The Annals of Statistics
2000, Vol. 28, No. 6, 1638–1669

ASYMPTOTIC APPROXIMATIONS FOR ERROR PROBABILITIES OF SEQUENTIAL OR FIXED SAMPLE SIZE TESTS IN EXPONENTIAL FAMILIES

BY HOCK PENG CHAN[1] AND TZE LEUNG LAI[2]

National University of Singapore and Stanford University

Asymptotic approximations for the error probabilities of sequential tests of composite hypotheses in multiparameter exponential families are developed herein for a general class of test statistics, including generalized likelihood ratio statistics and other functions of the sufficient statistics. These results not only generalize previous approximations for Type I error probabilities of sequential generalized likelihood ratio tests, but also provide a unified treatment of both sequential and fixed sample size tests and of Type I and Type II error probabilities. Geometric arguments involving integration over tubes play an important role in this unified theory.

1. Introduction. Let $X_1, X_2, \ldots$ be i.i.d. random variables. Wald (1945) introduced the sequential probability ratio test to test a simple null versus a simple alternative hypothesis and developed approximations to its error probabilities under both hypotheses by ignoring overshoots. Making use of renewal theory to approximate the distribution of the excess over the boundary, Siegmund (1975) found more refined approximations that are asymptotically equivalent to the actual error probabilities. Subsequently nonlinear renewal theory was developed by Woodroofe (1976) and Lai and Siegmund (1977), and asymptotic approximations to the error probabilities were derived for sequential generalized likelihood ratio (GLR) tests of composite hypotheses in exponential families. The monographs by Woodroofe (1982) and Siegmund (1985) give a systematic introduction to the basic results and methods, while Hu (1988) summarizes the different methods and extends the "backward method" of Siegmund (1985) to the multiparameter case.

A basic feature of this literature is that the approximations depend crucially on the fact that stopping occurs at the first time T when the likelihood ratio or GLR statistic l_T exceeds some threshold c. Thus l_T is equal to c plus an excess over the boundary whose limiting distribution can be obtained using renewal theory. When the test statistic used is not l_T, the arguments break down. Since they are based on the fact that $l_T = c + O_p(1)$, where the $O_p(1)$ term is the overshoot, these arguments are also not applicable when T is replaced by a fixed sample size n. Moreover, whereas the role of l_T in change-of-measure arguments is quite easy to see when the null hypothesis is simple,

Received May 1999; revised June 2000.
[1]Research supported by the National University of Singapore (RP3982708).
[2]Research supported by the NSF and the NSA.
AMS 2000 *subject classifications*. Primary 62L10, 62L15, 62E20; secondary 60F10, 49Q15.
Key words and phrases. Sequential generalized likelihood ratio tests, Bayes sequential tests, multiparameter exponential families, boundary crossing probabilities, integration over tubes.

it becomes increasingly difficult to work with l_T when the region defining a composite null hypothesis becomes increasingly complex. In this paper we describe a unified approach that can be applied not only to likelihood ratio or GLR statistics but also to other functions of the sufficient statistics in a multiparameter exponential family, and that is applicable to both sequential and fixed sample size tests.

The essential ideas of our approach are given in Section 2. Section 3 presents asymptotic approximations to the error probabilities of sequential and fixed sample size tests when these probabilities are of the large deviation type. Most of the error probability approximations in the literature on truncated sequential GLR tests deal with large deviations, and Section 3 gives a general form of these results, extending their applicability to other truncated sequential and nonsequential tests. In particular, in the case of t-tests applied to possibly nonnormal distributions, we show how our results provide more precise large deviation approximations for self-normalized sums than those recently developed by Shao (1997). Motivated by applications to the asymptotic theory of Bayes sequential tests and to composite hypotheses in which the underlying distributions other than their means and variances are not specified, Section 4 develops approximations to error probabilities that are of moderate deviation type. Section 5 considers moderate deviation approximations to the error probabilities of fixed sample size tests and also provides higher-order asymptotic expansions. Thus the approach presented herein is applicable to both large and moderate deviations, and can also yield higher-order asymptotic expansions.

2. Overview of the method and some preliminaries. Let $X_1, X_2, \dots$ be i.i.d. d-dimensional nonlattice random vectors whose common moment generating function is finite in some neighborhood of the origin. Let $S_n = X_1 + \cdots + X_n$, $\mu_0 = EX_1$ and $\Theta = \{\theta\colon Ee^{\theta' X} < \infty\}$, where the prime denotes transpose. Assume that $\operatorname{cov}(X_1)$ is positive definite. Let $\psi(\theta) = \log(Ee^{\theta' X})$ denote the cumulant generating function of X_1. Let Λ be the closure of $\nabla\psi(\Theta)$ and let Λ^o be its interior. Denote the boundary of Λ by $\partial\Lambda\,(= \Lambda - \Lambda^o)$. As noted by Lalley (1983), $\nabla\psi$ is a diffeomorphism from Θ^o onto Λ^o. Let $\theta_\mu = (\nabla\psi)^{-1}(\mu)$. For $\mu \in \Lambda^o$, define

$$\phi(\mu) = \sup_{\theta\in\Theta}\{\theta'\mu - \psi(\theta)\} = \theta_\mu'\mu - \psi(\theta_\mu). \tag{2.1}$$

The function ϕ is the convex dual of ψ and is also known as the *rate function* in large deviations theory. As will be shown in Section 3, we can reduce the analysis of error probabilities of sequential and fixed sample size tests based on the X_i to the following generic problem. Let $g\colon \Lambda \to \mathbf{R}$ and define the stopping time

$$T_c = \inf\{k \ge n_0 : kg(S_k/k) > c\}, \tag{2.2}$$

where n_0 corresponds to a prescribed minimal sample size. The generic problems are to evaluate

$$P\{T_c \le n\}, \quad P\{ng(S_n/n) > c\}, \quad P\Big\{\min_{k\le n}[(n-k)\beta + kg(S_k/k)] > c\Big\}, \tag{2.3}$$

with $n \sim ac$ and $n_0 \sim \delta c$ as $c \to \infty$, for some $a > \delta > 0$ such that $g(\mu_0) < 1/a$ for the first two probabilities, and $\beta > 1/a$ and $g(\mu_0) = 0$ for the third probability.

Consider the first probability in (2.3). For $\delta c \le n \le ac$, it follows from the large deviation principle that under certain conditions, $\log P\{ng(S_n/n) > c\}$ is asymptotically equivalent to $-n \inf\{\phi(\mu) : g(\mu) > c/n\}$ as $c \to \infty$ [cf. Dembo and Zeitouni (1998), Theorem 2.2.30]. Therefore, $\log(\sum_{\delta c\le n\le ac} P\{ng(S_n/n) > c\})$ is asymptotically equivalent to $-\min_{\delta c\le n\le ac} \inf_{g(\mu)>c/n} n\phi(\mu)$, which, upon interchanging the min and inf signs, is asymptotically equivalent to

$$-\inf_{g(\mu)>1/a} \frac{c\phi(\mu)}{\min(1/\delta, g(\mu))} = -\frac{c}{r},$$

where

$$r = \sup_{g(\mu)>1/a} \frac{\min(\delta^{-1}, g(\mu))}{\phi(\mu)}.$$

Hence $P\{T_c \le ac\} = e^{-c/r+o(c)}$ as $c \to \infty$. To obtain a more precise asymptotic approximation, we assume the following regularity conditions:

(A1) g is continuous on Λ^o and there exists $\varepsilon_0 > 0$ such that

$$\sup_{a^{-1}<g(\mu)<\delta^{-1}+\varepsilon_0} g(\mu)/\phi(\mu) = r < \infty.$$

(A2) $M_\varepsilon := \{\mu\colon a^{-1} < g(\mu) < \delta^{-1} + \varepsilon \text{ and } g(\mu)/\phi(\mu) = r\}$ is a q-dimensional oriented manifold for all $0 \le \varepsilon \le \varepsilon_0$, where $q \le d$.

(A3) $\liminf_{\mu\to\partial\Lambda} \phi(\mu) > (\delta r)^{-1}$ and there exists $\varepsilon_1 > 0$ such that $\phi(\mu) > (\delta r)^{-1} + \varepsilon_1$ if $g(\mu) > \delta^{-1} + \varepsilon_0$.

(A4) g is twice continuously differentiable in some neighborhood of M_{ε_0} and $\sigma(\{\mu\colon g(\mu) = \delta^{-1} \text{ and } g(\mu)/\phi(\mu) = r\}) = 0$, where σ is the volume element measure of M_{ε_0}.

Assumptions (A1)–(A3) imply that $\sup_{g(\mu)>a^{-1}} \min(\delta^{-1}, g(\mu))/\phi(\mu)$ can be attained on the q-dimensional manifold M_0. The first part of (A3) implies that there exists $\varepsilon^* > 0$ such that

$$M^* := \{\mu\colon a^{-1} \le g(\mu) \le \delta^{-1} + \varepsilon^*, g(\mu)/\phi(\mu) = r\} \tag{2.4}$$

is a compact subset of Λ; it clearly holds if $\phi(\mu) \to \infty$ as $\mu \to \partial\Lambda$, which is usually the case. The asymptotic formula for $P\{T_c \le ac\}$ in Theorem 1 (in Section 3) involves an integral (with respect to $d\sigma$) over $\{\mu\colon \alpha^{-1} < g(\mu) < \delta^{-1} + \varepsilon_0\}$, whose dimensionality q appears in the formula as the power of $\sqrt{c}$. Spivak (1965) provides a concise introduction to integration on q-dimensional oriented manifolds in $\mathbf{R}^d$. Following Lalley's (1983) notation, for $\mu \in M_0$, let

$TM_0(\mu)$ denote the tangent space of M_0 at μ and let $TM_0^{\perp}(\mu)$ denote its orthogonal complement [i.e., $TM_0^{\perp}(\mu)$ is the normal space of M_0 at μ]. Let $\rho(\mu) = \phi(\mu) - g(\mu)/r$. By (A1) and (A3), ρ attains on M_{ε_0} its minimum value 0 over $\{\mu\colon \alpha^{-1} < g(\mu) < \delta^{-1} + \varepsilon_0\}$ and, therefore,

$$\nabla\rho(\mu) = 0 \text{ and } \nabla^2\rho(\mu) \text{ is nonnegative definite for } \mu \in M_0. \tag{2.5}$$

Let $\Pi_\mu^{\perp}$ denote the $d\times(d-q)$ matrix whose column vectors form an orthonormal basis of $TM_0^{\perp}(\mu)$. Then the matrix $\nabla_\perp^2\rho(\mu) := (\Pi_\mu^{\perp})'\nabla^2\rho(\mu)\Pi_\mu^{\perp}$ is nonnegative definite for $\mu \in M_0$. Letting $|\cdot|$ denote the determinant of a nonnegative definite matrix, we shall also assume that

$$\inf_{\mu\in M_0} |\nabla_\perp^2\rho(\mu)| > 0 \quad \text{with } \rho = \phi - g/r, \tag{A5}$$

where we set $|\nabla_\perp^2\rho(\mu)| = 1$ in the case $d - q = 0$.

Under (A1)–(A5), our method to evaluate $P\{T_c \le ac\}$ consists of the following steps. First consider the case where X_1 has a bounded continuous density function (with respect to Lebesgue measure) so that S_n/n has the saddlepoint approximation

$$P\{S_n/n \in d\mu\} = (1 + o(1))(n/2\pi)^{d/2}|\Sigma(\mu)|^{-1/2}e^{-n\phi(\mu)}\, d\mu, \tag{2.6}$$

where $\Sigma(\mu) = \nabla^2\psi(\theta)|_{\theta=\theta_\mu}$ and the $o(1)$ term is uniform over compact subsets of Λ^o [cf. Borovkov and Rogozin (1965) and Barndorff-Nielsen and Cox (1979)]. Let

$$\begin{aligned} f(\mu)\, d\mu &= P\{T_c \le ac,\ S_{T_c}/T_c \in d\mu\} \\ &= \sum_{\delta c \le n \le ac} P\{S_n/n \in d\mu\} I_{\{ng(\mu) > c\}} \\ &\qquad \times\ P\{kg(S_k/k) < c \text{ for all } \delta c \le k < n | S_n/n \in d\mu\}. \end{aligned} \tag{2.7}$$

Making use of (2.6) and (2.7), we first show that

$$P\{T_c \le ac\} = \int_{\mathbf{R}^d} f(\mu)\, d\mu \sim \int_{U_{c^{-1/2}\log c}} f(\mu) d\mu, \tag{2.8}$$

where U_η is a tubular neighborhood of M_0 with radius η, and then perform the integration in (2.8) over $U_{c^{-1/2}\log c}$. The volumes of tubes around smooth closed curves and more general manifolds, first derived by Hotelling (1939) and Weyl (1939), have played a prominent role in recent developments in inference on nonlinear regression models [cf. Johansen and Johnstone (1990), Johnstone and Siegmund (1989), Knowles and Siegmund (1989), Naiman (1986, 1990), Siegmund and Zhang (1993)]. In the foregoing papers, the tube is a union of disks of radius η around all points of the manifold, and Weyl's tube formula is used to express its volume as a polynomial of degree $[q/2]$ in η. The integral in (2.8) uses a slightly different formulation of the tubular neighborhood and certain differential geometric results on the so-called "infinitesimal change of volume function." Specifically, we say that

$$U_\eta = \{y + z\colon y \in M_0, z \in TM_0^{\perp}(y) \text{ and } \|z\| \le \eta\} \tag{2.9}$$

is a tubular neighborhood of M_0 with radius η if the representation of the elements of U_η in (2.9) is unique. For the existence of tubular neighborhoods when η is sufficiently small, see Theorem 5.1 in Chapter 4 of Hirsch (1976) and its proof. A comprehensive treatment of the infinitesimal change of volume function is given by Gray (1990). From Lemmas 3.13 and 3.14 and Theorem 3.15 of Gray (1990), it follows that as $\eta := c^{-1/2}\log c \to 0$,

$$\int_{U_\eta} f(\mu)d\mu \sim \int_{M_0}\left\{\int_{z\in TM_0^{\perp}(y),\,\|z\|\le\eta} f(y+z)\,dz\right\}d\sigma(y). \tag{2.10}$$

The inner integral in (2.10) can be evaluated asymptotically by making use of (2.6) and (2.7), leading to an asymptotic formula for $P\{T_c \le ac\}$ in view of (2.8).

We have assumed in the preceding analysis that X_1 has a bounded continuous density function. We next replace this assumption by the much weaker assumption that X_1 be nonlattice. By partitioning Λ into suitably small cubes, we use exponential tilting and a refinement of Stone's (1965) local limit theorem to modify the preceding analysis, replacing "$\in d\mu$" by "$\in I_\mu$", where I_μ denotes a cube centered at μ.

For the second probability in (2.3), the large deviation principle yields

$$P\{ng(S_n/n) > c\} = \exp\{-c \inf_{g(\mu)>a^{-1}} \phi(\mu) + o(c)\},$$

since $n \sim ac$. To obtain a more precise approximation, we use integration over some tubular neighborhood of a manifold as in (2.8), under certain geometric assumptions that are analogous to (A1)–(A4). The third probability in (2.3) can again be treated by similar arguments. The moderate deviation results in Sections 4 and 5 involve considerably simpler geometric arguments since the relevant manifolds are contained in a small neighborhood of μ_0. However, we need to replace (2.6) and (2.7) by deeper and more refined versions in Section 4 and by higher-order expansions in Section 5.

3. Large deviation approximations to error probabilities. In this section we first state in Theorems 1–3 the asymptotic formulas for the probabilities in (2.3). We then apply these results to derive large deviation approximations to error probabilities of sequential and fixed sample size tests, giving extensions and refinements of the results of Woodroofe (1978), Lalley (1983), Hu (1988), Chernoff (1952), Hoeffding (1965) and Bahadur (1967). The theorems, which are proved at the end of this section, use the same notation and assumptions as in the first paragraph of Section 2. In addition, the following notation is also used. Let $X_i^{(\mu)}$ be i.i.d. such that $P\{X_i^{(\mu)} \in dx\} = e^{\theta_\mu' x - \psi(\theta_\mu)}dF(x)$, where F is the distribution of X_1, and let $S_n(\mu) = \sum_{i=1}^n\{\theta_\mu' X_i^{(\mu)} - \psi(\theta_\mu)\}$. Let $\Sigma(\mu) = \nabla^2\psi(\theta)|_{\theta=\theta_\mu}$.

THEOREM 1. *Suppose X_1 is nonlattice and g: $\Lambda \to \mathbf{R}$ satisfies* (A1)–(A5) *with $a > \delta$, $g(\mu_0) < a^{-1}$ and $n_0 \sim \delta c$. Let $\gamma(\mu) = \int_0^\infty e^{-y}P\{\min_{n\ge1} S_n(\mu) >$*

$y\}\,dy$. Then as $c \to \infty$,

$$P\{T_c \le ac\} \sim \left(\frac{c}{2\pi r}\right)^{q/2} e^{-c/r} \times \int_{M_0} \gamma(\mu)(\phi(\mu))^{-(q/2+1)}|\Sigma(\mu)|^{-1/2}|\nabla_\perp^2\rho(\mu)|^{-1/2}\, d\sigma(\mu),$$

where $\nabla_\perp^2\rho$ is introduced in (A5).

In the next two theorems, $g(\mu_0) < b$ and instead of (A1)–(A5), we impose the following conditions on g:

(B1) g is continuous on Λ^o and $\inf\{\phi(\mu): g(\mu) \ge b\} = b/r$.

(B2) g is twice continuously differentiable on $\{\mu \in \Lambda^o: b - \varepsilon_0 < g(\mu) < b + \varepsilon_0\}$ for some $\varepsilon_0 > 0$.

(B3) $\nabla g(\mu) \ne 0$ on $N := \{\mu \in \Lambda^o: g(\mu) = b\}$, and $M := \{\mu \in \Lambda^o: g(\mu) = b, \phi(\mu) = b/r\}$ is a smooth p-dimensional manifold (possibly with boundary) for some $0 \le p \le d-1$.

(B4) $\liminf_{\mu \to \partial\Lambda} \phi(\mu) > br^{-1}$ and $\inf_{g(\mu) > b+\delta} \phi(\mu) > br^{-1}$ for every $\delta > 0$.

For the notion of smooth submanifolds (with or without boundaries), see Hirsch (1976). Under (B2) and (B3), N is a $(d-1)$-dimensional manifold and $TN^\perp(\mu)$ is a one-dimensional linear space with basis vector $\nabla g(\mu)$; see Theorem 5-1 and Problem 5-13(c) of Spivak (1965). Moreover, making use of (B1)-(B4), it is shown in the Appendix that

$$(3.1)\quad \inf_{\mu \in M} \|\nabla\phi(\mu)\| > 0, \quad (\nabla g(\mu))'\nabla\phi(\mu) > 0 \text{ and } \nabla\phi(\mu) \in TN^\perp(\mu) \text{ for all } \mu \in M.$$

Hence $\nabla\phi(\mu) = s\nabla g(\mu)$ with $s = \|\nabla\phi(\mu)\|/\|\nabla g(\mu)\|$. Let $e_1(\mu) = \nabla\phi(\mu)/\|\nabla\phi(\mu)\|$ and let $\{e_1(\mu), e_2(\mu), \dots, e_{d-p}(\mu)\}$ be an orthonormal basis of $TM^\perp(\mu)$. Define the $d \times (d-p-1)$ matrix Π_μ (in the case $d > p+1$) and the positive number $\xi(\mu)$ by

$$(3.2)\qquad \Pi_\mu = (e_2(\mu) \cdots e_{d-p}(\mu)),$$

$$(3.3)\quad \xi(\mu) = \begin{cases} 1/\|\nabla\phi(\mu)\|, & \text{if } d = p+1, \\ |\Pi_\mu'\{\Sigma^{-1}(\mu) - s\nabla^2 g(\mu)\}\Pi_\mu|^{-1/2}/\|\nabla\phi(\mu)\|, & \text{if } d > p+1, \end{cases}$$

where analogous to (A5) we assume that

(B5) $\inf_{\mu \in M} |\Pi_\mu'\{\Sigma^{-1}(\mu) - s\nabla^2 g(\mu)\}\Pi_\mu| > 0$ if $d > p+1$.

THEOREM 2. *Suppose X_1 is nonlattice and $g: \Lambda \to \mathbf{R}$ satisfies* (B1)–(B5). *Let $b > g(\mu_0)$. Then as $n \to \infty$,*

$$P\{g(S_n/n) > b\} \sim P\{g(S_n/n) \ge b\} \sim (2\pi)^{-(p+1)/2} n^{(p-1)/2} e^{-bn/r} \int_M \xi(\mu)|\Sigma(\mu)|^{-1/2}\, d\sigma(\mu).$$

THEOREM 3. *Suppose X_1 is nonlattice, $g:\Lambda \to \mathbf{R}$ satisfies* (B1)–(B5) *and* $g(\mu_0) = 0$. *Let $\beta > b > 0$. Define $X_i^{(\mu)}$ as in Theorem* 1 *and let* $W_n(\mu) = \sum_{i=1}^n \{\theta_\mu'(X_i^{(\mu)} - \mu) + s(b-\beta)\}$. *Let* $w(\mu) = \int_0^\infty e^{-y} P\{\max_{n\ge 1} W_n(\mu) < y\}\, dy$. *Then as* $n \to \infty$,

$$
\begin{aligned}
&P\Big\{\min_{k\le n}[(n-k)\beta + kg(S_k/k)] > bn\Big\}\\
&\quad\sim P\Big\{\min_{k\le n}[(n-k)\beta + kg(S_k/k)] \ge bn\Big\} \qquad (3.4)\\
&\quad\sim (2\pi)^{-(p+1)/2} n^{-(p-1)/2} e^{-bn/r} \int_M \xi(\mu) w(\mu) |\Sigma(\mu)|^{-1/2}\, d\sigma(\mu).
\end{aligned}
$$

3.1. *Applications to Type I and Type II error probabilities of sequential and fixed sample size tests.* We now apply Theorems 1–3 to analyze the error probabilities of a variety of sequential and fixed sample size tests.

(A) Consider the multiparameter exponential family with density function $\exp(\theta' x - \psi(\theta))$ with respect to some probability measure F. The natural parameter space is Θ. Let Θ_1 be a q_1-dimensional smooth submanifold of Θ^o and let Θ_0 be a q_0-dimensional smooth submanifold of Θ_1 with $0 \le q_0 < q_1 \le d$. The GLR statistics for testing the null hypothesis $H_0: \theta \in \Theta_0$ versus the alternative hypothesis $H_1: \theta \in \Theta_1 - \Theta_0$ are of the form $ng(S_n/n)$, where

$$g(x) = \phi_1(x) - \phi_0(x) \quad \text{with } \phi_i(x) = \sup_{\theta\in\Theta_i} (\theta' x - \psi(\theta)). \qquad (3.5)$$

Then $g(x) \le \phi(x)$ and equality is attained if and only if $\phi_1(x) = \phi(x)$ and $\phi_0(x) = 0$. Since $\nabla\psi$ is a diffeomorphism, $\Lambda_i = \nabla\psi(\Theta_i)$ is a q_i-dimensional submanifold of Λ^o. Note that $\phi(x) = \phi_1(x)$ iff $x \in \Lambda_1$. Consider the sequential GLR test with stopping rule $T_c \wedge [ac]$, where T_c is defined in (2.2) with g given by (3.5) and $n_0 \sim \delta c$. To evaluate the Type I error probability at θ_0, we can assume, by choosing the underlying probability measure F as that associated with θ_0 and by replacing X_i by $X_i - \nabla\psi(\theta_0)$, that

$$\theta_0 = 0, \qquad \psi(0) = 0 \quad \text{and} \quad \nabla\psi(0) = 0. \qquad (3.6)$$

Then (A1)–(A5) hold with $r = 1$ and $q = q_1 - q_0$ under certain regularity conditions [cf. Woodroofe (1978)], and therefore we can apply Theorem 1 to approximate the Type I error probability $P_0\{T_c \le ac\}$. Woodroofe (1978) also obtained a similar approximation to the error probability in the case $\Theta_1 = \Theta$ under the additional assumption that S_m has a bounded continuous density function with respect to Lebesgue measure for some $m \ge 1$. However, the constant in the asymptotic formula of Theorem 1 is represented more directly as an integral over the manifold M_0 than as Woodroofe's representation, which involves a two-stage procedure that first integrates over $M_0^{(t)} := M_0 \cap \{\mu: \phi(\mu) = t\}$ and then integrates over t in the interval $(1/a, 1/\delta)$. The integral representation in Theorem 1 is also simpler than that of Lalley (1983), which involves the ratio of determinants of some complicated matrices and which replaces $\gamma(\mu)$

by the right-hand side of the following formula [cf. Theorem 2.7 of Woodroofe (1982), noting that $ES_1^{(\mu)} = \theta_\mu'\mu - \psi(\theta_\mu) = \phi(\mu)$]:

$$\gamma(\mu) = \phi(\mu) \lim_{c\to\infty} E_{\theta_\mu}[\exp\{-T_c g(S_{T_c}/T_c) + c\}].$$

Moreover, Lalley's derivation of the approximation to $P_0\{T_c \le ac\}$ is very different from our approach, and his change-of-measure arguments lead to somewhat different assumptions and representations of both the integrand and the manifold over which integration is performed. We next give two concrete examples of the applications of Theorem 1. In the examples, it is more convenient to denote the elements of $\mathbf{R}^d$ as row (instead of column) vectors.

EXAMPLE 1. Consider the repeated t-test of the null hypothesis H_0 that the common mean of i.i.d. normal observations $Y_1, Y_2, \dots$ is 0 when the variance is unknown. Here $X_i = (Y_i, Y_i^2)$, $S_n/n = (n^{-1}\sum_1^n Y_i, n^{-1}\sum_1^n Y_i^2)$, $\Lambda^o = \{(y,v): v > y^2\}$ and $\Lambda_0 = \{(0,v) : v > 0\}$. The GLR statistics are of the form $ng(S_n/n)$, where $g(y,v) = \frac{1}{2}\log(v/(v-y^2))$, and the repeated t-test rejects H_0 if $ng(S_n/n) > c$ for some $\delta c \le n \le ac$. The test is invariant under scale changes, so we can consider the Type I error probability when $\mathrm{Var}(Y_i) = 1$. Since $\phi(y,v) = [v - 1 - \log(v - y^2)]/2$, (A1)–(A4) are satisfied with $r = 1$ and

$$M_\varepsilon = \{(y,1): 1 - \exp(-2(\delta^{-1}+\varepsilon)) > y^2 > 1 - \exp(-2a^{-1})\}.$$

Moreover, (A5) holds since $\nabla_\perp^2 \rho(y,1) = 1/2$ for $(y,1) \in M_0$.

EXAMPLE 2. Let $Y_1, Y_2, \dots$ and $Z_1, Z_2, \dots$ be independent exponential random variables such that Y_i has mean λ_1 and Z_i has mean λ_2 for all i. Consider the truncated sequential GLR test of $H_0: \lambda_1 = \lambda_2$. Here $S_n/n = (n^{-1}\Sigma_1^n Y_i, n^{-1}\Sigma_1^n Z_i)$ and the GLR statistics are of the form $ng(S_n/n)$, where

$$g(y,z) = 2\log\{[(y+z)/2]/\sqrt{yz}\}.$$

This test rejects H_0 when $kg(S_k/k) > c$ for some $ac \ge k \ge \delta c$. Consider the Type I error probability when $\lambda_1 = \lambda_2 = 1$. We can take 1 to be the common value of λ_1 and λ_2 under H_0 because the problem is invariant under scale changes. Since $\phi(y,z) = -\log y - \log z - 2 + y + z$, $g(y,z) \le \phi(y,z)$ with equality if and only if $y + z = 2$. For $0 < y < 2$, $g(y, 2-y)$ is strictly convex and attains the minimum value 0 at $y = 1$. Thus $r = 1$ and for all $\varepsilon \ge 0$, $M_\varepsilon = \{(y, 2-y): 0 < y < 2 \text{ and } a^{-1} < g(y,2-y) < \delta^{-1} + \varepsilon\}$ is a one-dimensional manifold, so (A1) and (A2) are satisfied. Moreover, (A3) and (A4) can be easily shown to hold. To check (A5), note that $\rho(y,z) = -2\log[(y+z)/2] - 2 + y + z$ and that for $(y,z) \in M_0$, $\Pi_{y,z}^\perp = (1,1)/\sqrt{2}$, yielding $\nabla_\perp^2\rho(y,z) = 1$. Letting $\nu(\mu) = \gamma(\mu)/\phi(\mu)$ and noting that $d\sigma(y, 2-y) = \sqrt{2}\,dy$, it therefore follows from Theorem 1 that the Type I error probability of the test is

$$(1+o(1))\sqrt{c/\pi}e^{-c} \int_{e^{-1/a} \ge y(2-y) \ge e^{-1/\delta}} \nu(y, 2-y) \times \{-\log(y(2-y))\}^{-1/2}\{y(2-y)\}^{-1}\,dy. \tag{3.7}$$

(B) With the same notation as in (A), suppose that instead of the stopping rule $T_c \wedge [ac]$, the GLR test of H_0 is based on a sample of fixed size n. The test rejects H_0 if $g(S_n/n) > b$, where g is defined by (3.5). To evaluate the Type I error probability at θ_0, there is no loss of generality in assuming (3.6). Then (B1)–(B5) hold with $r = 1$ and $p = q_1 - q_0 - 1$ under certain regularity conditions [cf. Woodroofe (1978)], so Theorem 2 can be used to approximate the Type I error probability $P_0\{g(S_n/n) > b\}$. Not only does Theorem 2 remove Woodroofe's assumption that S_n/n has a bounded continuous density with respect to Lebesgue measure, but a different choice of g in Theorem 2 also gives an approximation to the Type II error probability $P_\theta\{g(S_n/n) \le b\}$ with $g(\nabla\psi(\theta)) > b$. Specifically, let $\tilde{g}(\mu) = g(\nabla\psi(\theta)) - g(\mu)$, $\tilde{b} = g(\nabla\psi(\theta)) - b$, and apply Theorem 2 with g, b replaced by $\tilde{g}$, $\tilde{b}$.

Following the seminal work of Chernoff (1952), Hoeffding (1965) and Bahadur (1967) on asymptotic efficiencies of tests at nonlocal alternatives, most of the papers in the literature on large deviation approximations of the Type I and Type II error probabilities give only the order of magnitude of the logarithms of the probabilities. Theorem 2 provides a much more precise approximation. For linear hypotheses about a multivariate normal mean, these more refined large deviation approximations have been derived from well developed exact distribution theory in the normal case [cf. Groeneboom (1980), pages 71–90].

EXAMPLE 3. Consider the case of two independent exponential distributions in Example 2. As before, $H_0\colon \lambda_1 = \lambda_2$, but here the sample size n is fixed. We first consider the Type I error probability and assume without loss of generality $\lambda_1 = \lambda_2 = 1$. Let $b > 0 = g(1, 1)$. As shown in Example 2, (B1) holds with $r = 1$ and (B2), (B4), (B5) also hold. Since $g(y, z)$ is a function of z/y, the manifold N is a disjoint union of two rays of the form $z = \alpha y$ $(y > 0)$. Since M is the intersection of N with the line $y + z = 2$, it is simply a set containing two points, $(\tilde{y}, 2 - \tilde{y})$ and $(2 - \tilde{y}, \tilde{y})$, where $\tilde{y} = 1 - \sqrt{1 - e^{-b}}$. In fact, $\tilde{y}$ and $2 - \tilde{y}$ are solutions of the equation $g(y, 2 - y) = b$. Hence (B3) is satisfied with $p = 0$. Moreover, $|\Sigma(\tilde{y}, 2 - \tilde{y})|^{1/2} = e^{-b}$. It can be shown that $\xi(\tilde{y}, 2 - \tilde{y}) = \xi(2 - \tilde{y}, \tilde{y}) = \xi_b$ and, therefore, by Theorem 2,

$$P\{g(S_n/n) > b\} \sim \sqrt{2/\pi}\ \xi_b n^{-1/2} e^{-bn+b}, \quad \xi_b = e^{-b}\{2(1 - e^{-b})\}^{-1/2}. \tag{3.8}$$

Except for the difference between the multiplicative constants $\xi_b e^b$ and $(2b)^{-1/2}$, the large deviation approximation (3.8) to the tail probability of $2ng(S_n/n)$ agrees with that of the chi-square distribution with 1 degree of freedom:

$$P\{\chi_1^2 > 2bn\} \sim \sqrt{2/\pi}\ (2bn)^{-1/2} e^{-bn} \quad \text{as } n \to \infty. \tag{3.9}$$

This is in contrast to the case of simple null hypotheses in one-parameter exponential families, for which $g = \phi$ and $P\{\phi(S_n/n) > b\} \sim P\{\chi_1^2 > 2bn\}$

[cf. Jensen (1995), page 118]. On the other hand, conditional large deviation probabilities of GLR statistics given the values of ancillary statistics have valid chi-square approximations after suitable Bartlett-type adjustments [cf. Section 5.2 and, in particular, pages 123–124 of Jensen (1995)].

We next consider the Type II error probability at (λ_1, λ_2) such that $\lambda_2 > \lambda_1$ and $g(\lambda_1, \lambda_2) > b$. Without loss of generality we shall take $\lambda_1 = 1$. Let $\lambda = \lambda_2 (> 1)$. Since the underlying density function is $\lambda^{-1} e^{-z/\lambda} e^{-y}$, the rate function ϕ now has the form

$$\phi(y, z) = -\log\ y - \log\ z - 2 + \log\ \lambda + y + z/\lambda.$$

For fixed $\alpha > 0$, $\phi(y, \alpha y)$ attains its minimum value $m(\alpha) := 2\ \log(1 + \alpha/\lambda) - \log\ (\alpha/\lambda) - 2\log 2$ at $y = 2/(1 + \alpha/\lambda)$. Moreover, $m(\alpha)$ is decreasing in $\alpha < \lambda$ and $g(1, \alpha) = 2\log((1+\alpha)/\sqrt{\alpha}) - 2\log 2$ is decreasing for $\alpha < 1$ and increasing for $\alpha > 1$. Let $z = \alpha_1 y$ and $z = \alpha_2 y$ $(y > 0)$ be the two rays of $N = \{(y, z): g(y, z) = b\}$, noting that $g(y, z) = g(1, z/y)$, where $\alpha_1 = e^b (2 - \tilde{y})^2$ and $\alpha_2 = e^b \tilde{y}^2$ are the roots of the equation $\alpha + 1 = 2e^{b/2}\sqrt{\alpha}$, or equivalently of $g(1, \alpha) = b$. Note that $\alpha_1 > \alpha_2$ and $1 < \alpha_1 < \lambda$, since $g(1, \lambda) > b$. It then follows that

$$\inf\{\phi(y, z):\ g(y, z) \le b\} = \inf\{\phi(y, z):\ g(y, z) = b\} = m(\alpha_1).$$

Let $\tilde{g}(y, z) = g(1, \lambda) - g(y, z)$ and $\tilde{b} = g(1, \lambda) - b$. Then (B1)–(B5) hold with $\tilde{g}, \tilde{b}$ replacing g, b and with $\tilde{b}/r = m(\alpha_1)$, $p = 0$ and M consisting of the single point $\tilde{\mu} = (2/(1+\alpha_1/\lambda), 2\alpha_1/(1+\alpha_1/\lambda))$. Hence Theorem 2 yields the following approximation to the Type II error probability:

$$P_{\lambda_1=1,\ \lambda_2=\lambda}\{g(S_n/n) \le b\} \sim (2\pi n)^{-1/2} \xi(\tilde{\mu}) |\Sigma(\tilde{\mu})|^{-1/2} e^{-m(\alpha_1)n}.$$

(C) Theorems 1 and 2 can also be applied to analyze error probabilities of tests that are not based on likelihood ratio statistics. For example, consider the repeated t-test of Example 1 when the underlying distribution is actually nonnormal. Here $g(y, v) = -\frac{1}{2}\log(1 - y^2/v)$ is an increasing function of $|y|/\sqrt{v}$, which increases as v decreases. Thus exponential tilting for the probabilities in (2.3) can be restricted to $\{(\theta_1, \theta_2): \theta_2 < 0\}$, on which $Ee^{\theta_1 Y + \theta_2 Y^2} < \infty$ without any moment conditions on Y. We shall therefore assume only that Y is nondegenerate and nonlattice. In this general setting,

$$\phi(y, v) = \sup_{\substack{\gamma \in \mathbf{R} \\ \lambda > 0}} \{\gamma y - \lambda v - \log E e^{\gamma Y - \lambda Y^2}\} \quad \text{for } v \ge y^2.$$

Write $g(y, v) = G(|y|/\sqrt{v})$. For $0 \le t \le 1$, define $F(t) = \inf_{v > 0} \phi(t\sqrt{v}, v) = \phi(t\sqrt{v_t}, v_t)$. Then

$$\sup_{v \ge y^2} g(y, v)/\phi(y, v) = \sup_{0 \le t \le 1} [G(t)/\min\{F(t), F(-t)\}].$$

In the normal case considered in Example 1, $G = F$ since $v - 1 - \log v$ has minimum value 0. For non-normal Y, suppose $r = \sup_{0 \le t \le 1}[G(t)/\min\{F(t), F(-t)\}]$ is attained at $t^* \in (0, 1)$ and $a^{-1} < G(t^*) < \delta^{-1}$. Then (A1)–(A5) hold with $q = 0$ and $M_\varepsilon = \{(t^* \sqrt{v_{t^*}}, v_{t^*})\}$, or $\{(-t^* \sqrt{v_{t^*}}, v_{t^*})\}$, or $\{(t^* \sqrt{v_{t^*}}, v_{t^*}),$

$(-t^*\sqrt{v_{t^*}}, v_{t^*})\}$ according as $F(t^*) < F(-t^*)$, or $F(-t^*) < F(t^*)$, or $F(t^*) = F(-t^*)$. Hence application of Theorem 1 yields a large deviation approximation to $P\{T_c \le ac\}$ even when the underlying distribution to which the repeated t-test is applied does not have finite pth absolute moment for any $p > 0$.

Results of this kind have been obtained recently for fixed sample size tests by Shao (1997). With the same notation as in Theorem 1, consider the one-sided t-test that rejects H_0 if $\sqrt{n}\overline{Y}_n/\{\sum_1^n(Y_i - \overline{Y}_n)^2\}^{1/2} > b/(1-b^2)^{1/2}$, or equivalently if $\tilde{g}(S_n/n) > b$, where $\tilde{g}(y, v) = y/\sqrt{v}$ and $0 < b < 1$. When the t-test is applied to a nondegenerate distribution satisfying either $EY = 0$ or $EY^2 = \infty$, Shao (1997) showed that $n^{-1}\log P\{\tilde{g}(S_n/n) > b\}$ converges, as $n \to \infty$, to

$$\kappa := \sup_{\sigma>0}\inf_{\gamma>0}\log E\exp\{\gamma\sigma Y - \gamma b(Y^2+\sigma^2)/2\}.$$

He derived this result from large deviation bounds for

$$P\left\{\sup_{\sigma>0}\sum_{i=1}^n(\sigma Y_i - b(Y_i^2+\sigma^2)/2) > 0\right\} \qquad (= P\{\tilde{g}(S_n/n) > b\}).$$

Assume that Y is nonlattice, with $EY^2 < \infty$ and $EY = 0$. We can apply Theorem 2 to give a more transparent derivation of Shao's result and strengthen it into

$$P\{\tilde{g}(S_n/n) > b\} \sim (2\pi n)^{-1/2}Ce^{\kappa n}, \tag{3.10}$$

where the constant C is given by the integral in Theorem 2 that will be made more specific later. First note that for $1 \ge t > 0$,

$$\begin{aligned}
-F(t) &= \log\left\{\sup_{v>0}\inf_{\gamma\in\mathbf{R}}\inf_{\lambda>0}Ee^{\gamma Y-\lambda Y^2-\gamma t\sqrt{v}+\lambda v}\right\}\\
&= \log\left\{\sup_{v>0}\inf_{\gamma>0}E\exp\left(\gamma Y - \frac{\gamma t}{2\sqrt{v}}Y^2 - \gamma t\sqrt{v} + \frac{\gamma t}{2}\sqrt{v}\right)\right\}\\
&= \log\left\{\sup_{\sigma>0}\inf_{\gamma>0}E\exp\left(\gamma\sigma Y - \gamma t\frac{Y^2+\sigma^2}{2}\right)\right\},
\end{aligned}$$

where the third equality follows by setting $v = \sigma^2$ and replacing γ by $\gamma\sigma$ and the second inequality comes from solving the equations $\partial/\partial\lambda = 0$, $\partial/\partial\gamma = 0$, $\partial/\partial v = 0$ (yielding $\lambda = \gamma t/2\sqrt{v}$), associated with the optimization problem that defines $-F(t)$. Clearly $-F(t)$ is decreasing in t and, therefore,

$$\inf\{\phi(y, v): \tilde{g}(y, v) \ge b\} = \inf\{F(t): t \ge b\} = F(b) = -\kappa.$$

Assumptions (B1)–(B5) hold with $d = 2$, $p = 0$ and $M = \{(b\sqrt{v_b}, v_b)\}$, where v_t is the minimizer of $\phi(t\sqrt{v}, v)$ defined earlier. Hence Theorem 2 gives the desired conclusion (3.10). When Y is normal, $C = b^{-1}(1-b^2)^{-1/2}$ and $\kappa = \frac{1}{2}\log(1-b^2)$, which can also be derived from the tail probability of Student's t-distribution with $n-1$ degrees of freedom:

$$P\{t_{n-1} > (n-1)^{1/2}b/(1-b^2)^{1/2}\} \sim (2\pi n)^{-1/2}b^{-1}(1-b^2)^{(n-1)/2} \qquad \text{as } n \to \infty.$$

(D) Theorem 3 can be used to evaluate the Type II error probability of the sequential test that rejects H_0 if $kg(S_k/k) > c$ for some $k \le ac$. Suppose $g(\mu_0) > a^{-1}$. Then the Type II error probability of the test at the alternative with $EX_1 = \mu_0$ can be expressed in the form

$$P_{\mu_0}\left\{\max_{k\le n} kg(S_k/k) \le c\right\} = P_{\mu_0}\left\{\min_{k\le n}[ng(\mu_0) - kg(S_k/k)] \ge ng(\mu_0) - c\right\}$$
$$= P_{\mu_0}\left\{\min_{k\le n}[(n-k)\beta + k\tilde{g}(S_k/k)] \ge bn\right\},$$

to which Theorem 3 is applicable, where $\beta = g(\mu_0)$, $\tilde{g} = g(\mu_0) - g$ and $b = g(\mu_0) - a^{-1}$.

EXAMPLE 4. With the same notation and assumptions as Example 1, consider the Type II error probability of the repeated t-test at the alternative where $E(Y_i) = \gamma \neq 0$ and $\mathrm{Var}(Y_i) = 1$. Thus $E(Y_i^2) = 1 + \gamma^2$. Suppose $\gamma > 0$ and $g(\gamma, 1+\gamma^2) > a^{-1}$. Let $b = g(\gamma, 1+\gamma^2) - a^{-1}$,

$$\tilde{g}(y, v) = g(\gamma, 1+\gamma^2) - g(y, v) = \{\log(1+\gamma^2) - \log(v/(v-y^2))\}/2.$$

Since the logarithm of the underlying density function is $-(y-\gamma)^2/2 - \log(\sqrt{2\pi})$, the rate function now takes the form

$$\phi(y, v) = [v - 1 - \log(v - y^2) - 2\gamma y + \gamma^2]/2.$$

Since ϕ is strictly convex with its global minimum at $(\gamma, 1+\gamma^2)$ and since $g(\gamma, 1+\gamma^2) > a^{-1}$, the minimum of ϕ over the region $\{(y, v) : g(y, v) \le a^{-1}\}$ occurs at $v = \alpha y^2$ with α satisfying $g(1, \alpha) = a^{-1}$, or equivalently $\alpha/(\alpha-1) = e^{2/a}$. Since $\phi(y, \alpha y^2) = \{\alpha y^2 - 1 - \log(\alpha - 1) - \log y^2 - 2\gamma y + \gamma^2\}/2$ is minimized at $y_a := (\gamma + \sqrt{\gamma^2 + 4\alpha})/2\alpha$, (B1) holds with $\tilde{g}$ in place of g and $b/r = \phi(\mu_a)$ and (B3) holds with M consisting of the single point $\mu_a := (y_a, \alpha y_a^2)$. Moreover, (B2), (B4) and (B5) also hold (with $\tilde{g}$ in place of g). Hence Theorem 3 can be applied to give

$$P_\gamma\left\{\max_{2\le k\le ac} kg(S_k/k) \le c\right\} \sim (2\pi ac)^{-1/2}\xi(\mu_a)w(\mu_a)|\Sigma(\mu_a)|^{-1/2}e^{-ac\phi(\mu_a)}$$

as $c \to \infty$. The proof of Theorem 3 actually shows that $P_\gamma\{\max_{\delta c\le k\le ac} kg(S_k/k) \le c\} \sim P_\gamma\{\max_{2\le k\le ac} kg(S_k/k) \le c\}$ and, therefore, the preceding asymptotic formula also yields the Type II error probability of the repeated t-test.

(E) In the case $X_i = \log(f_1(Y_i)/f_0(Y_i))$ and $g(x) = x$, Theorems 1 and 3 can be used to give large deviation approximations to the error probabilities of truncated sequential probability ratio tests of $H_0: f = f_0$ versus $H_1: f = f_1$, where f is the common density function of the Y_i with respect to some measure ν. In particular, suppose H_0 is true. Let $e^{\psi(\theta)} = Ee^{\theta X_1} (= \int f_1^\theta f_0^{1-\theta} d\nu)$. Then $\psi(0) = \psi(1) = 0$. Let $\mu^* = d\psi/d\theta|_{\theta=1}$. If $a^{-1} < \mu^* < \delta^{-1}$, then (A1)–(A5) hold with $r = 1$, $q = 0$ and $M_\varepsilon = \{\mu^*\}$.

3.2. *Proof of Theorems* 1–3. We preface the proof of the theorems by the following lemma, which gives uniform convergence over compact subsets of Λ^o in the local limit theorem for $\sum_{i=1}^n X_i^{(\mu)}$, where $X_i^{(\mu)}$ is defined in the sentence before Theorem 1. The proof of the lemma is a refinement of Stone's (1965) arguments, and the details are given in Chan (1998).

LEMMA 1. *For $h > 0$ and $x = (x_1, \dots, x_d)' \in \mathbf{R}^d$, let $K(x, h) = \{y \in \mathbf{R}^d: x_i \le y_i \le x_i + h$ for all $1 \le i \le d\}$. Suppose F is nonlattice. Let C be a compact subset of Λ^o. Then there exist positive numbers $h_0(n)$ with $\lim_{n\to\infty} h_0(n) = 0$ such that as $n \to \infty$, $\varepsilon \to 0$ and $h \to 0$ with $h \ge h_0(n)$,*

$$P\left\{\sum_{i=1}^n X_i^{(\mu)} \in K(\mu n + \varepsilon\sqrt{n}, h)\right\} = \{(2\pi)^{-d/2}|\Sigma(\mu)|^{-1/2} + o(1)\}(h/\sqrt{n})^d,$$

where the $o(1)$ term is uniform in $\mu \in C$.

PROOF OF THEOREM 1. Replacing g by g/r and c by c/r, we shall assume without loss of generality that $r = 1$. Suppose first that X_1 has a bounded continuous density. Then $P\{T_c \le ac\} = \int_{\mathbf{R}^d} f(\mu)\, d\mu$, where f is defined in (2.7). Let $t = [c^{1/4}]$ and

$$\begin{aligned} f_n(\mu) &= P\{S_n/n \in d\mu\} \\ &\quad \times P\big\{jg(S_j/j) \le c \text{ for all } n - t \le j < n | S_n/n = \mu\big\}. \end{aligned} \tag{3.11}$$

Let $U_{\eta,\varepsilon}$ be the tubular neighborhood of M_ε with radius η, that is, $U_{\eta,\varepsilon}$ is defined by (2.9) with M_ε in place of M_0. It will be shown in the Appendix that

$$P\{T_c \le ac\} = \int_{U_{c^{-1/2}\log c,\, c^{-1/2}}} \left\{\sum_{\substack{n > c/g(\mu) \\ \delta c + t \le n \le ac}} f_n(\mu)\right\} d\mu + o(c^{q/2}e^{-c}). \tag{3.12}$$

Note that for $\mu \in M_\varepsilon$, $g(\mu) = \phi(\mu)$ and $\mu'\nabla g(\mu) - g(\mu) = \psi(\theta_\mu)$. Let $S_{n,k} = \sum_{i=n-k+1}^n X_i$, so $S_n = S_{n-k} + S_{n,k}$. It follows from (A4) that $\max_{\delta c + t \le n \le ac,\, 1 \le k \le t} S_{n,k}^2/n \xrightarrow{P} 0$ and that

$$\begin{aligned} &(n-k)g\big((n\mu - S_{n,k})/(n-k)\big) \\ &\quad = ng(\mu) - \big\{(\nabla g(\mu))'S_{n,k} - k(\mu'\nabla g(\mu) - g(\mu))\big\} + O(S_{n,k}^2/n), \end{aligned} \tag{3.13}$$

and therefore an argument similar to Lemmas 1 and 2 of Woodroofe (1978) can be used to show that uniformly in $\delta c + t \le n \le ac$ and in $\mu \in U_{c^{-1/2}\log c,\, c^{-1/2}}$,

$$\begin{aligned} &P\{(n-k)g(S_{n-k}/(n-k)) \le c \text{ for all } 1 \le k \le t | S_n/n = \mu\} \\ &\quad = p(\mu; ng(\mu) - c) + o(1), \end{aligned} \tag{3.14}$$

where $p(\mu; x) = P\{\min_{k\geq 1} S_k(\mu) \geq x\}$. Putting (2.6) and (3.14) into (3.11) and setting $x = ng(\mu) - c$, it follows from (3.12) that $P\{T_c \leq ac\}$ is equal to

$$\int_{U_{c^{-1/2}\log c,\, c^{-1/2}}} \left\{ \int_0^\infty (2\pi)^{-d/2} [(x+c)/g(\mu)]^{d/2} |\Sigma(\mu)|^{-1/2} \times \exp[-(x+c)\phi(\mu)/g(\mu)]\; p(\mu; x)\, dx/g(\mu) \right\} d\mu + o(c^{q/2}e^{-c}). \tag{3.15}$$

Since M^* defined in (2.4) is a compact subset of Λ and $g = \phi$ on $M^* \supset M_{\varepsilon^*}$, it follows that uniformly in $\mu \in U_{c^{-1/2}\log c,\, c^{-1/2}}$, $g(\mu) = \phi(\mu) + o(1)$ and therefore $\int_0^\infty p(\mu; x)e^{-x\phi(\mu)/g(\mu)}\, dx = \int_0^\infty p(\mu; x)e^{-x}\, dx + o(1) = \gamma(\mu) + o(1)$, by the definition of $p(\mu; x)$ and $\gamma(\mu)$. Hence, (3.15) can be written as

$$(1+o(1))(c/2\pi)^{d/2} e^{-c} \int_{U_{c^{-1/2}\log c,\, c^{-1/2}}} (\phi(\mu))^{-d/2-1} |\Sigma(\mu)|^{-1/2} \times \gamma(\mu) e^{-c[\phi(\mu)/g(\mu) - 1]}\, d\mu + o(c^{q/2}e^{-c}). \tag{3.16}$$

We next evaluate the integral in (3.16) by making use of (2.10). For $\mu \in U_{c^{-1/2}\log c,\, c^{-1/2}}$, we can write $\mu = y + z$ with $y \in M_{c^{-1/2}}$, $z \in TM^{\perp}_{c^{-1/2}}(y)$ and $\|z\| \leq c^{-1/2}\log c$; see (2.9). Since $\nabla\phi(y) = \nabla g(y)$ and $\phi(y) = g(y)$ for $y \in M_{c^{-1/2}}$, Taylor's expansion around $y \in M_{c^{-1/2}}$ yields

$$\phi(\mu) - g(\mu) \;=\; z'\nabla^2\rho(y)z/2 + o(\|z\|^2) \tag{3.17}$$

uniformly in $\mu \in U_{c^{-1/2}\log c,\, c^{-1/2}}$. Using the change of variables $z = \Pi_y^{\perp} v$ with $v \in \mathbf{R}^{d-q}$ and applying (2.10), we can express the integral in (3.16) as

$$\begin{aligned}
&(1+o(1)) \int_{M_{c^{-1/2}}} (\phi(y))^{-d/2-1} |\Sigma(y)|^{-1/2} \gamma(y) \\
&\qquad \times \int_{v\in\mathbf{R}^{d-q},\, \|v\|\leq c^{-1/2}\log c} \exp\{-cv'(\Pi_y^{\perp})'\nabla^2\rho(y)\Pi_y^{\perp} v/2g(y)\}\, dv\, d\sigma(y) \\
&\sim \int_{M_0} (\phi(y))^{-d/2-1} |\Sigma(y)|^{-1/2} \gamma(y) (2\pi g(y)/c)^{(d-q)/2} \\
&\qquad \times |(\Pi_y^{\perp})'\nabla^2\rho(y)\Pi_y^{\perp}|^{-1/2}\, d\sigma(y)
\end{aligned}$$

in view of (A4). Since $P\{T_c \leq ac\}$ is equal to (3.16), this proves Theorem 1 under the assumption that X_1 has a bounded continuous density.

To replace this assumption by the much weaker one that X_1 be nonlattice, we shall use a tilting argument and Lemma 1 instead of the saddlepoint approximation (2.6). Let $h_c = h_0([ac])$ and $K_x = K(x, h_c)$, where h_0 and $K(x, h)$ are given in Lemma 1. Let P_θ denote the probability measure under which $X_1, X_2, \ldots$ are i.i.d. with

$$P_\theta\{X_i \in A\} = \int_A e^{\theta' x - \psi(\theta)} dF(x) \tag{3.18}$$

for all Borel subsets A of $\mathbf{R}^d$; in particular, $P_0 = P$. Thus P_{θ_μ} is the distribution of $X_1^{(\mu)}, X_2^{(\mu)}, \ldots$. For $\mu \in h_c\mathbf{Z}^d$, where $\mathbf{Z}$ denotes the set of integers, let

$$\bar{f}_n(\mu) = P\{S_n \in K_{n\mu},\ \ jg(S_j/j) \le c \text{ for all } n-t \le j < n\}.$$

From the fact that $dP_{\theta_\mu}/dP = e^{n\phi(\mu)+\theta'_\mu(S_n-n\mu)}$, where P_{θ_μ} and P are restricted to the σ-field generated by $X_1, \ldots, X_n$, it follows that

$$\begin{aligned}\bar{f}_n(\mu) &= (1+o(1))e^{-n\phi(\mu)}\\ &\quad\times P_{\theta_\mu}\bigl\{S_n \in K_{n\mu},\ \ jg(S_j/j) \le c \text{ for all } n-t \le j < n\bigr\}\end{aligned}$$

uniformly in $\mu \in h_c\mathbf{Z}^d \cap U_{c^{-1/2}\log c},\, c_{-1/2}$ and $\delta c + t \le n \le ac$. In analogy with (3.12), note that

$$\begin{aligned}\sum_{\mu\in h_c\mathbf{Z}^d}\ \sum_{\delta c+t\le n\le ac} \bar{f}_n(\mu) I_{\{\inf_{\nu\in K_{n\mu}} ng(\nu/n)>c\}} &\le P\{T_c \le ac\}(1+o(1))\\ &\le \sum_{\mu\in h_c\mathbf{Z}^d}\ \sum_{\delta c\le n\le ac} \bar{f}_n(\mu)\, I_{\{\sup_{\nu\in K_{n\mu}} ng(\nu/n)>c\}}.\end{aligned} \tag{3.19}$$

Moreover, an argument similar to that in the Appendix to prove (3.12) can be used to show that replacing $\sum_{\mu\in h_c\mathbf{Z}^d}$ in (3.19) by $\sum_{\mu\in h_c\mathbf{Z}^d\cap U_{c^{-1/2}\log c}, c_{-1/2}}$ leads to an error of at most $o(c^{q/2}e^{-c})$. Define the events

$$\begin{aligned}\Omega_{n,\mu} &= \bigl\{(n-k)g((\nu - S_{n,k})/(n-k))\\ &\qquad \le c \text{ for all } \nu \in K_{n\mu} \text{ and } 1 \le k \le t,\ \|S_{n,t}\| \le c^{1/3}\bigr\},\\ \Omega^*_{n,\mu} &= \bigl\{(n-k)g((\nu - S_{n,k})/(n-k))\\ &\qquad \le c \text{ for some } \nu \in K_{n\mu} \text{and all } 1 \le k \le t,\ \|S_{n,t}\ \| \le c^{1/3}\bigr\}.\end{aligned}$$

Applying Lemma 1 to $P_{\theta_\mu}\{S_{n-t} \in K_{(n-t)\mu+\varepsilon\sqrt{n-t}}\}$, we obtain that uniformly in $\delta c \le n \le ac$ and $\mu \in U_{c^{-1/2}\log c,\, c^{-1/2}}$,

$$\begin{aligned}&P_{\theta_\mu}\{S_n \in K_{n\mu} |\ X_n, \ldots, X_{n-t+1}\}\\ &\qquad = (1+o(1))h_c^d(n/2\pi)^{d/2}|\Sigma(\mu)|^{-1/2}\end{aligned} \tag{3.20}$$

on $\Omega_{n,\mu}$ (or $\Omega^*_{n,\mu}$). Using the fact that $\sup_{\nu\in K_{n\mu}} n|g(\nu/n) - g(\mu)| = O(h_c)$, it can be shown by a Taylor expansion as in (3.13) that uniformly in $\delta c \le n \le ac$ and $\mu \in U_{c^{-1/2}\log c,\, c^{-1/2}} \cap h_c\mathbf{Z}^d$,

$$\begin{aligned}|P_{\theta_\mu}(\Omega_{n,\mu}) - p(\mu; ng(\mu) - c)| I_{\{\inf_{\nu\in K_{n\mu}} ng(\nu/n)>c\}} &= o(1),\\ |P_{\theta_\mu}(\Omega^*_{n,\mu}) - p(\mu; ng(\mu) - c)| I_{\{\sup_{\nu\in K_{n\mu}} ng(\nu/n)>c\}} &= o(1),\end{aligned} \tag{3.21}$$

where $p(\mu; x)$ is defined after (3.14). Combining (3.20) with (3.21), we obtain from (3.19) [in which $\sum_{\mu\in h_c\mathbf{Z}^d}$ is replaced by $\Sigma_{\mu\in h_c\mathbf{Z}^d\cap U_{c^{-1/2}\log c}, c_{-1/2}}$ with an error of $o(c^{q/2}e^{-c})$] that $P\{T_c \le ac\}$ is again equal to (3.15) so that we can proceed as before to obtain the desired conclusion. □

PROOF OF THEOREM 2. Without loss of generality, we shall assume that $r = 1$. Recall that $e_1(y), \dots, e_{d-p}(y)$ form an orthonormal basis of $TM^{\perp}(y)$ and that $\nabla g(y)$ is a scalar multiple of $e_1(y)$, for every $y \in M$. For $y \in M$ and $\max_{1 \le i \le d-p} |v_i| \le (\log n)^{-1}$, since $g(y) = b$ and $(\nabla g(y))' \sum_{i=1}^{d-p} v_i e_i(y) = v_1 \|\nabla \phi(y)\|/s$, Taylor's expansion yields

$$
\begin{aligned}
& g\left(y + \sum_{i=1}^{d-p} v_i e_i(y)\right) \\
& \quad = b + v_1 \|\nabla \phi(y)\|/s + O(v_1^2) + v' \Pi_y' \nabla^2 g(y) \Pi_y / 2 + o(\|v\|^2) \\
& \quad > b \text{ if } v_1 \|\nabla \phi(y)\|/s > c(v) + o(\|v\|^2) + O(v_1^2),
\end{aligned}
\tag{3.22}
$$

where $v = (v_2, \dots, v_{d-p})'$ and $c(v) = -v' \Pi_y' \nabla^2 g(y) \Pi_y v/2$. Let

$$
V_n = \left\{ y + \sum_{i=1}^{d-p} v_i e_i(y) \colon y \in M, \max_{1 \le i \le d-p} |v_i| \le (\log n)^{-1}, v_1 \|\nabla \phi(y)\|/s > c(v) \right\}.
$$

First assume that X_1 has a bounded continuous density. Then by (2.6), $P\{S_n/n \in V_n\}$ is equal to

$$
(1 + o(1))(n/2\pi)^{d/2} \int_{V_n} |\Sigma(\mu)|^{-1/2} e^{-n\phi(\mu)} d\mu. \tag{3.23}
$$

Note that $\nabla^2 \phi(y) = \Sigma^{-1}(y)$. Let $(a_{11}(y), \dots, a_{1,d-p}(y))$ be the first row of $J_y' \nabla^2 \phi(y) J_y$, where $J_y = (e_1(y) \dots e_{d-p}(y))$. For $y \in M$, $\phi(y) = b$ and Taylor's expansion yields

$$
\begin{aligned}
& \phi\left(y + \sum_{i=1}^{d-p} v_i e_i(y)\right) \\
& \quad = b + v_1 \|\nabla \phi(y)\| + \{a_{11}(y)/2 + o(1)\} v_1^2 + \sum_{i=2}^{d-p} \{a_{1i}(y) + o(1)\} v_1 v_i \\
& \quad \quad - sc(v) + v' \Pi_y' \left\{\Sigma^{-1}(y) - s \nabla^2 g(y)\right\} \Pi_y v/2 + o(\|v\|^2).
\end{aligned}
\tag{3.24}
$$

Putting (3.24) in (3.23) and using the change of variables

$$
u = n\{v_1 - sc(v)/\|\phi(y)\|\}, \qquad w = \sqrt{n} v,
$$

we can apply the infinitesimal change of volume function over tubular neighborhoods as in (2.10) to show that (3.23) is equal to

$$
\begin{aligned}
& (1+o(1))(n/2\pi)^{d/2} n^{-1-(d-p-1)/2} e^{-bn} \int_{y \in M} |\Sigma(y)|^{-1/2} \\
& \quad \times \left\{ \int_0^\infty \exp\left(-u \|\nabla \phi(y)\| - n^{-1} a_{11}(y) u^2/2 - n^{-1/2} u \sum_{i=2}^{d-p} a_{1i}(y) w_i\right) du \right\} \\
& \quad \times \left\{ \int_{\mathbf{R}^{d-p-1}} \exp\left(-w' \Pi_y' (\Sigma^{-1}(y) - s \nabla^2 g(y)) \Pi_y w/2\right) dw_2 \cdots dw_{d-p} \right\} d\sigma(y),
\end{aligned}
\tag{3.25}
$$

setting $\int_{\mathbf{R}^{d-p-1}} = 1$ if $d = p + 1$. Making use of (B1), (B4), (2.6) and the Taylor expansions (3.22) and (3.24), it can be shown that

$$P\{g(S_n/n) > b\} = P\{S_n/n \in V_n\} + o(n^{-q}e^{-bn}) \tag{3.26}$$

for every $q > 0$. Combining (3.26) with (3.23) and (3.25) yields the desired result when X_1 has a bounded continuous density. To prove the result for general nonlattice X_1, we can proceed as in the proof of Theorem 1 by using a tilting argument in conjunction with Lemma 1 to replace the saddlepoint approximation (2.6). □

PROOF OF THEOREM 3. First assume that X_1 has a continuous density and use the same notation as in the proof of Theorem 2. For $\mu = y + n^{-1}\{u + sc(w)/\|\nabla\phi(y)\|\} + \sum_{i=2}^{d-p} n^{-1/2} w_i e_i(y) \in V_n$, $ng(\mu) = bn + u\|\nabla g(y)\| + O(n^{-1/2})$ as in (3.22), and therefore analogous to (3.14),

$$\begin{aligned} &P\Big\{k\beta + (n-k)g(S_{n-k}/(n-k)) > bn \\ &\qquad\qquad \text{for all } 0 \le k < n \mid S_n/n = \mu\Big\} \\ &\quad = q(\mu; u\|\nabla\phi(y)\|) + o(1), \\ &\qquad \text{where } q(\mu; x) = P\Big\{\max_{k\ge 1} W_k(\mu) < x\Big\}. \end{aligned} \tag{3.27}$$

To derive (3.27), note that since $n\mu = S_n$ and $\theta_\mu = \nabla\phi(\mu) = s\nabla g(\mu)$,

$$\begin{aligned} &k\beta + (n-k)g(S_{n-k}/(n-k)) \\ &\quad \doteq k\beta + (n-k)g(\mu) + (\nabla g(\mu))'(S_{n-k} - (n-k)\mu) \\ &\quad \doteq k\beta + (n-k)g(\mu) - \theta'_\mu \sum_{i=n-k+1}^{n} (X_i - \mu)/s, \end{aligned}$$

which exceeds $ng(\mu) - u\|\nabla g(y)\| (\doteq bn)$ if and only if

$$k(g(\mu) - \beta) + \theta'_\mu \sum_{i=n-k+1}^{n} (X_i - \mu)/s < u\|\nabla\phi(y))\|/s.$$

The rest of the argument is similar to that of the proof of Theorem 2, except that the integral $\int_0^\infty$ in (3.25) is now replaced by

$$\begin{aligned} &\int_0^\infty \exp\bigg(-u\|\nabla\phi(y)\| - n^{-1}a_{11}(y)u^2/2 - n^{-1/2}u\sum_{i=2}^{d-p} a_{1i}(y)w_i\bigg) \\ &\quad \times q(y; u\|\nabla\phi(y)\|)\, du, \end{aligned}$$

which leads to the term $w(\mu)$ in (3.4). To prove the result for general nonlattice X_1, we can proceed as in the proof of Theorem 1. □

4. Moderate deviation approximations and nearly optimal sequential GLR tests. In the large deviation approximations to error probabilities of sequential tests in Section 3, we have considered events of the type $\{T_c \le K\}$ with K/c approaching a positive limit as $c \to \infty$. The asymptotic formula in Theorem 1 involves g and the rate function ϕ in the region $\{a^{-1} < g < \delta^{-1}\}$, and it no longer holds if $K/c \to \infty$. On the other hand, as $a \to \infty$ (or $a^{-1} \to 0$), Theorem 1 suggests that the asymptotic result would involve the behavior of g and ϕ in a neighborhood of μ_0, noting that ϕ assumes its minimum value 0 at μ_0. Since $\phi(\mu) \sim (\mu-\mu_0)'\Sigma^{-1}(\mu_0)(\mu-\mu_0)/2$ as $\mu \to \mu_0$, conditions of the type (A1) and (A2) would require the following: There exist $r > 0$, $1 \le q < d$ and an open neighborhood D of μ_0 such that

(C1) g is twice continuously differentiable on D and

$$g(\mu_0) = 0, \qquad \nabla g(\mu_0) = 0, \qquad \nabla^2 g(\mu_0) = rQ\{Q\Sigma(\mu_0)Q\}^{-1}Q,$$

(C2) $\sup_{\mu\in D, \mu\neq\mu_0} g(\mu)/\phi(\mu) = r$, and $M{:}\ = \{\mu \in D{:}\ g(\mu) = r\phi(\mu)\}$ is a smooth q-dimensional manifold,

where Q is a $d \times d$ projection matrix (symmetric and idempotent) so that Qx is the orthogonal projection of $x \in \mathbf{R}^d$ into a q-dimensional linear subspace L of $\mathbf{R}^d$, and the inverse of the singular matrix $Q\Sigma(\mu_0)Q$ refers to the Moore–Penrose generalized inverse [cf. Rao (1973)].

In this section we consider the case $K/c \to \infty$. Besides T_c, we also consider

$$N_C = \inf\{n \ge m_C{:}\, ng(S_n/n) \ge h(n/C)\}, \tag{4.1}$$

in which $C = e^c$ and $h(t) \sim \alpha \log t^{-1}$ as $t \to 0$ for some $\alpha > 0$. Note that if we replace $h(n/C)$ in (4.1) by a function $h(n, C)$ of both n and C, then T_c is indeed a special case of (4.1) with $h(n, C) = \log C$ $(= c)$. On the other hand, since the large deviation approximation in Theorem 1 restricts n to be $\le ac$ $(= a \log C)$, $\log(C/n)$ is asymptotically equivalent to $\log C$ and therefore there is little to gain by replacing the constant boundary c in T_c by $h(n/C)$. However, for the moderate deviation approximation considered in this section, we take n to be much larger than $c = \log C$ but still to be $o(C)$, for which $(n/C, (S_n - n\mu_0)/\sqrt{C})$ behaves like (t, B_t) and therefore $ng(S_n/n)$ behaves like $(r/2)(QB_t/\sqrt{t})'(Q\Sigma(\mu_0)Q)^{-1}QB_t/\sqrt{t}$ in view of (C2), where $\Sigma^{-1/2}(\mu_0)B_t$ is Brownian motion. This explains why we choose a time-varying boundary of the form $h(n/C)$ in (4.1).

In fact, such time-varying boundaries in N_C with $g(x)$ given by (3.5) that corresponds to GLR statistics for testing $H_0{:}\, \theta \in \Theta_0$ have been shown to yield asymptotically optimal sequential tests in exponential families from both Bayesian and frequentist viewpoints [cf. Lai (1988a, b, 1997) and Lai and Zhang (1994a, b)]. By considering optimal stopping problems associated with corresponding sequential testing problems for Brownian motion, Lai (1988a, 1997) showed that the optimal stopping boundaries satisfy an asymptotic expansion of the form

$$h(t) = \alpha \log t^{-1} - \beta \log\log t^{-1} + \lambda + o(1) \quad \text{as } t \to 0 \tag{4.2}$$

for some $\alpha > 0$ and real numbers β, λ that depend on the loss function for the wrong decision and the prior distribution, assuming a constant sampling cost per unit time. Part (i) of the following theorem gives an asymptotic approximation to the moderate deviation probability $P\{b^{-1}m \le N_C \le m\}$ as $m/\log C \to \infty$ but $m/C \to 0$ when h satisfies (4.2) and g satisfies (C1) and (C2). Part (ii) of the theorem gives an analogous result for T_c. Unlike Theorems 1–3, X_1 is no longer assumed to be nonlattice. Instead of integrating the saddlepoint approximation (2.6) over a tubular neighborhood of a manifold and using conditional arguments like (3.13) and (3.14) to handle excess over the boundary, its proof uses a change-of-measure argument that involves a mixture measure obtained by integrating P_{θ_μ} over a tubular neighborhood of μ_0. Additional remarks on this approach and some applications of the theorem will be given after its proof.

THEOREM 4. *Suppose that* $g: \Lambda \to \mathbf{R}$ *satisfies* (C1) *and* (C2). *Let* $b > 1$.

(i) *Suppose* h *satisfies* (4.2) *with* $\alpha > 0$. *Define* N_C *by* (4.1). *Then as* $C \to \infty$ *and* $m/\log C \to \infty$ *but* $m/C \to 0$ *with* $b^{-1}m \ge m_C$,

$$P\{b^{-1}m < N_C \le m\} \sim (m/C)^{\alpha/r} |\log(m/C)|^{\beta/r+q/2} \pi^{-q/2} (\alpha/r)^{q/2} e^{-\lambda/r} \times \int_{u \in \mathbf{R}^q : 1 \le \|u\|^2 \le b} \|u\|^{-2\alpha/r - q}\, du.$$

Consequently, if $m_C/\log C \to \infty$ *and* $m/m_C \to \infty$ *but* $m/C \to 0$ *as* $C \to \infty$, *then*

$$P\{N_C \le m\} \sim [(\alpha/r)^{q/2-1}/\Gamma(q/2)] e^{-\lambda/r} (m/C)^{\alpha/r} |\log(m/C)|^{\beta/r+q/2}.$$

(ii) *Define* T_c *by* (2.2), *with* $n_0/c \to \infty$ *as* $c \to \infty$. *Then uniformly in* $b^{-1}m \ge n_0$,

$$P\{b^{-1}m < T_c \le m\} \sim (\log b)(c/r)^{q/2} e^{-c/r}/\Gamma(q/2)$$

as $c \to \infty$. *Consequently, as* $m/n_0 \to \infty$ *but* $\log\log(m/n_0) = o(c)$,

$$P\{T_c \le m\} \sim (\log(m/n_0))(c/r)^{q/2} e^{-c/r}/\Gamma(q/2).$$

PROOF. Replacing (g, h) by $(g/r, h/r)$, we can assume without loss of generality that $r = 1$. We shall also assume that

$$\Sigma(\mu_0) = I_d, \qquad Q\{Q\Sigma(\mu_0)Q\}^{-1}Q = \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix}, \quad \mu_0 = 0, \tag{4.3}$$

where I_k denotes the $k \times k$ identity matrix. By using a nonsingular linear transformation of $\mu - \mu_0$, there is no loss of generality in assuming (4.3); see

Rao [(1973), 1c.3(ii)] and Chandra [(1985), page 101]. To prove part (i) of the theorem, let

$$U_\xi = \{\mu_1 + \mu_2\colon \mu_1 \in M, \mu_2 \in TM^\perp(\mu_1), \max(\|\mu_1\|^2, \|\mu_2\|^2) < \xi \alpha m^{-1} \log(C/m)\}. \tag{4.4}$$

Let $U = U_{4b}$, $U^* = U_{3b}$. Standard exponential tilting can be used to show that

$$\begin{aligned} &P\{\overline{X}_n \notin U^* \text{ for some } b^{-1}m \le n \le m\} \\ &\le \sum_{i=1}^d P\left\{ \max_{b^{-1}m \le n \le m} | \overline{X}_{ni} - \mu_0 | \ge (3\alpha b m^{-1} \log(C/m))^{1/2} \right\} \\ &= o((m/C)^{6\alpha/5}); \end{aligned} \tag{4.5}$$

see Lai and Zhang [(1994b), proof of Lemma 1] for a similar argument.

Let $\Omega_m = \{b^{-1}m < N_C \le m, \overline{X}_{N_C} \in U^*\}$. We shall evaluate $P(\Omega_m)$ by using a change of measures,

$$P(\Omega_m) = \int_{\Omega_m} L_{N_C}^{-1}\, d\widetilde{P}, \tag{4.6}$$

where $\widetilde{P}(B) = \int_U P_{\theta_\mu}(B)\, d\mu$ for all B in the σ-field generated by $\{X_n\colon n \le m\}$, P_{θ_μ} is defined in (3.18) and L_n is the mixture likelihood ratio

$$L_n = \int_U \exp(\theta_\mu' S_n - n\psi(\theta_\mu))\, d\mu. \tag{4.7}$$

Define the Kullback–Leibler information number

$$\mathscr{I}(\theta_\mu, \theta) = (\theta_\mu - \theta)'\mu - (\psi(\theta_\mu) - \psi(\theta)). \tag{4.8}$$

Letting $\hat\theta_n = (\nabla\psi)^{-1}(\overline{X}_n)$, we can write

$$\begin{aligned} \theta_\mu' \overline{X}_n - \psi(\theta_\mu) &= \phi(\overline{X}_n) - \{(\hat\theta_n - \theta_\mu)'\overline{X}_n - (\psi(\hat\theta_n) - \psi(\theta_\mu))\} \\ &= \phi(\overline{X}_n) - \mathscr{I}(\hat\theta_n, \theta_\mu). \end{aligned}$$

Moreover, since $\Sigma(\mu_0) = I_d$ and $\mu_0 = 0$, $\mathscr{I}(\hat\theta_n, \theta_\mu) \sim (\overline{X}_n - \mu)'(\overline{X}_n - \mu)/2$ as $\|\mu\| + \|\overline{X}_n\| \to 0$. Hence on Ω_m, as $C \to \infty$ and $m/\log C \to \infty$,

$$\begin{aligned} L_{N_C} &= e^{N_C \phi(\overline{X}_{N_C})} \int_U \exp(-N_C \mathscr{I}(\hat\theta_{N_C}, \theta_\mu))\, d\mu \\ &\sim (2\pi/N_C)^{d/2} e^{N_C g(\overline{X}_{N_C})} \exp\{N_C(\phi(\overline{X}_{N_C}) - g(\overline{X}_{N_C}))\}, \end{aligned} \tag{4.9}$$

recalling the definitions of U and U^*. Since $\nabla^2\phi(\mu_0) = \Sigma^{-1}(\mu_0) = I_d$, it follows from (C1), (C2) and (4.3) that for $\mu = \mu_1 + \mu_2 \in U$,

$$\phi(\mu) - g(\mu) = \|\mu_2\|^2/2 + o(\|\mu_2\|^2), \tag{4.10}$$

analogous to (3.17). Let $t_{\mu,m} - \{2\alpha \log(C/m)\}/\|\mu_1\|^2$ for $\mu = \mu_1 + \mu_2 \in U$.

We have the representation $\overline{X}_n = \overline{X}_n^{(1)} + \overline{X}_n^{(2)}$ if $\overline{X}_n \in U$, where $\overline{X}_n^{(1)} \in M$ and $\overline{X}_n^{(2)} \in TM^{\perp}(\overline{X}_N^{(1)})$. As $C \to \infty$ and $m/\log C \to \infty$ such that $m = o(C)$, it follows from (C1) and (4.3) that

$$N_C g(\overline{X}_{N_C}) = N_C\{\|\overline{X}_{N_C}^{(1)}\|^2/2 + o(\|\overline{X}_{N_C}\|^2)\}.$$

Combining this with (4.1), (4.2) and (4.5) yields that for every $\delta > 0$,

$$\sup_{\mu\in U} P_{\theta_\mu}\{|N_C/t_{\mu,m} - 1| \geq \delta\} \to 0. \tag{4.11}$$

Moreover, $\sum_{b^{-1}m\leq n\leq m} P_{\theta_\mu}\{|(n+1)\overline{X}_{n+1}^{(1)'}\overline{X}_{n+1}^{(1)} - n\overline{X}_n^{(1)'}\overline{X}_n^{(1)}| \geq \delta\} \to 0$ and therefore

$$\sum_{b^{-1}m\leq n\leq m} P_{\theta_\mu}\{|(n+1)g(\overline{X}_{n+1}) - ng(\overline{X}_n)| \geq \delta\} \to 0,$$

uniformly in $\mu \in U$. Hence

$$\sup_{\mu\in U} P_{\theta_\mu}\{N_C g(\overline{X}_{N_C}) - h(N_C/C) \geq \delta\} \to 0. \tag{4.12}$$

Let $\eta = 4b\alpha m^{-1}\log(C/m)$ and σ be the volume element measure of M. By (4.6) and (2.10), as $\eta \to 0$,

$$\begin{aligned} P(\Omega_m) &= \int_U E_{\theta_\mu}(L_{N_C}^{-1} I_{\Omega_m})\, d\mu \\ &\sim \int_{M\cap\{\|\mu_1\|^2\leq\eta\}} \int_{TM^{\perp}(\mu_1)\cap\{\|\mu_2\|^2\leq\eta\}} E_{\theta_\mu}(L_{N_C}^{-1} I_{\Omega_m})\, d\mu_2\, d\sigma(\mu_1). \end{aligned} \tag{4.13}$$

Note that $N_C g(\overline{X}_{N_C}) - h(N_C/C) \geq 0$. Let $B_m = \{y \in \mathbf{R}^q\colon 2\alpha\log(C/m) \leq m\|y\|^2 \leq 2\alpha b\log(C/m)\}$ and let $\tau_{y,m} = \{2\alpha\log(C/m)\}/\|y\|^2$ for $y \in B_m$. From (4.9)–(4.13), it follows that $P(\Omega_m)$ is asymptotically equivalent to

$$\begin{aligned} &(2\pi)^{-d/2} \int_{M\cap\{\|\mu_1\|^2\leq\eta\}} \int_{TM^{\perp}(\mu_1)\cap\{\|\mu_2\|^2\leq\eta\}} \\ &\quad\times E_{\theta_\mu}\{e^{-h(N_C/C)} N_C^{d/2} \exp[-(1/2 + o(1))N_C\|\overline{X}_{N_C}^{(2)}\|^2]\}\, d\mu_2\, d\sigma(\mu_1) \\ &\quad\sim (2\pi)^{-d/2} e^{-\lambda} \int_{B_m} (\tau_{y,m}/C)^{\alpha}\{\log(C/\tau_{y,m})\}^{\beta} \tau_{y,m}^{d/2} \\ &\quad\times \left\{\int_{z\in\mathbf{R}^{d-q},\|z\|^2\leq 4b\alpha m^{-1}\log(C/m)} \exp[-(1/2 + o(1))\tau_{y,m}\|z\|^2] dz\right\} dy. \end{aligned}$$

Using the change of variables $u = \sqrt{m}\{2\alpha\log(C/m)\}^{-1/2} y$, $v = \sqrt{m} z$ in the above double integral, it then follows that $P(\Omega_m)$ is asymptotically equivalent to

$$\begin{aligned} &(2\pi)^{-d/2} e^{-\lambda}(m/C)^{\alpha}(\log(C/m))^{\beta+q/2}(2\alpha)^{q/2} \int_{1\leq\|u\|^2\leq b} \|u\|^{-2\alpha-d} \\ &\quad\times \left\{\int_{\|v\|^2\leq 4b\alpha\log(C/m)} \exp[-(1 + O(1))\|v\|^2/(2\|u\|^2)] dv\right\} du \end{aligned}$$

$$\sim (2\pi)^{-d/2} e^{-\lambda} (m/C)^{\alpha} (\log(C/m))^{\beta+q/2} (2\alpha)^{q/2} (2\pi)^{(d-q)/2} \times \int_{1 \le \|u\|^2 \le b} \|u\|^{-2\alpha-q} du,$$

yielding the desired conclusion for $P\{b^{-1}m < N_C \le m\}$.

Let J be the positive integer defined by $b^{-(J+1)}m < m_C \le b^{-J}m$. Since

$$P\{N_C \le m\} = P\{b^{-1}m < N_C \le m\} + P\{b^{-2}m < N_C \le b^{-1}m\} + \cdots + P\{m_C \le N_C \le b^{-J}m\}$$

and since $m/m_C \to \infty$, the desired conclusion for $P\{N_C \le m\}$ then follows by choosing b arbitrarily large, noting that $\int_{\|u\|^2 \ge 1} \|u\|^{-2\alpha-q}\, du = \pi^{q/2}/(\alpha\Gamma(q/2))$.

The proof of part (ii) of the theorem is similar, replacing $\alpha \log(C/m)$ throughout the preceding argument by c. In particular, the bound in (4.5) now becomes $o(e^{-6c/5})$ and (4.11) holds with T_c in place of N_C and $t_{\mu,m} = 2c/\|\mu_1\|^2$. Letting $\tau_{y,m} = 2c/\|y\|^2$, the analogue of (4.13) in the present case is

$$\begin{aligned} &P\{b^{-1}m < T_c \le m,\ \overline{X}_{T_c} \in U^*\} \\ &\sim (2\pi)^{-d/2} e^{-c} \int_{B_m} \int_{z \in \mathbf{R}^{d-q}, m\|z\|^2 \le 4bc} \tau_{y,m}^{d/2} \exp[-(1+o(1))\tau_{y,m}\|z\|^2/2]\, dy\, dz \\ &\sim (2\pi)^{-d/2} e^{-c} (2c)^{q/2} \\ &\quad \times \int_{1 \le \|u\|^2 \le b} \|u\|^{-d} \int_{\|v\|^2 \le 4c} \exp[-(1+o(1))\|v\|^2/(2\|u\|^2)]\, dv\, du \\ &\sim \pi^{-q/2} c^{q/2} e^{-c} \int_{1 \le \|u\| \le \sqrt{b},\, u \in \mathbf{R}^q} \|u\|^{-q} du = 2(\log \sqrt{b}) c^{q/2} e^{-c}/\Gamma(q/2). \end{aligned}$$

Note that unlike the situation in (i), $\int_{\|u\|^2 \ge 1,\, u \in \mathbf{R}^q} \|u\|^{-q}\, du = \infty$. Taking $b = e$ in the preceding asymptotic formula, the desired conclusion for $P\{T_c \le m\}$ follows from

$$P\{T_c \le m\} = P\{e^{-1}m < T_c \le m\} + \cdots + P\{n_0 \le T_c \le b^{-J}m\},$$

where $e^{-(J+1)}m < n_0 \le e^{-J}m$ so that $J \sim \log(m/n_0)$. □

The moderate deviation approximation in Theorem 4(ii) is a diffusion approximation that approximates $(n/m, (S_n - n\mu_0)/\sqrt{m})$ by (t, B_t), where $\Sigma^{-1/2}(\mu_0)B_t$ is Brownian motion, for which

$$\begin{aligned} &P\{(QB_t)'(Q\Sigma(\mu_0)Q)^{-1}QB_t \ge 2ct \text{ for some } b^{-1} \le t \le 1\} \\ &\quad \sim (\log b) c^{q/2} e^{-c}/\Gamma(q/2) \end{aligned} \tag{4.14}$$

as $c \to \infty$. (If c were fixed as $m \to \infty$, then the diffusion approximation would be a consequence of the functional central limit theorem.) The same change-of-measure argument involving a suitably chosen mixture measure can be used to prove (4.14) for Brownian motion, showing the versatility of the method for such moderate deviation calculations. On the other hand, the method of Section 3, which cannot be applied to Brownian motion, is also not directly applicable to moderate deviation probabilities. As noted in the second paragraph of this section, the moderate deviation approximation in Theorem 4(i) is a diffusion approximation that approximates $(n/C, (S_n - n\mu_0)/\sqrt{C})$ by (t, B_t). Putting $\varepsilon = m/C$ in Theorem 4(i) suggests the following counterpart for Brownian motion:

$$\begin{aligned} &P\{tg(B_t/t) \geq h(t) \text{ for some } t \leq \varepsilon\} \\ &\quad \sim [(\alpha/r)^{q/2-1}/\Gamma(q/2)]e^{-\lambda/r}\varepsilon^{\alpha/r}|\log \varepsilon|^{\beta/r+q/2} \end{aligned} \tag{4.15}$$

as $\varepsilon \to 0$, which can be proved by the same method used to prove Theorem 4(i).

5. Asymptotic expansions for GLR and perturbed χ^2 statistics. In this section we consider fixed sample size tests and give a moderate deviation approximation to the probability $P\{2ng(S_n/n) \geq rc\}$, where g satisfies (C1) and (C2) and $c \to \infty$ but $c = o(n)$. The moderate deviation probabilities of $2ng(S_n/n)$ can be approximated by those of the chi-square distribution with q degrees of freedom, in contrast with (3.8) and (3.9) in which the large deviation probabilities of $2ng(S_n/n)$ and of the chi-square distribution differ by a constant factor. The chi-square approximation can be derived by a tilting argument similar to that used in the proof of Theorem 1. Moreover, if $c = o(n^{1/3})$ and X_1 satisfies Cramér's condition while g is sufficiently smooth in a neighborhood of μ_0, then we can refine the chi-square approximation into asymptotic expansions for the moderate deviation probabilities. For the case where c is assumed to be fixed (instead of tending to ∞ with n), Chandra (1985) derived an Edgeworth expansion [under (C1)] for $P\{2ng(S_n/n) \geq rc\}$ as $n \to \infty$.

THEOREM 5. *Suppose that* g: $\Lambda \to \mathbf{R}$ *satisfies* (C1). *Let* $1 \leq q < d$ *and let* $\chi^2(q)$ *denote a chi-square random variable with* q *degrees of freedom.*

(i) *Suppose* (C2) *also holds. Then as* $n \to \infty$ *and* $c \to \infty$ *such that* $c = o(n)$,

$$P\{2ng(S_n/n) \geq rc\} \sim P\{\chi^2(q) \geq c\}. \tag{5.1}$$

(ii) *Suppose* g *is of class* C^{2J+2} *in a neighborhood of* μ_0 *for some positive integer* J. *Let* $\phi_Q(\mu) = (\mu - \mu_0)' Q\{Q\Sigma(\mu_0)Q\}^{-1}Q(\mu - \mu_0)/2$, *where* Q *is a* $d \times d$ *projection matrix so that* Qx *is the orthogonal projection of* $x \in \mathbf{R}^d$ *into a* q*-dimensional linear subspace of* $\mathbf{R}^d$. *Suppose there exists a polynomial* $p(\mu)$ *in the components of* $\mu - \mu_0$ *such that* $p(\mu_0) = 1$ *and*

$$\begin{aligned} g(\mu) &= rp(\mu)\phi_Q(\mu) + O(\|\phi_Q(\mu - \mu_0)\|^2\|\mu - \mu_0\|^{2J+1}), \\ \|\nabla g(\mu) - (\nabla pq)(\mu)\| &= O(\|Q(\mu - \mu_0)\|\|\mu - \mu_0\|^{2J+1}) \end{aligned} \tag{5.2}$$

as $\mu \to \mu_0$. *Assume furthermore Cramér's condition*

$$\limsup_{\substack{y \in L \\ \|y\| \to \infty}} |E\exp(\sqrt{-1}y' Q(X_1 - \mu_0))| < 1. \tag{5.3}$$

Take any $c_0 > 0$ *and* $c_n \to \infty$ *such that* $c_n = o(n^{1/3})$. *Then there exist constants* a_{ij} *and nonnegative integers* k_j $(j = 1, \dots, J)$ *such that as* $n \to \infty$,

$$P\{2ng(S_n/n) \geq rc\} = P\{\chi^2(q) \geq c\} + \sum_{j=1}^{J} n^{-j} \sum_{i=0}^{k_j} a_{ij} P\{\chi^2(q+i) \geq c\} + o(n^{-J} P\{\chi^2(q+J) \geq c\})$$

uniformly in $c_0 \leq c \leq c_n$.

PROOF. Without loss of generality, we shall assume that $r = 1$ and that (4.3) holds. Moreover, we shall also assume that $\mu_0 = 0$ for notational simplicity.

To prove (i), first consider the case in which X_1 is nonlattice. Let $h_n = h_0(n)$ and $K_x = K(x, h_n)$, where h_0 and $K(x, h)$ are given in Lemma 1, and define the probability measure P_μ by (3.18). Let U be the tubular neighborhood of $M \cap \{\mu\colon \|\mu\| < (c/n)^{1/3}\}$ with radius $(c/n)^{1/3}$. By Lemma 1 applied to P_μ together with a tilting argument similar to that used in the proof of Theorem 1, we obtain that as $n \to \infty$ and $c/n \to 0$,

$$P\{S_n \in K_{n\mu}\} = (1 + o(1)) e^{-n\phi(\mu)} (n^{-1} h_n)^d (n/2\pi)^{d/2}$$

uniformly in $\mu \in n^{-1} h_n \mathbf{Z}^d \cap U$, noting that $|\Sigma(\mu)| = 1 + o(1)$ by (4.3). Hence

$$\sum_{\substack{\mu \in n^{-1} h_n \mathbf{Z}^d \cap U \\ 2ng(\mu) \geq c}} P\{S_n \in K_{n\mu}\} \sim (n/2\pi)^{d/2} \int_U e^{-n\phi(\mu)} I_{\{2ng(\mu) \geq c\}} d\mu. \tag{5.4}$$

The integral in (5.4) can be evaluated by making use of (2.10) and an argument similar to that in the proof of Theorem 4 to analyze the double integral in (4.13), together with the same composite transformation as in Chandra and Ghosh (1979):

$$\begin{aligned} (\sqrt{n} y_1, \dots, \sqrt{n} y_q, \sqrt{n} z') &\to (\sqrt{n} \|y\|, \theta_1, \dots, \theta_{q-1}, v') \\ &\to (2ng(\mu), \theta_1, \dots, \theta_{q-1}, v'), \end{aligned} \tag{5.5}$$

in which $y = (y_1, \dots, y_q)'$ and z, v belong to $\mathbf{R}^{d-q}$. The first transformation in (5.5) is simply the polar coordinate representation of $\sqrt{n} y$. Letting $w = 2ng(\mu)$, we can therefore write $e^{-n\phi(\mu)} = e^{-w/2} \exp\{-(\frac{1}{2} + o(1)) \|v\|^2\}$. Making use of the Jacobian of the transformation, we then obtain that the right-hand side of (5.4) is asymptotically equivalent to

$$\int_c^\infty w^{q/2-1} e^{-w/2} \, dw / \{2^{q/2} \Gamma(q/2)\} = P\{\chi^2(q) \geq c\}.$$

Since $P\{\overline{X}_n \notin U\} = \exp\{-(1/2 + o(1))(n/c)^{1/3} c\}$, (5.1) follows. In the case where X_1 is lattice with span h, we can replace h_n by h in the preceding argument and (5.4) still holds.

To prove (ii) under Cramér's condition (5.3), we first assume that X_1 has a bounded continuous density. Then the density function of S_n has a tilted

Edgeworth expansion [cf. Section 4.3 and the Appendix of Barndoff-Nielsen and Cox (1979)] and we can apply the change of variables (5.5) to this saddlepoint (tilted Edgeworth) expansion. The same calculations as in Chandra and Ghosh (1979) for the direct Edgeworth expansion then yield the desired conclusion, noting that since $c = o(n^{1/3})$, the Taylor expansion of $\phi(\mu)$ in $e^{-n\phi(\mu)}$ combined with the Taylor expansion of $|\Sigma(\mu)|^{-1/2}$ basically reduces the saddlepoint expansion to the Edgeworth expansion. □

6. Computational issues and concluding remarks. The functions $\gamma(\mu)$ in Theorem 1 and $w(\mu)$ in Theorem 3 can be computed by making use of the fluctuation theory of random walks; see Siegmund [(1985), Chapter VIII] and Woodroofe [(1982), Chapter 2], where these functions are expressed either as infinite series involving the marginal distributions of $S_n(\mu)$ [or $W_n(\mu)$] or integrals involving the characteristic function of $S_1(\mu)$ [or $W_1(\mu)$]. When M_0 (or M) is of low dimension, standard numerical integration techniques can be used to compute the integrals that appear as multiplicative constants in the asymptotic approximations of Theorems 1–3. On the other hand, numerical integration is computationally expensive for higher dimensions, particularly in view of the fact that $\gamma(\mu)$ or $w(\mu)$ does not have closed-form expressions and requires an extra layer of numerical integration (or series summation). For multidimensional integrals, Monte Carlo or quasi-Monte Carlo methods are widely used numerical methods. Instead of using Monte Carlo methods to evaluate the integrals that appear in the large deviation approximations in Theorems 1–3, it is more natural to consider Monte Carlo simulation of these probabilities. However, since these (large deviation) probabilities are very small, direct simulation can be prohibitively difficult and importance sampling is needed. Specifically, for any stopping rule τ (which also includes the fixed sample size n in Theorem 2), Wald's likelihood ratio identity yields

$$P(A) = E_Q(L_\tau^{-1} I_A) \quad \text{for all } A \in \mathscr{F}_\tau, \tag{6.1}$$

where $\mathscr{F}_n$ is the σ-field generated by $X_1, \dots, X_n$ and L_n is the Radon–Nikodyn derivative (likelihood ratio) of $Q|_{\mathscr{F}_n}$ with respect to $P|_{\mathscr{F}_n}$. The basic idea behind importance sampling is to choose the measure Q suitably so that $L_\tau^{-1} I_A$ has a smaller variance under Q than that of I_A under P.

For the event A considered in Theorem 2 or Theorem 3, our asymptotic formula and its proof suggest choosing Q as the mixture $\int_M P_{\theta_\mu} d\sigma(\mu)$. Since σ has compact support, $\sigma(M)$ is finite. We can therefore choose $Q = \int_M P_{\theta_\mu} d\tilde{\sigma}(\mu)$, where $\tilde{\sigma} = \sigma/\sigma(M)$ is a probability measure on M. Similarly, for the event A considered in Theorem 1, define $Q = \int_{M_0} P_{\theta_\mu} d\tilde{\sigma}(\mu)$. Corresponding to this choice of Q, the mixture likelihood ratio L_n in (6.1) can be expressed as

$$L_n = \int \exp(\theta_\mu' S_n - n\psi(\theta_\mu))\, d\tilde{\sigma}(\mu), \tag{6.2}$$

where the integral is over M_0 or M. The integral in (6.2) may be difficult to compute when M_0 or M is a high-dimensional manifold. To get around the numerical integration involved in (6.1), we can treat μ as a random variable with distribution $\tilde{\sigma}$ and replace (6.1) by

$$P(A) = E_{\tilde{\sigma}}\{E_{\theta_\mu}[\exp(-\theta'_\mu S_\tau + \tau\psi(\theta_\mu))I_A]\}, \tag{6.3}$$

which suggests the following Monte Carlo method to compute $P(A)$:

(i) Generate μ from the distribution $\tilde{\sigma}$, treating it as a random variable.

(ii) Given μ, generate i.i.d. d-dimensional vectors $X_1, X_2, \dots, X_\tau$ from P_{θ_μ}.

(iii) Compute $P(A)$ as the average $B^{-1}\sum_{i=1}^{B} I_i \exp\{-\theta'_{\mu(i)} S^{(i)}_{\tau(i)} + \tau(i)\times \psi(\theta_{\mu(i)})\}$ over B independent samples $(\mu(i), X_1^{(i)}, \dots, X^{(i)}_{\tau(i)})$, $i = 1, \dots, B$, where $S^{(i)}_{\tau(i)} = X_1^{(i)} + \cdots + X^{(i)}_{\tau(i)}$ and $I_i = 1$ if A occurs for the ith sample and $I_i = 0$ otherwise.

As explained in Chan and Lai (1999), it is sometimes useful to decompose the event A into a union of disjoint sets $A_1, \dots, A_k$ and to evaluate $P(A)$ as a sum $\sum_{j=1}^{k} P(A_j)$, so that the manifold $M^{(j)}$ associated with A_j is a connected subset of $\mathbf{R}^d$ and is considerably smaller than M (or M_0). Further details of this importance sampling procedure are given in Chan and Lai (1999), where a theory similar to that of Siegmund (1976) for the sequential probability ratio test is also developed. The following two examples illustrate the implementation of this procedure and use it as a benchmark to compare the numerical results computed from the analytic approximations of Theorems 1 and 2.

EXAMPLE 2 (Continued). Woodroofe (1979) used an alternative approach to derive an asymptotic approximation to the Type I error probability of the sequential GLR test of the equality of two exponential means. When $\lambda_1 = \lambda_2 = 1$, his asymptotic approximation has the form

$$\sqrt{c}\, e^{-c} \int_{a^{-1}}^{\delta^{-1}} t^{-1}\kappa(t)\, dt, \tag{6.4}$$

where $t = -\log(y(2-y))$, $u = (1-e^{-t})^{1/2}$ and $\kappa(t) = \{\pi t^{-1} u^2(1+u^2)\}^{-1/2}\times \nu(y, 2-y)$. He has also shown that $\nu(y, 2-y) = (1+u)^{-1}\{2t^{-1}(1-e^{-t}) - e^{-t}\}$. Comparing (6.4) with (3.7) shows that Woodroofe's formula has an extra factor $(1+u^2)^{-1/2} = (2-e^{-t})^{-1/2} = (2-y(2-y))^{-1/2}$ in his integrand. This explains the discrepancy between his results and ours in Table 1, where we take $\delta_0 = (n_0 - \frac{1}{2})/c$ and $a = (N + \frac{1}{2})/c$ to introduce a continuity correction in (3.7) and (6.4) for greater accuracy. Also given in Table 1 are the Monte Carlo estimates of the Type I error probability using importance sampling (mean $\pm$ standard error) and the direct Monte Carlo estimates. Each Monte Carlo estimate is based on 10,000 simulation runs. Table 1 shows that (3.7) is markedly nearer to the importance sampling result than Woodroofe's approximation (6.4). It also shows the advantage of importance sampling over direct Monte Carlo, especially when the actual probability is very small, for which

TABLE 1
Asymptotic approximations and Monte Carlo estimates of $P\{n_0 \le T_c \le N\}$

c	n_0	N	(6.4)	(3.7)	Importance sampling	Direct MC
5	5	25	8.50×10^{-3}	9.96×10^{-3}	$(11.23 \pm 0.21) \times 10^{-3}$	$(11.40 \pm 1.06) \times 10^{-3}$
5	5	50	1.26×10^{-2}	1.44×10^{-2}	$(1.59 \pm 0.02) \times 10^{-2}$	$(1.63 \pm 0.13) \times 10^{-2}$
5	5	100	1.72×10^{-2}	1.91×10^{-2}	$(1.94 \pm 0.03) \times 10^{-2}$	$(2.17 \pm 0.15) \times 10^{-2}$
5	3	50	1.49×10^{-2}	1.74×10^{-2}	$(1.81 \pm 0.03) \times 10^{-2}$	$(1.72 \pm 0.13) \times 10^{-2}$
5	13	50	7.99×10^{-3}	8.70×10^{-3}	$(10.56 \pm 0.21) \times 10^{-3}$	$(9.10 \pm 0.95) \times 10^{-3}$
10	10	50	7.84×10^{-5}	9.16×10^{-5}	$(9.57 \pm 0.16) \times 10^{-5}$	0
10	10	100	1.18×10^{-4}	1.34×10^{-4}	$(1.41 \pm 0.02) \times 10^{-4}$	$(1.00 \pm 1.00) \times 10^{-4}$
10	10	200	1.62×10^{-4}	1.79×10^{-4}	$(1.83 \pm 0.02) \times 10^{-4}$	$(2.00 \pm 1.41) \times 10^{-4}$
10	5	100	1.45×10^{-4}	1.65×10^{-4}	$(1.71 \pm 0.03) \times 10^{-4}$	$(1.00 \pm 1.00) \times 10^{-4}$
10	25	100	7.68×10^{-5}	8.37×10^{-5}	$(9.20 \pm 0.16) \times 10^{-5}$	0

the standard error of direct Monte Carlo is too large relative to the actual probability. Following Chan and Lai (1999), we have decomposed $A = \{T_c \le N\}$ prior to performing importance sampling into 10 subsets of the form

$$A_j = \left\{T_c \le N,\ \alpha_{j-1} \le \left(\sum_{i=1}^{T_c} Z_i\right) \Big/ \left(\sum_{i=1}^{T_c} Y_i\right) < \alpha_j\right\},$$

$$A_{j+5} = \left\{T_c \le N,\ \alpha_{j-1} \le \left(\sum_{i=1}^{T_c} Y_i\right) \Big/ \left(\sum_{i=1}^{T_c} Z_i\right) < \alpha_j\right\} \quad \text{for } 1 \le j \le 5,$$

where $\alpha_0 = 0$, $\alpha_j = (1-\beta_j)/(1+\beta_j)$ and

$$\beta_j = (j/5)(1-e^{-1/a})^{1/2} + (1-j/5)(1-e^{-1/\delta})^{1/2} \qquad 1 \le j \le 5.$$

EXAMPLE 3 (Continued). Table 2 compares the asymptotic formula (3.8) with the result computed by Monte Carlo using importance sampling after

TABLE 2
Asymptotic approximations and Monte Carlo estimates of $P\{g(S_n/n) > b\}$

n	b	(6.5)	(3.8)	Importance sampling	Direct MC
50	0.1	1.67×10^{-3}	1.74×10^{-3}	$(1.58 \pm 0.03) \times 10^{-3}$	$(1.90 \pm 0.44) \times 10^{-3}$
50	0.2	7.83×10^{-6}	8.51×10^{-6}	$(8.28 \pm 0.18) \times 10^{-6}$	0
50	0.5	1.50×10^{-12}	1.77×10^{-12}	$(1.70 \pm 0.05) \times 10^{-12}$	0
50	1	1.52×10^{-23}	1.94×10^{-23}	$(1.90 \pm 0.06) \times 10^{-23}$	0
50	2	2.34×10^{-45}	3.19×10^{-45}	$(3.37 \pm 0.11) \times 10^{-45}$	0
100	0.1	7.93×10^{-6}	8.30×10^{-6}	$(7.80 \pm 0.17) \times 10^{-6}$	0
100	0.2	2.51×10^{-10}	2.73×10^{-10}	$(2.89 \pm 0.07) \times 10^{-10}$	0
100	0.5	1.47×10^{-23}	1.73×10^{-23}	$(1.73 \pm 0.06) \times 10^{-23}$	0
100	1	2.07×10^{-45}	2.64×10^{-45}	$(2.46 \pm 0.09) \times 10^{-45}$	0
100	2	6.15×10^{-89}	8.40×10^{-89}	$(8.33 \pm 0.33) \times 10^{-89}$	0

decomposing $A = \{g(S_n/n) > b\}$ into two subsets of the form

$$A_1 = A \cap \{(\Sigma_1^n Y_i)/(\Sigma_1^n Z_i) > (2-\tilde{y})/\tilde{y}\},$$
$$A_2 = A \cap \{(\Sigma_1^n Y_i)/(\Sigma_1^n Z_i) \le (2-\tilde{y})/\tilde{y}\},$$

so that their associated manifolds are $M^{(1)} = \{(2-\tilde{y}, \tilde{y})\}$ and $M^{(2)} = \{(\tilde{y}, 2-\tilde{y})\}$. Example 1 of Woodroofe (1978) derived from his Theorem 1 on large deviation probabilities of GLR statistics the following formula for this example:

$$P\{g(S_n/n) > b\} \sim \{\pi b^{-1}(1-e^{-b})(2-e^{-b})\}^{-1/2}(nb)^{-1/2}e^{-nb}, \tag{6.5}$$

which differs from (3.8) by an extra factor of $(2-e^{-b})^{-1/2}$. Table 2 also compares (6.5) with the exact values computed by Monte Carlo using importance sampling, showing that this extra factor should not be included. Also given in Table 2 are the direct Monte Carlo estimates. Each Monte Carlo result (mean $\pm$ standard error) is based on 10,000 simulations.

APPENDIX: PROOF OF (3.1) AND (3.12)

PROOF OF (3.1). Fix $y \in M$. Then $g(y) = b$ and $\phi(y) = b/r$. For all sufficiently small $\varepsilon > 0$, $g(y+\varepsilon\nabla g(y)) = b+\varepsilon\|\nabla g(y)\|^2 + O(\varepsilon^2) > b$, and therefore by (B1),

$$b/r \le \phi(y+\varepsilon\nabla g(y)) = b/r + \varepsilon(\nabla g(y))'\nabla\phi(y) + O(\varepsilon^2),$$

showing that $(\nabla g(y))'\nabla\phi(y) \ge 0$. By (B3), $\nabla g(y) \ne 0$. We next show that $\inf_{\mu\in M}\|\nabla\phi(\mu)\| > 0$. In view of (B2), M is a compact subset of Λ^o and therefore it suffices to show that $\nabla\phi(\mu) \ne 0$ for all $\mu \in M$, which follows from the fact that $\nabla\phi$ is a diffeomorphism with $\nabla\phi(\mu_0) = 0$ and $\phi(\mu_0) = 0$ (so $\mu_0 \notin M$).

Suppose $\nabla\phi(y) \notin TN^{\perp}(y)$. Then there exists $z \in TN(y)$ such that $z'\nabla\phi(y) < 0$. Since $z'\nabla g(y) = 0$,

$$g(y+\varepsilon\nabla g(y)+\varepsilon^{2/3}z) = b + \varepsilon\|\nabla g(y)\|^2 + O(\varepsilon^{4/3}) > b$$

for all sufficiently small $\varepsilon > 0$. Therefore by (B1), $\phi(y+\varepsilon\nabla g(y)+\varepsilon^{2/3}z) \ge b/r$, which contradicts

$$\phi(y+\varepsilon\nabla g(y)+\varepsilon^{2/3}z) = b/r + \varepsilon(\nabla g(y))'\nabla\phi(y) + \varepsilon^{2/3}z'\nabla\phi(y) + O(\varepsilon^{4/3}) < b/r.$$

□

PROOF OF (3.12). Since (A1) holds with $r = 1$, $g \le \phi$ on $\{a^{-1} < g < \delta^{-1}+\varepsilon_0\}$. Moreover, $\phi > \delta^{-1}+\varepsilon_1$ on $\{g > \delta^{-1}+\varepsilon_0\}$ by (A3). Since

$$P\{T_c \le ac\} = \sum_{\delta c \le n \le ac} P\big\{ng(S_n/n) > c,\ mg(S_m/m) \le c \quad \text{for all } \delta c \le m < n\big\},$$

it then follows that for all large c,

$$\begin{aligned}
&\Big| P\{T_c \le ac\} - \int_{\mathbf{R}^d} \sum_{\substack{n>c/g(\mu)\\ \delta c+t\le n\le ac}} f_n(\mu)\, d\mu \Big| \\
&\quad \le \sum_{\delta c\le n\le \delta c+t} P\{ng(S_n/n) > c\} \\
&\qquad + \sum_{\delta c\le k\le ac} P\{k\phi(S_k/k) > c + c^{1/5}\} \\
&\qquad + \sum_{\delta c+t\le n\le ac} P\{n\phi(S_n/n) > c \text{ and } c + c^{1/5} \\
&\quad \ge k\phi(S_k/k) > c \text{ for some } \delta c \le k \le n-t\}.
\end{aligned} \tag{A.1}$$

Making use of the saddlepoint approximation (2.6) (recalling that X_1 is assumed to have a bounded continuous density in this part of the proof of Theorem 1), it can be shown that

$$\sum_{\delta c\le k\le ac} P\{k\phi(S_k/k) > c + c^{1/5}\} = o(e^{-c}). \tag{A.2}$$

The last term in (A.1) can be expressed as the left-hand side of

$$\begin{aligned}
&\sum_{\delta c+t\le n\le ac}\ \sum_{\delta c\le k\le n-t} \int_{c<k\phi(\mu)\le c+c^{1/5}} P\{n\phi(n^{-1}S_n) > c | k^{-1}S_k = \mu\} \\
&\qquad \times P\{k^{-1}S_k \in d\mu\} = o(e^{-c}).
\end{aligned} \tag{A.3}$$

To prove (A.3), first note that by (A3) there exists $b > \delta^{-1}$ such that $B := \{\mu : \phi(\mu) \le b\}$ is a compact subset of Λ^o. From large deviation bounds [cf. Dembo and Zeitouni (1998)], it follows that $\sum_{n\ge k\ge \delta c} P\{n^{-1}S_n \notin B$ or $k^{-1}S_k \notin B\} = o(e^{-c})$. In view of (2.6), it suffices to show that for every $p > 1$,

$$\begin{aligned}
&\sup\big\{P[n^{-1}S_n \in B,\ n\phi(n^{-1}S_n) > c | k^{-1}S_k = \mu] : \\
&\qquad \mu \in B,\ n-k \ge t, k \ge \delta c,\ k\phi(\mu) \le c + c^{1/5}\big\} = o(c^{-p}).
\end{aligned} \tag{A.4}$$

Without loss of generality we shall assume that $\mu_0 = 0$. Note that

$$\begin{aligned}
&P\{n^{-1}S_n \in B,\ n\phi(n^{-1}S_n) > c | k^{-1}S_k = \mu\} \\
&\qquad = P\{Y_\mu \in B,\ n\phi(Y_\mu) > c\},
\end{aligned} \tag{A.5}$$

where $Y_\mu = (k\mu + \sum_{i=k+1}^n X_i)/n$. Moreover, $\mu - Y_\mu = (nY_\mu - \sum_{i=k+1}^n X_i)/k - Y_\mu = k^{-1}(n-k)Y_\mu - k^{-1}\sum_{i=k+1}^n X_i$, and $Y_\mu' \nabla\phi(Y_\mu) = Y_\mu' \theta_{Y_\mu} = \phi(Y_\mu) + \psi(\theta_{Y_\mu})$ by (2.1). Since ϕ is convex,

$$\begin{aligned}
\phi(\mu) &\ge \phi(Y_\mu) + (\mu - Y_\mu)' \nabla\phi(Y_\mu) \\
&= \phi(Y_\mu) + k^{-1}(n-k)\{\phi(Y_\mu) + \psi(\theta_{Y_\mu})\} \\
&\qquad - k^{-1}\bigg(\sum_{i=k+1}^n X_i\bigg)' \nabla\phi(\theta_{Y_\mu}).
\end{aligned} \tag{A.6}$$

Since $\psi(0) = 0$ and $\nabla\psi(0) = \mu_0 = 0$ and ψ is convex, $\psi_{\min} := \inf\{\psi(\theta_y) : a^{-1} \le \phi(y) \le b\} > 0$. It then follows from (A.6) that

$$k\phi(\mu) \ge n\phi(Y_\mu) + (n-k)\psi_{\min} - \left\| \sum_{i=k+1}^{n} X_i \right\| \max_{y \in B} \|\nabla\phi(y)\| \quad \text{on } \{Y_\mu \in B\}.$$

From this and (A.5), it follows that the left-hand side of (A.4) is majorized by

$$\sup_{n-k \ge c^{1/4}} P\left\{ \left\| \sum_{i=k+1}^{n} X_i \right\| \max_{y\in B} \|\nabla\phi(y)\| \ge (n-k)\psi_{\min} - c^{1/5} \right\},$$

which is $o(\exp(-\zeta c^{1/4}))$ for some $\zeta > 0$, since $EX_i = \mu_0 = 0$, proving (A.4).

Noting (A.2) and that $f_n(\mu)\, d\mu \le P\{S_n/n \in d\mu\}$, we next show that

$$\text{(A.7)} \quad \sum_{\delta c \le n \le ac} P\{c + c^{1/5} \ge ng(S_n/n) > c,\ n^{-1}S_n \notin U_{c^{-1/2}\log c,\, c^{-1/2}}\} = o(e^{-c}).$$

From (A5) together with continuity and compactness arguments [noting that (2.4) is compact], it follows that there exist $\alpha > 0$ and $\varepsilon > 0$ (sufficiently small) such that $\liminf_{\mu\to\partial\Lambda} \phi(\mu) > \delta + 2\varepsilon$ and $\beta := \inf_{\eta\in U_{\alpha,\varepsilon}} \lambda_{\min}(\nabla^2\rho(\eta))/2 > 0$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a nonnegative definite matrix. Analogous to (3.17), for every $\mu\, (= y+z) \in U_{\alpha,\varepsilon}$, there exists $\eta_\mu \in U_{\alpha,\varepsilon}$ such that $\phi(\mu) - g(\mu) = z'\nabla^2\rho(\eta_\mu)z/2$ and therefore $\phi(\mu) \ge g(\mu) + \beta\|z\|^2$. Moreover, $\inf_{\mu\notin U^o_{\alpha,\varepsilon}, a^{-1}\le g(\mu)\le\delta^{-1}+\varepsilon}(\phi(\mu) - g(\mu)) > 0$ by (A1) and compactness arguments. Hence the saddlepoint approximation (2.6) yields

$$\text{(A.8)} \quad \sum_{\delta c \le n \le ac} P\{c + c^{1/5} \ge ng(S_n/n) > c,\ n^{-1}S_n \notin U_{\alpha,\, c^{-1/2}}\} = o(e^{-c}),$$

noting that $c + c^{1/5} \ge ng(\mu) > c$ and $\delta c \le n \le ac \Rightarrow \delta^{-1} + \delta^{-1}c^{-4/5} \ge g(\mu) > a^{-1}$. It also yields

$$\text{(A.9)} \quad \sum_{\delta c \le n \le ac} P\{n^{-1}S_n \in U_{\alpha,\, c^{-1/2}} \text{ but } n^{-1}S_n \notin U_{c^{-1/2}\log c,\, c^{-1/2}}\} = o(e^{-c}),$$

since $\|z\|^2 \ge c^{-1}(\log c)^2$ if $\mu (= y+z) \notin U_{c^{-1/2}\log c,\, c^{-1/2}}$.

The first term on the right-hand side of (A.1) can be analyzed by making use of (A.2), (A.7) and arguments similar to (3.15)–(3.17), in which the manifold $M_{c^{-1/2}}$ is now replaced by $M_{c^{-1/2}} \cap \{\mu: \delta^{-1} + \delta^{-1}c^{-4/5} \ge g(\mu) > c/(\delta c + t)\}$. Since $t = [c^{1/4}]$, (A4) then yields

$$\begin{aligned} &\sum_{\delta c\le n\le \delta c+t} P\{ng(S_n/n) > c\} \\ \text{(A.10)} \quad &= \sum_{\delta c \le n \le \delta c + t} P\big\{c + c^{1/5} \ge ng(S_n/n) > c, n^{-1}S_n \in U_{c^{-1/2}\log c,\, c^{-1/2}}\big\} \\ &\quad + o(e^{-c}) \\ &= o(c^{q/2}e^{-c}). \end{aligned}$$

From (A.1), (A.2), (A.3), (A.7) and (A.10), (3.12) follows. □

REFERENCES

BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

BARNDORFF-NIELSEN, O. and COX, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 279–312.

BOROVKOV, A. A. and ROGOZIN, B. A. (1965). On the multidimensional central limit theorem. *Theory Probab. Appl.* **10** 55–62.

CHAN, H. P. (1998). Boundary crossing theory in change-point detection and its applications. Ph.D. dissertation, Stanford Univ.

CHAN, H. P. and LAI, T. L. (1999). Importance sampling for Monte Carlo evaluation of boundary crossing probabilities in hypothesis testing and changepoint detection. Technical report, Dept. Statistics Stanford Univ.

CHANDRA, T. K. (1985). Asymptotic expansion of perturbed chi-squared variables. *Sankyā Ser. A* **47** 100–110.

CHANDRA, T. K. and GHOSH, J. K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and the perturbed chi-square variables. *Sankyā Ser. A* **41** 22–47.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York.

GROENEBOOM, P. (1980). *Large Deviations and Asymptotic Efficiencies*. Math. Centrum, Amsterdam.

GRAY, A. (1990). *Tubes*. Addison–Wesley, Reading, MA.

HIRSCH, M. (1976). *Differential Topology*. Springer, New York.

HOEFFDING, W. (1965). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–405.

HOTELLING, H. (1939). Tubes and spheres in n-spaces, and a class of statistical problems. *Amer. J. Math.* **61** 440–460.

HU, I. (1988). Repeated significance tests for exponential families. *Ann. Statist.* **16** 1643–1666.

JENSEN, J. L. (1995). *Saddlepoint Approximations*. Oxford Univ. Press, London.

JOHANSEN, S. and JOHNSTONE, I. (1990). Hotelling's theorem on the volume of tubes: Some illustrations in simultaneous inference and data analysis. *Ann. Statist.* **18** 652–684.

JOHNSTONE, I. and SIEGMUND, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality. *Ann. Statist*. **17** 184–194.

KNOWLES, M. and SIEGMUND, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *Internat. Statist. Rev.* **57** 205–220.

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. (1997). On optimal stopping problems in sequential hypothesis testing. *Statist. Sinica* **7** 33–51.

LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and ZHANG, L. M. (1994a). Nearly optimal generalized sequential likelihood ratio tests in multivariate exponential families. In *Multivariate Analysis and Its Applications* (T. W. Anderson, K. T. Fang and I. Olkin, eds.). 331–346. IMS, Hayward, CA.

LAI, T. L. and ZHANG, L. M. (1994b). A modification of Schwarz's sequential likelihood ratio tests in multivariate sequential analysis. *Sequential Anal.* **13** 79–96.

LALLEY, S. P. (1983). Repeated likelihood ratio tests for curved exponential families. *Z. Wahrsch. Verw. Gebiete* **62** 293–321.

NAIMAN, D. Q. (1986). Conservative confidence bands in curvilinear regression. *Ann. Statist.* **14** 896–906.

NAIMAN, D. Q. (1990). Volumes of tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.* **18** 685–716.

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

SHAO, Q. (1997). Self-normalized large deviations. *Ann. Probab.* **25** 285–328.

SIEGMUND, D. (1975). Error probabilities and average sample number of the sequential probability ratio test. *J. Roy. Statist. Soc. Ser. B* **37** 394–401.

SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4** 673–684.

SIEGMUND, D. (1985). *Sequential Analysis*. Springer, New York.

SIEGMUND, D. and ZHANG, H. (1993). The expected number of local maxima of a random field and the volume of tubes. *Ann. Statist.* **21** 1948–1966.

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

STONE, C. (1965). A local limit theorem for nonlattice multi-dimensional distribution functions. *Ann. Math. Statist.* **36** 546–551.

WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Statist.* **16** 117–186.

WEYL, H. (1939). On the volume of tubes. *Amer. J. Math.* **61** 461–472.

WOODROOFE, M. (1976). A renewal theorem for curved boundaries and moments of estimation. *Ann. Probab.* **4** 67–80.

WOODROOFE, M. (1978). Large deviations of the likelihood ratio statistics with applications to sequential testing. *Ann. Statist.* **6** 72–84.

WOODROOFE, M. (1979). Repeated likelihood ratio tests. *Biometrika* **66** 453–463.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 119260
E-MAIL: stachp@nus.edu.sg

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
E-MAIL: lait@stat.stanford.edu