The Annals of Statistics
1988, Vol. 16, No. 2, 856–886

NEARLY OPTIMAL SEQUENTIAL TESTS OF COMPOSITE HYPOTHESES[1]

BY TZE LEUNG LAI

Stanford University

A simple class of sequential tests is proposed for testing the one-sided composite hypotheses H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ for the natural parameter θ of an exponential family of distributions under the 0–1 loss and cost c per observation. Setting $\theta_1 = \theta_0$ in these tests also leads to simple sequential tests for the hypotheses H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ without assuming an indifference zone. Our analytic and numerical results show that these tests have nearly optimal frequentist properties and also provide approximate Bayes solutions with respect to a large class of priors. In addition, our method gives a unified approach to the testing problems of H versus K and also of H_0 versus H_1 and unifies the different asymptotic theories of Chernoff and Schwarz for these two problems.

1. Introduction and background. Let $X_1, X_2, \ldots$ be i.i.d. random variables whose common density $f_\theta(x)$ (with respect to some nondegenerate measure ν) belongs to the exponential family

$$f_\theta(x) = e^{\theta x - \psi(\theta)}. \tag{1.1}$$

Thus, $E_\theta X_1 = \psi'(\theta)$ is increasing in θ, and the Kullback–Leibler information number $I(\theta, \lambda) = E_\theta\{\log[f_\theta(X_1)/f_\lambda(X_1)]\}$ is given by

$$I(\theta, \lambda) = (\theta - \lambda)\psi'(\theta) - (\psi(\theta) - \psi(\lambda)) = \int_\theta^\lambda (\lambda - t)\psi''(t)\, dt. \tag{1.2}$$

Letting $S_n = X_1 + \cdots + X_n$ and $\overline{X}_n = S_n/n$, the maximum likelihood estimate of θ after n observations is obtained by solving the equation $\psi'(\theta) = \overline{X}_n$, which may not have a solution in the natural parameter space Θ. Throughout the sequel, we shall assume that θ is known to lie in an open interval A ($\subset \Theta$) with end-points $-\infty \le a_1 < a_2 \le \infty$ such that

$$\inf_{a_1 - \eta < \theta < a_2 + \eta} \psi''(\theta) > 0, \qquad \sup_{a_1 - \eta < \theta < a_2 + \eta} \psi''(\theta) < \infty$$
$$\text{and } \psi'' \text{ is uniformly continuous on } (a_1 - \eta, a_2 + \eta) \text{ for some } \eta > 0. \tag{1.3}$$

For example, when the X_i are normal, we can take $A = \Theta = (-\infty, \infty)$. Since θ is known to lie in A, the maximum likelihood estimate of θ based on $X_1, \ldots, X_n$ is

Received February 1986; revised October 1987.
[1]Research supported by the National Science Foundation, the National Institutes of Health and the Army Research Office.
AMS 1980 *subject classifications*. Primary 62L10; secondary 62L05, 62L15.
Key words and phrases. Bayes sequential tests, diffusion approximations, exponential family, Kullback–Leibler information, generalized sequential likelihood ratio tests, boundary crossing probabilities.

given by

$$\begin{aligned}\hat{\theta}_n &= (\psi')^{-1}(\overline{X}_n), \quad &&\text{if } \psi'(a_1) < \overline{X}_n < \psi'(a_2),\\ &= a_1, &&\text{if } \overline{X}_n \le \psi'(a_1),\\ &= a_2, &&\text{if } \overline{X}_n \ge \psi'(a_2).\end{aligned} \tag{1.4}$$

Let $\theta_0, \theta_1 \in A$ with $\theta_0 < \theta_1$. The problem of testing sequentially the simple null hypothesis H_0: $\theta = \theta_0$ versus the simple alternative H_1: $\theta = \theta_1$ was solved definitively by Wald and Wolfowitz (1948). They showed that among all tests for which

$$P_{\theta_0}\{\text{reject } H_0\} \le \alpha \quad \text{and} \quad P_{\theta_1}\{\text{reject } H_1\} \le \beta \tag{1.5}$$

and which have finite expected sample sizes under both H_0 and H_1, Wald's (1945) sequential probabilty ratio test, SPRT(α, β) with error probabilities α and β under P_{θ_0} and P_{θ_1}, respectively, minimizes the expected sample sizes both under H_0 and under H_1.

The theory of optimal sequential tests of composite hypotheses, however, is much less complete, and many basic problems still remain to be settled. First consider the problem of testing sequentially the composite null hypothesis H_0: $\theta \le \theta_0$ versus the composite alternative H_1: $\theta \ge \theta_1$ subject to the error constraints

$$\begin{aligned}P_\theta\{\text{reject } H_0\} &\le \alpha \quad \text{for } \theta \le \theta_0,\\ P_\theta\{\text{reject } H_1\} &\le \beta \quad \text{for } \theta \ge \theta_1.\end{aligned} \tag{1.6}$$

Although SPRT(α, β) also satisfies the error constraints (1.6) and has minimal $E_\theta T$ at $\theta = \theta_0$ or θ_1, where T denotes the sample size, its $E_\theta T$ may be far from being optimal at other values of θ, as was pointed out by Kiefer and Weiss (1957) who also suggested the minimax approach of finding a sequential test which minimizes $\sup_\theta E_\theta T$ over all tests satisfying (1.6). This minimax problem has been studied by Weiss (1962), Lai (1973) and Lorden (1976, 1980).

Instead of this frequentist minimax approach, one may adopt a Bayesian approach putting a prior distribution π on A and introducing a cost of c for each observation together with a loss of 1 (if the decision is wrong) and 0 (otherwise) —the so-called "0–1 loss." Thus, c represents the ratio of the sampling cost to the cost due to wrong decision. The problem then is to find a stopping rule T and a terminal decision rule δ to minimize

$$\begin{aligned}r(T, \delta) = c\int_A E_\theta T\, d\pi(\theta) &+ \int_{\theta \le \theta_0} P_\theta\{\delta \text{ accepts } H_1\}\, d\pi(\theta)\\ &+ \int_{\theta \ge \theta_1} P_\theta\{\delta \text{ accepts } H_0\}\, d\pi(\theta).\end{aligned} \tag{1.7}$$

A well known asymptotic solution to this problem is due to Schwarz (1962). Let $\mathscr{B}(c)$ denote the continuation region of the Bayes rule, i.e., the Bayes rule continues sampling at stage $n + 1$ iff $(n, S_n) \in \mathscr{B}(c)$. Assuming that $\pi(I) > 0$ for every open interval $I \subset A$, Schwarz's (1962) theory of asymptotic shapes leads to the following simple limiting continuation region of the Bayes rule: As $c \to 0$,

(1.8) $$\mathscr{B}(c)/|\log c| \to \left\{(t, w): 1 + \min_{i=0,1} [\theta_i w - t\psi(\theta_i)] > \sup_{\theta \in A} [\theta w - t\psi(\theta)]\right\}.$$

Thus, writing $n = t|\log c|$ and $S_n = w|\log c|$, an asymptotic approximation to the Bayes rule is to continue sampling at stage $n + 1$ iff

$$\log c^{-1} + \min_{i=0,1} [\theta_i S_n - n\psi(\theta_i)] > \sup_{\theta \in A} [\theta S_n - n\psi(\theta)]$$

or, equivalently, to stop sampling at stage

(1.9) $$N_c = \inf\left\{n \geq 1: \max\left[\prod_{i=1}^n f_{\hat{\theta}_n}(X_i) \Big/ \prod_{i=1}^n f_{\theta_0}(X_i), \prod_{i=1}^n f_{\hat{\theta}_n}(X_i) \Big/ \prod_{i=1}^n f_{\theta_1}(X_i)\right] \geq c^{-1}\right\}.$$

The terminal decision rule δ^* is to accept H_1 (or H_0) if

$$\prod_{i=1}^{N_c} f_{\theta_1}(X_i) > (\text{or } \leq) \prod_{i=1}^{N_c} f_{\theta_0}(X_i).$$

In view of the integral representation in (1.2), the function

(1.10) $$J(\theta) = \max\{I(\theta, \theta_0), I(\theta, \theta_1)\}$$

is minimized at θ^* (with $\theta_0 < \theta^* < \theta_1$) defined by

(1.11) $$\psi'(\theta^*) = \{\psi(\theta_1) - \psi(\theta_0)\}/(\theta_1 - \theta_0) \quad \text{or, equivalently,} \quad I(\theta^*, \theta_0) = I(\theta^*, \theta_1).$$

As shown by Wong (1968), as $c \to 0$,

(1.12) $$N_c \leq |\log c|/J(\theta^*),$$

(1.13) $$E_\theta N_c \sim |\log c|/J(\theta) \quad \text{for every } \theta,$$

(1.14) $$\sup_{\theta \leq \theta_0} P_\theta\{\delta^* \text{ accepts } H_1\} = o(c|\log c|) = \sup_{\theta \geq \theta_1} P_\theta\{\delta^* \text{ accepts } H_0\},$$

and the Bayes risk of the test (N_c, δ^*) satisfies

(1.15) $$r(N_c, \delta^*) \sim c|\log c| \int_A \frac{d\pi(\theta)}{J(\theta)} \sim \inf_{(T,\delta)} r(T, \delta).$$

A basic problem with Schwarz's theory of asymptotic shapes is whether the simple attractive rule (1.9) provides an adequate approximation to the actual Bayes rule. The theory assumes that $c \to 0$ while the indifference zone (θ_0, θ_1) remains fixed. To see how this assumption may lead to difficulties in applying the theory, consider the simple case where $\theta_1 = \Delta = -\theta_0$ and the X_i are normal with mean θ and variance 1, so that $J(\theta) = \frac{1}{2}(|\theta| + \Delta)^2$. Although c may be very small (say $c = 10^{-20}$) and Δ appears much larger to justify its being considered as "fixed" (say $\Delta = 0.1$), $|\log c|$ may turn out to be smaller than $1/J(\theta)$ for $|\theta| \le \Delta$. Since the asymptotic formula for the Bayes risk (1.15) involves $|\log c|/J(\theta)$ [see also (1.13)], if one uses the "$c \to 0$" approximation in this case, it seems more reasonable not to consider Δ as "fixed" but to also let $\Delta \to 0$.

Another important development in the area of Bayes sequential tests of composite hypotheses is Chernoff's (1961, 1965a, b) work on testing sequentially H_0: $\theta < 0$ versus H_1: $\theta > 0$ in the case where the X_i are normal with mean θ and variance 1. Instead of assuming an indifference zone as in Schwarz's theory, Chernoff's theory assumes a loss of $|\theta|$ for the wrong decision and considers the problem of finding a stopping rule T to minimize

$$r(T) = c\int_{-\infty}^{\infty} E_\theta T\, d\pi(\theta) + \int_{-\infty}^{0} |\theta| P_\theta\{S_T > 0\}\, d\pi(\theta) + \int_0^{\infty} \theta P_\theta\{S_T \le 0\}\, d\pi(\theta), \tag{1.16}$$

where the prior distribution π is assumed to be normal with mean 0 and variance σ^2, noting that the Bayes terminal decision rule accepts H_0 (or H_1) according as $S_n \le 0$ (or $S_n > 0$) when stopping occurs at stage n.

While Schwarz (1962) applied the transformation $t = n/|\log c|$ and $w = S_n/|\log c|$ to obtain the limiting region (1.8) in the (t, w) plane as $c \to 0$, Chernoff (1961) introduced a different normalization for the problem (1.16),

$$t = c^{2/3}(n + \sigma^{-2}), \quad w = c^{1/3} S_n, \tag{1.17}$$

and obtained also a limiting continuation region of the form $\{(t, w): |w| < f(t)\}$ as $c \to 0$. The stopping boundary $f(t)$ arises as the solution of the corresponding continuous-time stopping problem involving the Wiener process, and an asymptotic analysis of the free boundary problem associated with the optimal stopping problem leads to

$$f(t) = \{3t[\log 1/t - (\log 8\pi)/3 + o(1)]\}^{1/2} \quad \text{as } t \to 0, \tag{1.18}$$

$$f(t) \sim (4t)^{-1} \quad \text{as } t \to \infty \tag{1.19}$$

[cf. Chernoff (1965a) and Breakwell and Chernoff (1964)]. It is interesting to compare this with the boundary $|w| = f^*(t)$ in Schwarz's limiting region (1.8) in the normal case where $\theta_1 = \Delta = -\theta_0$ [$\psi(\theta) = \frac{1}{2}\theta^2$ and $A = (-\infty, \infty)$]:

$$f^*(t) = \left[(2t)^{1/2} - \Delta t\right]^+. \tag{1.20}$$

There is, therefore, a big difference in the asymptotic solutions as $c \to 0$ between Schwarz's theory, which assumes a fixed indifference zone (θ_0, θ_1), and Chernoff's theory, which assumes (instead of an indifference zone) a loss of $|\theta|$ for the wrong decision. Not only are the normalizations and limiting regions different in the two theories, but the methods of deriving these results are also very different. Schwarz's (1962) theory is based on approximating the Bayes rule by simple upper and lower bounds that are associated with stopping when the posterior risk falls below c (for the upper bound) or falls below constant times

$c|\log c|$ (for the lower bound). Chernoff's (1961) theory deals only with the normal case and is based on replacing the discrete-time stopping problem by a continuous-time stopping problem which can in turn be reduced to a free boundary problem.

In this paper, adopting the 0–1 loss function, we provide a unified approach to Bayes sequential tests of H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ (as considered by Schwarz) and also of H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ (as considered by Chernoff) for the parameter θ of the exponential family (1.1). Note by (1.4) that if $\overline{X}_n \in \psi'(A)$, then

$$\log\left\{ \prod_{i=1}^{n} f_{\hat{\theta}_n}(X_i) \bigg/ \prod_{i=1}^{n} f_{\theta}(X_i) \right\} = nI(\hat{\theta}_n, \theta). \tag{1.21}$$

Therefore, Schwarz's stopping rule (1.9) is essentially equivalent to

$$N = \inf\left\{ n \ge 1 \colon \max\left[I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1) \right] \ge n^{-1}|\log c| \right\}. \tag{1.22}$$

We propose to replace the factor $|\log c|$ in (1.22) by $g(cn)$, where g is a nonnegative function on $(0, \infty)$ with $g(t) \sim |\log t|$ as $t \to 0$ and satisfying some other conditions that will be described in Section 3. Thus, to test H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ with the 0–1 loss and a cost of c for each observation, we propose to use the stopping rule

$$N(g, c) = \inf\left\{ n \ge 1 \colon \max\left[I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1) \right] \ge n^{-1}g(cn) \right\} \tag{1.23}$$

and the terminal decision rule δ^* that accepts H_0 or H_1 according as $\hat{\theta}_{N(g,c)} \le \theta^*$ or $\hat{\theta}_{N(g,c)} > \theta^*$, where θ^* is defined in (1.11). In Sections 3 and 6, it will be shown that the test is asymptotically Bayes [in the sense of minimizing (1.7)] as $c \to 0$ for a large class of priors on A, not only for fixed θ_0 and θ_1 (as in Schwarz's theory), but also as $\theta_1 - \theta_0 \to 0$. Moreover, from the frequentist viewpoint of optimality, it will be shown that as $c \to 0$ the test asymptotically minimizes not only the maximal expected sample size $\sup_\theta E_\theta T$, but also $\int (E_\theta T) p(\theta)\, d\theta$ for a large class of weight functions p, among all tests that satisfy the error constraints (1.6) with $\alpha = P_{\theta_0}\{\hat{\theta}_{N(g,c)} > \theta^*\}$ and $\beta = P_{\theta_1}\{\hat{\theta}_{N(g,c)} \le \theta^*\}$.

Letting $\theta_1 = \theta_0$ in (1.23) leads to the stopping rule

$$T(g, c) = \inf\left\{ n \ge 1 \colon I(\hat{\theta}_n, \theta_0) \ge n^{-1}g(cn) \right\}. \tag{1.24}$$

In Section 4, we propose to use this stopping rule for the problem of testing H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ with the 0–1 loss and a cost of c for each observation. The terminal decision rule is to accept H or K according as $\hat{\theta}_{T(g,c)} \le \theta_0$ or $\hat{\theta}_{T(g,c)} > \theta_0$. It will be shown that by a proper choice of g (developed in Section 2), the test is asymptotically Bayes as $c \to 0$ for a large class of priors on A.

The stopping rules (1.23) and (1.24), therefore, provide a unified treatment of testing H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ (with an indifference zone) and of testing H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ (without an indifference zone). Intuitively, as $\theta_1 \to \theta_0$, the Bayes sequential test of H_0 versus H_1 should approach the Bayes procedure for testing H versus K. Section 2 presents several basic results on the continuous-time stopping problems in Bayes sequential tests of H_0 versus H_1

and of H versus K for the drift coefficient θ of a Wiener process. These results show that the stopping boundary for testing H_0 versus H_1 indeed approaches that for testing H versus K as $\theta_1 \to \theta_0$. Moreover, the asymptotic behavior of the stopping boundaries near the origin is essentially the same for both problems.

Section 5 gives some numerical results on the performance of the tests proposed herein. Their risks $cE_\theta T + P_\theta\{\text{wrong decision}\}$ are compared at various values of θ with the risk of the "fictitious" optimal fixed sample size test that assumes the value of θ to be known. A discussion of these results and of the adaptive character of the stopping rules (1.22) and (1.24) is also provided in Section 5.

2. Bayes sequential tests for the drift of a Wiener process. Let $w(t)$, $t \geq 0$, be a Wiener process with $E(w(t)) = \mu t$ and $\text{Var}(w(t)) = t$ (μ is the drift coefficient). Consider the problem of testing sequentially H: $\mu < 0$ versus K: $\mu > 0$ with the 0–1 loss and a cost of t for observing the process for a period of length t. Assuming a flat prior (i.e., Lebesgue measure) on $\mu \in (-\infty, \infty)$, the Bayes terminal decision rule is to accept H or K according as $w(t) < 0$ or $w(t) > 0$ when stopping occurs at time t. With this terminal decision rule, the posterior loss $L_0(t, w)$ at time t if $w(t) = w$ is observed and we decide to stop can be easily shown to be

$$L_0(t,w) = t + \Phi(-|w|t^{-1/2}), \tag{2.1}$$

where $\Phi(x) = \int_{-\infty}^{x}(2\pi)^{-1/2}\exp(-\frac{1}{2}y^2)\,dy$. Introduce the transformation

$$\begin{aligned} s &= 1/t, \qquad Y(s) = w(t)/t, \\ y &= w/t, \qquad v(s,y) = L_0(t,w) = s^{-1} + \Phi(-|y|s^{-1/2}). \end{aligned} \tag{2.2}$$

Then $Y(s)$ is the posterior mean at time $t = s^{-1}$, and the optimal stopping rule is to stop as soon as $|Y(s)| \geq y^*(s)$, where the optimal boundary $y^*(s)$ can be determined by the free boundary problem

$$\frac{1}{2}\frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial s} \quad \text{for } |y| < y^*(s), \tag{2.3a}$$

$$u(s,y) = v(s,y) \quad \text{for } |y| = y^*(s), \tag{2.3b}$$

$$\frac{\partial u}{\partial y} = \frac{\partial u}{\partial s} \quad \text{for } |y| = y^*(s) \tag{2.3c}$$

[cf. Chernoff (1968)].

An asymptotic analysis of the free boundary problem (2.3) along the lines of Breakwell and Chernoff (1964) and Chernoff (1965a) shows that

$$s^{-1/2}y^*(s) = \tfrac{1}{4}(2/\pi)^{1/2}\{s - 5s^3/48\pi + \cdots\} \quad \text{as } s \to 0,$$

$$s^{-1/2}y^*(s) = \{2[\log s + \tfrac{1}{2}\log\log s - \tfrac{1}{2}\log 4\pi + o(1)]\}^{1/2} \quad \text{as } s \to \infty.$$

Since $ys^{-1/2} = wt^{-1/2}$ by (2.2), we therefore obtain

LEMMA 1. *For the preceding problem of testing H: $\mu < 0$ versus K: $\mu > 0$ with 0–1 loss and unit sampling cost per unit time, the Bayes stopping rule (with respect to the flat prior) is*

$$\tau_0 = \inf\{t > 0: |w(t)| \geq h_0(t)\}, \tag{2.4}$$

where h_0 is a positive function on $(0, \infty)$ such that

$$h_0(t) = \{2t[\log t^{-1} + \tfrac{1}{2}\log\log t^{-1} - \tfrac{1}{2}\log 4\pi + o(1)]\}^{1/2} \quad \text{as } t \to 0 \tag{2.5}$$

and

$$h_0(t) = \tfrac{1}{4}(2/\pi)^{1/2}\{t^{-1/2} - 5t^{-5/2}/48\pi + \cdots\} \quad \text{as } t \to \infty. \tag{2.6}$$

We now introduce an indifference zone $(-\gamma, \gamma)$ and consider the problem of testing H_0: $\mu \leq -\gamma$ versus H_1: $\mu \geq \gamma$ with unit sampling cost per unit time and a loss of 1 for the wrong terminal decision (and 0, otherwise). Assuming a flat prior on μ, the Bayes terminal decision rule accepts H_0 or H_1 according as $w(t) < 0$ or $w(t) > 0$ when stopping occurs at time t, and the posterior loss at this time is given by

$$L_\gamma(t, w) = t + \Phi(-|w|t^{-1/2} - \gamma t^{1/2}). \tag{2.7}$$

An asymptotic analysis of the corresponding free boundary problem leads to

LEMMA 2. *For the problem of testing H_0: $\mu \leq -\gamma$ versus H_1: $\mu \geq \gamma$ with 0–1 loss and unit sampling cost per unit time, the Bayes stopping rule (with respect to the flat prior) is*

$$\tau_\gamma = \inf\{t > 0: |w(t)| \geq h_\gamma(t)\}, \tag{2.8}$$

where h_γ is a positive function on $(0, \infty)$.

(i) *For fixed $0 < \gamma < \infty$, as $t \to 0$,*

$$h_\gamma(t) = \{2t[\log t^{-1} + \tfrac{1}{2}\log\log t^{-1} - \tfrac{1}{2}\log 4\pi + o(1)]\}^{1/2}, \tag{2.9}$$

and

$$h_\gamma(t) \sim \tfrac{1}{4}(2/\pi)^{1/2}t^{-1/2}\exp(-\tfrac{1}{2}\gamma^2 t) \quad \text{as } t \to \infty. \tag{2.10}$$

(ii) *Let $\gamma \to \infty$. Then the asymptotic expansion* (2.9) *still holds as $t \to 0$ such that $t = o((\gamma^2\log\gamma^2)^{-1})$.*

(iii) *Let $\gamma \to \infty$. Then for every fixed $0 < \rho < 2$, as $t \to 0$ such that $t \leq \rho\gamma^{-2}\log\gamma^2$,*

$$h_\gamma(t) = -\gamma t + \{2t[\log t^{-1} + \tfrac{1}{2}\log\log t^{-1} + O(1)]\}^{1/2}. \tag{2.11}$$

(iv) *As $\gamma \to \infty$, $\sup_{t \geq \lambda\gamma^{-2}\log\gamma^2} h_\gamma(t) \to 0$ for every $\lambda > 2$.*

Details of the asymptotic analysis of the free boundary problem (2.3) [with $v(s, y) = s^{-1} + \Phi(-|y|s^{-1/2} - \gamma s^{-1/2})$, where $s = t^{-1}$ and $y = w/t$] associated

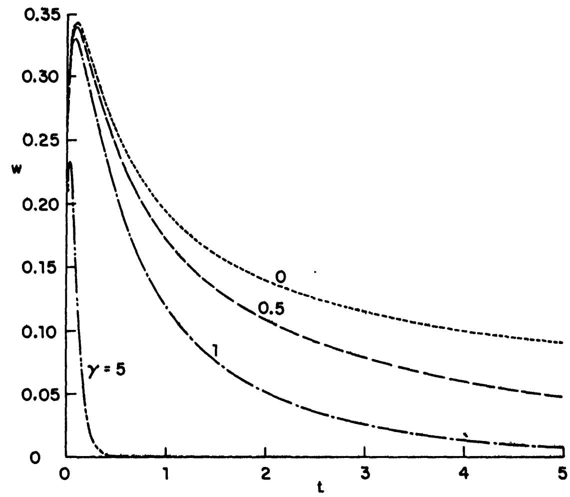

FIG. 1.

with Lemmas 1 and 2 are given elsewhere. Comparison of (2.9) with (2.5) shows that the optimal boundary for the problem of testing H_0: $\mu \le -\gamma$ versus H_1: $\mu \ge \gamma$ has the same asymptotic behavior for small t as the optimal boundary for testing H: $\mu < 0$ versus K: $\mu > 0$. Moreover, even when $\gamma \to \infty$ (large indifference zones), (2.9) still holds for $t = o((\gamma^2 \log \gamma^2)^{-1})$ and a modified form (2.11) also holds for $t < (2 - \varepsilon)\gamma^{-2} \log \gamma^2$. Thus, in the case of 0–1 loss, there is a unified asmptotic theory as $t \to 0$ for continuous-time Bayes sequential tests of H versus K and also of H_0 versus H_1. In this connection, also note the difference between (2.5) corresponding to the 0–1 loss and Chernoff's asynmptotic expansion (1.18) corresponding to the loss $|\mu|$ due to wrong decision. Since setting $\gamma = 0$ in (2.10) gives the leading term in (2.6), there is also a unified asymptotic theory as $t \to \infty$ for Bayes sequential tests of H versus K and also of H_0 versus H_1 in the case of 0–1 loss.

Using the numerical methods developed by Chernoff and Petkau (1986), we computed the continuous-time optimal stopping boundaries h_γ for a variety of values of γ. The graphs of h_γ for $\gamma = 0, 0.5, 1, 5$ are given in Figure 1. Our numerical and analytic results suggest the following simple approximation to h_0:

$$
\begin{aligned}
h_0^*(t) &= \tfrac{1}{4}(2/\pi)^{1/2}\left(t^{-1/2} - 5t^{-5/2}/48\pi\right), \quad \text{if } t \ge 0.8,\\
&= \exp(-0.69t - 1), \quad \text{if } 0.1 \le t < 0.8,\\
&= 0.39 - 0.015t^{-1/2}, \quad \text{if } 0.01 \le t < 0.1, \qquad (2.12)\\
&= \left\{t\left[2\log t^{-1} + \log\log t^{-1} - \log 4\pi \right.\right.\\
&\qquad \left.\left. - 3\exp(-0.016t^{-1/2})\right]\right\}^{1/2}, \quad \text{if } t < 0.01.
\end{aligned}
$$

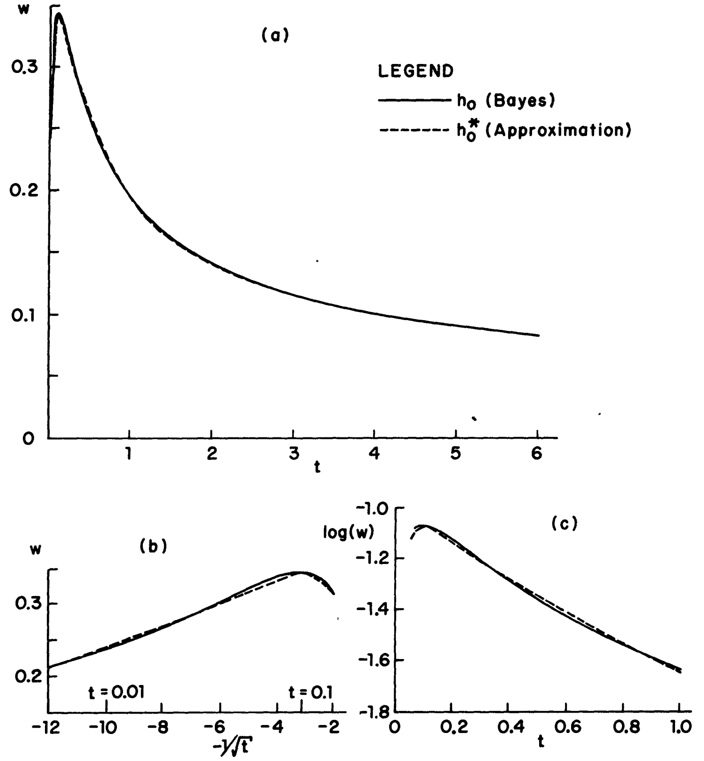

FIG. 2.

Figure 2 shows that h_0 is closely approximated by h_0^*. The approximation h_0^* is based on (i) the asymptotic expansion (2.6) which holds remarkably well for $t \geq 0.8$ (Figure 2a), (ii) the approximate linearity of $\log h_0(t)$ for $0.1 \leq t \leq 1$ (Figure 2c), (iii) the approximate linearity of h_0 as a function of $t^{-1/2}$ over the range $0.01 \leq t \leq 0.1$ (Figure 2b) and (iv) the asymptotic expansion (2.5) as $t \to 0$.

For $0 < \gamma \leq 20$, our numerical results and the asymptotic relation (2.10) suggest the following simple approximation to h_γ:

$$
(2.13) \qquad \begin{aligned} h_\gamma^*(t) &= h_0^*(t)\exp\left(-\tfrac{1}{2}\gamma^2 t\right), \qquad t \geq 1, \\ &= h_0^*(t)\exp\left(-\tfrac{1}{2}\gamma^2 t^{1.125}\right), \qquad 0 < t < 1. \end{aligned}
$$

Figure 3 shows that this provides a good approximation to h_γ; moreover, for $\gamma \geq 20$, we can use the following approximation which is suggested by Lemma 2(ii)–(iv):

$$
(2.14) \qquad h_\gamma^{**}(t) = \left[t^{1/2}\left(2\log t^{-1} + \log\log t^{-1} - \log 4\pi\right)^{1/2} - \gamma t\right]^+.
$$

3. Asymptotically optimal sequential tests of H_0: $\theta \leq \theta_0$ versus H_1: $\theta \geq \theta_1$ for the parameter of an exponential family. Let $X_1, X_2, \ldots$ be i.i.d. random variables having a common density $f_\theta(x) = e^{\theta x - \psi(\theta)}$ with respect to

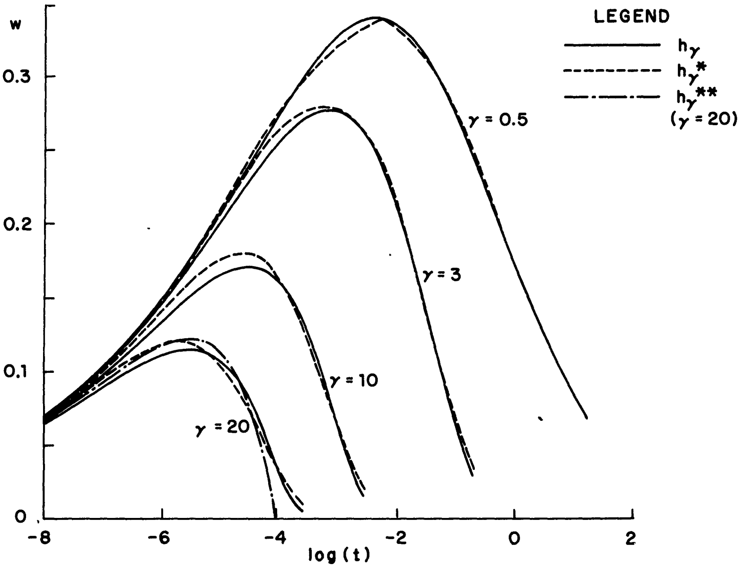

FIG. 3.

some nondegenerate measure ν. Let $S_n = X_1 + \cdots + X_n$, $\overline{X}_n = S_n/n$. In this section, we study the problem of testing H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ ($> \theta_0$) with a cost c for each observation and the 0–1 loss.

To begin with, suppose that c is small and the width $\theta_1 - \theta_0$ of the indifference zone is considerably larger than $c^{1/2}$. Specifically, let

$$c \to 0 \quad \text{and} \quad (\theta_1 - \theta_0)^2/c \to \infty. \tag{3.1}$$

Moreover, assume that the X_i are normal with mean θ and variance 1, and that $\theta_0 = -\theta_1$. Define

$$t = cn, \qquad w(t) = c^{1/2}S_n, \qquad \mu = c^{-1/2}\theta, \qquad \gamma = c^{-1/2}\theta_1. \tag{3.2}$$

Since $c^{1/2}\theta n = \mu t$, $w(t)$ is a Wiener process with drift coefficient μ and with t restricted to the set $I_c = \{c, 2c, \dots\}$. As $c \to 0$, I_c becomes dense in $[0, \infty)$. This suggests using for small c the Bayes stopping rule τ_γ defined in (2.8) for testing H_0: $\mu \le -\gamma$ versus H_1: $\mu \ge \gamma$ based on the $w(t)$ defined in (3.2).

Let $g_\gamma(t) = \{h_\gamma(t) + \gamma t\}^2/2t$, where h_γ is defined in Lemma 2. Then in view of (3.2),

$$\begin{aligned} |w(t)| \ge h_\gamma(t) &\Leftrightarrow \big(|w(t)| + \gamma t\big)^2/2t \ge g_\gamma(t) \\ &\Leftrightarrow (|S_n| + \theta_1 n)^2/2n \ge g_\gamma(cn) \\ &\Leftrightarrow \max\big\{I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1)\big\} \ge n^{-1}g_\gamma(cn), \end{aligned} \tag{3.3}$$

where $\hat{\theta}_n = \overline{X}_n$ and $I(\theta, \lambda) = \frac{1}{2}(\theta - \lambda)^2$ is the Kullback–Leibler information number. Hence τ_γ leads to a stopping rule of the form (1.23). By (3.1), $\gamma \to \infty$

and, therefore, by Lemma 2(iii), for every fixed $0 < \varepsilon < 1$,

$$(3.4)\quad \begin{aligned} g_\gamma(t) &= \log t^{-1} + \tfrac{1}{2}\log\log t^{-1} + O(1) \quad \text{if } 0 < t \le (2-\varepsilon)\gamma^{-2}\log\gamma^2, \\ &= o(1) \quad \text{if } t \ge (2+\varepsilon)\gamma^{-2}\log\gamma^2. \end{aligned}$$

We now extend the rule τ_γ from the normal case to the general exponential family. We assume that θ is known to lie in an open interval $A \subset \Theta$ satisfying (1.3) and that $\theta_0, \theta_1 \in A$. Define the maximum likelihood estimate $\hat{\theta}_n$ by (1.4). For every real number ξ, let $\mathscr{C}_\xi$ denote the class of all nonnegative functions on $(0, \infty)$ such that

$$(3.5)\quad g(t) \sim \log t^{-1} \quad \text{and} \quad g(t) \ge \log t^{-1} + \xi \log\log t^{-1} \quad \text{as } t \to 0.$$

For $g \in \mathscr{C}_\xi$ and $c > 0$, define the stopping rule

$$(3.6)\quad N(g, c) = \inf\{n \ge 1: \max[I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1)] \ge n^{-1}g(cn)\}$$

as in (1.23). In conjunction with this stopping rule, use the terminal decision rule δ^* which accepts H_0 or H_1 according as $\hat{\theta}_{N(g,c)} \le \theta^*$ or $\hat{\theta}_{N(g,c)} > \theta^*$, where θ^* is defined in (1.11). The following theorem shows that the test $(N(g,c), \delta^*)$ is asymptotically Bayes [in the sense of minimizing (1.7)] as $c \to 0$ for a large class of priors, not only for fixed θ_0, θ_1 (as in Schwarz's theory), but also as $\theta_1 - \theta_0 \to 0$.

THEOREM 1. *Let π be a probability distribution on A and let $r(T, \delta)$ denote the Bayes risk* (1.7) *of a test (T, δ) of H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ with the* 0–1 *loss and cost c per observation. Let $g \in \mathscr{C}_\xi$ with $\xi > -\frac{1}{2}$.*

(i) *Assume that $\pi([\theta_0 - t, \theta_0]) > 0$ and $\pi([\theta_1, \theta_1 + t]) > 0$ for all $t > 0$ and that for some $\rho > 0$ and $\varepsilon > 0$,*

$$(3.7)\quad \pi([x, y]) \le \rho(y - x) \quad \text{for all } x, y \in [\theta_0 - \varepsilon, \theta_0] \cup [\theta_1, \theta_1 + \varepsilon] \quad \text{with } x < y.$$

Then for fixed θ_0 and θ_1, as $c \to 0$,

$$(3.8)\quad r(N(g,c), \delta^*) \sim c|\log c| \int_A \frac{d\pi(\theta)}{J(\theta)} \sim \inf_{(T,\delta)} r(T, \delta),$$

where J is defined in (1.10). *Moreover, we can drop the assumption* (3.7) *if $\xi > \frac{1}{2}$.*

(ii) *Assume that π has a positive continuous density π' in some neighborhood of θ_0. Then as $c \to 0$ and $\theta_1 \to \theta_0$ such that $(\theta_1 - \theta_0)^2/c \to \infty$,*

$$(3.9)\quad \begin{aligned} r(N(g,c), \delta^*) &\sim (8\pi'(\theta_0)/\psi''(\theta_0))c(\theta_1 - \theta_0)^{-1}\log[(\theta_1 - \theta_0)^2/c] \\ &\sim \inf_{(T,\delta)} r(T, \delta). \end{aligned}$$

(iii) *Let $0 \le \gamma < \infty$ and let $g_\gamma(t) = \{h_\gamma(t) + \gamma t\}^2/2t$, where h_γ is the optimal boundary for the continuous-time stopping problem introduced in Lemma* 1 (*for $\gamma = 0$*) *and Lemma* 2 (*for $\gamma > 0$*). *Then $g_\gamma \in \mathscr{C}_{1/2}$. Assume that π has a positive continuous density π' in some neighborhood of θ_0. Then as $c \to 0$ and $\theta_1 \to \theta_0$*

such that $(\theta_1 - \theta_0)/(2c^{1/2}) \to \gamma$,

$$r\big(N(g_\gamma, c), \delta^*\big) \sim \inf_{(T,\delta)} r(T,\delta)$$

$$\sim \frac{c^{1/2}\pi'(\theta_0)}{(\psi''(\theta_0))^{1/2}}\bigg\{\int_{-\infty}^{\infty} E(\tau_\gamma|\mu)\,d\mu + \int_{-\infty}^{-\gamma} P\big[w(\tau_\gamma) > 0|\mu\big]\,d\mu + \int_{\gamma}^{\infty} P\big[w(\tau_\gamma) < 0|\mu\big]\,d\mu\bigg\}, \tag{3.10}$$

where $w(t)$, $t \geq 0$, *denotes the Wiener process with drift coefficient* μ *under* $P(\cdot|\mu)$, *and* τ_γ *is defined in* (2.8).

The proof of Theorem 1 is given in Section 6, which also studies the power function and the expected sample size of the test $(N(g,c), \delta^*)$. Let

$$\alpha = P_{\theta_0}\big\{\hat{\theta}_{N(g,c)} > \theta^*\big\}, \qquad \beta = P_{\theta_1}\big\{\hat{\theta}_{N(g,c)} \leq \theta^*\big\} \tag{3.11}$$

and let $\mathscr{I}(\alpha,\beta)$ denote the class of all tests that satisfy the error constraints (1.6). Making use of the asymptotic properties of the test $(N(g,c),\delta^*)$, together with Hoeffding's (1960) lower bound for the expected sample size of any test in $\mathscr{I}(\alpha,\beta)$, we prove in Section 6 the following asymptotically optimal frequentist properties of the test $(N(g,c),\delta^*)$.

THEOREM 2. *Let* $g \in \mathscr{C}_\xi$ *for some* ξ *and define* α, β *by* (3.11). *Then* $(N(g,c),\delta^*) \in \mathscr{I}(\alpha,\beta)$.

(i) *For fixed* θ_0 *and* θ_1, *as* $c \to 0$,

$$\log\alpha \sim \log\beta \sim \log c, \tag{3.12}$$

and for every bounded subset B *of* A,

$$E_\theta N(g,c) \sim |\log c|/J(\theta) \sim \inf_{(T,\delta)\in\mathscr{I}(\alpha,\beta)} E_\theta T, \tag{3.13}$$

uniformly in $\theta \in B$, *where* J *is defined in* (1.10).

(ii) *As* $c \to 0$ *and* $\theta_1 \to \theta_0$ *such that* $(\theta_1 - \theta_0)^2/c \to \infty$,

$$\log\alpha \sim \log\beta \sim \log(c/d^2), \tag{3.14}$$

where $d = \theta_1 - \theta_0$. *Moreover,*

$$\sup_\theta E_\theta N(g,c) \sim 8d^{-2}(\log c^{-1}d^2)/\psi''(\theta_0) \sim \inf_{(T,\delta)\in\mathscr{I}(\alpha,\beta)} \sup_\theta E_\theta T \tag{3.15}$$

and for every distribution function π *on* A *having a positive continuous derivative* π' *in some neighborhood of* θ_0,

$$\int_A E_\theta N(g,c)\,d\pi(\theta) \sim (8\pi'(\theta_0)/\psi''(\theta_0))d^{-1}\log(d^2/c) \sim \inf_{(T,\delta)\in\mathscr{I}(\alpha,\beta)} \int_A E_\theta T\,d\pi(\theta). \tag{3.16}$$

4. Asymptotically optimal tests of H: $\theta < \theta_0$ versus K: $\theta > \theta_0$. In this section we study the problem of testing H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ for the parameter θ of the exponential family (1.1), with a cost c for each observation and the 0–1 loss. Assuming a prior distribution π on A, the Bayes risk of a test (T, δ) is given by

$$r(T,\delta) = c\int_A E_\theta T\,d\pi(\theta) + \int_{\theta<\theta_0} P_\theta\{\delta \text{ accepts } K\}\,d\pi(\theta) + \int_{\theta>\theta_0} P_\theta\{\delta \text{ accepts } H\}\,d\pi(\theta). \tag{4.1}$$

Define h_0 as the optimal boundary for the stopping problem in Lemma 1 involving the Wiener process and let $g_0(t) = h_0^2(t)/(2t)$. Consider the stopping rule

$$T_c = \inf\{n \ge 1\colon I(\hat{\theta}_n, \theta_0) \ge n^{-1} g_0(cn)\}, \tag{4.2}$$

and the terminal decision rule δ^* such that

$$\delta^* \text{ accepts } H \text{ or } K \text{ according as } \hat{\theta}_{T_c} < \theta_0 \text{ or } \hat{\theta}_{T_c} > \theta_0. \tag{4.3}$$

Note that $\hat{\theta}_{T_c} \ne \theta_0$ by (4.2) and that the stopping rule $N(g, c)$ defined in (3.6) for testing H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$ reduces to T_c when $\theta_1 = \theta_0$ and $g = g_0$. Theorem 3 shows that (T_c, δ^*) is asymptotically Bayes, with respect to a general class of priors π, for testing H versus K.

THEOREM 3. *Let π be a prior distribution on A such that π has a positive continuous density π' in some neighborhood of θ_0 $(\in A)$. Define $r(T, \delta)$ as in* (4.1) *and define T_c by* (4.2) *and δ^* by* (4.3). *Then as $c \to 0$,*

$$r(T_c, \delta^*) \sim \inf_{(T,\delta)} r(T, \delta)$$

$$\sim \frac{c^{1/2}\pi'(\theta_0)}{(\psi''(\theta_0))^{1/2}}\left\{\int_{-\infty}^{\infty} E(\tau_0|\mu)\,d\mu + \int_{-\infty}^{0} P[w(\tau_0) > 0|\mu]\,d\mu + \int_0^\infty P[w(\tau_0) < 0|\mu]\,d\mu\right\}, \tag{4.4}$$

where $w(t)$, $t \ge 0$, denotes the Wiener process with drift coefficient μ under $P(\cdot|\mu)$, and τ_0 is defined in (2.4).

We preface the proof of Theorem 3 by Lemmas 3 and 4.

LEMMA 3. *Let $w(t)$, $t \ge 0$, be the Wiener process with drift coefficient μ under $P(\cdot|\mu)$. Let h be a bounded positive function on $(0, \infty)$ such that*

$$h^2(t) \sim 2t \log t^{-1} \quad \textit{and} \quad h^2(t)/2t \ge \log t^{-1} + \xi \log\log t^{-1} \quad \textit{as } t \to 0 \tag{4.5}$$

for some ξ. *Define* $\tau = \inf\{t > 0: |w(t)| \ge h(t)\}$. *Then*

$$\int_{-\infty}^{\infty} E(\tau|\mu)\, d\mu + \int_{-\infty}^{0} P[w(\tau) > 0|\mu]\, d\mu + \int_{0}^{\infty} P[w(\tau) < 0|\mu]\, d\mu < \infty. \tag{4.6}$$

PROOF. By (4.5) and Lemma 1 of Lai, Robbins and Siegmund (1983), as $|\mu| \to \infty$,

$$E(\tau|\mu) \sim (2\log \mu^2)/\mu^2, \quad P[w(\tau)\operatorname{sgn}(\mu) < 0|\mu] = O\left(\mu^{-2}(\log \mu^2)^{-\xi-1/2}\right). \tag{4.7}$$

From (4.7), the desired conclusion (4.6) follows. □

LEMMA 4. *Let* $\theta_0 \in A$. *For* $c > 0$, *let* $w_c(t) = (c/\psi''(\theta_0))^{1/2}(S_n - n\psi'(\theta_0))$ *for* $t = cn$ $(n = 1, 2, \dots)$ *and define* $w_c(t)$ *by linear interpolation for* $cn \le t \le c(n+1)$. *Then for* $B > 0$ *and* $M > 0$, *as* $c \to 0$, *the process* $\{w_c(t), 0 \le t \le B\}$ *under* $P_{\theta_0+\mu(c/\psi''(\theta_0))^{1/2}}$ *converges weakly to the Wiener process* $\{w(t), 0 \le t \le B\}$ *with drift coefficient* μ, *the convergence being uniform in* $-M \le \mu \le M$.

PROOF. Note that for $t \in \{c, 2c, \dots\}$,

$$E_{\theta_0+\mu(c/\psi''(\theta_0))^{1/2}}(w_c(t)) = t(c\psi''(\theta_0))^{-1/2}\left\{\psi'\left(\theta_0 + \mu(c/\psi''(\theta_0))^{1/2}\right) - \psi'(\theta_0)\right\} \to \mu t,$$

$$\operatorname{Var}_{\theta_0+\mu(c/\psi''(\theta_0))^{1/2}}(w_c(t)) \sim \operatorname{Var}_{\theta_0}(w_c(t)) = t,$$

as $c \to 0$, uniformly in $t \le B$ and $-M \le \mu \le M$. Since

$$\sup_{\theta \in A} E_\theta |X - E_\theta X|^r < \infty \quad \text{for all } r > 0$$

[cf. Lai (1988), (3.3)] the desired conclusion then follows from the uniform version of Donsker's invariance principle [cf. Freedman (1971), pages 90–93 and Lai (1977), Theorem 2]. □

LEMMA 5. *Let* $\theta_0 \in A$. *Define* T_c *by* (4.2) *and* δ^* *by* (4.3). *Then as* $c \to 0$,

$$E_\theta T_c = O\left((\theta - \theta_0)^{-2}\log\{c^{-1}(\theta - \theta_0)^2\}\right) \quad \text{uniformly in } \theta \in A \text{ with } (\theta - \theta_0)^2 \ge 2c, \tag{4.8}$$

$$P_\theta\{\hat\theta_{T_c} > \theta_0\} = O\left(c(\theta - \theta_0)^{-2}\right) \quad \text{uniformly in } \theta \in A \text{ with } \theta - \theta_0 \le -(2c)^{1/2}, \tag{4.9}$$

$$P_\theta\{\hat\theta_{T_c} < \theta_0\} = O\left(c(\theta - \theta_0)^{-2}\right) \quad \text{uniformly in } \theta \in A \text{ with } \theta - \theta_0 \ge (2c)^{1/2}. \tag{4.10}$$

PROOF. We make use of the boundary crossing theory developed in Lai (1988). Since $g_0(t) = h_0^2(t)/2t$, it follows from Lemma 1 that

$$(4.11)\qquad g_0(t) \sim \log t^{-1} \quad \text{as } t \to 0 \quad \text{and} \quad g_0(t) \to 0 \quad \text{as } t \to \infty.$$

Let $L_c = \sup\{n: I(\hat{\theta}_n, \theta_0) \le n^{-1} g_0(cn)\}$. In view of (4.11), an argument similar to the proof of Theorem 3 of Lai (1988) shows that

$$E_\theta L_c = O\Big((\theta - \theta_0)^{-2} \log\big\{c^{-1}(\theta - \theta_0)^2\big\}\Big)$$

uniformly in $\theta \in A$ with $(\theta - \theta_0)^2 \ge 2c$. Since $T_c - 1 \le L_c$, (4.8) follows. Moreover, (4.9) and (4.10) follow from Theorem 1(iii) of Lai (1988) [see also Lai (1988), Lemma 7 and Lai (1987), Lemma 1].

PROOF OF THEOREM 3. Let $M \ge 2$. By Lemma 4, as $c \to 0$,

$$P_{\theta_0 + \mu(c/\psi''(\theta_0))^{1/2}}\{cT_c \le x\} \to P\{\tau_0 \le x | \mu\} \quad \text{for all } x > 0,$$

$$P_{\theta_0 + \mu(c/\psi''(\theta_0))^{1/2}}\big\{\hat{\theta}_{T_c} < \theta_0\big\} \to P\{w(\tau_0) < 0 | \mu\}$$

uniformly in $-M \le \mu \le M$, noting that for $\hat{\theta}_n$ near θ_0,

$$\begin{aligned} nI(\hat{\theta}_n, \theta_0) &\sim \tfrac{1}{2}\psi''(\theta_0)(\hat{\theta}_n - \theta_0)^2 n \sim \tfrac{1}{2}(\psi''(\theta_0))^{-1}\{\psi'(\hat{\theta}_n) - \psi'(\theta_0)\}^2 n \\ &= \tfrac{1}{2}(cn)^{-1}(c/\psi''(\theta_0))(S_n - n\psi'(\theta_0))^2 \end{aligned}$$

and that $\tau_0 = \inf\{t: w^2(t)/2t \ge g_0(t)\}$. It therefore follows from Lemma 4 that as $c \to 0$,

$$\begin{aligned} &\int_{\theta_0 - M(c/\psi''(\theta_0))^{1/2}}^{\theta_0 + M(c/\psi''(\theta_0))^{1/2}} \{E_\theta(cT_c) + P_\theta[(T_c, \delta^*) \text{ makes the wrong decision}]\}\, d\pi(\theta) \\ (4.12)\quad &\sim (c/\psi''(\theta_0))^{1/2} \pi'(\theta_0) \Big\{ \int_{-M}^{M} E(\tau_0|\mu)\, d\mu + \int_{-M}^{0} P[w(\tau_0) > 0|\mu]\, d\mu \\ &\qquad + \int_0^M P[w(\tau_0) < 0|\mu]\, d\mu \Big\}. \end{aligned}$$

By choosing M arbitrarily large, it then follows from Lemma 5 and (4.12) that

$$\begin{aligned} r(T_c, \delta^*) &\sim (c/\psi''(\theta_0))^{1/2} \pi'(\theta_0) \\ (4.13)\quad &\quad \times \Big\{ \int_{-\infty}^{\infty} E(\tau_0|\mu)\, d\mu + \int_{-\infty}^{0} P[w(\tau_0) > 0|\mu]\, d\mu \\ &\qquad + \int_0^\infty P[w(\tau_0) < 0|\mu]\, d\mu \Big\}. \end{aligned}$$

Let $M > 2$. As $c \to 0$,

$$\inf_{(T,\delta)} r(T,\delta) \ge (\pi'(\theta_0) + o(1))(c/\psi''(\theta_0))^{1/2}$$

$$\times \inf_{(T,\delta)} \int_{-M}^{M} \{E_{\theta_0+\mu(c/\psi''(\theta_0))^{1/2}}(cT)$$

$$+ P_{\theta_0+\mu(c/\psi''(\theta_0))^{1/2}}[(T,\delta) \text{ makes the wrong decision}]\} \, d\mu \tag{4.14}$$

$$\sim \pi'(\theta_0)(c/\psi''(\theta_0))^{1/2} \inf_{\tau} \left\{ \int_{-M}^{M} E(\tau|\mu)\, d\mu + \int_{-M}^{0} P[w(\tau) > 0|\mu]\, d\mu + \int_{0}^{M} P[w(\tau) < 0|\mu]\, d\mu \right\},$$

by Lemma 4, noting that for any given stopping rule τ of the Wiener process $w(t)$ with drift coefficient μ, the Bayes terminal decision rule for testing H': $\mu < 0$ versus K': $\mu > 0$, with respect to 0–1 loss and uniform prior distribution on $[-M, M]$, accepts H' or K' according as $w(\tau) < 0$ or $w(\tau) > 0$. By the definition of τ_0 and Lemma 3 [noting that h_0 satisfies (4.5)],

$$\lim_{M\to\infty} \inf_{\tau} \left\{ \int_{-M}^{M} E(\tau|\mu)\, d\mu + \int_{-M}^{0} P[w(\tau) > 0|\mu]\, d\mu + \int_{0}^{M} P[w(\tau) < 0|\mu]\, d\mu \right\}$$

$$= \int_{-\infty}^{\infty} E(\tau_0|\mu)\, d\mu + \int_{-\infty}^{0} P[w(\tau_0) > 0|\mu]\, d\mu + \int_{0}^{\infty} P[w(\tau_0) < 0|\mu]\, d\mu. \tag{4.15}$$

From (4.13)–(4.15), the desired conclusion (4.4) follows. □

5. Some numerical results and discussion. In this section, we report some simulation results for the sequential tests introduced in the preceding sections. Instead of the exact continuous-time optimal stopping boundaries h_0 and h_γ of Lemmas 1 and 2, it is much more convenient to use their approximations h_0^* defined in (2.12), h_γ^* defined in (2.13) for $\gamma \le 20$, and h_γ^{**} defined in (2.14) for $\gamma > 20$. These approximations lead to the stopping rules

$$T_c^* = \inf\{n \ge 1: I(\hat{\theta}_n, \theta_0) \ge n^{-1} g_0^*(cn)\} \tag{5.1}$$

and

$$N_c^* = \inf\{n \ge 1: \max[I(\hat{\theta}_n, \theta_0), I(\hat{\theta}_n, \theta_1)] \ge n^{-1} g_\gamma^*(cn)\}, \tag{5.2}$$

where $g_\gamma^*(t) = \{h_\gamma^*(t) + \gamma t\}^2/2t$ for $0 \le \gamma \le 20$, $g_\gamma^*(t) = \{h_\gamma^{**}(t) + \gamma t\}^2/2t$ for $\gamma > 20$ and

$$\gamma = (\theta_1 - \theta_0)/(2c^{1/2}) \tag{5.3}$$

[cf. (3.2), (3.3) and (3.6)].

TABLE 1

Expected sample size ($E_\theta T_c^*$), error probability (p_θ) and risk $R_c(\theta)$ of the test (T_c^*, δ^*) of H: $\theta < 0$ versus K: $\theta > 0$ for the mean θ of a normal distribution with unit variance. Cost per observation = c. Each result is based on 1600 *simulation runs using importance sampling* [*Siegmund* (1976)].

	$c = 10^{-2}$			$c = 10^{-3}$			$c = 10^{-4}$		
$\mu = c^{-1/2}\theta$	$E_\theta T_c^*$	p_θ	$R_c(\theta)$	$E_\theta T_c^*$	p_θ	$R_c(\theta)$	$E_\theta T_c^*$	p_θ	$R_c(\theta)$
100	1.0	0.000	0.010	1.7	0.0000	0.0017	12	0.0000	0.0012
80	1.0	0.000	0.010	2.2	0.0000	0.0022	17	0.0000	0.0017
60	1.0	0.000	0.010	3.2	0.0000	0.0032	26	0.0000	0.0026
40	1.1	0.000	0.011	5.3	0.0000	0.0053	46	0.0000	0.0046
20	1.9	0.000	0.019	13.3	0.0002	0.0135	121	0.0004	0.0125
10	3.8	0.001	0.039	30.8	0.0032	0.0340	288	0.0049	0.0337
5	7.3	0.024	0.097	61.0	0.0382	0.0992	553	0.0470	0.1023
2.5	11.6	0.132	0.248	91.3	0.1593	0.2506	850	0.1718	0.2568
0.5	13.8	0.407	0.545	110.4	0.4173	0.5277	995	0.4215	0.5210

For the problem of testing H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ with the 0–1 loss and cost c for each observation, the risk function of a test is given by

$$(5.4)\qquad \begin{aligned} R_c(\theta) &= cE_\theta T + P_\theta\{\text{reject } H\}, \qquad \theta < \theta_0,\\ &= cE_\theta T + P_\theta\{\text{reject } K\}, \qquad \theta > \theta_0, \end{aligned}$$

where T is the stopping time of the test. We will use the terminal decision rule δ^* such that when stopping occurs at stage n,

$$(5.5)\qquad \begin{aligned} &\delta^* \text{ accepts } H \text{ or } K \text{ according as } \hat{\theta}_n < \theta_0 \text{ or } \hat{\theta}_n > \theta_0,\\ &\text{and accepts } H \text{ or } K \text{ with probability } \tfrac{1}{2} \text{ if } \hat{\theta}_n = \theta_0. \end{aligned}$$

Table 1 considers the case of a normal distribution with mean θ and variance 1 and tabulates the risk function of (T_c^*, δ^*) for $c = 10^{-2}, 10^{-3}, 10^{-4}$. Note that for every stopping rule T, the terminal decision rule δ^* is Bayes with respect to prior distributions that are symmetric about θ_0. Assuming without loss of generality that $\theta_0 = 0$ (as in Table 1), note that the risk function of (T_c^*, δ^*) is an even function ($R_c(\theta) = R_c(-\theta)$). Table 1 shows relative constancy of $R_c(c^{1/2}\mu)$ for the test (T_c^*, δ^*) as c varies from 10^{-2} to 10^{-4}, particularly when μ is not too large, in agreement with the fact that $R_c(c^{1/2}\mu)$ converges, as $c \to 0$, uniformly over bounded μ intervals (Lemma 4).

Table 2 compares the risk of (T_c^*, δ^*) with that of the approximate Bayes test (T_c, δ^*) defined in (4.2) using the continuous-time Bayes boundary g_0. It shows that there is little difference in risk performance between the two tests. Table 2 also compares the risk at θ of the test (T_c^*, δ^*) with that of the (fictitious) optimal nonsequential test $(n_{c,\theta}, \delta^*)$ that chooses the best fixed sample size, assuming knowledge of θ and allowing fractional observations. First note that for $\theta > 0$,

$$cn + P_\theta\{(n, \delta^*) \text{ rejects } K\} = cn + \Phi(-\theta n^{1/2})$$

TABLE 2

Risk function (= sampling cost + error probability) of the fictitious optimal fixed sample size test $(n_{c,\theta}, \delta^*)$ assuming knowledge of θ, compared with the risk functions of (T_c^*, δ^*) and of (T_c, δ^*) in testing H: $\theta < 0$ versus K: $\theta > 0$ for the mean θ of a normal distribution with unit variance. Cost (c) per observation $= 10^{-4}$.

$\mu\ (= c^{-1/2}\theta)$	Sampling Cost t_μ	Error $\Phi(-\mu t_\mu^{1/2})$	Risk of $(n_{c,\theta}, \delta^*)$	Risk of (T_c^*, δ^*)	Risk of (T_c, δ^*)
100	0.00126	0.00019	0.00145	0.0012	0.0012
80	0.00185	0.00029	0.00214	0.0017	0.0017
60	0.00299	0.00051	0.00350	0.0026	0.0026
40	0.00581	0.00115	0.00696	0.0046	0.0046
20	0.01709	0.00447	0.02156	0.0125	0.0125
10	0.04485	0.01709	0.06194	0.0337	0.0337
5	0.09426	0.06238	0.15664	0.1023	0.1027
2.5	0.11854	0.19469	0.31323	0.2568	0.2552
0.5	0.00992	0.48014	0.49006	0.5210	0.5240

is minimized when $\frac{1}{2}\theta(2\pi n)^{-1/2}\exp(-\theta^2 n/2) = c$, treating n as a continuous variable. Hence, letting t_μ be the solution of the equation

$$(2\pi t)^{1/2} = \tfrac{1}{2}\mu \exp\left(-\tfrac{1}{2}\mu^2 t\right), \tag{5.6}$$

the optimal fixed sample size $n_{c,\theta}$ is given by

$$n_{c,\theta} = c^{-1}t_\mu \quad \text{with } \mu = c^{-1/2}\theta, \tag{5.7}$$

and the risk of the test $(n_{c,\theta}, \delta^*)$ is given by

$$cn_{c,\theta} + \Phi\left(-\theta n_{c,\theta}^{1/2}\right) = t_\mu + \Phi\left(-\mu t_\mu^{1/2}\right). \tag{5.8}$$

Note that as $\mu \to \infty$,

$$t_\mu = \mu^{-2}\left\{4\log\mu - \log\log\mu - \log 8 - \tfrac{1}{2}\log(2\pi) + o(1)\right\},$$

so $\Phi(-\mu t_\mu^{1/2}) = O(\mu^{-2}) = o(t_\mu)$. Therefore, the risk (5.8) of the fictitious optimal fixed sample size test $(n_{c,\theta}, \delta^*)$ assuming knowledge of θ is asymptotically equivalent to $(2\log\mu^2)/\mu^2$. From (2.12) and (4.7), it follows that the risk at θ of the test (T_c^*, δ^*) is also asymptotically equivalent to $(2\log\mu^2)/\mu^2$ as $c \to 0$ such that $|\mu| = c^{-1/2}|\theta| \to \infty$, but $\theta = o(|\log c|^{1/2})$ [cf. Lai, Robbins and Siegmund (1983), Lemma 2]. Table 2 shows that (T_c^*, δ^*) compares favorably with $(n_{c,\theta}, \delta^*)$ over the entire range of values of μ, not only when μ is large.

The fictitious optimal fixed sample size $n_{c,\theta}$ serves as a benchmark for comparison of the test (T_c, δ^*), or its approximation (T_c^*, δ^*), with the class of fixed sample size tests. We now compare (T_c, δ^*) with a realizable fixed sample size test of H: $\theta < 0$ versus K: $\theta > 0$ that is optimal in some sense. Since (T_c, δ^*) is an asymptotic solution, in view of Theorem 3, to the Bayes problem of testing sequentially H versus K under the 0–1 loss, cost c per observation and a prior distribution F of θ that has a positive continuous density f in some neighborhood of 0, it is natural to consider the corresponding Bayes nonsequential test.

Assuming F to be symmetric about 0, the sample size n_F of the Bayes nonsequential test can be determined by minimizing

$$cn + \int_0^\infty P_\theta\{S_n < 0\}\, dF(\theta) + \int_{-\infty}^0 P_\theta\{S_n > 0\}\, dF(\theta)$$
$$= cn + 2\int_0^\infty \Phi(-n^{1/2}\theta)\, dF(\theta).$$

Treating n as a continuous variable, an asymptotic analysis shows that as $c \to 0$,

$$n_F \sim \left(f^2(0)/2\pi\right)^{1/3} c^{-2/3} \tag{5.9}$$

and that the Bayes risk of this Bayes nonsequential test is of the order $3(f^2(0)/2\pi)^{1/3}c^{1/3}$ as $c \to 0$, which is larger than the order of constant times $c^{1/2}$ for the Bayes risk of the test (T_c, δ^*) given in (4.4).

We next compare the test (T_c, δ^*) with the class of SPRTs for testing the sign of a normal mean under the 0–1 loss. Again taking any prior distribution F which is symmetric about 0 and which has a positive continuous density f in some neighborhood of 0, we note that the Bayes SPRT stops sampling at stage

$$\tau(b) = \inf\{n: |S_n| \ge b\} \tag{5.10}$$

and rejects H: $\theta < 0$ or K: $\theta > 0$ according as $S_{\tau(b)} > 0$ or $S_{\tau(b)} < 0$, where b is the value of $x \ge 0$ that minimizes the Bayes risk

$$2\int_0^\infty \left\{cE_\theta\tau(x) + P_\theta\left[S_{\tau(x)} < 0\right]\right\} dF(\theta). \tag{5.11}$$

Standard computations show that $b \to \infty$ as $c \to 0$ and

$$\begin{aligned}\int_0^\infty P_\theta[S_{\tau(b)} < 0]\, dF(\theta) &\sim \int_0^\infty e^{-2b\theta}/(1 + e^{-2b\theta})\, dF(\theta)\\ &\sim \tfrac{1}{2}b^{-1}f(0)\log 2,\end{aligned} \tag{5.12}$$

$$\begin{aligned}\int_0^\infty E_\theta\tau(b)\, dF(\theta) &\sim b\int_0^\infty \theta^{-1}(1 - e^{-2b\theta})/(1 + e^{-2b\theta})\, dF(\theta)\\ &\sim f(0) b\log b.\end{aligned} \tag{5.13}$$

From (5.12) and (5.13), it then follows that (5.11) is minimized at

$$x = b \sim (\log 2)^{1/2}(c|\log c|)^{-1/2} \quad \text{as } c \to 0. \tag{5.14}$$

Therefore, the Bayes risk (5.11) of $(\tau(b), \delta^*)$, the Bayes SPRT, is of the order $2(\log 2)^{1/2}f(0)(c|\log c|)^{1/2}$ as $c \to 0$, which is larger than the order of constant times $c^{1/2}$ for the Bayes risk of the test (T_c, δ^*) given in Theorem 3.

In Table 3, we apply the rule (T_c^*, δ^*) to test H: $p < \frac{1}{2}$ versus K: $p > \frac{1}{2}$ for the mean p of the Bernoulli distribution $P\{X = 1\} = p = 1 - P\{X = 0\}$. The natural parameter is $\theta = \log[p/(1-p)]$, and the hypotheses can be written as H: $\theta < 0$ versus K: $\theta > 0$. Note that $\psi(\theta) = \log(1 - p) = -\log(1 + e^\theta)$, so (1.3) does not hold if A is the entire parameter space Θ. We assume, therefore, that p is known to lie between 0.05 and 0.95, so that the maximum likelihood estimate of p at stage n is $\hat{p}_n = 0.05 \vee (\bar{X}_n \wedge 0.95)$. In terms of the parameter p, the

TABLE 3

Expected sample size, error probability and risk of the test (T_c^*, δ^*) *of* H: $p < \frac{1}{2}$ *versus* K: $p > \frac{1}{2}$ *for the mean* p *of a Bernoulli distribution. Cost* (c) *per observation equals* 10^{-4}. *Each result is based on* 1600 *simulation runs. Also given for comparison are the fictitious optimal fixed sample size test* ($n_{c,p}, \delta^*$) *assuming knowledge of* p *and the test* ($\hat{T}_c, \delta^*$).

		The test (T_c^*, δ^*)			Risk of	The test ($n_{c,p}, \delta^*$)		
μ	p	$E_pT_c^*$	Error probability	Risk	the test ($\hat{T}_c, \delta^*$)	$n_{c,p}$	Error probability	Risk
−100	0.119	16	0.0000	0.0016	0.0015	13	0.00030	0.00160
−80	0.158	20	0.0000	0.0020	0.0020	19	0.00038	0.00228
−60	0.231	28	0.0000	0.0028	0.0028	31	0.00053	0.00363
−40	0.310	48	0.0000	0.0048	0.0048	59	0.00116	0.00706
−20	0.401	123	0.0004	0.0127	0.0128	71	0.00452	0.02162
−10	0.450	280	0.0063	0.0343	0.0350	449	0.01707	0.06197
−5	0.475	565	0.0489	0.1054	0.1058	941	0.06253	0.15663
−2.5	0.488	824	0.1699	0.2523	0.2519	1143	0.19897	0.31327
−0.5	0.498	1013	0.4384	0.5397	0.5397	99	0.48011	0.49001

Kullback–Leibler information number can be written as

$$I(\tilde{p}, p) = \tilde{p}\log(\tilde{p}/p) + (1-\tilde{p})\log[(1-\tilde{p})/(1-p)].$$

Thus, $T_c^* = \inf\{n \geq 1: I(\hat{p}_n, \frac{1}{2}) \geq n^{-1}g_0(cn)\}$. For comparison with the normal case in Tables 1 and 2, define $\mu = \frac{1}{2}c^{-1/2}\theta$, the factor $\frac{1}{2}$ being the standard deviation of X at $\theta = 0$ (or $p = \frac{1}{2}$).

Without assuming prior knowledge that p lies between 0.05 and 0.95, we consider also the stopping rule $\hat{T}_c = \inf\{n \geq 1: I(\overline{X}_n, \frac{1}{2}) \geq n^{-1}g_0(cn)\}$, where we define $I(x, \frac{1}{2}) = \log 2$ (by continuity) for $x = 0$ or 1. The performance of ($\hat{T}_c, \delta^*$) is very close to that of (T_c^*, δ^*) as its risk function reported in Table 3 shows. Table 3 also compares the risk at p of the test (T_c^*, δ^*) with that of the (fictitious) optimal nonsequential test ($n_{c,p}, \delta^*$) that chooses the best fixed sample size, assuming knowledge of p. This best fixed sample size is obtained by choosing n to minimize

$$R_c(p) = cn + P_p\{S_n > \tfrac{1}{2}n\} + \tfrac{1}{2}P_p\{S_n = \tfrac{1}{2}n\}$$

for $p < \frac{1}{2}$, noting that $R_c(p)$ is symmetric about $p = \frac{1}{2}$ and that S_n has the binomial $B(n, p)$ distribution. The optimal values of n for various values of p are given in Table 3 for $c = 10^{-4}$. Note that the risk, as a function of μ, of this fictitious optimal fixed sample size test in the Bernoulli case is quite well approximated by that in the normal case of Table 2 over the wide range of μ values. Moreover, Table 3 shows that the risk of the test (T_c^*, δ^*) compares favorably with that of ($n_{c,p}, \delta^*$).

The results of Tables 1–3 show that the sequential tests (T_c, δ^*) and (T_c^*, δ^*) are able to adapt to the wide range of possible values of the unknown θ in striking an optimal balance between sampling cost and error probability. Theorem 4, whose proof is given in Section 6, establishes for the general exponential family that the risk at θ of (T_c^*, δ^*) or of (T_c, δ^*) is asymptotically equivalent to that of the fictitious optimal fixed sample size test assuming knowledge of θ as $c \to 0$ and $\theta \to \theta_0$ such that $c^{-1}(\theta - \theta_0)^2 (= \mu^2/\psi''(\theta_0)) \to \infty$.

THEOREM 4. *Let $X_1, X_2, \ldots$ be i.i.d. random variables whose common density $f_\theta(x)$ belongs to the exponential family* (1.1). *Let $\theta_0 \in A$, where A is an open interval satisfying* (1.3). *Define $\hat{\theta}_n$ by* (1.4) *and $I(\theta, \lambda)$ by* (1.2). *Let $R_c(\theta; T)$ denote the risk function* (5.4) *of the test (T, δ^*) of H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ with the* 0–1 *loss and cost c per observation, where T is the stopping rule and δ^* is the terminal decision rule* (5.5). *Let $d_c < D_c$ be positive numbers such that as $c \to 0$,*

$$(5.15)\quad d_c \to 0 \quad and \quad d_c/c^{1/2} \to \infty, \qquad D_c \to \infty \quad and \quad D_c = o\left(|\log c|^{1/2}\right).$$

(i) *Given $\theta \in A$, let $n_{c,\theta}$ be the value of n that minimizes $R_c(\theta; n)$ over all nonrandom sample sizes n. Then as $c \to 0$,*

$$(5.16)\qquad R_c(\theta; n_{c,\theta}) \sim \{c/I(\theta_0, \theta)\}\log\left[c^{-1}(\theta - \theta_0)^2\right]$$

uniformly in $\theta \in A$ such that $d_c \le |\theta - \theta_0| \le D_c$.

(ii) *Let g be a nonnegative function on $(0, \infty)$ satisfying condition* (3.5) *with $\xi > -\frac{3}{2}$ and such that $\sup_{t \ge a} g(t)/t < \infty$ for all $a > 0$. Define the stopping rule $T(g, c)$ by* (1.24). *Then as $c \to 0$,*

$$(5.17)\qquad R_c(\theta; T(g, c)) \sim \{c/I(\theta, \theta_0)\}\log\left[c^{-1}(\theta - \theta_0)^2\right]$$

uniformly in $\theta \in A$ such that $d_c \le |\theta - \theta_0| \le D_c$.

(iii) *As $\theta \to \theta_0$, $I(\theta_0, \theta) \sim I(\theta, \theta_0) \sim \frac{1}{2}\psi''(\theta_0)(\theta - \theta_0)^2$.*

We now consider the problem of testing H_0: $\theta \le -\varepsilon$ versus H_1: $\theta \ge \varepsilon$ for the mean θ of a normal distribution with variance 1. As in Section 3 (noting that $\theta^* = 0$ in this case), we will use the terminal decision rule which accepts the null or alternative hypothesis according as $\overline{X}_n < 0$ or $\overline{X}_n > 0$ when stopping occurs at stage n. Note that this is the same as the terminal decision rule δ^* in Tables 1 and 2. Assuming the 0–1 loss and cost c per observation, the risk $R(\theta)$ of a test (T, δ^*) at $\theta > 0$ is given by

$$\begin{aligned} R(\theta) &= cE_\theta T + P_\theta\{\overline{X}_T < 0\}, && \theta \ge \varepsilon, \\ &= cE_\theta T, && 0 \le \theta < \varepsilon. \end{aligned}$$

Note that $R(\theta) = R(-\theta)$ and that γ [defined in (5.3)] $= c^{-1/2}\varepsilon$.

Denote the stopping rule $N(g_\gamma, c)$ of Theorem 1(iii) by N_c^γ. Table 4 tabulates the risk of the test (N_c^γ, δ^*) for $\varepsilon = 0.05, 0.25$ and $c = 10^{-4}$. It shows that the risk is closely approximated by that of the approximating test (N_c^*, δ^*) defined in (5.2), whose expected sample size, error probability and risk are also tabulated. Also given for comparison in Table 4 is Schwarz's test (N, δ^*), where N is the stopping rule defined in (1.22). Note that the risk of Schwarz's test is considerably larger than that of (N_c^*, δ^*) within the indifference zone in both cases. Note also the steep change in error probability for (N_c^*, δ^*) within the narrow range

TABLE 4

Testing H_0: $\theta \leq -\varepsilon$ versus H_1: $\theta \geq \varepsilon$ for a normal mean with $c = 10^{-4}$. Each result is based on 1600 *simulations.*

(i) $\varepsilon = 0.05$ ($\gamma = 5$)

θ	Risk of (N_c^γ, δ^*)	The test (N_c^*, δ^*)			Schwarz's test (N, δ^*)		
		$E_\theta N_c^*$	Error	Risk	$E_\theta N$	Error	Risk
1	0.0011	12	0.0000	0.0012	17	0.0000	0.0017
0.6	0.0024	25	0.0000	0.0025	43	0.0000	0.0043
0.4	0.0043	44	0.0000	0.0044	89	0.0000	0.0089
0.2	0.0109	108	0.0006	0.0114	282	0.0000	0.0282
0.1	0.0346	227	0.0113	0.0340	798	0.0001	0.0799
0.05	0.1380	352	0.0987	0.1339	1725	0.0006	0.1731
0.03	0.0407	415	—	0.0415	2673	—	0.2673
0	0.0440	457	—	0.0457	4309	—	0.4309

(ii) $\varepsilon = 0.25$ ($\gamma = 25$)

θ	Risk of (N_c^γ, δ^*)	The test (N_c^*, δ^*)			Risk of Schwarz's (N, δ^*)
		$E_\theta N_c^*$	Error	Risk	
1	0.0009	9	0.0000	0.0009	0.0012
0.6	0.0017	17	0.0000	0.0017	0.0026
0.4	0.0029	27	0.0003	0.0030	0.0043
0.3	0.0066	35	0.0029	0.0064	0.0060
0.25	0.0129	39	0.0083	0.0122	0.0075
0.1	0.0060	60	—	0.0060	0.0139
0	0.0067	69	—	0.0069	0.0178

$0.05 \leq \theta \leq 0.2$ in the case $\varepsilon = 0.05$ and also within the range $0.25 \leq \theta \leq 0.4$ in the case $\varepsilon = 0.25$. Such a feature not only provides a very good balance between error probability and sampling cost for this range of θ values, but also enables the expected sample size to increase much more slowly than Schwarz's test as θ decreases from ε to 0, as shown in Table 4. For small values of c and ε, this slower rate of increase in expected sample size as θ decreases to 0 provides substantial reduction in Bayes risk over Schwarz's test, as will be discussed later.

Note that in the present normal case, Schwarz's stopping rule (1.22) reduces to

$$N = \inf\{n: |S_n| \geq (2n|\log c|)^{1/2} - \varepsilon n\}, \tag{5.18}$$

while (1.13) and (1.15) reduce to

$$E_\theta N \sim 2|\log c|/(|\theta| + \varepsilon)^2 \tag{5.19}$$

for every fixed θ and

$$r(N, \delta^*) \sim 2c|\log c| \int_{-\infty}^{\infty} (|\theta| + \varepsilon)^{-2}\, d\pi(\theta) \sim \inf_{(T,\delta)} r(T, \delta) \tag{5.20}$$

as $c \to 0$. The asymptotic relations (5.19) and (5.20) assume fixed $\varepsilon > 0$, as pointed out in Section 1, and they also hold when N is replaced by N_c^γ or by its approximation N_c^*, in view of Lemma 6 (in Section 6) and Theorem 1.

We now show that (5.19), in fact, holds uniformly in $0 < |\theta| + \varepsilon \le B$ for every $B > 0$. First assume that $\theta \ge 0$ and consider $N_+ = \inf\{n: S_n \ge (2n|\log c|)^{1/2} - \varepsilon n\}$ instead of N. Introduce the transformation

$$t = (\theta + \varepsilon)^2 n, \qquad w(t) = (\theta + \varepsilon)(S_n + \varepsilon n).$$

Treating n as a continuous variable, $w(t)$ is a Wiener process with drift coefficient 1 and $(\theta + \varepsilon)^2 N_+$ has the same distribution as $\tau = \inf\{t: w(t) \ge (2t|\log c|)^{1/2}\}$. Since $E\tau \sim 2|\log c|$ as $c \to 0$, it then follows that (5.19) holds uniformly in $0 \le \theta \le B$ and $0 < \varepsilon \le B$. The case $\theta \le 0$ can be treated similarly.

Since

$$\begin{aligned} P_\theta\left\{S_n \le -(2n|\log c|)^{1/2} + \varepsilon n\right\} &= \Phi\left(-(2|\log c|)^{1/2} - (\theta - \varepsilon)n^{1/2}\right) \\ &= o\left(c \exp\left\{-\tfrac{1}{2}(\theta - \varepsilon)^2 n\right\}\right) \end{aligned}$$

uniformly in $\theta \ge \varepsilon$ and $n \ge 1$ as $c \to 0$, and since N is bounded by $2\varepsilon^{-2}|\log c| + 1$, it then follows that as $c \to 0$ and $\varepsilon \to 0$,

$$\sup_{2\varepsilon \ge \theta \ge \varepsilon} \left(P_\theta\{S_N < 0\} + P_{-\theta}\{S_N > 0\}\right) = o(c\varepsilon^{-2}|\log c|), \tag{5.21}$$

$$P_\theta\{S_N < 0\} + P_{-\theta}\{S_N > 0\} = o(c(\theta - \varepsilon)^{-2}) \quad \text{uniformly in } \theta \ge 2\varepsilon. \tag{5.22}$$

Let π be a prior distribution of θ that has a positive continuous density π' in some neighborhood of 0. Then from (5.21), (5.22) and (5.19), it follows that as $c \to 0$ and $\varepsilon \to 0$,

$$r(N, \delta^*) \sim 2c|\log c| \int_{-\infty}^{\infty} (|\theta| + \varepsilon)^{-2}\, d\pi(\theta) \sim 4\pi'(0)\varepsilon^{-1}c|\log c|. \tag{5.23}$$

Comparing this order of magnitude for the Bayes risk of Schwarz's test (N, δ^*) with that of the test (N_c^γ, δ^*) given by Theorem 1, we find that Schwarz's test is no longer asymptotically Bayes as $\varepsilon \to 0$. In fact, by Theorem 1(iii), as $c \to 0$ and $\varepsilon \to 0$ such that $\gamma = c^{-1/2}\varepsilon$ converges to a finite limit, $r(N_c^\gamma, \delta^*)$ is of the order of constant times $c^{1/2}$. Note that in this case, $\varepsilon^{-1}c = \gamma^{-1/2}c^{1/2}$, so the Bayes risk (5.23) of Schwarz's test is of a larger order of magnitude than $r(N_c^\gamma, \delta^*)$. Moreover, by Theorem 1(ii), as $c \to 0$ and $\varepsilon \to 0$ such that $c^{-1/2}\varepsilon \to \infty$,

$$r(N_c^\gamma, \delta^*) \sim 4\pi'(0)\varepsilon^{-1}c(\log c^{-1} - \log \varepsilon^{-2}),$$

while the Bayes risk $r(N, \delta^*)$ of Schwarz's test is of the order $4\pi'(0)\varepsilon^{-1}c \log c^{-1}$ by (5.23).

Theorem 2 says that the test (N_c^γ, δ^*), or its approximation (N_c^*, δ^*), is asymptotically optimal as $c \to 0$ and $\gamma = c^{-1/2}\varepsilon \to \infty$ for minimizing the maximum expected sample size subject to error constraints under H_0: $\theta \le -\varepsilon$ and H_1: $\theta \ge \varepsilon$. Moreover, it also asymptotically minimizes certain weighted averages $\int E_\theta T\, d\pi(\theta)$ of the expected sample size function $E_\theta T$. As an illustration of this even when γ is only of moderate size, consider the case $\varepsilon = 0.25$ and $\gamma = 25$

studied in Table 4(ii). Subject to the error constraints $P_{-\varepsilon}\{(T,\delta)$ rejects $H_0\} \le \alpha$ and $P_\varepsilon\{(T,\delta)$ rejects $H_1\} \le \alpha$, Hoeffding (1960) showed that

$$E_0 T \ge 2\varepsilon^{-2}\{1 - \log 2\alpha - (1 - 2\log 2\alpha)^{1/2}\}. \tag{5.24}$$

For $\varepsilon = 0.25$, Hoeffding's lower bound for $E_0 T$ ($\le \max_\theta E_\theta T$) is 66.1 when $\alpha = 0.0083$ and is 64.4 when $\alpha = 0.009$. From Table 4(ii), the maximum expected sample size of the test (N_c^*, δ^*) is 69 (at $\theta = 0$), which does not differ much from Hoeffding's lower bound. The values of $E_\theta N_c^*$ given in Table 4(ii) also show substantial savings in sample size for θ near 0 when compared to SPRTs, fixed sample size tests and Schwarz's tests that have approximately the same error probabilities at $\pm\varepsilon$. In particular, the optimal fixed sample size test with error probability 0.0082 at ± 0.25 takes 92 observations. The SPRT that stops sampling as soon as $|S_n| \ge 9$ has error probability 0.0082 at ± 0.25, while Schwarz's stopping rule (5.18) with $c = 1/250$ (instead of $c = 10^{-4}$ as in Table 4) gives an error probability of 0.0081 at ± 0.25. The expected sample sizes $E_\theta T$ of these two tests are listed next and are compared with the values of $E_\theta N_c^*$ given in Table 4(ii). Each of these results is based on 1600 simulations:

θ	1	0.6	0.4	0.3	0.25	0.1	0
SPRT	10	16	24	32	38	71	92
Schwarz ($c = 1/250$)	8	15	25	36	42	72	85
N_c^* ($c = 10^{-4}$)	9	17	27	35	39	60	69

6. Asymptotic approximations of risk and proofs of Theorems 1, 2 and 4. To prove Theorems 1 and 2, we first derive asymptotic approximations to the expected sample size and error probabilities of the test $(N(g,c), \delta^*)$ proposed in Section 3 for testing H_0: $\theta \le \theta_0$ versus H_1: $\theta \ge \theta_1$. Lemma 6 shows that the test has a bounded stopping rule with sample size bounded by

$$n_c = \inf\{n: nJ(\theta^*) \ge g(cn)\}, \qquad \inf\varnothing = \infty, \tag{6.1}$$

where J and θ^* are defined in (1.10) and (1.11). Note that

$$N(g,c) = \inf\{n: nJ(\hat\theta_n) \ge g(cn)\}. \tag{6.2}$$

Throughout the sequel, let $d_c < D_c$ be positive numbers satisfying (5.15).

LEMMA 6. *Let g be a nonnegative function on $(0,\infty)$ such that $g(t) \sim \log t^{-1}$ as $t \to 0$. Then $N(g,c) \le n_c$. Moreover, as $c \to 0$:*

(i) $(\infty >) n_c \sim \langle \log(c^{-1}(\theta_1 - \theta_0)^2)/J(\theta^*)\rangle$ *uniformly in $\theta_0, \theta_1 \in A$ with $\theta_1 - \theta_0 \ge d_c$, where $\langle x\rangle$ denotes the smallest positive integer greater than or equal to x.*

(ii) $E_\theta N(g,c) \sim \{\log(c^{-1}J(\theta))\}/J(\theta)$ *uniformly in $\theta, \theta_0, \theta_1 \in A$ with $d_c^2 \le J(\theta) \le D_c^2$.*

(iii) $\sup\{E_\theta N(g,c)$: $\theta, \theta_0, \theta_1 \in A$ *with* $J(\theta) \ge k\} = O(|\log c|)$ *for every* $k > 0$, PROOF. Let $b = \frac{1}{2}\inf_{\theta \in A}\psi''(\theta)$ and $B = \frac{1}{2}\sup_{\theta \in A}\psi''(\theta)$. Then by (1.2),

$$b(\theta - \lambda)^2 \le I(\theta, \lambda) \le B(\theta - \lambda)^2 \quad \text{for all } \theta, \lambda \in A, \tag{6.3}$$

and since $\theta_0, \theta_1 \in A$,

$$Bc^{-1}(\theta_1 - \theta_0)^2 \ge J(\theta^*)/c \ge bc^{-1}(\theta_1 - \theta_0)^2/4. \tag{6.4}$$

By (6.4), as $c \to 0$ such that $(\theta_1 - \theta_0)^2/c \to \infty$,

$$\inf\{t > 0: tJ(\theta^*)/c \ge g(t)\} \sim (c/J(\theta^*))\log\left(c^{-1}(\theta_1 - \theta_0)^2\right). \tag{6.5}$$

Writing $t = cn$, we obtain (i) from (6.5). Since $J(\theta) \ge J(\theta^*)$ for all θ, $N(g, c) \le n_c$ by (6.1) and (6.2). Finally, an argument similar to the proof of Theorem 3 of Lai (1988) can be used to prove (ii) and (iii). □

We now make use of the boundary crossing theory developed in Lai (1988) to study the error probabilities of the test. Define α, β by (3.11) and, more generally, define

$$\begin{aligned} \alpha(\theta) &= P_\theta\left\{\hat{\theta}_{N(g,c)} > \theta^*\right\} \text{ if } \theta \le \theta_0, \\ \beta(\theta) &= P_\theta\left\{\hat{\theta}_{N(g,c)} \le \theta^*\right\} \text{ if } \theta \ge \theta_1, \end{aligned} \tag{6.6}$$

so that $\alpha = \alpha(\theta_0)$ and $\beta = \beta(\theta_1)$. For $\theta \le \theta_0 (< \theta^*)$,

$$\begin{aligned} \alpha(\theta) = \int_{\{\hat{\theta}_{N(g,c)} > \theta^*\}} & \exp\Big\{-N(g, c)\big[(\theta_0 - \theta)\big(\overline{X}_{N(g,c)} - \psi'(\theta^*)\big) \\ & + \psi'(\theta^*)(\theta_0 - \theta) - (\psi(\theta_0) - \psi(\theta))\big]\Big\}\, dP_{\theta_0} \\ \le P_{\theta_0}\left\{\hat{\theta}_{N(g,c)} > \theta^*\right\} &= \alpha(\theta_0). \end{aligned}$$

By a similar argument for $\beta(\theta)$, we therefore have

$$\alpha(\theta) \le \alpha \quad \text{for } \theta \le \theta_0, \qquad \beta(\theta) \le \beta \quad \text{for } \theta \ge \theta_1. \tag{6.7}$$

The following two lemmas study the asymptotic behavior of $\beta(\theta)$ for $\theta \ge \theta_1$ as $c \to 0$. Similar results hold for $\alpha(\theta)$ with $\theta \le \theta_0$.

LEMMA 7. *Let $g \in \mathscr{C}_\xi$. Then as $c \to 0$,*

$$\beta(\theta) = O\left(\left\{c/(\theta - \theta_1)^2\right\}\left\{\log\left(c^{-1}(\theta - \theta_1)^2\right)\right\}^{-\xi - 1/2}\right) \tag{6.8}$$

uniformly in $\theta, \theta_0, \theta_1 \in A$ with $\theta - \theta_1 \ge d_c$ and $\theta_1 - \theta_0 \ge d_c$.

PROOF. Note that for $\theta \in A$ with $\theta \ge \theta_1 + d_c$,

$$\begin{aligned} & P_\theta\left\{\hat{\theta}_{N(g,c)} \le \theta^*\right\} \\ & \quad \le P_\theta\left\{\hat{\theta}_n < \theta_1 \text{ and } I\left(\hat{\theta}_n, \theta_1\right) \ge n^{-1}g(cn) \text{ for some } n \le n_c\right\}, \end{aligned} \tag{6.9}$$

since $J(\lambda) = I(\lambda, \theta_1)$ is $\lambda \le \theta^*$. Hence the desired conclusion follows from Theorem 1(iii) of Lai (1988) and Lemma 1 of Lai (1987). □

LEMMA 8. *Let* $g \in \mathscr{C}_\xi$. *Then as* $c \to 0$,

$$\beta(\theta_1) = O\Big(\{c/(\theta_1 - \theta_0)^2\}\{\log(c^{-1}(\theta_1 - \theta_0)^2)\}^{-\xi + 3/2}\Big) \tag{6.10}$$

uniformly in $\theta_0, \theta_1 \in A$ *with* $\theta_1 - \theta_0 \ge d_c$.

PROOF. Let $\tau(\theta_1) = \inf\{n: I(\hat{\theta}_n, \theta_1) \ge n^{-1}g(cn)\}$. From (4.27) of Lai (1988) and Lemma 1 of Lai (1987), it follows that as $cm \to 0$,

$$P_{\theta_1}\{m \le \tau(\theta_1) < 2m\} = O(cm|\log cm|^{-\xi + 1/2}) \text{ uniformly in } \theta_1 \in A. \tag{6.11}$$

By (6.9) and (6.11),

$$\beta(\theta_1) = O\Big(c \sum_{j:\, 2^j \le n_c} 2^j |\log(2^j c)|^{-\xi + 1/2}\Big). \tag{6.12}$$

Let $d = \theta_1 - \theta_0$. Since $n_c = O(d^{-2}\log(c^{-1}d^2))$ by Lemma 6(i), the desired conclusion follows from (6.12) [see Lai (1988), Proof of Theorem 1(iii)]. □

Lemma 9 provides more precise estimates for $\alpha(\theta_0)$ and $\beta(\theta_1)$ when $\theta_1 - \theta_0 \to 0$.

LEMMA 9. *Let g be a nonnegative function on* $(0, \infty)$ *such that*

$$g(t) = \log t^{-1} + \xi \log\log t^{-1} + \rho + o(1) \quad \text{as } t \to 0 \tag{6.13}$$

for some constants ξ *and* ρ. *Then as* $c \to 0$ *and* $\theta_1 \to \theta_0$ *such that* $(\theta_1 - \theta_0)^2/c \to \infty$,

$$\alpha(\theta_0) \sim \beta(\theta_1) \sim 4\pi^{-1/2}e^{-\rho}(\psi''(\theta_0))^{-1}(c/d^2)\{\log(d^2/c)\}^{-\xi + 3/2}, \tag{6.14}$$

where $d = \theta_1 - \theta_0$.

PROOF. Let $0 < \zeta < 1$. For simplicity write N instead of $N(g, c)$. Since $I(\lambda, \theta_0) \sim \frac{1}{2}\psi''(\theta_0)(\lambda - \theta_0)^2$ as $\lambda \to \theta_0$, $\theta^* - \theta_0 \sim \frac{1}{2}d$ and, by Lemma 6(i),

$$n_c \sim 8\log(c^{-1}d^2)/(d^2\psi''(\theta_0)). \tag{6.15}$$

Therefore, $\{|\log(cn_c)|/n_c\}^{1/2} \sim (\psi''(\theta_0)/8)^{1/2}d$. Hence by Lemma 4 of Lai (1988) and Lemma 1 of Lai (1987), we can choose k sufficiently large so that

$$P_{\theta_0}\{|\hat{\theta}_N - \theta_0| \ge kd \text{ and } \zeta n_c \le N \le n_c\} = O((cn_c)^2). \tag{6.16}$$

Let $E = \{\hat{\theta}_N > \theta^*, |\hat{\theta}_N - \theta_0| \le kd \text{ and } \zeta n_c \le N \le n_c\}$. Defining the measure Q by

$$Q(B) = \int_{\theta_0 - 2kd}^{\theta_0 + 2kd} P_u(B)\, du$$

and choosing k sufficiently large, an argument similar to the proof of Lemma 8 of Lai (1988) shows that as $c \to 0$ and $d \to 0$ such that $d^2/c \to \infty$,

$$P_{\theta_0}(E) = \int_E (dP_{\theta_0}/dQ)\, dQ$$

$$\sim (\psi''(\theta_0)/2\pi)^{1/2} e^{-\rho} |\log cn_c|^{-\xi} cd \int_{-2k}^{2k} \left\{ \int_E N^{3/2} dP_{\theta_0 + td} \right\} dt \tag{6.17}$$

$$\sim (\psi''(\theta_0)/2\pi)^{1/2} e^{-\rho} |\log cn_c|^{-\xi} c\, dn_c^{3/2} \int_{1/2}^{(4\zeta)^{-1/2}} (4t^2)^{-3/2}\, dt.$$

By choosing ζ arbitrarily small, the desired conclusion follows from (6.15)–(6.17) and the analog of (6.11) with θ_0 replacing θ_1. □

While Lemma 9 shows that the order for $\beta(\theta_1)$ given by Lemma 8 is minimal when $\theta_1 \to \theta_0$, Lemma 10, which can be proved by a similar method, shows that it is also minimal for fixed θ_0, θ_1 as $c \to 0$.

LEMMA 10. *Let g be a positive function on $(0, \infty)$ such that*

$$g(t) = \log t^{-1} + (\xi + o(1)) \log\log t^{-1} \quad \text{as } t \to 0. \tag{6.18}$$

Then with θ_0, θ_1 fixed, as $c \to 0$,

$$c|\log c|^{-\xi' + 3/2} \le \beta(\theta_1) \le c|\log c|^{-\xi'' + 3/2} \tag{6.19}$$

for every $\xi' > \xi > \xi''$. The same inequality also holds for $\alpha(\theta_0)$.

We now proceed to the proof of Theorems 1 and 2. This makes use of the preceding lemmas and Hoeffding's lower bound for the expected sample size stated in Lemma 11 in the context of the exponential family [cf. Hoeffding (1960), (1.4)].

LEMMA 11. *For the exponential family* (1.1), *let (T, δ) be a test of H_0: $\theta \in \Theta_0$ versus H_1: $\theta \notin \Theta_0$ such that $P_{\lambda_1}[(T, \delta)$ rejects $H_0] \le p_1$ and $P_{\lambda_2}[(T, \delta)$ rejects $H_1] \le p_2$, where $0 < p_1 + p_2 < 1$, Θ_0 is a given subset of the natural parameter space Θ and $\lambda_1 \in \Theta_0$, $\lambda_2 \notin \Theta_0$. Then for every $\theta \in \Theta$,*

$$E_\theta T \ge \zeta^{-1}\{|\log(p_1 + p_2)| - \tfrac{1}{2}(\sigma/\zeta^{1/2})|\log(p_1 + p_2)|^{1/2}\},$$

where $\zeta = \max\{I(\theta, \lambda_1), I(\theta, \lambda_2)\}$, $\sigma^2 = \mathrm{Var}_\theta[(\lambda_2 - \lambda_1)X] = (\lambda_2 - \lambda_1)^2 \psi''(\theta)$.

PROOF OF THEOREM 2. From Lemmas 7–10, (3.12) and (3.14) follow. Moreover, the desired conclusions (3.13), (3.15) and (3.16) on $E_\theta N(g, c)$ follow from Lemma 6 and Lemma 11 (with $\lambda_1 = \theta_0$, $\lambda_2 = \theta_1$, $p_1 = \alpha$, $p_2 = \beta$), together with (3.12), (3.14) and the following two asymptotic relations as $c \to 0$ and $\theta_1 - \theta_0 = d \to 0$ such that $d^2/c \to \infty$:

$$J(\theta^*) = \inf_\theta J(\theta) \sim \psi''(\theta_0) d^2/8, \tag{6.20}$$

$$\int_{\theta_0 - \rho}^{\theta_0 + \rho} \{\log(c^{-1} J(\theta))\} \pi'(\theta)\, d\theta / J(\theta) \sim (8\pi'(\theta_0)/\psi''(\theta_0)) d^{-1} \log(d^2/c), \tag{6.21}$$

where $\rho > 0$ is such that π' is positive and continuous on $[\theta_0 - \rho, \theta_0 + \rho]$. □

Proof of Theorem 1. (i) For fixed $\theta_0 < \theta_1$, it follows from Theorem 2(i) that as $c \to 0$,

$$c \int_A E_\theta N(g, c)\, d\pi(\theta) \sim c|\log c| \int_A d\pi(\theta)/J(\theta), \tag{6.22}$$

noting that $N(g, c) \le n_c = O(|\log c|)$. If $\xi > \frac{1}{2}$, then from (6.7) and Lemma 8,

$$\int_{\theta \le \theta_0} \alpha(\theta)\, d\pi(\theta) + \int_{\theta \ge \theta_1} \beta(\theta)\, d\pi(\theta) = o(c|\log c|). \tag{6.23}$$

Now assume that $\xi > -\frac{1}{2}$ and that (3.7) holds. Then by (3.7), (6.7) and Lemma 8,

$$\int_{\theta_0 - |\log c|^{-1}}^{\theta_0} \alpha(\theta)\, d\pi(\theta) + \int_{\theta_1}^{\theta_1 + |\log c|^{-1}} \beta(\theta)\, d\pi(\theta)$$
$$\le \rho |\log c|^{-1}\{\alpha(\theta_0) + \beta(\theta_1)\} = o(c|\log c|).$$

Moreover, since $-\xi - \frac{1}{2} < 0$, it follows from (3.7) and Lemma 7 that

$$\int_{\theta \le \theta_0 - |\log c|^{-1}} \alpha(\theta)\, d\pi(\theta) + \int_{\theta \ge \theta_1 + |\log c|^{-1}} \beta(\theta)\, d\pi(\theta)$$
$$= O\left(\left\{c|\log c|^{-\xi - 1/2} \int_{|\log c|^{-1}}^{\varepsilon} t^{-2}\, dt + c\right\}\right) = o(c|\log c|).$$

Hence (6.23) still holds under (3.8) when $\xi > -\frac{1}{2}$. From (6.22) and (6.23), $r(N(g, c), \delta^*) \sim c|\log c| \int_A d\pi(\theta)/J(\theta)$. It therefore remains to show that as $c \to 0$

$$\inf_{(T, \delta)} r(T, \delta) \ge (1 + o(1)) c|\log c| \int_A d\pi(\theta)/J(\theta). \tag{6.24}$$

To prove (6.24), it suffices to restrict to tests (T, δ) such that

$$r(T, \delta) \le c|\log c|^2, \tag{6.25}$$

when c is sufficiently small. Let $\eta > 0$. Since

$$\int_{\theta_0 - \eta}^{\theta_0} P_x[(T, \delta) \text{ rejects } H_0]\, d\pi(x) + \int_{\theta_1}^{\theta_1 + \eta} P_y[(T, \delta) \text{ rejects } H_1]\, d\pi(y)$$
$$\le c|\log c|^2$$

and since $\pi([\theta_0 - \eta, \theta_0]) > 0$ and $\pi([\theta_1, \theta_1 + \eta]) > 0$, there exist x, y with $\theta_0 \ge x \ge \theta_0 - \eta$, $\theta_1 \le y \le \theta_1 + \eta$ such that

$$\begin{aligned} &P_x[(T, \delta) \text{ rejects } H_0] \le c|\log c|^2/\pi([\theta_0 - \eta, \theta_0]), \\ &P_y[(T, \delta) \text{ rejects } H_1] \le c|\log c|^2/\pi([\theta_1, \theta_1 + \eta]). \end{aligned} \tag{6.26}$$

From (6.26) and Lemma 11, it then follows that

$$\begin{aligned} &\inf\{E_\theta T: (T, \delta) \text{ satisfies } (6.25)\} \\ &\ge (1 + o(1))|\log c| \Big/ \max_{\theta_0 \ge x \ge \theta_0 - \eta,\, \theta_1 \le y \le \theta_1 + \eta} \max\{I(\theta, x), I(\theta, y)\}. \end{aligned} \tag{6.27}$$

Since $r(T, \delta) \geq c \int E_\theta T \, d\pi(\theta)$ and since η can be chosen arbitrarily small, (6.24) follows from (6.27).

(ii) Let $d = \theta_1 - \theta_0$. By Theorem 2(ii), as $c \to 0$ and $\theta_1 \to \theta_0$ such that $d^2/c \to \infty$,

$$(6.28) \qquad c \int_A E_\theta N(g, c) \, d\pi(\theta) \sim (8\pi'(\theta_0)/\psi''(\theta_0)) c d^{-1} \log(d^2/c).$$

Moreover, since $\xi > -\frac{1}{2}$, it follows from (6.7) and Lemma 8 that

$$(6.29) \qquad \begin{aligned} &\int_{\theta_0 - d(\log d^2/c)^{-1}}^{\theta_0} \alpha(\theta) \, d\pi(\theta) + \int_{\theta_1}^{\theta_1 + d(\log d^2/c)^{-1}} \beta(\theta) \, d\pi(\theta) \\ &\quad = o\big(c d^{-1} \log(d^2/c)\big) \end{aligned}$$

and from Lemma 7 that

$$(6.30) \qquad \begin{aligned} &\int_{\theta \leq \theta_0 - d(\log d^2/c)^{-1}} \alpha(\theta) \, d\pi(\theta) + \int_{\theta \geq \theta_1 + d(\log d^2/c)^{-1}} \beta(\theta) \, d\pi(\theta) \\ &\quad = O\left(\left\{ c \int_{d(\log d^2/c)^{-1}}^{\varepsilon} t^{-2} \{\log(c^{-1} t^2)\}^{-\xi - 1/2} \, dt + c \right\}\right) \\ &\quad = o\big(c d^{-1} \log(d^2/c)\big). \end{aligned}$$

By (6.28)–(6.30), $r(N(g, c), \delta^*) \sim (8\pi'(\theta_0)/\psi''(\theta_0)) c d^{-1} \log(d^2/c)$.

Let $\eta > 0$. Let (T, δ) be a test such that $r(T, \delta) \leq c d^{-1} (\log d^2/c)^2$. An argument similar to the proof of (6.26) shows that there exist x, y with $\theta_0 \geq x \geq \theta_0 - \eta d$, $\theta_1 \leq y \leq \theta_1 + \eta d$ such that

$$P_x\big[(T, \delta) \text{ rejects } H_0\big] \leq 2 c d^{-2} (\log d^2/c)^2 / \eta \pi'(\theta_0),$$

$$P_y\big[(T, \delta) \text{ rejects } H_1\big] \leq 2 c d^{-2} (\log d^2/c)^2 / \eta \pi'(\theta_0),$$

provided that d is sufficiently small. Hence an application of Lemma 11 as before shows that

$$\inf_{(T, \delta)} r(T, \delta) \geq \{8\pi'(\theta_0)/\psi''(\theta_0) + o(1)\} c d^{-1} \log(d^2/c).$$

(iii) The desired conclusion (3.10) can be proved by an argument similar to the proof of Theorem 3 in Section 4. □

We now derive asymptotic properties of the risk functions (5.4) for the class of tests $(T(g, c), \delta^*)$ of H: $\theta < \theta_0$ versus K: $\theta > \theta_0$ considered in Theorem 4. For $\theta > \theta_0$, it follows from Theorem 1(i) of Lai (1988) that as $c \to 0$,

$$(6.31) \qquad E_\theta T(g, c) \sim \big\{\log\big[c^{-1}(\theta - \theta_0)^2\big]\big\} / I(\theta, \theta_0)$$

and from Theorem 1(iii) of Lai (1988) and Lemma 1 of Lai (1987) that

$$(6.32) \qquad P_\theta\big\{\hat{\theta}_{T(g, c)} < \theta_0\big\} = O\Big(c(\theta - \theta_0)^{-2} \big\{\log\big[c^{-1}(\theta - \theta_0)^2\big]\big\}^{-\xi - 1/2}\Big),$$

the convergence in (6.31) and (6.32) being uniform in $\theta \in A$ such that

$d_c \le \theta - \theta_0 \le D_c$. Since $\xi > -\frac{3}{2}$, the asymptotic formula (5.17) on the risk of $(T(g, c), \delta^*)$ follows from (6.31), (6.32) and (6.3). This proves Theorem 4(ii), and we now proceed to

Proof of Theorem 4(i). For $\theta < \theta_0$,

$$(6.33)\quad \begin{aligned} P_\theta\{\hat{\theta}_n > \theta_0\} &= \int_{\{\hat{\theta}_n > \theta_0\}} \exp\{(\theta - \theta_0)S_n - n(\psi(\theta) - \psi(\theta_0))\}\, dP_{\theta_0} \\ &= \int_{\{\hat{\theta}_n > \theta_0\}} \exp\{-nI(\theta_0, \theta) - (\theta_0 - \theta)(S_n - n\psi'(\theta_0))\}\, dP_{\theta_0}. \end{aligned}$$

Let $\Delta = (n\psi''(\theta_0))^{1/2}(\theta_0 - \theta)$ and $Z_n = \{S_n - n\psi'(\theta_0)\}/(n\psi''(\theta_0))^{1/2}$. Under P_{θ_0}, Z_n converges to the standard normal distribution. Since $\Delta > 0$ and since $\{\hat{\theta}_n > \theta_0\} = \{Z_n > 0\}$, it then follows from (6.33) that

$$e^{-nI(\theta_0, \theta)} \ge P_\theta\{\hat{\theta}_n > \theta_0\} \ge e^{-nI(\theta_0,\theta) - \Delta} P_{\theta_0}\{Z_n \ge 1\}.$$

Note that $P_{\theta_0}\{Z_n \ge 1\} \ge \frac{1}{2}\Phi(-1)$ for all large n, say $n \ge n_0$, and that there exists $C > 0$ such that $\Delta \le C\{nI(\theta_0, \theta)\}^{1/2}$ for all $n \ge 1$ and $\theta \in A \cap (-\infty, \theta_0)$, in view of (6.3). Using a similar argument for $P_\theta\{\hat{\theta}_n < \theta_0\}$ if $\theta > \theta_0$, it then follows that

$$(6.34)\quad \begin{aligned} \exp\{-nI(\theta_0, \theta)\} \ge p_\theta(n) \ge \tfrac{1}{2}\Phi(-1)\exp\Big\{ &-nI(\theta_0, \theta) \\ &- C\big[nI(\theta_0, \theta)\big]^{1/2}\Big\} \end{aligned}$$

for all $n \ge n_0$ and $\theta \in A$, where

$$\begin{aligned} p_\theta(n) &= P_\theta\{\hat{\theta}_n > \theta_0\} \quad \text{if } \theta < \theta_0, \\ &= P_\theta\{\hat{\theta}_n < \theta_0\} \quad \text{if } \theta > \theta_0. \end{aligned}$$

From (6.34), it follows that the value $n_{c,\theta}$ of n that minimizes $cn + p_\theta(n)$ satisfies

$$(6.35)\quad n_{c,\theta} \sim \log[I(\theta_0, \theta)/c]/I(\theta_0, \theta) \quad \text{and} \quad p_\theta(n_{c,\theta}) = o(cn_{c,\theta})$$

as $c \to 0$, uniformly in $\theta \in A$ such that $d_c \le |\theta - \theta_0| \le D_c$. From (6.35), the desired conclusion (5.16) follows. □

Acknowledgment. The author wishes to express his gratitude to Fridrik Baldursson for helpful discussions and valuable assistance in the numerical work.

REFERENCES

Breakwell, J. and Chernoff, H. (1964). Sequential tests for the mean of a normal distribution. II (large t). *Ann. Math. Statist.* **35** 162–173.

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

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

Chernoff, H. (1965b). Sequential tests for the mean of a normal distribution. IV (discrete case). *Ann. Math. Statist.* **36** 55–68.

CHERNOFF, H. (1968). Optimal stochastic control. *Sankhyā Ser. A* **30** 221–252.

CHERNOFF, H. and PETKAU, J. (1986). Numerical solutions for Bayes sequential decision problems. *SIAM J. Sci. Statist. Comput.* **7** 46–59.

FREEDMAN, D. (1971). *Brownian Motion and Diffusion*. Holden-Day, San Francisco.

HOEFFDING, W. (1960). Lower bounds for the expected sample size and average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.

KIEFER, J. and WEISS, L. (1957). Some properties of generalized sequential probabilty ratio tests. *Ann. Math. Statist.* **28** 57–75.

LAI, T. L. (1973). Optimal stopping and sequential tests which minimize the maximum expected sample size. *Ann. Statist.* **1** 659–673.

LAI, T. L. (1977). First exit times from moving boundaries for sums of independent random variables. *Ann. Probab.* **5** 210–221.

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

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

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

LORDEN, G. (1976). 2-SPRT's and the modified Kiefer–Weiss problem of minimizing an expected sample size. *Ann. Statist.* **4** 281–291.

LORDEN, G. (1980). Structure of sequential tests minimizing an expected sample size. *Z. Wahrsch. verw. Gebiete* **51** 291–302.

SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

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

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

WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.

WEISS, L. (1962). On sequential tests which minimize the maximum expected sample size. *J. Amer. Statist. Assoc.* **57** 551–566.

WONG, S. P. (1968). Asymptotically optimum properties of certain sequential tests. *Ann. Math. Statist.* **39** 1244–1263.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4065