Statistica Sinica **7**(1997), 33-51

ON OPTIMAL STOPPING PROBLEMS IN SEQUENTIAL HYPOTHESIS TESTING

Tze Leung Lai

Stanford University

Abstract: After a brief survey of a variety of optimal stopping problems in sequential testing theory, we give a unified treatment of these problems by introducing a general class of loss functions and prior distributions. In the context of a one-parameter exponential family, this unified treatment leads to relatively simple sequential tests involving generalized likelihood ratio statistics or mixture likelihood ratio statistics. The latter have been used by Robbins in his development of power-one tests, whose optimality properties are also discussed in this connection.

Key words and phrases: Bayes sequential tests, generalized likelihood ratio statistics, mixture likelihood ratios, optimal stopping, Wiener process approximations.

1. Introduction

Probability theory began with efforts to calculate the odds and to develop strategies in games of chance. Optimal stopping problems arose naturally in this context, determining when one should stop playing a sequence of games to maximize one's expected fortune. A systematic theory of optimal stopping emerged with the seminal papers of Wald and Wolfowitz (1948) and Arrow, Blackwell and Girschick (1949) on the optimality of the sequential probability ratio test (SPRT). The monographs by Chow, Robbins and Siegmund (1971), Chernoff (1972) and Shiryayev (1978) provide comprehensive treatments of optimal stopping theory, which has subsequently developed into an important branch of stochastic control theory. The subject of sequential hypothesis testing has also developed far beyond its original setting of a simple null versus a simple alternative hypothesis assumed by the SPRT. Although it is not difficult to formulate optimal stopping problems associated with optimal tests of composite hypotheses, these optimal stopping problems no longer have explicit solutions that are easily interpretable as in the case of the SPRT. Moreover, numerical solutions of the optimal stopping problems require precise specification of prior distributions, loss functions for wrong decisions and sampling costs, which may be difficult to come up with in practice. In Sections 2 and 3 we develop an asymptotic approach to solve approximately a general class of optimal stopping problems associated with sequential tests of composite hypotheses. The asymptotic solutions provide natural analogues of the classical likelihood ratio or generalized likelihood ratio (GLR) tests in nonsequential testing theory. In addition, we also show that the different procedures developed by Schwarz (1962), Chernoff (1961, 1965a,b) and Robbins (1970) are in fact asymptotic solutions corresponding to different loss functions and prior distributions within this general class of optimal stopping problems. In the remainder of this section we give a brief survey of various optimal stopping problems in sequential hypothesis testing theory.

1.1. Optimality of SPRTs and approximate optimality of 2-SPRTs

Let $X_1, X_2, \ldots,$ be i.i.d. with common density f with respect to some measure ν. To test the simple null hypothesis $H_0 : f = f_0$ versus the simple alternative $H_1 : f = f_1$, Wald's SPRT uses the likelihood ratio statistics $R_n = \prod_{i=1}^n (f_1(X_i)/f_0(X_i))$ and stops sampling at stage $N = \inf\{n \geq 1 : R_n \leq A \text{ or } R_n \geq B\}$. The Wald-Wolfowitz theorem states that this SPRT minimizes both $E_0(T)$ and $E_1(T)$ among all tests whose sample size T has a finite expectation under both H_0 and H_1 and whose type I and type II error probabilities are less than or equal to those of the SPRT. To prove this theorem, a Lagrange-multiplier-type approach to handle the error probability constraints leads to the so-called "auxiliary Bayes problem" of minimizing over all stopping rules T and terminal decision rules δ

$$p\{w_0 P_0[\delta \text{ rejects } H_0] + cE_0(T)\} + (1-p)\{w_1 P_1[\delta \text{ rejects } H_1] + cE_1(T)\}. \quad (1.1)$$

The optimal solution to this problem turns out to be an SPRT and the Wald-Wolfowitz theorem then follows by varying the parameters p, c, w_0 and w_1 (cf. Ferguson (1967)). This solution to the auxiliary Bayes problem also implies further refinements of the Wald-Wolfowitz theorem (cf. Simons (1976)).

Although the SPRT for testing θ_0 versus θ_1 can still be used to test one-sided composite hypotheses of the form $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1 (> \theta_0)$ in the case of parametric families with monotone likelihood ratio in θ and although it has minimum expected sample size at $\theta = \theta_0$ and $\theta = \theta_1$, its expected sample size may be quite unsatisfactory at other parameter points, particularly those between θ_0 and θ_1. In the case of an exponential family $f_\theta(x) = e^{\theta x - \psi(\theta)}$, Kiefer and Weiss (1957) considered the problem of minimizing $E_\lambda(T)$ at a fixed parameter λ subject to $P_{\theta_0}[\text{Reject } H_0] \leq \alpha$ and $P_{\theta_1}[\text{Reject } H_1] \leq \beta$. Again a Lagrange-multiplier-type argument leads to the Bayes problem of minimizing

$$pP_{\theta_0}[\text{Reject } H_0] + qP_{\theta_1}[\text{Reject } H_1] + c(1-p-q)E_\lambda(T). \quad (1.2)$$

Lai (1973) and Lorden (1980) studied this optimal stopping problem in the normal case and the general one-parameter exponential family, respectively, and showed that the asymptotic shape of the continuation region (as $c \to 0$) agrees with that of the 2-SPRT with stopping rule

$$T^*(B,B') = \inf\Big\{n \geq 1 : \prod_{i=1}^{n}(f_\lambda(X_i)/f_{\theta_0}(X_i)) \geq B \text{ or } \prod_{i=1}^{n}(f_\lambda(X_i)/f_{\theta_1}(X_i)) \geq B'\Big\}. \tag{1.3}$$

The 2-SPRT rejects H_0 if $\prod_1^{T^*}(f_\lambda(X_i)/f_{\theta_0}(X_i)) \geq B$ and rejects H_1 if $\prod_1^{T^*}(f_\lambda(X_i)/f_{\theta_1}(X_i)) \geq B'$. Letting $n(B,B')$ denote the infinum of $E_\lambda(T)$ over all sequential tests whose type I and type II error probabilities are less than or equal to those of the 2-SPRT with stopping rule (1.3), Lorden (1976) also showed that $E_\lambda T^*(B,B') - n(B,B') \to 0$ as $\min(B,B') \to \infty$.

1.2. Bayes sequential tests of one-sided hypotheses and their GLR approximations

Sobel (1953) and Schwarz (1962) studied Bayes sequential tests of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$ for the natural parameter θ of an exponential family, assuming a cost c per observation, a prior distribution G on the natural parameter space Θ and a loss function $\ell(\theta)$ for the wrong decision. Letting $\Theta_0 = (-\infty, \theta_0] \cap \Theta$ and $\Theta_1 = [\theta_1, \infty) \cap \Theta$, the Bayes risk of a sequential test (T, δ) with stopping rule T and terminal decision rule δ is given by

$$r(T,\delta) = c\int_\Theta E_\theta T dG + \int_{\Theta_0} \ell(\theta)P_\theta\{\delta \text{ rejects } H_0\}dG + \int_{\Theta_1} \ell(\theta)P_\theta\{\delta \text{ rejects } H_1\}dG. \tag{1.4}$$

Sobel showed that the optimal stopping rule for this problem has the form $\inf\{n : S_n \geq a_n(c) \text{ or } S_n \leq b_n(c)\}$, where $S_n = \sum_1^n X_i$. Assuming that $\ell(\theta) > 0$ for $\theta \notin (\theta_0, \theta_1)$ and that $G(I) > 0$ for every open interval $I \subset \Theta$, Schwarz proved the following asymptotic shape for the continuation region $\mathcal{B}(c)$ of the Bayes rule: As $c \to 0$,

$$\mathcal{B}(c)/|\log c| \to \Big\{(t,w) : 1 + \min_{i=0,1}(\theta_i w - t\psi(\theta_i)) > \sup_\theta(\theta w - t\psi(\theta))\Big\}. \tag{1.5}$$

Thus, writing $n = t|\log c|$ and $S_n = w|\log c|$, Schwarz's approximation to the Bayes rule stops sampling at stage

$$N_c = \inf\Big\{n \geq 1 : \max[\prod_{i=1}^{n}(f_{\widehat{\theta}_n}(X_i)/f_{\theta_0}(X_i)),\ \prod_{i=1}^{n}(f_{\widehat{\theta}_n}(X_i)/f_{\theta_1}(X_i))] \geq a_c\Big\} \tag{1.6}$$

with $\log a_c \sim \log c^{-1}$, and uses the terminal decision rule $\widetilde{\delta}$ that accepts H_0 iff $\widehat{\theta}_n < \theta^*$, where $\widehat{\theta}_n$ is the maximum likelihood estimate of θ, $\theta^* \in (\theta_0, \theta_1)$ is such that $I(\theta^*, \theta_0) = I(\theta^*, \theta_1)$ and

$$I(\theta,\lambda) = E_\theta \log(f_\theta(X_1)/f_\lambda(X_1)) = (\theta - \lambda)\psi'(\theta) - (\psi(\theta) - \psi(\lambda)) \tag{1.7}$$

denotes the Kullback-Leibler information number. Wong (1968) showed that as $c \to 0$,

$$\inf_{(T,\delta)} r(T,\delta) \sim c|\log c| \int_\Theta \{\max[I(\theta,\theta_0), I(\theta,\theta_1)]\}^{-1} dG(\theta) \sim r(N_c, \widetilde{\delta}), \qquad (1.8)$$

with $a_c = c^{-1}$ in (1.6). Note that Schwarz's stopping rule (1.6) can be regarded as replacing the λ in the 2-SPRT stopping rule (1.3) by its maximum likelihood estimate $\widehat{\theta}_n$ at every stage n and putting $B = B' = a_c$ in (1.3).

Schwarz's asymptotic theory assumes a fixed indifference zone (θ_0, θ_1) as $c \to 0$. Optimal stopping problems associated with sequential tests of one-sided hypotheses of the form $H_0 : \theta < \theta_0$ versus $H_1 : \theta > \theta_0$ without an indifference zone were first considered by Chernoff (1961, 1965a,b) and Moriguti and Robbins (1962). Moriguti and Robbins studied the Bayes test of $H_1 : p \le \frac{1}{2}$ versus $H_1 : p > \frac{1}{2}$, with respect to a Beta prior on a Bernoulli parameter p, so the posterior distribution of p is also a Beta distribution that can be expressed in terms of the number of successes and failures so far observed. They assumed a loss of $|p - \frac{1}{2}|$ for the wrong decision and derived numerical solutions and analytic approximations of the dynamic programming equations defining the value function and stopping boundary. Lindley and Barnett (1965) and Simons and Wu (1986) carried out further analysis of this optimal stopping problem for Bernoulli random variables. Chernoff (1961, 1965a,b) and Breakwell and Chernoff (1964) studied a similar problem of testing $H_0 : \theta < 0$ versus $H_1 : \theta > 0$ for the mean θ of a normal distribution with unit variance, assuming a loss of $|\theta|$ for the wrong decision, cost c per observation and a normal prior distribution on θ. Instead of the absolute value loss function, Bather (1962), Bickel and Yahav (1972) and Bickel (1973) considered the case of 0-1 loss for testing the sign of a normal mean.

Lai (1988a,b) gave a unified treatment of (i) the 0-1 loss and the absolute value loss as special cases of more general loss functions, (ii) the Bernoulli and normal distributions as special cases of exponential families, and (iii) one-sided hypotheses separated by an indifference zone considered by Schwarz (1962) and one-sided hypotheses without an indifference zone considered by Chernoff (1961, 1965a,b) and by Moriguti and Robbins (1962). The basic idea is to approximate the highly complex Bayes procedures by relatively simple sequential tests involving GLR statistics and easily interpretable and implementable time-varying boundaries for these test statistics. The methods and results will be discussed in greater detail in Section 2 where they are further extended to more general loss functions in a unified framework within which properly chosen constant or time-varying stopping boundaries for mixture or generalized likelihood ratio statistics are shown to provide asymptotically optimal solutions.

Extensions of Schwarz's approximation (1.5) of the Bayes rule to multiparameter testing problems were considered by Kiefer and Sacks (1963), Schwarz

(1968), Fortus (1979) and Woodroofe (1980). Recently Lai and Zhang (1994a,b) extended the ideas of Lai (1988a) to construct asymptotically optimal GLR tests in multiparameter exponential families under the 0-1 loss, with an indifference zone separating the null and alternative hypotheses and also without an indifference zone, in the case of one-sided hypotheses concerning some smooth scalar functions of the parameters (such as testing the sign of a normal mean in the presence of an unknown variance which can be regarded as a nuisance parameter).

1.3. Approximate optimality of sequential mixture likelihood ratio tests

To test a simple null hypothesis $H_0 : f = f_0$ based on i.i.d. observations $X_1, X_2, \ldots$ with a common density function f with respect to some σ-finite measure ν, suppose that one continues sampling until one has enough evidence against H_0, whereupon one rejects H_0. Assuming a simple alternative f_1 and a cost c per observation under f_1, Chow, Robbins and Siegmund (1971), pages 107-108, showed that the optimal stopping rule that minimizes $P_0(T < \infty) + cE_1(T)$ is of the form

$$\tau = \inf\Big\{n \geq 1 : (\prod_{i=1}^{n} f_1(X_i))/(\prod_{i=1}^{n} f_0(X_i)) \geq B\Big\} \ (\inf \emptyset = \infty), \tag{1.9}$$

which is the stopping rule of a one-sided SPRT. Choosing B such that $P_0(\tau < \infty) = \alpha$, the stopping rule (1.9) also minimizes $E_1(T)$ among all stopping times T with $P_0(T < \infty) \leq \alpha$, by a Lagrange-multiplier-type argument.

Instead of a simple alternative f_1, suppose that one has a parametric family of densities $\{f_\theta, \theta \in \Theta\}$ with respect to the measure $f_0 d\nu$, where the parameter space Θ is some subset of the real line. To treat composite alternatives, Robbins (1970) proposed to generalize (1.9) to

$$\tau_G = \inf\Big\{n \geq 1 : \int_\Theta \prod_{i=1}^{n} f_\theta(X_i) dG(\theta) / \prod_{i=1}^{n} f_0(X_i) \geq B\Big\}, \tag{1.10}$$

where G is a probability measure on Θ. Since $\{\prod_{i=1}^n (f_\theta(X_i)/f_0(X_i)), n \geq 1\}$ is a martingale under P_0, so is $\{\int_\Theta \prod_{i=1}^n (f_\theta(X_i)/f_0(X_i)) dG(\theta), n \geq 1\}$ and therefore

$$P_0\{\tau_G < \infty\} = P_0\Big\{\int_\Theta \prod_{i=1}^{n} (f_\theta(X_i)/f_0(X_i)) dG(\theta) \geq B \ \text{ for some } \ n \geq 1\Big\} \leq B^{-1},$$

which gives a simple upper bound for the type I error $P_0\{\tau_G < \infty\}$ of the sequential mixture likelihood ratio test. For the case of an exponential family $f_\theta(x) = e^{\theta x - \psi(\theta)}$ and Θ a closed interval $[a, b]$ not containing 0, Pollak (1978) proved the following asymptotic optimality property of the rule (1.10) in which

G has a positive continuous density function on $[a, b]$ with respect to Lebesgue measure: As $B \to \infty$,

$$\sup_{a \le \theta \le b} I(\theta, 0) E_\theta(\tau_G) = \log B + (\log \log B)/2 + O(1)$$
$$= \inf \Big\{ \sup_{a \le \theta \le b} I(\theta, 0) E_\theta(T) : T \text{ is a stopping time and } P_0(T < \infty) \le B^{-1} \Big\} + O(1), \quad (1.11)$$

where $I(\theta, \lambda)$ denotes the Kullback-Leibler information number (1.7). His proof uses certain bounds in the Bayesian optimal stopping problem of minimizing $\omega \int_a^b c E_\theta(T) dG(\theta) + (1 - \omega) P_0(T < \infty)$, which corresponds to cost c per observation under P_θ for $\theta \in [a, b]$ and unit loss upon stopping under P_0 and to a prior distribution $\omega G + (1 - \omega) G_0$ in which $0 < \omega < 1$ and G_0 puts all its mass at 0.

Without assuming the set of alternatives to be bounded away from the simple null hypothesis, Lerche (1986b) assumed a cost of $c\mu^2$ per observation under P_μ for $\mu \ne 0$ in testing $H_0 : \mu = 0$ for the drift coefficient μ of a Wiener process $w(t), t \ge 0$. Specifically he studied the Bayesian optimal stopping problem of minimizing

$$\rho_c(T) = \omega \int_{-\infty}^{\infty} c\mu^2 E_\mu(T)\, g(\mu) d\mu \;+\; (1 - \omega) P_0(T < \infty), \quad (1.12)$$

where g is the density function of a zero-mean normal distribution. Let $\mathcal{B}(c)$ be the continuation region of the Bayes rule and let $\mathcal{R}(a) = \{(x, t) : \beta(x, t) > a\}$, where $\beta(x, t)$ is the posterior risk if stopping occurs at time t when $w(t) = x$. Lerche showed that there exists $M > 2$ such that for every $c > 0$, $\mathcal{R}(Mc/(1 + Mc)) \subset \mathcal{B}(c) \subset \mathcal{R}(2c/(1 + 2c))$. He also used this to show that the Bayes risk (1.11) of a suitably chosen sequential mixture likelihood ratio test is asymptotically minimal up to an $o(c)$-term, i.e.,

$$\rho_c(\tau(c)) = \inf_T \rho_c(T) + o(c)$$
$$= \omega c \Big\{ 2 \log a_c + \log \log a_c + 2 + \log 2 - 4 \int_0^\infty \phi(x) \log x dx + o(1) \Big\} \quad (1.13)$$

as $c \to 0$, where $a_c = (1 - \omega)/(2\omega c)$, ϕ denotes the standard normal density and

$$\tau(c) = \inf \Big\{ t \ge 0 : \int_{-\infty}^{\infty} g(\theta) \exp(\theta w(t) - t\theta^2/2) d\theta \ge a_c \Big\}. \quad (1.14)$$

Moreover, in addition to the normal prior distribution on the real line as in (1.12), he also obtained similar results for the case of a half-normal prior distribution on $(0, \infty)$, i.e., with g in (1.12) given by $g(\theta) = \sqrt{2/\pi} \exp(-\theta^2/2)$ if $\theta > 0$ and $g(\theta) = 0$ if $\theta \le 0$.

In Section 3 we give a more general theory concerning the optimality of Robbins' test (1.10) and other "power-one tests", which Neyman (1971) described as a "remarkable achievement" in the subject of hypothesis testing.

2. A Unified Asymptotic Theory for One-sided Hypotheses

The preceding survey of optimal stopping problems in sequential hypothesis testing shows a wide diversity of loss/cost functions and asymptotically optimal solutions. A common property, however, of these asymptotically optimal sequential tests is that they all involve (simple, generalized or mixture) likelihood ratio statistics and can be represented as boundary crossing times of these likelihood ratio statistics. In this section we give a unified treatment which yields relatively simple approximations to the stopping boundaries of these problems and their generalizations.

We begin with a brief review of the different asymptotic theories of Chernoff (1961, 1965a,b) and Schwarz (1962) for testing one-sided composite hypotheses $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_1$ for the natural parameter of an exponential family, assuming cost c per observation. Chernoff only considered the normal case (with mean θ and variance 1) for the special loss function $\ell(\theta) = |\theta|$ for the wrong decision and under a normal prior distribution G with mean 0 and variance σ^2. The Bayes terminal decision rule δ^* accepts $H_0 : \theta < 0$ iff $S_n < 0$ when stopping occurs at stage n, and the optimal stopping problem is to find the stopping rule T that minimizes (1.4) with $\Theta_0 = (-\infty, 0)$, $\Theta_1 = (0, \infty)$ and $\delta = \delta^*$. To study the optimal stopping problem, Chernoff introduced the normalization $t = c^{2/3}(n + \sigma^{-2})$, $w = c^{1/3} S_n$ (which is different from Schwarz's normalization $t = n/|\log c|$ and $w = S_n/|\log c|$). With this normalization, Chernoff obtained a limiting continuation region of the form $\{(t, w) : |w| < h(t)\}$ as $c \to 0$. The stopping boundary $h(t)$ arises as the solution of the corresponding continuous-time stopping problem involving the Wiener process and an asymptotic analysis yields

$$h(t) \sim (4t)^{-1} \text{ as } t \to \infty, \; h(t) = \sqrt{t}\{3 \log t^{-1} - \log 8\pi + o(1)\}^{1/2} \text{ as } t \to 0. \quad (2.1)$$

Instead of the absolute value loss, Lai (1988a) considered the 0-1 loss and gave a unified treatment of the problem of testing (i) $H_0 : \theta < 0$ versus $H_1 : \theta > 0$ and (ii) $H_0' : \theta \leq -\Delta$ versus $H_1' : \theta \geq \Delta$ (with an indifference zone) for the mean θ of i.i.d. normal random variables with unit variance, assuming cost c per observation and a prior distribution G on θ. Letting $t = cn, w(t) = \sqrt{c} S_n, \mu = \theta/\sqrt{c}$ and $\gamma = \Delta/\sqrt{c}$, note that $w(t)$ is a Wiener process with drift coefficient μ and with τ restricted to the set $I_c = \{c, 2c, \ldots\}$. As $c \to 0$, I_c becomes dense in $[0, \infty)$. Moreover, for any prior distribution G that has a positive continuous density G', the density function of $\mu = \theta/\sqrt{c}$ is $\sqrt{c} G'(\sqrt{c} x) \sim \sqrt{c} G'(0)$. This leads to the problem of testing $H_0 : \mu < 0$ versus $H_1 : \mu > 0$ (or $H_0' : \mu \le -\gamma$ versus $H_1' : \mu \ge \gamma$) for the drift coefficient μ of a Wiener process $\{w(t), t \ge 0\}$, assuming the 0-1 loss, a flat prior on μ, and a cost of t for observing the process for a period of length t. The optimal stopping rule is of the form $\tau_\gamma = \inf\{t > 0 : |w(t)| \ge h_\gamma(t)\}$, in which h_γ has the following asymptotic behavior: For fixed $0 \le \gamma < \infty$, as $t \to 0$,

$$h_\gamma(t) = \Big\{2t[\log t^{-1} + \frac{1}{2}\log\log t^{-1} - \frac{1}{2}\log 4\pi + o(1)]\Big\}^{1/2}, \tag{2.2}$$

and $h_\gamma(t) \sim \frac{1}{4}\sqrt{2/\pi}\,t^{-1/2}\exp(-\frac{1}{2}\gamma^2 t)$ as $t \to \infty$. Moreover, (2.2) still holds as $\gamma \to \infty$ and $t \to 0$ such that $t = o((\gamma^2 \log\gamma^2)^{-1})$ (cf. Section 2 of Lai (1988a)).

The problem of Bayes sequential tests of $H_0 : \mu < 0$ versus $H_1 : \mu > 0$ for the drift coefficient μ of a Wiener process has also been studied by Lerche (1986a) under the 0-1 loss and a normal prior distribution G with mean 0 and variance σ^2, but with a cost proportional to $\mu^2 t$ (depending quadratically on μ) for observing the process for a period of length t. Although this cost structure is unconventional and appears somewhat artificial, it is closely related to the cost in (1.12) under which Lerche (1986b) established the asymptotic optimality of the sequential mixture likelihood ratio rule (1.14). Moreover, the problem of minimizing

$$\begin{aligned} r(T) = \int_{-\infty}^{\infty} c\mu^2 E(T|\mu)dG(\mu) + \int_{-\infty}^{0} P\{w(T) > 0|\mu\}dG(\mu) \\ + \int_{0}^{\infty} P\{w(T) < 0|\mu\}dG(\mu) \end{aligned} \tag{2.3}$$

turns out to have a simple exact solution. The optimal stopping rule is that of the "repeated significance test"

$$\tau_c^* = \inf\Big\{t \ge 0 : |w(t)| \ge \lambda_c\sqrt{t + \sigma^{-2}}\Big\}, \tag{2.4}$$

where λ_c is the solution of $\phi(\lambda) = 2c\lambda$. Note that the repeated significance test rejects H_0 iff $w(\tau_c^*) > 0$.

For the problem of testing the sign of a normal mean, instead of an artificial sampling cost that is proportional to the square of the unknown mean θ, we shall consider a given cost c per observation but allow the loss function for a wrong decision to behave like $|\theta|^p$ as $\theta \to 0$, with $-\infty < p < \infty$. We shall show that this incorporates the essence of Lerche's (1986a,b) theory on the optimality of repeated significance or mixture likelihood ratio tests and unifies it with the asymptotic theory of Chernoff (1961, 1965a,b) and its extension in Lai (1988b) for testing one-sided hypotheses in an exponential family.

2.1. A general class of loss functions and prior distributions

Let $X_1, X_2, \ldots$ be i.i.d. random variables whose common density function belongs to an exponential family $f_\theta(x) = e^{\theta x - \psi(\theta)}$ with respect to some σ-finite measure ν. To test the one-sided hypotheses $H_0 : \theta < \theta_0$ versus $H_1 : \theta > \theta_0$, we assume a sampling cost c for each observation and a loss $\ell(\theta) \geq 0$ for the wrong decision. Let A be an open interval containing θ_0 and having endpoints $(-\infty \leq) a_1 < a_2 (\leq \infty)$ such that for some $\zeta > 0, \inf_{a_1 - \zeta < \theta < a_2 + \zeta} \psi''(\theta) > 0, \sup_{a_1 - \zeta < \theta < a_2 + \zeta} \psi''(\theta) < \infty$ and ψ'' is uniformly continuous on $(a_1 - \zeta, a_2 + \zeta)$. Let G be a probability distribution on A having a continuous density function G' in some neighborhood of θ_0 such that for some $\eta > 0$ and $q > -1$,

$$G'(\theta) \sim \eta|\theta - \theta_0|^q \text{ as } \theta \to \theta_0. \tag{2.5}$$

Assume that for some $\xi > 0$ and $-\infty < p < \infty$ with $p + q > -1$,

$$\ell(\theta) \sim \xi|\theta - \theta_0|^p \text{ as } \theta \to \theta_0 \text{ and } \int_A \ell(\theta) dG(\theta) < \infty. \tag{2.6}$$

Since θ is known to lie in A, the maximum likelihood estimate $\widehat{\theta}_n$ at stage n is defined by

$$\psi'(\widehat{\theta}_n) = \max\{\psi'(a_1),\ \min[S_n/n, \psi'(a_2)]\}. \tag{2.7}$$

In the case $p > -1$, the assumptions (2.5) and (2.6) with $q = 0$ are the same as those in Theorem 1 of Lai (1988b). More generally, if $q < 1$, then since $p > -q - 1$, it follows that $p > -2$. Part (i) of the following theorem assumes $q < 1$ and generalizes Theorem 1 of Lai (1988b) (which only considers $p > -1$ and $q = 0$) to cover the case $p > -2$ in (2.6). It uses a result of Brezzi and Lai (1996) stated in Lemma 1 below. Part (ii) of the theorem deals with $q > 1$ and contains the special case $p = -2$ which is closely related to Lerche's (1986a,b) optimality theory, as will be explained in Section 3. Note that in the case $q > 1$, (2.5) implies that $\int_A (\theta - \theta_0)^{-2-\epsilon} dG(\theta) < \infty$ for every $0 \leq \epsilon < q - 1$, and therefore in particular that $\int_A (I(\theta, \theta_0))^{-1} dG(\theta) < \infty$.

Lemma 1. *Let $w(t), t \geq 0$, be a Wiener process with mean function μt and variance function t. Let $1 > q > -1, s > -1$ and $a > 0$. Then the optimal stopping rule that minimizes the Bayes risk*

$$\begin{aligned}\rho(\tau; q, s, a) = &\int_{-\infty}^{\infty} |\mu|^q E(\tau|\mu) d\mu + a \int_{-\infty}^{0} |\mu|^s P\{w(\tau) > 0|\mu\} d\mu \\ &+ a \int_0^{\infty} \mu^s P\{w(\tau) < 0|\mu\} d\mu \end{aligned} \tag{2.8}$$

is of the form $\tau^* = \inf\{t \geq 0 : |w(t)| \geq b_{q,s,a}(t)\}$, where $\lim_{t\to\infty} t^{-1/2} b_{q,s,a}(t) = 0$ and

$$b_{q,s,a}(t) = \{t[(s - q + 2)\log t^{-1} + (1 - s - q)\log\log t^{-1} + O(1)]\}^{1/2} \text{ as } t \to 0. \tag{2.9}$$

Theorem 1. *Under the assumptions* (2.6) *on the loss function and* (2.5) *on the prior distribution* G (*with* $q > -1$ *and* $p+q > -1$), *let* $r(T,\delta)$ *be the Bayes risk* (1.4) *of a sequential test* (T,δ), *with stopping rule* T *and terminal decision rule* δ, *of* $H_0 : \theta \in \Theta_0 = (-\infty, \theta_0) \cap A$ *versus* $H_1 : \theta \in \Theta_1 = (\theta_0, \infty) \cap A$. *Let* δ^* *be the terminal decision rule that accepts* H_0 *iff* $\widehat{\theta}_n < \theta_0$ *when stopping occurs at stage* n, *where* $\widehat{\theta}_n$ *is the maximum likelihood estimator defined by* (2.7). *Let* $I(\theta,\lambda)$ *be the Kullback-Leibler information number defined in* (1.7).

(i) *Suppose that* $q < 1$. *Then* $p > -2$. *Let* $a = \xi(\psi''(\theta_0))^{-p/2}$ *and* $B(t) = b^2_{q,p+q,a}(t)/2t$, *where* $b_{q,s,a}(t)$ *is given in Lemma* 1 *and* ξ, η *are given by* (2.5) *and* (2.6). *Then as* $c \to 0$,

$$\inf_{(T,\delta)} r(T,\delta) \sim \eta(\psi''(\theta_0))^{-(q+1)/2}\rho(\tau^*; q, p+q, a)c^{(p+q+1)/(p+2)} \sim r(T_c^*, \delta^*),$$

where $\rho(\tau; q, s, a)$ *and* τ^* *are given in Lemma* 1 *and*

$$T_c^* = \inf\{n \geq 1 : nI(\widehat{\theta}_n, \theta_0) \geq B(c^{2/(p+2)}n)\}. \tag{2.10}$$

(ii) *Suppose that* $q > 1$. *Then* $\int_{|\theta-\theta_0|<1}[(\log|\theta-\theta_0|^{-1})/(\theta-\theta_0)^2]dG(\theta) < \infty$, *and as* $c \to 0$,

$$\inf_{(T,\delta)} r(T,\delta) \sim c|\log c| \int_A (I(\theta,\theta_0))^{-1} dG(\theta) \sim \inf r(T_{c,h}, \delta^*),$$

where $T_{c,h} = \inf\{n \geq 1 : \int_{a_1-\zeta}^{a_2+\zeta} e^{(\theta-\theta_0)S_n - n(\psi(\theta)-\psi(\theta_0))} h(\theta)d\theta \geq c^{-1}\}$ *and* h *is a positive continuous function such that* $\int_{a_1-\zeta}^{a_2+\zeta} h(\theta)d\theta < \infty$.

2.2. Proof of Theorem 1(i) and Wiener process approximation for the case $q < 1$

For the case $1 > q(> -1)$, since $p+q > -1$, it follows from the asymptotic behavior (2.9) of the stopping boundary $b_{q,p+q,a}$ and Lemma 1 of Lai (1988c) that as $|\mu| \to \infty$,

$$E(\tau^*|\mu) \sim (p+2)\mu^{-2}(\log\mu^2),\ P\{w(\tau^*)\mathrm{sgn}(\mu) < 0|\mu\} = O(|\mu|^{-(p+2)}(\log\mu^2)^{q-1}). \tag{2.11}$$

Hence $\int_{-\infty}^{\infty} |\mu|^q E(\tau^*|\mu)d\mu < \infty$ and $\int_{-\infty}^{\infty} |\mu|^{p+q} P\{w(\tau^*)\mathrm{sgn}(\mu) < 0|\mu\}d\mu < \infty$, recalling that $q < 1$ and $p+q > -1$. Let $r = (p+2)^{-1}$. By Theorem 1 of Lai (1988c) and an argument similar to that of Lemma 5 of Lai (1988a),

$$E_\theta T_c^* = O((\theta-\theta_0)^{-2}\log\{c^{-2r}(\theta-\theta_0)^2\}) \text{ uniformly in } \theta \in A \text{ with } (\theta-\theta_0)^2 \geq 2c^{2r}, \tag{2.12}$$

$$P_\theta\{(\widehat{\theta}_{T_c}-\theta_0)(\theta-\theta_0)<0\}=O(\{c^{-2r}(\theta-\theta_0)^2\}^{-(p+2)/2}\{\log[c^{-2r}(\theta-\theta_0)^2]\}^{q-1}) \tag{2.13}$$

uniformly in $\theta \in A$ with $|\theta-\theta_0|\geq c^r$, in view of (2.9) and (2.10).

Define $w_c(t)=c^r(S_n-n\psi'(\theta_0))/\sqrt{\psi''(\theta_0)}$ at $t=c^{2r}n, n\in\{1,2,\ldots\}, w_c(0)=0$, and define $w_c(t)$ by linear interpolation at other positive values of t. Then for every $M\geq 2$ and $B>0$, the process $\{w_c(t), 0\leq t\leq B\}$ under $P_{\theta_0+\mu c^r/\sqrt{\psi''(\theta_0)}}$ converges weakly to the Wiener process $\{w(t), 0\leq t\leq B\}$ with drift coefficient μ, the convergence being uniform in $-M\leq\mu\leq M$ (cf. Lemma 4 of Lai (1988a)). Making use of this weak convergence and the estimates (2.11)–(2.13) together with (2.5) and (2.6), we can proceed in the same way as the proof of Theorem 3 of Lai (1988a) to prove part (i) of Theorem 1. In particular, as in Lai (1988a), p. 870, we have for $\widehat{\theta}_n$ near θ_0,

$$\begin{aligned} nI(\widehat{\theta}_n,\theta_0) &\sim n\psi''(\theta_0)(\widehat{\theta}_n-\theta_0)^2/2 \sim n(\psi''(\theta_0))^{-1}\{\psi'(\widehat{\theta}_n)-\psi'(\theta_0)\}^2/2 \\ &= (2c^{2r}n)^{-1}(c^r/\sqrt{\psi''(\theta_0)})^2(S_n-n\psi'(\theta_0))^2. \end{aligned}$$

If the X_i are i.i.d. $N(\theta,1)$ random variables and $\theta_0=0$, then the preceding transformation $(t,w_c(t),\mu)=(c^{2r}n, c^rS_n, \theta/c^r)$ produces a Wiener process $w_c(t)$ with drift coefficient μ and with t restricted to the set $\{c^{2r},2c^{2r},\ldots\}$ which becomes dense in $[0,\infty)$ as $c\to 0$. Thus the discrete-time Bayes sequential testing problem of minimizing the Bayes risk $r(T,\delta)$ in Theorem 1 can be approximated by the continuous-time optimal stopping problem in Lemma 1, as was first noted by Chernoff (1961) in the case $q=0$ and $p=1$. For the general exponential family, a key observation in the preceding argument is that the stopping rule in the normal case can be re-expressed in terms of the generalized likelihood ratio statistics $S_n^2/2n$, which generalize to $nI(\widehat{\theta}_n,\theta_0)$ in an exponential family, with $I(\theta,\lambda)$ the Kullback-Leibler information number (1.7). Although the $w_c(t)$ is now only approximately a Wiener process with drift coefficient $\mu=(\theta-\theta_0)\sqrt{\psi''(\theta_0)}/c^r$ under P_θ for $-M\leq\mu\leq M$, approximation of the discrete-time Bayes sequential testing problem in Theorem 1(i) by the Wiener process optimal stopping problem in Lemma 1 is still valid because of (2.12) and (2.13), which imply that the contribution of the integral of the risk function over the region $\{\theta\in A: |\theta-\theta_0|\geq Mc^r\}$ to the Bayes risk of (T_c^*,δ^*) becomes negligible as $M\to\infty$. In the special case $q=0$ and $p=0$, this was noted by Lai (1988a) who also showed that similar Wiener process approximations are still valid under the 0-1 loss even when there is an indifference zone which separates the null and alternative hypotheses, thus providing a refinement of Schwarz's (1962) asymptotic theory so that it can yield the preceding asymptotic solution to the testing problem without an indifference zone by letting the size of the indifference zone in Schwarz's problem shrink to 0.

Specifically, to test $H_0:\theta\leq\theta_0$ versus $H_1:\theta\geq\theta_1$ under the 0-1 loss, let $\gamma=\frac{1}{2}(\theta_1-\theta_0)\sqrt{\psi''(\theta_0)/c}$ and let h_γ be the optimal stopping boundary (2.2) for the corresponding problem of testing $H_0' : \mu \le -\gamma$ versus $H_1' : \mu \ge \gamma$ for the drift coefficient μ of a Wiener process. Let $B_\gamma = \{h_\gamma(t) + \gamma t\}^2/2t$ and let

$$N_{\gamma,c} = \inf\{n \ge 1 : \max[nI(\widehat{\theta}_n, \theta_0), nI(\widehat{\theta}_n, \theta_1)] \ge B_\gamma(cn)\}. \tag{2.14}$$

Let $\widetilde{\delta}$ be the same terminal decision rule as in (1.8). Making use of (2.2) to obtain analogues of (2.12) and (2.13) for the rule (2.14), Lai (1988a) showed that not only is the test $(N_{\gamma,c}, \widetilde{\delta})$ asymptotically Bayes risk efficient in the sense of (1.8) with N_c replaced by $N_{\gamma,c}$ when the indifference zone parameters θ_0 and θ_1 are fixed, but that it is also asymptotically Bayes risk efficient as $c \to 0$ and $\theta_1 - \theta_0 \to 0$ such that $(\theta_1 - \theta_0)^2/c$ converges to a positive constant or to ∞. Moreover, letting $\theta_1 = \theta_0$ in (2.14) yields the stopping rule (2.10) with $q = p = 0, B = B_\gamma$ and $\gamma = 0$.

In Lai (1988a), simple closed-form approximations to the stopping boundaries $B_\gamma(\gamma \ge 0)$ are given for easy implementation of these tests, and simulation studies of their risk functions show that these stopping rules not only provide approximate Bayes solutions with respect to a large class of priors but also have nearly optimal frequentist properties. In the case $A = \Theta$, note that Schwarz's stopping rule (1.6) for testing $H_0 : \theta \le \theta_0$ versus $H_1 : \theta \ge \theta_1$ can be expressed in terms of the Kullback-Leibler information number $I(\theta, \lambda)$ as $N_c = \inf\{n \ge 1 : \max[nI(\widehat{\theta}_n, \theta_0), nI(\widehat{\theta}_n, \theta_1)] \ge \log a_c\}$. Thus, the stopping rule (2.14) simply modifies Schwarz's rule by replacing the constant boundary $\log a_c \sim \log c^{-1}$ by a time-varying boundary $B_\gamma(cn)$ that is of the order of $\log(cn)^{-1}$. Alternatively we can regard (2.14) as an adaptive 2-SPRT in which the λ in the 2-SPRT stopping rule (1.3) is replaced by $\widehat{\theta}_n$ and the constant threshold $B = B'$ is replaced by $B_\gamma(cn)$ that incorporates the time-varying uncertainties in the parameter estimates $\widehat{\theta}_n$.

3. The Case $q > 1$ and Asymptotic Optimality of Power-one Tests

In this section we prove Theorem 1(ii) and show that the special case $q = 2 = -p$ is closely related to Lerche's (1986a,b) optimal stopping problems that involve a cost $c\theta^{-2}$ per observation. In this connection we also study certain properties of the power-one tests developed by Robbins (1970) and Robbins and Siegmund (1970, 1973).

Let $q > 1$. To begin with, consider the special case of Theorem 1(ii) in which $\theta_0 = 0, A = (a_1, a_2)$ is an open interval containing $0, G'(\theta) = \eta|\theta|^q \exp(-\theta^2/2\sigma^2)$ and $\ell(\theta) = \theta^{-2}$ for $a_1 < \theta < a_2$. Then $K := \int_{a_1}^{a_2} |\theta|^{q-2} e^{-\theta^2/2\sigma^2} d\theta < \infty$ and therefore $\pi(\theta) := K^{-1}|\theta|^{q-2}e^{-\theta^2/2\sigma^2}$ is a density function on (a_1, a_2). Moreover, the Bayes risk (1.4) of (T, δ) under the prior distribution G, cost c per observation and loss $\ell(\theta)$ for the wrong decision can be expressed as

$$\begin{aligned} r(T,\delta) &= \eta \int_{a_1}^{a_2} \{c|\theta|^q E_\theta T + |\theta|^{q-2} P_\theta(\delta \text{ is wrong})\} e^{-\theta^2/2\sigma^2} d\theta \\ &= \eta K \int_{a_1}^{a_2} \pi(\theta)\{c\theta^2 E_\theta T + P_\theta(\delta \text{ is wrong})\} d\theta, \end{aligned} \tag{3.1}$$

which is ηK times the Bayes risk of (T,δ) under the 0-1 loss, cost $c\theta^2$ per observation and prior density π. Hence the 0-1 loss and cost function considered by Lerche (1986a) can be tansformed to a special case of Theorem 1(ii).

In particular, for $q = 2$ and $A = (-\infty,\infty)$, π is the density function of a normal distribution with mean 0 and variance σ^2. Lerche (1986a) used this choice of π but considered a continuous-time Wiener process $w(t)$ with drift coefficient μ instead of i.i.d. $N(\theta,1)$ random variables X_i in (3.1). He showed that in this case the optimal stopping rule τ_c^* for testing the sign of μ has the simple form (2.4) and that its Bayes risk (2.3) is given by $r(\tau_c^*) = \Phi(-\lambda_c) + c\lambda_c^2$ with $\phi(\lambda_c) = 2c\lambda_c$. Here and in the sequel we let Φ and ϕ denote the standard normal distribution and density functions. Since $\lambda_c^2 \sim 2\log c^{-1}$ and $\Phi(-\lambda_c) \sim \phi(\lambda_c)/\lambda_c = 2c$ as $c \to 0$, it follows that $r(\tau_c^*) \sim 2c\log c^{-1}$ as $c \to 0$. This asymptotically optimal Bayes risk can also be attained by the mixture likelihood ratio stopping rule $\tau(c)$ defined by (1.14), in which we take $g = \phi$ and $a_c = (2c)^{-1}$ for definiteness so that (1.14) reduces to

$$\tau(c) = \inf\{t \geq 0 : |w(t)| \geq (t+1)^{1/2}[\log(t+1) + 2\log(2c)^{-1}]^{1/2}\} \tag{3.2}$$

(cf. Robbins (1970)). In fact, it will be shown later that $\int_{-\infty}^{\infty} \mu^2 E(\tau(c)|\mu) d\Phi(\mu) \sim 2\log c^{-1}$ as $c \to 0$. Moreover, letting $w^*(t)$ denote driftless Brownian motion, we have for $\mu > 0$,

$$\begin{aligned} &P\{w(\tau(c)) < 0|\mu\} \\ &\leq P\{w^*(t) \leq -(t+1)^{1/2}[\log(t+1) + 2\log(2c)^{-1}]^{1/2} \text{ for some } t \geq 0\} \leq 2c \end{aligned}$$

(cf. Robbins (1970), Eq. (18)). Hence $r(\tau(c)) = 2c|\log c|(1 + o(1)) + O(c) \sim 2c\log c^{-1}$ as $c \to 0$. We next use similar arguments to prove part (ii) of Theorem 1.

Proof of Theorem 1(ii). Let $\kappa > \int_A (I(\theta,\theta_0))^{-1} dG(\theta)$ and let $\mathcal{F}_c$ be the class of all sequential tests (T,δ) such that $r(T,\delta) \leq \kappa c|\log c|$. We first show that

$$\inf_{(T,\delta)\in\mathcal{F}_c} r(T,\delta) \geq c|\log c| \Big\{ \int_A (I(\theta_0,\theta))^{-1} dG(\theta) + o(1) \Big\} \ \text{ as } \ c \to 0. \tag{3.3}$$

For $(T,\delta) \in \mathcal{F}_c$, since $\int_{|\theta-\theta_0|\leq|\log c|^{-1}} \ell(\theta) P_\theta\{\delta \text{ is wrong}\} dG(\theta) \leq \kappa c|\log c|$, it follows from (2.5) and (2.6) that for all sufficiently small c there exist θ_c' and θ_c'' belonging to A such that $0 < \theta_c' - \theta_0 < |\log c|^{-1}, 0 < \theta_0 - \theta_c'' < |\log c|^{-1}$ and

$$\max(P_{\theta_c'}\{\delta \text{ accepts } H_0\}, P_{\theta_c''}\{\delta \text{ accepts } H_1\}) \leq 2\kappa(p+q+1)c|\log c|^{p+q+1}/(\xi\eta),$$

recalling that $p+q>-1$. Hence by Hoeffding's (1960) lower bound for the expected sample size of a sequential test (cf. Lemma 11 of Lai (1988a)), as $c \to 0$,

$$\inf_{(T,\delta)\in\mathcal{F}_c} E_\theta T \geq (1+o(1))|\log c|/\max\{I(\theta,\theta_c'),I(\theta,\theta_c'')\}$$

uniformly in $\theta \in A$. Since $I(\theta,\theta_c') \to I(\theta,\theta_0)$ and $I(\theta,\theta_c'') \to I(\theta,\theta_0)$, it then follows that

$$\inf_{(T,\delta)\in\mathcal{F}_c} r(T,\delta) \geq \inf_{(T,\delta)\in\mathcal{F}_c} c\int_A E_\theta T dG(\theta) \geq (1+o(1))c|\log c|\int_A (I(\theta,\theta_0))^{-1}dG(\theta).$$

Hence $\inf_{(T,\delta)} r(T,\delta) \geq (1+o(1))c|\log c|\int_A (I(\theta,\theta_0))^{-1}dG(\theta)$, noting that $r(T,\delta) > \kappa c|\log c|$ if $(T,\delta)\notin\mathcal{F}_c$.

We next show that this asymptotic lower bound is attained by the test $(T_{c,h},\delta^*)$. For $\theta>\theta_0$,

$$\begin{aligned} P_\theta\{\widehat{\theta}_{T_{c,h}} \leq \theta_0\} &= \int_{\{\widehat{\theta}_{T_{c,h}}\leq\theta_0\}} \exp\{(\theta-\theta_0)S_{T_{c,h}} - T_{c,h}(\psi(\theta)-\psi(\theta_0))\}dP_{\theta_0} \\ &\leq P_{\theta_0}\{\widehat{\theta}_{T_{c,h}} \leq \theta_0\}, \end{aligned}$$

since on $\{\widehat{\theta}_{T_{c,h}}\leq\theta_0\}$, $S_{T_{c,h}}/T_{c,h} \leq \psi'(\widehat{\theta}_{T_{c,h}}) \leq \psi'(\theta_0) \leq (\psi(\theta)-\psi(\theta_0))/(\theta-\theta_0)$ by (2.7) and the convexity of ψ. Moreover, since $\{\int_{a_1-\zeta}^{a_2+\zeta} e^{(\theta-\theta_0)S_n - n(\psi(\theta)-\psi(\theta_0))}h(\theta)d\theta$, $n\geq 1\}$ is a martingale under P_{θ_0}, $P_{\theta_0}\{\widehat{\theta}_{T_{c,h}}\leq\theta_0\} \leq P_{\theta_0}\{T_{c,h}<\infty\} \leq c\int_{a_1-\zeta}^{a_2+\zeta} h(t)dt$. A similar argument for $P_\theta\{\widehat{\theta}_{T_{c,h}}\geq\theta_0\}$ in the case $\theta<\theta_0$ then yields $P_\theta\{\delta^*\text{is wrong}\} \leq c\int_{a_1-\zeta}^{a_2+\zeta} h(t)dt$ for all $\theta\neq\theta_0$. From (2.5) (with $q>1$) and Lemma 2 below it follows that $\int_A E_\theta T_{c,h}dG(\theta) \sim |\log c|\int_A (I(\theta,\theta_0))^{-1}dG(\theta)$ as $c\to 0$. Since $\int_A \ell(\theta)dG(\theta)<\infty$, it then follows that $r(T_{c,h},\delta^*) = c|\log c|\{\int_A (I(\theta,\theta_0)^{-1}dG(\theta) + o(1)\} + O(c)$ as $c\to 0$, completing the proof of Theorem 1(ii).

Lemma 2. *With the same notation as in Theorem* 1(ii), *for any* $\epsilon>0$, $E_\theta T_{c,h} \sim |\log c|/I(\theta,\theta_0)$ *as* $c\to 0$, *uniformly in* $\theta\in A$ *with* $|\theta-\theta_0|\geq\epsilon$. *Moreover, as* $c\to 0$ *and* $\theta\to\theta_0$,

$$E_\theta T_{c,h} \sim 2(\log c^{-1} + \log|\theta-\theta_0|^{-1})/\{(\theta-\theta_0)^2\psi''(\theta_0)\}. \tag{3.4}$$

For the stopping rule (3.2) of the continuous-time Wiener process $w(t)$ with drift coefficient μ, a straightforward modification of the proof of Lemma 2 in the Appendix shows that analogous to Lemma 2, we again have $E(\tau(c)|\mu) \sim 2|\log c|/\mu^2$ uniformly in $|\mu|\geq\epsilon$ and $E(\tau(c)|\mu) \sim 2\mu^{-2}(\log c^{-1} + \frac{1}{2}\log\mu^{-2})$ as $c\to 0$ and $\mu\to 0$. This result is an extension of Theorem 2 of Lai (1977) and Theorem 4 of Jennen and Lerche (1982) where c is assumed to be fixed as $\mu\to 0$.

Example. Let $X_1, X_2, \ldots$ be i.i.d. $N(\theta, 1)$ random variables and consider the problem of testing $H_0 : \theta < 0$ versus $H_1 : \theta > 0$. Here $\theta_0 = 0, A = (-\infty, \infty)$, $\psi(\theta) = \theta^2/2$ and $nI(\widehat{\theta}_n, \theta_0) = S_n^2/2n$. Taking $h = \phi$ in the stopping rule $T_{c,h}$ of Theorem 1(ii), we can express it as

$$\begin{aligned} T_{c,\phi} &= \inf\{n \geq 1 : (n+1)^{-1/2} \exp(S_n^2/2(n+1)) \geq c^{-1}\} \\ &= \inf\{n \geq 1 : nI(\widehat{\theta}_n, \theta_0) \geq (1 + n^{-1})[\log c^{-1} + \frac{1}{2}\log(n+1)]\}, \end{aligned} \tag{3.5}$$

which has the property $P_{\theta_0}\{T_{c,\phi} < \infty\} \leq c$ so that $P_\theta\{\delta^* \text{ is wrong}\} \leq c$ for all $\theta \neq \theta_0$, as shown in the proof of Theorem 1(ii). Note that the time-varying boundary for $nI(\widehat{\theta}_n, \theta_0)$ in the stopping rule (3.5) tends to ∞ as $n \to \infty$, with the consequence that $E_\theta T_{c,\phi} \sim 2\theta^{-2}(\log c^{-1} + \log|\theta|^{-1})$ as $c \to 0$ and $\theta \to 0$ by (3.4). This order of magnitude for $E_\theta T_{c,\phi}$ when θ is near 0 does not cause difficulty if $q > 1$ in (2.5), under which $\int_{|\theta|<1} \theta^{-2}|\log|\theta||dG(\theta) < \infty$. On the other hand, if $q < 1$ in (2.5), then $\int_{-\infty}^{\infty} \theta^{-2} dG(\theta) = \infty$ and we therefore need a much smaller order of magnitude for $E_\theta T$ than $E_\theta T_{c,\phi}$ when θ is near 0. Hence the boundary $B(c^{2/(p+2)}n)$ of the stopping rule $T_c^* = \inf\{n \geq 1 : nI(\widehat{\theta}_n, \theta_0) \geq B(c^{2/(p+2)}n)\}$ in Theorem 1(i) is markedly different from the boundary $(1+n^{-1})[\log c^{-1} + \frac{1}{2}\log(n+1)]$ in (3.5). First note how c appears in both boundaries. It appears as $\log c^{-1}$ in (3.5) but as a factor of the time n in T_c^*. Moreover, while the boundary in (3.5) tends to ∞ with n, the boundary $B(t) = b_{q,p+q,a}^2(t)/2t$ tends to 0 as $t \to \infty$. By (2.9), $B(c^{2/(p+2)}n) \sim \frac{1}{2}(p+2)|\log(c^{2/(p+2)}n)|$ as $c^{2/(p+2)}n \to 0$. In particular, if $\log n = o(|\log c|)$, then $B(c^{2/(p+2)}n) \sim \log c^{-1} \sim \log c^{-1} + \frac{1}{2}\log(n+1)$. Note from the proof of Theorem 1(i) that unlike (3.4), the asymptotic behavior of $E_\theta T_c^*$ is given by (2.12) (with $r = 1/(p+2)$) and by the weak convergence approximation

$$c^{2/(p+2)} E_{\mu c^{1/(p+2)}}(T_c^*) \to E(\tau^*|\mu) \quad \text{as} \quad c \to 0,$$

uniformly for μ in compact sets, where τ^* is defined in Lemma 1.

Instead of the one-sided hypotheses $H_0 : \theta < \theta_0$ versus $H_1 : \theta > \theta_0$, the preceding proof of Theorem 1(ii) can be modified to show that the mixture likelihood stopping rule $T_{c,h}$ is also asymptotically optimal for testing the simple hypothesis $H : \theta = \theta_0$ versus the composite two-sided alternative hypothesis $K : \theta \neq \theta_0$. The terminal decision rule δ' of the test is to reject H upon stopping. The test has type I error probability $\alpha := P_{\theta_0}(T_{c,h} < \infty) \leq c\int_{a_1-\zeta}^{a_2+\zeta} h(t)dt$. Moreover, $\log \alpha \sim \log c$ as $c \to 0$ (cf. Pollak (1986)). On the other hand, $E_\theta T_{c,h} < \infty$ and therefore $P_\theta(T_{c,h} < \infty) = 1$ for $\theta \neq \theta_0$. Hence the test has power one. Subject to the type I error constraint $P_{\theta_0}(T < \infty) \leq \alpha, E_\theta T$ is minimized by a one-sided SPRT whose stopping rule is of the form (1.9) with $f_1 = f_\theta$ and $f_0 = f_{\theta_0}$, with minimal expected sample of the order $E_\theta \tau \sim |\log c|/I(\theta, \theta_0)$ as

$|\log c|(\sim |\log \alpha|) \to \infty$. Lemma 2 shows that the mixture likelihood rule $T_{c,h}$ attains this minimal order of magnitude for the expected sample size under every fixed $\theta \neq \theta_0$ as $c \to 0$. Moreover, the following theorem shows that under the 0-1 loss and cost c per observation if $\theta \neq \theta_0$, the test $(T_{c,h}, \delta')$ is asymptotically Bayes with respect to prior distributions of the form $\omega G + (1-\omega)G_0$, where $0 < \omega < 1$, G_0 puts all its probability mass at θ_0 and G is a probability distribution satisfying condition (2.5) with $q > 1$. Note that there is no sampling cost under H_0, and therefore if H_0 should be true one would ideally continue to collect data forever. To illustrate this point, suppose a drug is licensed for use based on a clinical trial indicating a positive treatment effect of the drug, but some concern exists that it may have deleterious side effects which will only become apparent after much more extensive use. Monitoring of the side effects may continue indefinitely until there is enough evidence for such side effects. Hence, under the null hypothesis of no deleterious side effects for the licensed drug, no sampling cost is incurred in administering the drug to patients. The sampling cost is only relevant when the alternative hypothesis is true and can be interpreted as an ethical cost of administering to each patient a drug whose deleterious side effects have not been found. This is the motivation behind the open-ended, power-one tests of Robbins (1970) and Robbins and Siegmund (1970, 1973).

Theorem 2. *Let* $\rho(T,\delta) = \omega \int_A \{cE_\theta T + P_\theta(\delta \text{ accepts } H)\} dG(\theta) + (1-\omega)P_{\theta_0}(\delta$ *rejects* $H)$ *be the Bayes risk of a sequential test* (T,δ) *of* $H : \theta = \theta_0$ *for the natural parameter* θ *of an exponential family under the* 0-1 *loss, where* $0 < \omega < 1$ *and* G *is a probability distribution on* A *satisfying* (2.5) *for some* $q > 1$ *and* $\eta > 0$ *(so* $G(\{\theta_0\}) = 0$*). Define the stopping rule* $T_{c,h}$ *as in Theorem* 1(ii) *and let* δ' *be the terminal decision rule that rejects* H *upon stopping. Then as* $c \to 0$,

$$\inf_{(T,\delta)} \rho(T,\delta) \sim \omega c |\log c| \int_A (I(\theta,\theta_0))^{-1} dG(\theta) \sim \rho(T_{c,h}, \delta').$$

Proof. Straightforward modification of the proof of Theorem 1(ii) shows that $\inf_{(T,\delta)} \rho(T,\delta) \geq \{\omega \int_A (I(\theta,\theta_0))^{-1} dG(\theta) + o(1)\} c|\log c|$ as $c \to 0$. For the test $(T_{c,h}, \delta')$, $P_{\theta_0}\{\delta'$ rejects $H\} \leq c \int_{a_1-\zeta}^{a_2+\zeta} h(t)dt$ and $P_\theta\{\delta'$ accepts $H\} = 0$ for $\theta \neq 0$. Hence from (2.5) (with $q > 1$) and Lemma 2, $\rho(T_{c,h}, \delta') \sim \omega c |\log c| \int_A (I(\theta,\theta_0))^{-1}$ $dG(\theta)$ as $c \to 0$.

Acknowledgement

This research was supported by the National Science Foundation under grant DMS-9403794.

Appendix: Proof of Lemma 2

For fixed $\theta \neq \theta_0$, Pollak and Siegmund (1975) have proved that $\infty > E_\theta T_{c,h} \sim |\log c| / I(\theta_0, \theta)$. Simple refinements of their proof of this result show that $E_\theta T_{c,h} \sim$

$|\log c|/I(\theta_0, \theta)$ uniformly in $\theta \in A$ with $|\theta - \theta_0| \geq \epsilon$. To prove (3.4), we can use arguments similar to those in Section 3 and Lemma 6 of Lai (1988c). The basic idea is as follows. Let $\bar{X}_n = S_n/n$ and $L = \sup\{n : |\bar{X}_n - \psi'(\theta)| \geq |\theta - \theta_0|\}$. Then there exists $M > 0$ such that $E_\theta L \leq M(\theta - \theta_0)^{-2}$ for all $\theta \in A$ with $0 < |\theta - \theta_0| \leq \epsilon$ (sufficiently small) (cf. (3.14) of Lai (1988c)). For $n > \max\{L, (\theta - \theta_0)^{-2}(\log|\theta - \theta_0|^{-1})^{1/2}\}$, application of Laplace's method for asymptotic evaluation of integrals as in Lemma 6 of Lai (1988c) yields

$$\begin{aligned}&\int_{a_1-\zeta}^{a_2+\zeta} e^{(t-\theta_0)S_n - n(\psi(t)-\psi(\theta_0))} h(t)dt \\ &\sim (2\pi/n\psi''(\theta_0))^{1/2} h(\theta_0) \exp(nI(\widehat{\theta}_n, \theta_0)) \\ &= \exp\{nI(\widehat{\theta}_n, \theta_0) - (\log n)/2 + O(1)\} \quad \text{as} \quad \theta \to \theta_0.\end{aligned}$$

Noting that $nI(\theta, \theta_0) - \frac{1}{2}\log n \sim \log c^{-1} \Leftrightarrow \frac{1}{2}\psi''(\theta_0)(\theta-\theta_0)^2 n \sim \log c^{-1} + \frac{1}{2}\log|\theta - \theta_0|^{-2}$ as $c \to 0$ and $\theta \to \theta_0$, we can proceed as in the proof of Theorem 3 of Lai (1988c) to complete the proof.

References

Arrow, K. J., Blackwell, D. and Girshick, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica* **17**, 213-244.

Bather, J. A. (1962). Bayes procedures for deciding the sign of a normal mean. *Proc. Cambridge Philos. Soc.* **58**, 599-620.

Bickel, P. J. (1973). On the asymptotic shape of Bayesian sequential tests of $\theta \leq 0$ versus $\theta > 0$ for exponential families. *Ann. Math. Statist.* **1**, 231-240.

Bickel P. J. and Yahav, Y. A. (1972). On the Wiener process approximation to Bayesian sequential testing problems. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1**, 57-84. Univ. California Press.

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

Brezzi, M. and Lai. T. L. (1996). Optimal stopping for Brownian motion in bandit problems and sequential analysis. Technical Report, Department of Statistics, Stanford University.

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. (1972). *Sequential Analysis and Optimal Design.* Soc. Industr. Appl. Math., Philadelphia.

Chow, Y. S., Robbins, H. and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Houghton Mifflin, Boston.

Ferguson, T. (1967). *Mathematical Statistics — A Decision Theoretic Approach.* Academic Press, New York.

Fortus, R. (1979). Approximations to Bayesian sequential tests of composite hypotheses. *Ann. Statist.* **7**, 579-591.

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

Jennen, C. and Lerche, H. R. (1982). Asymptotic densities of stopping times associated with tests of power one. *Z. Wahrsch. Verw. Gebiete* **61**, 501-511.

Kiefer, J. and Sacks, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34**, 705-750.

Kiefer, J. and Weiss, L. (1957). Some properties of generalized sequential probability ratio tests. *Ann. Math. Statist.* **28**, 57-74.

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). Power-one tests based on sample sums. *Ann. Statist.* **5**, 866-880.

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

Lai, T. L. (1988b). On Bayes sequential tests. In *Statistical Decision Theory and Related Topics IV* **2** (Edited by S. S. Gupta and J. O. Berger), 131-143. Springer-Verlag, New York.

Lai, T. L. (1988c). Boundary crossing problems for sample means. *Ann. Probab.* **16**, 375-396.

Lai, T. L. and Zhang, L. (1994a). A modification of Schwarz's sequential likelihood ratio tests in multivariate sequential analysis. *Sequential Anal.* **13**, 79-96.

Lai, T. L. and Zhang, L. (1994b). Nearly optimal generalized sequential likelihood ratio tests in multivariate exponential families. In *Multivariate Analysis and Its Applications* (Edited by T. W. Anderson, K. T. Fang and I. Olkin), 331-346. Inst. Math. Statist., Hayward.

Lerche, H. R. (1986a). An optimal property of the repeated significance test. *Proc. Natl. Acad. Sci. USA* **83**, 1546-1548.

Lerche, H. R. (1986b). The shape of Bayes tests of power one. *Ann. Statist.* **14**, 1030-1048.

Lindley, D. V. and Barnett, B. N. (1965). Sequential sampling: two decision problems with linear losses for binomial and normal random variables. *Biometrika* **52**, 507-532.

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.

Moriguti, S. and Robbins, H. (1962). A Bayes test of $p \leq 1/2$ versus $p > 1/2$. *Rep. Statist. Appl. Res., Un. Japan. Sci. Engrs.* **9**, 39-60.

Neyman, J. (1971). Foundations of behavioristic statistics. In *Foundations of Statistical Inference* (Edited by V. P. Godambe and D. A. Sprott), 1-19. Holt, Reinhart & Winston, Toronto.

Pollak, M. (1978). Optimality and almost optimality of mixture stopping rules. *Ann. Statist.* **6**, 910-916.

Pollak, M. (1986). On the asymptotic formula for the probability of a type I error of mixture type power one tests. *Ann. Statist.* **14**, 1012-1029.

Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3**, 1267-1282.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41**, 1397-1409.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and partial sums. *Ann. Math. Statist.* **41**, 1410-1429.

Robbins, H. and Siegmund, D. (1973). Statistical tests of power one and the integral representation of solutions of certain partial differential equations. *Bull. Inst. Math., Academia Sinica* **1**, 93-120.

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

Schwarz, G. (1968). Asymptotic shapes for sequential testing of truncation parameters. *Ann. Math. Statist.* **39**, 2038-2043.

Shiryayev, A. N. (1978). *Optimal Stopping Rules.* Springer-Verlag, New York.

Simons, G. (1976). An improved statement of optimality for sequential probability ratio tests. *Ann. Statist.* **4**, 1240-1243.

Simons, G. and Wu, X. (1986). On Bayes tests for $p \leq 1/2$ versus $p > 1/2$: analytic approximations. In *Adaptive Statistical Procedures and Related Topics* (Edited by J. Van Ryzin), 79-95. Inst. Math. Statist., Hayward.

Sobel, M. (1953). An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Statist.* **24**, 319-337.

Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19**, 326-339.

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

Woodroofe, M. (1980). On the Bayes risk incurred by using asymptotic shapes. *Comm. Statist. Ser.A* **9**, 1727-1748.

Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received May 1995; accepted January 1996)