text{ uniformly in } y \in y(A), \tag{2.5b}$$

$\pi_G(z_0, y)$ is continuous in $y \in y(A)$ and

$$0 < \int_{y(A)} \pi_G(z_0, y)dy < \infty. \tag{2.5c}$$

THEOREM 2. *Let G be a probability distribution on A satisfying (2.5a)–(2.5c). Let $r(T, \delta)$ be the Bayes risk (1.9) of a test (T, δ) of $H_0 : z \leq z_0$ versus $H_1 : z \geq z_0 + \varepsilon u(y)$ with the $0-1$ loss and cost c per observation, where $\zeta = (z, y)$ is a reparameterization satisfying (2.1) and u is a real-valued function on $y(A)$ such that $\sup_{y \in y(A)} u(y) < \infty$ and $\inf_{y \in y(A)} u(y) > 0$. For $y \in y(A)$, define*

$$\sigma(y) = \|(\nabla^2\psi(\theta))^{-1/2}\nabla z(\theta)\|_{z(\theta)=z_0, y(\theta)=y}. \tag{2.6}$$

Then $\sup_{y\in y(A)} \sigma(y) < \infty$ *and* $\inf_{y\in y(A)} \sigma(y) > 0$.

(i) *Let* g *be a nonnegative function on* $(0,\infty)$ *satisfying* (1.4) *for some* $\xi > p/2$. *Then as* $c \to 0$ *and* $\varepsilon \to 0$ *such that* $\varepsilon^2/c \to \infty$,

$$r(N(g,c),\delta^*) \sim \inf_{T,\delta} r(T,\delta) \sim \left\{4\int_{y(A)} (\sigma^2(y)/u(y))\pi_G(z_0,y)dy\right\} c\varepsilon^{-1}\log(\varepsilon^2/c).$$

(ii) *Suppose that* $u(\cdot) = \sigma(\cdot)$. *Let* $0 \le \gamma < \infty$ *and define* $b(\gamma)$ *and* g_γ *by* (2.2) *and* (2.4). *Then as* $c \to 0$ *and* $\varepsilon \to 0$ *such that* $\frac{1}{2}\varepsilon/\sqrt{c} \to \gamma$,

$$r(N(g_\gamma,c),\delta^*) \sim \inf_{T,\delta} r(T,\delta) \sim \sqrt{c}\, b(\gamma)\int_{y(A)} \sigma(y)\pi_G(z_0,y)\,dy.$$

THEOREM 3. *Suppose that* G *is a probability distribution on* A *satisfying* (2.5a)–(2.5c) *and that* $\zeta = (z,y)$ *is a reparameterization satisfying* (2.1). *Let* $r(T,\delta)$ *be the Bayes risk* (1.9) *of a test* (T,δ) *of* $H_0 : z < z_0$ *versus* $H_1 : z > z_0$ *with the* $0-1$ *loss and cost* c *per observation. Define* g_0 *by* (2.4) *and*

$$T(g_0,c) = \inf\Big\{n \ge 1 : \widehat{\theta}_n \in A_\rho \text{ and } \sum_{i=1}^n \log f_{\widehat{\theta}_n}(X_i) - \sup_{\theta\in A, z(\theta)=z_0} \sum_{i=1}^n \log f_\theta(X_i) \ge g_0(cn)\Big\}. \tag{2.7}$$

Let $\widetilde{\delta}$ *be the terminal decision rule that accepts* H_0 *iff* $z(\widehat{\theta}_n) < z_0$ *when stopping occurs at stage* n. *Then as* $c \to 0$,

$$r(T(g_0,c),\widetilde{\delta}) \sim \inf_{T,\delta} r(T,\delta) \sim \sqrt{c}\, b(0)\int_{y(A)} \sigma(y)\pi_G(z_0,y)\,dy,$$

where $b(\cdot)$ *is defined in* (2.2) *and* $\sigma(\cdot)$ *is defined in* (2.6).

3. Proof of Theorems 2 and 3. For $x \in \boldsymbol{R}^p$, define

$$\begin{aligned} \phi(x) &= \sup_{\theta\in\Theta}(\theta' x - \psi(\theta)), \\ \phi_0(x) &= \sup_{\theta\in A: z(\theta)=z_0} (\theta' x - \psi(\theta)), \\ L(x) &= \phi(x) - \phi_0(x). \end{aligned} \tag{3.1}$$

In view of (2.1), given any $(z,y')' \in \zeta(A)$, there exists a unique θ such that $\zeta(\theta) = (z,y')'$. This θ will be denoted by $\theta_{y,z}$. Since $I(\theta,\lambda) = \int_0^1 (1-t)(\lambda - \theta)'\{\nabla^2\psi(t\lambda + (1-t)\theta)\}(\lambda-\theta)dt$ by (1.2), it follows that

$$b\|\theta-\lambda\|^2 \le I(\theta,\lambda) \le B\|\theta-\lambda\|^2 \qquad \text{for all} \quad \theta,\ \lambda \in A_\rho, \tag{3.2}$$

where

$$b = \frac{1}{2} \inf_{\theta \in A_\rho} \lambda_{\min}(\nabla^2 \psi(\theta)), \qquad B = \sup_{\theta \in A_\rho} \lambda_{\max}(\nabla^2 \psi(\theta)).$$

LEMMA 1. *For $y \in y(A)$, let $\mu_y = \nabla \psi(\theta_{y,z_0})$, $\Sigma_y = \nabla^2 \psi(\theta_{y,z_0})$, $J_y = \partial\theta/\partial\zeta|_{\theta=\theta_{y,z_0}}$. Then*

$$J_y' \Sigma_y (\nabla^2 L(\mu_y)) \Sigma_y J_y = \begin{pmatrix} 1/\sigma^2(y) & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.3}$$

where $\sigma(y)$ is defined in (2.6). Consequently, uniformly in $y \in y(A)$, as $x \to \mu_y$ with $x \in A_\rho$,

$$L(x) = \{(\nabla z(\theta_{y,z_0}))' \Sigma_y^{-1} (x - \mu_y)\}^2 / (2\sigma^2(y)) + o(\|x - \mu_y\|^2). \tag{3.4}$$

PROOF. Let $U_y = \partial\theta/\partial y|_{\theta=\theta_{y,z_0}}$. Using Taylor expansions and the inverse function theorem, it can be shown that

$$\nabla^2 \phi_0(\mu_y) = U_y (U_y' \Sigma_y U_y)^{-1} U_y', \tag{3.5}$$

cf. Lemma 3.2 of Zhang (1992). In the remainder of the proof we shall fix y and denote Σ_y, U_y, J_y simply by Σ, U, J. Letting I_q denote the $q \times q$ identity matrix, note that

$$U = J \begin{pmatrix} 0 \\ I_{p-1} \end{pmatrix}, \qquad U' \Sigma U = \Gamma_{22}, \tag{3.6}$$

where

$$J' \Sigma J = \Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}.$$

We first show that the $(1,1)$ element of Γ^{-1} is $\sigma^2(y)$. Since Γ^{-1} is the covariance matrix of $J^{-1}\Sigma^{-1}X_1$, its $(1,1)$ element is the variance of the first component of $J^{-1}\Sigma^{-1}X_1$. Since $J^{-1} = \partial\zeta/\partial\theta|_{\theta=\theta_{y,z_0}}$ by the inverse function theorem, the first component of $J^{-1}\Sigma^{-1}X_1$ is $(\nabla z(\theta_{y,z_0}))'\Sigma^{-1}X_1$, which has variance $\sigma^2(y)$. By (3.5) and (3.6),

$$\begin{aligned} J' \Sigma (\nabla^2 \phi_0(\mu_y)) \Sigma J &= \begin{pmatrix} \Gamma_{12} \\ \Gamma_{22} \end{pmatrix} \Gamma_{22}^{-1} (\Gamma_{21} \ \Gamma_{22}) \\ &= \begin{pmatrix} \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}. \end{aligned} \tag{3.7}$$

Since $\nabla^2 \phi(\mu_j) = \Sigma^{-1}$ and $L = \phi - \phi_0$, (3.3) follows from (3.6) and (3.7), noting that $(\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21})^{-1}$ is the $(1,1)$ element of Γ^{-1}, cf. Rao (1973).

Since $L(\mu_y) = 0$ and $\nabla L(\mu_y) = 0$ by Lemma 3.1 of Zhang (1992), it follows from (3.3) and Taylor's expansion around μ_y that as $x \to \mu_y$ with $x \in A_\rho$,

$$\begin{aligned}
L(x) =&(x-\mu_y)'\nabla L^2(\mu_y)(x-\mu_y)/2 + o(\|x-\mu_y\|^2)\\
=&\{(\Sigma J)^{-1}(x-\mu_y)\}'\{J'\Sigma(\nabla L^2(\mu_y))\Sigma J\}\{(\Sigma J)^{-1}(x-\mu_y)\}/2\\
&+ o(\|x-\mu_y\|^2)\\
=&\{\text{first component of } J^{-1}\Sigma^{-1}(x-\mu_y)\}^2/(2\sigma^2(y))\\
&+ o(\|x-\mu_y\|^2).
\end{aligned} \tag{3.8}$$

Let $\widetilde{x} = \Sigma^{-1}(x-\mu_y)$. Since $J^{-1} = \partial\zeta/\partial\theta|_{\theta=\theta_{y,z_0}}$, it follows that the first component of the vector $J^{-1}\widetilde{x}$ is simply $(\nabla z(\theta_{y,z_0}))'\widetilde{x}$. Hence (3.4) follows from (3.8).

LEMMA 2. *With the same notation as in Lemma 1, define $v_y = (\sigma(y))^{-1}\ \Sigma_y^{-1/2}\nabla z(\theta_{y,z_0})$ and let Z_y be a $p\times(p-1)$ matrix whose column vectors are orthonormal and are orthogonal to v_y. Let $S_n = X_1 + \cdots + X_n$. For $c > 0$ and $y \in y(A)$, define*

$$W_{c,y}(t) = \sqrt{c}\begin{pmatrix} v_y' \\ Z_y' \end{pmatrix}\Sigma_y^{-1/2}(S_n - n\mu_y) \quad \text{if} \quad t = cn \quad (n = 1,2,\cdots), \tag{3.9}$$

and define $W_{c,y}(t)$ by linear interpolation for $cn < t < c(n+1)$. Let $\{w_\eta(t), t \geq 0\}$ denote a one-dimensional Wiener process with drift coefficient η and let $\{B(t), t \geq 0\}$ be a $1 \times (p-1)$ vector Brownian motion, with $EB(t) = 0$ and $\operatorname{Cov} B(t) = tI_{p-1}$, that is independent of $\{w_\eta(t), t \geq 0\}$. Then for every $T > 0$ and $M > 0$, the process $\{W_{c,y}(t), 0 \leq t \leq T\}$ converges weakly to $\{(w_\eta(t), B(t))', 0 \leq t \leq T\}$ under P_θ with $\nabla\psi(\theta) = \mu_y + (\sqrt{c}\eta/\sigma(y))\nabla z(\theta_{y,z_0})$, the convergence being uniform in $-M \leq \eta \leq M$ and $y \in y(A)$.

PROOF. First note that for $t = cn$,

$$\begin{aligned}
\operatorname{Cov}_\theta(W_{c,y}(t)) &= cn\begin{pmatrix} v_y' \\ Z_y' \end{pmatrix}(v_y, Z_y) = tI_p,\\
E_\theta(W_{c,y}(t)) &= cn\begin{pmatrix} v_y' \\ Z_y' \end{pmatrix}\Sigma_y^{-1/2}\eta\nabla z(\theta_{y,z_0})/\sigma(y) = t\begin{pmatrix} \eta \\ 0 \end{pmatrix},
\end{aligned}$$

since $Z_y'v_y = 0$ by definition and $v_y'\Sigma_y^{-1/2}\nabla z(\theta_{y,z_0})/\sigma(y) = v_y'v_y = 1$. The desired conclusion then follows by an argument similar to that used in the proof of Lemma 4 of Lai (1988a).

LEMMA 3. (i) *For* $\Theta_0 = \{\theta \in A : z(\theta) < z_0\}$ *and* $\Theta_1 = \{\theta \in A : z(\theta) > z_0\}$, $\inf_{\lambda\in\Theta_j} I(\theta,\lambda) = \inf_{\lambda\in A:z(\lambda)=z_0} I(\theta,\lambda)$ *for any* $\theta \in A_\rho - \Theta_j (j = 0,1)$.

(ii) *For* $\Theta_0 = \{\theta \in A : z(\theta) \leq z_0\}$ *and* $\Theta_1 = \{\theta \in A : z(\theta) \geq z_0 + \varepsilon u(y(\theta))\}$, $\inf_{\lambda \in \Theta_0} I(\theta, \lambda) = \inf_{\lambda \in A: z(\lambda) = z_0} I(\theta, \lambda)$ *for* $\theta \in A_\rho - \Theta_0$, *and* $\inf_{\lambda \in \Theta_1} I(\theta, \lambda) = \inf_{\lambda \in A: z(\lambda) = z_0 + \varepsilon u(y(\lambda))} I(\theta, \lambda)$ *for* $\theta \in A_\rho - \Theta_1$.

PROOF. For $\Theta_0 = \{\theta \in A : z(\theta) < z_0\}$, suppose that for some $\theta \in A_\rho - \Theta_0, \inf_{\lambda \in \Theta_0} I(\theta, \lambda) < \inf_{\lambda \in A: z(\lambda) = z_0} I(\theta, \lambda)$. Then there exist $\lambda_1 \in \Theta_0 (\subset A)$ and $0 < t_1 < 1$ such that $\{t\theta + (1-t)\lambda_1 : 0 \leq t \leq t_1\} \subset \Theta_0$ and $I(\theta, \lambda_1) = \min_{0 \leq t \leq t_1} I(\theta, t\theta + (1-t)\lambda_1)$, recalling that A is convex and that z is continuous with $z(\lambda_1) < z_0$. Using (1.2) and a differentiation argument, it can be shown that $I(\theta, t\theta + (1-t)\lambda_1)$ is a decreasing function of $t \in [0,1]$, which contradicts that $I(\theta, \lambda_1) = \min_{0 \leq t \leq t_1} I(\theta, t\theta + (1-t)\lambda_1)$. Similarly we can prove the other assertions of the lemma.

PROOF OF THEOREM 3. Let $\mu(y,z) = \nabla\psi(\theta_{y,z}) = E_{\theta_{y,z}}(X_1)$. We shall use the change of variables $\mathcal{Y} = \mathcal{Y}(y,z) \in \boldsymbol{R}^{p-1}$, $\eta = \eta(y,z) \in \boldsymbol{R}$ defined by

$$\begin{aligned} &\mu(y,z) = \mu(\mathcal{Y}, z_0) + \sqrt{c}\,\eta \nabla z(\theta_{\mathcal{Y},z_0})/\sigma(\mathcal{Y}), \quad \text{or equivalently,} \\ &\begin{pmatrix} z \\ y \end{pmatrix} = \zeta\left((\nabla\psi)^{-1}\left\{ \nabla\psi(\theta_{\mathcal{Y},z_0}) + \frac{\sqrt{c}\,\eta}{\sigma(\mathcal{Y})} \nabla z(\theta_{\mathcal{Y},z_0}) \right\} \right), \end{aligned} \tag{3.10}$$

where

$$\zeta(\theta) = \begin{pmatrix} z(\theta) \\ y(\theta) \end{pmatrix}.$$

Let $M \geq 2$. From (3.10), (3.2) and (2.1), it follows that uniformly in $\mathcal{Y} \in y(a)$ and $|\eta| \leq M$,

$$\begin{aligned} z =& z_0 + (\sqrt{c}\,\eta/\sigma(\mathcal{Y}))(\nabla z(\theta_{\mathcal{Y},z_0}))' \Sigma_{\mathcal{Y}}^{-1} \nabla z(\theta_{\mathcal{Y},z_0}) + O(c) \\ =& z_0 + \sqrt{c}\,\eta\sigma(\mathcal{Y})(1 + O(\sqrt{c})), \end{aligned} \tag{3.11}$$

$$y = \mathcal{Y} + O(\sqrt{c}), \quad \partial(z,y)/\partial(\eta,\mathcal{Y}) = \sqrt{c}\,\sigma(\mathcal{Y})(1 + O\sqrt{c}). \tag{3.12}$$

For any sequential test (T,δ), define its risk function $R_{T,\delta}(z,y)$ by

$$\begin{aligned} R_{T,\delta}(z,y) =& cE_\theta T + P_\theta\{(T,\delta) \text{ accepts } H_1\} \quad \text{if} \quad z < z_0, \\ =& cE_\theta T + P_\theta\{(T,\delta) \text{ accepts } H_0\} \quad \text{if} \quad z > z_0, \end{aligned} \tag{3.13}$$

where

$$\zeta(\theta) = (z, y')'.$$

Consider the risk function of the test $(T(g_0, c), \widetilde{\delta})$. By Lemma 1,

$$nL(\bar{X}_n) = \left\{ \sqrt{c}\, v_{\mathcal{Y}}' \Sigma_{\mathcal{Y}}^{-1/2} (S_n - n\mu_{\mathcal{Y}}) \right\}^2 / (2cn) + o(n \|\bar{X}_n - \mu_{\mathcal{Y}}\|^2),$$

where $v_{\mathcal{Y}}$ is defined in Lemma 2 and $(\mathcal{Y}, \eta)$ is defined from (y, z) via (3.10). Moreover,

$$\begin{aligned} z(\widehat{\theta}_n) - z_0 =& z\big((\nabla\psi)^{-1}(\bar{X}_n)\big) - z\big((\nabla\psi)^{-1}(\mu_{\mathcal{Y}})\big) \\ =& \big(\nabla z(\theta_{\mathcal{Y}, z_0})\big)' \Sigma_{\mathcal{Y}}^{-1}(\bar{X}_n - \mu_{\mathcal{Y}}) + O\big(\|\bar{X}_n - \mu_{\mathcal{Y}}\|^2\big). \end{aligned}$$

Hence by Lemma 2 together with (3.11), (3.12) and (2.5), as $c \to 0$,

$$\begin{aligned} &\int_{|z-z_0| \le M\sqrt{c}\,\sigma(y)} R_{T(g_0,c),\widetilde{\delta}}(z, y)\, dG \\ &= \int_{y \in y(A), |z-z_0| \le M\sqrt{c}\,\sigma(y)} R_{T(g_0,c),\widetilde{\delta}}(z, y) \pi_G(z, y)\, dz dy \\ &\sim \sqrt{c} \int_{\mathcal{Y} \in y(A)} \sigma(\mathcal{Y}) \pi_G(z_0, \mathcal{Y}) \Big\{ \int_{-M}^{M} E(\tau_{0,\eta})\, d\eta \\ &\quad + \int_{-M}^{0} P(w_\eta(\tau_{0,\eta}) > 0)\, d\eta + \int_{0}^{M} P(w_\eta(\tau_{0,\eta}) < 0)\, d\eta \Big\}\, d\mathcal{Y}, \end{aligned} \tag{3.14}$$

for every $M \ge 2$, where $\tau_{0,\eta} = \inf\{t > 0 : w_\eta^2(t)/2t \ge g_0(t)\} = \inf\{t > 0 : |w_\eta(t)| \ge h_0(t)\}$.

Let $H(\theta) = \inf_{\lambda \in A: z(\lambda) = z_0} I(\theta, \lambda)$. By an argument similar to the proof of Theorem 3 of Lai (1988b), it can be shown that

$$\begin{aligned} E_\theta T(g_0, c) =& O(\{\log(c^{-1}H(\theta))\}/H(\theta)) \text{ uniformly in} \\ & \theta \in A \quad \text{with} \quad 2c \le H(\theta) \le |\log c|^{3/4}. \end{aligned} \tag{3.15}$$

By (2.1), there exist $K > \kappa > 0$ such that

$$\kappa\|\theta - \lambda\| \le |z(\theta) - z(\lambda)| \le K\|\theta - \lambda\| \quad \text{for all} \quad \theta,\ \lambda \in A_\rho. \tag{3.16}$$

Let $\theta \in A$ be such that $z(\theta) = z_0 + s$ with $s \ne 0$. If $\lambda \in A_\rho$ is such that $z(\lambda) - z_0$ and s have different signs (i.e., $s(z(\lambda) - z_0) \le 0$), then $|s| \le |z(\theta) - z(\lambda)| \le K\|\theta - \lambda\|$ by (3.16). Hence by (3.2), $H(\theta) \ge bK^{-2}s^2$. Moreover, $H(\theta) \le B\kappa^{-2}s^2$ by (3.2) and (3.16). Therefore (3.15) yields

$$E_{\theta_{y,z}} T(g_0, c) = O((z - z_0)^{-2} \log((z - z_0)^2/c)) \tag{3.17}$$

uniformly in $\theta_{y,z} \in A$ with $Mc \le (z - z_0)^2 \le |\log c|^{2/3}$, for every sufficiently large M.

By Lemma 1 of Lai (1988a), g_0 satisfies (1.4) with $\xi = 1/2$. Let $\theta \in A$ be such that $z(\theta) = z_0 + s$ with $s > 0$. Then by the preceding argument, $\{\lambda \in A_\rho : z(\lambda) \le z_0\} \subset \{\lambda \in A_\rho : \|\theta - \lambda\| \ge K^{-1}s\}$. Therefore, if $\widehat{\theta}_n \in A_\rho$ and $z(\widehat{\theta}_n) \le z_0$, then $\inf_{\lambda \in A_\rho, \|\theta - \lambda\| \ge K^{-1}s} I(\widehat{\theta}_n, \lambda) = 0$. Moreover, $I(\widehat{\theta}_n, \theta) \ge \inf_{\lambda \in \Theta_1} I(\widehat{\theta}_n, \lambda) = \inf_{\lambda \in A: z(\lambda) = z_0} I(\widehat{\theta}_n, \lambda)$ by Lemma 3(i). Therefore by Lemma

2 of Lai and Zhang (1993) (with a slight modification of the statement but using the same proof), as $c \to 0$,

$$
\begin{aligned}
&P_{\theta_{y,z}}\{z(\widehat{\theta}_{T(g_0,c)}) \le z_0\} \\
=&O(c(z-z_0)^{-2}\{\log((z-z_0)^2/c)\}^{1+(p-1)/2}),
\end{aligned}
\tag{3.18}
$$

uniformly in $\theta_{y,z} \in A$ with $M\sqrt{c} \le z - z_0 \le |\log c|^{1/3}$, for every sufficiently large M. Clearly a similar result holds for $P_{\theta_{y,z}}\{z(\widehat{\theta}_{T(g_0,c)}) \ge z_0\}$ with $z < z_0$.

By choosing M arbitrarily large, it follows from (3.14), (3.17), (3.18) and (2.5) that $r(T(g_0,c),\widetilde{\delta}) \sim \sqrt{c}b(0)\int_{y(A)}\sigma(y)\pi_G(z_0,y)dy$. Note in this connection that $\sup_{\theta\in A_\rho}\|\nabla z(\theta)\| < \infty$ and $\inf_{\theta\in A_\rho}\|\nabla z(\theta)\| > 0$ by (2.1), and therefore (3.2) in turn implies that $\sup_{y\in y(A)}\sigma(y) < \infty$ and $\inf_{y\in y(A)}\sigma(y) > 0$.

To show that $\inf_{T,\delta} r(T,\delta) \sim \sqrt{c}\,b(0)\int_{y(A)}\sigma(y)\pi_G(z_0,y)dy$, take any $M > 1$ and note that by (2.5), (3.11) and (3.12),

$$
\begin{aligned}
\inf_{T,\delta} r(T,\delta) \ge&(1+o(1))\sqrt{c}\int_{\mathcal{Y}\in y(A)}\sigma(\mathcal{Y})\pi_G(z_0,\mathcal{Y}) \\
&\cdot\left\{\inf_{T,\delta}\int_{-M}^{M} R_{T,\delta}(z(\mathcal{Y},\eta),y(\mathcal{Y},\eta))d\eta\right\}d\mathcal{Y}.
\end{aligned}
\tag{3.19}
$$

By (3.13) and Lemma 2, uniformly in $\mathcal{Y} \in y(A)$,

$$
\begin{aligned}
&\inf_{T,\delta}\int_{-M}^{M} R_{T,\delta}(z(\mathcal{Y},\eta),y(\mathcal{Y},\eta))\,d\eta \\
\sim&\inf_{\tau}\left\{\int_{-M}^{M}E(\tau)d\eta + \int_0^M P(w_\eta(\tau)<0)\,d\eta\right. \\
&\left.+\int_{-M}^{0} P(w_\eta(\tau)>0)\,d\eta\right\},
\end{aligned}
\tag{3.20}
$$

noting that for any stopping time τ of the Wiener process $w_\eta(\cdot)$ with drift coefficient η, the Bayes terminal decision rule for testing $H' : \eta < 0$ versus $K' : \eta > 0$ with respect to $0-1$ loss and uniform prior distribution on $[-M, M]$ accepts H' and K' according as $w_\eta(\tau) < 0$ or $w_\eta(\tau) > 0$. Letting $M \to \infty$ in (3.19) and (3.20) and making use of (2.2), we obtain the desired conclusion on $\inf_{T,\delta} r(T,\delta)$.

Proof of Theorem 2(ii). Define ϕ, ϕ_0, L by (3.1) and let

$$
\begin{aligned}
\phi_\varepsilon(x) &= \sup_{\theta\in A: z(\theta)=z_\theta+\varepsilon u(y(\theta))}(\theta' x - \psi(\theta)), \\
L_\varepsilon(x) &= \max\{\phi(x)-\phi_0(x),\ \phi(x)-\phi_\varepsilon(x)\}.
\end{aligned}
\tag{3.21}
$$

Here $u(\cdot) = \sigma(\cdot)$. A simple extension of the argument used in the proof of Lemma 1 and (3.11), (3.12) can be used to show that uniformly in $y \in y(A)$, as $\varepsilon \to 0$ and $x \to \mu_y$ with $x \in A_\rho$,

$$\begin{aligned} L_\varepsilon(x) =& \left\{ |\nabla z(\theta_{y,z_0})'\Sigma_y^{-1}(x-\mu_y) + \varepsilon\sigma(y)/2| + \varepsilon\sigma(y)/2 \right\}^2 / \{2\sigma^2(y)\} \\ &+ o\big(\|x-\mu_y\|^2 + \varepsilon^2\big) \\ =& \left\{ |v_y'\Sigma_y^{-1/2}(x-\mu_y) + \varepsilon/2| + \varepsilon/2 \right\}^2/2 \\ &+ o\big(\|x-\mu_y\|^2 + \varepsilon^2\big), \end{aligned} \tag{3.22}$$

where v_y is defined in Lemma 2. Using the transformations (3.9), (3.10) and $\gamma = c^{-1/2}\varepsilon/2 + o(1)$ in conjunction with Lemma 2, the rest of the proof is similar to that of Theorem 3.

PROOF OF THEOREM 2(I). To evaluate the Bayes risk $r(N(g,c),\delta^*)$, we shall use the change of variables $\mathcal{Y} = \mathcal{Y}(y,z) \in \boldsymbol{R}^{p-1}, s = s(y,z) \in \boldsymbol{R}$ defined by

$$\mu(y,z) = \mu(\mathcal{Y}, z_0) + s\nabla z(\theta_{\mathcal{Y},z_0})/\sigma(\mathcal{Y}), \tag{3.23}$$

which is the same as (3.10) except that we replace $\sqrt{c}\eta$ by s. Define $J(\theta)$ by (1.8) with Θ_0 and Θ_1 given by Lemma 3(ii), and define $L_\varepsilon(x)$ by (3.21). Then under the transformation (3.23),

$$J(\theta_{y,z}) = L_\varepsilon(\mu(y,z)) \sim \left\{ |s + \varepsilon u(\mathcal{Y})/2\sigma(\mathcal{Y})| + \varepsilon u(\mathcal{Y})/2\sigma(\mathcal{Y}) \right\}^2/2 \tag{3.24}$$

as $\varepsilon \to 0$ and $z \to z_0$, uniformly in $y \in y(A)$, by Lemma 3(ii) and an argument similar to that used to establish (3.22), noting that $(\nabla z(\theta_{\mathcal{Y},z_0}))'\Sigma_{\mathcal{Y}}^{-1}\nabla z(\theta_{\mathcal{Y},z_0}) = \sigma^2(\mathcal{Y})$ by (2.6). As shown by Lai and Zhang (1993) in the proof of Theorem 1,

$$\begin{aligned} E_\theta N(g,c) \sim& \left\{ \log(c^{-1}J(\theta)) \right\}/J(\theta) \quad \text{uniformly in} \\ & \theta \in A \quad \text{and} \quad d_c \le J(\theta) \le D_c, \end{aligned} \tag{3.25}$$

for any positive numbers $d_c \to 0$, $D_c \to \infty$ such that $d_c/c \to \infty$ and $D_c = o(|\log c|)$ as $c \to 0$. Combining (3.24) with (3.25), (2.5) and an argument similar to (3.14) yields that as $\varepsilon \to 0$,

$$\begin{aligned} &\int_{|z-z_0|\le\varepsilon|\log\varepsilon|\sigma(y)} E_{\theta_{y,z}} N(g,c)\, dG \\ &\sim \int_{\mathcal{Y}\in y(A)} \sigma(\mathcal{Y})\pi_G(z_0,\mathcal{Y}) \\ &\quad \cdot \left\{ \int_{|s|\le\varepsilon|\log\varepsilon|} \frac{\log(c^{-1}\varepsilon^2)\, ds}{\left\{|s+\varepsilon u(\mathcal{Y})/2\sigma(\mathcal{Y})| + \varepsilon u(\mathcal{Y})/2\sigma(\mathcal{Y})\right\}^2} \right\} d\mathcal{Y}, \end{aligned} \tag{3.26}$$

noting that $\sup_{y\in y(A)}(\sigma(y)+u(y)) < \infty$ and $\inf_{y\in y(A)} \min(\sigma(y),u(y)) > 0$. From (3.16) and (3.2), it follows that

$$\begin{aligned} J(\theta_{y,z}) &\geq \inf_{\lambda\in A: z(\lambda)=z_0} I(\theta_{y,z},\lambda) \\ &\geq bK^{-2}(z-z_0)^2 \qquad \text{for all} \quad \theta_{y,z}\in A, \end{aligned} \tag{3.27}$$

$$\begin{aligned} &\inf_{\lambda\in\Theta_0} \|\theta-\lambda\| \geq K^{-1}(z(\theta)-z_0) && \text{if } \theta\in\Theta_1, \\ &\inf_{\lambda\in\Theta_1} \|\theta-\lambda\| \geq K^{-1}(z_0+\varepsilon \inf_{y\in y(A)} u(y) - z(\theta)) && \text{if } \theta\in\Theta_0. \end{aligned} \tag{3.28}$$

By (2.5), (3.25) and (3.27), as $\varepsilon\to 0$,

$$\begin{aligned} &\int_{|z-z_0|\geq\varepsilon|\log\varepsilon|\sigma(y)} E_{\theta_{y,z}} N(g,c)\, dG \\ &= O\big(\{\varepsilon|\log\varepsilon|\}^{-1}\{\log(c^{-1}\varepsilon^2)+\log|\log\varepsilon|\}\big). \end{aligned} \tag{3.29}$$

Noting that $I(\widehat{\theta}_n,\theta) \geq \inf_{\lambda\in\Theta_0} I(\widehat{\theta}_n,\lambda)$ if $\theta\in\Theta_0$, it follows from (3.16) that for $\theta\in\Theta_0$,

$$\begin{aligned} &P_\theta\big\{(N(g,c),\delta^*) \text{ rejects } H_0\big\} \\ &\leq P_\theta\Big\{\widehat{\theta}_n \in A_\rho, I(\widehat{\theta}_n,\theta) \geq n^{-1}g(cn) \text{ and} \\ &\qquad\qquad I(\widehat{\theta}_n,\theta) \geq \inf_{\lambda\in\Theta_1} I(\widehat{\theta}_n,\lambda) \text{ for some } n\geq 1\Big\}. \end{aligned} \tag{3.30}$$

Moreover, Lemma 2 of Lai and Zhang (1993) (which is used to prove Theorem 1) says that

$$\begin{aligned} &P_\theta\Big\{\widehat{\theta}_n\in A_\rho, I(\widehat{\theta}_n,\theta)\geq n^{-1}g(cn) \text{ and} \\ &\qquad I(\widehat{\theta}_n,\theta) \geq \inf_{\lambda\in A_\rho, \|\lambda-\theta\|\geq a} I(\widehat{\theta}_n,\lambda) \quad \text{for some} \quad n\geq 1\Big\} \\ &= O((c/a^2)(\log(a^2/c))^{1-\xi+p/2}) \text{ as } c\to 0, \\ &\qquad \text{uniformly in } \sqrt{d_c}\leq a\leq\sqrt{D_c} \text{ and } \theta\in A_\rho, \end{aligned} \tag{3.31}$$

where d_c and D_c are the same as in (3.25). By (3.28), (3.30) and (3.31), as $\varepsilon\to 0$,

$$\begin{aligned} &\int_{\theta_{y,z}\in\Theta_0} P_{\theta_{y,z}}\big\{(N(g,c),\delta^*) \text{ rejects } H_0\big\}\, dG \\ &= O\Big(c\int_{\theta_{y,z}\in\Theta_0}\Big(z_0+\varepsilon\inf_{y\in y(A)} u(y)-z\Big)^{-2} \\ &\qquad \cdot\Big\{\log\Big(\Big(z_0+\varepsilon\inf_{y\in y(A)}u(y)-z\Big)^2/c\Big)\Big\}^{1-\xi+p/2} dG\Big) \\ &= o\big(c\varepsilon^{-1}\log(c^{-1}\varepsilon^2)\big), \qquad \text{by (2.5) and since } \xi > p/2, \end{aligned} \tag{3.32}$$

and a similar result holds for $\int_{\Theta_1} P_\theta\{(N(g,c),\delta^*)$ rejects $H_1\}dG$. By (3.26), (3.29) and (3.32), $r(N(g,c),\delta^*) \sim 4c\varepsilon^{-1}\log(\varepsilon^2/c)\int_{y(A)}(\sigma^2(y)/u(y))\pi_G(z_0,y)\,dy$.

To prove the desired conclusion for $\inf_{T,\delta} r(T,\delta)$, it suffices to restrict to tests (T,δ) such that $r(T,\delta) \le c\varepsilon^{-1}\{\log(\varepsilon^2/c)\}^2$. For such tests, as $\varepsilon \to 0$,

$$(1+o(1))\int_{\substack{y\in y(A)\\ |z-z_0|\le\varepsilon|\log\varepsilon|}} \pi_G(z_0,y)P_{\theta_{y,z}}\Big\{(T,\delta)\text{ makes wrong decision}\Big\}\,dydz$$
$$\le c\varepsilon^{-1}\Big\{\log(\varepsilon^2/c)\Big\}^2. \tag{3.33}$$

Using the change of variables $\mathcal{Y} = \mathcal{Y}(y,z)$, $s = s(y,z)$ defined in (3.23), we obtain from (3.33) by calculations analogous to (3.14) and (3.26) that for any $0<\alpha<1$, as $\epsilon \to 0$,

$$\int_{\mathcal{Y}\in y(A)} \sigma(\mathcal{Y})\pi_G(z_0,\mathcal{Y})\left[\int_{-\alpha\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})}^{0} + \int_{\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})}^{(1+\alpha)\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})}\right]$$
$$\cdot\, P_{\mathcal{Y},s}\{(T,\delta)\text{ errs}\}\,dsd\mathcal{Y}$$
$$\le (1+o(1))c\varepsilon^{-1}\{\log(\varepsilon^2/c)\}^2, \tag{3.34}$$

where $P_{\mathcal{Y},s}\{(T,\delta)$ errs $\}$ denotes $P_\theta\{(T,\delta)$ rejects $H_0\}$ if $\theta(=\theta_{y(\mathcal{Y},s),z(\mathcal{Y},s)}) \in \Theta_0$, and denotes $P_\theta\{(T,\delta)$ rejects $H_1\}$ if $\theta\in\Theta_1$. We can choose $s(\mathcal{Y}) \in [-\alpha\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y}),0]$ and $s_1(\mathcal{Y}) \in [\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y}),\ (1+\alpha)\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})]$ such that

$$\left[\int_{-\alpha\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})}^{0} + \int_{\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})}^{(1+\alpha)\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})}\right] P_{\mathcal{Y},s}\{(T,\delta)\text{ errs}\}\,ds$$
$$\ge \{\alpha\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})\}p_\varepsilon(\mathcal{Y}),$$

where $p_\varepsilon(\mathcal{Y}) = P_{\mathcal{Y},s(\mathcal{Y})}\{(T,\delta)$ rejects $H_0\}+P_{\mathcal{Y},s_1(\mathcal{Y})}\{(T,\delta)$ rejects $H_1\}$. Putting this in (3.34) yields

$$\alpha\varepsilon\int_{\mathcal{Y}\in y(A)} u(\mathcal{Y})\pi_G(z_0,\mathcal{Y})p_\varepsilon(\mathcal{Y})\{M\sigma^2(\mathcal{Y})/u^2(\mathcal{Y})\}\,d\mathcal{Y}$$
$$\le(1+o(1))c\varepsilon^{-1}\{\log(\varepsilon^2/c)\}^2, \tag{3.35}$$

where $M = \{\sup_{y\in y(A)}\sigma(y)/u(y)\}^{-2}$. Define a probability distribution F on $y(A)$ by

$$dF(y) = m^{-1}(\sigma^2(y)/u(y))\pi_G(z_0,y)\,dy, \tag{3.36}$$

where

$$m = \int_{y(A)}(\sigma^2(y)/u(y))\pi_G(z_0,y)\,dy.$$

We can rewrite (3.35) in the form

$$\log\left(\int_{y(A)} p_\varepsilon(y)dF(y)\right) \le \log\left(\{(\alpha Mm)^{-1}+o(1)\}c\varepsilon^{-2}\{\log(\varepsilon^2/c)\}^2\right)$$
$$\sim\log(c\varepsilon^{-2}). \tag{3.37}$$

Let $\theta[\mathcal{Y}, s] = \theta_{y,z}$ with $\mathcal{Y} = \mathcal{Y}(y, z)$, $s = s(y, z)$. By Lemma 3 of Lai and Zhang (1993), which is a restatement of Hoeffding's lower bound for $E_\theta T$ in the context of (1.1),

$$
\begin{aligned}
E_{\theta[\mathcal{Y},s]}T \geq & \frac{(1+o(1))|\log p_\varepsilon(\mathcal{Y})|}{\max\{I(\theta[\mathcal{Y},s],\theta[\mathcal{Y},s(\mathcal{Y})]),\ I(\theta[\mathcal{Y},s],\theta[\mathcal{Y},s_1(\mathcal{Y})])\}} \\
\sim & \frac{2|\log p_\varepsilon(\mathcal{Y})|}{\max\{(s-s(\mathcal{Y}))^2,(s-s_1(\mathcal{Y}))^2\}} \qquad \text{as} \quad s \to 0 \quad \text{and} \quad \varepsilon \to 0,
\end{aligned}
$$

uniformly in $\mathcal{Y} \in y(A)$, where the last relation above follows from $I(\theta[\mathcal{Y}, s], \theta[\mathcal{Y}, s^*]) \sim \frac{1}{2}(s-s^*)^2(\nabla z(\theta_{\mathcal{Y},z_0})/\sigma(\mathcal{Y}))'\Sigma_y^{-1}(\nabla z(\theta_{\mathcal{Y},z_0})/\sigma(\mathcal{Y}))$ as $s - s^* \to 0$ in view of (3.23). Hence, noting that $s_1(\mathcal{Y}) - s(\mathcal{Y}) \leq (1+2\alpha)\varepsilon u(\mathcal{Y})/\sigma(\mathcal{Y})$, we obtain that analogous to (3.14) and (3.26),

$$
\begin{aligned}
&\int_{|z-z_0|\leq\varepsilon|\log\varepsilon|\sigma(y)} E_{\theta_{y,z}}T\, dG \\
\geq&(2+o(1))\int_{\mathcal{Y}\in y(A)} \sigma(\mathcal{Y})\pi_G(z_0,\mathcal{Y})|\log p_\varepsilon(\mathcal{Y})| \\
&\cdot\int_{|s|\leq\varepsilon|\log\varepsilon|} \{\max((s-s(\mathcal{Y}),(s-s_1(\mathcal{Y}))\}^{-2}\, ds\, d\mathcal{Y} \\
\geq&(4+o(1))(1+2\alpha)^{-1}\varepsilon^{-1}\int_{\mathcal{Y}\in y(A)} (\sigma^2(\mathcal{Y})/u(\mathcal{Y}))\pi_G(z_0,\mathcal{Y})|\log p_\varepsilon(\mathcal{Y})|\, d\mathcal{Y} \\
=&(4+o(1))(1+2\alpha)^{-1}\varepsilon^{-1}m\int_{y(A)}(-\log p_\varepsilon(\mathcal{Y}))dF(\mathcal{Y}), \qquad \text{(by (3.36))} \\
\geq&-(4+o(1))(1+2\alpha)^{-1}\varepsilon^{-1}m\log\left(\int_{y(A)} p_\varepsilon(\mathcal{Y})\, dF(\mathcal{Y})\right), \\
&\text{(by Jensen's inequality).} \qquad (3.38)
\end{aligned}
$$

Combining (3.37) with (3.38) yields

$$
r(T,\delta) \geq c\int E_{\theta_{y,z}}T\, dG \geq (4+o(1))(1+2\alpha)^{-1}c\varepsilon^{-1}m\log(\varepsilon^2/c).
$$

Since α can be arbitrarily small, $r(T,\delta) \geq (4+o(1))mc\varepsilon^{-1}\log(\varepsilon^2/c)$. As has already been shown, $r(N(g,c),\delta^*) \sim 4mc\varepsilon^{-1}\log(\varepsilon^2/c)$. Hence the desired conclusion follows.

REFERENCES

LAI, T. L. (1988a). Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* **16**, 856–886.

LAI, T. L. (1988b). Boundary crossing problems for sample means. *Ann. Probab.* **16**, 375–396.

LAI, T. L. (1988c). Bayes sequential tests. In *Statistical Decision Theory and Related Topics IV* **2** (S. S. Gupta and J. O. Berger, eds.), 131–143. Springer-Verlag, New York–Berlin–Heidelberg.

LAI, T. L. and ZHANG, L. (1993). A modification of Schwarz's sequential likelihood ratio tests in multivariate sequential analysis. To appear in *Sequential Analysis* .

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications. 2nd Edition*. Wiley, New York.

ZHANG, L. (1992). Asymptotically optimal sequential tests of linear hypotheses in multiparameter exponential families. Ph.D. Dissertation, Stanford Univ.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CA 94305
USA